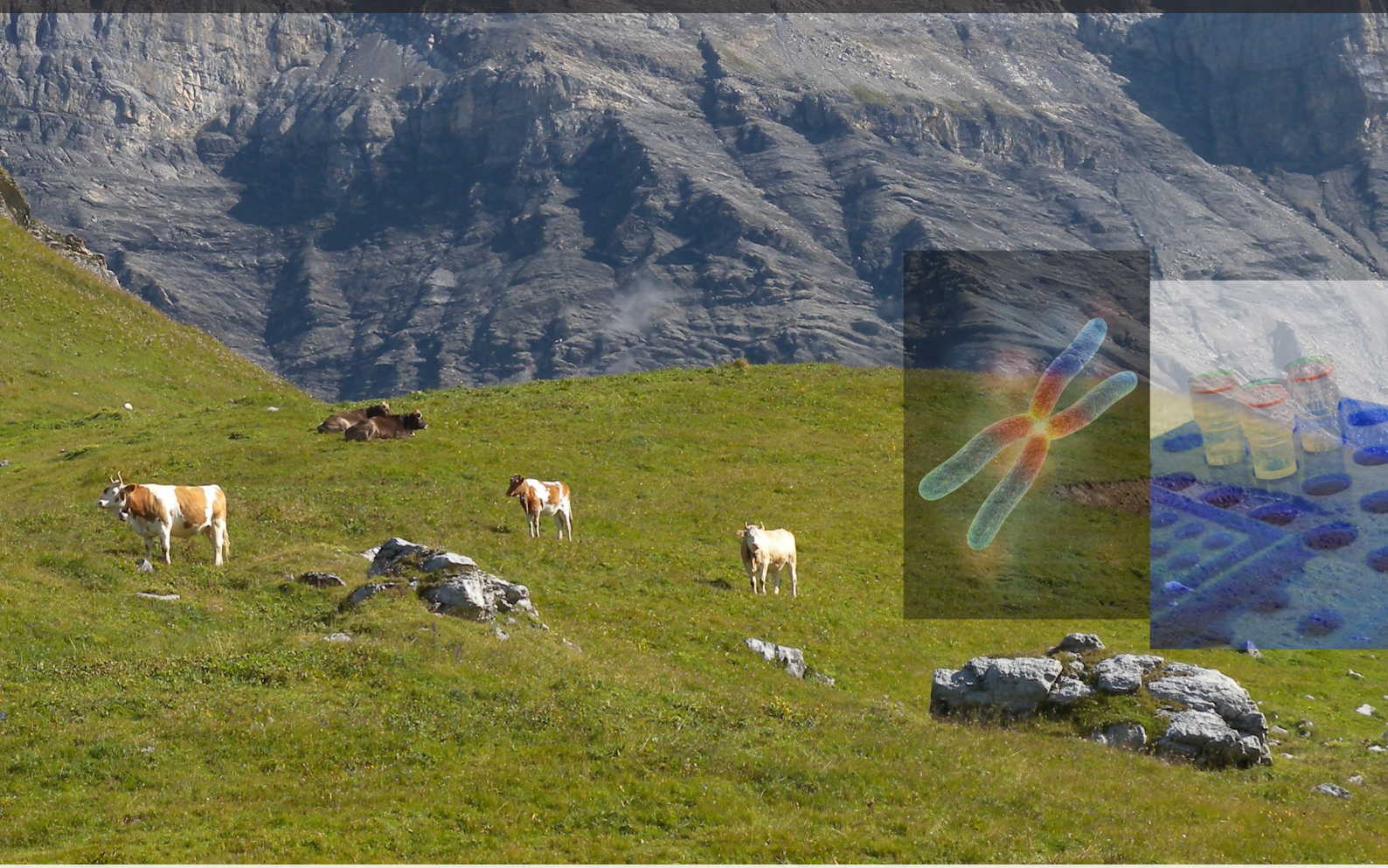
ADVANCES IN FARM ANIMAL GENOMIC RESOURCES

EDITED BY: Stéphane Joost, Michael W. Bruford, Ino Curik, Juha Kantanen, Johannes A. Lenstra, Johann Sölkner, Göran Andersson, Philippe V. Baret, Nadine Buys, Jutta Roosen, Michèle Tixier-Boichard and Paolo Ajmone Marsan

PUBLISHED IN: Frontiers in Genetics



frontiers Research Topics

frontiers

Frontiers Copyright Statement

© Copyright 2007-2016 Frontiers Media SA. All rights reserved.

All content included on this site, such as text, graphics, logos, button icons, images, video/audio clips, downloads, data compilations and software, is the property of or is licensed to Frontiers Media SA ("Frontiers") or its licensees and/or subcontractors. The copyright in the text of individual articles is the property of their respective authors, subject to a license granted to Frontiers.

The compilation of articles constituting this e-book, wherever published, as well as the compilation of all other content on this site, is the exclusive property of Frontiers. For the conditions for downloading and copying of e-books from Frontiers' website, please see the Terms for Website Use. If purchasing Frontiers e-books from other websites or sources, the conditions of the website concerned apply.

Images and graphics not forming part of user-contributed materials may not be downloaded or copied without permission.

Individual articles may be downloaded and reproduced in accordance with the principles of the CC-BY licence subject to any copyright or other notices. They may not be re-sold as an e-book.

As author or other contributor you grant a CC-BY licence to others to reproduce your articles, including any graphics and third-party materials supplied by you, in accordance with the Conditions for Website Use and subject to any copyright notices which you include in connection with your articles and materials.

All copyright, and all rights therein, are protected by national and international copyright laws.

The above represents a summary only. For the full conditions see the Conditions for Authors and the Conditions for Website Use.

ISSN 1664-8714
ISBN 978-2-88919-735-4
DOI 10.3389/978-2-88919-735-4

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view.
By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: **researchtopics@frontiersin.org**

ADVANCES IN FARM ANIMAL GENOMIC RESOURCES

Topic Editors:
Stéphane Joost, École Polytechnique Fédérale de Lausanne, Switzerland
Michael W. Bruford, Cardiff University, UK
Ino Curik, University of Zagreb, Croatia
Juha Kantanen, Natural Resources Institute Finland, Finland
Johannes A. Lenstra, Utrecht University, Netherlands
Johann Sölkner, University of Natural Resources and Life Sciences Vienna, Austria
Göran Andersson, Swedish University of Agricultural Sciences, Sweden
Philippe V. Baret, Université catholique de Louvain, Belgium
Nadine Buys, KU Leuven, Belgium
Jutta Roosen, Technische Universität München, Germany
Michèle Tixier-Boichard, INRA, France
Paolo Ajmone Marsan, Università Cattolica del S. Cuore, Italy

Images 1-3 taken from: http://www.photolibre.fr (retrieved in 2010), image 4 by Stéphane Joost.
Cover image by Stéphane Joost, EPFL and http://www.photolibre.fr (retrieved in 2010).

The history of livestock started with the domestication of their wild ancestors: a restricted number of species allowed to be tamed and entered a symbiotic relationship with humans. In exchange for food, shelter and protection, they provided us with meat, eggs, hides, wool and draught power, thus contributing considerably to our economic and cultural development. Depending on the species, domestication took place in different areas and periods. After domestication, livestock spread over all inhabited regions of the earth, accompanying human migrations and becoming also trade objects. This required an adaptation to different climates and varying styles of husbandry and resulted in an enormous phenotypic diversity.

Approximately 200 years ago, the situation started to change with the rise of the concept of breed. Animals were selected for the same visible characteristics, and crossing with different phenotypes was reduced. This resulted in the formation of different breeds, mostly genetically isolated from other populations. A few decades ago, selection pressure was increased again with intensive production focusing on a limited range of types and a subsequent loss of genetic

diversity. For short-term economic reasons, farmers have abandoned traditional breeds. As a consequence, during the 20th century, at least 28% of farm animal breeds became extinct, rare or endangered. The situation is alarming in developing countries, where native breeds adapted to local environments and diseases are being replaced by industrial breeds. In the most marginal areas, farm animals are considered to be essential for viable land use and, in the developing world, a major pathway out of poverty.

Historic documentation from the period before the breed formation is scarce. Thus, reconstruction of the history of livestock populations depends on archaeological, archeo-zoological and DNA analysis of extant populations. Scientific research into genetic diversity takes advantage of the rapid advances in molecular genetics. Studies of mitochondrial DNA, microsatellite DNA profiling and Y-chromosomes have revealed details on the process of domestication, on the diversity retained by breeds and on relationships between breeds. However, we only see a small part of the genetic information and the advent of new technologies is most timely in order to answer many essential questions.

High-throughput single-nucleotide polymorphism genotyping is about to be available for all major farm animal species. The recent development of sequencing techniques calls for new methods of data management and analysis and for new ideas for the extraction of information. To make sense of this information in practical conditions, integration of geo-environmental and socio-economic data are key elements. The study and management of farm animal genomic resources (FAnGR) is indeed a major multidisciplinary issue.

The goal of the present Research Topic was to collect contributions of high scientific quality relevant to biodiversity management, and applying new methods to either new genomic and bioinformatics approaches for characterization of FAnGR, to the development of FAnGR conservation methods applied ex-situ and in-situ, to socio-economic aspects of FAnGR conservation, to transfer of lessons between wildlife and livestock biodiversity conservation, and to the contribution of FAnGR to a transition in agriculture (FAnGR and agro-ecology).

Citation: Joost, S., Bruford, M. W., Curik, I., Kantanen, J., Lenstra, J. A., Sölkner, J., Andersson, G., Baret, P. V., Buys, N., Roosen, J., Tixier-Boichard, M., Marsan, P. A., eds. (2016). Advances in Farm Animal Genomic Resources. Lausanne: Frontiers Media. doi: 10.3389/978-2-88919-735-4

Table of Contents

08 ***Editorial: Advances in Farm Animal Genomic Resources***
Stéphane Joost, Michael W. Bruford and The Genomic-Resources Consortium

Challenges for the conservation of farm animal genomic resources

13 ***Prospects and challenges for the conservation of farm animal genomic resources, 2015-2025***
Michael W. Bruford, Catarina Ginja, Irene Hoffmann, Stéphane Joost, Pablo Orozco-terWengel, Florian J. Alberto, Andreia J. Amaral, Mario Barbato, Filippo Biscarini, Licia Colli, Mafalda Costa, Ino Curik, Solange Duruz, Maja Ferenčaković, Daniel Fischer, Robert Fitak, Linn F. Groeneveld, Stephen J. G. Hall, Olivier Hanotte, Faiz-ul Hassan, Philippe Helsen, Laura Iacolina, Juha Kantanen, Kevin Leempoel, Johannes A. Lenstra, Paolo Ajmone-Marsan, Charles Masembe, Hendrik-Jan Megens, Mara Miele, Markus Neuditschko, Ezequiel L. Nicolazzi, François Pompanon, Jutta Roosen, Natalia Sevane, Anamarija Smetko, Anamaria Štambuk, Ian Streeter, Sylvie Stucki, China Supakorn, Luis Telo Da Gama, Michèle Tixier-Boichard, Daniel Wegmann and Xiangjiang Zhan

24 ***What can livestock breeders learn from conservation genetics and* vice versa?**
Torsten N. Kristensen, Ary A. Hoffmann, Cino Pertoldi and Astrid V. Stronen

36 ***Changing values of farm animal genomic resources. from historical breeds to the Nagoya Protocol***
Sakari Tamminen

45 ***Genetic resources and genomics for adaptation of livestock to climate change***
Paul J. Boettcher, Irene Hoffmann, Roswitha Baumung, Adam G. Drucker, Concepta McManus, Peer Berg, Alessandra Stella, Linn B. Nilsen, Dominic Moran, Michel Naves and Mary C. Thompson

FAnGR in Africa

49 ***Characterizing neutral genomic diversity and selection signatures in indigenous populations of Moroccan goats (*Capra hircus*) using WGS data***
Badr Benjelloun, Florian J. Alberto, Ian Streeter, Frédéric Boyer, Eric Coissac, Sylvie Stucki, Mohammed BenBati, Mustapha Ibnelbachyr, Mouad Chentouf, Abdelmajid Bechchari, Kevin Leempoel, Adriana Alberti, Stefan Engelen, Abdelkader Chikhi, Laura Clarke, Paul Flicek, Stéphane Joost, Pierre Taberlet, François Pompanon and NextGen Consortium

63 ***Trypanosomosis: Potential driver of selection in African cattle***
Anamarija Smetko, Albert Soudre, Katja Silbermayr, Simone Müller, Gottfried Brem, Olivier Hanotte, Paul J. Boettcher, Alessandra Stella, Gábor Mészáros, Maria Wurzinger, Ino Curik, Mathias Müller, Jörg Burgstaller and Johann Sölkner

71 ***Genomic adaptation of admixed dairy cattle in East Africa***
Eui-Soo Kim and Max F. Rothschild

81 ***Population genetic structure, linkage disequilibrium and effective population size of conserved and extensively raised village chicken populations of Southern Africa***
Khulekani S. Khanyile, Edgar F. Dzomba and Farai C. Muchadeyi

92 ***Genetic diversity and population structure among six cattle breeds in South Africa using a whole genome SNP panel***
Sithembile O. Makina, Farai C. Muchadeyi, Este van Marle-Köster, Michael D. MacNeil and Azwihangwisi Maiwashe

The role of social science in the management of FAnGR

100 ***Comparing decision-support systems in adopting sustainable intensification criteria***
Bouda Vosough Ahmadi, Dominic Moran, Andrew P. Barnes and Philippe V. Baret

105 ***Citizens' preferences for the conservation of agricultural genetic resources***
Eija Pouta, Annika Tienhaara and Heini Ahtiainen

115 ***Farm animal genetic and genomic resources from an agroecological perspective***
Michèle Tixier-Boichard, Etienne Verrier, Xavier Rognon and Tatiana Zerjal

118 ***Utilization of farm animal genetic resources in a changing agro-ecological environment in the Nordic countries***
Juha Kantanen, Peter Løvendahl, Erling Strandberg, Emma Eythorsdottir, Meng-Hua Li, Anne Kettunen-Præbel, Peer Berg and Theo Meuwissen

Demographic events and diversity in cattle

128 ***Genomic data as the "hitchhiker's guide" to cattle adaptation: Tracking the milestones of past selection in the bovine genome***
Yuri T. Utsunomiya, Ana M. Pérez O'Brien, Tad S. Sonstegard, Johann Sölkner and José F. Garcia

141 ***Revisiting demographic processes in cattle with genome-wide population genetic analysis***
Pablo Orozco-terWengel, Mario Barbato, Ezequiel Nicolazzi, Filippo Biscarini, Marco Milanesi, Wyn Davies, Don Williams, Alessandra Stella, Paolo Ajmone-Marsan and Michael W. Bruford

156 ***Microsatellite genotyping of medieval cattle from central Italy suggests an old origin of Chianina and Romagnola cattle***
Maria Gargani, Lorraine Pariset, Johannes A. Lenstra, Elisabetta De Minicis, European Cattle Genetic Diversity Consortium and Alessio Valentini

162 ***Hybrid origin of European commercial pigs examined by an in-depth haplotype analysis on chromosome 1***
Mirte Bosse, Ole Madsen, Hendrik-Jan Megens, Laurent A. F. Frantz, Yogesh Paudel, Richard P. M. A. Crooijmans and Martien A. M. Groenen

171 **SNeP*: A tool to estimate trends in recent effective population size trajectories using genome-wide SNP data***
Mario Barbato, Pablo Orozco-terWengel, Miika Tapio and Michael W. Bruford

Local breeds

178 ***Challenges and opportunities in genetic improvement of local livestock breeds***
Filippo Biscarini, Ezequiel L. Nicolazzi, Alessandra Stella, Paul J. Boettcher and Gustavo Gandini

185 ***Recent advances in understanding the genetic resources of sheep breeds locally-adapted to the UK uplands: Opportunities they offer for sustainable productivity***
Dianna Bowles

190 ***A case study on strains of Buša cattle structured into a metapopulation to show the potential for use of single-nucleotide polymorphism genotyping in the management of small, cross-border populations of livestock breeds and varieties***
Elli T. Broxham, Waltraud Kugler and Ivica Medugorac

192 ***Genomic analysis for managing small and endangered populations: A case study in Tyrol Grey cattle***
Gábor Mészáros, Solomon A. Boison, Ana M. Pérez O'Brien, Maja Ferenčaković, Ino Curik, Marcos V. Barbosa Da Silva, Yuri T. Utsunomiya, Jose F. Garcia and Johann Sölkner

204 ***Characterization of genetic diversity and gene mapping in two Swedish local chicken breeds***
Anna M. Johansson and Ronald M. Nelson

212 ***Morphological and genetic characterization of an emerging Azorean horse breed: The Terceira Pony***
Maria S. Lopes, Duarte Mendonça, Horst Rojer, Verónica Cabral, Sílvia X. Bettencourt and Artur da Câmara Machado

219 ***Fecal egg counts for gastrointestinal nematodes are associated with a polymorphism in the MHC-DRB1 gene in the Iranian Ghezel sheep breed***
Rahman Hajializadeh Valilou, Seyed A. Rafat, David R. Notter, Djalil Shojda, Gholamali Moghaddam and Ahmad Nematollahi

Approaches and tools for breeding programs

231 ***An interpretive review of selective sweep studies in* Bos taurus *cattle populations: Identification of unique and shared selection signals across breeds***
Beatriz Gutiérrez-Gil, Juan J. Arranz and Pamela Wiener

251 ***Artificial selection with traditional or genomic relationships: Consequences in coancestry and genetic diversity***
Silvia Teresa Rodríguez-Ramilo, Luis Alberto García-Cortés and María Ángeles Rodríguez de Cara

262 ***Genome-wide association and pathway analysis of feed efficiency in pigs reveal candidate genes and pathways for residual feed intake***
Duy N. Do, Anders B. Strathe, Tage Ostersen, Sameer D. Pant and Haja N. Kadarmideen

272 ***Changes in variance explained by top SNP windows over generations for three traits in broiler chicken***
Breno de Oliveira Fragomeni, Ignacy Misztal, Daniela Lino Lourenco, Ignacio Aguilar, Ronald Okimoto and William M. Muir

279 ***Genetic differentiation of Mexican Holstein cattle and its relationship with Canadian and U.S. Holsteins***

Adriana García-Ruiz, Felipe de J. Ruiz-López, Curtis P. Van Tassell, Hugo H. Montaldo and Heather J. Huson

286 ***Assessment of autozygosity in Nellore cows (*Bos indicus*) through high-density SNP genotypes***

Ludmilla B. Zavarez, Yuri T. Utsunomiya, Adriana S. Carmo, Haroldo H. R. Neves, Roberto Carvalheiro, Maja Ferenčaković, Ana M. Pérez O'Brien, Ino Curik, John B. Cole, Curtis P. Van Tassell, Marcos V. G. B. da Silva, Tad S. Sonstegard, Johann Sölkner and José F. Garcia

frontiers in Genetics

EDITORIAL
published: 24 November 2015
doi: 10.3389/fgene.2015.00333

Editorial: Advances in Farm Animal Genomic Resources

Stéphane Joost[1]*, *Michael W. Bruford*[2,3] *and The Genomic-Resources Consortium*[†]

[1] *Laboratory of Geographic Information Systems, School of Architecture, Civil and Environmental Engineering, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland,* [2] *School of Biosciences, Cardiff University, Cardiff, UK,* [3] *Sustainable Places Research Institute, Cardiff University, Cardiff, UK*

Keywords: genomic resources, conservation of genomic diversity, data integration, GIS, next generation sequencing, social sciences, disease resistance, sustainable breeding

Livestock conservation is changing rapidly in light of policy developments, climate change, and diversifying market demands. The last decade has seen a step change in technology and analytical approaches available to define and manage Farm Animal Genetic Resources (FAnGR). However, these rapid changes pose challenges for FAnGR management in terms of technological continuity, analytical capacity, and integrative methodologies. Indeed, high-throughput single-nucleotide polymorphism genotyping is available for all major farm animal species and beyond the technological challenge to deal with these large molecular datasets, their integration with geo-environmental and socio-economic information is key to making sense of the data in practical conditions.

In this context, a 4-year (2010–2014) European Science Foundation (http://www.esf.org) Research Networking Programme "Advances in Farm Animal Genomic Resources" (Genomic-Resources) proposed an action dedicated to the education of young scientists in cutting edge approaches to the characterization, analysis, evaluation, management, and conservation of FAnGR. The RNP funded three summer schools (Italy, Croatia, Austria), three workshops (Switzerland, Iceland, Finland), two conferences (Belgium, United Kingdom), and 26 exchange grants. These actions directly connected a community of 350 researchers to develop activities with the goal to meet two major challenges: (i) training in the use of novel methods able to manage and analyse high-throughput molecular data, and (ii) promoting collaboration between the animal science and social science communities to more efficiently manage FAnGR.

In addition to the activities described above, Genomic-Resources has, in this issue of *Frontiers in Genetics*, fostered scientific contributions applying new methods to genomic and bioinformatics approaches for characterization of FAnGR, enhancing *ex-situ* and *in-situ* FAnGR conservation methods, promoting socio-economic elements of FAnGR conservation, transferring lessons between wildlife and livestock biodiversity conservation and has evaluated the contribution of FAnGR to a transition in agriculture (agro-ecology). The 31 articles can be broadly attributed to six different topic areas.

The first topic area contains general papers dealing with the identification of questions of highest priority for FAnGR research during the coming decade (Bruford et al., 2015), the common management challenges shared by livestock breeds and threatened natural populations (Kristensen et al., 2015), the transformation of FAnGR from economic, ecological, and scientific into political entities (Tamminen, 2015), and on the impact of climate change on genetic resources, which constitute the livelihoods of around 1 billion people worldwide (Boettcher et al., 2015).

The latter paper links to a section of papers dedicated to Africa, a continent where livestock genetic resources are particularly endangered and where climate change poses a major threat. The need to increase short-term productivity is prompting the substitution of local breeds by cosmopolitan ones, with the consequence of breeders becoming dependent on expensive external inputs (Colli et al., 2014) instead of making greater use of the well-adapted livestock already

OPEN ACCESS

Edited and reviewed by:
Guilherme J. M. Rosa, University of Wisconsin, USA

***Correspondence:**
Stéphane Joost
stephane.joost@epfl.ch

[†] *http://genomic-resources.epfl.ch*

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: *14 September 2015*
Accepted: *04 November 2015*
Published: *24 November 2015*

Citation:
Joost S, Bruford MW and The Genomic-Resources Consortium (2015) Editorial: Advances in Farm Animal Genomic Resources. Front. Genet. 6:333. doi: 10.3389/fgene.2015.00333

living in Africa (Hanotte et al., 2010). Here Benjelloun et al. (2015) characterize neutral genomic diversity and selection signatures in local Moroccan goat populations illustrating the use of whole genome sequence data. To the South, in Burkina Faso, Trypanosomosis transmitted by tsetse-flies is a cause of productivity reduction in cattle. To better understand resistance to the disease, Smetko et al. (2015) compare levels of zebu and taurine admixture in genomic regions possibly involved in trypanotolerance. Eastwards, in Kenya, Kim and Rothschild (2014) analyze the ancestry of local cattle admixed with imported breeds including Guernsey, Norwegian Red, and Holstein to provide useful information for dairy breeding. In Malawi, Zimbabwe, and South Africa, Khanyile et al. (2015) focus on village chicken production and investigated the genetic structure and diversity in more than 300 individuals with a High-density SNP assay to extract valuable information useful for indigenous animal genetic resources management. Finally, in South Africa Makina et al. (2014) investigate genetic diversity and population structure among six cattle breeds using a whole genome SNP panel to examine the possible valuable distinctiveness of indigenous South African breeds likely to cope with climate change.

To stress the important role of social science and its links with animal science in FAnGR management, Genomic-Resources encouraged interaction between the fields (see http://www.genresandpit.eu/) and solicited contributions to this special issue to illustrate the inputs of these disciplines. Key contributions within this scope constitute part of section 3. Ahmadi et al. (2015) highlight the role of decision support systems—used for FAnGR prioritization for example—to integrate technical and social aspects of farming practices. The social dimension of FAnGR conservation is also exemplified by citizen's willingness to pay for conservation programs according to their preferences for native breeds, as shown by Pouta et al. (2014). Section 3 is complemented by papers dealing with agro-ecology. As argued by Tixier-Boichard et al. (2015), the application of agro-ecology to livestock production requires a change of scale in breed management, and represents a social rather than a genetic challenge. Then, concerned with the physical dimension of agro-ecology, Kantanen et al. (2015) review the main changes in Nordic agro-climatic conditions caused in part by livestock production, stressing the importance of animals' ability to adapt.

The fourth section is dedicated to the analysis of demographic events that have shaped cattle diversity. Utsunomiya et al. (2015) track the milestones of past selection in the bovine genome, Orozco-terWengel et al. (2015) revisit demographic processes in cattle with genome-wide data, and Gargani et al. (2015) show how DNA from archeological remains can be used to interpret the history of ancient populations and their supposed relationship with Chianina and Romagnola, two modern central-Italian breeds. Bosse et al. (2015) investigate different events in the history of the domestication of the Eurasian wild boar (*Sus scrofa*), comparing the genomes of European commercial pigs to their wild ancestors. The section closes with a description of SNeP, a tool to estimate changes in effective population size using genome-wide SNP data which can improve our understanding of population demography in the recent past (Barbato et al., 2015).

The next section in the special issue focuses on local breeds. It is introduced by two mini-review papers on the relevance of genetic improvement for these breeds (Biscarini et al., 2015), and on the opportunity represented by locally-adapted livestock breeds in the United Kingdom as valuable reservoirs of adaptive fitness to face productivity issues under changing climate (Bowles, 2015). Also highlighting the characteristics of local breeds, Broxham et al. (2015) describe the BushaLive project targeting the autochthonous Buša cattle of the Balkans. The use of genomic analysis to manage small endangered populations (Mészáros et al., 2015), the determination of comb color in two Swedish local chicken breeds (Johansson and Nelson, 2015), the genetic characterization of the Terceira Pony from the Azores (Lopes et al., 2015), and resistance to gastrointestinal nematodes in the Iranian Ghezel sheep (Valilou et al., 2015) comprise the other research papers presented in this section.

The final section combines research on tools and approaches used in the context of (industrial) breeding programs. Gutiérrez-Gil et al. (2015) compile a review of studies with more than 1000 selection signatures in mainly beef and dairy breeds, and propose a characterization of these selective sweeps. The other contributions illustrate different applications related to estimated breeding value (Rodríguez-Ramilo et al., 2015), genomic selection (Do et al., 2014; Fragomeni Bde et al., 2014), and genetic differentiation (García-Ruiz et al., 2015).

This research topic has provided a valuable set of papers taking stock of the current advances in farm animal genomic resources worldwide. However, it inevitably lacks contributions in some areas, such as incorporation of Geographic Information Systems (GIS) to integrate complementary data on population genetics, animal husbandry practices, socio-economic and environmental characteristics (Joost et al., 2010) needed to enable the "landscape approach" advocated (Boettcher et al., 2015), and featuring in several of the top 20 questions in farm animal genomics research (Bruford et al. 2015, Table 1, questions #4, #12, #13 and #14). While the integrative function they provide is likely to identify potentially valuable genetic material (Hanotte et al., 2010), GIS and related approaches remain underexploited in FAnGR management.

AUTHOR CONTRIBUTION

SJ initiated the research topic and chaired the corresponding European Science Foundation project, SJ and MB wrote and revised the manuscript, members of the Genomic-Resources consortium managed the ESF project and participated in the editorial process of this research topic.

ACKNOWLEDGMENTS

GENOMIC-RESOURCES was supported by: Fonds zur Förderung der wissenschaftlichen Forschung (FWF), FWF Austrian Science Fund, Austria; Fonds National de la Recherche Scientifique (FNRS), Belgium; Fonds voor Wetenschappelijk Onderzoek-Vlaanderen (FWO), The Research Foundation, Flanders, Belgium; Nacionalna zaklada za znanost, visoko školstvo i tehnologijski razvoj Republike Hrvatske, Croatian

Science Foundation, Republic of Croatia; Suomen Akatemia, Biotieteiden ja ympäristön tutkimuksen toimikunta, Academy of Finland, Research Council for Biosciences and Environment, Finland; Deutsche Forschungsgemeinschaft (DFG), German Research Foundation, Germany; Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO), The Netherlands Organisation for Scientific Research, The Netherlands; Norges Forskningsråd, The Research Council of Norway, Norway; Forskningsrådet för miljö, areella näringar och samhällsbyggande, Swedish Council for Environment, Agricultural Sciences and Spatial Planning (FORMAS), Sweden; Schweizerischer Nationalfonds (SNF), Swiss National Science Foundation, Switzerland; Biotechnology and Biological Sciences Research Council (BBSRC), United Kingdom.

REFERENCES

Ahmadi, B. V., Moran, D., Barnes, A. P., and Baret, P. V. (2015). Comparing decision-support systems in adopting sustainable intensification criteria. *Front Genet*. 6:23. doi: 10.3389/fgene.2015.00023

Barbato, M., Orozco-terWengel, P., Tapio, M., and Bruford, M. W. (2015). SNeP: a tool to estimate trends in recent effective population size trajectories using genome-wide SNP data. *Front Genet*. 6:109. doi: 10.3389/fgene.2015.00109

Benjelloun, B., Alberto, F. J., Streeter, I., Boyer, F., Coissac, E., Stucki, S., et al. (2015). Characterizing neutral genomic diversity and selection signatures in indigenous populations of Moroccan goats (*Capra hircus*) using WGS data. *Front Genet*. 6:107. doi: 10.3389/fgene.2015.00107

Biscarini, F., Nicolazzi, E. L., Stella, A., Boettcher, P. J., and Gandini, G. (2015). Challenges and opportunities in genetic improvement of local livestock breeds. *Front Genet*. 6:33. doi: 10.3389/fgene.2015.00033

Boettcher, P. J., Hoffmann, I., Baumung, R., Drucker, A. G., McManus, C., Berg, P., et al. (2015). Genetic resources and genomics for adaptation of livestock to climate change. *Front Genet*. 5:461. doi: 10.3389/fgene.2014.00461

Bosse, M., Madsen, O., Megens, H.-J., Frantz, L. A. F., Paudel, Y., Crooijmans, R. P. M. A., et al. (2015). Hybrid origin of European commercial pigs examined by an in-depth haplotype analysis on chromosome 1. *Front Genet*. 5:442. doi: 10.3389/fgene.2014.00442

Bowles, D. (2015). Recent advances in understanding the genetic resources of sheep breeds locally-adapted to the UK uplands: opportunities they offer for sustainable productivity. *Front Genet*. 6:24. doi: 10.3389/fgene.2015.00024

Broxham, E. T., Kugler, W., and Medugorac, I. (2015). A case study on strains of Buša cattle structured into a metapopulation to show the potential for use of single-nucleotide polymorphism genotyping in the management of small, cross-border populations of livestock breeds and varieties. *Front Genet*. 6:73. doi: 10.3389/fgene.2015.00073

Bruford, W. M., Ginja, C., Hoffmann, I., Joost, S., Orozco-terWengel, P., Alberto, F., et al. (2015). Prospects and challenges for the conservation of farm animal genomic resources, 2015-2025. *Front Genet*. 6:314. doi: 10.3389/fgene.2015.00314

Colli, L., Joost, S., Negrini, R., Nicoloso, L., Crepaldi, P., and Ajmone-Marsan, P. (2014). Assessing the spatial dependence of adaptive loci in 43 European and Western Asian goat breeds using AFLP markers. *PLoS ONE* 9:e86668. doi: 10.1371/journal.pone.0086668

Do, D. N., Strathe, A. B., Ostersen, T., Pant, S. D., and Kadarmideen, H. N. (2014). Genome-wide association and pathway analysis of feed efficiency in pigs reveal candidate genes and pathways for residual feed intake. *Front Genet*. 5:307. doi: 10.3389/fgene.2014.00307

Fragomeni Bde, O., Misztal, I., Lourenco, D. L., Aguilar, I., Okimoto, R., and Muir, W. M. (2014). Changes in variance explained by top SNP windows over generations for three traits in broiler chicken. *Front Genet*. 5:332. doi: 10.3389/fgene.2014.00332

García-Ruiz, A., Ruiz-López Fde, J., Van Tassell, C. P., Montaldo, H. H., and Huson, H. J. (2015). Genetic differentiation of Mexican Holstein cattle and its relationship with Canadian and U.S. Holsteins. *Front Genet*. 6:7. doi: 10.3389/fgene.2015.00007

Gargani, M., Pariset, L., Lenstra, J. A., De Minicis, E., European Cattle Genetic Diversity Consortium and Valentini, A. (2015). Microsatellite genotyping of medieval cattle from central Italy suggests an old origin of Chianina and Romagnola cattle. *Front Genet*. 6:68. doi: 10.3389/fgene.2015.00068

Gutiérrez-Gil, B., Arranz, J. J., and Wiener, P. (2015). An interpretive review of selective sweep studies in Bos taurus cattle populations: identification of unique and shared selection signals across breeds. *Front Genet*. 6:167. doi: 10.3389/fgene.2015.00167

Hanotte, O., Dessie, T., and Kemp, S. (2010). Time to Tap Africa's Livestock Genomes. *Science* 328, 1640–1641. doi: 10.1126/science.1186254

Johansson, A. M., and Nelson, R. M. (2015). Characterization of genetic diversity and gene mapping in two Swedish local chicken breeds. *Front Genet*. 6:44. doi: 10.3389/fgene.2015.00044

Joost, S., Colli, L., Baret, P. V., Garcia, J. F., Boettcher, P. J., Tixier-Boichard, M., et al. (2010). Integrating geo-referenced multiscale and multidisciplinary data for the management of biodiversity in livestock genetic resources. *Anim. Genet.* 41, 47–63. doi: 10.1111/j.1365-2052.2010.02037.x

Kantanen, J., Løvendahl, P., Strandberg, E., Eythorsdottir, E., Li, M.-H., Kettunen-Præbel, A., et al. (2015). Utilization of farm animal genetic resources in a changing agro-ecological environment in the Nordic countries. *Front Genet*. 6:52. doi: 10.3389/fgene.2015.00052

Khanyile, K. S., Dzomba, E. F., and Muchadeyi, F. C. (2015). Population genetic structure, linkage disequilibrium and effective population size of conserved and extensively raised village chicken populations of Southern Africa. *Front Genet*. 6:13. doi: 10.3389/fgene.2015.00013

Kim, E.-S., and Rothschild, M. F. (2014). Genomic adaptation of admixed dairy cattle in East Africa. *Front Genet*. 5:443. doi: 10.3389/fgene.2014.00443

Kristensen, T. N., Hoffmann, A. A., Pertoldi, C., and Stronen, A. V. (2015). What can livestock breeders learn from conservation genetics and vice versa? *Front Genet*. 6:38. doi: 10.3389/fgene.2015.00038

Lopes, M. S., Mendonça, D., Rojer, H., Cabral, V., Bettencourt, S. X., and da Câmara Machado, A. (2015). Morphological and genetic characterization of an emerging Azorean horse breed: the Terceira Pony. *Front Genet*. 6:62. doi: 10.3389/fgene.2015.00062

Makina, S. O., Muchadeyi, F. C., van Marle-Köster, E., MacNeil, M. D., and Maiwashe, A. (2014). Genetic diversity and population structure among six cattle breeds in South Africa using a whole genome SNP panel. *Front Genet*. 5:333. doi: 10.3389/fgene.2014.00333

Mészáros, G., Boison, S. A., Pérez O'Brien, A. M., Ferenčaković, M., Curik, I., Da Silva, M. V. B., et al. (2015). Genomic analysis for managing small and endangered populations: a case study in Tyrol Grey cattle. *Front Genet*. 6:173. doi: 10.3389/fgene.2015.00173

Orozco-terWengel, P., Barbato, M., Nicolazzi, E., Biscarini, F., Milanesi, M., Davies, W., et al. (2015). Revisiting demographic processes in cattle with genome-wide population genetic analysis. *Front Genet*. 6:191. doi: 10.3389/fgene.2015.00191

Pouta, E., Tienhaara, A., and Ahtiainen, H. (2014). Citizens' preferences for the conservation of agricultural genetic resources. *Front Genet*. 5:440. doi: 10.3389/fgene.2014.00440

Rodríguez-Ramilo, S. T., García-Cortés, L. A., and de Cara, M. Á. R. (2015). Artificial selection with traditional or genomic relationships: consequences in coancestry and genetic diversity. *Front Genet*. 6:127. doi: 10.3389/fgene.2015.00127

Smetko, A., Soudre, A., Silbermayr, K., Müller, S., Brem, G., Hanotte, O., et al. (2015). Trypanosomosis: potential driver of selection in African cattle. *Front Genet*. 6:137. doi: 10.3389/fgene.2015.00137

Tamminen, S. (2015). Changing values of farm animal genomic resources. from historical breeds to the Nagoya Protocol. *Front. Genet*. 6:279. doi: 10.3389/fgene.2015.00279

Tixier-Boichard, M., Verrier, E., Rognon, X., and Zerjal, T. (2015). Farm animal genetic and genomic resources from an agroecological perspective. *Front. Genet*. 6:153. doi: 10.3389/fgene.2015.00153

Utsunomiya, Y. T., Pérez O'Brien, A. M., Sonstegard, T. S., Sölkner, J., and Garcia, J. F. (2015). Genomic data as the "hitchhiker's guide" to cattle adaptation: tracking the milestones of past selection in the bovine genome. *Front Genet*. 6:36. doi: 10.3389/fgene.2015.00036

Valilou, R. H., Rafat, S. A., Notter, D. R., Shojda, D., Moghaddam, G., and Nematollahi, A. (2015). Fecal egg counts for gastrointestinal nematodes are associated with a polymorphism in the MHC-DRB1 gene in the Iranian Ghezel sheep breed. *Front Genet*. 6:105. doi: 10.3389/fgene.2015.00105

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Joost, Bruford and The Genomic-Resources Consortium. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

CHALLENGES FOR THE CONSERVATION OF FARM ANIMAL GENOMIC RESOURCES

frontiers in Genetics

HYPOTHESIS AND THEORY
published: 21 October 2015
doi: 10.3389/fgene.2015.00314

Prospects and challenges for the conservation of farm animal genomic resources, 2015-2025

Michael W. Bruford[1,2], Catarina Ginja[3,4], Irene Hoffmann[5], Stéphane Joost[6], Pablo Orozco-terWengel[1], Florian J. Alberto[7], Andreia J. Amaral[8], Mario Barbato[1], Filippo Biscarini[9], Licia Colli[10], Mafalda Costa[1], Ino Curik[11], Solange Duruz[6], Maja Ferenčaković[11], Daniel Fischer[12], Robert Fitak[13], Linn F. Groeneveld[14], Stephen J. G. Hall[15], Olivier Hanotte[16], Faiz-ul Hassan[16,17], Philippe Helsen[18], Laura Iacolina[19], Juha Kantanen[12,20], Kevin Leempoel[6], Johannes A. Lenstra[21], Paolo Ajmone-Marsan[10], Charles Masembe[22], Hendrik-Jan Megens[23], Mara Miele[24], Markus Neuditschko[25], Ezequiel L. Nicolazzi[9], François Pompanon[7], Jutta Roosen[26], Natalia Sevane[27], Anamarija Smetko[28], Anamaria Štambuk[29], Ian Streeter[30], Sylvie Stucki[6], China Supakorn[16,31], Luis Telo Da Gama[32], Michèle Tixier-Boichard[33], Daniel Wegmann[34] and Xiangjiang Zhan[35,36]*

[1] School of Biosciences, Cardiff University, Cardiff, UK, [2] Sustainable Places Research Institute, Cardiff University, Cardiff, UK, [3] Faculdade de Ciências, Centro de Ecologia, Evolução e Alterações Ambientais (CE3C), Universidade de Lisboa, Lisboa, Portugal, [4] Centro de Investigação em Biodiversidade e Recursos Genéticos (CIBIO-InBIO), Universidade do Porto, Campus Agrário de Vairão, Portugal, [5] Food and Agriculture Organization of the United Nations, Animal Genetic Resources Branch, Animal Production and Health Division, Rome, Italy, [6] Laboratory of Geographic Information Systems (LASIG), School of Civil and Environmental Engineering (ENAC), Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, [7] Laboratoire d'Ecologie Alpine, Université Grenoble Alpes, Grenoble, France, [8] Faculty of Sciences, BioISI- Biosystems and Integrative Sciences Institute, University of Lisbon, Campo Grande, Portugal, [9] Parco Tecnologico Padano, Lodi, Italy, [10] BioDNA Centro di Ricerca sulla Biodiversità a sul DNA Antico, Istituto di Zootecnica, Università Cattolica del Sacro Cuore di Piacenza, Italy, [11] Faculty of Agriculture, University of Zagreb, Zagreb, Croatia, [12] Natural Resources Institute Finland (Luke), Green Technology, Jokioinen, Finland, [13] Institut für Populationsgenetik, Vetmeduni, Vienna, Austria, [14] NordGen -The Nordic Genetic Resource Center, Ås, Norway, [15] Livestock Diversity Ltd., Lincoln, UK, [16] School of Life Sciences, University of Nottingham, Nottingham, UK, [17] Department of Animal Breeding and Genetics, University of Agriculture, Faisalabad, Pakistan, [18] Centre for Research and Conservation, Royal Zoological Society of Antwerp, Antwerp, Belgium, [19] Department of Chemistry and Bioscience, Aalborg University, Aalborg, Denmark, [20] Department of Biology, University of Eastern Finland, Kuopio, Finland, [21] Faculty of Veterinary Medicine, Utrecht University, Utrecht, Netherland, [22] Institute of the Environment and Natural Resources, Makerere University, Kampala, Uganda, [23] Animal Breeding and Genomics Centre, Wageningen University, Wageningen, Netherlands, [24] School of Planning and Geography, Cardiff University, Cardiff, UK, [25] Agroscope, Swiss National Stud Farm, Avenches, Switzerland, [26] TUM School of Management, Technische Universität München, Munich, Germany, [27] Department of Animal Production, Veterinary Faculty, Universidad Complutense de Madrid, Madrid, Spain, [28] Croatian Agricultural Agency, Zagreb, Croatia, [29] Department of Biology, Faculty of Science, University of Zagreb, Zagreb, Croatia, [30] European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK, [31] School of Agricultural Technology, Walailak University, Tha Sala, Thailand, [32] Centre of Research in Animal Health (CIISA) – Faculty of Veterinary Medicine, University of Lisbon, Lisbon, Portugal, [33] INRA, AgroParisTech, UMR GABI, Jouy-en-Josas, France, [34] Department of Biology, University of Fribourg, Fribourg, Switzerland, [35] Key Laboratory of Animal Ecology and Conservation Biology, Institute of Zoology, Chinese Academy of Sciences, Beijing, China, [36] Cardiff University – Institute of Zoology, Joint Laboratory for Biocomplexity Research, Beijing, China

OPEN ACCESS

Edited by:
Peter Dovc,
University of Ljubljana, Slovenia

Reviewed by:
Juan Steibel,
Michigan State University, USA
John B. Cole,
United States Department of Agriculture, USA

***Correspondence:**
Michael W. Bruford
brufordmw@cardiff.ac.uk

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: *26 May 2015*
Accepted: *05 October 2015*
Published: *21 October 2015*

Livestock conservation practice is changing rapidly in light of policy developments, climate change and diversifying market demands. The last decade has seen a step change in technology and analytical approaches available to define, manage and conserve Farm Animal Genomic Resources (FAnGR). However, these rapid changes pose challenges for FAnGR conservation in terms of technological continuity, analytical capacity and integrative methodologies needed to fully exploit new, multidimensional data. The final conference of the ESF Genomic Resources program aimed to address

these interdisciplinary problems in an attempt to contribute to the agenda for research and policy development directions during the coming decade. By 2020, according to the Convention on Biodiversity's Aichi Target 13, signatories should ensure that "*…the genetic diversity of …farmed and domesticated animals and of wild relatives …is maintained, and strategies have been developed and implemented for minimizing genetic erosion and safeguarding their genetic diversity*." However, the real extent of genetic erosion is very difficult to measure using current data. Therefore, this challenging target demands better coverage, understanding and utilization of genomic and environmental data, the development of optimized ways to integrate these data with social and other sciences and policy analysis to enable more flexible, evidence-based models to underpin FAnGR conservation. At the conference, we attempted to identify the most important problems for effective livestock genomic resource conservation during the next decade. Twenty priority questions were identified that could be broadly categorized into challenges related to methodology, analytical approaches, data management and conservation. It should be acknowledged here that while the focus of our meeting was predominantly around genetics, genomics and animal science, many of the practical challenges facing conservation of genomic resources are societal in origin and are predicated on the value (e.g., socio-economic and cultural) of these resources to farmers, rural communities and society as a whole. The overall conclusion is that despite the fact that the livestock sector has been relatively well-organized in the application of genetic methodologies to date, there is still a large gap between the current state-of-the-art in the use of tools to characterize genomic resources and its application to many non-commercial and local breeds, hampering the consistent utilization of genetic and genomic data as indicators of genetic erosion and diversity. The livestock genomic sector therefore needs to make a concerted effort in the coming decade to enable to the democratization of the powerful tools that are now at its disposal, and to ensure that they are applied in the context of breed conservation as well as development.

Keywords: farm animal genetic resources, livestock genetic resources, genomic diversity, livestock population prioritization, effective conservation policy

INTRODUCTION

Understanding current technical, infrastructural and policy challenges and assessing the likely benefits of overcoming them in the future is essential for any field of scientific endeavor and especially those with clear societal consequences and potential benefits. In this context, the concept of horizon scanning has been developed and applied annually in the field of biodiversity conservation since 2009 (Sutherland and Woodroof, 2009), using a variety of systematic and semi-systematic methods to mine trending issues from web engines and social media and by analyzing focused questionnaires. Similar approaches have also been taken to identify emerging issues in agriculture (Pretty et al., 2010) and related fields such as soil science, food systems and pollination (Dicks et al., 2013; Ingram et al., 2013; Adewopo et al., 2014). Such exercises have identified a number of issues of relevance to the conservation of FAnGR, such as genetic control of invasive species (Sutherland et al., 2014) and sustainable intensification of high yielding agriculture (Sutherland et al., 2015). In 2010, Pretty et al.'s article pinpointing the "*Top 100 questions of importance to the future of global agriculture*" identified genetic issues in crop improvement (e.g., gains in improvement that could result from breeding for stress tolerance) but identified no such pressing agendas for livestock genomic resources. Since Cardellino and Boyazoglu (2009) no attempt has been published to identify research priorities for FAnGR conservation, despite genetic erosion (*sensu* Aichi Target 13) continuing apace (e.g., Berthouly-Salazar et al., 2012; FAO, 2015a) and the step-change that has occurred in molecular breed characterization since the routine implementation of livestock Single Nucleotide Polymorphism (SNP) arrays. To fill this gap, a central activity of the Final Conference of the European Science Foundation's Genomic Resources program, held at Cardiff University June 17th–19th 2014 was to pick out a series of pressing questions that could form part of a research and policy agenda for FAnGR conservation for the next decade. While not following the standard systematic approaches adopted by conventional Horizon Scanning exercises, all 43 attendees of this focused meeting took part in the exercise, including scientists and policy-makers from South and East Asia, North America, Europe

and Africa involved in a range of disciplines from genomics to animal breeding, genetic resource management, economic and social sciences and global agricultural policy development.

METHODS AND RESULTS

During the course of the conference, attendees were asked to contribute up to five questions of highest priority for research, infrastructure and policy development during the coming decade. Eighty-six suggestions were received. The issue identified with highest frequency (18 times) was the need for "next generation phenotyping" (i.e., high-throughput methods to collect and summarize detailed phenotypic data from domestic animals). A summary of the top 20 questions is found in **Table 1**, a subset of which are presented below (some are amalgamated). All responses were categorized into four major groups, "Methodological Challenges," "Analytical Challenges," "Data Management," and "Conservation Management and Prioritization." Four working groups were convened to cover these categories and their findings are presented below.

Methodological Challenges

Next Generation Phenotyping

The need for high-resolution phenotypic data to be collected for in-depth characterization of FAnGR was identified, especially in light of the rapid advances that have been made in molecular breed characterization. Developing methods for phenotypic characterization was also identified by Cardellino and Boyazoglu (2009) following from FAO recommendations (FAO, 2007a) and has clearly remained an under-explored research area. However, with the richness of molecular data increasing dramatically since 2009, the mismatch between molecular and phenotypic data is widening for all except highly commercial transboundary breeds and lines with genomic breeding values. Inherent in high-resolution breed characterization is a need to define key phenotypic traits and characteristics (particularly those potentially involved in local adaptation) based on guidelines that can be used as common measures for such studies with stringent field protocols for their collection. FAO published guidelines on phenotypic characterization (FAO, 2012a). In this way more comparable data can be generated, and breed characterization can have a more functional basis, especially with the urgent need to understand breed characteristics in the face of climate change (Hoffmann, 2010). Also an improved description of the specific production environment and epidemiological history in which populations of a breed are kept would allow better comparison of phenotypes and performances (e.g., FAO, 2009). Since breed characterization can be a costly exercise, especially for remote regions of the world, as many phenotypic traits as possible should be collected following well documented and reproducible procedures, a process that calls for the need for standardized methods to measure/collect data and ultimately for training of people on how to do it. Where possible, data should be made publicly available through a repository such as FAO's global Domestic Animal Diversity Information System DAD-IS (http://dad.fao.org) for comparative purposes. The establishment of a working group to define guidelines, protocols and tools for collecting such data under the auspices of the FAO, International Society for Animal Genetics or the International Committee for Animal Recording (www.icar.org) would accelerate this process.

Omics Data and Association Studies

The dramatic acceleration in genome sequencing means that all domesticated species and their few remaining wild relatives will become genome-enabled in the coming decade (e.g., Qiu et al., 2012; Wu et al., 2014). Reference genomes provide the basis for development of genome-wide assays for variation in less commonly farmed and/or more regionally distributed livestock species and populations using SNP arrays, as have been developed and made available for commercial livestock in the past 5 years (e.g., Matukumalli et al., 2009). The choice of SNPs for inclusion in arrays for less commercial populations may be expected to focus on a wider array of traits than for commercial/transboundary breeds, such as those related to local adaptation, disease resistance, drought tolerance and niche product characters, but in practice this could be hampered by a lack of reliable phenotypic data. To enable SNP arrays to be developed in a rapid, cost effective and widely applicable manner, the identification of common reference genomes and test panels of individuals for array development and diversity studies is key. However, it is important to note that with the rapidly falling cost of whole genome resequencing (e.g., Lee et al., 2013; Zhang et al., 2015) using next generation technologies and the availability of even lower cost genotyping by sequencing (GBS: De Donato et al., 2013) being available, the problem of ascertainment bias can be mitigated against since they allow the identification and direct estimation of SNP diversity for FAnGR populations, breeds or species at reasonable prices. Indeed these methods are sufficiently cost-effective now, that they can be in principle used as standard assaying approaches, with a cost in the low tens of dollars for GBS now feasible for analysis of tens of thousands of SNPs.

A major issue identified for genome-wide association studies (GWAS) is experimental design including, but not confined to, sample size considerations (Kadarmideen, 2014) and the availability of different SNP genotyping arrays for some species and their compatibility or lack thereof (Nicolazzi et al., 2015). Characterization of environmental parameters in extensive production systems is another key challenge for GWAS but may be assisted by the application of E(environment)WAS methodologies as applied in humans (e.g., Patel et al., 2010). Additionally, understanding the role of the epigenome and its role in environment-dependent phenotypic diversity and plasticity is becoming an increasing focus in livestock genetics (e.g., Jammes and Renard, 2010; Magee et al., 2011, 2014). Ultimately, the integration of genomic, epigenomic, transcriptomic, and environmental data will be required if meaningful large-scale studies are to be successful in identifying selection and conservation targets in heterogeneous environments (Jones et al., 2013; Wu et al., 2014) and in scrutinizing the biological basis for adaptation, resilience, and even animal improvement.

Non-autosomal Inheritance

Non-autosomal inheritance (Y-chromosomal, X-chromosomal, and mitochondrial) is a comparatively neglected area of

TABLE 1 | Summary of the Top 20 questions in farm animal genomics research identified by the participants of the Cardiff symposium.

Question #	
1. Next generation phenotyping	The mismatch between molecular and phenotypic data has increased dramatically. Which key phenotypic traits should be used as common measures for diversity studies to define breed characteristics in the face of climate change?
2. Genome-wide SNP assays	The identification of common reference genomes and test panels of individuals for SNP array development in less commercial and/or local populations is key. Which strategy shall be used to enable SNP arrays to be developed in a rapid, cost effective and widely applicable manner?
3. Reference genomes	Which common reference genomes and test panels of individuals should be used for array development and diversity studies?
4. E(environment)WAS	How to characterize environmental parameters in extensive production systems?
5. Epigenetics	How can epigenomic information be integrated with phenotypic and genomic data to scrutinize the biological basis for adaptation and plasticity/resilience in livestock populations?
6. Male-mediated genetic diversity	Which methodological approach can be applied to promote reliable assembly of the Y-chromosome, still lacking for many livestock species, as well as to develop polymorphic Y-chromosome markers?
7. Ancient DNA and paleoenvironmental analyses	Which strategies should be followed to collect zooarchaeological specimens from critical geographic sites and promote the analysis of ancient genomes?
8. Conservation of genomic diversity	How to design a management program that evaluates genomic regions for conservation?
9. Polygenic adaptive and economic traits	Haplotypes vs. SNPs: in which situations do one or the other provide a more efficient unit of diversity in QTL regions?
10. Microsatellites (STRs) vs. SNPs	How to integrate data from the STRs and SNPs, and how to manage the transition from STR- to SNP-based characterization of FAnGR?
11. GW diversity statistics	Which combination of parameters will be required to adequately summarize genome diversity?
12. Data management	How can links between major FAnGR databases be promoted to be able to federate resources and act as an educational central point?
13. Data availability	Which format should be used to make NGS, phenotyping and GIS data publicly available, and how can industry contribute toward population and maintenance of such database?
14. Participatory projects	How can participatory projects, including citizen science, for example, the use of smart-phone technologies be encouraged to enable data collection on FAnGR at a large scale?
15. Prioritization for conservation	Why are prioritization methods not being applied by policy makers and managers and is there a lack of dissemination or penetrance?
16. Genomic prioritization	How to implement genomic approaches systematically in conservation prioritization to include genes important in functionally valuable traits?
17. Utilization in practice	How to reconcile the cost of genomic analysis vs. the economic returns on genotyped stock to allow for a wider use of genomic data to assist conservation, production and management of FAnGR? What is the demand and willingness to pay within the sector?
18. Systematic collection	How to ensure that genetic and genomic data are collected sufficiently systematically to be applied to new indicators?
19. Defining goals	Which indicators can be applied to most efficiently monitor genetic trends in domestic populations?
20. Strategic approach	How will the latest advances in 'omics technology contribute to achieve the ultimate goal of halting the loss of biodiversity of FAnGR?

Frequencies are not included for each question and the questions are not listed in rank order.

research in livestock conservation. While studies of non-autosomal genetic markers have been extensively used in studies of evolutionary history, both singly and combined (e.g., Götherström et al., 2005; Meadows and Kijas, 2008; Svensson and Götherström, 2008; Pereira et al., 2009; Ramírez et al., 2009; Ginja et al., 2010; Groeneveld et al., 2010), their exploitation in genomic studies has been somewhat overlooked in comparison to autosomal markers in many livestock species. This oversight is surprising given the well-documented links between mitochondrial sequence variation and fitness in human populations (e.g., Wallace, 2005) and the increasingly recognized role that Y-chromosomal variation plays in male fertility in livestock (e.g., Chang et al., 2013; Yue et al., 2014). Technical challenges have long been acknowledged with finding polymorphic markers on the Y-chromosome in mammals and W-chromosome in birds, however such markers, although elusive, have been shown to provide novel insights into livestock diversity when available (e.g., Edwards et al., 2011; Wallner et al., 2013), and should be used as a matter of course to provide a male/female perspective on livestock genomic diversity.

Ancient DNA Studies

Although firmly established as a major route into a deeper understanding of livestock evolution and diversity (e.g., Larson et al., 2010), ancient DNA (aDNA) studies have been hampered by a number of constraints. These include limited access to samples from geographic areas where (local) domestication may have taken place (e.g., Africa, Near East, Asia, South America), limited data sharing among those groups working on samples from critical sites (but see Arbuckle et al., 2014) and limited success rates, especially for genome-wide studies. Nonetheless, recently developed methodological and bioinformatics tools allowed for increased accuracy in the analysis of high-throughput ancient DNA data and even the characterization of complete

genomes of Pleistocene horses (Orlando et al., 2013). Also, alternative sources of material such as parchment are, however, providing promising outcomes (Teasdale et al., 2015). Exciting opportunities have recently been opened up by the discovery of livestock DNA in lake sediment samples in Lake Anterne, Switzerland (Giguet-Covex et al., 2014), which enabled a direct comparison to be made of the paleoenvironment with changes in this environment due to the arrival of farming and domestic livestock, and could be applied to describe historic fluctuations in agricultural intensity and practice and, excitingly, may even allow the possibility of predictive modeling for the presence/absence of suitable agri-habitat under future climate change scenarios.

Analytical Challenges

Conservation of Genomic Diversity

The concept of genome conservation has been discussed extensively in the literature but advances in genome data and technologies only now allow the development of breed management programs able to achieve this aim. For example, Herrero-Medrano et al. (2014), using genome resequencing and SNP arrays discovered almost 100 non-synonymous polymorphic nucleotides nearly fixed in commercial pig breeds but with an alternative allele in non-commercial populations, affecting 65 genes in total. Such genomic polymorphisms could fall into a category of those that "*cannot afford to be lost*" from less commercial local breeds, given their distinctiveness and the value they potentially represent as a genetic resource for alternative selection should the production environment change (Kristensen et al., 2015). However, to design a management program that evaluates genomic regions for conservation, not only do polymorphisms need to be identified, the functional architecture of those genomic regions and the genes they contain needs to be assessed and the interaction among those genes needs to be considered. Recently, a study of chicken breeds examined functional variation in copy number variants (CNV) at over 200 genes overlapping 1000 quantitative trait loci, including some putatively involved in traits such as skin color and skeletal characteristics (Han et al., 2014).

Haplotype Blocks vs. Individual SNPs

Obtaining an accurate description of the genetic polymorphisms explaining a trait of evolutionary, adaptive and/or economic importance is not a trivial task, as traits substantially vary in the number of polymorphisms involved in their phenotype and where these occur across the genome (Goddard and Hayes, 2009; Olson-Manning et al., 2012). For example, many of such traits are polygenic and distributed around the genome, making whole-genome resequencing, and medium and high-density SNP arrays a powerful approach to locating them and elucidating their variation (e.g., Huang et al., 2010). However, for certain linked traits, haplotypes may provide a more efficient unit of assessing diversity in QTL regions than individual SNPs (e.g., Kijas et al., 2013; Bosse et al., 2014a,b; Mokry et al., 2014), reflecting local genomic architecture in a more accurate fashion. Consequently, at the initial stages of studies aiming to identify the genetic basis of phenotypic variation, general genome-wide SNP analyses may be more suitable. It is worth noting, however, that phasing haplotypes in divergent populations lacking complementary pedigree data presents a non-trivial challenge. Haplotype analysis can provide an especially powerful tool to investigate the hybrid origin of domesticated populations. For instance, modern Western commercial pig genomes are a mosaic of Eastern and Western Eurasian biogeographic origin. Admixture mapping allows the "sorting" of haplotype segments for their putative origin. In addition, this strategy has been shown to be powerful to infer selection on specific haplotypes post-hybridization (Bosse et al., 2014a,b).

Managing the Transition from Microsatellite to SNP Data

The transition from microsatellite markers to SNPs has happened rapidly in FAnGR for commercial/transboundary breeds due to the availability of relatively inexpensive 50K SNP genotyping arrays for most common livestock species (Matukumalli et al., 2009). However, SNP arrays are not yet affordable tools for much of the world's FAnGR and are not yet available for all species (see above). This therefore raises the immediate problem of how to integrate data from the two marker types and how to manage the transition from microsatellite-based FAnGR characterization (much of which has been carried out using markers recommended by ISAG, FAO, 2011) to SNP-based characterization. One option is to re-genotype many of the breeds that already have microsatellite genotypes with SNPs (Ajmone-Marsan et al., 2014), but this would be expensive and if implemented would raise the question as to whether the new data would again be replaced by a newer technology (e.g., whole-genome resequencing). Pragmatically, it seems that microsatellite data are perfectly adequate for estimating genetic diversity and describing demographic relationships (e.g., Ferrando et al., 2014). However, for cost reasons the full set of microsatellite markers was frequently not applied, especially in developing countries. Also, microsatellite data will not be as efficient for enabling the identification and targeted conservation of genomic regions under selection since data are usually produced with a few tens of quasi-neutral markers (e.g., Herrero-Medrano et al., 2013).

Nevertheless, it is becoming clear that data produced using SNP arrays are more repeatable and do not suffer from scoring differences that have made the combination of microsatellite datasets sometimes problematic and requiring statistical evaluation (Lenstra et al., 2012). Paradoxically, whole genome resequencing may become the most reliable and cost effective way to analyse genomic diversity in the future, even for non-commercial breeds, if the cost comes down by another order of magnitude (as may happen with portable sequencers such as Oxford Nanopore's MiniION system), providing the advantage of no longer needing to use a set of SNP markers ascertained from commercial populations.

Genome-wide Diversity Statistics

The emergence of whole genome sequencing and medium-high density SNP arrays means that summarizing genetic diversity can now be a more nuanced and genomic region-specific exercise. It is well known that ascertainment bias of SNP arrays can strongly underestimate the diversity of the (usually autochthonous and

less commercial) breeds not used to design the arrays (Porto Neto and Barendse, 2010). This phenomenon does not impact on whole-genome resequencing as all polymorphisms are captured provided sufficient sequence depth is achieved. A combination of parameters will be required to adequately summarize genome diversity (e.g., heterozygosity and effective population size and inbreeding), as no single all-encompassing statistic to summarize all of a population's genomic diversity and history exists, despite of how tempting it may be to define such statistic (e.g., for policy makers). Effective population size (Ne) estimates can be obtained with as little as a single genome using methods such as the Pairwise Sequential Markovian Coalescent, although these analyses can prove inconclusive if genome coverage is insufficient or if admixture pertains (Li and Durbin, 2011; Orozco-terWengel and Bruford, 2014; Schiffels and Durbin, 2014; Frantz et al., 2015). For recently evolved populations, such as many domestic species, linkage disequilibrium-based (LD) estimates may be more accurate and methods are now emerging to carry out these analysis (e.g., Barbato et al., 2015). Runs of homozygosity (ROH; e.g., Bosse et al., 2012; Scraggs et al., 2014) functions describing the distribution of homozygosity throughout the genome may also serve as a robust genome-scale Ne estimator in the future, although interpretation and scaling depends on the local recombination. ROH are already used as a genomic proxy for inbreeding (e.g., Purfield et al., 2012; Curik et al., 2014), including for specific genome-located traits (Pryce et al., 2014). This approach promises to be an efficient way to avoid the production of offspring homozygous for deleterious alleles at specific genomic regions that are associated with inbreeding depression (Pryce et al., 2014).

Data Management

Data Accessibility

As also identified by Cardellino and Boyazoglu (2009) there remains a major need to provide much better *links* between the major FAnGR databases, which have largely been set up independently and are breed-focused (Groeneveld et al., 2010). The livestock genomics community needs either to build on an existing platform (such as the ARKDB, http://www.thearkdb.org/arkdb/ and the European Nucleotide Archive, http://www.ebi.ac.uk/ena), that have some level of connectivity, e.g., with Ensembl (http://www.ensembl.org/index.html) or to establish an independent community-based initiative(s) under the form of a user-friendly global web portal and would include web services able to federate resources and act as an *educational* central point. Such resources are already being developed, including the Adaptmap project for goats (http://www.goatadaptmap.org/). Information on livestock related data should be made available and useful recommendations are required to inform stakeholders on how to record data, and where to store what type of information. In particular, it is important to promote within the community of users that raw and meta-data are key components and that they should be made available in public datasets together with elaborated datasets. When there are existing public resources for a given datatype such as those listed above, they should be used for their ability to set standards and centralize data access. For other data types, open digital repositories such as Dryad (http://datadryad.org/), Zenodo (https://zenodo.org/), or figshare (http://figshare.com/) comprise invaluable tools acting as incentives for people to maintain and upgrade their datasets as data can be submitted and authors are provided with a reference which can be cited. This data ecosystem becomes especially important with the myriad of SNP array datasets that are now available and the incompatibility among different versions of these arrays within the same species (Nicolazzi et al., 2015). Moreover, to add value to genetic resources, federating gene bank resources is one step that needs to be completed by explicit connection—through geographical coordinates—with phenotypic data, but also with socio-economic, socio-demographic, climatic, environmental, and policy information. This requires links to existing online digital resources (Joost et al., 2010) that are currently rarely used by the FAnGR community and need to be listed on such a global portal.

Data Availability

While many genotyping projects on commercial livestock breeds are funded by industry, rendering all except summary data unavailable in many cases, in principle raw data from publicly funded projects should be made publicly available. Indeed, when data are open, it first makes the information more credible, makes data re-usable, and also enables reproducibility an important scientific principle (Ertz et al., 2014). Increasingly, international consortia, such as FAANG on animal functional genomics follow the Toronto protocol and immediately place data in the public domain (http://www.faang.org; The Toronto International Data Release Workshop Authors, 2009; Andersson et al., 2015). A next generation phenotyping database should also be established, including GIS and anonymized farm level data, animal photographs and meta-data—this could partly follow the format of the EU FP5 project Econogene (http://www.econogene.eu) and would be most efficiently linked with FAO's DAD-IS and EFABIS (http://efabis.tzv.fal.de). The ownership and hosting of such a resource would be logistically and financially challenging, and could provide an opportunity for the agri-industry to contribute toward conservation of the genetic resources it has utilized in the past and may need again in the future. This could also be part of the community-based action mentioned above, with many advantages (logistic and funding), but requiring a strong leadership. An approach to data resourcing such has been exemplified with human data by the 1000 Genomes project (http://www.1000genomes.org) and the 1001 Arabidopsis genomes resource (http://1001genomes.org with data being publicly available either immediately or after an agreed embargo period, could be very applicable to livestock studies. For example, the resequencing data from the EU Framework 7 Nextgen project was made available shortly after the project's completion at the European Bioinformatics Institute's FTP site (ftp://ftp.ebi.ac.uk/pub/databases/nextgen/).

Participatory Projects

Many individuals who are interested in FAnGR are involved in agriculture as smallholders, farmers, breeders, and producers

and many of these are not formally involved in breeding programs and livestock conservation, yet maintain an interest through agricultural shows and farmers' markets (e.g., Zimmerer, 2010; Johns et al., 2013). At the same time, the role of participatory approaches and mobile technology potentially enables robust data collection on a previously unimaginable scale (Lisson et al., 2010; Teacher et al., 2013; Sambo et al., 2015). Use of crowdsourcing should therefore be encouraged in FAnGR as should use of smart-phone apps and technologies for photography, data storage and sampling (e.g., "do-forms" http://www.doforms.com). A logical combination of these initiatives lies in the possibility of a livestock community independent initiative, including web services to federate these data sources, to carry out quality control and providing a central access point for data but also information to educate people on how to record FAnGR data. Such approaches could also help in securing funds for projects in FAnGR populations and breeds, which often face the problem of securing funds to carry out this necessary research.

Conservation, Management, and Prioritization

Is Prioritization a Priority?

A paradigm within FAnGR for the past 15 years concerns the use of genetic data, alongside other information in prioritization of livestock populations and breeds for conservation (Weitzman, 1992; Simianer et al., 2003; Boettcher et al., 2010; Ginja et al., 2013). However, there is limited evidence that this approach is being applied systematically across countries reporting to the FAO, although the second report on the State of the World's Animal Genetic Resources has documented activities to some extent (FAO, 2015b,c). If, however, prioritization methods are not being applied by managers and policy-makers, the question needs to be asked as to why? A number of explanations may pertain: first, the method(s) may have not gained enough traction with policy makers to ensure its/their implementation, which may indeed be because genomic methods, which have yet to be systematically implemented, will largely supersede the microsatellite-based approaches implemented thus far and enable conservation prioritization to include genes important in functionally valuable traits (e.g., Toro et al., 2014). Furthermore, prioritization on the basis of genetic distances (Weitzman, 1992) is confounded by genetic isolation of breeds (European Cattle Genetic Diversity Consortium, 2006). Second, prioritization may not actually be needed, at least in certain regions, where breed societies are active and all or most of the breeds can be maintained. However, recent animal health emergencies (e.g., outbreaks of transmissible spongiform encephalopathies, TSEs) have cast doubt on this simplistic scenario and required the application of careful genetic management during and after the outbreak. While prioritization may be less of a priority in the world's richest regions, it is not expected to be the case in developing countries, where extinction may take a number of forms, including genetic erosion (e.g., Berthouly-Salazar et al., 2012; FAO, 2015a,b). Finally, the methods developed may not have been applied because policy makers and managers are unaware of their availability, which could be due to a lack of dissemination or penetrance of educational material to the decision makers.

Utilization in Practice

While research and application of genomic tools in livestock is occurring in many commercial/transboundary breeds (e.g., Pryce et al., 2014; Scraggs et al., 2014), its application in less commercial populations is sporadic and the scientific basis of decisions on management of indigenous livestock, for example in which germplasm to store, assessing the effects of upgrading or evaluating ongoing genetic management is therefore highly variable (e.g., Brown et al., 2014; FAO, 2015b). This points to the reality that genetically-based prioritization is unlikely to be operational in the absence of other considerations, including commercial reality and the ecosystem/production environment (e.g., Sanderson et al., 2013). The use of genomic data to manage FAnGR *within* breeds is however, continuing apace (see above) and can be demonstrated to be assisting conservation, production and management in many cases (e.g., Herrero-Medrano et al., 2014; Scraggs et al., 2014). However, for many breeds the cost of genetic/genomic analysis vs. the potential economic returns on genotyped stock (with a few exceptions such as TSE resistance) makes its application uneconomic, and therefore it is often not applied. It is unlikely that genotyping costs will reach the level of economic viability for many FAnGR, however this assumption should be tested by some targeted research across the sector.

Defining Goals

The Convention on Biological Diversity's Aichi Target 13, which recommends that: "*strategies have been developed and implemented for minimizing genetic erosion and safeguarding genetic diversity*" is reflected in the Target for Strategic Priority Area 4 of the Global Plan of Action for Animal Genetic Resources (FAO, 2007b). These resource indicators contribute to the measurement of progress toward Aichi Target 13 (FAO, 2012b) and are calculated at national, regional and global levels, based on data entered by National Coordinators for the Management of Animal Genetic Resources[1] (172 countries had nominated a National Coordinator as of July, 2014) into the Domestic Animal Diversity Information System (DAD-IS). The following indicators have been agreed by the Commission on Genetic Resources for Food and Agriculture:

- the number of locally adapted breeds;
- the proportion of the total population accounted for by locally adapted and exotic breeds; and
- the number of breeds classified as at risk, not at risk and unknown.

The Global Databank for Animal Genetic Resources, the backbone of DAD-IS, enables National Coordinators to enter breed-specific data, including data on the size and structure of breed populations, required to calculate their risk status. FAO

[1]The list of National Coordinators for the Management of Animal Genetic Resources is found at dad.fao.org/cgi-bin/EfabisWeb.cgi?sid=-1,contacts.

produces biannual Status and Trends Reports (FAO, 2015a). For the first report on The State of the Worlds Animal Genetic Resources, a risk status classification based on population size data was used. The (lack of) availability of global data currently makes a more elaborate system involving, for example, molecular diversity indices, population structure/fragmentation, pedigree data, number and size of herds, and geographic distribution inoperable. While genomic methods might help to overcome these data deficiencies, if they are to be applied to livestock conservation, it is important to define the goals of such approaches and how the data could be used to improve or augment the current set of indicators using data that could be collected on trends in effective population size, admixture, inbreeding and genome-wide diversity. The wider application of such data hinges on their applicability to autochtonous, less-commercial breeds. Unfortunately, the data currently provided to FAO does not even allow the reliable calculation of basic trends currently measured via the above indicators (Tittensor et al., 2014; FAO, 2015a), yet the livestock genetics and conservation community possess many of the tools needed to directly evaluate whether signatories to the CBD are "…*minimizing genetic erosion*" and "*safeguarding genetic diversity*" (CBD Target 13). Two key developments are required to enable the current approach to more directly use genetic or genomic data in the future: first, the livestock conservation genetics community must therefore insist that data are collected and analyzed in such a way that results are directly comparable and second, to help develop better indicators applied to monitoring genetic trends in domestic populations.

CONCLUSION

Any exercise designed to assess the state-of-the-art in a scientific field only manages to capture a brief moment in time, which is why the Horizon scanning exercises carried out in biodiversity conservation are repeated every year (see Sutherland et al., 2015). Here, we attempted to take a longer-term (decadal) view of genomic resources conservation, and during this period, some major milestones will be passed. Chief among these is the imminent release of the Second Report on the State of the World's Animal Genetic Resources (FAO, 2015b,d) and the Convention on Biological Diversity's 2020 deadline halting the loss of biodiversity Aichi targets. In the context of the dramatic advances in 'omics technology that are expected during the next decade, the field is expected to move fast. But structural changes in the livestock sector that will bring further erosion during this period are likely to be equally rapid. However, this makes it critically important that a strategic approach is taken to incorporating these technological advances into real world FAnGR conservation. Such an approach has been taken in the past (e.g., with the implementation of approved microsatellite marker sets) and, we would argue, is needed now to ensure that practical conservation of farm animal agricultural biodiversity is not left behind. The FAnGR community therefore needs to make best use of new genomic tools, and at the same time continue and augment its classical phenotyping efforts. Both, genomic and phenotypic tools need to be applied more consistently, at a much wider scale and for more breeds, to describe, utilize and conserve the world's genomic/breed diversity for future generations.

ACKNOWLEDGMENTS

The European Science Foundation (ESF) GENOMIC—RESOURCES Research Networking Programme (RNP) was supported by: Fonds zur Förderung der wissenschaftlichen Forschung (FWF), FWF Austrian Science Fund, Austria—Fonds National de la Recherche Scientifique (FNRS), Belgium—Fonds voor Wetenschappelijk Onderzoek—Vlaanderen (FWO), The Research Foundation—Flanders, Belgium—Nacionalna zaklada za znanost, visoko školstvo i tehnologijski razvoj Republike Hrvatske, Croatian Science Foundation, Republic of Croatia—Suomen Akatemia, Biotieteiden ja ympäristön tutkimuksen toimikunta, Academy of Finland, Research Council for Biosciences and Environment, Finland—Deutsche Forschungsgemeinschaft (DFG), German Research Foundation, Germany—Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO), The Netherlands Organisation for Scientific Research, The Netherlands—Norges Forskningsråd, The Research Council of Norway, Norway—Forskningsrådet för miljö, areella näringar och samhällsbyggande, Swedish Council for Environment, Agricultural Sciences and Spatial Planning (FORMAS), Sweden—Schweizerischer Nationalfonds (SNF), Swiss National Science Foundation, Switzerland—Biotechnology and Biological Sciences Research Council (BBSRC), United Kingdom.

REFERENCES

Adewopo, J. B., VanZomeren, C., Bhomia, R. K., Almaraz, M., Bacon, A., Eggleston, E., et al. (2014). Top-ranked priority research questions for soil science in the 21st century. *Soil Sci. Soc. Am. J.* 78, 337–347. doi: 10.2136/sssaj2013.07.0291

Ajmone-Marsan, P., Colli, L., Han, J., Achilli, A., Lancioni, H., Joost, S., et al. (2014). The characterization of goat genetic diversity: towards a genomic approach. *Small Ruminant. Res.* 121, 58–72. doi: 10.1016/j.smallrumres.2014.06.010

Andersson, L., Archibald, A. L., Bottema, C. D., Brauning, R., Burgess, S. C., Burt, D. W., et al. (2015). Coordinated international action to accelerate genome-to-phenome with FAANG, the Functional Annotation of Animal Genomes project. *Genome Biol.* 16, 57. doi: 10.1186/s13059-015-0622-4

Arbuckle, B. S., Kansa, S. W., Kansa, E., Orton, D., Çakirlar, C., Gourichon, L., et al. (2014). Data sharing reveals complexity in the westward spread of domestic animals across Neolithic Turkey. *PLoS ONE* 9:e99845. doi: 10.1371/journal.pone.0099845

Barbato, M., Orozco-terWengel, P., Tapio, M., and Bruford, M. W. (2015). SNeP: a tool to estimate trends in recent effective population size trajectories using genome-wide SNP data. *Front. Genet.* 6:109. doi: 10.3389/fgene.2015.00109

Berthouly-Salazar, C., Thévenon, S., Van, T. N., Nguyen, B. T., Pham, L. D., Chi, C. V., et al. (2012). Uncontrolled admixture and loss of genetic diversity in a local Vietnamese pig breed. *Ecol. Evol.* 2, 962–975. doi: 10.1002/ece3.229

Boettcher, P. J., Tixier-Boichard, M., Toro, M. A., Simianer, H., Eding, H., Gandini, G., et al. (2010). Objectives, criteria and methods for using molecular genetic

data in priority setting for conservation of animal genetic resources. *Anim. Genet.* 41, 64–77. doi: 10.1111/j.1365-2052.2010.02050.x

Bosse, M., Megens, H. J., Frantz, L. A., Madsen, O., Larson, G., Paudel, Y., et al. (2014b). Genomic analysis reveals selection for Asian genes in European pigs following human-mediated introgression. *Nat. Commun.* 5, 4392. doi: 10.1038/ncomms5392

Bosse, M., Megens, H. J., Madsen, O., Frantz, L. A., Paudel, Y., Crooijmans, R. P., et al. (2014a). Untangling the hybrid nature of modern pig genomes: a mosaic derived from biogeographically distinct and highly divergent *Sus scrofa* populations. *Mol. Ecol.* 23, 4089–4102. doi: 10.1111/mec.12807

Bosse, M., Megens, H. J., Madsen, O., Paudel, Y., Frantz, L. A., Schook, L. B., et al. (2012). Regions of homozygosity in the porcine genome: consequence of demography and the recombination landscape. *PLoS Genet.* 8:e1003100. doi: 10.1371/journal.pgen.1003100

Brown, C., Orford, M., Tzamaloukas, O., Mavrogenis, A. P., and Miltiadou, D. (2014). Assessment of inbreeding resulting from selection for scrapie resistance: a model for rare sheep breeds. *Vet. Rec.* 175, 624. doi: 10.1136/vr.102503

Cardellino, R. A., and Boyazoglu, J. (2009). Research opportunities in the field of animal genetic resources. *Livestock Sci.* 120, 166–173. doi: 10.1016/j.livsci.2008.07.002

Chang, T. C., Yang, Y., Retzel, E. F., and Liu, W. S. (2013). Male-specific region of the bovine Y chromosome is gene rich with a high transcriptomic activity in testis development. *Proc. Natl. Acad. Sci. U.S.A.* 110, 12373–12378. doi: 10.1073/pnas.1221104110

Curik, I., Ferencakovic, M., and Soelkner, J. (2014). Inbreeding and runs of homozygosity: a possible solution to an old problem. *Livestock Sci.* 166, 26–34. doi: 10.1016/j.livsci.2014.05.034

De Donato, M., Peters, S. O., Mitchell, S. E., Hussain, T., and Imumorin, I. G. (2013). Genotyping-by-sequencing (GBS): a novel, efficient and cost-effective genotyping method for cattle using next-generation sequencing. *PLoS ONE* 8:e62317. doi: 10.1371/journal.pone.0062137

Dicks, L. V., Abrahams, A., Atkinson, J., Biesmeijer, J., Bourn, N., Brown, C., et al. (2013). Identifying key knowledge needs for evidence-based conservation of wild insect pollinators: a collaborative cross-sectoral exercise. *Insect Conserv. Divers.* 6, 435–446. doi: 10.1111/j.1752-4598.2012.00221.x

Edwards, C. J., Ginja, C., Kantanen, J., Perez-Pardal, L., Tresset, A., Stock, F., et al. (2011). Dual origins of dairy cattle farming - evidence from a comprehensive survey of European Y-chromosomal variation. *PLoS ONE* 6:e15922. doi: 10.1371/journal.pone.0015922

Ertz, O., Rey, S. J., and Joost, S. (2014). The open source dynamics in geospatial research and education. *J. Spatial Inf. Sci.* 8, 67–71. doi: 10.5311/josis.2014.8.182

European Cattle Genetic Diversity Consortium. (2006). Marker-assisted conservation of European cattle breeds: an evaluation. *Anim. Genet.* 37, 475–481. doi: 10.1111/j.1365-2052.2006.01511.x

FAO (2007a). *The Global Plan of Action for Animal Genetic Resources and the Interlaken Declaration*. Rome. Available online at: www.fao.org/docrep/010/a1404e/a1404e00.htm

FAO (2007b). *The State of the World's Animal Genetic Resources for Food and Agriculture*, eds B. Rischkowsky and D. Pilling. Rome: Food and Agriculture Organization of the United Nations, 2007. 511. Available online at: www.fao.org/3/a-a1250e.pdf

FAO (2009). *Report of the FAO/WAAP Workshop on Production Environment Descriptors for Animal Genetic Resources*. Available online at: dad.fao.org/cgi-bin/getblob.cgi?sid=-1,593

FAO (2011). *Molecular Genetic Characterization of Animal Genetic Resources*. FAO Animal Production and Health Guidelines. No. 9. Rome.

FAO (2012a). *Phenotypic Characterization of Animal Genetic Resources*. FAO Animal Production and Health Guidelines No. 11. Rome.

FAO (2012b). "Targets and indicators for animal genetic resource," in *Seventh Session of the Intergovernmental Technical Working Group on Animal Genetic Resources for Food And Agriculture, (CGRFA/WG-AnGR-7/12/7)* (Rome). Available online at: www.fao.org/docrep/meeting/026/me514e.pdf

FAO (2015a). "Status and trends of animal genetic resources – 2014," in *Fifteenth Regular Session of the Commission on Genetic Resources for Food and Agriculture, (CGRFA-15/15/Inf.18)* (Rome). Available online at: www.fao.org/3/a-mm278e.pdf

FAO (2015b). *Draft Second Report on the State of the World's Animal Genetic Resources for Food and Agriculture (Part 3), Commission on Genetic Resources for Food and Agriculture, CGRFA-15/15/Inf.17.2*. Available online at: www.fao.org/3/a-mm310e.pdf

FAO (2015c). "Synthesis progress report on the implementation of the Global Plan of Action for Animal Genetic Resources – 2014," in *Fifteenth Regular Session of the Commission on Genetic Resources for Food and Agriculture, (CGRFA-15/15/Inf.19)* (Rome). Available online at: www.fao.org/3/a-mm282e.pdf

FAO (2015d). *Draft Second Report on the State of the World's Animal Genetic Resources for Food and Agriculture (Part 1 and 2), Commission on Genetic Resources for Food and Agriculture, CGRFA-15/15/Inf.17.1*. Available online at: www.fao.org/3/a-mm313e.pdf

Ferrando, A., Goyache, F., Pares, P.-M., Carrion, C., Miro, J., and Jordana, J. (2014). Genetic relationships between six eastern Pyrenean sheep breeds assessed using microsatellites. *Spanish J. Agric. Res.* 12, 1029–1037. doi: 10.5424/sjar/2014124-6173

Frantz, L. A. F., Madsen, O., Megens, H.-J., Schraiber, J. G., Paudel, Y., Bosse, M., et al. (2015). Evolution of Tibetan wild boars. *Nat. Genet.* 47, 188–189. doi: 10.1038/ng.3197

Giguet-Covex, C., Pansu, J., Arnaud, F., Rey, P.-J., Griggo, C., Gielly, L., et al. (2014). Long livestock farming history and human landscape shaping revealed by lake sediment DNA. *Nat. Commun.* 5, 3211. doi: 10.1038/ncomms4211

Ginja, C., Gama, L. T., Cortes, O., Vicente, D. J., Dunner, S., et al. (2013). Analysis of conservation priorities of Iberoamerican cattle based on autosomal microsatellite markers. *Genet. Sel. Evol.* 45:35. doi: 10.1186/1297-9686-45-35

Ginja, C., Penedo, M. C. T., Melucci, L., Quiroz, J., Martínez López, O. R., Revidatti, M. A., et al. (2010). Origins and genetic diversity of New World Creole cattle: inferences from mitochondrial and Y chromosome polymorphisms. *Anim. Genet.* 41, 128–141. doi: 10.1111/j.1365-2052.2009.01976.x

Goddard, M. T., and Hayes, B. J. (2009). Mapping genes for complex traits in domestic animals and their use in breeding programs. *Nat. Rev. Genet.* 10, 381–391. doi: 10.1038/nrg2575

Götherström, A., Anderung, C., Hellborg, L., Elburg, R., Smith, C., Bradley, D. G., et al. (2005). Cattle domestication in the Near East was followed by hybridization with aurochs bulls in Europe. *Proc. R. Soc. Lond. B* 272, 2345–2350. doi: 10.1098/rspb.2005.3243

Groeneveld, L. F., Lenstra, J. A., Eding, H., Toro, M. A., Scherf, B., Pilling, D., et al. (2010). Genetic diversity in livestock breeds. *Anim. Genet.* 41(Suppl. 1), 6–31. doi: 10.1111/j.1365-2052.2010.02038.x

Han, R., Yang, P., Tian, Y., Wang, D., Zhang, Z., Wang, L., et al. (2014). Identification and functional characterization of copy number variation in diverse chicken breeds. *BMC Genomics* 15:934. doi: 10.1186/1471-2164-15-934

Herrero-Medrano, J. M., Megens, H. J., Crooijmans, R. P., Abellaneda, J. M., and Ramis, G. (2013). Farm-by-farm analysis of microsatellite, mtDNA and SNP genotype data reveals inbreeding and crossbreeding as threats to the survival of a native Spanish pig breed. *Anim. Genet.* 44, 259–266. doi: 10.1111/age.12001

Herrero-Medrano, J.-M., Megens, H.-J., Groenen, M. A. M., Boss, M., Pérez-Enciso, M., and Crooijmans, R. P. M. A. (2014). Whole-genome sequence analysis reveals differences in population management and selection of European low-input pig breeds. *BMC Genomics* 15:601. doi: 10.1186/1471-2164-15-601

Hoffmann, I. (2010). Climate change and the characterization, breeding and conservation of animal genetic resources. *Anim. Genet.* 41(Suppl. 1), 32–46. doi: 10.1111/j.1365-2052.2010.02043.x

Huang, W., Kirkpatrick, B. W., Rosa, G. J. M., and Khatib, H. (2010). A genome-wide association study using selective DNA pooling identifies candidate markers for fertility in Holstein cattle. *Anim. Genet.* 41, 570–578. doi: 10.1111/j.1365-2052.2010.02046.x

Ingram, J. S., Wright, H. L., Foster, L., Aldred, T., Barling, D., Benton, T. G., et al. (2013). Priority research questions for the UK food system. *Food Secur.* 5, 617–636. doi: 10.1007/s12571-013-0294-4

Jammes, H., and Renard, J.-P. (2010). "Epigénétique et construction du phénotype, un enjeu pour les productions animales? (Epigenetics and phenotype construction, a challenge for livestock production?)," in *Robustesse, Rusticité, Flexibilité, Plasticité, Résilience... les Nouveaux Critères de Qualité des Animaux et des Systèmes D'élevage*, eds D. Sauvant and J. M. Perez (Paris: Institut National de la Recherche Agronomique), 23–42.

Johns, T., Powell, B., Maundu, P., and Eyzaguirre, P. B. (2013). Agricultural biodiversity as a link between traditional food systems and contemporary development, social integrity and ecological health. *J. Sci. Food Agric.* 93, 3433–3442. doi: 10.1002/jsfa.6351

Jones, M. R., Forester, B. R., Teufel, A. I., Adams, R. V., Anstett, D. N., Goodrich, B. A., et al. (2013). Integrating landscape genomics and spatially explicit explicit approaches to detect loci under selection in clinal populations. *Evolution* 67, 3455–3468. doi: 10.1111/evo.12237

Joost, S., Colli, L., Baret, P. V., Garcia, J. F., Boettcher, P. J., Tixier-Boichard, M., et al. (2010). Integrating geo-referenced multiscale and multidisciplinary data for the management of biodiversity in livestock genetic resources. *Anim. Genet.* 41, 47–63. doi: 10.1111/j.1365-2052.2010.02037.x

Kadarmideen, H. N. (2014). Genomics to systems biology in animal and veterinary sciences: progress, lessons and opportunities. *Livestock Sci.* 166, 232–246. doi: 10.1016/j.livsci.2014.04.028

Kijas, J. W., Ortiz, J. S., McCulloch, R., James, A., Brice, B., Swain, B., et al. (2013). Genetic diversity and investigation of polledness in divergent goat populations using 52088 SNPs. *Anim. Genet.* 44, 325–335. doi: 10.1111/age.12011

Kristensen, T. N., Hoffmann, A. A., Pertoldi, C., and Stronen, A. V. (2015). What can livestock breeders learn from conservation genetics and vice versa? *Front. Genet.* 6:38. doi: 10.3389/fgene.2015.00038

Larson, G., Lui, R., Zhao, X., Yuan, J., Fuller, D., Barton, L., et al. (2010). Patterns of East Asian pig domestication, migration, and turnover revealed by modern and ancient DNA. *Proc. Natl. Acad. Sci. U.S.A.* 107, 7686–7691. doi: 10.1073/pnas.0912264107

Lee, K. T., Chung, W. H., Lee, S. Y., Choi, J. W., Kim, J., Lim, D., et al. (2013). Whole-genome resequencing of Hanwoo (Korean cattle) and insight into regions of homozygosity. *BMC Genomics* 14:519. doi: 10.1186/1471-2164-14-519

Lenstra, J. A., Groeneveld, L. F., Eding, H., Kantanen, J., Williams, J. L., Taberlet, P., et al. (2012). Molecular tools and analytical approaches for the characterization of farm animal genetic diversity. *Anim. Genet.* 43, 483–502. doi: 10.1111/j.1365-2052.2011.02309.x

Li, H., and Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature* 475, 493–496. doi: 10.1038/nature10231

Lisson, S., MacLeod, N., McDonald, C., Corfield, J., Pengelly, B., Wirajaswadi, L., et al. (2010). A participatory, farming systems approach to improving Bali cattle production in the smallholder crop–livestock systems of Eastern Indonesia. *Agric. Syst.* 103, 486–497. doi: 10.1016/j.agsy.2010.05.002

Magee, D. A., Sikora, K. M., Berkowicz, E. W., Berry, D. P., Howard, D. J., Mullen, M. P., et al. (2011). DNA sequence polymorphisms in a panel of eight candidate bovine imprinted genes and their association with performance traits in Irish Holstein-Friesian cattle. *BMC Genet.* 11:93. doi: 10.1186/1471-2156-11-93

Magee, D. A., Spillane, C., Berkowicz, E. W., Sikora, K. M., and MacHugh, D. E. (2014). Imprinted loci in domestic livestock species as epigenomic targets for artificial selection of complex traits. *Anim. Genet.* 45(Suppl. 1), 25–39. doi: 10.1111/age.12168

Matukumalli, L. K., Lawley, C. T., Schnabel, R. D., Taylor, J. F., Allan, M. F., Heaton, M. P., et al. (2009). Development and characterization of a high density SNP genotyping assay for cattle. *PLoS ONE* 4:e5350. doi: 10.1371/journal.pone.0005350

Meadows, J. R. S., and Kijas, J. W. (2008). Re-sequencing regions of the ovine Y chromosome in domestic and wild sheep reveals novel paternal haplotypes. *Anim. Genet.* 40, 119–123. doi: 10.1111/j.1365-2052.2008.01799.x

Mokry, F. B., Buzanskas, M. E., de Alvarenga Mudadu, M., do Amaral Grossi, D., Higa, R. H., Ventura, R. V., et al. (2014). Linkage disequilibrium and haplotype block structure in a composite beef cattle breed. *BMC Genomics* 15:S6. doi: 10.1186/1471-2164-15-S7-S6

Nicolazzi, E. L., Caprera, A., Nazzicari, N., Cozzi, P., Strozzi, F., Lawley, C., et al. (2015). SNPchiMp v.3: integrating and standardizing single nucleotide polymorphism data for livestock species. *BMC Genomics* 16:283. doi: 10.1186/s12864-015-1497-1

Olson-Manning, C. F., Wagner, M. R., and Mitchell-Olds, T. (2012). Adaptive evolution: evaluating empirical support for theoretical predictions. *Nat. Rev. Genet.* 13, 867–877. doi: 10.1038/nrg3322

Orlando, L., Ginolhac, A., Zhang, G., Froese, D., Albrechtsen, A., Stiller, M., et al. (2013). Recalibrating *Equus* evolution using the genome sequence of an early Middle Pleistocene horse. *Nature* 499, 74–78. doi: 10.1038/nature12323

Orozco-terWengel, P. A., and Bruford, M. W. (2014). Mixed signals from hybrid genomes. *Mol. Ecol.* 23, 3941–3943. doi: 10.1111/mec.12863

Patel, C. J., Bhattacharya, J., and Butte, A. J. (2010). An environment-wide association study (EWAS) on type 2 diabetes mellitus. *PLoS ONE* 5:e10746. doi: 10.1371/journal.pone.0010746

Pereira, F., Queirós, S., Gusmão, L., Nijman, I. J., Cuppen, E., Lenstra, J. A., et al. (2009). Tracing the history of goat pastoralism: new clues from mitochondrial and Y chromosome DNA in North Africa. *Mol. Biol. Evol.* 26, 2765–2773. doi: 10.1093/molbev/msp200

Porto Neto, L. R., and Barendse, W. (2010). Effect of SNP origin on analyses of genetic diversity in cattle. *Anim. Prod. Sci.* 50, 792–800. doi: 10.1071/AN10073

Pretty, J., Sutherland, W. J., Ashby, J., Auburn, J., Baulcombe, D., Bell, M., et al. (2010). Top 100 questions of importance to the future of global agriculture. *Int. J. Agric. Sustainability* 8, 219–238. doi: 10.3763/ijas.2010.0534

Pryce, J. E., Haile-Mariam, M., Goddard, M. E., and Hayes, B. J. (2014). Identification of genomic regions associated with inbreeding depression in Holstein and Jersey dairy cattle. *Genet. Sel. Evol.* 46:71. doi: 10.1186/s12711-014-0071-7

Purfield, D. C., Berry, D. P., McParland, S., and Bradley, D. G. (2012). Runs of homozygosity and population history in cattle. *BMC Genet.* 13:70. doi: 10.1186/1471-2156-13-70

Qiu, Q., Zhang, G., Ma, T., Qian, W., Wang, J., Ye, Z., et al. (2012). The yak genome and adaptation to life at high altitude. *Nat. Genet.* 44, 946. doi: 10.1038/ng.2343

Ramírez, O., Ojeda, A., Tomàs, A., Gallardo, D., Huang, L. S., Folch, J. M., et al. (2009). Integrating Y-chromosome, mitochondrial, and autosomal data to analyze the origin of pig breeds. *Mol. Biol. Evol.* 26, 2061–2072. doi: 10.1093/molbev/msp118

Sambo, E., Bettridge, J., Dessie, T., Amare, A., Habte, T., Wigley, P., et al. (2015). Participatory evaluation of chicken health and production constraints in Ethiopia. *Prevent. Vet. Med.* 118, 117–127. doi: 10.1016/j.prevetmed.2014.10.014

Sanderson, M. A., Archer, D., Hendrickson, J., Kronberg, S., Liebig, M., Nichols, K., et al. (2013). Diversification and ecosystem services for conservation agriculture: outcomes from pastures and integrated crop-livestock systems. *Renew. Agric. Food Syst.* 26, 129–144. doi: 10.1017/S1742170512000312

Schiffels, S., and Durbin, R. (2014). Inferring human population size and separation history from multiple genome sequences. *Nat. Genet.* 46, 919–925. doi: 10.1101/005348

Scraggs, E., Zanella, R., Wojtowicz, A., Taylor, J. F., Gaskins, C. T., Reeves, J. J., et al. (2014). Estimation of inbreeding and effective population size of full-blood wagyu cattle registered with the American Wagyu Cattle Association. *J. Anim. Breed. Genet.* 131, 3–10. doi: 10.1111/jbg.12066

Simianer, H., Mart, S. B., Gibson, J., Hanotte, O., and Rege, J. E. O. (2003). An approach to the optimal allocation of conservation funds to minimize loss of genetic diversity between livestock breeds. *Ecol. Econ.* 45, 377–392. doi: 10.1016/S0921-8009(03)00092-2

Sutherland, W. J., Aveling, R., Brooks, T. M., Clout, M., Dicks, L. V., Fellman, L., et al. (2014). A horizon scan of global conservation issues for 2014. *Trends Ecol. Evol.* 29, 1. doi: 10.1016/j.tree.2013.11.004x

Sutherland, W. J., Clout, M., Depledge, M., Dicks, L. V., Dinsdale, J., Entwistle, A. C., et al. (2015). A horizon scan of global conservation issues for 2015. *Trends Ecol. Evol.* 30, 1. doi: 10.1016/j.tree.2014.11.002

Sutherland, W. J., and Woodroof, H. J. (2009). The need for environmental horizon scanning. *Trends Ecol. Evol.* 24, 523–527. doi: 10.1016/j.tree.2009.04.008

Svensson, E., and Götherström, A. (2008). Temporal fluctuations of Y-chromosomal variation in *Bos taurus*. *Biol. Lett.* 4, 752–754. doi: 10.1098/rsbl.2008.0342

Teacher, A. G. F., Griffiths, D. J., Hodgson, D. J., and Inger, R. (2013). Smartphones in ecology and evolution: a guide for the app-rehensive. *Ecol. Evol.* 3, 5268–5278. doi: 10.1002/ece3.888

Teasdale, M. D., van Doorn, N. L., Fiddyment, S., Webb, C. C., O'Connor, T., Hofreiter, M., et al. (2015). Paging through history: parchment as a reservoir of ancient DNA for next generation sequencing. *Philos. Trans. R. Soc. Lond. B* 370:20130379. doi: 10.1098/rstb.2013.0379

The Toronto International Data Release Workshop Authors. (2009). Prepublication data sharing. *Nature* 461, 168–170. doi: 10.1038/461168a

Tittensor, D. P., Walpole, M., Hill, S. L. L., Boyce, D. G., Britten, G. L., Burgess, N. D., et al. (2014). A mid-term analysis of progress toward international biodiversity targets. *Science* 346, 241–244. doi: 10.1126/science.1257484

Toro, M. A., Villanueva, B., and Fernandez, J. (2014). Genomics applied to management strategies in conservation programmes. *Livestock Sci.* 166, 48–53. doi: 10.1016/j.livsci.2014.04.020

Wallace, D. C. (2005). A mitochondrial paradigm of metabolic and degenerative disease, aging and cancer: a dawn for evolutionary medicine. *Ann. Rev. Genet.* 39, 359–407. doi: 10.1146/annurev.genet.39.110304.095751

Wallner, B., Vogl, C., Shukla, P., Burgstaller, J. P., Druml, T., and Brem, G. (2013). Identification of genetic variation on the Horse Y chromosome and the tracing of male founder lineages in modern breeds. *PLoS ONE* 8:e60015. doi: 10.1371/journal.pone.0060015

Weitzman, M. L. (1992). On diversity. *Q. J. Econ.* CVII, 363–405. doi: 10.2307/2118476

Wu, H., Guang, X., Al-Fageeh, M. B., Cao, J., Pan, S., Zhou, H., et al. (2014). Camelid genomes reveal evolution and adaptation to desert environments. *Nat. Commun.* 5, 5188. doi: 10.1038/ncomms6188

Yue, X. P., Dechow, C., Chang, T. C., DeJarnette, J. M., Marshall, C. E., Lei, C. Z., et al. (2014). Copy number variations of the extensively amplified Y-linked genes, HSFY and ZNF280BY, in cattle and their association with male reproductive traits in Holstein bulls. *BMC Genomics* 15:113. doi: 10.1186/1471-2164-15-113

Zhang, Q., Calus, M. P. L., Guldbrandtsen, B., Lund, M. S., and Sahana, G. (2015). Estimation of inbreeding using pedigree, 50k SNP chip genotypes and full sequence data in three cattle breeds. *BMC Genetics* 16:88. doi: 10.1186/s12863-015-0227-7

Zimmerer, K. (2010). Biological diversity in agriculture and global change. *Annu. Rev. Environ. Resour.* 35, 137–166. doi: 10.1146/annurev-environ-040309-113840

Disclaimer: The views expressed in this information product are those of the author and do not necessarily reflect the views or policies of FAO.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Citation: Bruford MW, Ginja C, Hoffmann I, Joost S, Orozco-terWengel P, Alberto FJ, Amaral AJ, Barbato M, Biscarini F, Colli L, Costa M, Curik I, Duruz S, Ferenčaković M, Fischer D, Fitak R, Groeneveld LF, Hall SJG, Hanotte O, Hassan F, Helsen P, Iacolina L, Kantanen J, Leempoel K, Lenstra JA, Ajmone-Marsan P, Masembe C, Megens H-J, Miele M, Neuditschko M, Nicolazzi EL, Pompanon F, Roosen J, Sevane N, Smetko A, Štambuk A, Streeter I, Stucki S, Supakorn C, Telo Da Gama L, Tixier-Boichard M, Wegmann D and Zhan X (2015) Prospects and challenges for the conservation of farm animal genomic resources, 2015-2025. Front. Genet. 6:314. doi: 10.3389/fgene.2015.00314

Copyright © 2015 Bruford, Ginja, Hoffmann, Joost, Orozco-terWengel, Alberto, Amaral, Barbato, Biscarini, Colli, Costa, Curik, Duruz, Ferenčaković, Fischer, Fitak, Groeneveld, Hall, Hanotte, Hassan, Helsen, Iacolina, Kantanen, Leempoel, Lenstra, Ajmone-Marsan, Masembe, Megens, Miele, Neuditschko, Nicolazzi, Pompanon, Roosen, Sevane, Smetko, Štambuk, Streeter, Stucki, Supakorn, Telo Da Gama, Tixier-Boichard, Wegmann and Zhan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in
GENETICS

REVIEW ARTICLE
published: 10 February 2015
doi: 10.3389/fgene.2015.00038

What can livestock breeders learn from conservation genetics and *vice versa*?

Torsten N. Kristensen[1]*, Ary A. Hoffmann[2], Cino Pertoldi[1,3] and Astrid V. Stronen[1]

[1] Section of Biology and Environmental Science, Department of Chemistry and Bioscience, Aalborg University, Aalborg, Denmark
[2] Department of Zoology and Department of Genetics, Bio21 Institute, The University of Melbourne, Melbourne, VIC, Australia
[3] Aalborg Zoo, Aalborg, Denmark

Edited by:
Juha Kantanen, Natural Resources Institute Finland, Finland

Reviewed by:
Evangelina López De Maturana, Centro Nacional de Investigaciones Oncológicas, Spain
Miika Tapio, MTT Agrifood Research Finland, Finland

***Correspondence:**
Torsten N. Kristensen, Section of Biology and Environmental Science, Department of Chemistry and Bioscience, Aalborg University, Fredrik Bajers Vej 7H, DK-9220 Aalborg East, Denmark
e-mail: tnk@bio.aau.dk

The management of livestock breeds and threatened natural population share common challenges, including small effective population sizes, high risk of inbreeding, and the potential benefits and costs associated with mixing disparate gene pools. Here, we consider what has been learnt about these issues, the ways in which the knowledge gained from one area might be applied to the other, and the potential of genomics to provide new insights. Although there are key differences stemming from the importance of artificial versus natural selection and the decreased level of environmental heterogeneity experienced by many livestock populations, we suspect that information from genetic rescue in natural populations could be usefully applied to livestock. This includes an increased emphasis on maintaining substantial population sizes at the expense of genetic uniqueness in ensuring future adaptability, and on emphasizing the way that environmental changes can influence the relative fitness of deleterious alleles and genotypes in small populations. We also suspect that information gained from cross-breeding and the maintenance of unique breeds will be increasingly important for the preservation of genetic variation in small natural populations. In particular, selected genes identified in domestic populations provide genetic markers for exploring adaptive evolution in threatened natural populations. Genomic technologies in the two disciplines will be important in the future in realizing genetic gains in livestock and maximizing adaptive capacity in wildlife, and particularly in understanding how parts of the genome may respond differently when exposed to population processes and selection.

Keywords: effective population size, inbreeding, indigenous breeds, genetic rescue, genomics, selection

INTRODUCTION

The effective population size, Ne, is a measure of fundamental importance for understanding the potential of species and populations to evolve and adapt to natural and artificial selection pressures. Quantitative genetic theory predict that Ne is positively associated with the level of additive genetic variation and that the capacity of a population to respond to selection depends on the level of genetic variation for the trait(s) undergoing selection (see Falconer and Mackay, 1996). However, the association between Ne, genetic variation and evolutionary potential is complex and depends on factors such as the number of loci underlying a trait, the presence of dominance or epistasis, the effects of new mutations, and selection mode and intensity (reviewed in Willi et al., 2006).

In natural populations, there are numerous examples showing that a small Ne can reduce adaptive potential as a consequence of reduced genetic variation (e.g., Markert et al., 2010; Siol et al., 2010; Strasburg et al., 2011; Gossmann et al., 2012; Phifer-Rixey et al., 2012) and also as a consequence of a reduction in fitness due to inbreeding depression (Mattila et al., 2012; Hoffman et al., 2014). However, some small populations remain capable of adapting through evolutionary changes to shifting environmental conditions (references in Merilä, 2014).

Populations of domestic animals with a small Ne can also exhibit reduced genetic variation compared to ancestral ones (Ghafouri-Kesbi et al., 2008; Kim et al., 2013; Freedman et al., 2014; Quaresma et al., 2014). For example Freedman et al. (2014) found that the domestication process of dogs resulted in reduced genomic variation consistent with at least a 16-fold reduction in population size. On a shorter time scale Kim et al. (2013) showed that selection in US Holstein dairy cattle during the last 50 years has led to increased autozygosity across the genome. Inbreeding depression due to low Ne occur in domestic animal populations as well as in natural populations, decreasing milk yield and fat and protein content in the milk in dairy cattle and growth rates in sheep (Croquet et al., 2006; Pedrosa et al., 2010) and increasing the incidence of diseases such as mastitis in dairy cattle (Croquet et al., 2006; Sørensen et al., 2006).

However, a small Ne in domesticated animal populations helps produce phenotypic uniformity within populations (and even more so in domesticated crops) that results in products that can be more easily processed and marketed (Allard, 1999; Aslam et al., 2012; Janhunen et al., 2013). Intense directional artificial selection (and strong selection responses) in livestock is partly attained by use of a few selected animals, with superior genetic profiles, making the genetic contribution of animals in a population highly

skewed. This is one reason why Ne below 100 is observed within many intensively managed modern breeds (Leroy et al., 2013). Despite low Ne, large and ongoing genetic gains for production traits are typically achieved in commercial livestock (Hill and Kirkpatrick, 2010).

Although the implications of small Ne in natural populations and domestic breeds are somewhat different, there are insights to be gained from combining knowledge of these disparate areas. Genetic studies on small and fragmented populations in nature, such as island populations or populations at the brink of extinction, provide ideas and concepts that could be applied to the management of small domestic populations. This includes applying genetic rescue to boost the adaptive potential of small populations, and using genotype by environment interactions to ensure populations maintain a high fitness across environments (Ingvarsson, 2001; Vilá et al., 2003; Tallmon et al., 2004; Armbruster and Reed, 2005; Edmands, 2007; Adams et al., 2011; Reed et al., 2012). On the other hand, animal and plant breeders have shown how deep pedigrees, detailed phenotypic information, large sample sizes and reproductive technologies can be effectively used to meet genetic challenges in small populations. Information from pedigrees is already being applied to natural populations of birds and mammals where individuals can be tracked and followed across generations (e.g., Reid et al., 2006; Nielsen et al., 2012; Hedrick et al., 2014), and there is also potential to apply other approaches from livestock management to small and threatened natural populations.

While Ne is important from a genetic perspective, the census size of domestic and natural populations is obviously also important for predicting extinction risk. Populations at a small census size are more likely to go extinct due to factors such as demographic and environmental stochasticity even if genetic considerations are not taken into account. However, in this paper we focus only on the consequences of low Ne, providing examples of genetic rescue in natural populations, and considering prospects of using genomics as a tool that could advance long-term conservation management in both natural and domestic populations.

CONSEQUENCES OF LOW Ne – INBREEDING AND EVOLUTIONARY POTENTIAL

Most domestic breeds are small with Ne typically counted in tens or a few 100s and some of them are threatened by extinction (Hill and Kirkpatrick, 2010; Leroy et al., 2013; **Figure 1**). Likewise a large and increasing number of wild populations are threatened and also have low Ne (references in Frankham et al., 1998; IUCN, 2014). Populations with small Ne are prone to inbreeding and loss of genetic variation due to genetic drift (Falconer and Mackay, 1996; Willi et al., 2006).

INBREEDING

Mating of related individuals is unavoidable in populations of finite sizes but occurs also when there is preferential mating of relatives in large populations. Inbreeding leads to a reduced number of heterozygotes and ultimately to complete homozygosity in the genome (Falconer and Mackay, 1996). Inbred individuals typically have lower fitness (inbreeding depression) although

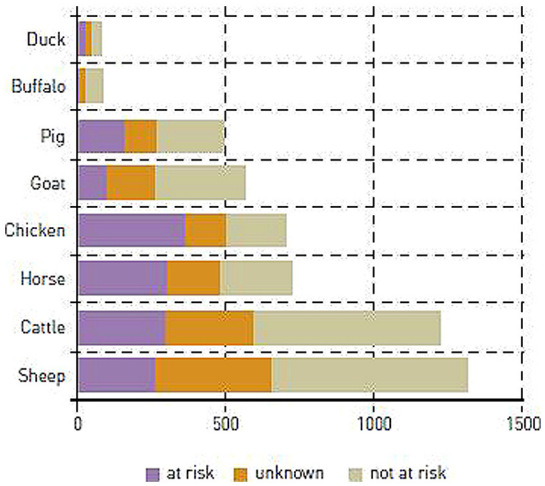


FIGURE 1 | Estimated numbers of domestic animal breeds at risk worldwide (FAO Livestock report, 2005).

effects are highly trait and population specific (Frankham et al., 1998; Kristensen and Sørensen, 2005). Frankham et al. (1998) summarized estimates of inbreeding depression across different components of fitness in 15 domesticated and non-domesticated animals and plant species, and concluded that a 25% increase in inbreeding reduced mean fitness of inbred compared to outbred individuals by 15%. Levels of inbreeding depression are on average higher for fitness-related traits compared to morphological or behavioral traits (DeRose and Roff, 1999; Willi et al., 2006). Further there is evidence that inbred individuals are less robust when exposed to stressful environmental conditions compared to outbred individuals (i.e., inbreeding effects are exacerbated by stressful conditions, resulting in inbreeding by environment interactions; Armbruster and Reed, 2005; Reed et al., 2012; **Figure 2**). When estimates of inbreeding depression from Frankham et al. (1998) are separated into estimates from domestic and non-domesticated species, it appears that inbreeding depression may be somewhat less pronounced in the former group (12 versus 17%). Although this needs further investigation, this difference might reflect the relatively benign and less variable environments experienced by domestic animals. The environmental dependency of inbreeding depression has been shown in numerous studies in natural populations (e.g., Jimenez et al., 1994; Hauser and Loeschcke, 1996; Keller et al., 2002; Szulkin and Sheldon, 2007) but surprisingly mostly in older literature on animal and plant breeding (e.g., Finlay, 1963; Hull, 1963).

There is evidence in wild as well as domestic populations that recessive deleterious alleles can be purged with inbreeding (Hedrick and Kalinowski, 2000; Crnokrak and Barrett, 2002; Charlesworth, 2009). This results in inbreeding depression being diminished in *Drosophila* and bird populations with a long history of inbreeding (Swindell and Bouzat, 2006; Laws and Jamieson, 2011). Purging, however, can be environment-specific; thus *Drosophila* studies have shown that purging performed in one environment might not be efficient in reducing inbreeding

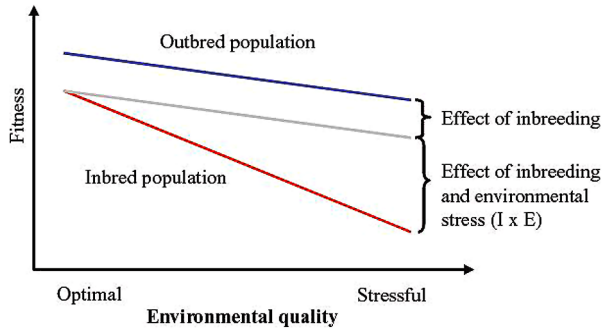


FIGURE 2 | Schematic illustration of fitness effects of inbreeding by environment interactions. Assuming the effect of inbreeding is independent of the environment, the reduction in fitness as a result of reduced environmental quality will be equal for outbred and inbred populations. The blue and gray lines illustrate fitness of an outbred and an inbred population, respectively, in the absence of inbreeding by environment interactions. Inbreeding depression is, however, often more severe under stressful environmental conditions. Thus, the red line illustrates fitness of an inbred population taking into account the effect of inbreeding by environment interactions (redrawn from Kristensen et al., 2010).

depression in another environment (Bijlsma et al., 1999; Dahlgaard and Hoffmann, 2000; Mikkelsen et al., 2010).

Both environment-specific purging and inbreeding by environment interactions suggest that an inbred population that does not currently suffer from inbreeding depression is nevertheless likely to do so if the environment changes and particularly if it becomes more stressful. This is expected to contribute to the extinction risk in small natural populations facing environmental shifts including dramatic climate changes. Liao and Reed (2009) determined that inclusion of fitness effects stemming from inbreeding by environment interactions reduced persistence times of natural populations by 17.5–28.5% across a wide range of scenarios. This might be an overlooked reason for poor performance of domestic livestock moved across areas where production systems differ, such as dairy cattle breeds exported from temperate to tropical climates and *vice versa* (Zwald et al., 2003).

Genomic tools will increase our understanding of the genetic architecture of inbreeding depression in natural and domestic populations (Charlesworth and Willis, 2009; Kristensen et al., 2010; Ouborg et al., 2010). Genomics is already used to accurately estimate levels of inbreeding, detect loci that contribute significantly to inbreeding depression, and dissect the history of inbreeding in a population (old or recent; Charlesworth, 2009; Purfield et al., 2012; Curik et al., 2014; Pertoldi et al., 2014). Genomic tools should also help untangle the population and trait specific nature of inbreeding effects by identifying the nature of interactions between identified loci and their genetic background. So far genomic tools used to dissect the genetics of inbreeding have mainly been applied to model organisms and livestock but genomic resources are also rapidly emerging for wild animals and plants (Ouborg et al., 2010; Ekblom and Galindo, 2011). Dissecting the genomic architecture of inbreeding in natural populations is helped if there are genomic resources available for a related species; this facilitates assembly and functional annotation (Ekblom and Galindo, 2011). Hence a mapped and annotated cattle or chicken genome can facilitate studies on small and inbred natural populations of closely related mammal and bird species (Romanov et al., 2009).

GENETIC DRIFT AND LOSS OF GENETIC VARIATION

Genetic drift represents the random change of allele frequencies over generations due to finite population size, and a small Ne is expected to increase the rate of genetic drift and associated loss of genetic variation across generations as alleles become fixed (Wright, 1929; Falconer and Mackay, 1996; Willi et al., 2006; Merilä, 2014). These theoretical predictions are generally supported by empirical data from model organisms and wildlife (and to a lesser degree in livestock) showing less genetic variation and selection responses in populations with small Ne (Jones et al., 1968; Weber, 1990; Caballero et al., 1991; Kristensen et al., 2005; Hill and Kirkpatrick, 2010; Olson-Manning et al., 2012; Kim et al., 2013). The high heterozygosity values sometimes reported in domestic breeds despite low Ne may, at least in part, result from biased selection of hypervariable microsatellite markers in chromosome regions not under selection (Taberlet et al., 2008). Kim et al. (2013) provide support for this hypothesis showing that Holstein dairy cattle selected intensively for increased milk yield do have increased overall autozygosity when assessed across the genome; this is likely a consequence of both genetic drift and selection.

What practical implications do expected reductions in genetic variation in small populations have for the management of small breeds? Breeding in commercial dairy cattle breeds like Holstein Frisian and Jersey with Ne's below 100 has been very efficient and no apparent signs of selection plateaus have been seen (Chikhi et al., 2004; Sørensen et al., 2005; Hill and Kirkpatrick, 2010; Leroy et al., 2013). With very high selection intensity, most traits can be changed through directional selection (Hill and Kirkpatrick, 2010) and genetic variation may remain relatively stable within the time frames typically considered by a breeding company or an individual farmer. Many domestic species have long generation length, and intense selection within domestic breeds has (from an evolutionary point of view) not been practiced for long. The breed concept is not more than approximately 200 years old, and reproductive technologies enabling intense selection has only been in common use since the 1960s (Taberlet et al., 2011).

Genomic selection applied to animal breeding will reduce generation intervals which is one reason for expected increased genetic gain using this technique (Schaeffer, 2006; Meuwissen, 2007). However, assuming that the high level of genetic variation observed in some domestic breeds despite low Ne is partly a consequence of the long generation length (genetic variation is simply not lost yet), genomic selection may speed up loss of genetic variation. It is estimated that generation intervals in dairy cattle will be reduced by 50% in the future because individuals will have breeding values at birth (in contrast to the situation where an individual's performance has to be assessed first; Blasco and Toro, 2014). Thus genomic selection could potentially double the speed at which genetic variation is lost within breeds, even though genomic selection can also be used to optimize heterozygosity (see, e.g., Pedersen et al., 2009; Pertoldi et al., 2014).

Ne RECOMMENDATIONS

Rules of thumb that have influenced management of domestic and wild populations since the early 1980s state that (1) an Ne of at least 50 is needed for avoiding inbreeding depression in the short term (five generations), and (2) an Ne of at least 500 are sufficient to retain long-term evolutionary potential (Franklin, 1980; Soulé, 1980; Franklin and Frankham, 1998). It has been suggested that these numbers are much too low (Willi et al., 2006; Frankham et al., 2014) for several reasons, including the fact that they are not based on realistic N/Ne ratios, they do not take into account inbreeding depression and that environments are changing at an unprecedented speed. The value of Ne recommendations in a conservation management context has been the subject of extensive discussion (see, e.g., Willi et al., 2006; Jamieson and Allendorf, 2012; Frankham et al., 2014). Such recommendations can be criticized because the potential to adapt (1) depends on the type of environmental change and the type of traits, and (2) the evolutionary potential of small populations is compromised by processes that do not directly alter genetic variation such as suboptimal environmental conditions (which may lower heritability estimates due to increased environmental variance), inbreeding depression (Willi et al., 2006), and inbreeding by environment interactions (Armbruster and Reed, 2005; Reed et al., 2012). Despite these limitations, recommendations remain useful because the alternative might be unscientific conservation decisions made at the political and bureaucratic levels (Frankham et al., 2014; **Box 1**).

For domesticated animals, Leroy et al. (2013) found that the large majority of domestic breeds have Ne's below 500 when estimated as the increase in homozygosity over generations by measuring identity by descent probabilities. Thus Ne estimates were below 500 in all but a few of the 20 cattle breeds, 40 sheep breeds, 20 horse breeds, and 60 dog breeds investigated, and several had Ne estimates below 50 (Leroy et al., 2013). The low Ne estimates observed in livestock are mirrored in many wildlife species and populations, although precise estimates are difficult to obtain and tend to vary depending on the method employed (Frankham, 1995; Luikart et al., 2010; Leroy et al., 2013). According to the International Union for Conservation of Nature (IUCN, 2014), more than 22,000 species are currently threatened in nature. Effective population sizes of these species are rarely known, but assuming a Ne/N ratio of 0.1 the huge majority of these species have effective population sizes below 500 individuals. Thus these numbers match findings in livestock breeds.

However, these estimates may be revised as more accurate estimates of Ne based on genome-wide markers emerge. A recent study on Chinook salmon (*Oncorhynchus tshawytscha*) estimated Ne from RAD-seq data and found surprisingly high Ne estimates given the census size of these populations (Larson et al., 2014). This approach will also be valuable in Ne estimates for livestock and possibly shed light on the paradox that many commercial breeds have low Ne (typically based on pedigree data) but high and continued selection responses (Hill and Kirkpatrick, 2010).

With Ne being low for many breeds and natural populations, a challenge is to take steps to ensure that Ne is maintained or even increased. Equalizing sex ratios of those animals contributing to the next generation, and reducing variation in the number of offspring, directional selection, inbreeding and variation in Ne across generations are all important for increasing Ne (or diminishing its further reduction; Charlesworth, 2009). In many European countries farmers are subsidized for keeping indigenous breeds. Ways to encourage stabilization or increases in breed Ne would be to support initiatives such as allowing more males to contribute to the next generation to equalize sex ratios, to include cryopreserved genetic material in the breeding plan (Sonesson et al., 2002), or by controlling inbreeding in the herd efficiently such as by the use of software programs (Sørensen et al., 2008). However, increasing Ne is a daunting task in a small population where there is no possibility of increasing genetic variation through immigration. The harmonic means of Ne across generations describe the impact of fluctuations in population size on overall Ne (Caballero, 1994; Falconer and Mackay, 1996). A population that has been through a genetic bottleneck will suffer long lasting consequences in terms of reduced Ne despite the fact that the population size might have increased following the bottleneck (Charlesworth, 2009). Genetic rescue of a population or breed might then be required as discussed below.

According to Ne recommendations mentioned above, the long-term evolutionary potential for the majority of livestock breeds and threatened wild species is expected to be severely diminished. The number of generations required to reach selection plateaus in small populations obviously depends on numerous factors such as the level of standing genetic variation for the trait in question, number of loci contributing to the variation, and the selection intensity. Results from selection experiments are diverse but suggest that selection plateaus are typically reached in less

BOX 1. General recommendations for conservation of populations of species with conservation concerns.

Scenario 1: A genetically unique population that is highly adapted to local conditions and has high levels of genetic variation. *Recommendation:* If Ne >500–1000 and not decreasing there may be no need to change management strategies. *Conservation priority:* High.

Scenario 2: A genetically unique population (that is) highly adapted to local conditions but with limited genetic variation. *Recommendation:* Despite currently being successful in its environment some gene flow from populations is recommended to increase Ne and genetic variation. *Conservation priority:* High.

Scenario 3: A population that is unique but maladapted and has high levels of genetic variation. *Recommendation:* Characterize why this population is maladapted and start to select (if the population is managed) for increased local adaptation. Gene flow from populations adapted to similar environments is recommended if the population does not respond to selection. *Conservation priority:* Intermediate.

Scenario 4: A population that is locally adapted, but not genetically unique and with low levels of genetic variation. *Recommendation:* Management should prioritize increasing Ne and genetic variation. Gene flow from populations adapted to similar environments is recommended. *Conservation priority:* Low assuming that there are other conspecific populations available.

Scenario 5: A population that is not unique, maladapted, and with low levels of genetic variation. *Recommendation:* Gene flow from populations adapted to similar environments is needed. *Conservation priority:* Low.

than 30 generations in mice and *Drosophila* (references in Falconer and Mackay, 1996). With a typical 3–4 years generation interval (defined as the average age of parents when offspring are born) for cattle, this corresponds to ∼90–120 years of selection. Still there is no evidence suggesting selection plateaus in the majority of commercial livestock breeds (Hill and Kirkpatrick, 2010). Genomics data on livestock are likely to help understand how selection has shaped genetic variation in populations, such as by enabling more accurate Ne estimates and pinpointing which parts of the genome have been under selection and remain variable.

CONSEQUENCES OF LOW Ne IN SMALL INDIGENOUS LIVESTOCK BREEDS

Many indigenous livestock breeds have low Ne and several are considered threatened, endangered or already extinct (Hoffmann, 2013; Leroy et al., 2013; DAD-IS, 2014; **Figure 1**). Intense and structured directional selection is typically not performed in these breeds, and many of them have been through more extreme bottlenecks compared to the commercial breeds, with documented low levels of genetic variation (Melis et al., 2012; Herrero-Medrano et al., 2014; Pertoldi et al., 2014). Accordingly, they are more likely to suffer from evolutionary constraints, inbreeding and drift load due to low Ne compared to modern commercial breeds. From an evolutionary point of view this has two obvious and immediate consequences of relevance for the management of indigenous breeds with small Ne: (1) in breeds with no remaining genetic variation, evolution is constrained no matter how intense the selection pressure, unless it involves removal of newly arisen deleterious alleles, and (2) even if genetic variation is present, only small responses to selection are expected. This is because when Ne decreases, the impact of genetic drift increases and loci under selection start to behave as neutral when selection coefficients become equal or smaller than 1/[2Ne] (Wright, 1931). Hence, to prevent the loss of rare beneficial alleles by genetic drift in small populations, stronger selection is required.

There will sometimes be good cultural, historical and also genetic reasons for managing indigenous (and commercial) breeds as separate and closed breeds. However, this will rarely be the most efficient management practice if the goal is to conserve breeds, adaptive genetic variation and the scope for local adaptation in the long run. *In situ* conservation is sometimes referred to as the golden standard in the management of domestic breeds. We argue that for *in situ* conservation to be efficient in maintaining and generating locally adapted breeds, Ne should be increased so that it counts 100s and not tens of animals. Until this goal is achieved, what matters from a genetic perspective is to increase genetic variation so that evolution is not constrained. The generation of new mutations is a very slow process, and genetic rescue by bringing in new variation from other populations will often be the only option (see **Box 1** for recommendations regarding the management of small populations).

LOW Ne AND MALADAPTATION

In the literature it is sometimes taken for granted that a population in the wild or an indigenous livestock breed is locally adapted to the habitat or geographical region in which it is present. Although some well-documented cases exists, especially from the tropics (Bayer et al., 1987; Ayantunde et al., 2002; references in Hoffmann, 2010, 2013), it is often not known whether a given phenotype represent an adaptation to the local environment (Hoffmann, 2013). This issue also applies to natural populations where there is often good evidence for local adaptation (Dobzhansky, 1956) but also many cases where it does not seem to occur (references in Crespi, 2000). Thus proper, standardized and continuous phenotypic and genotypic characterization according to guidelines such as those suggested by FAO (2011, 2012) is highly recommended.

Maladaptation can have many genetic causes, including mutation, inbreeding, random genetic drift, gene flow leading to breakdown of co-adapted gene complexes, heterozygote advantage and pleiotropy (Crespi, 2000). Space does not allow reviewing these causes in detail but we discuss some aspects relevant for the management of domestic breeds and small natural populations (see also **Box 1**).

Gene flow and genetic drift might prevent or disrupt local adaptation and lead to maladaptation or outbreeding depression. Due to genetic drift, genes of adaptive value may behave neutrally, potentially leading to maladaptation in small populations. Gene flow may prevent local adaptation (Stearns and Sage, 1980), but this depends on whether gene flow is constrained by geographical distance or the environment, and cases of maladaptive gene flow in wild populations are relatively uncommon (Sexton et al., 2014). Introgression may have deleterious effects if there is outbreeding depression, however, the probability of outbreeding depression in crosses between two populations of the same species appears to be low for populations with the same karyotype, isolated for <500 years, and that occupy similar environments (Frankham et al., 2011). These are important aspects when considering crossbreeding as an option in domestic breeds. Amador et al. (2014) present two genomic selection strategies, using genome-wide DNA markers, to recover the genomic content of the original endangered population from admixtures. Such tools will be useful in the conservation of domestic populations where crossing between breeds has occurred intentionally or unintentionally and it is worthwhile recovering the breeds.

GENETIC RESCUE – EXAMPLES FROM THE WILD

The merits and challenges of genetic rescue – augmenting genetic variation and limiting inbreeding depression – have been evaluated for over a decade in wildlife biology (e.g., Ingvarsson, 2001; Vilá et al., 2003; Tallmon et al., 2004; Edmands, 2007; Adams et al., 2011). The lessons learned may offer important insights for conservation of threatened livestock, where many breeds are facing threats parallel to those of wild species including inbreeding and rapid loss of genetic variation (Kristensen and Sørensen, 2005; Leroy et al., 2013).

Examples of genetic rescue include the arrival of new wolves (*Canis lupus*) into the isolated and highly inbred populations on Isle Royale, MI, USA (Adams et al., 2011) and the Scandinavian Peninsula (Vilá et al., 2003; Hagenblad et al., 2009). These have led to a subsequent increase in genetic diversity, and potentially in fitness although the latter is challenging to measure in wild populations (Ingvarsson, 2001, 2002), possibly short-lived (Hedrick and Fredrickson, 2010; Hedrick et al., 2014), and may be masked

by environmental factors such as availability of food and space (Adams et al., 2011).

A consideration for relocation programs in augmenting genetic diversity has been the introduction of donor animals from environments as similar as possible to the new location (e.g., Hedrick and Fredrickson, 2010), to help avoid maladaptation/outbreeding depression. This can occur when donor and recipient populations are adapted to different environmental conditions and the resulting hybrid offspring are ill-suited to either habitat (Templeton, 1994). Examples reported in the past include translocations of Arabian oryx (*Oryx leucoryx*; Marshall and Spalton, 2000), as well as ibex (*Capra ibex*) from Sinai and collared lizards (*Crotaphytus collaris*) from the US Ozark mountains (Templeton, 1994). However, the relationship between population divergence and hybrid fitness is highly variable among taxa and may be difficult to predict without experimental crosses, which, although recommended, are not always feasible in particular for long-lived species with long generation times (Edmands, 2007 and references therein; Hedrick and Fredrickson, 2010).

In some instances, genetic rescue occurs naturally and without human intervention or knowledge, and is later confirmed by genetic investigations (e.g., Scandinavian wolves; Ingvarsson, 2002; Vilá et al., 2003; Hagenblad et al., 2009). In these situations, the arrival of the rescuing individual(s) typically poses no ethical or conservation dilemma. In contrast, there is discussion concerning the merits and ethical implications of human-mediated genetic rescue. The small and isolated wolf population on Isle Royale in Lake Superior, USA, represents a valuable illustration (Vucetich et al., 2012, 2013; Cochrane, 2013; Mech, 2013). Wolves arrived unassisted on the island by crossing the ice, and the population has experienced bottlenecks, in part owing to suspected human introduction of parvovirus in dogs that caused a crash in the wolf population (Peterson et al., 1998). Inbreeding is reported to have caused problems with bone malformations, which could limit movement and potentially the ability to hunt large prey such as moose (*Alces alces*) and reduce general life expectancy (Räikkönen et al., 2009). The arrival of a more recent immigrant appears to have produced a selective sweep and genetic rescue of the population, but isolation and environmental conditions (such as reduced ice cover with climate change and thus reduced chance of new immigrants) pose continuing long-term challenges (Hedrick and Fredrickson, 2010; Adams et al., 2011).

At times, it may be necessary to crossbreed with individuals from another relatively similar population to augment genetic diversity. In such situations there may be no ideal solution for the introduction of new genetic material. Genetic rescue of the Florida subspecies of panther (*Puma concolor coryi*) implied relocating individuals from the closest wild population, a subspecies from Texas (*Puma concolor stanleyana*), which was a controversial decision (Pimm et al., 2006; Hedrick and Fredrickson, 2010; Johnson et al., 2010). The remnant Florida population suffered from several problems believed to be associated with genetic drift and inbreeding, such as undescended testicles and morphological abnormalities (reviewed in Pimm et al., 2006; Hedrick and Fredrickson, 2010). The relocation appears to have improved the survival, genetic diversity and range of the Florida panther and augmented the population, at least for the moment (Pimm et al., 2006; Hedrick and Fredrickson, 2010; Johnson et al., 2010).

Management decisions will often have to be made with incomplete knowledge of all relevant scientific data, and requires explicit acknowledgment of the ethical norms that are guiding principles in conservation biology (Soulé, 1985). The social and biological sciences are therefore both important, as is the acknowledgment that human intervention may play an essential and often necessary role. Additionally, Templeton (1994) and Weeks et al. (2011) recommend prioritizing genetic diversity and the potential for evolutionary change, and note that conservation efforts should aim to preserve processes such as evolution rather than specific genetic variants. In this context, human-induced fragmentation and subsequent genetic drift may have had a major influence on wildlife such as Texas and Florida panthers that in the past likely exchanged genes via intermediary populations (Hedrick and Fredrickson, 2010).

GENETIC RESCUE IN DOMESTIC BREEDS

A situation parallel to these wild populations, involving a small population with few breeders, and ensuing risks of reduced fitness and long-term survival, is the endangered Norwegian Lundehund, a Spitz breed native to coastal Norway where it was historically used to hunt puffins (*Fratercula arctica*). The remaining individuals have extremely low genetic diversity and are highly inbred (Melis et al., 2012; Pfahler and Distl, 2013). Efforts are currently underway to evaluate similar Nordic Spitz breeds for possible cross-breeding[1]. Although the genetic background for the condition(s) remains unknown, the Lundehund is affected by serious gastrointestinal problems that seem particularly prevalent for this breed (Landsverk and Gamlem, 1984; Berghoff et al., 2007; Qvigstad et al., 2008) and for which cross-breeding may be beneficial. Whereas cross-breeding may alter the breed's morphology and behavior, there appears to be no alternative means of increasing genetic diversity.

Native livestock species may be well-adapted to their regions and be of historical and economic importance (Joost et al., 2007; Pariset et al., 2009; Hoffmann, 2013). The population declines and isolation experienced by many native breeds following the expansion of modern agriculture (Taberlet et al., 2008, 2011) typically occur because they are less productive – in terms of, e.g., milk, wool, or meat production – than the commercial breeds. The native breeds may be adapted to a more stringent environment and climate, such as that of high mountains and northern coastal areas that are relatively marginal for agriculture. Hence, their commercial disadvantage might protect important genetic variants that promote survival in harsher climates (e.g., Hoffmann and Parsons, 1997; Joost et al., 2007). Preservation of genetic diversity in native breeds may have key evolutionary applications (Kantanen et al., 2000; Taberlet et al., 2008), including adaptation to climate change and in promoting sustainable agriculture (Hoffmann, 2013).

Environmental conditions for domestic animals are typically benign, increasing the probability of survival for crosses that might not otherwise survive in the wild. However, where livestock

[1] http://lundehund.no/

are maintained for free-roaming grazing and landscape management, their survival under difficult environmental conditions may be paramount. For a thorough examination of inbreeding and outbreeding it is necessary to study the entire life cycle and not focus on any single component of fitness (Edmands, 2007), and also to consider these effects on organisms in their natural environment (Kristensen et al., 2008; Enders and Nunney, 2012). Furthermore, Edmands (2007) highlights the critical importance of understanding how hybridization affects the generations beyond F2 and the initial back-crosses, although any such effects will be ameliorated to some extent by ongoing selection against poorly adapted F2 genotypes (Weeks et al., 2011). Genetic rescue may be possible when the only available donors are from other inbred populations, and reciprocal translocations or gene flow between such populations may provide an important short-term measure for conservation (Heber et al., 2013). However, inbreeding and outbreeding depression may occur at the same time and their effects can be difficult to distinguish, especially in managed populations (Edmands, 2007 and references therein). These findings merit additional attention for endangered species of wild and domestic species, and the latter may provide important opportunities for experimental and controlled study with benefits for both wildlife and livestock at risk.

HOW CAN GENOMICS BENEFIT CONSERVATION OF LIVESTOCK BREEDS AND WILD POPULATIONS?

Current management guidelines for populations at risk frequently emphasize genetic uniqueness over genetic diversity (Funk et al., 2012). Such practices may need review, as trade-offs between genetic diversity and genetic uniqueness has been observed (Coleman et al., 2013). For example, the endangered dwarf galaxias *Galaxiella pusilla* populations reported to be genetically most unique also have the least amount of genetic variation when assessed with microsatellite genetic markers (Coleman et al., 2013). This relationship between uniqueness and diversity may be a general observation in small and threatened populations where genetic structure is strongly affected by genetic drift. Prioritizing populations that are genetically unique may therefore, at times, decrease overall genetic diversity and, accordingly, reduce evolutionary potential. These issues have been recognized more widely in genetic management of domestic species (Barker, 2001; Caballero and Toro, 2002).

Genomic tools can contribute to genetic resource management through accurate estimation of genetic uniqueness and control of inbreeding (Li et al., 2011; Amador et al., 2014; Pertoldi et al., 2014). Breeding schemes based on information from commercial single nucleotide polymorphism (SNP) chips for example can be used to select efficiently against deleterious alleles in a population and/or select for increased heterozygosity in genes of adaptive significance provided these can be identified (for an example see, Marcos-Carcavilla et al., 2010). Thus revision of breeding plans based on genome-wide study of variation within populations is now a practical option. For example Pertoldi et al. (2014) showed how genome-wide SNP data can be used to design breeding programs aiming at reducing the loss of genetic variability within a small population of Danish cattle by prioritizing matings between individuals with relatively low pairwise identity-by-state.

Another element of genetic rescue could involve adaptive management toward climate change (Aitken and Whitlock, 2013), where native breeds well-adapted to local conditions such as high precipitation (Joost et al., 2007) might provide genetic material to commercial breeds and isolated populations of the same or similar breeds. This may be necessary in view of future climate change, where adaptations to factors such as hot, arid, and saline conditions (Hoffmann, 2010; Hoffmann et al., 2013) may be increasingly essential for survival. Caballero and Toro (2002) developed an approach that balances genetic diversity against uniqueness. However, this approach does not account for the fact that some unique populations may contain private alleles of adaptive value, which can arise as a consequence of either random genetic drift or strong selection. Genomic tools that permit identification of such genetic variants will therefore be highly valuable for identifying populations, and individuals within populations, of special importance for long-term conservation. Using genomic selection, these genes can be introduced very rapidly into populations.

Incorporating more individuals in breeding management will augment Ne, and candidate profiles should be evaluated with consideration to long-term evolutionary potential. This could involve weighing genetic diversity against uniqueness (Coleman et al., 2013); in some cases donor individuals may have overall low genetic diversity, but carry unique genetic variants that can benefit the more genetically diverse recipient population for specific traits. This is relevant if genetic variability in a commercial breed is considered for augmentation from a small local breed. In the short term, donor individuals may not have the highest breeding values for production traits, but in a broader perspective they could contribute valuable material with respect to increased robustness when exposed to diseases and environmental variability associated with climate change. Here, genome-wide profiles can assist in selecting individuals with the desired features of the local breed such as alleles associated with parasite resistance (Coltman et al., 2001), while matching the purpose of the commercial breed as closely as possible (production of, e.g., milk, meat, or wool).

An important difference between wild and domestic species is that not only fitness – survival, growth, fecundity – but also production and output in economic terms will be important in domestic species. That is, natural selection for independent survival in the local environment is traded for traits such as high yield of milk and meat that may be unsustainable under natural conditions [an extreme example is selection for cattle muscle mass necessitating high frequencies of Cesarean sections, e.g., Pirottin et al. (2005)]. Native and relatively naturally selected breeds may contribute genetic material to commercial breeds, and these could in turn contribute genetic variation to native breeds. The direction(s) and scale of gene flow will depend on the specific conservation breeding objectives, such as avoiding extinction of a small native breed, or increasing the frequency of disease resistance-associated alleles in a large commercial herd.

Genome-wide profiles in the form of SNP markers permit investigation of neutral and functional genes, providing insight

into a broader range of evolutionary processes (e.g., Rice et al., 2011). This can help optimize breeding management in the form of detailed information on variation present within a herd, across a given breed, or within a species. A small native breed may have few or no unrelated individuals and may be considered for limited cross-breeding with a more variable breed to improve the probability of persistence. Genomics offers a tool to help screen and prioritize contributions, where traditional use of phenotypic information can be combined with accurate data on individual relatedness and genetic diversity.

The highly managed and small Ne of many domestic breeds provide valuable learning opportunities to help bridge the gap between laboratory model organisms and wild species in understanding the strength of selection in small populations. One example is work to identify the chondrodystrophy (dwarfism) locus in California condor which is being aided by comparisons with the DNA sequence of the domestic chicken (Romanov et al., 2009). Such examples can benefit conservation of endangered wildlife and domestic species, and inform future work on model organisms.

Another example is the work currently underway to cross the dog breeds Lundehund and Buhund to augment the genetic diversity and conservation (including health status) of the Lundehund. Hybrid pups were born in 2014[2], and will be evaluated with respect to factors such as morphology and behavior. Careful selection of future animals used for breeding should be possible through genomic analyses of hybrid profiles, by following backcrosses between F1-hybrids and Lundehund. An important objective will be to retain physical characteristics of the Lundehund breed whereas there is urgent need to reduce the prevalence and severity of the breed's gastrointestinal problems. The methods proposed by Amador et al. (2014) on the use of genomic selection to recover the original genetic background from hybrids may be highly applicable in this respect. Identification of genes involved in similar gastrointestinal disorders is advancing for humans (see, e.g., McGovern et al., 2010 on ulcerative colitis and Crohn's disease). Even though other genes may be implicated in gastrointestinal diseases of dogs, breeding individuals may be selected that contribute new genetic variation to the Lundehund for the genome regions known to be affected in humans.

With domestic species it is relatively easy to perform controlled experimental breeding and evaluate offspring characteristics, helping to assess the consequences of mixing two breeds, and optimizing strategies for cross-breeding. Thus genomic data can help expand conservation management of farm animals by permitting a careful, adaptive process with a long-term perspective whereby breeds are explicitly considered as an evolutionary work-in-progress (**Box 2**).

BOX 2. Insights/tools for genetic management of small populations.

- Many hereditary diseases are caused by recessive deleterious alleles. Genomic tools are useful for identifying the causal mutation and for detection of carriers (heterozygotes). *Example:* Identification and efficient selection against the mutation causing complex vertebral malformation in Holstein dairy cattle (Thomsen et al., 2006; Whitlock et al., 2008).
- Cryopreservation of biological material including semen, oocytes, and embryos is possible for most livestock species (Mara et al., 2013) but not developed for most species of conservation concern. *Perspectives:* The use of frozen semen or embryos in breeding plans for small populations can decrease inbreeding and increase genetic variation (Sonesson et al., 2002; Woelders et al., 2012).
- Genomes of most species can now be sequenced for relatively reasonable costs (Hayden, 2014). *Perspectives:* Novel applications of genomic data in domestic and non-domestic populations include: (1) genome-wide profiles comprising neutral and adaptive genetic variation enabling selection for specific adaptive alleles and genetic variation, (2) estimates of Ne that are accurate and comparable, and (3) description of the demographic history of populations.
- Inbreeding depression is a trait, population and environment-specific phenomena. *Perspectives:* Generalizations are problematic but simple guidelines suggest that Ne >50 is sufficient to avoid significant inbreeding depression in livestock (Meuwissen and Woolliams, 1994) whereas this number likely needs to be doubled in wild populations (Frankham et al., 2014).
- The effective population size, Ne, is below 500 in the vast majority of livestock breeds and threatened species in nature. *Perspectives:* Long-term evolutionary change is likely constrained in these populations. Genetic rescue through hybridization is a tool to increase Ne and genetic variance and reduce inbreeding in small natural populations (Hogg et al., 2006) and domestic breeds (Schaeffer et al., 2011).

FUTURE OUTLOOK

The above considerations indicate that genomic data provide highly useful information about the applicability and likely success of strategies aimed at increasing and maintaining the adaptability of threatened breeds and natural populations. By providing information on the genetic architecture of complex traits, these approaches provide tools that are changing the way we approach conservation of genetic variation in livestock and natural populations. Whereas genomics is not a magic bullet that will instantly permit full overview of the genome and the interactions among its parts, the use of an increased number of genetic markers through next generation sequencing approaches will augment the accuracy of estimating diversity and population demographic parameters of conservation relevance (Allendorf et al., 2010; Shafer et al., in press). Furthermore it opens up the possibility to screen individuals and populations for adaptive loci and use genomic information to guide breeding decisions including which populations and individuals to use in genetic rescue programs (Amador et al., 2014; Toro et al., 2014). Genome-scale data will also be increasingly used to document the demographic impact resulting from genetic rescue programs. In one example Miller et al. (2012) showed that migrant alleles (from translocated individuals) increased over time in an insular population of bighorn sheep (*Ovis canadensis*). More generally, Amador et al. (2014) provide a framework on how to use genomic selection for recovery of original genetic background from hybrids. At this early stage methods still need to mature, pipelines for dealing with assembly and annotation in non-model organisms need further development, and clear examples of practical applications need to be disseminated to practitioners (Shafer et al., in press).

[2]http://lundehund.no/om-norsk-lundehund-klubb/ras

For conservation of livestock breeds and small populations in nature, we advocate that long-term genetic and phenotypic monitoring is needed, and based on such data management decisions are taken (**Box 1**). Current and past performance of a population may represent poor predictors of future performance due to factors such as genotype by environment interactions, inbreeding, genetic drift, changed breeding objectives, and environmental changes. In the long run genetic variation will be depleted in populations with small Ne. These populations will become constrained in their evolutionary responses and are likely to go extinct in environments that change rapidly and may become stressful, although the experience with livestock suggests that as long as environments are relatively constant phenotypic changes remain possible under strong selection. Nevertheless, threatened populations should not be regarded as museum specimens, and an increased focus on Ne and genetic variation in populations of conservation concern should help promote the potential for adaptive evolution. For small and threatened breeds one way to achieve this is through limited cross-breeding and active use of genomic data to guide breeding decisions.

ACKNOWLEDGMENTS

This research was funded by the Danish Natural Research Council with a *Sapere aude* stipend to TNK (DFF – 4002-00036) and a *post doc* grant to AVS (DFF – 1337-00007). CP thank the Aalborg Zoo Conservation Foundation (AZCF) for providing financial support. The Australian Research Council and Science and Industry Endowment Fund provided financial support to AAH.

REFERENCES

Adams, J. R., Vucetich, L. M., Hedrick, P. W., Peterson, R. O., and Vucetich, J. A. (2011). Genomic sweep and potential rescue during limiting environmental conditions in an isolated wolf population. *Proc. R. Soc. B Biol. Sci.* 278, 3336–3344. doi: 10.1098/rspb.2011.0261

Aitken, S. N., and Whitlock, M. C. (2013). Assisted gene flow to facilitate local adaptation to climate change. *Annu. Rev. Ecol. Evol. Syst.* 44, 367–388. doi: 10.1146/annurev-ecolsys-110512-135747

Allard, R. W. (1999). *Principles of Plant Breeding*, 2nd Edn. New York: John Wiley & Sons.

Allendorf, F. W., Hohenlohe, P. A., and Luikart, G. (2010). Genomics and the future of conservation genetics. *Nat. Rev. Genet.* 11, 697–709. doi: 10.1038/nrg2844

Amador, C., Hayes, B. J., and Daetwyler, H. D. (2014). Genomic selection for recovery of original genetic background from hybrids of endangered and common breeds. *Evol. Appl.* 7, 227–237. doi: 10.1111/eva.12113

Armbruster, P., and Reed, D. H. (2005). Inbreeding depression in benign and stressful environments. *Heredity* 95, 235–242. doi: 10.1038/sj.hdy.6800721

Aslam, M. L., Bastiaansen, J. W. M., Elferink, M. G., Megens, H. J., Crooijmans, R. P. M. A., Blomberg, L., et al. (2012). Whole genome SNP discovery and analysis of genetic diversity in Turkey (*Meleagris gallopavo*). *BMC Genomics* 13:391. doi: 10.1186/1471-2164-13-391

Ayantunde, A. A., Fernandez-Rivera, S., Hiernaux, P. H., Van Keulen, H., and Udo, H. M. J. (2002). Day and night grazing by cattle in the Sahel. *J. Range. Manage.* 55, 144–149. doi: 10.2307/4003350

Barker, J. S. F. (2001). Conservation and management of genetic diversity: a domestic animal perspective. *Can. J. Forest. Res.* 31, 588–595. doi: 10.1139/x00-180

Bayer, W., Suleiman, H., Kaufmann, R., and Waters-Bayer, A. (1987). Resource use and strategies for development of pastoral systems in sub-humid West Africa. The case of Nigeria. *Q. J. Int. Agric.* 26, 58–71.

Berghoff, N., Ruaux, C. G., Steiner, J. M., and Williams, D. A. (2007). Gastroenteropathy in Norwegian Lundehunds. *Compendium. Contin. Educ. Vet.* 29, 456–465.

Bijlsma, R., Bundgaard, J., and Van Putten, W. F. (1999). Environmental dependence of inbreeding depression and purging in *Drosophila melanogaster*. *J. Evol. Biol.* 12, 1125–1137. doi: 10.1046/j.1420-9101.1999.00113.x

Blasco, A., and Toro, M. A. (2014). A short critical history of the application of genomics to animal breeding. *Livest. Sci.* 166, 4–9. doi: 10.1016/j.livsci.2014.03.015

Caballero, A. (1994). Developments in the prediction of effective population size. *Heredity* 73, 657–679. doi: 10.1038/hdy.1994.174

Caballero, A., and Toro, M. A. (2002). Analysis of genetic diversity for the management of conserved subdivided populations. *Conserv. Genet.* 3, 289–299. doi: 10.1023/A:1019956205473

Caballero, A., Toro, M. A., and Lopez-Fanjul, C. (1991). The response to artificial selection from new mutations in *Drosophila melanogaster*. *Genetics* 128, 89–102.

Charlesworth, B. (2009). Effective population size and patterns of molecular evolution and variation. *Nat. Rev. Genet.* 10, 195–205. doi: 10.1038/nrg2526

Charlesworth, D., and Willis, J. H. (2009). The genetics of inbreeding depression. *Nat. Rev. Genet.* 10, 783–796. doi: 10.1038/nrg2664

Chikhi, L., Goossens, B., Treanor, A., and Bruford, M. W. (2004). Population genetic structure of and inbreeding in an insular cattle breed, the Jersey, and its implications for genetic resource management. *Heredity* 92, 396–401. doi: 10.1038/sj.hdy.6800433

Cochrane, T. (2013). Island complications: should we retain wolves on Isle Royale? *George Wright Forum* 30, 313–325.

Coleman, R. A., Weeks, A. R., and Hoffmann, A. A. (2013). Balancing genetic uniqueness and genetic variation in determining conservation and translocation strategies: a comprehensive case study of threatened dwarf galaxias, *Galaxiella pusilla* (Mack) (Pisces: Galaxiidae). *Mol. Ecol.* 22, 1820–1835. doi: 10.1111/mec.12227

Coltman, D. W., Wilson, K., Pilkington, J. G., Stear, M. J., and Pemberton, J. M. (2001). A microsatellite polymosphism in the gamma interferon gene is associated with resistance in a naturally-parasitized population of soy sheep. *Parasitology* 122, 571–582. doi: 10.1017/S0031182001007570

Crespi, B. J. (2000). The evolution of maladaptation. *Heredity* 84, 623–629. doi: 10.1046/j.1365-2540.2000.00746.x

Crnokrak, P., and Barrett, S. C. H. (2002). Purging the genetic load: a review of the experimental evidence. *Evolution* 56, 2347–2358. doi: 10.1111/j.0014-3820.2002.tb00160.x

Croquet, C., Mayeres, P., Gillon, A., Vanderick, S., and Gengler, N. (2006). Inbreeding depression for global and partial economic indexes, production, type, and functional traits. *J. Dairy Sci.* 89, 2257–2267. doi: 10.3168/jds.S0022-0302(06)72297-4

Curik, I., Ferencakovic, M., and Soelkner, J. (2014). Inbreeding and runs of homozygosity: a possible solution to an old problem. *Livest. Sci.* 166, 26–34. doi: 10.1016/j.livsci.2014.05.034

DAD-IS. (2014). *Intergovernmental Technical Working Group on Animal Genetic Resources*, Eight Session, Rome. Available at: http://dad.fao.org/ (assessed October 1, 2014).

Dahlgaard, J., and Hoffmann, A. A. (2000). Stress resistance and environmental dependency of inbreeding depression in *Drosophila melanogaster*. *Conserv. Biol.* 14, 1187–1192. doi: 10.1046/j.1523-1739.2000.99206.x

DeRose, M. A., and Roff, D. A. (1999). A comparison of inbreeding depression in life-history and morphological traits in animals. *Evolution* 53, 1288–1292. doi: 10.2307/2640831

Dobzhansky, T. (1956). What is an adaptive trait? *Am. Nat.* 90, 337–347. doi: 10.1086/281944

Edmands, S. (2007). Between a rock and a hard place: evaluating the relative risks of inbreeding and outbreeding for conservation and management. *Mol. Ecol.* 16, 463–475. doi: 10.1111/j.1365-294X.2006.03148.x

Ekblom, R., and Galindo, J. (2011). Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity* 107, 1–15. doi: 10.1038/hdy.2010.152

Enders, L. S., and Nunney, L. (2012). Seasonal stress drives predictable changes in inbreeding depression in field-tested captive populations of *Drosophila melanogaster*. *Proc. R. Soc. B Biol. Sci.* 279, 3756–3764. doi: 10.1098/rspb.2012.1018

Falconer, D. S., and Mackay, T. F. C. (1996). *Introduction to Quantitative Genetics*. Harlow: Longman.

FAO. (2011). Molecular genetic characterization of animal genetic resources. *FAO Anim. Product. Health Guidel.* 9, 7–25.

FAO. (2012). Phenotypic characterization of animal genetic resources. *FAO Anim. Product. Health Guidel.* 11, 27–80.

FAO Livestock report. (2005). *Animal genetic resources – time to worry*? Available at: ftp://ftp.fao.org/docrep/fao/010/a1404e/a1404e00.pdf [assessed October 01, 2014].

Finlay, K. W. (1963). "Adaptation: its measurement and significance in barley breeding," in *Proceedings of the 1st International Barley Genetics Symposium*, Wageningen.

Frankham, R. (1995). Effective population size/adult population size ratios in wildlife: a review. *Genet. Res.* 66, 95–107. doi: 10.1017/S0016672300034455

Frankham, R., Ballou, J. D., and Briscoe, D. A. (1998). *An Introduction to Conservation Genetics*. Cambridge: Cambridge University Press.

Frankham, R., Ballou, J. D., Eldridge, M. D. B., Lacy, R. C., Ralls, K., Dudash, M. R., et al. (2011). Predicting the probability of outbreeding depression. *Conserv. Biol.* 25, 465–475. doi: 10.1111/j.1523-1739.2011.01662.x

Frankham, R., Bradshaw, C. J. A., and Brook, B. W. (2014). Genetics in conservation management: revised recommendations for the 50/500 rules, red list criteria and population viability analyses. *Biol. Conserv.* 170, 56–63. doi: 10.1016/j.biocon.2013.12.036

Franklin, I. R. (1980). "Evolutionary change in small populations," *in Conservation Biology: An Evolutionary-Ecological Perspective*, eds M. E. Soulé and B. A. Wilcox (Sunderland, MA: Sinauer), 135–149.

Franklin, I. R., and Frankham, R. (1998). How large must populations be to retain evolutionary potential? *Anim. Conserv.* 1, 69–70. doi: 10.1111/j.1469-1795.1998.tb00228.x

Freedman, A. H., Gronau, I., and Schweizer, R. M. (2014). Genome sequencing highlights the dynamic early history of dogs. *PLoS Genet.* 10:e1004016. doi: 10.1371/journal.pgen.1004016

Funk, C. F., McKay, J. K., Hohenlohe, P. A., and Allendorf, F. W. (2012). Harnessing genomics for delineating conservation units. *Trends Ecol. Evol.* 27, 489–496. doi: 10.1016/j.tree.2012.05.012

Ghafouri-Kesbi, F., Eskandarinasab, M., and Hassanabadi, A. (2008). Investigation of genetic variability and inbreeding characteristics in a population of Zandi sheep. *Can. J. Anim. Sci.* 88, 409–417. doi: 10.4141/CJAS07126

Gossmann, T. I., Keightley, P. D., and Eyre-Walker, A. (2012). The effect of variation in the effective population size on the rate of adaptive molecular evolution in eukaryotes. *Genome Biol. Evol.* 4, 658–667. doi: 10.1093/gbe/evs027

Hagenblad, J., Olsson, M., Parker, H. G., Ostrander, E. A., and Ellegren, H. (2009). Population genomics of the inbred Scandinavian wolf. *Mol. Ecol.* 180, 1341–1351. doi: 10.1111/j.1365-294X.2009.04120.x

Hauser, T. P., and Loeschcke, V. (1996). Drought stress and inbreeding depression in *Lychnis flos-cuculi* (*Caryophyllaceae*). *Evolution* 50, 1119–1126. doi: 10.2307/2410652

Hayden, E. C. (2014). The $1,000 genome. *Nature* 507, 294–295 doi: 10.1038/507294a

Heber, S., Varsani, A., Kuhn, S., Girg, A., Kempernaers, B., and Briskie, J. (2013). The genetic rescue of two bottlenecked South Island robin populations using translocations of inbred donors. *Proc. R. Soc. B Biol. Sci.* 280, 20122228. doi: 10.1098/rspb.2012.2228

Hedrick, P. W., and Fredrickson, R. (2010). Genetic rescue guidelines with examples from Mexican wolves and *Florida panthers*. *Conserv. Genet.* 11, 615–626. doi: 10.1007/s10592-009-9999-5

Hedrick, P. W., and Kalinowski, S. T. (2000). Inbreeding depression in conservation biology. *Annu. Rev. Ecol. Evol. Syst.* 31, 139–162. doi: 10.1146/annurev.ecolsys.31.1.139

Hedrick, P. W., Peterson, R. O., Vucetich, L. M., Adams, J. R., and Vucetich, J. A. (2014). Genetic rescue in Isle Royale wolves: genetic analysis and the collapse of the population. *Conserv. Genet.* 15, 1111–1121. doi: 10.1007/s10592-014-0604-1

Herrero-Medrano, J. M., Megens, H.-J., Groenen, M. A. M., Bosse, M., Perez-Enciso, M., and Crooijmans, R. P. M. A. (2014). Whole-genome sequence analysis reveals differences in population management and selection of European low-input pig breeds. *BMC Genomics* 15:601. doi: 10.1186/1471-2164-15-601

Hill, W. G., and Kirkpatrick, M. (2010). What animal breeding has taught us about evolution. *Annu. Rev. Ecol. Evol. Syst.* 41, 1–19. doi: 10.1146/annurev-ecolsys-102209-144728

Hoffman, J. I., Simpson, F., David, P., Rijks, J. M., Kuiken, T., Thorne, M. A. S., et al. (2014). High-throughput sequencing reveals inbreeding depression in a natural population. *Proc. Natl. Acad. Sci. U.S.A.* 111, 3775–3780. doi: 10.1073/pnas.1318945111

Hoffmann, A. A., Chown, S. L., and Clusella-Trullas, S. (2013). Upper thermal limits in terrestrial ectotherms: how constrained are they? *Funct. Ecol.* 27, 934–949. doi: 10.1111/j.1365-2435.2012.02036.x

Hoffmann, A. A., and Parsons, P. A. (1997). *Extreme Environmental Change and Evolution*. Cambridge: Cambridge University Press.

Hoffmann, I. (2010). Climate change and the characterization, breeding, and conservation of animal genetic resources. *Anim. Genet.* 41(Suppl. 1), 32–46. doi: 10.1111/j.1365-2052.2010.02043.x

Hoffmann, I. (2013). Adaptation to climate change – exploring the potential of locally adapted breeds. *Animal* 7, 346–362. doi: 10.1017/S1751731113000815

Hogg, J. T., Forbes, S. H., Steele, B. M., and Luikart, G. (2006). Genetic rescue of an insular population of large mammals. *Proc. R. Soc. B Biol. Sci.* 273, 1491–1499.

Hull, R. (1963). Influence of disease on yield of sugar beet. *Ann. Appl. Biol.* 51, 516–518.

Ingvarsson, P. K. (2001). Restoration of genetic variation lost – the genetic rescue hypothesis. *Trends Ecol. Evol.* 16, 62–63. doi: 10.1016/S0169-5347(00)02065-6

Ingvarsson, P. K. (2002). Conservation biology: lone wolf to the rescue. *Nature* 420, doi: 10.1038/420472a

IUCN. (2014). *The IUCN Red List of Threatened Species. Version 2014.2.* Available at: http://www.iucnredlist.org. [assessed October 01, 2014].

Jamieson, I. G., and Allendorf, F. W. (2012). How does the 50/500 rule apply to MVPs? *Trends Ecol. Evol.* 27, 578–584. doi: 10.1016/j.tree.2012.07.001

Janhunen, M., Kause, A., Mäntysaari, E. A., Vehviläinen, H., Praebel, A. K., Järvisalo, O., et al. (2013). A novel breeding design to produce genetically protected homogenous fish populations for on-growing. *Aquac. Res.* 44, 1847–1859. doi: 10.1111/j.1365-2109.2012.03190.x

Jimenez, J. A., Hughes, K. A., Alaks, G., Graham, L., and Lacy, R. C. (1994). An experimental study of inbreeding depression in a natural habitat. *Science* 266, 271–273. doi: 10.1126/science.7939661

Johnson, W. E., Onorato, D. P., Roelke, M. E., Land, E. D., Cunningham, M., Belden, R. C., et al. (2010). Genetic restoration of the *Florida Panther*. *Science* 329, 1641–1645. doi: 10.1126/science.1192891

Jones, L. P., Frankham, R., and Barker, J. S. F. (1968). The effects of population size and selection intensity in selection for a quantitative character in *Drosophila*. II. Long-term response to selection. *Genet. Res.* 12, 237–248. doi: 10.1017/S001667230001185X

Joost, S., Bonin, A., Bruford, M. W., Després, L., Conord, C., Erhardt, G., et al. (2007). A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. *Mol. Ecol.* 16, 3955–3969. doi: 10.1111/j.1365-294X.2007.03442.x

Kantanen, J., Olsaker, I., Holm, L.-E., Lien, S., Vilkki, J., Brusgaard, K., et al. (2000). Genetic diversity and population structure of 20 North European cattle breeds. *J. Hered.* 91, 446–457. doi: 10.1093/jhered/91.6.446

Keller, L. F., Grant, P. R., Grant, B. R., and Petren, K. (2002). Environmental conditions affect the magnitude of inbreeding depression in survival of Darwin's finches. *Evolution* 56, 1229–1239. doi: 10.1111/j.0014-3820.2002.tb01434.x

Kim, E.-S., Cole, J. B., Huson, H., Wiggans, G. R., Van Tassell, C. P., Crooker, B. A., et al. (2013). Effect of artificial selection on runs of homozygosity in US Holstein cattle. *PLoS ONE* 8:e80813. doi: 10.1371/journal.pone.0080813

Kristensen, T. N., Barker, J. S. F., Pedersen, K. S., and Loeschcke, V. (2008). Extreme temperatures increase the deleterious consequences of inbreeding under laboratory and semi-natural conditions. *Proc. R. Soc. B Biol. Sci.* 275, 2055–2061.

Kristensen, T. N., Pedersen, K. S., Vermeulen, C. J., and Loeschcke, V. (2010). Research on inbreeding in the 'omic' Era. *Trends Ecol. Evol.* 25, 44–52. doi: 10.1016/j.tree.2009.06.014

Kristensen, T. N., and Sørensen, A. C. (2005). Inbreeding – lessons from animal breeding, evolutionary biology and conservation genetics. *Anim. Sci.* 80, 121–133. doi: 10.1079/ASC41960121

Kristensen, T. N., Sørensen, A. C., Sorensen, D., Pedersen, K., Sørensen, J. G., and Loeschcke, V. (2005). A test of quantitative genetic theory using *Drosophila* – effects of inbreeding and rate of inbreeding on heritabilities and variance components. *J. Evol. Biol.* 18, 763–770. doi: 10.1111/j.1420-9101.2005.00883.x

Landsverk, T., and Gamlem, H. (1984). Intestinal lymphangiectasia in the Lundehund. Scanning electron microscopy of intestinal mucosa. *Acta Pathol. Microbiol. Immunol. Scand. A.* 92, 353–362.

Larson, W. A., Seeb, L. W., Everett, M. V., Waples, R. K., Templin, W. D., and Seeb, J. E. (2014). Genotyping by sequencing resolves shallow population structure to inform conservation of Chinook salmon (*Oncorhynchus tshawytscha*). *Evol. Appl.* 7, 355–369. doi: 10.1111/eva.12128

Laws, R. J., and Jamieson, I. G. (2011). Is lack of evidence of inbreeding depression in a threatened New Zealand robin indicative of reduced genetic load? *Anim. Conserv.* 14, 47–55. doi: 10.1111/j.1469-1795.2010.00388.x

Leroy, G., Mary-Huard, T., Verrier, E., Danvy, S., Charvolin, E., and Danchin-Burge, C. (2013). Methods to estimate effective population size using pedigree data: examples in dog, sheep, cattle, and horse. *Genet. Sel. Evol.* 45, 1. doi: 10.1186/1297-9686-45-1

Li, M.-H., Stranden, I., Tiirikka, T., Sevon-Aimonen, M.-L., and Kantanen, J. (2011). A comparison of approaches to estimate the inbreeding coefficient and pairwise relatedness using genomic and pedigree data in a sheep population. *PLoS ONE* 6:e26256. doi: 10.1371/journal.pone.0026256

Liao, W., and Reed, D. H. (2009). Inbreeding-environment interactions increase extinction risk. *Anim. Conserv.* 12, 54–61. doi: 10.1111/j.1469-1795.2008.00220.x

Luikart, G., Ryman, N., Tallmon, D. A., Schwartz, M. K., and Allendorf, F. W. (2010). Estimation of census and effective population sizes: the increasing usefulness of DNA-based approaches. *Conserv. Genet.* 11, 355–373. doi: 10.1007/s10592-010-0050-7

Mara, L., Casu, S., Carta, A., and Dattena, M. (2013). Cryobanking of farm animal gametes and embryos as a means of conserving livestock genetics. *Anim. Reprod. Sci.* 138, 25–38. doi: 10.1016/j.anireprosci.2013.02.006

Marcos-Carcavilla, A., Moreno, C., Serrano, M., Laurent, P., Cribiu, E. P., Andréoletti, O., et al. (2010). Polymorphisms in the HSP90AA1 5' flanking region are associated with scrapie incubation period in sheep. *Cell Stress Chaperones* 15, 343–349. doi: 10.1007/s12192-009-0149-2

Markert, J. A., Champlin, D. M., and Gutjahr-Gobell, R. (2010). Population genetic diversity and fitness in multiple environments. *BMC Evol. Biol.* 10:205. doi: 10.1186/1471-2148-10-205

Marshall, T. C., and Spalton, J. A. (2000). Simultaneous inbreeding and outbreeding depression in reintroduced Arabian oryx. *Anim. Conserv.* 3, 241–248. doi: 10.1111/j.1469-1795.2000.tb00109.x

Mattila, A. L. K., Duplouy, A., Kirjokangas, M., Lehtonen, R., Rastas, P., and Hanski, I. (2012). High genetic load in an old isolated butterfly population. *Proc. Natl. Acad. Sci. U.S.A.* 109, E2496–E2505. doi: 10.1073/pnas.1205789109

McGovern, D. P. B., Gardet, A., Törkvist, L., Goyette, P., Essers, J., Taylor, K. D., et al. (2010). Genome-wide association identifies multiple ulcerative colitis susceptibility loci. *Nat. Genet.* 42, 332–339. doi: 10.1038/ng.549

Mech, L. D. (2013). The case for watchful waiting with Isle Royale's wolf population. *George Wright Forum* 30, 326–332.

Melis, C., Borg, A. A., Espelien, I. S., and Jensen, H. (2012). Low neutral genetic variability in a specialist puffin hunter: the Norwegian Lundehund. *Anim. Genet.* 44, 348–351. doi: 10.1111/age.12000

Merilä, J. (2014). Lakes and ponds as model systems to study parallel evolution. *J. Limnol.* 73, 33–45.

Meuwissen, T. (2007). Genomic selection: marker assisted selection on a genome wide scale. *J. Anim. Breed. Genet.* 124, 321–322. doi: 10.1111/j.1439-0388.2007.00708.x

Meuwissen, T. H. E., and Woolliams, J. A. (1994). Effective sizes of livestock populations to prevent a decline in fitness. *Theor. Appl. Genet.* 89, 1019–1026. doi: 10.1007/BF00224533

Mikkelsen, K., Loeschcke, V., and Kristensen, T. N. (2010). Trait specific consequences of fast and slow inbreeding: lessons from captive populations of *Drosophila melanogaster*. *Conserv. Genet.* 11, 479–488. doi: 10.1007/s10592-009-0030-y

Miller, J. M., Poissant, J., Hogg, J. T., and Coltman, D. W. (2012). Genomic consequences of genetic rescue in an insular population of bighorn sheep (*Ovis canadensis*). *Mol. Ecol.* 21, 1583–1596. doi: 10.1111/j.1365-294X.2011.05427.x

Nielsen, J. F., English, S., Goodall-Copestake, W. P., Wang, J., Walling, C. A., Bateman, A. W., et al. (2012). Inbreeding and inbreeding depression of early life traits in a cooperative mammal. *Mol. Ecol.* 21, 2788–2804. doi: 10.1111/j.1365-294X.2012.05565.x

Olson-Manning, C. F., Wagner, M. R., and Mitchell-Olds, T. (2012). Adaptive evolution: evaluating empirical support for theoretical predictions. *Nat. Rev. Genet.* 13, 867–877. doi: 10.1038/nrg3322

Ouborg, N. J., Pertoldi, C., Loeschcke, V., Bijlsma, R. K., and Hedrick, P. W. (2010). Conservation genetics in transition to conservation genomics. *Trends Genet.* 26, 177–187. doi: 10.1016/j.tig.2010.01.001

Pariset, L., Joost, S., Marsan, P. A., Valentini, A., and Econogene Consortium (EC). (2009). Landscape genomics and biased Fst approaches reveal single nucleotide polymorphisms under selection in goat breeds of North-East Mediterranean. *BMC Genet.* 10:7. doi: 10.1186/1471-2156-10-7

Pedersen, L. D., Sørensen, A. C., and Berg, P. (2009). Marker-assisted selection can reduce true as well as pedigree-estimated inbreeding. *J. Dairy Sci.* 92, 2214–2223. doi: 10.3168/jds.2008-1616

Pedrosa, V. B., Santana, M. L., Oliveira, P. S., Eler, J. P., and Ferraz, J. B. S. (2010). Population structure and inbreeding effects on growth traits of Santa Ines sheep in Brazil. *Small Ruminant. Res.* 93, 135–139. doi: 10.1016/j.smallrumres.2010.05.012

Pertoldi, C., Purfield, D. C., Berg, P., Jensen, T. H., Bach, O. S., Vingborg, R., et al. (2014). Genetic characterization of a herd of the endangered Danish Jutland cattle. *J. Anim. Sci.* 92, 2372–2376. doi: 10.2527/jas.2013-7206

Peterson, R. O., Thomas, N. J., Thurber, J. M., Vucetich, J. A., and Waite, T. A. (1998). Population limitation and the wolves of Isle Royale. *J. Mammal.* 79, 828–841. doi: 10.2307/1383091

Pfahler, S., and Distl, O. (2013). A massive reduction of the genetic diversity in the Lundehund. *Anim. Genet.* 45, 154. doi: 10.1111/age.12084

Phifer-Rixey, M., Bonhomme, F., Boursot, P., Churchill, G. A., Pialek, J., Tucker, P. K., et al. (2012). Adaptive evolution and effective population size in wild house mice. *Mol. Biol. Evol.* 29, 2949–2955. doi: 10.1093/molbev/mss105

Pimm, S. L., Dollar, L., and Bass, O. L. (2006). The genetic rescue of the *Florida panther*. *Anim. Conserv.* 9, 115–122. doi: 10.1111/j.1469-1795.2005.00010.x

Pirottin, D., Grobet, L., Adamantidis, A., Farnir, F., Herens, C., Schroder, H. D., et al. (2005). Transgenic engineering of male-specific muscular hypertrophy. *Proc. Natl. Acad. Sci. U.S.A.* 102, 6413–6418. doi: 10.1073/pnas.0502426102

Purfield, D. C., Berry, D. P., McParland, S., and Bradley, D. G. (2012). Runs of homozygosity and population history in cattle. *BMC Genet.* 13:70. doi: 10.1186/1471-2156-13-70

Quaresma, M., Martins, A. M. F., Rodrigues, J. B., Colaço, J., and Payan-Carreira, R. (2014). Pedigree and herd characterization of a donkey breed vulnerable to extinction. *Animal* 8, 354–359. doi: 10.1017/S1751731113002218

Qvigstad, G., Kolbjørnsen, Ø., Skancke, E., and Waldum, H. L. (2008). Gastric neuroendocrine carcinoma associated with atrophic gastritis in the norwegian lundehund. *J. Comp. Pathol.* 139, 194–201. doi: 10.1016/j.jcpa.2008.07.001

Räikkönen, J., Vucetich, J. A., Peterson, R. O., and Nelson, M. P. (2009). Congenital bone deformities and the inbred wolves (*Canis lupus*) of Isle Royale. *Biol. Conserv.* 142, 1025–1031. doi: 10.1016/j.biocon.2009.01.014

Reed, D. H., Fox, C. W., Enders, L. S., and Kristensen, T. N. (2012). Inbreeding–stress interactions: evolutionary and conservation consequences. *Ann. N. Y. Acad. Sci.* 1256, 33–48. doi: 10.1111/j.1749-6632.2012.06548.x

Reid, J. M., Arcese, P., and Keller, L. F. (2006). Intrinsic parent-offspring correlation in inbreeding level in a song sparrow (*Melospiza melodia*) population open to immigration. *Am. Nat.* 168, 1–13. doi: 10.1086/504852

Rice, A. M., Rudh, A., Ellegren, H., and Qvarnström, A. (2011). A guide to the genomics of ecological speciation in natural animal populations. *Ecol. Lett.* 14, 9–18. doi: 10.1111/j.1461-0248.2010.01546.x

Romanov, M. N., Tuttle, E. M., Houck, M. L., Modi, W. S., Chemnick, L. G., Korody, M. L., et al. (2009). The value of avian genomics to the conservation of wildlife. *BMC Genomics* 10:S10. doi: 10.1186/1471-2164-10-S2-S10

Schaeffer, L. R. (2006). Strategy for applying genome-wide selection in dairy cattle. *J. Anim. Breed. Genet.* 123, 218–223. doi: 10.1111/j.1439-0388.2006.00595.x

Schaeffer, L. R., Burnside, E. B., Glover, P., and Fatehi, J. (2011). Crossbreeding results in Canadian dairy cattle for production, reproduction and conformation. *TOASJ* 5, 63–72. doi: 10.2174/1874331501105010063

Sexton, J. P., Hangartner, S. B., and Hoffmann, A. A. (2014). Genetic isolation by environment or distance: which pattern of gene flow is most common? *Evolution* 68, 1–15. doi: 10.1111/evo.12258

Shafer, A. B. A., Wolf, J. B. W., and Alves, P. C. (in press). Genomics and the challenging translation into conservation practice. *TREE* doi: 10.1016/j.tree.2014.11.009

Siol, M., Wright, S. I., and Barrett, S. C. H. (2010). The population genomics of plant adaptation. *New Phytol.* 188, 313–332. doi: 10.1111/j.1469-8137.2010.03401.x

Sonesson, A. K., Goddard, M. E., and Meuwissen, T. H. E. (2002). The use of frozen semen to minimize inbreeding in small populations. *Genet. Res. Camb.* 80, 27–30. doi: 10.1017/S0016672302005712

Sørensen, A. C., Madsen, P., Sørensen, M. K., and Berg, P. (2006). Udder health shows inbreeding depression in Danish Holsteins. *J. Dairy Sci.* 89, 4077–4082. doi: 10.3168/jds.S0022-0302(06)72452-3

Sørensen, A. C., Sørensen, M. K., and Berg, P. (2005). Inbreeding in Danish dairy cattle breeds. *J. Dairy Sci.* 88, 1865–1872. doi: 10.3168/jds.S0022-0302(05)72861-7

Sørensen, M. K., Sørensen, A. C., Baumung, R., Borchersen, S., and Berg, P. (2008). Optimal genetic contribution selection in Danish Holstein depends on pedigree quality. *Livest. Sci.* 118, 212–222. doi: 10.1016/j.livsci.2008.01.027

Soulé, M. E. (1980). "Thresholds for survival: maintaining fitness and evolutionary potential," in *Conservation Biology: An Evolutionary-Ecological Perspective*, eds M. E. Soulé and B. A. Wilcox (Sunderland, MA: Sinauer), 151–169.

Soulé, M. E. (1985). What is conservation biology? *Bioscience* 35, 727–734. doi: 10.2307/1310054

Stearns, S. C., and Sage, R. D. (1980). Maladaptation in a marginal population of the mosquito fish, Gambusia affinis. *Evolution* 34, 65–75. doi: 10.2307/2408315

Strasburg, J. L., Kane, N. C., Raduski, A. R., Bonin, A., Michelmore, R., and Rieseberg, L. H. (2011). Effective population size is positively correlated with levels of adaptive divergence among annual sunflowers. *Mol. Biol. Evol.* 28, 1569–1580. doi: 10.1093/molbev/msq270

Swindell, W. R., and Bouzat, J. L. (2006). Reduced inbreeding depression due to historical inbreeding in *Drosophila melanogaster*: evidence for purging. *J. Evol. Biol.* 1257–1264. doi: 10.1111/j.1420-9101.2005.01074.x

Szulkin, M., and Sheldon, B. (2007). The environmental dependence of inbreeding depression in a wild bird population. *PLoS ONE* 2:e1027. doi: 10.1371/journal.pone.0001027

Taberlet, P., Coissac, E., Pansu, J., and Pompanon, F. (2011). Conservation genetics of cattle, sheep, and goats. *C. R. Biol.* 334, 247–254. doi: 10.1016/j.crvi.2010.12.007

Taberlet, P., Valentini, A., Rezaei, H. R., Naderi, S., Pompanon, F., Negrini, R., et al. (2008). Are cattle, sheep, and goats endangered species? *Mol. Ecol.* 17, 275–284. doi: 10.1111/j.1365-294X.2007.03475.x

Tallmon, D. A., Luikart, G., and Waples, R. S. (2004). The alluring simplicity and complex reality of genetic rescue. *Trends Ecol. Evol.* 19, 489–496. doi: 10.1016/j.tree.2004.07.003

Templeton, A. R. (1994). "Coadaptation, local adaptation, and outbreeding depression," in *Principles of Conservation Biology*, eds G. K. Meffe and C. R. Carroll (Sunderland, MA: Sinauer), 152–153.

Thomsen, B., Horn, P., Panitz, F., Bendixen, E., Petersen, A. H., Holm, L. E., et al. (2006). A missense mutation in the bovine SLC35A3 gene, encoding a UDP-N-acetylglucosamine transporter, causes complex vertebral malformation. *Genome Res.* 16, 97–105. doi: 10.1101/gr.3690506

Toro, M. A., Villanueva, B., and Fernandez, J. (2014). Genomics applied to management strategies in conservation programmes. *Livest. Sci.* 166, 48–53. doi: 10.1016/j.livsci.2014.04.020

Vilá, C., Sunquist, A. K., Flagstad, Ø., Seddon, J., Björnerfeldt, K. I., Casuli, A., et al. (2003). Rescue of a severely bottlenecked wolf (*Canis lupus*) population by a single immigrant. *Proc. R. Soc. B Biol. Sci.* 270, 91–97. doi: 10.1098/rspb.2002.2184

Vucetich, J. A., Peterson, R. O., and Nelson, M. P. (2012). Should Isle Royale wolves be reintroduced? A case study on wilderness management in a changing world. *George Wright Forum* 29, 126–147.

Vucetich, J. A., Peterson, R. O., and Nelson, M. P. (2013). Discernment and precaution: a response to Cochrane and Mech. *George Wright Forum* 30, 333–340.

Weber, K. E. (1990). Increased selection response in larger populations. I. Selection for wing-tip height in *Drosophila melanogaster* at three population sizes. *Genetics* 125, 579–584.

Weeks, A. R., Sgro, C. M., Young, A. G., Frankham, R., Mitchell, N. J., Miller, K. A., et al. (2011). Assessing the benefits and risks of translocations in changing environments: a genetic perspective. *Evol. Appl.* 4, 709–725. doi: 10.1111/j.1752-4571.2011.00192.x

Whitlock, B. K., Kaiser, L., and Maxwell, H. S. (2008). Heritable bovine fetal abnormalities. *Theriogenology* 70, 535–549. doi: 10.1016/j.theriogenology.2008.04.016

Willi, Y., Van Buskirk, J., and Hoffmann, A. A. (2006). Limits to the adaptive potential of small populations. *Annu. Rev. Ecol. Syst.* 17, 433–458. doi: 10.1146/annurev.ecolsys.37.091305.110145

Woelders, H., Windig, J., and Hiemstra, S. J. (2012). How developments in cryobiology, reproductive technologies and conservation genomics could shape gene banking strategies for (farm) animals. *Reprod. Dom. Anim.* 47, 264–273. doi: 10.1111/j.1439-0531.2012.02085.x

Wright, S. (1929). The evolution of dominance. *Am. Nat.* 63, 556–561. doi: 10.1086/280290

Wright, S. (1931). Evolution in mendelian populations. *Genetics* 16, 97–159.

Zwald, N. R., Weigel, K. A., Fikse, W. F., and Rekaya, R. (2003). Identification of factors that cause genotype by environment interaction between herds of Holstein cattle in seventeen countries. *J. Dairy Sci.* 86, 1009–1018. doi: 10.3168/jds.S0022-0302(03)73684-4

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 07 October 2014; accepted: 26 January 2015; published online: 10 February 2015.

*Citation: Kristensen TN, Hoffmann AA, Pertoldi C and Stronen AV (2015) What can livestock breeders learn from conservation genetics and vice versa? Front. Genet. **6**:38. doi: 10.3389/fgene.2015.00038*

This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics.

Copyright © 2015 Kristensen, Hoffmann, Pertoldi and Stronen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

HYPOTHESIS AND THEORY
published: 08 September 2015
doi: 10.3389/fgene.2015.00279

Changing values of farm animal genomic resources. from historical breeds to the Nagoya Protocol

*Sakari Tamminen**

Department of Social Research, University of Helsinki, Helsinki, Finland

The paper reviews the history of Animal genetic resources (AnGRs) and claims that over the course of history they have been conceptually transformed from economic, ecologic and scientific life forms into political objects, reflecting in the way in which any valuation of AnGRs is today inherently imbued with national politics and its values enacted by legally binding global conventions. Historically, the first calls to conservation were based on the economic, ecological and scientific values of the AnGR. While the historical arguments are valid and still commonly proposed values for conservation, the AnGR have become highly politicized since the adoption of the Convention of Biological Diversity (CBD), the subsequent Interlaken Declaration, the Global Plan for Action (GPA) and the Nagoya Protocol. The scientific and political definitions of the AnGRs were creatively reshuffled within these documents and the key criteria by which they are now identified and valued today were essentially redefined. The criteria of "*in situ* condition" has become the necessary starting point for all valuation efforts of AnGRs, effectively transforming their previous nature as natural property and global genetic commons into objects of national concern pertaining to territorially discrete national genetic landscapes, regulated by the sovereign powers of the parties to the global conventions.

Keywords: values, animal genetic resources, convention on biological diversity, interlaken declaration, global plan of action, national genetic landscapes

OPEN ACCESS

Edited by:
Juha Kantanen, Natural Resources Institute Finland, Finland

Reviewed by:
Philippe V. Baret, Université Catholique de Louvain, Belgium
Michèle Tixier-Boichard, Institut National de la Recherche Agronomique, France

***Correspondence:**
Sakari Tamminen, Department of Social Research, University of Helsinki, Unioninkatu 37, 2nd Floor, 00014 Helsingin, Finland
sakari.tamminen@helsinki.fi

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: *02 February 2015*
Accepted: *20 August 2015*
Published: *08 September 2015*

Citation:
Tamminen S (2015) Changing values of farm animal genomic resources. from historical breeds to the Nagoya Protocol. Front. Genet. 6:279. doi: 10.3389/fgene.2015.00279

Animal Genetic Resources as a Global Matter of Concern

Animal genetic resources (AnGRs) have become a topic of renewed interest in international politics taking agricultural species as its object of concern for the last decade. The reasons for this point to a number of intertwined reasons. The first has to do with the unclear legal status and the scope of regulation stemming from biodiversity agreements targeting a wide range of animal species from wild to agricultural, their biological materials and genetic resources within the global politics of late 20th and early 21st century. Here, states, nations and indigenous communities have become new key stakeholders of genetic resources as they have been granted sovereign rights over territorially bound, native "*in situ*" resources within the text of the convention on biological diversity (CBD)[1], signed by over 150 states in 1992. The sovereign rights over GRs have been subsequently re-enforced with the adoption of the Nagoya Protocol on Access and Benefit-sharing[2] entering into force on October 2014. The Nagoya Protocol is the legally binding protocol guiding how to interpret and act upon the genetic resources issues presented in the CBD over 20 years earlier.

[1]https://www.cbd.int/
[2]http://www.cbd.int/abs/

The second, and inherently nested interest relates to the way in which access and animal could, or should, be regulated internationally under biodiversity frameworks, including debates on how different kinds of genetic resources from plants to animals and from wild to agricultural species differ from each other and how the difference might have to inform the practical execution of their global governance. The Food and Agriculture Organization of the United Nations (FAO) established a Commission on Plant Genetic Resources in 1983 to deal with policy, access and benefit sharing issues related plant genetic resources. FAO did broaden the mandate of the Commission in 1995 to cover all aspects after the CBD entered into force and after recognizing "that broadening the coverage of the Commission would allow the Organization to deal in a more integrated manner with agrobiodiversity issues" (Food and Agriculture Organization, 1995, p. 66). 2 years later in 1997, the Commission also established separate working groups for animal and plant genetic resources, followed by one expert group for forest genetic resources[3]. All these Committees—specifically established for different types of genetic resources—demonstrate how difficult the policy, ownership and access and benefit sharing issues related to GRs, and especially to AnGRs, are to understand, let alone to manage in practice.

Three examples from international analyses from the last 10 years will clarify some of the difficult aspects related to global agreements and governance on agricultural AnGRs and open up good questions how did the AnGRs become so politically contested objects of agricultural nature.

Consider, for example, a report from 2006 exploring policy options for the "Exchange, Use and Conservation of Animal Genetic Resources," commissioned by the FAO and funded by the Government of the United Kingdom of Great Britain and Northern Ireland, recognized that fundamental tension between the traditional ownership of AnGRs and new global conventions had emerged (Hiemstra et al., 2006) and this tension needs to be resolved on international level. Going through a number of options for AnGR regulation, tellingly the report ended with the summarizing paragraph claiming that "*[c]lassical ownership' of AnGR includes physical ownership and communal 'law of the land' affecting livestock keeping and breeding. There is an increasing tension with developments in the realms of biodiversity law and intellectual property rights protection. Demarcation of these different rights systems and maintaining equity among different stakeholders is crucial to avoiding conflict and increased transaction costs. In this context, it is important to consider the rights of livestock keepers/breeders vis-à-vis national level sovereign rights, as well as obligations between patent holders and breeders/livestock keepers.*" (Hiemstra et al., 2006, p. 37).

The report had been commissioned as FAO wanted to clarify the options on how to navigate the world of new political and legal frameworks after CBD for AnGR management. 3 years later another expert report on AnGR raised a concern that relates to the fact that all different types ownership relations faced now a potential disruptive element. "*Private or communal ownership of AnGR, is potentially at least, challenged by national sovereignty over genetic resources. Individual owners may find that their rights to sell breeding animals or other genetic material, particularly across national boundaries, are restricted. Those seeking to buy specific AnGR may find that they are unable to do so, or that they can only do so on terms that are acceptable not only to the owner of the resources but also in compliance with national legislation.*"(Commission on Genetic Resources for Food and Agriculture, 2009, p. 29).

Finally, in November 2014, the Intergovernmental Technical Working Group on Animal Genetic Resources for Food and Agriculture concluded after its meeting that more work on AnGR is needed, but that at least the different types of *utilization* of AnGR, the criteria and ways in which the *country of origin* of AnGRs is assessed, and *access and benefit sharing policies* all need further clarification, although at global level a number of internationally binding legal treaties exist (Commission on Genetic Resources for Food and Agriculture, 2014, pp. 19–25).

For a long time, animals, breeds and their genetic resources were solely governed by rights that were based on physical access and use rights to animals as animals and breeds were seen as "wholes," either as living animals or recorded breeds. A mix of private, semi-private and common ownership models for agricultural and farm animals have been in use, and these have also generated much discussion about the forms of entitlement over the life of the animals and the best possible ways to organize these relations (Hardin, 1968; David, 2011). However, as biotechnologies used in animal production have developed—increasing animal growth rates and carcass composition, enhancing disease resistance and improving hair and fiber production (Wilmut et al., 1992; Wheeler et al., 2010; Wheeler, 2013)—the value of individual farm animal, or even the value of a breed, is not only solely calculated in direct relation to its output of agricultural goods (e.g., meat, milk) but also by its value within the social system of breeding. Thus, farm animals and breeds are valuable also because their capacity to produce particular kind of offspring, or to transmit valuable features encoded within the DNA. This capacity can be either codified in rough ideas of maintaining a pure breed type or within the sophisticated algorithms calculating the Estimated Breeding Value in modern farms based on the development of the herdbook, an innovation that enables population management through exact recordings innovated in late 18th and early 19th centuries[4].

Given that the two sources of value in farm animals have been recognized for over 200 years, it is surprising that at the present, in 2015, the global community dealing with AnGRs has ended up in a situation where the access, ownership rights and benefit sharing issues—issues that for a long time remained

[3]The names of these expert bodies are: the Intergovernmental Technical Working Group on Animal Genetic Resources for Food Agriculture, the Intergovernmental Technical Working Group on Plant Genetic Resources for Food and Agriculture to deal with specific matters in their areas of expertise, and the Intergovernmental Technical Working Group on Forest Genetic Resources.

[4]General herdbooks emerged in Europe in the late eighteenth—and early nineteenth century—the first one for cattle was the "*Short-Horned Cattle Herd*" book, published in 1822 in England. Elsewhere, general registries were published in France (1855), Germany (1864), Holland (1874), and Denmark (1881) (Derry, 2003, p. 8; Walton, 1999, p. 153; Ritvo, 1995, p. 420). The idea of "Estimated Breeding Value" is based on this herdbook keeping but introduces a more refined statistical modeling into the calculation of breeding value of a individual animal at the end of C21th.

unchallenged—have become a matter of global concern and source of slowly proceeding political processes where there are no easy resolutions. The issue of "sovereignty" over all types of genetic resources described in the major international agreements gives the signatory states relatively free hands to develop and implement national laws and regulations. In fact, to fulfill their sovereignty over AnGRs, for example, the states must decide what types of entitlements and relationships over AnGRs they should implement, and how this relates to the national rights that farmers have over their animals, for example.

This paper presents two questions and two hypotheses on nature and status of the AnGRs in the post-CBD world:

(1) First, why are the issues of AnGRs for agriculture debated alongside more general issues of biological diversity;

and

(2) Second, how should we understand that idea of sovereignty over genetic resources given to the signatory states of the CBD and the Nagoya Protocol in the context of AnGRs?

The first hypothesis builds on the cultural history of AnGRs movement in the political institutions, most notably in the FAO. I claim that early warnings about the need for the conservation and coordinated management of AnGRs for agriculture did not lead to action and resulted in a failure to mobilize larger communities to action. This, in turn, lead the animal geneticists affiliated with FAO and other interests parties to join forces with environmental conservation movement, especially United Nations' Environmental Programme (UNEP), to gain international support to the issue of conservation then seen as a agenda priority.

Second, and following from the historical reason explained above, the way in which CBD and to certain extent the subsequent Nagoya Protocol defined and understood genetic resources owes much to the world of plant genetic resources (PGR). Defining the right and obligations of signatories through the PGR leads implicitly to the world of plant breeding, which operates differently from animal breeding practices, the key economic relations, and related biological processes. This is also why the key articles and provisions in the biodiversity conventions are couched in strong terms under national governments' sovereign powers.

I claim that this has resulted in a world, where we have moved from a system where animals once were part of a seamless universal nature without political boundaries, to a world that is a collection of discrete "national genetic landscapes" safeguarded by state policies and legal provisions.

The Short Institutional History of AnGR Concerns

The management of farm AnGRs become a topic immediately after the establishment of the United Nation's Food and Agriculture Organization (FAO), however, the concerns related to genetic resources were first identified as a challenge for developing countries. The negative consequences of modern animal production aiming at increasing animal productivity started to raise doubts within the scientific community, and the first calls for genetic conservation followed quickly. Phillips, the first Deputy Director-General for the FAO remembers how he, as the first employeed of the animal section for FAO, had all his career "*already carried out activities relating to animal genetic resources. . .and the Organization's involvement in this work dates back to 1946*" (Phillips, 1981, p. 5). From early on the worry was about losing local breeds to extinction in developing countries. Local animals were replaced with globally homogenized and more productive breeds that became easily available and were adopted at fast pace. Despite the early warning calls, little to no action aimed at conservation ensued at global level even if FAO produced a number of scientific reports and hosted a series of meetings around the issue between the early 1950s and 1960s.

It was only after the widespread negative impacts of Green Revolution became evident in the 1960's that AnGRs became truly a global matter of concern also for scientists working within developed countries. This was the direct result of unmanaged use of new breeding techniques combined with shrinking and homogenized ecological habitats. For example, in the 1969 regional meeting of the European Association for Animal Production, the issue for "gene pool losses" was already clearly articulated by Maijala (1971) who also identified the root cause for these losses: "*The present era of frozen semen. . .has reactualized the problem of gene losses. . .The problem arises mainly from the fact that an effective utilization of the best animals of today automatically means setting aside the poorer animals, strains, breeds and even species*" (Maijala, 1971, pp. 403–444).

In response to these developments, FAO and the United Nations Environment Programme (UNEP) launched a joint project in 1974 with the title of "*Conservation of animal genetic resources*." It had the key objective to "prepare a list of breeds of farm animals in danger of extinction together with an account of any measures which have been recommended or taken to prevent this extinction" (Mason, 1981, p. 17). A Consultation Report followed in Mason (1981) with a review of the work achieved by the project through the participating regional and national organizations, and made recommendations for future action.

This report was presented in a workshop for animal geneticists working with genetic resources and was framed with Phillip's opening words, that simultaneously exhibited hope and exasperation on the current state of affairs. He proclaimed: " *I am pleased to bid you welcome here, on behalf of the Director—General. It is indeed heartening to see such a distinguished group of animal geneticists assembled to consider the problems of identification, conservation and effective management of animal genetic resources. This is matter critical to man's future, yet it has had little recognition and little real attention* (Phillips, 1981, p. 2)." Phillips' opening speech betrays how, by the early 1980's, the animal scientists had been awakened to the dire straits of genetic resources but the political support of the issue was still weak and more generally, unrecognized as an global political issue. It did not appear in the general global agendas as did other issues related to modernization and increase of production, such as the environmental movement which had started already in 1970s to attract more political attention and gained fast political weight in the international political arenas. As a result, the issue for farm AnGRs did not spur action nor attract funding for conservation efforts (Boyazoglu and Chupin, 1991).

Fast forward a decade into the early 1990's, and one finds more explicit frustration toward the slow progress on conservation efforts and lack of coordinated international action. Explaining the issue and need for AnGRs conservation Hodges (1990), a Senior Officer at FAO wrote that "*the time for technical talk is over. The issues are clear. What is now needed is an effective international decision to provide funds to do what all agree is now necessary the global, regional and national levels*" (Hodges, 1990, p. 153). AnGRs needed more political support but this proved to be hard to gain without rethinking and reframing the issue, and joining forces with other institutional actors. International action did finally follow a few years later in 1992 when the FAO joined forces with the UNEP, and co-organized the Rio Earth Summit in Rio de Janeiro, Brazil. This was also the historical moment for AnGRs. This is the place and the time where genetic resources became newly articulated as parts of nature as they were linked directly to the recently introduced concept of "biodiversity," the key theme of the global meeting for the world's leaders.

In the meeting, UNEP and FAO introduced the global CBD, a convention aimed at saving biodiversity, for larger public and opened it for signatures. It was undersigned by some 160 countries at Rio de Janeiro and over 30 other countries followed suit during the upcoming years. Several of the Articles included in the Convention, addressed the issue of genetic resources directly and introduced an obligation to identify, report and take appropriate actions to conserve genetic resources. The long follow-up work finally resulted in The Global Plan of Action for Animal Genetic Resources (GPA), adopted in 2007, and the Guidelines on the Preparation of national strategies and action plans for AnGRs, published in 2009.

Given the half-century history of AnGR's as a matter of concern for animal geneticists and long and idle wait for political action, the key question is why did the wide scale political traction to save genetic resources only emerge with the introduction of the CBD, leading to the global and national action plans and guidelines specific to AnGRs over a decade later?

Early Failures in Valuation

The reason why the FAO and the regional institutions such as the EAAP failed in gaining political traction with their early alarms about the need for conservation measures relates to two shortcomings in the definition and the valuation of AnGRs.

First shortcoming was the lack of consensus in scientific definition, valuation and prioritization of AnGRs that would lead into simple and uniform action recommendations. The question on what is it exactly that needs to be conserved and how to prioritize the required conservation actions was left open, or at best was illustrated through a case of a few particular breeds. The second, and more important shortcoming was the failure of global political and legal identification of the responsible parties and beneficiaries of any value deriving from the costly conservation actions. This, in turn, relates to the fact that up until the CBD, AnGRs were treated as a mixture of "private" and "commons," or as "club commons" (David, 2011) to be shared and used, subject only to individual farmers' and breeding associations' property right regimes and explicit regulations at country level.

After the introduction of the CBD the legal status of AnGRs changed globally as they were politically identified falling under the sovereign power of the signatory parties to the Convention—a major change and complication in access and benefit sharing relations that was later affirmed by the GPA in 2007 and later by the Nagoya Protocol. Understanding the latter is especially important as this understanding exposes the new overarching paradigm under which the value most of the AnGRs today are to be governed.

First, the failure to provide a clear direction for conservation relates to the arguments about the overall role of different kinds of AnGRs in animal production. When the concept of AnGRs were first introduced among the animal scientist, they were framed in and through two different ways (both scientifically informed) of demonstrating the role of AnGRs in animal production. The two ways—the "utilizationist" and "conservationist" standpoints—literally attributed the value of AnGRs in animal production in two incommensurable ways (and to some extent this debate still continues even today). Hodges (1984) report on genetic resources explained the main differences between the two approaches:

"*The utilizationist's primary concern is the immediate usefulness of available genetic resources to improve livestock populations. . .The loss of breeds as distinct identities is not generally a concern, as long as the genes that make these breeds potentially useful are retained in the commercial stocks. . .The preservationist's primary objective is long-term conservation of genetic resources for future use. This view emphasizes the value of preserving the widest possible spectrum of genetic diversity to be prepared for unpredictable changes of future needs. The greatest possible number of breeds are to be preserved as purebreds.*" (Hodges, 1984).

The differences of these two views boil down to conserving "the known useful genes" in one form or another versus conserving the "genetic diversity of whole animal breeds" to hedge the uncertainty deriving from unknown future needs. The first approach aims to save the sliced and diced, functionally valuable component of animals regardless of its "breed"; the other also the animal breeds in the purebred form and to maximize diversity as an insurance policy against future unknowns. Although analytically distinct from animals or breeds, the animal scientists first presented the issue of AnGR conservation as a choice between *isolated genetic components* immediately useful in the production of high performance animals or as the maximization of genetic diversity by the conservation of *local breeds in their animal forms*. In these two approaches, AnGR's are conceptually presented as different objects of conservation and seen valuable for different purposes[5].

Second, the failure to identify parties responsible for the conservation irked the conversations as this related directly to the economics of conservation, or more generally, to the political

[5]There are a number of ways to maximize diversity. Conserving a sum of isolates of pure inbred populations will allow saving rare genetic combinations adapted to specific environmental conditions but might result in losing overall diversity. Other options, such as maintaining a large out bred population resulting from crossbreeding, would also provide a large diversity but is not usually the overall aim of conservation programs. It is today generally recognized that a combination of ex-situ and in-situ measures are complementary strategies.

economy of global animal production. The problem was captured in a report produced by the United States' board of Agriculture in its "Managing global livestock genetic diversity":

"*The concept of conservation. . .is complex. One can think of live animals, being preserved in situ, or in some semi-artificial situation; alternatively one may think of cryogenic storage of sperm or fertilized ova or other tissues or gene segments. The economic problems are difficult with both live animals and with haploid or diploid cells. Who is to pay? There are also questions of how many to preserve, for how long, and where*" (National Academy of Science, 1993, p. 3).

Although plant varieties and their genetic material were protected by various intellectual property systems since 1930s (see Kloppenburg, 1988), AnGRs were used by farmers and breeder associations alike without generalized and specified rights or restrictions imposed at global level. Since there was no definition on the ownership rights over the genetic materials of animals, the global attribution of conservation responsibility through political processes proved to be impossible without more specific consideration. Yet, for pigs and chicken, ownership and responsibility questions have been more straightforward. This reflects what Tvedt et al., 2007, p. 8) note of the legal protection farm animals and their protection in general, and chicken and pig in particular:

"*For farm animals there are strong biological and physical means of protection available: The owner of the animal can more easily than the plant breeder have an overview and control over who is receiving genetic material from his animals or his population. For poultry and pig breeding, however, where farmers often buy hybrids whose genetics are more difficult to reproduce. The sale of hybrids is thus an important strategy for maintaining physical control over the genetic material by physical control over the material. For other breeds, in particular cattle, the physical ownership is often combined with a register, a herd book that maintains a protocol for the generations of animals fulfilling the criteria for registration.*"

This is why the different claims about the value of AnGRs and the need for their conservation, made by both the "utilizationists" and "conservationists," rang to deaf ears outside the animal scientist circles. The failure to spur action was not based on the scientific challenge to demonstrate the value of AnGRs in animal production or the lack of consensus in setting the priorities of conservation. Instead, and above all, it was a problem of political economy: who is to pay? And even more importantly, who is to benefit?

Global Re-framing of AnGRs

Food and Agriculture Organization remained active on AnGRs since the FAO/UNEP consultation program in 1980, established a Committee on Agriculture that kept reminding about the issue at the FAO Council level; designed a FAO expert consultation round on AnGRs in 1989 and in 1992 (Food and Agriculture Organization, 1990, 1999; Steane, 1992). What become clear over the years was that a global binding framework was needed.

Anticipating the global political agreement on AnGRs, de Haen (1992), Assistant Director-General of the Agriculture Department of FAO wrote in 1992 that "*it is clear that there is a greater awareness that a framework for the management of global animal genetic resources must be established. It is most appropriate that this Expert Consultation is taking place now in the context and timing of the Earth Summit, the United Nations Conference on Environment and Development (UNCED) to be held in Brazil in about eight weeks time*" (de Haen, 1992, p. 3).

The first reframing of the AnGR's came in the form of the global CBD a few months later. The Convention had been long in preparation and FAO had been involved in its drafting phases influencing, among other issues, the inclusion of genetic resources to and their definition in the Convention text. There were two important re-framings in the Convention. First was the definition of the genetic resources, as genetic material of "actual or potential value" (CBD, Article 2). This definition bridged the two different views on the valuable material to conserve, or the "utilitizationist" and "preservationist" standpoints. Genetic resources become genetic material that could be attributed with demonstrable or imaginable value. But the question then arises: who has the right to attribute any value claims to AnGRs?

The other reframing answered to this question. Under the definitions of the Article 2 and the Article 15, genetic resources found "*in situ*" within the territory of a signatory were identified as belonging under the sovereign power of signatory states representing the nations of the world, reframing their ownership relationships globally. This is how CBD enacted an important political redefinition of genetic resources: previous problems in the definition of the value of nonhuman life were re-articulated through the politics of nationhood, in the idea of national differences found within the CBD's vision of genetic nature.

With the convention, also AnGRs became tightly nested within the sovereignty of nation-states and their geography. A reversal of the old idea of nations being rooted in natural differences of human populations took place—nonhuman populations, conceptualized as "genetic resources," could now be identified and placed under national or international jurisdiction in terms of their geographical location and the political powers that represented the nationhood that governed that geographical area. A global cartographic demarcation of nonhuman life took place as these novel objects of nature were grafted to the foundations of national sovereignty. They became a new part of the body of nations, a novel form of nonhuman nationhood.

The convention assumes significant amount of power over AnGR and their governance to signatory nation-states. Tvedt et al., 2007, p. 24) interpret the convention and its provisions in the following manner: "*The CBD presupposes the right of a country to exercise sovereign control over its AnGR (accompanied by a number of responsibilities). From the perspective of an exporting country, one of its main concerns is to maintain any property rights it may wish to retain over the AnGR after the resources have left the country. Similarly, it may wish to ensure that the rights of the exporter are respected by the buyer/importer of the AnGR. The most prominent rationale for a country to regulate export of AnGR would be to secure a right over that particular material in the future, including preventing that countries or companies gain control over these resources (e.g., through patenting or other forms of intellectual property rights), which might reduce the value of it in the exporting country.*"

This reframing introduced a whole new system where the value of any animal breed will be decided by the nations signatory to the parties but without any common reference what consists legitimate value claim over the material, except the condition of "*in situ*." In the CBD, these are the "conditions where genetic resources exist within ecosystems and natural habitats, and, in the case of domesticated or cultivated species, in the surroundings where they have developed their distinctive properties." (CBD Article 2). This "*in situ*" condition of valuable genetic resources has tremendous effects into how AnGRs are seen within the post-CBD world, especially as AnGRs were now removed from the idea of being freely circulated or tradable objects of nature. They stopped being global commons and instead become subject to the political powers of the Convention parties, many who did not have, and still today do not have, a clear stance what are "valuable" AnGRs to them, and how they will enact their sovereign powers over the access and benefit sharing to the valuable AnGRs. A definitional and legal disorientation followed.

The third re-framing of AnGR's happened as they were presented through ideas derived from the plant and crop worlds. The FAO background study in the CBD on the "Exchange, Use and Conservation of Animal Genetic Resources" acknowledged this as a major problem. It explained:

"*[a]lthough current debates regarding agricultural genetic resources have largely had a crop/plant focus, these discussions, and the international instruments or agreements that are emerging have tended to frame the debate for AnGR as well. At first sight plant breeding does not differ much from animal breeding. The genetics of plants and animals are based on the same principles. Plant and animal breeders both need genetic diversity in order to advance and the genetics determine adaptation to particular agro-ecological circumstances, as well as product qualities to a large extent. However, plant varieties can be protected by plant breeder's rights (UPOV), which is not the case for animal breeds/strains. Plant breeders aim at the development of new uniform varieties that are defined by certain phenotypic traits that can identify them from other varieties. Farm animal breeding is largely based on the selection of individuals within populations rather than selection between populations or strains. Farm animal breeders are interested in individual animals (within populations/breeds), while the whole population of a plant variety (clones) is the main focus of plant breeders*." (Hiemstra et al., 2006, p. 22).

The third reframing, then, pointed to the difference of animal and plant genetic resources as biological bred resource and legal protected asset: animals might carry interesting genetic traits but it is difficult to exploit one unique genetic characteristic, there are no large international breeding centers but most breeding happens in farms—except for poultry and partly for pigs—and the centers of origin or diversity for AnGR are not as clearly defined as for plants. Most importantly, farmers are not protected by internationally binding rights frameworks—plant breeders, however, are by the International Union for the Protection of New Varieties of plants (UPOV)[6]. The differences between plant and AnGRs make it hard to enforce only system for the two, however, the CBD does exactly this by enforcing the sovereignty of the signatory states as its starting point for rights and obligations via the discourses mostly appropriate to plant genetic resources.

These re-framings of the AnGR dictate much of how global action now unfolds. 15 years after the CBD, in 2007, the state representatives adopted the first "Global Plan of Action for Animal Genetic Resources" (GPA) at the Interlaken Conference held in Switzerland, something that was called a "historical breakthrough" by the FAO Director General Jacques Diouf (Food and Agriculture Organization, 2007, p. iii). The GPA includes the "Interlaken Declaration on Animal Genetic Resources," in which the sovereign rights of states over their AnGRs for food and agriculture was restated (declaration point 2).

In situ, Transboundary, and Domestic Applications

The fact that animals can move across politically established boundaries created a potential problem to these sovereign rights, however, and led to new politically innovated categories of AnGRs, such as "transboundary" species for criss-cross institutionalized country borders. The GPA explained: "*Assessing the status of animal genetic resources on a global scale presents some methodological difficulties. In the past, analysis of the Global Databank to identify breeds that are globally at risk was hampered by the structure of the system, which is based on breed populations at the national level. To address this problem. . .a new breed classification system was developed. Breeds are now classified as either local or transboundary, and further as regional or international transboundary*" (Food and Agriculture Organization, 2007, p. 13).

With these political documents not only did animals considered as genetic resources become "national" pertaining to a state, but some of them also became "transboundary," regionally and internationally. The result of this is that political categories infuse with conservation science categories because of the political economy involved in the ownership rights over the actually or potentially valuable genetic resources.

These categories are as much politically informed as they are scientifically true. The definitions of "*in situ*" or "transboundary" are inherently related to the political cartographic demarcation of the natural ecologies of domesticated animals, pointing to the deep connection between politics of value and the science of conservation of farm genetic animal resources. This is what eventually created the incentive for nation-states to act upon the issue of genetic erosion of animal populations, but is now, at the same token, generating new challenges that are beyond the scope of animal scientists or even international organizations to solve.

This complexity is reflected on how the national legislations have been drafted and implemented. Writing about the challenges in the implementation of the CBD legal experts Buck and Hamilton claim that "*[t]he complex subject matter of ABS, its potentially far reaching impact on uses of genetic resources and related information as well as the lack of detail in Articles. . .have all combined to result in a very low level of domestic implementation by Contracting Parties to the CBD. By 2007, only 39 of the then 189 Contracting Parties had established domestic legislation or were in the process of doing so*." (Buck and Hamilton, 2011, p. 48).

Reporting back on the negotiations of the Nagoya Protocol in Japan in 2010, the protocol that is to meant to clarify the

[6]http://www.upov.int

TABLE 1 | The historical changes in the value system for animal genetic resources.

Animal genetic resources	Pre-CBD/GPA/Nagoya	Post-CBD/GPA/Nagoya
Nature of animal genetic resources	• Natural breeds or functional genetic components • Identification based on scientific definition and evaluation of conservation need	• Natural-Cultural objects • Identification based on the politically agreed "*in situ*" condition • Genetic material found within the geographically bounded territories of the nation-states, with the exception of politically innovated "transboundary" conditions and categories of animals
Access and benefit sharing	• Private access, or club commons • Freely usable for those who have physical access and local permission to use • No enforced benefit sharing	• National sovereignty to decide • Usable only under the rule of the sovereign party to the conventions • Local political decision on access and benefit sharing principles.
Criteria for valuation	• Actual or potential value, no consensus in general	• Actual AND potential value • Based on territorial *in situ* condition and local political valuation of important national genetic landscapes to conserve

initial CBD, they point out that the key to really "understanding" the real effects of the CBD and Nagoya is dependent on how national governments use their sovereign powers: "*The adoption of the Nagoya Protocol was a major achievement in international biodiversity policy making in 2010. . .Further international work preparing the entry into force of the Nagoya Protocol will be needed. However, most efforts over the coming years will need to be at domestic level, developing implementing rules to prepare ratification. In all Parties with well-developed or emerging research and development systems this will require significant awareness-raising with stakeholders from research and industry and will result in quite some discussions*." (Buck and Hamilton, 2011, p. 60).

Most importantly, the national implementation has to take into account that access should take place on "mutually agreed terms" and "be subject to prior informed consent," conditions found in the original CBD and all subsequent treaties. However, other aspects of AnGR can also be regulated, and some of the countries have enacted already requirements for animal genetic material import and export. FAO's Technical Working Group on AnGR Access and Benefit sharing issues explained in its recent report in 2014 that "*[t]he sovereign right of states to determine access to genetic resources should not be confused with other categories of entitlement, such as the private ownership of an animal. A farmer's ownership of an animal may be conditioned by certain laws. For example, animal welfare legislation may regulate the handling, husbandry and transport of the animal. Other laws may require the animal to be vaccinated against specific diseases, and so on. In a similar way, ABS measures may require that, even though an animal is the private property of a farmer or the collective property of a community, certain conditions (e.g., related to the need for "prior informed consent") must be met before it can be provided to a third party for research and development*" (Commission on Genetic Resources for Food and Agriculture, 2014, Item 18).

Indeed, some of the countries have already exercised their sovereign rights. For example China has adopted a set of rules to AnGR, or "*Measures of examination and approval of the entry and exit of animal genetic resources and the research in cooperation with foreign entities in their utilization*," in 2008. These include a set of import and export rules, such as prohibition "*on the export of newly discovered and unverified*" AnGR in cooperation with "*any foreign institution of individual*." Also, any research and use of AnGR involving foreign collaborators requires permission from the Chinese authorities. South-Africa, on the other hand, now requires a "*genetic impact assessment*" before the import of new breeds. These studies need to be prepared by reputable South African animal scientists and submitted to the relevant authorities (see Commission on Genetic Resources for Food and Agriculture, 2009, p. 34). The national implementation of the sovereign rights over genetic resources can happen in many ways, not only by regulating of access or benefit sharing but also by use and impact, as the examples from China and South Africa demonstrate.

Conclusion

The challenges are located now within the realm of national politics where the *in situ* condition of genetic resources are turning animals into collections of nationally valuable animals, governed not by the previous ideals of global commons but by the logic of "actual and potential" value, by innovated political re-categorizations of natural beings, and by national restrictions to the access, use and benefit sharing of AnGRs. We have moved from a world where animals once were part of a seamless universal nature without boundaries to a world that is a collection of discrete "national genetic landscapes."

Over the course of the short history of AnGR conservation, the natural identities of farm animals have been refashioned from being objects of breeding to boost the productivity of individual animals and breeds to objects that can be defined actually or potentially valuable as nationally recognized genetic resources. The change in their identity is a creative outcome product of the animal breeding and conservation sciences that have argued the value of animals on the basis of scientific evidence as well as the global politics surrounding the ownership rights over genetic resources considered valuable. AnGRs, including farm animals, are now as much political as they are scientific, as much "cultural" than they are "natural" by their essence.

Table 1 above summarizes the key changes in the conceptualization and valuation of AnGRs before and after the introduction and ratification of the CBD, the Global Plan of Action and the Nagoya Protocol. What becomes clear while looking at the key changes in the value system of AnGRs is that AnGRs have become increasingly complex objects for breeders,

scientists and politicians alike, with no easy answers to how the balancing of rights, responsibilities and benefit sharing in the near future. While AnGRs have finally become a global issue with high political priority and action, so have the political conditions under which the animals live become inherently global entanglement of science and politics, culture and nature.

At the same time, the status of AnGRs that reside outside of the CBD system—either owned by private companies or breeding societies before the entry into force of the CBD in 1993—is unclear. Although they are not objects of CBD's articles, they might still be affected by and become targets of legal interventions by way in which for example China and South Africa have applied the sovereignty over genetic resources within their respective AnGR regulations. This makes the global system even more complicated, and most likely with a number of unforeseen challenging cases in the future.

The CBD, the Global Plan of Action and the Nagoya Protocol present a global value system framing AnGRs in a way that is finally generating conservation action at national level. But on the global level, the system is more muddled than ever calling for a great deal of conceptual, political and legal analysis to bring more clarity to the current condition that requires the generation of discrete genetic landscapes and marks AnGRs with their nationally correct *in situ* location as their political condition of existence. Given the complex history of AnGRs as a global matter of concern, creating clarity to the present situation will not be easy.

At least three key questions need to be clarified with regards to AnGR and the different claims laid over them in order to move on in the global politics, in the creation and implementation of legal frameworks at national level, and in the reflection of the true impact of the CBD and Nagoya Protocol.

(1) What is the true scope of the CBD and Nagoya Protocol in terms of AnGR types? Are there types of AnGRs that remain totally unaffected by and reside outside of the scope of the global conventions?

(2) Do the signatory parties (nation-states) have prototypical reactions to—or at least broadly identifiable patterns in—the implementation of their sovereign powers over AnGR?

(3) If the signatory parties do exhibit identifiable patterns, a guiding typology of CBD and Nagoya Protocol implementation at national level would help to make sense of how governments are adopting the global agreements (e.g., types of entitlement claims, access regulations etc) at large.

Addressing these three points would already give a much richer and much more coherent overview of AnGRs' status in the post-CBD and post-Nagoya Protocol world than is currently available for public. We do suggest that the institutions driving the global framework on genetic resources provide it soon.

References

Boyazoglu, J., and Chupin, D. (1991). *"Editorial," Animal Genetic Resources Information Bulletin*. Rome: Food and Agriculture Organization of the United Nations.

Buck, M., and Hamilton, C. (2011). The nagoya protocol on access to genetic resources and the fair and equitable sharing of benefits arising from their utilization to the convention on biological diversity. *Reciel* 20, 47–61. doi: 10.1111/j.1467-9388.2011.00703.x

Commission on Genetic Resources for Food and Agriculture. (2009). *The Use and Exchange of Animal Genetic Resources for Food and Agriculture*. Rome: FAO.

Commission on Genetic Resources for Food and Agriculture. (2014). *Report of the Eighth Session of the Intergovernmental Technical Working Group on Animal Genetic resources for Food and Agriculture*. Report CGRFA/WG-angr-8/14/REPORT. Rome: FAO.

David, P. A. (2011). "Breaking anti-commons constraints on global scientific research: some new moves in "Legal Jujitsu"," in *Designing the Microbial Research Commons: Proceedings of an International Symposium*, ed. P. F. Uhlir (Washington, DC: National Academies Press), 13–24.

de Haen, H. (1992). "Opening statement," in *The Management of Global Animal Genetic Resources. Proceedings of an FAO Expert Consultation*, ed. J. Hodges (Rome: FAO), 23–25.

Derry, M. E. (2003). *Bred for Perfection: Shorthorn Cattle, Collies, and Arabian Horses Since 1800*. Baltimore: Johns Hopkins University Press.

Food and Agriculture Organization. (1990). "Animal genetic resources: a global programme for sustainable development," in *Animal Production and Health Paper No. 80. Proceedings of an FAO Expert Consultation*, Rome.

Food and Agriculture Organization. (1995). *Major Trends and Policies in Food and Agriculture*. Rome: FAO.

Food and Agriculture Organization. (1999). *The Global Strategy for the Management of Farm Animal Genetic Resources*. Rome: FAO.

Food and Agriculture Organization. (2007). *Global Plan of Action for Animal Genetic Resources and the Interlaken Declaration. Commission on Genetic Resources for Food and Agriculture*. Rome: FAO, Commission on Genetic Resources for Food and Agriculture.

Hardin, G. (1968). The tragedy of the commons. *Science* 162, 1243–1248. doi: 10.1126/science.162.3859.1243

Hiemstra, S. J. A. G., Drucker, M. W., Tvedt, N., Louwaars, J. K., Oldenbroek, K., Awgichew, S., et al. (2006). *Exchange, Use and Conservation of Animal Genetic Resources Policy and Regulatory Options*. CGN Report 2006/06. Wageningen University and Research Centre: Centre for Genetic Resources, the Netherlands (CGN).

Hodges, J. (1984). "Review of the fao/unep programme on animal genetic resources conservation and management," in *FAO Animal Production and Health Paper 44/1: Animal Genetic Resources Conservation by Management, Data Banks and Training Rome*. 2–9.

Hodges, J. (1990). Animal genetic resources. *Impact Sci. Soc.* 40, 143–154.

Kloppenburg, J. R. (1988). *First the Seed: The Political Economy of Plant Biotechnology, 1492–2000*. Cambridge: Cambridge University Press.

Maijala, K. (1971). Need and methods of gene conservation in animal breeding. *Ann. Fenet. Sel. Anim.* 2, 403–415.

Mason, I. L. (1981). "Cooperative work by FAO and UNEP on the conservation of animal genetic resources," in *FAO Animal Production and Health Paper 24: Animal Genetic Resources Conservation and Management. Proceedings of the FAO/UNEP Technical Consultation* (Rome: FAO), 16–29.

National Academy of Science. (1993). "Preface," in *Managing Global Livestock Genetic Resources Committee on Managing Global Genetic Resources: Agricultural Imperatives Board on Agriculture National Research Council*. ed. Barbara J. Rice (Washington, DC: National Academy Press), 1–20.

Phillips, R. W. (1981). "The identification, conservation and effective use of valuable animal genetic resources," in *FAO Animal Production and Health Paper 24: Animal Genetic Resources Conservation and Management. Proceedings of the fao/unep Technical Consultation in 1980*. Rome: FAO, 2–6.

Ritvo, H. (1995). "Possessing Mother Nature. Genetic Capital in Eighteenth-Century Britain," in *Early Modern Conceptions of Property*, eds J. Brewer and S. Staves (London and New York: Routledge), 413–426.

Steane, D. E. (1992). Note on the fao expert consultation on management of global animal genetic resources rome, 7-10 APRIL 1992. *Anim. Genet. Resour. Informat.* 9, 3–6.

Tvedt, M., Hiemstra, S. J., Drucker, A. G., Louwaars, N., and Oldenbroek, K. (2007). *Legal Aspects of Exchange, Use and Conservation of Farm Animal Genetic Resources*. FNI Report 1/2007. Lysaker, FNI.

Walton, J. (1999). Pedigree and productivity in the British and North American cattle kingdoms. *J. Hist. Geogr.* 25, 441–462. doi: 10.1006/jhge.1999.0161

Wheeler, M. B. (2013). Transgenic animals in agriculture. *Nat. Educ. Knowl.* 4, 1.

Wheeler, M. B., Monaco, E., Bionaz, M., and Tanaka, T. (2010). The role of existing and emerging biotechnologies for livestock production: toward holism. *Acta Sci. Vet.* 38, 463–484.

Wilmut, I., Haley, C. S., and Woolliams, J. A. (1992). Impact of biotechnology on animal breeding. *Anim. Reprod. Sci.* 28, 149–162. doi: 10.1016/0378-4320(92)90101-I

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Tamminen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in GENETICS

GENERAL COMMENTARY
published: 19 January 2015
doi: 10.3389/fgene.2014.00461

Genetic resources and genomics for adaptation of livestock to climate change

Paul J. Boettcher[1]*, Irene Hoffmann[1], Roswitha Baumung[1], Adam G. Drucker[2], Concepta McManus[3], Peer Berg[4], Alessandra Stella[5,6], Linn B. Nilsen[7,8], Dominic Moran[9], Michel Naves[10] and Mary C. Thompson[2,11]

[1] *Animal Production and Health Division, Food and Agriculture Organization of the United Nations, Rome, Italy*
[2] *In Situ and On Farm Conservation, Bioversity International, Maccarese, Italy*
[3] *INCT Pecuaria, Universidade de Brasília, Brasilia, Brazil*
[4] *NordGen Farm Animals, Nordic Genetic Resource Center, Ås, Norway*
[5] *Parco Tecnologico Padano, Lodi, Italy*
[6] *Institute of Agricultural Biology and Biotechnology, National Research Council, Milan, Italy*
[7] *Plant Production and Protection Division, Food and Agriculture Organization of the United Nations, Rome, Italy*
[8] *NADEL – Center for Development and Cooperation, Swiss Federal Institute of Technology in Zürich, Zürich, Switzerland*
[9] *Sustainable Rural Systems, Scotland's Rural College, Edinburgh, UK*
[10] *UR143, Unité de Recherches Zootechniques, Institut National de la Recherche Agronomique, Petit-Bourg, Guadeloupe, France*
[11] *Basque Center for Climate Change, Bilbao, Spain*
**Correspondence: paul.boettcher@fao.org*

Edited by:
Juha Kantanen, MTT Agrifood Research Finland, Finland

Reviewed by:
Dirk-Jan De Koning, Swedish University of Agricultural Sciences, Sweden
Jarmo Juga, University of Helsinki, Finland

Keywords: animal genetic resources, adaptation, biological, genetic diversity, livestock genomics, climate change

A commentary on

Global plan of action for animal genetic resources and the interlaken declaration *by FAO. (2007). Rome. ISBN: 978-92-5-105848-0*

INTRODUCTION

Animal genetic resources (AnGR) are critical for global food security and livelihoods. Livestock products have high densities of energy, protein, and other critical nutrients, which are particularly beneficial for infants and expectant mothers. Around a billion people rely directly on livestock for their livelihoods, many of which are among the rural poor (FAO, 2009). Demand for animal products is foreseen to increase significantly in the future while competition for resources will intensify, dictating that livestock systems must increase both productivity and efficiency. Maintaining sufficient diversity of AnGR is necessary to ensure adaptation potential in times of uncertainty. In the future, climate change is expected to be a major force testing resilience of global food production systems (Thornton et al., 2009; Renaudeau et al., 2012). Ensuring that livestock systems remain productive and efficient while maintaining their flexibility will be a major challenge.

Adaptation to climate change is unlikely to be achieved with a single strategy (Hoffmann, 2010). Clearly, modifications will be needed in animals' housing, reproduction, nutrition, and health care. Genetic changes in the animals (both within and across species) will also play a role. Preparation for these transformations will require a significant research commitment and genomics will play a role in the genetic measures taken for adaptation of livestock to climate change.

INFORMATION NEEDS

The first step in this process will be gathering relevant information. There is currently a shortage of specific knowledge explaining why certain AnGR are adapted to a given environment and in which other environments they can survive and flourish. This deficit calls for greater effort in characterization of AnGR, including their production environments, by using the most modern tools available. To date, many livestock breeds have been genetically characterized (e.g., Groeneveld et al., 2010), but the value of these data for study of adaptation is questionable. For climate change adaptation, the breeds of greatest interest may be those reared today in harsh environments. Past studies have mostly addressed breeds from developed countries, where climate-control is widely practiced. The study of adaptation implies the use of a "landscape approach," with detailed information describing the production system (e.g., FAO, 2012), including socio-economic information (e.g., Drucker, 2010) and indigenous knowledge about management of the breed in its environment as well as geographic coordinates to incorporate climatic data and soil, vegetation, and water resources. Collection of such detailed complementary data is a relatively recent trend. Past studies have emphasized pure breeds, whereas crossbreeding can be a valuable strategy for achieving increased productivity and adaptability, so those populations also require characterization. Finally, many studies have assayed only small numbers of selectively-neutral markers. The rapid development of genomic tools now allows analysis of functional genomic regions with potential associations with adaptation (e.g., Qian et al., 2013).

Table 1 | Numbers of local[a] breeds recorded per species and region for the six main livestock species[b] for food and agriculture, according to the Domestic Animal Diversity Information System (DAD-IS) of FAO (http://dad.fao.org – November 2014).

Species	Africa	Asia	Europe and the Caucasus	Latin America and the Caribbean	Near and Middle East	North America	Southwest Pacific	World
Cattle	176	241	369	141	43	17	32	1019
Chicken	129	305	912	88	35	15	30	1514
Sheep	117	262	613	51	53	21	38	1155
Horse	40	138	371	84	14	22	25	694
Goat	96	183	218	28	34	6	11	576
Pig	53	214	188	60	1	12	15	543
Duck	15	92	107	22	4	1	12	253
Others	171	328	568	123	67	37	27	1321
Total	797	1763	3346	597	251	131	190	7075

[a] *Local breed, a breed reported in only one country, according to DAD-IS.*

[b] *Main species determined based on the number of breeds.*

SELECTION OBJECTIVES AND STRATEGIES

A second research need is the elaboration of the genetic objective for adaptation. Increasing productivity and efficiency will be fundamental, but maintenance of genetic diversity will also be of importance. Having diverse AnGR will allow for more opportunities to match breeds to a changing climate or to replace populations hit by severe climatic events such as droughts and floods. Within breed, broad genetic diversity will clearly allow for greater opportunities for selection for adaptation, but there is evidence from wild populations that increased genetic diversity is also selectively advantageous on the individual level (Fourcada and Hoffman, 2014). Directional selection for adaptive traits will likely accompany maintenance of diversity, but questions remain about indicators of adaptation and resilience and hence the breeding goal. An obvious option is to breed for traits associated with superior productivity and resilience in conditions expected to be prevalent as a result of climate change, such as heat and drought tolerance and resistance to certain diseases. However, care must be taken when defining such traits. In dairy cattle, the ability to maintain high milk production with increasing ambient temperatures seems like a logical definition of heat tolerance, but research at the physiological level has shown that such cattle direct their energy toward milk production, making them vulnerable to extremely high temperatures (Dikmen et al., 2012). An alternative to breeding for specific traits is to target general robustness; the ability of animals to adjust to a range of environmental conditions. The Brown Swiss dairy cow, developed under comparatively cool, but rugged conditions in the Alps, seems to show greater heat tolerance than the Holstein (Correa-Calderon et al., 2005), which originated in more temperate lowland regions. The Domestic Animal Diversity Information system of FAO (http://dad.fao.org) lists numerous breeds, particularly from mountainous and arid areas, that are adapted to extreme ranges in temperature and such breeds may merit further research (Hoffmann, 2013). **Table 1** has the numbers of local breeds by region and species for the major livestock species, as recorded in DAD-IS. Within-breed selection for robustness would likely require the development of an index involving multiple traits.

A third category of research involves the genetic strategy for adaptation. Options comprise purebreeding, crossbreeding (including introgression), and breed or species substitution. Among the key influential factors is the expected rate of climate change and the speed with which genetic change can realistically occur with the various strategies. Substitution and crossbreeding expedite genetic change, but their implementation may be more complex than purebreeding and thus involve additional research needs (e.g., on genotype-by-environment interaction).

TOOLS FOR THE GENETICS OF ADAPTATION

A final research topic will be the development of tools required for the aforementioned topics. Genomics will surely play a role in all three of these areas, as well as in implementation of the results obtained. Increased characterization with high-throughput single nucleotide polymorphism (SNP) assays or genome sequencing will be necessary for unraveling the physiological basis for adaptation. Species-wide HapMap studies (e.g., Gibbs et al., 2009; Jiang et al., 2014) and multi-species studies (e.g., Stella, 2014) have represented a valuable first step in understanding the genome and its function in adaptation, but must be expanded to more breeds and geographical areas and augmented with more information on production environments. Metagenomics can provide insight regarding the co-adaptation of AnGR with other organisms in their production environments. Genomic selection has the potential to expedite both pure- and crossbreeding programmes for adaptation, assuming phenotypes are available (Hayes et al., 2012); programs for performance recording in developing countries are thus needed. Given the importance of the landscape approach, tools and methods for improved integration of geographical information will be critical. The information gathered in all of these processes will be of little value if they are not properly organized, stored, and disseminated to stakeholders through new

and improved databases and information systems. Finally, cooperative efforts between all stakeholders will be needed to achieve the final goal, the optimal utilization of the genetic and genomic resources for the adaptation of livestock to climate change.

ACKNOWLEDGMENTS

This report comprises conclusions from the Expert Workshop on Crop and Livestock Diversity for Climate Change Adaptation held in Rome from 8 to 12 October 2013. The workshop was a collaborative effort among FAO, Bioversity International, the Basque Center for Climate Change and DIVERSITAS and was financially supported by the Government of Sweden through FAO project GCP/GLO/287/MUL and by the CGIAR program on Climate Change, Agriculture, and Food Security (CCAFS).

REFERENCES

Correa-Calderon, A., Armstrong, D. V., Ray, D. E., DeNise, S. K., Enns, R. M., and Howison, C. M. (2005). Productive and reproductive response of Holstein and Brown Swiss heat stressed dairy cows to two different cooling systems. *J. Anim. Vet. Adv.* 4, 572–578.

Dikmen, S., Cole, J. B., Null, D. J., and Hansen, P. J. (2012). Heritability of rectal temperature and genetic correlations with production and reproduction traits in dairy cattle. *J. Dairy Sci.* 95, 3401–3405. doi: 10.3168/jds.2011-4306

Drucker, A. G. (2010). Where's the beef? The economics of AnGR conservation and its influence on policy design and implementation. *Anim. Genetic Resour.* 47, 85–90. doi: 10.1017/S2078633610000913

FAO. (2009). *The State of Food and Agriculture – Livestock in the Balance*. Rome. Available online at: http://www.fao.org/docrep/012/i0680e/i0680e.pdf

FAO. (2012). *Phenotypic Characterization of Animal Genetic Resources*. FAO Animal Production and Health Guidelines No. 11. Rome.

Fourcada, J., and Hoffman, J. I. (2014). Climate change selects for heterozygosity in a declining fur seal population. *Nature* 511, 462–465. doi: 10.1038/nature13542

Gibbs, R. A., Taylor, J. F., Van Tassell, C. P., Barendse, W., Eversole, K. A., Gill, C. A., et al. (2009). Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Science* 324, 528–532. doi: 10.1126/science.1167936

Groeneveld, L. F., Lenstra, J. A., Eding, H., Toro, M. A., Scherf, B., Pilling, D., et al. (2010). Genetic diversity in farm animals–a review. *Anim. Genet.* 41, 6–31. doi: 10.1111/j.1365-2052.2010.02038.x

Hayes, B. J., Lewin, H. A., and Goddard, M. E. (2012). The future of livestock breeding: genomic selection for efficiency, reduced emissions intensity, and adaptation. *Trends Genet.* 29, 206–214. doi: 10.1016/j.tig.2012.11.009

Hoffmann, I. (2010). Climate change and the characterization, breeding and conservation of animal genetic resources. *Anim. Genet.* 41, 32–46. doi: 10.1111/j.1365-2052.2010.02043.x

Hoffmann, I. (2013). Adaptation to climate change – exploring the potential of locally adapted breeds. *Animal* 7, 346–362. doi: 10.1017/S1751731113000815

Jiang, Y., Xie, M., Chen, W., Talbot, R., Maddox, J. F., Faraut, T., et al. (2014). The sheep genome illuminates biology of the rumen and lipid metabolism. *Science* 344, 1168–1173. doi: 10.1126/science.1252806

Qian, W., Deng, L., Lu, D., and Xu, S. (2013). Genome-wide landscapes of human local adaptation in Asia. *PLoS ONE* 8:e54224. doi: 10.1371/journal.pone.0054224

Renaudeau, D., Collin, A., Yahav, S., de Basilio, V., Gourdine, J. L., and Collier, R. J. (2012). Adaptation to hot climate and strategies to alleviate heat stress in livestock production. *Animal* 6, 707–728. doi: 10.1017/S1751731111002448

Stella, A. (2014). "Insights into the interaction of goat breeds and their environment," in *Proceedings of the 10th World Congress on Genetics Applied to Livestock Production* (Vancouver, BC).

Thornton, P. K., van de Steeg, J., Notenbaert, A., and Herrero, M. (2009). The impacts of climate change on livestock and livestock systems in developing countries: a review of what we know and what we need to know. *Agricult. Syst.* 101, 113–127. doi: 10.1016/j.agsy.2009.05.002

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 12 November 2014; paper pending published: 26 November 2014; accepted: 15 December 2014; published online: 19 January 2015.

Citation: Boettcher PJ, Hoffmann I, Baumung R, Drucker AG, McManus C, Berg P, Stella A, Nilsen LB, Moran D, Naves M and Thompson MC (2015) Genetic resources and genomics for adaptation of livestock to climate change. Front. Genet. ***5****:461. doi: 10.3389/fgene.2014.00461*

This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics.

Copyright © 2015 Boettcher, Hoffmann, Baumung, Drucker, McManus, Berg, Stella, Nilsen, Moran, Naves and Thompson. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

FAnGR IN AFRICA

frontiers in Genetics

ORIGINAL RESEARCH
published: 07 April 2015
doi: 10.3389/fgene.2015.00107

Characterizing neutral genomic diversity and selection signatures in indigenous populations of Moroccan goats (*Capra hircus*) using WGS data

Badr Benjelloun[1,2,3], Florian J. Alberto[1,2], Ian Streeter[4], Frédéric Boyer[1,2], Eric Coissac[1,2], Sylvie Stucki[5], Mohammed BenBati[3], Mustapha Ibnelbachyr[6], Mouad Chentouf[7], Abdelmajid Bechchari[8], Kevin Leempoel[5], Adriana Alberti[9], Stefan Engelen[9], Abdelkader Chikhi[6], Laura Clarke[4], Paul Flicek[4], Stéphane Joost[5], Pierre Taberlet[1,2], François Pompanon[1,2] and NextGen Consortium[10]*

[1] Laboratoire d'Ecologie Alpine, Université Grenoble-Alpes, Grenoble, France, [2] Laboratoire d'Ecologie Alpine, Centre National de la Recherche Scientifique, Grenoble, France, [3] National Institute of Agronomic Research (INRA Maroc), Regional Centre of Agronomic Research, Beni-Mellal, Morocco, [4] European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton, UK, [5] Laboratory of Geographic Information Systems (LASIG), School of Civil and Environmental Engineering (ENAC), École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, [6] Regional Centre of Agronomic Research Errachidia, National Institute of Agronomic Research (INRA Maroc), Errachidia, Morocco, [7] Regional Centre of Agronomic Research Tangier, National Institute of Agronomic Research (INRA Maroc), Tangier, Morocco, [8] Regional Centre of Agronomic Research Oujda, National Institute of Agronomic Research (INRA Maroc), Oujda, Morocco, [9] Centre National de Séquençage, CEA-Institut de Génomique, Genoscope, Évry, France, [10] NextGen Consortium, http://nextgen.epfl.ch/

Since the time of their domestication, goats (*Capra hircus*) have evolved in a large variety of locally adapted populations in response to different human and environmental pressures. In the present era, many indigenous populations are threatened with extinction due to their substitution by cosmopolitan breeds, while they might represent highly valuable genomic resources. It is thus crucial to characterize the neutral and adaptive genetic diversity of indigenous populations. A fine characterization of whole genome variation in farm animals is now possible by using new sequencing technologies. We sequenced the complete genome at 12× coverage of 44 goats geographically representative of the three phenotypically distinct indigenous populations in Morocco. The study of mitochondrial genomes showed a high diversity exclusively restricted to the haplogroup A. The 44 nuclear genomes showed a very high diversity (24 million variants) associated with low linkage disequilibrium. The overall genetic diversity was weakly structured according to geography and phenotypes. When looking for signals of positive selection in each population we identified many candidate genes, several of which gave insights into the metabolic pathways or biological processes involved in the adaptation to local conditions (e.g., panting in warm/desert conditions). This study highlights the interest of WGS data to characterize livestock genomic diversity. It illustrates the valuable genetic richness present in indigenous populations that have to be sustainably managed and may represent valuable genetic resources for the long-term preservation of the species.

Keywords: ***Capra hircus*, WGS, genomic diversity, population genomics, selection signatures, indigenous populations, Morocco**

OPEN ACCESS

Edited by:
Johannes Arjen Lenstra, Utrecht University, Netherlands

Reviewed by:
Luca Fontanesi, University of Bologna, Italy
Mathieu Gautier, Institut National de la Recherche Agronomique, France

****Correspondence:***
Badr Benjelloun, Laboratoire d'Ecologie Alpine (LECA), Centre National de la Recherche Scientifique UMR 5553, Univ. Joseph Fourier, BP 53, 38041 Grenoble Cedex 9, France; Regional Centre of Agronomic Research, National Institute of Agronomic Research (INRA Maroc), B.P. 567, Beni-Mellal 23000, Morocco badr.benjelloun@gmail.com; badr.benjelloun@ujf-grenoble.fr

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: *17 December 2014*
Accepted: *02 March 2015*
Published: *07 April 2015*

Citation:
Benjelloun B, Alberto FJ, Streeter I, Boyer F, Coissac E, Stucki S, BenBati M, Ibnelbachyr M, Chentouf M, Bechchari A, Leempoel K, Alberti A, Engelen S, Chikhi A, Clarke L, Flicek P, Joost S, Taberlet P, Pompanon F and NextGen Consortium (2015) Characterizing neutral genomic diversity and selection signatures in indigenous populations of Moroccan goats (Capra hircus) using WGS data. Front. Genet. 6:107. doi: 10.3389/fgene.2015.00107

Introduction

Livestock species play a major socio-economic role in the world since they provide many goods and services to human populations. Goats (*Capra hircus*) in particular are one of the more important livestock species, because of their high potential of adaptation to harsh environments. They had a worldwide population of about 1006 million in 2013 (http://faostat3.fao.org/browse/Q/QA/E) and, together with cattle and sheep, they represent the most important source of meat, milk, and skin.

Goats are considered to be the first ungulate to be domesticated, about 10,500 to 9900 years ago near the Fertile Crescent (Zeder, 2005; Naderi et al., 2008). Following human migrations and trade routes, goats rapidly spread over the rest of the world, mainly in Eurasia and Africa (Taberlet et al., 2008; Tresset and Vigne, 2011). During this expansion, they became adapted to different climatic conditions and husbandry practices. In response to these environmental and anthropic selection pressures, a large variety of locally-adapted populations emerged. These populations were managed in a traditional way, *i.e.,* with moderate selection for traits of interest and reproduction allowing important gene flows among them, thus maintaining high levels of phenotypic diversity (Taberlet et al., 2008). However, the rise of the breed concept during mid-1800s (Porter, 2002), and its application to husbandry practices, led to the creation of well-defined breeds. This process aimed at standardizing phenotypic traits mainly associated with morphological aspects (e.g., coat color). Selection of animals for these traits was generally moderated, while crossing among different phenotypes was reduced (Taberlet et al., 2008). More recently, since mid-1900s, industrial breeding has become more widespread, backed by the progress of husbandry practices including the introduction of artificial insemination, embryo transfer, the improvements in feed technology and the use of vaccines and therapeutics against endemic diseases. This has led breeders to progressively substitute the many locally-adapted indigenous breeds for very few highly productive cosmopolitan ones for short-term economic reasons (Taberlet et al., 2008). Thus, FAO in 2013 estimated that 18% of local goat breeds over the world were threatened or already extinct (http://faostat.fao.org/). Consequently, a part of the highly valuable genetic resources captured from the wilds and gradually accumulated over 98% of their common history with humans is now threatened (Taberlet et al., 2008).

Thus, it appears crucial to assess the genetic resources of indigenous populations in order to manage them sustainably and to propose zootechnical approaches that take into account the preservation of genetic resources. This might be critical in the current context of global environmental changes. To accurately characterize genetic resources, it is necessary to access variation data across the whole genome. This would allow the identification of alleles related to contrasted environmental conditions and those potentially playing an adaptive role. Recent progress in sequencing technologies has opened new perspectives toward the magnitude of genetic analysis that is possible. Sequencing cost and time have dramatically decreased (Snyder et al., 2010) and it is now possible to obtain Whole Genome Sequencing (WGS) data for several dozen individuals, which allows access to variation data sets of the whole genome (Altshuler et al., 2012; Kidd et al., 2012). It is thus possible to combine WGS data and population genomic approaches to characterize neutral and adaptive variation in an unprecedented way. This allows an accurate characterisation of genetic resources and their geographic distribution. The Moroccan territory represents an ideal case-study for evaluating the potential of indigenous breeds for constituting neutral and adaptive genomic resources. Despite the massive introduction of "cosmopolitan" breeds to improve goat milk production in some areas, indigenous populations still represent about 95% of Moroccan goats. This proportion has been continually decreasing and this could lead in a mid-long term to the complete absorption of some indigenous populations by cosmopolitan breeds. In Morocco there are more than 6.2 Million goats (http://faostat3.fao.org/browse/Q/QA/E). Direct anthropic selection was relatively modest and until recently it was difficult to distinguish well-defined breeds. However, several phenotypic groups displaying specific morphological and adaptive characteristics have been identified. They will be referred hereafter here as populations. The three major groups are: (i) the Black goats with three sub-populations that have been recently officially recognized (Atlas, Barcha and Ghazalia), (ii) the Draa population, (iii) and the Northern population. Besides these three main populations/breeds, the major proportion of Moroccan goats presents intermediate phenotypes and non-recognized local populations. The Black population is characterized by its dark color, long hair, a low water turnover and thus good resistance to water stress (Hossainihilali et al., 1993). It presents a good acclimation to various environmental conditions in Morocco (from the Eastern plateaus to Atlas Mountains and the Souss valley more in the South). The Northern population displays some phenotypic similarities with Spanish breeds such as the Murciana-Granadina, Malaguena or Andalusia breeds (Benlekhal and Tazi, 1996). It is bred for milk and meat production although it presents a lower level of production than cosmopolitan industrial dairy breeds (Analla and Serradilla, 1997). It shows a substantial reproductive seasonality related to photoperiod variation (Chentouf et al., 2011). Following an extensive breeding system, it is the preferred breed to be raised in the harsh mountains of the extreme North of Morocco with oceanic influence and a milder climate. The Draa population is bred in the oasis in Southern Morocco, which is characterized by arid/desert climate conditions. Its water turnover is low compared to European goat breeds studied in similar environments. The Draa goat also has the ability to maintain an unchanged food intake during periods of water deprivation (Hossaini-Hilali and Mouslih, 2002). It displays relatively higher performances of reproduction (i.e., prolificacy, earliness; Ibnelbachyr et al., 2014) and hornless individuals represent about 54.1% of the total (Ibnelbachyr et al., in preparation). In this study, we applied a population genomic framework using WGS data to (i) describe neutral genomic diversity and population structure in the main Moroccan indigenous goat populations (ii) identify potential genomic regions differentially selected among the main populations according to their specific traits. To address these issues, we sequenced at 12× coverage 44 goats representing the Moroccan-wide geographic

diversity of the three main goat indigenous populations in the country.

Material and Methods

Sampling

Sample collection was performed in a wide part of Morocco [~400,000 km^2; Northern part of Morocco in latitude range (28°−36°)]. A total of 44 individuals unambiguously assigned to one of the three main indigenous populations (i.e., Black, Draa and Northern) were sampled (Table S1) in a way that maximized individuals' spread over the sampling area. This resulted in sampling spatially distant unrelated individuals, ensuring a spatial representativeness of all regions (**Figure 1**). For each individual, tissue samples were collected from the distal part of the ear and placed in alcohol for 1 day, and then transferred to a silica-gel tube until DNA extraction.

Production Of WGS Data

DNA extractions were done using the Puregene Tissue Kit from Qiagen® following the manufacturer's instructions. Then, 500 ng of DNA were sheared to a 150–700 bp range using the Covaris® E210 instrument (Covaris, Inc., USA). Sheared DNA was used for Illumina® library preparation by a semi-automatized protocol. Briefly, end repair, A tailing and Illumina® compatible adaptors (BiooScientific) ligation were performed using the SPRIWorks Library Preparation System and SPRI TE instrument (Beckmann Coulter), according to the manufacturer protocol. A 300–600 bp size selection was applied in order to recover the most of fragments. DNA fragments were amplified by 12 cycles PCR using Platinum Pfx Taq Polymerase Kit (Life® Technologies) and Illumina® adapter-specific primers. Libraries were purified with 0.8× AMPure XP beads (Beckmann Coulter). After library profile analysis by Agilent 2100 Bioanalyzer (Agilent® Technologies, USA) and qPCR quantification, the libraries were sequenced using 100 base-length read chemistry in paired-end flow cell on the Illumina HiSeq2000 (Illumina®, USA).

WGS Data Processing

Paired-end reads were mapped to the goat reference genome (CHIR v1.0, GenBank assembly GCA_000317765.1) (Dong et al., 2013) using BWA mem (Li and Durbin, 2009). The BAM files produced were then sorted using Picard SortSam and improved using Picard MarkDuplicates (http://picard.sourceforge.net), GATK RealignerTargetCreator, GATK IndelRealigner (Depristo et al., 2011), and Samtools calmd (Li et al., 2009). Variant calling was done using three different algorithms: Samtools mpileup (Li et al., 2009), GATK UnifiedGenotyper (McKenna et al., 2010), and Freebayes (Garrison and Marth, 2012).

There were two successive rounds of filtering variant sites. Filtering stage 1 merged together calls from the three algorithms, whilst filtering out the lowest-confidence calls. A variant site passed if it was called by at least two different calling algorithms with variant phred-scaled quality >30. An alternate allele at a site passed if it was called by any one of the calling algorithms, and the genotype count >0. Filtering stage 2 used Variant Quality Score Recalibration by GATK. First, we generated a training set of the highest-confidence variant sites where (i) the site is called by all three variant callers with variant phred-scaled quality >100, (ii) the site is biallelic (iii) the minor allele count is at least 3 while counting only samples with genotype phred-scaled quality >30. The training set was used to build a Gaussian model using the tool GATK VariantRecalibrator using the following variant annotations from UnifiedGenotyper: QD, HaplotypeScore, MQRankSum, ReadPosRankSum, FS, DP, InbreedingCoefficient. The Gaussian model was applied to the full data set, generating a VQSLOD (log odds ratio of being a true variant). Sites were filtered out if VQSLOD < cutoff value. The cutoff value was set for each population by the following: Minimum VQSLOD = {the median value of VQSLOD for training set variants} − 3* {the median absolute deviation VQSLOD

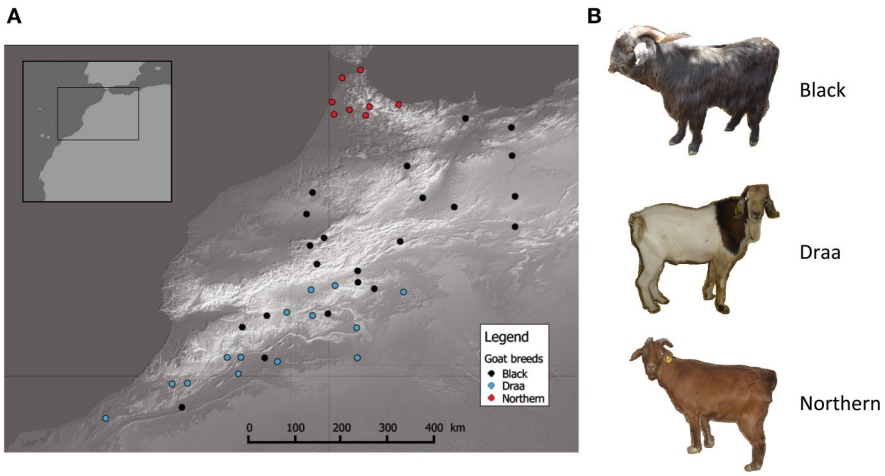


FIGURE 1 | Distribution of goats sampled. (A) Geographic map showing the distribution of the 44 goats sampled in this study. Each point represents one individual and different colors illustrate different populations. **(B)** Striking phenotypic differences between the 3 main goat populations in Morocco.

of training set variants}. Measures of the transition / transversion ratio of SNPs suggest that this chosen cutoff criterion gives the best balance between selectivity and sensitivity. Genotypes were improved and phased by Beagle 4 (Browning and Browning, 2013), and then filtered out where the genotype probability calculated by Beagle is less than 0.95.

The whole mitochondrial genome (mtDNA) was assembled from a subset of random 20,000,000 reads using the ORGASM tool (Coissac, unpublished). We then extracted the sequence of the HVI segment of the control region for each individual in order to compare with the haplogroup references discovered worldwide (see below).

Population Genomic Analyses

Characterisation of mtDNA Diversity

The number of polymorphic sites and the number of haplotypes were calculated from the whole mitochondrial sequences using DNAsp (Librado and Rozas, 2009). We also calculated these parameters for the hyper variable segment (HVI) of the control region, for which 22 reference sequences representing the diversity of the 6 haplogroups found over the world were available (Naderi et al., 2007). We were interested in the level of resolution of the HVI segment to discriminate the different haplotypes compared to the whole mitochondrion.

Then, using the sequences corresponding to the HVI segment for our dataset and the reference sequences, we drew a network of the haplotypes to identify the different haplogroups present in our dataset. The best evolutionary model was determined using jModelTest v 2.1.4 (Darriba et al., 2012). A median joining network representing the relationships between haplotypes was drawn using SplitsTree4 (Huson and Bryant, 2006).

Characterisation of Neutral Nuclear Diversity

Neutral nuclear genomic variations were characterized to evaluate the level of genetic diversity present in Morocco and within populations. The total number of variants and the number of variants within each population were calculated. Allele frequencies and the percentage of exclusive variants (i.e., variants polymorphic in only one population) were estimated at the population scale using the Perl module vcf-compare of Vcftools (Danecek et al., 2011). The level of nucleotide diversity (π) was calculated within each population and averaged over all of the biallelic and fully diploid variants for which all individuals had a called genotype. The observed percentage of heterozygote genotypes per individual (*Ho*) was calculated considering only the biallelic SNPs with no missing genotype calls. From *Ho,* the inbreeding coefficients (*F*) were calculated for each individual using population allelic frequencies over all 44 individuals. The relatedness among individuals was assessed using the pairwise identity-by-state (*IBS*) distances calculated as the average proportion of alleles shared using Vcftools.

Pairwise linkage disequilibrium (LD) was assessed through the correlation coefficient (r^2). It was estimated in 5 segments of 2 Mb on different chromosomes (physical positions between 5 and 7 Mb on chromosomes 6, 11, 16, 21, and 26). LD was estimated either by using the whole set of reliable variants or after discarding rare variants with a minor allele frequency (MAF) less than 0.05. For both estimations, we calculated r^2 values between all pairs of bi-allelic variants (SNPs and indels) on the same segment using Vcftools. Inter-SNP distances (kb) were binned into the following 7 classes: 0–0.2, 0.2–1, 1–2, 2–10, 10–30, 30–60, and 60–120 kb and observed pairwise *LD* was averaged for each inter-SNP distance class and used to draw *LD* decay. Due to the insufficient number of individuals per population we made these estimations for the whole set of individuals without considering each population individually.

Genetic structure was assessed using three different methods: (i) a principal component analysis (PCA) was done using an *LD* pruned subset of bi-allelic SNPs. *LD* between SNPs in windows containing 50 markers was calculated before removing one SNP from each pair where *LD* exceeded 0.95. Subsequently, only 12,543,534 SNPs among a total of 22,304,702 bi-allelic SNPs were kept for this analysis. The R package adegenet v1.3-1 (Jombart and Ahmed, 2011) was used to run PCA and Plink v1.90a (https://www.cog-genomics.org/plink2) was used for *LD* pruning. (ii) An analysis with the clustering method sNMF (Frichot et al., 2014) was carried-out. This method was specifically developed to analyse large genomic datasets in a fast, efficient and reliable way. It is based on sparse non-negative matrix factorization to estimate admixture coefficients of individuals. All biallelic variants were used and five runs for each *K* value from 1 to 10 were performed using a value of *alpha* parameter of 8. For each run, the cross-entropy criterion was calculated with 5% missing data to identify the most likely number of clusters. The run showing the lowest cross-entropy value for a given K was considered. (iii) Finally, the *Fst* index was estimated according to Weir and Cockerham (1984) for each polymorphic site and then weighted to obtain one value over the whole genome. The overall *Fst* between the three groups and the population pairwise values were calculated using Vcftools.

Detection of Selection Signatures

A genome scan approach was performed using the XP-CLR method (Chen et al., 2010) to identify potential regions differentially selected among the three populations. It is a likelihood method for detecting selective sweeps that involves jointly modeling the multi-locus allele frequency differentiation between two populations. This method is robust to detect selective sweeps and especially with regards to the uncertainty in the estimation of local recombination rate (Chen et al., 2010). Due to the absence of genomic position, the physical position (1 Mb $\approx$ 1 cM) was used. An in-house script based on overlapped segments of a maximum of 27 cM was designed to estimate and assemble XP-CLR scores using the whole set of bi-allelic variants. Overlapping regions of 2 cM were applied and the scores related to the extreme 1 cM were discarded, except at the starting and the end of chromosomes on the CHIR v1.0 assembly. XP-CLR scores were calculated using grid points spaced by 2500 bp with a maximum of 250 variants in a window of 0.5 cM and by down-weighting contributions of highly correlated variants ($r^2 > 0.95$) in the reference group.

To equilibrate the number of individuals per population, only 14 Black goats were randomly sampled among the 22. They were included with the 14 Draa and the 8 Northern individuals. Each

population was tested using a reference group including individuals from the two other populations. The 0.1% genomic regions with highest XP-CLR scores revealed by the analysis were identified and lists of genes partially or fully covered by these regions were then established. To ensure the coverage of short genes (i.e., genes shorter than the distance between adjacent grid points), two segments of 1500 bp each surrounding both sides of genes were also considered. NCBI databases were used to identify coordinates of the 20700 annotated autosomal genes on the CHIR v1.0 genome assembly (http://www.ncbi.nlm.nih.gov/genome?term=capra%20hircus).

Gene Ontology Enrichment Analyses

To explore the biological processes in which the top candidate genes are involved, Gene Ontology (GO) enrichment analyses were performed using the application GOrilla (Eden et al., 2009). The 12,669 goat genes associated with a GO term were used as background reference. Significance for each individual GO-identifier was assessed with *P*-values that were corrected using FDR *q*-value according to the Benjamini and Hochberg (1995) method. GO terms identified in each population were clustered into homogenous groups using REVIGO (Supek et al., 2011). Medium similarity among GO terms in a group was applied and the weight of each GO term was assessed by its *p*-value.

Results

Phylogeny of mtDNA Genomes

The whole mitochondrial genome was assembled successfully for 41 individuals and represented 16,651 bp length sequences. A total of 239 polymorphic sites were detected, which allowed discriminating 41 haplotypes. In an alternative complementary approach, the 481 bp length sequenced of the HVI segment of the control region was extracted, and this revealed 64 polymorphic sites identifying 40 single haplotypes. We constructed a network using the GTK + G + I model, which showed the best likelihood. The network (**Figure 2**) including the 22 reference haplotypes (i.e., haplogroups A, B, C, D, F, and G; Naderi et al., 2007) showed that the 40 haplotypes all belonged to the haplogroup A. We did not detect any coherent pattern of geographic structure among the haplotypes. There was also no clear differentiation of the haplotypes according to the three considered populations.

Neutral Diversity from WGS Data

The whole nuclear genomes were assembled on the goat reference genome CHIR1.0 along the 30 chromosomes. We mapped unambiguously 99.0% (±0.1%) of reads to the CHIR v1.0 assembly. However, the mapped reads properly paired constituted 90.3% (±0.1%) of reads in average. After the filtering processes, a total of 24,022,850 variants were found to be polymorphic in the total dataset among which 22,396,750 were SNPs and 1,626,100 were small indels. There were a total of 15,948,529 transitions and 6,540,478 transversions leading to a ts/tv ratio of 2.44. Due to differences in quality among individuals, the number of variants called per individual was at least 23,273,239 and 24,003,837 on average. As a consequence, a total of 23,059,968 variants showed no missing genotype over the 44 samples, among which 22,963,257 were biallelic.

Among the 24,022,850 polymorphic variants, only 12,024,778 variants were polymorphic within each of the three populations. The remaining variants were either polymorphic in only one or in two populations. When considering variants exclusive to each population, 3,704,299 were found polymorphic only in the Black population ($n = 22$), 1,887,724 only in the Draa population ($n = 14$) and 1,305,561 only in the Northern population ($n = 8$) (**Figure 3**). Rare variants (MAF < 0.05) represented a total of 10,892,203 (45.3%).

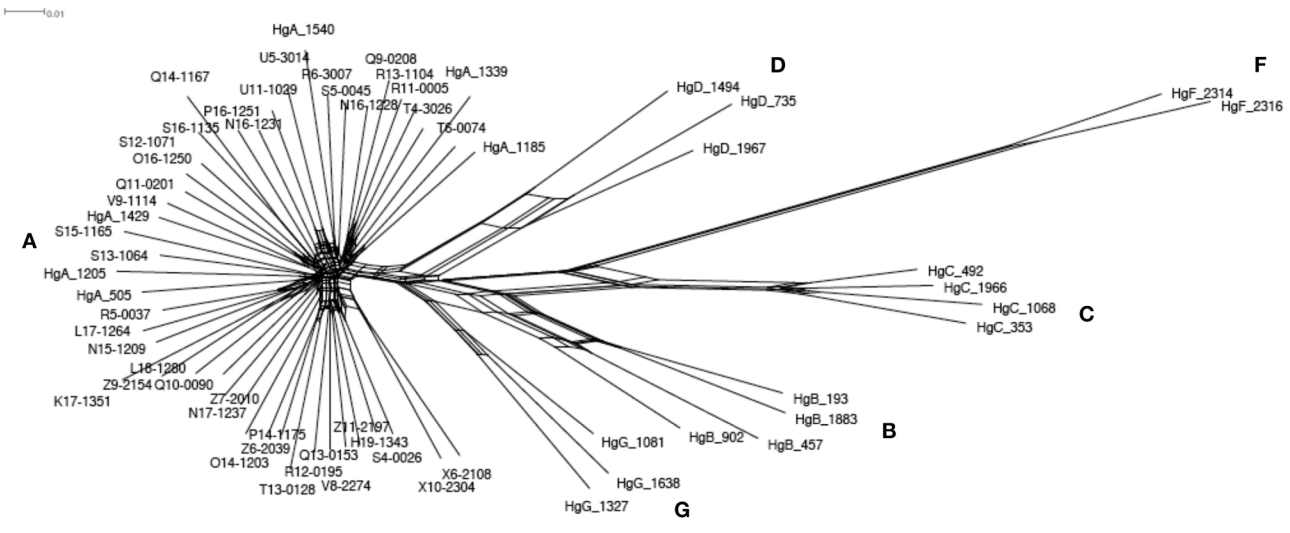


FIGURE 2 | Phylogenetic network based on the mitochondrial HVI segment of the control region. Sequences of 41 Moroccan goats and the 22 references representing the worldwide diversity (Naderi et al., 2007) were used. The 22 reference identifiers start with ≪ Hg ≫ and the following letter indicate which haplogroup each belongs to. The other identifiers correspond to the Moroccan goats. The red letters give the names of the 6 haplogroups.

Considering the 44 goats together, the average nucleotide diversity (π) calculated from 22,963,257 biallelic variants without missing genotype calls was 0.180. The Draa and the Black populations displayed similar π values amounting to 0.180 and 0.181 respectively. Among the 8 individuals representing the Northern population, π was slightly higher, amounting to 0.189. The observed percentage of heterozygote genotypes per individual (*Ho*) was 17.2% on average, ranging from 12.1% to 18.4%. The average inbreeding coefficient (F) was globally rather low (0.05 ± 0.07) and values were evenly distributed among populations. Similar average values were obtained for the Northern and Black populations (respectively 0.04 ± 0.07 and 0.04 ± 0.05). The Draa goats were slightly more inbred (average $F = 0.07 \pm 0.09$), particularly due to one individual showing $F = 0.32$.

We assessed LD by calculating the pairwise r^2 values between polymorphic sites for five chromosome regions. When withdrawing rare variants (i.e., MAF < 0.05), the average r^2 value was 0.40 for the first bin (0–0.2 kb) and decayed to less than 0.20 in 5.4 kb (**Figure 4**). Using the whole set of reliable variants, the average r^2 was 0.21 for the first bin and decreased rapidly to less than 0.20 in 239 bp of distance. Moreover, it decayed to less than 0.15 in about 1.33 kb distance (Figure S2).

Among the three populations, the level of genetic differentiation over the whole nuclear genome was extremely low (*Fst* = 0.0024). The pairwise *Fst* values varied from 0.001 for the Black-Draa comparison to 0.004 for the Northern-Draa comparison. Between the Black and Northern populations the pairwise *Fst* was 0.003.

The PCA analysis showed a very low population structure in the 44 Moroccan goats. The 3 main principal components (PCs) explained 5.8% of variance. The first PC tended to distinguish the Northern and Draa populations while the Black populations formed an in-between group. The second PC acted predominantly to distinguish individuals within the Draa and the Northern populations (Figure S1).

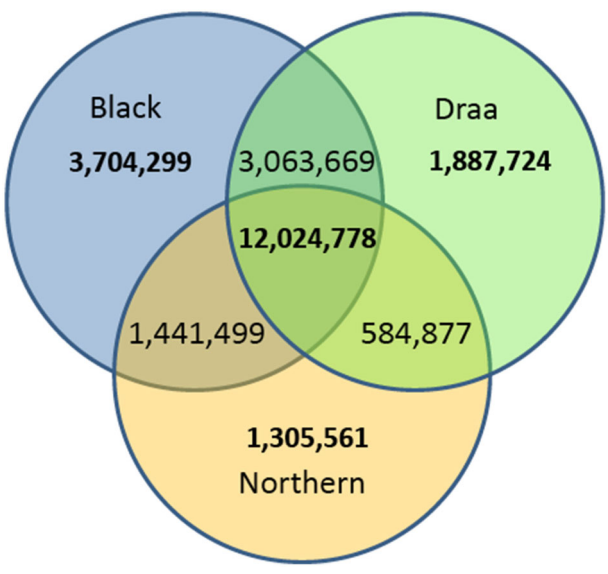


FIGURE 3 | Venn diagram of the number of polymorphic variants in the three Moroccan goat populations.

The clustering analysis of the genetic structure using sNMF (Frichot et al., 2014) showed that the 44 Moroccan goats belonging to the three populations were more likely represented by only one cluster according to the "crossentropy" criterion (lower values for $K = 1$). However, this criterion is not straightforward and when increasing until $K = 3$ we observed a weak pattern of genetic structure (**Figure 5**). At $K = 2$, the Northern goats were all strongly assigned to one distinct cluster. The second cluster was characterized by high assignment from the Draa population, except for two individuals that belong to the same cluster as the Northern goats. Finally, the Black goats showed variable levels of admixture between the two clusters (**Figure 5A**). When mapping the assignment results on a map we observed a geographic pattern with one cluster represented mainly in the north of Morocco (red component; **Figure 5B**) and the second cluster more present in the south (**Figure 5B**). At $K = 3$, the additional cluster was mostly represented in the Black goats which are located in the center of the sampling area (**Figure 5A**). The two other clusters still mostly represented the separation of Northern and Draa populations but the pattern was less evident. It was difficult to disentangle the relationship of genetic structure with populations and geography because the two factors were confounding.

Selection Signatures

We applied the XP-CLR genome scan method (Chen et al., 2010) on the whole genomes of 36 goats from the three phenotypic populations (14 Black, 14 Draa, and 8 Northern). We identified selective sweep genes in each population considering the top 0.1% genome-wide scores. Our approach highlighted respectively 142, 167, and 176 candidate genes in the Black, Draa, and Northern populations. The region showing the strongest XP-CLR score was located on chromosome 6 for the Black goats (Figure S3) and on chromosome 22 for the Northern goats (Figure S4), but they did not match any annotated gene. The annotated genes showing the strongest selective sweeps were *HTT*, *MSANTD1,* and *LOC102170765* in the Black goats, and *FOXP2, TRAP1* and *DNASE1* in the Northern goats (**Table 1**). In the Draa population, the highest XP-CLR scores corresponded to *LOC102190531, ADD3,* and *ASIP* genes (**Figure 6**). The enrichment categories of the identified candidate genes in the Black goats were associated with 15 GO terms (Table S2). They clustered into the following four differentiated categories by REVIGO (Supek et al., 2011): tube development, calcium ion transmembrane import into mitochondrion, negative regulation of transcription from RNA polymerase II promoter during mitosis and response to fatty acid. The enrichment of the identified candidate genes in Draa goats highlighted the significance of 25 GO terms (Table S3) clustering into five differentiated categories: regulation of respiratory gaseous exchange, behavior, postsynaptic membrane organization, protein localization to synapse, and neuron cell-cell adhesion. In the Northern goats, we did not find significant enrichment categories for the candidate genes identified.

TABLE 1 | Top-20 candidate genes under positive selection in each Moroccan goat population using the top-0.1% XP-CLR scores autosomal-wide cut-off level.

Black population					Draa					Northern population				
Gene	**Chr**	**Number of top-scores**	**Distance/ grid point**	**Higher score**	**Gene**	**Chr**	**Number of top-scores**	**Distance/ grid point**	**Higher score**	**Gene**	**Chr**	**Number of top-scores**	**Distance/ grid point**	**Higher score**
HTT	6	29	4739	82.6	*LOC102190531*	13	9	2493	94.0	*FOXP2*	4	14	33163	48.7
MSANTD1	6	3	5501	61.4	*ADD3*	26	24	5409	74.4	*TRAP1*	25	5	3977	42.8
LOC102170765	6	1	699	54.0	*ASIP*	13	2	995	71.3	*DNASE1*	25	4	2485	41.8
FAM160B1	26	2	42409	45.6	*VPS13B*	14	36	21697	70.1	*FAM227B*	10	9	25094	39.8
STRIP1	3	5	3069	43.6	*RALY*	13	9	5294	66.1	*CREBBP*	25	14	9497	38.6
NDUFA6	5	4	2786	41.8	*ICAM3*	7	5	1696	62.4	*PAPSS2*	26	1	43841	35.7
HNRNPA3	2	3	1183	40.3	*HIVEP2*	9	15	6353	61.4	*SLX4*	25	2	9472	32.4
KITLG	5	7	15223	39.7	*GGH*	14	10	2984	59.3	*PGM5*	8	9	23504	32.0
ALX3	3	1	7499	39.6	*PLSCR3*	19	2	1770	58.2	*BCAS3*	19	11	53928	31.4
IFT88	12	13	4139	39.6	*SOX6*	15	17	28387	54.8	*GALK2*	10	3	51372	31.2
XPO4	12	25	3141	39.5	*JARID2*	23	27	8506	52.9	*MAB21L1*	12	2	1213	31.0
VPS13B	14	16	48818	38.0	*NOL4*	24	6	78026	49.7	*NBEA*	12	31	21437	30.9
LOC102183160	14	1	298	37.5	*TIMP3*	5	3	4339	49.4	*LOC102182654*	25	1	2127	30.8
FLI1	29	6	10103	36.9	*EIF2S2*	13	2	7340	48.8	*LCOR*	26	7	8873	30.6
C4H7orf10	4	11	70095	35.7	*TTC39C*	24	10	10009	48.2	*RANBP10*	18	10	5817	29.5
TTC21A	22	3	11417	35.6	*PCBP3*	1	2	103547	46.0	*SLC12A4*	18	2	10049	28.0
LATS2	12	4	6149	34.3	*TTPA*	14	7	8967	45.0	*ROR1*	3	5	39379	27.8
NSMCE2	14	5	46323	33.7	*ASTN2*	8	11	79003	43.1	*MRPL54*	7	1	2005	27.7
ATG2B	21	3	25060	33.4	*GALNT7*	8	10	10949	42.2	*PDE1A*	2	2	146650	27.4
FAT2	7	2	42599	33.4	*MUC13*	1	7	3341	41.7	*KRT8*	5	2	3717	27.3

Coordinates of 20700 autosomal genes on the CHIR v1.0 goat assembly were used to identify candidate genes matching XP-CLR top scores. Genes were ranked according to the higher XP-CLR score. Chr, Chromosome. Number of top-scores, Number of grid points among the top-0.1% XP-CLR scores matching the gene. Distance/grid point: gene length/number of top-scores. Grid points in XP-CLR analysis were separated by 2.5 Kb.

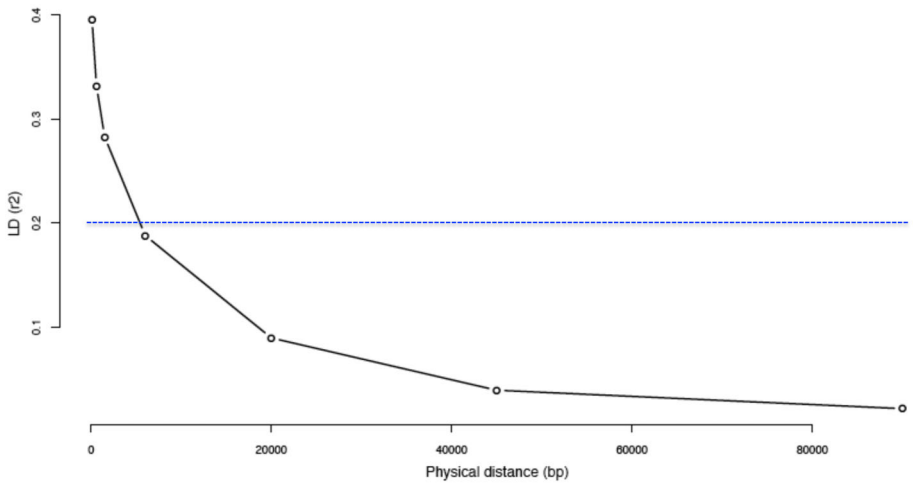


FIGURE 4 | Decay of linkage disequilibrium (r^2) as a function of physical distance by excluding "rare" variants. The Linkage Disequilibrium (*LD*) was calculated for the 44 Moroccan goats on 5 different segments of 2 Mb each on 5 different chromosomes. Inter-variant distances (bp) were binned and averaged into the classes: 0–0.2, 0.2–1, 1–2, 2–10, 10–30, 30–60, and 60–120 kb.

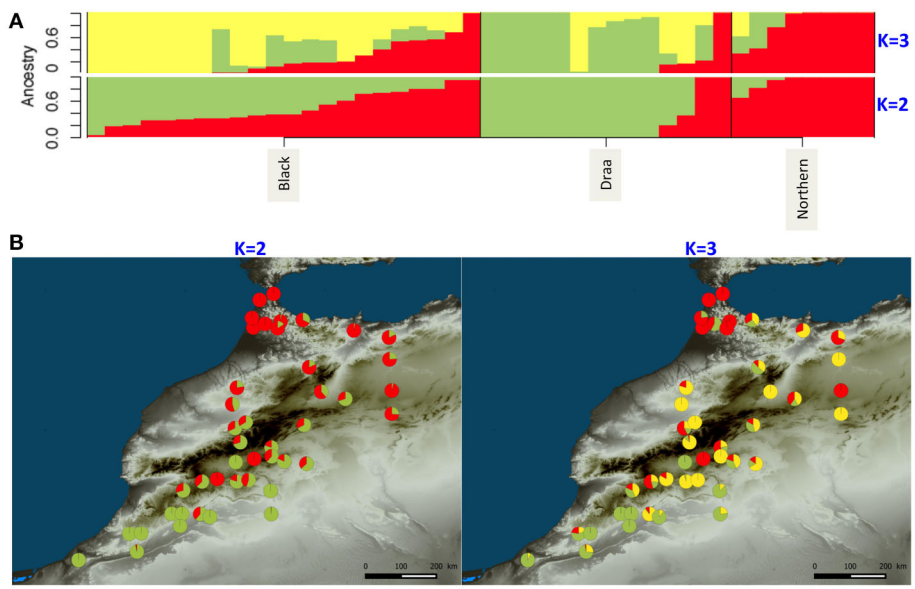


FIGURE 5 | WGS ancestry estimates for Moroccan goats for $K = 2$ and $K = 3$ clusters using sNMF (Frichot et al., 2014). (A) Each bar represents one individual. Different colors illustrate the assignment proportion (Q score) to each one of the assumed clusters. **(B)** Geographical distribution of individual *Q*-score values.

Discussion

Indigenous/traditional goats have been raised for a long time for various purposes and they have gradually accumulated several traits making them well adapted to their environments. The mechanisms underlying these adaptive traits have been poorly studied until now. The recent development of sequencing technologies has now made possible the sequencing of individuals' whole genomes and this may greatly expand our understanding of genomic diversity. Except for a few studies based on medium density SNP panels (about 50,000 SNPs) (Kijas et al., 2013; Tosser-Klopp et al., 2014), previous population genetic studies on goats have been limited to just a few dozens of markers (i.e., microsatellites). In this study we used variants spanning the whole genome to characterize indigenous goat populations of Morocco.

Mitochondrial Variation

Complete mitochondrial sequences were successfully assembled from a low portion of reads for 41 individuals. In terms of its ability to discriminate between the different haplotypes, the 481

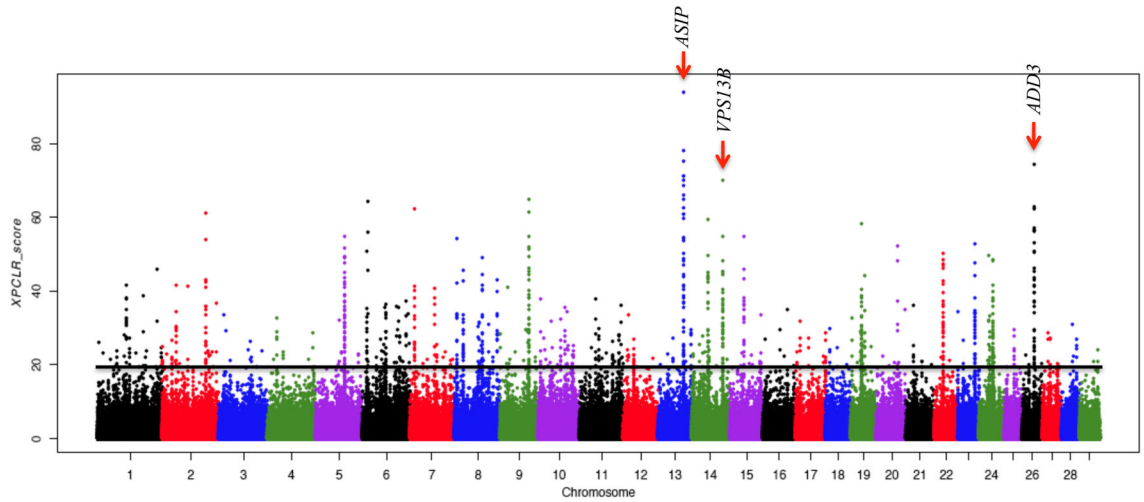


FIGURE 6 | Plot of XP-CLR scores along autosomes in selective sweep analysis for the Draa goat population. The horizontal line indicates a 0.1% autosomal-wide cut-off level. Red arrows and names indicate the top three candidate genes.

bp length of the HVI segment of the control region was almost as accurate as the whole mitochondrion sequence of 16,651 bp length from which it was extracted. Only a small difference in the total number of haplotypes defined was found (41 against 40 haplotypes respectively). This result shows that despite a low number of variable sites, the dense variability found in the control region (26.8% of the total number of variants for only 2.8% of the sequence length) concentrated most of the phylogenetic information. Thus, the HVI segment of the control region seems to be a good surrogate of the whole mitochondrial polymorphism. This study confirmed previous results based on the HVI segment of the control region (Pereira et al., 2009; Benjelloun et al., 2011) where Moroccan domestic goats showed only haplotypes from the A haplogroup (HgA). In a larger study using 2430 samples with a worldwide distribution, Naderi et al. (2007) found that most of the domestic goats displayed HgA (about 94%). Thus, it seems that the mitochondrial categorization in Morocco is rather representative of the rest of the world, even if the remaining haplogroups were not identified in our sampling. Besides this, the mtDNA diversity was weakly structured according to geography, as already reported by (Benjelloun et al., 2011) on the HVI region.

We did not find any clear structure of the mitochondrial haplotypes among the three populations. The high mitochondrial diversity characterizing these three populations probably indicates the diversity present in the first domesticated goats that arrived in Morocco and/or recurrent gene flows from diverse origins. According to (Pereira et al., 2009), Moroccan goat populations would have been established via two main colonization routes, one a North African land route and the other a Mediterranean maritime route across the Strait of Gibraltar. The high gene flows between populations, mediated by humans, would be ultimately responsible for the absence of structure across Morocco.

Nuclear Neutral Variation

Although the low percentage of the properly paired mapped reads (about 10%) in comparison with the percentage of mapped reads (about 99%) would illustrate a possible fragmentation of the genome assembly used, we identified many high confidence variants (approximately 24 million among which 6.8% were small indels) over the whole nuclear genomes of the 44 Moroccan goats studied. This is much higher than was found in all previous studies detecting variants in large sample cohorts from whole genome sequencing. For example, the human 1000 Genomes Project (Altshuler et al., 2012) detected approximately 15 million SNPs and 1 million short indels, while in the 1001 Genomes Project of *Arabidopsis thaliana* about 5 million SNPs and 81,000 small indels were found (Cao et al., 2011). The polymorphism detected in the Moroccan goats remains huge even when considered in proportion to the genome size of the species.

This huge number of variants did not show a strong genetic structure either among populations or over geographic space. The globally weak genetic structure suggests that extensive gene flows along with low level of selection have produced this pattern. Our findings contrast with most previous studies, which generally show a clear structure among goat breeds or populations (Cañon et al., 2006; Agha et al., 2008; Serrano et al., 2009; Di et al., 2011; Hassen et al., 2012; Kijas et al., 2013). Several reasons could explain this difference. First, most of the previous studies used microsatellite markers exhibiting high mutation rate. Thus, compared to SNP markers, microsatellites could more likely show imprints of recent demographic events such as differentiation between recently isolated populations. Moreover, the microsatellite markers generally used (Serrano et al., 2009; Di et al., 2011) were recommended by FAO and designed to exacerbate genetic differentiation among breeds, which was thus artificially inflated. In a more recent study, (Kijas et al., 2013) used a panel of SNP markers from a chip designed with animals representing industrial breeds for the SNP discovery (Tosser-Klopp et al., 2014).

In that case the results were certainly inflated by the ascertainment bias due to the chip design. However, it is also likely that in our case the demographic history of Moroccan goats differs from that of the breeds previously studied, and in particular from the ones compared at larger geographic scales such as Europe and Middle East (Cañon et al., 2006), or China, Iran and Africa (Di et al., 2011). The structured diversity found in these latter two studies would result from the strong isolation between countries. However, even at smaller scales the selection pressures exerted by breeding processes and husbandry practices may have increased isolation among breeds, and thus reinforced population differentiation compared to Morocco. The situation found in Morocco is close to the one described by Hassen et al. (2012) for six Ethiopian goat ecotypes, where even with microsatellite markers most of the diversity was found within populations, showing low levels of genetic differentiation. This result was explained by the existence of uncontrolled breeding strategies and agricultural extensive systems. In Morocco, it seems that goat populations have experienced moderate levels of selection and that most of the genetic diversity has been preserved during the breeding process which led to the three phenotypic populations. However, a weak genetic pattern was revealed by sNMF, which seems to be partially related to populations as well as geography. When mapping the clustering results (for $K = 3$, **Figure 4B**), a pattern appeared across Morocco, with Northern goats displaying a higher assignment probability to one distinct cluster. The Northern population is observably slightly more diverse than the others for which higher numbers of individuals were studied. This higher diversity and the slightly higher genetic differentiation of the Northern goats support the hypothesis of an influence of Iberian gene flows through the strait of Gibraltar in the North of Morocco (Analla and Serradilla, 1997).

The goal of our study was not to visualize the *LD* variations along chromosomes by covering all regions including centromeres and chromosomal inversions that are reportedly characterized by an elevated *LD* (Weetman et al., 2010; Marsden et al., 2014). Rather, we aimed to generate a global representation of *LD* across the genome by covering segments of 2 Mb in 5 different chromosomes taking all the reliable variants found from WGS data. Furthermore, knowing the effect of rare variants on *LD* estimation (Andolfatto and Przeworski, 2001) and to compare our findings with previous studies, we also estimated *LD* after discarding rare variants (MAF < 0.05). The extent of LD reported without rare variants ($r^2 < 0.20$ after 5.4 kb on average) is clearly shorter compared to all previous studies on farm animals, where it largely exceeds 10 kb for $r^2 = 0.20$ (Meadows et al., 2008; Villa-Angulo et al., 2009; Wade et al., 2009; McCue et al., 2012; Ai et al., 2013; Veroneze et al., 2013). In these studies, whole genome variants were not available and potential biases due to the use of SNP chips may partially explain the results. However, we consider that our finding would mainly result from the extensive breeding system favoring high gene flows among Moroccan goat populations/herds and low inbreeding and from the absence until now of strong selection during the breeding processes. Results on *LD* and genetic variability illustrate the important diversity present in indigenous populations in comparison with industrial breeds on which previous studies mainly focussed (e.g., Meadows et al., 2008; Villa-Angulo et al., 2009). This should be considered in the establishment of future programs aimed at improving these populations to preserve this highly valuable genetic diversity.

Beside this, when using the whole set of reliable variants we found a much lower *LD* ($r^{20}_{0.20} = 239$ bp). We do believe that this value should be considered in genome wide association and genome scan studies. Indeed most of studies remove rare variants for genotyping quality issues. In our case, the quality filtering produced reliable rare variants (about 45%) that would give a more realistic estimation of LD. To our knowledge, very few studies included rare variants to estimate *LD* (e.g., Mackay et al., 2012).

Selection Signatures in Moroccan Goat Populations

The weakly structured genetic diversity in Moroccan goats was suitable to detect selection signatures, avoiding possible false positives potentially generated by genetic structure. Despite a common genomic background and this weak population structure in Moroccan goats, the three main populations have been bred in various conditions and thereby have been subject to different anthropic and environmental selections in their recent history. As a result, they differ in their physiology, behavior and morphology. The observation of rapid phenotypic changes raises the question of the underlying genetic changes that would be shaped by selection. We identified numerous signatures of selection corresponding to genomic regions potentially under selection in each population.

A difficulty in identifying the genes or metabolic pathways under selection resides in the currently incomplete annotation of the goat genome. The stronger selective sweeps corresponded to regions in the Black population (chromosome 6) and in the Northern population (chromosome 22) matching un-annotated genes on the CHIR v1.0 assembly. This is probably due to either the incomplete annotation of the caprine genome or the fact that the selected functional mutations within each of these regions are not located within or close to a protein-coding gene. The incomplete genome annotation prevented us from identifying several known selected genes among Moroccan goat populations. For example, the *melanocortin-1 receptor* (*MC1R*) gene that is reported to be involved in coat color differentiation in goats (e.g., Fontanesi et al., 2009a) is not associated to any chromosome on the CHIR v1.0 assembly. Therefore, we were not able to detect its possible associated signal of selection in populations where the coat color is fixed knowing that we looked for selection signatures on autosomes only. Another problem consisted in the presence of several annotated genes that were not identified (i.e., no known orthologs, gene identifier starting with "LOC"). Thus, many genes potentially under selection could not be used in our GO enrichment analyses (e.g., the higher-score candidate gene in Draa population on Chromosome 13; **Table 1**). Despite these restrictions, we identified several sets of strong candidate genes in the three studied populations.

In the Black population the top-ranked candidate gene identified was *huntingtin* (*HTT;* **Table 1**). It has been comprehensively studied in humans where it is associated with Huntington's disease, an inherited autosomal dominant neurodegenerative disorder (Mende-Mueller et al., 2001; Sathasivam et al., 2013). The

HTT protein directly binds the endoplasmic reticulum (ER) and may play a role in autophagy triggered by ER stress (Atwal and Truant, 2008). Thus, we could speculate a possible involvement of this gene in the adaptation to physiological or pathological conditions leading to ER stress. This gene, among other candidates, was involved in the enrichment of GO terms *pattern specification process* (GO:0007389) and *organ development* (GO:0048513). These two categories were clustered together with the enriched *neuron maturation* term (GO:0042551) (Table S2). Hence, we could hypothesize a possible role of genes involved in these categories in some morphological traits specific to the Black goat population. Besides this, we noticed the enrichment of genes associated with the response to fatty acids GO terms (GO:0070542; GO:0071398). Candidate genes in these categories include *CPT1A* that encodes for a mitochondrial enzyme responsible for the formation of acyl carnitines that enables activated fatty acids to enter the mitochondria (van der Leij et al., 2000; Vaz and Wanders, 2002). The *SREBF1* gene encodes for a family of transcription factors (*SREBPs*) that regulate lipid homeostasis (Yokoyama et al., 1993; Eberle et al., 2004). The *GNPAT* gene encodes an essential enzyme to the synthesis of ether phospholipids. The last gene in these categories is *CPS1* and it encodes for a mitochondrial enzyme that catalyzes synthesis of carbamoyl phosphate (Aoshima et al., 2001). This suggests that selection acted upon the metabolism of fatty acids and lipids in the Black population, reflecting the possible development of an effective metabolism that could be linked to a higher amount of volatile fatty acids generated by the rumen microbial flora (Bergman, 1990).

In the Draa population, which is raised in oasis/desert areas and well adapted to high temperatures (Hossaini-Hilali and Mouslih, 2002), the enrichment of GO terms associated with the regulation of respiratory system and gaseous exchange categories (GO:0002087; GO:0043576; GO:0044065) would reflect the likely use of panting in evaporative heat loss. Goats could use panting as well as sweating for body thermo regulation according to the level of hydration and solar radiation (Dmiel and Robertshaw, 1983; Baker, 1989), and the type of regulatory system also depends on the breed/population (e.g., The Black Bedouin goats of Sinai Peninsula that use sweating in preference to panting) (Dmiel et al., 1979). Panting compared to sweating helps animals to better preserve their blood plasma volume (no losses of salt) and involves cooling of the blood passing the nasal area, which makes it possible to keep brain temperature lower than body temperature (Baker, 1989). Differences between Draa and Black populations in coat color, hair length and head size (larger in Black, Ibnelbachyr et al., in preparation) would support the hypothesis of different mechanisms of adaptation. Black goats would favor sweating and Draa panting as the more beneficial adaptation to warm environments. Mechanisms underlying dissipation should be further studied in these populations to elucidate the adaptive processes involved.

The enrichment of GO terms associated with lactate transport (GO:0015727; GO:0035873) (Table S3) in the Draa population could be linked to the stronger specific energetic demand associated with pregnancy and lactation in this population. The prolificacy in this population is much higher than in the rest of Moroccan goats (about 1.51 kids/birth vs. about 1 kid/birth; Ibnelbachyr et al., 2014). Thereby lactate transport may play a crucial role to meet this higher energetic requirement by shuttling lactate to a variety of sites where it could be oxidized directly, reconverted back to pyruvate or glucose and oxidized again, allowing the process of glycolysis to restart and ATP provision maintained (Brooks, 2000; Philp et al., 2005). This corroborates the higher concentration of lactate in cells during lactation than during dry-off period 5 weeks before parturition in cattle reported by Schwarm et al. (2013). Besides this, a top candidate gene in the Draa population was the *agouti signaling protein* (*ASIP*) (**Table 1**), which plays a key role in the modulation of hair and skin pigmentation in mammals (Lu et al., 1994; Furumura et al., 1996; Kanetsky et al., 2002) by antagonizing the effect of the *melanocortin-1 receptor gene* (*MC1R*) and promoting the synthesis of phaeomelanin, a yellow–red pigment (Hida et al., 2009). *ASIP* was associated with different coat colors in cattle and sheep (Seo et al., 2007; Norris and Whan, 2008). The strong selective sweep related to this gene could be linked to the higher variation in Draa's coat color when compared to other populations (Ibnelbachyr et al., in preparation). This variation in coat color was highly represented in the 14 Draa samples used in this study (Table S4). However, previous studies focussing on this gene identified an important polymorphism in worldwide goat breeds without any clear association with differences in coat color (Badaoui et al., 2011; Adefenwa et al., 2013). Fontanesi et al. (2009b) reported the presence of a copy number variation (CNV) affecting *ASIP* and *AHCY* genes, and might be associated to the white color in Girgentana and Saneen breeds. Nevertheless, the design of our study was not adapted to identify CNV and we cannot link the selection signature detected here in this gene to the findings of this study.

In the Northern population, no GO term was enriched but the second ranked candidate gene identified was *TRAP1*, which encodes a mitochondrial chaperone protein (Felts et al., 2000). Under stress conditions this gene was shown to protect cells from reactive oxygen species, (ROS)-induced apoptosis and senescence (Im et al., 2007; Pridgeon et al., 2007). Such regulation of the cellular stress response would play a role in the adaptation of this population to harsh environments (e.g., mountainous areas in the North of Morocco).

Finally, several strong signals of selection pointed to genes or pathways for which possible functions remained ambiguous. For example in the Northern population, the strong signal of selection associated with *FOXP2*, which encodes for a regulatory protein, is required for proper development of language in Humans (Lai et al., 2001), song learning in songbirds (Haesler et al., 2004), and learning of rapid movement sequences in mice (Groszer et al., 2008). This gene could be involved in learning but its possible functions in goats cannot be hypothesized easily. A similar case was found in the Draa population for which GO categories linked to behavior and vocalization behavior (GO:0071625; GO:0030534; GO:0007610) were enriched. We were not able to predict the possible functions of these genes. Furthermore, the *NR6A1* gene that was identified potentially under selection in Draa (within the top 0.1% XP-CLR scores) was previously associated with the number of vertebras in pigs (Mikawa et al., 2007; Rubin et al., 2012). Considering the larger body length and size in

this population in comparison with the Black population (Ibnelbachyr et al., in preparation), we could hypothesize a similar role of this gene in the body elongation in goats. A future characterization of this morphologic trait in Draa goats would confirm or refute this hypothesis.

Conclusion

Our study characterized whole genome variation in the main goat indigenous populations at a countrywide scale in an unprecedented way. The whole genome data and the wide geographic spread of animals allowed for a precise characterization of the distribution of genomic diversity in various populations. The position of Morocco has made it subject to various colonization waves for domestic animals. Additionally, previous and present management schemes have favored gene flow between goat populations. This created and maintained a very high level of total genetic diversity that is weakly structured according to geography and populations. A part of the overall diversity corresponded to potentially adaptive variation, as several genes appeared to be under selection. The different populations studied appeared to bear specific adaptations, even when submitted to similar conditions such as those related to a warm/desert context. This would demonstrate the potential of different indigenous livestock populations to constitute complementary reservoirs of possibly adaptive diversity that would be highly valuable in the context of global environmental changes. However, these populations are threatened due to their substitution by more productive cosmopolitan breeds that should not have the potential to become locally adapted to harsh environments. It is thus extremely important to promote the sustainable management of these genetic resources with emphasis on both overall neutral and adaptive diversity. This study has also identified several genes as potentially under selection and further studies are needed to depict the underlying mechanisms.

Accession numbers

The accession numbers of the 44 samples in the BioSamples archive, the accession numbers of the sequencing data and aligned bam files in the ENA archive are reported in the Table S1. The variant calls and genotype calls used in this paper are archived in the European Variation Archive with accession ID ERZ020631.

Author Contributions

PT, FP, SJ, PF designed the study. PT and FP supervised the study. BB, MB, MI, MC, AB, AC sampled individuals. AA, SE produced whole genome sequences. BB, FJA, IS, FB, EC, SS, KL, MI, LC analyzed the data and interpreted the results. BB, FJA, FP, KL, SJ, IS, AA wrote the Manuscript. All authors revised and accepted the final version of the manuscript.

Funding

This work was funded by the UE FP7 project *NEXTGEN* "Next generation methods to preserve farm animal biodiversity by optimizing present and future breeding options"; grant agreement no. *244356*.

Acknowledgments

We are greatful to R. Hadria, M. Laghmir, L. Haounou, E. Hafiani, E. Sekkour, M. ElOuatiq, A Dadouch, A. Lberji, C. Errouidi and M. Bouali for their great efforts in sampling goats in Morocco. We thank T. Benabdelouahab for his contribution in the production of some maps. We also thank the two reviewers for valuable suggestions to improve this paper.

Supplementary Material

The Supplementary Material for this article can be found online at: http://www.frontiersin.org/journal/10.3389/fgene.2015.00107/abstract

References

Adefenwa, M. A., Peters, S. O., Agaviezor, B. O., Wheto, M., Adekoya, K. O., Okpeku, M., et al. (2013). Identification of single nucleotide polymorphisms in the agouti signaling protein (ASIP) gene in some goat breeds in tropical and temperate climates. *Mol. Biol. Rep.* 40, 4447–4457. doi: 10.1007/s11033-013-2535-1

Agha, S. H., Pilla, F., Galal, S., Shaat, I., D'andrea, M., Reale, S., et al. (2008). Genetic diversity in Egyptian and Italian goat breeds measured with microsatellite polymorphism. *J. Anim. Breed. Genet.* 125, 194–200. doi: 10.1111/j.1439-0388.2008.00730.x

Ai, H., Huang, L., and Ren, J. (2013). Genetic diversity, linkage disequilibrium and selection signatures in chinese and Western pigs revealed by genome-wide SNP markers. *PLoS ONE* 8:e56001. doi: 10.1371/journal.pone.0056001

Altshuler, D. M., Durbin, R. M., Abecasis, G. R., Bentley, D. R., Chakravarti, A., Clark, A. G., et al. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature* 491, 56–65. doi: 10.1038/nature11632

Analla, M., and Serradilla, J. M. (1997). "Problems of selection criteria and genetic evaluations of the goat population in the north of Morocco," in *Data Collection and Definition of Objectives in Sheep and Goat Breeding Programmes: New Prospects. Zaragoza: CIHEAM. Options Méditerranéennes: Série A. Séminaires Méditerranéens; n. 33*, eds D. Gabiña and L. Bodin (Toulouse: Options Méditerranéennes CIHEAM), 153–156

Andolfatto, P., and Przeworski, M. (2001). Regions of lower crossing over harbor more rare variants in African populations of *Drosophila melanogaster*. *Genetics* 158, 657–665.

Aoshima, T., Kajita, M., Sekido, Y., Kikuchi, S., Yasuda, I., Saheki, T., et al. (2001). Novel mutations (H337R and 238-362del) in the CPS1 gene cause carbamoyl phosphate synthetase I deficiency. *Hum. Hered.* 52, 99–101. doi: 10.1159/000053360

Atwal, R. S., and Truant, R. (2008). A stress sensitive ER membrane-association domain in Huntingtin protein defines a potential role for Huntingtin in the regulation of autophagy. *Autophagy* 4, 91–93. doi: 10.4161/auto.5201

Badaoui, B., D'Andrea, M., Pilla, F., Capote, J., Zidi, A., Jordana, J., et al. (2011). Polymorphism of the goat agouti signaling protein gene and its relationship with coat color in Italian and Spanish breeds. *Biochem. Genet.* 49, 523–532. doi: 10.1007/s10528-011-9427-7

Baker, M. A. (1989). Effects of dehydration and rehydration on thermoregulatory sweating in goats. *J. Physiol. Lond.* 417, 421–435. doi: 10.1113/jphysiol.1989.sp017810

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate—a practical and powerful approach to multiple testing. *J. R. Statist. Soc. B* 57, 289–300.

Benjelloun, B., Pompanon, F., Ben Bati, M., Chentouf, M., Ibnelbachyr, M., El Amiri, B., et al. (2011). Mitochondrial DNA polymorphism in Moroccan goats. *Small Ruminant Res.* 98, 201–205. doi: 10.1016/j.smallrumres.2011.03.041

Benlekhal, A., and Tazi, M. S. (1996). "Situation du secteur caprin au Maroc," in *Les Perspectives de Développement de la Filière Lait de Chèvre Dans le Bassin Méditerranéen. Une Réflexion Collective Appliquée au cas Marocain. (Goat Sector in Morocco. in: Perspectives for Goat Milk Production in the Mediterranean Area. A Collective Reflexion Applied to the Moroccan Situation.)*, Rome: FAO Animal Production and Health paper, FAO.

Bergman, E. N. (1990). Energy contributions of volatile fatty-acids from the gastrointestinal-tract in various species. *Physiol. Rev.* 70, 567–590.

Brooks, G. A. (2000). Intra- and extra-cellular lactate shuttles. *Med. Sci. Sports Exerc.* 32, 790–799. doi: 10.1097/00005768-200004000-00011

Browning, B. L., and Browning, S. R. (2013). Detecting identity by descent and estimating genotype error rates in sequence data. *Am. J. Hum. Genet.* 93, 840–851. doi: 10.1016/j.ajhg.2013.09.014

Cañon, J., Garcia, D., Garcia-Atance, M. A., Obexer-Ruff, G., Lenstra, J. A., Ajmone-Marsan, P., et al. (2006). Geographical partitioning of goat diversity in Europe and the Middle East. *Anim. Genet.* 37, 327–334. doi: 10.1111/j.1365-2052.2006.01461.x

Cao, J., Schneeberger, K., Ossowski, S., Guenther, T., Bender, S., Fitz, J., et al. (2011). Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. *Nat. Genet.* 43, 956–U960. doi: 10.1038/ng.911

Chen, H., Patterson, N., and Reich, D. (2010). Population differentiation as a test for selective sweeps. *Genome Res.* 20, 393–402. doi: 10.1101/gr.100545.109

Chentouf, M., Bister, J. L., and Boulanouar, B. (2011). Reproduction characteristics of North Moroccan indigenous goats. *Small Ruminant Res.* 98, 185–188 doi: 10.1016/j.smallrumres.2011.03.037

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., Depristo, M. A., et al. (2011). The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158. doi: 10.1093/bioinformatics/btr330

Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* 9, 772–772. doi: 10.1038/nmeth.2109

Depristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* 43, 491. doi: 10.1038/ng.806

Di, R., Vahidi, S. M. F., Ma, Y. H., He, X. H., Zhao, Q. J., Han, J. L., et al. (2011). Microsatellite analysis revealed genetic diversity and population structure among Chinese cashmere goats. *Anim. Genet.* 42, 428–431. doi: 10.1111/j.1365-2052.2010.02072.x

Dmiel, R., and Robertshaw, D. (1983). The control of panting and sweating in the Black Bedouin goat—a comparison of 2 modes of imposing a heat load. *Physiol. Zool.* 56, 404–411.

Dmiel, R., Robertshaw, D., and Choshniak, I. (1979). Sweat gland secretion in the black bedouin goat. *Physiol. Zool.* 52, 558–564.

Dong, Y., Xie, M., Jiang, Y., Xiao, N., Du, X., Zhang, W., et al. (2013). Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (Capra hircus). *Nat. Biotechnol.* 31, 135–141. doi: 10.1038/nbt.2478

Eberle, D., Hegarty, B., Bossard, P., Ferre, P., and Foufelle, F. (2004). SREBP transcription factors: master regulators of lipid homeostasis. *Biochimie* 86, 839–848. doi: 10.1016/j.biochi.2004.09.018

Eden, E., Navon, R., Steinfeld, I., Lipson, D., and Yakhini, Z. (2009). GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* 10:48. doi: 10.1186/1471-2105-10-48

Felts, S. J., Owen, B. A. L., Nguyen, P., Trepel, J., Donner, D. B., and Toft, D. O. (2000). The hsp90-related protein TRAP1 is a mitochondrial protein with distinct functional properties. *J. Biol. Chem.* 275, 3305–3312. doi: 10.1074/jbc.275.5.3305

Fontanesi, L., Beretti, F., Riggio, V., Dall'Olio, S., Gonzalez, E. G., Finocchiaro, R., et al. (2009a). Missense and nonsense mutations in melanocortin 1 receptor (MC1R) gene of different goat breeds: association with red and black coat colour phenotypes but with unexpected evidences. *BMC Genet.* 10:47. doi: 10.1186/1471-2156-10-47

Fontanesi, L., Beretti, F., Riggio, V., Gonzalez, E. G., Dall'Olio, S., Davoli, R., et al. (2009b). Copy number variation and missense mutations of the agouti signaling protein (ASIP) gene in goat breeds with different coat colors. *Cytogenet. Genome Res.* 126, 333–347. doi: 10.1159/000268089

Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G., and Francois, O. (2014). Fast and efficient estimation of individual ancestry coefficients. *Genetics* 196, 973. doi: 10.1534/genetics.113.160572

Furumura, M., Sakai, C., Abdelmalek, Z., Barsh, G. S., and Hearing, V. J. (1996). The interaction of agouti signal protein and melanocyte stimulating hormone to regulate melanin formation in mammals. *Pigment Cell Res.* 9, 191–203. doi: 10.1111/j.1600-0749.1996.tb00109.x

Garrison, E., and Marth, G. (2012). *Haplotype-Based Variant Detection from Short-Read Sequencing*. arXiv.

Groszer, M., Keays, D. A., Deacon, R. M. J., De Bono, J. P., Prasad-Mulcare, S., Gaub, S., et al. (2008). Impaired synaptic plasticity and motor learning in mice with a point mutation implicated in human speech deficits. *Curr. Biol.* 18, 354–362. doi: 10.1016/j.cub.2008.01.060

Haesler, S., Wada, K., Nshdejan, A., Morrisey, E. E., Lints, T., Jarvis, E. D., et al. (2004). FoxP2 expression in avian vocal learners and non-learners. *J. Neurosci.* 24, 3164–3175. doi: 10.1523/JNEUROSCI.4369-03.2004

Hassen, H., Lababidi, S., Rischkowsky, B., Baum, M., and Tibbo, M. (2012). Molecular characterization of Ethiopian indigenous goat populations. *Trop. Anim. Health Prod.* 44, 1239–1246. doi: 10.1007/s11250-011-0064-2

Hida, T., Wakamatsu, K., Sviderskaya, E. V., Donkin, A. J., Montoliu, L., Lamoreux, M. L., et al. (2009). Agouti protein, mahogunin, and attractin in pheomelanogenesis and melanoblast-like alteration of melanocytes: a cAMP-independent pathway. *Pigment Cell Melanoma Res.* 22, 623–634. doi: 10.1111/j.1755-148X.2009.00582.x

Hossaini-Hilali, J., and Mouslih, Y. (2002). La chèvre Draa. Potentiel de production et caractéristiques d'adaptation aux contraintes de l'environnement aride. *Anim. Genet. Res. Inf.* 32, 49–56. doi: 10.1017/S1014233900001553

Hossainihilali, J., Benlamlih, S., and Dahlborn, K. (1993). Fluid balance and milk secretion in the fed and feed-deprived black moroccan goat. *Small Ruminant Res.* 12, 271–285. doi: 10.1016/0921-4488(93)90063-N

Huson, D. H., and Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* 23, 254–267. doi: 10.1093/molbev/msj030

Ibnelbachyr, M., Boujenane, I., Chikhi, A., and et Er-rouidi, C. (2014). "Le système de conduite de 3 chevrotages en 2 ans: Outil de gestion moderne de la conduite technique de la race caprine locale Draa," in *Technology Creation and Transfer in Small Ruminants: Roles of Research, Developement Services and Farmer Associations. Options Méditerranéennes: Série A. Séminaires Méditerranéens; n. 108*, eds M. Chentouf, A. López-Francos, M. Bengoumi, and D. Gabiña (Tangier: Options Méditerranéennes CIHEAM), 199–207.

Im, C. N., Lee, J. S., Zheng, Y., and Seo, J. S. (2007). Iron chelation study in a normal human hepatocyte cell line suggests that tumor necrosis factor receptor-associated protein 1 (TRAP1) regulates production of reactive oxygen species. *J. Cell. Biochem.* 100, 474–486. doi: 10.1002/jcb.21064

Jombart, T., and Ahmed, I. (2011). adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* 27, 3070–3071. doi: 10.1093/bioinformatics/btr521

Kanetsky, P. A., Swoyer, J., Panossian, S., Holmes, R., Guerry, D., and Rebbeck, T. R. (2002). A polymorphism in the agouti signaling protein gene is associated with human pigmentation. *Am. J. Hum. Genet.* 70, 770–775. doi: 10.1086/339076

Kidd, J. M., Gravel, S., Byrnes, J., Moreno-Estrada, A., Musharoff, S., Bryc, K., et al. (2012). Population genetic inference from personal genome data: impact of ancestry and admixture on human genomic variation. *Am. J. Hum. Genet.* 91, 660–671. doi: 10.1016/j.ajhg.2012.08.025

Kijas, J. W., Ortiz, J. S., McCulloch, R., James, A., Brice, B., Swain, B., et al. (2013). Genetic diversity and investigation of polledness in divergent goat populations using 52088 SNPs. *Anim. Genet.* 44, 325–335. doi: 10.1111/age.12011

Lai, C. S. L., Fisher, S. E., Hurst, J. A., Vargha-Khadem, F., and Monaco, A. P. (2001). A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature* 413, 519–523. doi: 10.1038/35097076

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

Librado, P., and Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25, 1451–1452. doi: 10.1093/bioinformatics/btp187

Lu, D. S., Willard, D., Patel, I. R., Kadwell, S., Overton, L., Kost, T., et al. (1994). Agouti protein is an antagonist of the melanocyte-stimulating-hormone receptor. *Nature* 371, 799–802. doi: 10.1038/371799a0

Mackay, T. F. C., Richards, S., Stone, E. A., Barbadilla, A., Ayroles, J. F., Zhu, D. H., et al. (2012). The Drosophila melanogaster genetic reference panel. *Nature* 482, 173–178. doi: 10.1038/nature10811

Marsden, C. D., Lee, Y., Kreppel, K., Weakley, A., Cornel, A., Ferguson, H. M., et al. (2014). Diversity, differentiation, and linkage disequilibrium: prospects for association mapping in the malaria vector Anopheles arabiensis. *G3* 4, 121–131. doi: 10.1534/g3.113.008326

McCue, M. E., Bannasch, D. L., Petersen, J. L., Gurr, J., Bailey, E., Binns, M. M., et al. (2012). A high density SNP array for the domestic horse and extant perissodactyla: utility for association mapping, genetic diversity, and phylogeny studies. *PLoS Genet.* 8:e1002451. doi: 10.1371/journal.pgen.1002451

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010). The genome analysis toolkit: a mapreduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303. doi: 10.1101/gr.107524.110

Meadows, J. R. S., Chan, E. K. F., and Kijas, J. W. (2008). Linkage disequilibrium compared between five populations of domestic sheep. *BMC Genet.* 9:61. doi: 10.1186/1471-2156-9-61

Mende-Mueller, L. M., Toneff, T., Hwang, S. R., Chesselet, M. F., and Hook, V. Y. H. (2001). Tissue-specific proteolysis of huntingtin (htt) in human brain: evidence of enhanced levels of N- and C-terminal htt fragments in Huntington's disease striatum. *J. Neurosci.* 21, 1830–1837.

Mikawa, S., Morozumi, T., Shimanuki, S. I., Hayashi, T., Uenishi, H., Domukai, M., et al. (2007). Fine mapping of a swine quantitative trait locus for number of vertebrae and analysis of an orphan nuclear receptor, germ cell nuclear factor (NR6A1). *Genome Res.* 17, 586–593. doi: 10.1101/gr.6085507

Naderi, S., Rezaei, H.-R., Pompanon, F., Blum, M. G. B., Negrini, R., Naghash, H.-R., et al. (2008). The goat domestication process inferred from large-scale mitochondrial DNA analysis of wild and domestic individuals. *Proc. Natl. Acad. Sci. U.S.A.* 105, 17659–17664. doi: 10.1073/pnas.0804782105

Naderi, S., Rezaei, H. R., Taberlet, P., Zundel, S., Rafat, S. A., Naghash, H. R., et al. (2007). Large-scale mitochondrial DNA analysis of the domestic goat reveals six haplogroups with high diversity. *PLoS ONE* 2:e1012. doi: 10.1371/journal.pone.0001012

Norris, B. J., and Whan, V. A. (2008). A gene duplication affecting expression of the ovine ASIP gene is responsible for white and black sheep. *Genome Res.* 18, 1282–1293. doi: 10.1101/gr.072090.107

Pereira, F., Queiros, S., Gusmao, L., Nijman, I. J., Cuppen, E., Lenstra, J. A., et al. (2009). Tracing the history of goat pastoralism: new clues from mitochondrial and y chromosome DNA in North Africa. *Mol. Biol. Evol.* 26, 2765–2773. doi: 10.1093/molbev/msp200

Philp, A., Macdonald, A. L., and Watt, P. W. (2005). Lactate - a signal coordinating cell and systemic function. *J. Exp. Biol.* 208, 4561–4575. doi: 10.1242/jeb.01961

Porter, V. (2002). *Mason's World Dictionary of Livestock Breeds, Types and Varieties, 5th Edn.* Wallingford: CABI Publishing.

Pridgeon, J. W., Olzmann, J. A., Chin, L. S., and Li, L. (2007). PINK1 protects against oxidative stress by phosphorylating mitochondrial chaperone TRAP1. *PLoS Biol.* 5, 1494–1503. doi: 10.1371/journal.pbio.0050172

Rubin, C. J., Megens, H. J., Barrio, A. M., Maqbool, K., Sayyab, S., Schwochow, D., et al. (2012). Strong signatures of selection in the domestic pig genome. *Proc. Natl. Acad. Sci. U.S.A.* 109, 19529–19536. doi: 10.1073/pnas.1217149109

Sathasivam, K., Neueder, A., Gipson, T. A., Landles, C., Benjamin, A. C., Bondulich, M. K., et al. (2013). Aberrant splicing of HTT generates the pathogenic exon 1 protein in Huntington disease. *Proc. Natl. Acad. Sci. U.S.A.* 110, 2366–2370. doi: 10.1073/pnas.1221891110

Schwarm, A., Viergutz, T., Kuhla, B., Hammon, H. M., and Schweigel-Rontgen, M. (2013). Fuel feeds function: energy balance and bovine peripheral blood mononuclear cell activation. *Comp. Biochem. Physiol. A Mol. Integr. Physiol.* 164:101–110. doi: 10.1016/j.cbpa.2012.10.009

Seo, K. S., Mohanty, T. R., Choi, T. J., and Hwang, I. (2007). Biology of epidermal and hair pigmentation in cattle: a mini-review. *Vet. Dermatol.* 18, 392–400. doi: 10.1111/j.1365-3164.2007.00634.x

Serrano, M., Calvo, J. H., Martinez, M., Marcos-Carcavilla, A., Cuevas, J., Gonzalez, C., et al. (2009). Microsatellite based genetic diversity and population structure of the endangered Spanish Guadarrama goat breed. *BMC Genet.* 10:61. doi: 10.1186/1471-2156-10-61

Snyder, M., Du, J., and Gerstein, M. (2010). Personal genome sequencing: current approaches and challenges. *Genes Dev.* 24, 423–431. doi: 10.1101/gad.1864110

Supek, F., Bosnjak, M., Skunca, N., and Smuc, T. (2011). REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS ONE* 6:e21800. doi: 10.1371/journal.pone.0021800

Taberlet, P., Valentini, A., Rezaei, H. R., Naderi, S., Pompanon, F., Negrini, R., et al. (2008). Are cattle, sheep, and goats endangered species? *Mol. Ecol.* 17, 275–284. doi: 10.1111/j.1365-294X.2007.03475.x

Tosser-Klopp, G., Bardou, P., Bouchez, O., Cabau, C., Crooijmans, R., Dong, Y., et al. (2014). Design and characterization of a 52K SNP chip for goats. *PLoS ONE* 9:e86227. doi: 10.1371/journal.pone.0086227

Tresset, A., and Vigne, J. D. (2011). Last hunter-gatherers and first farmers of Europe. *C. R. Biol.* 334, 182–189. doi: 10.1016/j.crvi.2010.12.010

van der Leij, F. R., Huijkman, N. C. A., Boomsma, C., Kuipers, J. R. G., and Bartelds, B. (2000). Genomics of the human carnitine acyltransferase genes. *Mol. Genet. Metab.* 71, 139–153. doi: 10.1006/mgme.2000.3055

Vaz, F. M., and Wanders, R. J. A. (2002). Carnitine biosynthesis in mammals. *Biochem. J.* 361, 417–429. doi: 10.1042/0264-6021:3610417

Veroneze, R., Lopes, P. S., Guimaraes, S. E. F., Silva, F. F., Lopes, M. S., Harlizius, B., et al. (2013). Linkage disequilibrium and haplotype block structure in six commercial pig lines. *J. Anim. Sci.* 91, 3493–3501. doi: 10.2527/jas.2012-6052

Villa-Angulo, R., Matukumalli, L. K., Gill, C. A., Choi, J., Van Tassell, C. P., et al. (2009). High-resolution haplotype block structure in the cattle genome. *BMC Genet.* 10:19. doi: 10.1186/1471-2156-10-19

Wade, C. M., Giulotto, E., Sigurdsson, S., Zoli, M., Gnerre, S., Imsland, F., et al. (2009). Genome sequence, comparative analysis, and population genetics of the domestic horse. *Science* 326, 865–867. doi: 10.1126/science.1178158

Weetman, D., Wilding, C. S., Steen, K., Morgan, J. C., Simard, F., and Donnelly, M. J. (2010). Association mapping of insecticide resistance in wild *Anopheles gambiae* populations: major variants identified in a low-linkage disequilbrium genome. *PLoS ONE* 5:e13140. doi: 10.1371/journal.pone.0013140

Weir, B. S., and Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population-structure. *Evolution* 38, 1358–1370. doi: 10.2307/2408641

Yokoyama, C., Wang, X. D., Briggs, M. R., Admon, A., Wu, J., Hua, X. X., et al. (1993). SREBP-1, a basic-helix-loop-helix-leucine zipper protein that controls transcription of the low-density-lipoprotein receptor gene. *Cell* 75, 187–197. doi: 10.1016/S0092-8674(05)80095-9

Zeder, M. A. (2005). "A view from the Zagros: new perspectives on livestock domestication in the Fertile Crescent," in *The First Steps of Animal Domestication. New Archaeological Approaches,* eds J. D. Vigne, J. Peters, and D. Helmer (Oxford: Oxbow Books), 125–146.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Benjelloun, Alberto, Streeter, Boyer, Coissac, Stucki, BenBati, Ibnelbachyr, Chentouf, Bechchari, Leempoel, Alberti, Engelen, Chikhi, Clarke, Flicek, Joost, Taberlet, Pompanon and NextGen Consortium. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 21 April 2015
doi: 10.3389/fgene.2015.00137

Trypanosomosis: potential driver of selection in African cattle

*Anamarija Smetko[1,2†], Albert Soudre[1,3†], Katja Silbermayr[4], Simone Müller[5], Gottfried Brem[5], Olivier Hanotte[6], Paul J. Boettcher[7,8], Alessandra Stella[9], Gábor Mészáros[1], Maria Wurzinger[1], Ino Curik[10], Mathias Müller[5], Jörg Burgstaller[5,11] and Johann Sölkner[1]**

[1] Division of Livestock Sciences, Department of Sustainable Agricultural Systems, BOKU-University of Natural Resources and Life Sciences Vienna, Vienna, Austria, [2] Croatian Agricultural Agency, Zagreb, Croatia, [3] Ecole Normale Supérieure, Université de Koudougou, Koudougou, Burkina Faso, [4] Institute of Parasitology, University of Veterinary Medicine, Vienna, Austria, [5] Institute of Animal Breeding and Genetics, University of Veterinary Medicine, Vienna, Austria, [6] School of Life Sciences, University of Nottingham, Nottingham, UK, [7] Animal Production and Health Division, Agriculture and Consumer Protection Department, Food and Agriculture Organization of the United Nations, Rome, Italy, [8] FAO/IAEA Joint Division on Nuclear Techniques in Food and Agriculture, Vienna, Austria, [9] Parco Tecnologico Padano, Lodi, Italy, [10] Faculty of Agriculture, University of Zagreb, Zagreb, Croatia, [11] Biotechnology in Animal Production, Department for Agrobiotechnology, IFA Tulln, Tulln, Austria

Trypanosomosis is a serious cause of reduction in productivity of cattle in tsetse-fly infested areas. Baoule and other local *Taurine* cattle breeds in Burkina Faso are trypanotolerant. Zebuine cattle, which are also kept there are susceptible to trypanosomosis but bigger in body size. Farmers have continuously been intercrossing Baoule and Zebu animals to increase production and disease tolerance. The aim of this study was to compare levels of zebuine and taurine admixture in genomic regions potentially involved in trypanotolerance with background admixture of composites to identify differences in allelic frequencies of tolerant and non-tolerant animals. The study was conducted on 214 animals (90 Baoule, 90 Zebu, and 34 composites), genotyped with 25 microsatellites across the genome and with 155 SNPs in 23 candidate regions. Degrees of admixture of composites were analyzed for microsatellite and SNP data separately. Average Baoule admixture based on microsatellites across the genomes of the Baoule- Zebu composites was 0.31, which was smaller than the average Baoule admixture in the trypanosomosis candidate regions of 0.37 ($P = 0.15$). Fixation index F_{ST} measured in the overall genome based on microsatellites or with SNPs from candidate regions indicates strong differentiation between breeds. Nine out of 23 regions had $F_{ST} \geq 0.20$ calculated from haplotypes or individual SNPs. The levels of admixture were significantly different from background admixture, as revealed by microsatellite data, for six out of the nine regions. Five out of the six regions showed an excess of Baoule ancestry. Information about best levels of breed composition would be useful for future breeding ctivities, aiming at trypanotolerant animals with higher productive capacity.

Keywords: trypanosome, tolerance, Baoule, Zebu, cross, composite

OPEN ACCESS

Edited by:
Paolo Ajmone Marsan, Università Cattolica del S. Cuore, Italy

Reviewed by:
Pablo Orozco-terWengel, Cardiff University, UK
Riccardo Negrini, Università Cattolica, Italy

****Correspondence:***
Johann Sölkner, Division of Livestock Sciences, Department of Sustainable Agricultural Systems, BOKU- University of Natural Resources and Life Sciences Vienna, Gregor Mendel Strasse 33, 1180 Vienna, Austria
johann.soelkner@boku.ac.at

† These authors have contributed equally to this work.

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: *15 November 2014*
Accepted: *22 March 2015*
Published: *21 April 2015*

Citation:
Smetko A, Soudre A, Silbermayr K, Müller S, Brem G, Hanotte O, Boettcher PJ, Stella A, Mészáros G, Wurzinger M, Curik I, Müller M, Burgstaller J and Sölkner J (2015) Trypanosomosis: potential driver of selection in African cattle. Front. Genet. 6:137. doi: 10.3389/fgene.2015.00137

Introduction

African animal trypanosomosis (AAT) is a severe disease caused by three species of *Trypanosoma* parasites: *T. congolence, T. vivax, and T. brucei*. Trypanosomosis is responsible for the deaths

of millions of livestock each year and a reduction in the productivity of many more. With no vaccine available, and with heavy expenditure on trypanocidal and vector control, trypanosomosis is estimated to cost over 4 billion US dollars each year in direct costs and lost production (Hanotte et al., 2003).

West African cattle have the ability to control parasitemia and anemia related to trypanosomosis and a greater ability to grow and produce in tsetse infested areas (Murray et al., 1984). This is thought to result from an adaptation process of indigenous cattle breeds. Trypanotolerant breeds represent a small proportion (6%) of the cattle population of Africa and 17% of the cattle in the tsetse challenged areas (Agyemang, 2005). The option of using those breeds in breeding systems thus reduces or eliminates the use of chemicals to control the trypanosomosis vector and other parasites and contributes positively to a balanced ecosystem health. The complex trypanotolerant trait that is present in indigenous West African taurine cattle is not present in the introgressed *Bos indicus* cattle and is dependent on admixture proportion in hybrid breeds (Freeman et al., 2004).

Humpless taurine populations (*Bos taurus*) are original indigenous cattle of Africa which entered into the African continent before Zebu cattle, around 8000 years ago for longhorn and around 4750–4500 years ago for shorthorn, while the humped zebu populations (*Bos indicus*) were brought only later into the African continent, *via* the Horn of Africa (Loftus et al., 1994; Bradley et al., 1996; MacHugh et al., 1997; Hanotte et al., 2002; Epstein, 1971). It is believed that taurine cattle penetrated West African forests about 4000 years ago (MacDonald and MacDonald, 2002; Freeman et al., 2004) while zebu populations arrived 1300–1000 years ago (Epstein, 1971; MacHugh et al., 1997; MacDonald and MacDonald, 2002).

There is evidence indicating that host genetic factors play a significant role in determining an individual's susceptibility/resistance status to trypanosoma infection (Murray et al., 1984; Hanotte et al., 2003; Courtin et al., 2006, 2007, 2008). Trypanotolerance of cattle can be associated with genomic regioms known as trypanotolerance candidate regions. Observations from Naessens et al. (2002) confirm that this trypanosoma tolerance encompasses at least two mechanisms: one that improves the control of parasitemia and another that limits anemia. The physiological and genetic mechanisms underlying trypanotolerance are being extensively investigated. Hanotte et al. (2003) performed experimental crossing of trypanotolerant N'Dama (*Bos taurus*) and trypanosusceptible improved Kenya Boran (*Bos indicus*) cattle, and mapped QTLs associated to trypanotolerance on 18 autosomes. Results suggest that selection for trypanotolerance within F2 cross between N'Dama and Boran cattle could produce a synthetic breed with higher trypanotolerance levels than currently exist in the parental breeds. Noyes et al. (2010) performed a genetic expression analysis to identify candidate genes in pathways responding to *T. congolense* infection.

Evidence for selective sweeps was observed at TICAM1 and ARHGAP15 loci in African taurine cattle, leading the authors to propose these genes as strong candidates to explain the QTL. Candidate QTL genes were identified in other QTL by their expression profile and the pathways in which they participate. Dayo et al. (2009) tested heterozygosity and variances in microsatellite allelic size among trypanotolerant and trypanosusceptible breeds which led to two significantly less variable microsatellite markers. One of these two outlier loci is located within the confidence interval of a previously described QTL underlying a trypanotolerance-related trait (Hanotte et al., 2003). Stella et al. (2010) analyzed selection signatures by contrasting 32,689 SNP genotypes of trypanotolerant African taurine N'Dama and Sheko cattle with those of all other breeds included in the Bovine HapMap study (Bovine HapMap Consortium, 2009). The overlap of candidate regions found in different studies is comparatively small. West African cattle have the ability to control parasitemia and anemia related to trypanosomosis and a greater ability to grow and produce in tsetse infested areas (Murray et al., 1984). This is thought to result from an adaptation process of indigenous cattle breeds.

Among the indigenous cattle in Burkina Faso, the Baoule, a taurine breed native to the tsetse-challenged southern part of the country, is known for its ability to cope with trypanosome infections. Pure Zebu (Bos indicus) is much more susceptible to the disease, but still preferred by farmers because of body size and suitability as draft animal. With the intention of having both big and tolerant animals, many farmers use composites, continuously mating Zebu, Baoule and their crosses. The preference for larger animals means that Zebu ancestry is predominant among the admixed animals. In genomic regions responsible for trypanotolerance however, higher levels of Baoule ancestry are expected. In a paper on approaches to detect signatures of selection from genome wide scans, Oleksyk et al. (2010) describe a way of detecting significant differences of local admixture levels in crossbred/admixed individuals compared to the average admixture across their genomes. This method can be applied to identify genome signatures of historic selective pressures on genes and gene regions.

Aim of this study was to compare levels of zebuine and taurine admixture in candidate regions for trypanotolerance with the "background" admixture levels, to identify differences in allelic frequencies of trypanotolerant and non-tolerant breeds, and to assess individual differences in admixture for particular animals. Regions potentially responsible for trypanotolerance were identified based on composite log-likelihoods of the differences in allelic frequencies of trypanotolerant and non-tolerant breeds, using Bovine HapMap data (Bovine HapMap Consortium, 2009). Individual admixture levels in these regions versus admixture levels of the background genome of composite Baoule x Zebu animals were compared.

Materials and Methods

Study Design and Animals

Blood was taken from 214 animals in total out of which 90 were Baoule from South West (SW) Burkina Faso, 90 were Zebu from the North ($n = 54$) and SW ($n = 36$) regions and 34 were Baoule-Zebu composites from SW. The North of Burkina Faso is part of the Sahelian region with no threat of trypanosomosis, while SW is a Sudanese region that is heavily tse-tse infested.

Designation of animals to breed was based on information by owners of the animals. Animals were from 23 different locations

in Burkina Faso. FTA cards were used for collection and storage of blood for all animals.

Discovery of Regions for Selective SNP Genotyping

Only a small number of SNPs could be selectively genotyped in this project. For the choice of these SNPs, the selection signature approach and sampling of animals by Stella et al. (2010) was employed. Data were from the International Bovine HapMap study (Bovine HapMap Consortium, 2009), including the trypanotolerant African taurine breeds N'Dama and Sheko. Baoule is very closely related to N'Dama (Decker et al., 2014). The 32,689 HapMap SNPs as well as 54,001 Illumina 50k bovine BeadChip SNPs were available for analysis, extending the study of Stella et al. (2010). These two sources of data were merged and after quality control, applying a minor allele frequency threshold of 0.05, a minimum call rate of 0.95 per SNP and removal of duplicate SNP, the final data set comprised 71,235 SNP.

To identify putative selection signatures, allelic frequencies of the N'Dama ($N = 22$) and Sheko ($N = 19$), either pooled or separate, were compared to the allelic frequencies of the entire population ($N = 497$) in the study and nominal *P*-values were calculated for the differences in frequencies at each SNP. The nomimal *P*-values were then used to calculate composite log-likelihoods (CLL) for sliding windows of 9 SNP across the genome. To determine statistical significance, permutation testing was employed by comparing the CLL to the distribution of 50,000 permutations of CLL obtained with random samples of animals (i.e., across all HapMap breeds).

The signals typically pointed to narrow regions (0.2–0.4 Mb), with an average of 254,841 bp. A total of 158 SNPs from 23 regions with strong signals (genome-wide $P < 0.01$ in each breed) were chosen. Within each region, 4–10 roughly equally spaced SNPs were selected from the Illumina data base for genotyping. The rationale for this approach was that signatures of selection are likely linked to trypanosome tolerance in the African taurine breeds. Also, the signals were narrow compared to the results of QTL analyses available at the time, allowing targeted SNP selection. Furthermore, the signatures targeted were observed in both the N'Dama and Sheko, perhaps suggesting that they arose in a past ancestral population of all trypanotolerant breeds and were likely to be present in the Baoule as well.

Choice of Microsatellites

To reflect the admixture levels in the background genome of the animals in this study, a total of 25 autosomal microsatellites were chosen, giving a preference to FAO recommended markers (FAO, 2011), without considering information about trypanosome candidate regions. For the autosomal chromosomes, a total of 31 microsatellite primers have been chosen for the amplification of the genomic DNA. 15 primers were donated by the International Livestock Research Institute, Nairobi, Kenya. PCR conditions were optimized and all the 31 microsatellites tested for polymorphism. A final panel of 25 microsatellites was selected for genotyping of the cattle populations. 22 microsatellites out of them (BM1818, BM1824, BM2113, CSSM066, ETH3, ETH10, ETH185, ETH225, HAUT24, HAUT27, HEL1, HEL5, HEL9, HEL13, ILSTSS005, ILSTS006, INRA023, INRA032, TGLA53, TGLA122, TGLA126, and TGLA227) were from a list recommended by the Food and Agriculture Organisation (FAO) and the International Society for Animal Genetics (ISAG) for use in cattle diversity studies. The others, namely AGLA293, ILSST033, and MGTG4B, were out of both the FAO and the ISAG list. The microsatellites were selected combining information from both National Centre for Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov/) database. The selected microsatellites covered 22 autosomal chromosomes.

Genotyping of Animals

Genomic DNA was isolated from white blood cells according to a modified protocol of Whatman (Whatman FTA Protocol BD09). Genotyping of the 25 microsatellites was performed on a MegaBACETM 500 genotyping device. The PCR reaction mixture with the final volume of 22 µl included 10 ng template genomic DNA was used in autosomes amplification. 8.05 µl of double distilled water, 3.20 µl of 10 × Buffer B (Mg2+ free containing 0.8 M Tris-HCl, 0.2 M (NH4)2SO4, 0.2% w/v Tween-20), 2 µl of 2 mM dNTP-Mix, 1.60 µl of 25 mM MgCl2, 0.5 µl of each forward and reverse primers and 0.15 µl of 5 U/µl FIREoL® DNA polymerase. One primer in each pair was labeled FAM or TET. The 155 selected SNPs were were multiplexed and genotyped on the Sequenom MassARRAY system. The choice of SNPs within the regions of interest was guided by a bioinformatic protocol optimizing the multiplexing strategy. We tried to space the SNPs equally across the 200–400 Kb regions of interest. The total number of SNPs targeted for genotyping was 150 with a minimum of 5 per region. The mastermix for multiplying SNPs comprised 0.50 µl of each forward and reverse primers and 0.20 µl of 5 U/µl HotStar Taq DNA polymerase plus 3 µl (3 µg) of Salmon sperm to the all mastermix. Total volume of tubes was 4 µl. A digestion was made after the PCR with shrimp alkaline phosphate (SAP). The SAP cleaves a phosphate from the unincorporated dNTPs, converting them to dNDPs and rendering them unavailable to future reaction. The SAP mix has been made from 1.5 µl of water (HPLC grade), 0.17 µl of 10 × SAP buffer, 0.30 µl of 1.7 U/µl SAP enzyme. 2 µl of SAP mix was added to the normal PCR product for digestion. The digestion was followed by ani PLEX PCR. The iPLEX mix was made of 0.619 µl of water, 0.20 µl of 10 × iPLEX buffer, 0.2 µl od iPLEX Termination mix, 0.041 µl of iPLEX Enzyme and 0.940 µl of the extent primer. 2 µl of the iPLEX mix was added to digested product. The following cycling program was run for amplification: 5 min initial denaturation at 95°C followed by 35 cycles of denaturation at 95°C for 1 min, annealing at 55°C for 1:30 min, extension at 65°C for 3 min and final extension step of 65°C for 5 min using Applied Biosystems 96-Well GeneAmp® PCR System 9700 thermal cycler.

The normal PCR of SNP study was run according the following protocol: 2 min initial denaturation at 95°C followed by 45 cycles of denaturation at 95°C for 0:30 min, annealing at 56°C for 0:30 min, extension at 72°C for 1 min and a final extension step of 72°C for 5 min and 4°C for 5 min. The digestion was run on 45 min: 40 min at 37°C and 5 min at 85°C. While the iPLEX PCR run as followed: 030 min of initial denaturation

at 94°C, denaturation again at 94°C for 0:05 min, annealing at 52°C for 0:05 min, extension at 80°C for 0:005 min. Annealing up to extension (80°C) was repeated 5 times, from the second denaturation to the extension 40 repeats as well. A final extension step of 72°C was run for 3 min ended by 4°C forever. The normal PCR, the SAP digestion and the iPLEX PCR were performed on using Applied Biosystems 384-Well GeneAmp® PCR System 9700 thermal cycler.

SNPs were positioned according to Btau 4.0. Monomorphic SNPs and SNPs with more than 10% of missing values were excluded, data analysis was performed with the remaining 135 SNPs. Average linkage disequilibrium levels, calculated as R-squared values of SNPs within candidate regions were 0.091, with 5% and 95% quantiles of 0.00003 and 0.478.

Data Analysis

Processing of raw data to formats usable in PLINK was done with SAS (SAS Institute Inc, 2009). Ancestry inferences were performed using STRUCTURE (Pritchard et al., 2000, 2010; Hubisz et al., 2009). STRUCTURE uses a model-based clustering algorithm to infer population structure using genotype data. The software clusters data according to allele frequencies into K populations. As there was linkage disequilibrium in our SNP data, we used version 2.3.4. We employed the admixture model using a burn-in period of 10,000 repeats followed by 10,000 Markov Chain Monte Carlo (MCMC) repeats and considering SNP frequencies correlated. Convergence of the MCMC was investigated with a several STRUCTURE runs on the same datasets. STRUCTURE analyses were all supervised, with added information on pure breed or a cross identity. The assumption of a two breed cross was confirmed with Admixture software (Alexander et al., 2009) with K from 2 to 7, the lowest cross validation error was at $K = 2$; $cv = 0.48$. PLINK (Purcell et al., 2007) was used to recode alleles for analysis. AlphaPhase (Hickey et al., 2011) was used for haplotype imputation of SNPs from candidate regions. To evaluate population differentiation, proc ALLELE of SAS/GENETICS 9.2 was used to calculate the fixation index (F_{ST}) for every microsatellite, SNP and haplotype derived from candidate regions. This calculation was based on variance in allele frequencies (Weir and Cockerham, 1984; Weir and Hill, 2002).

Results

Crosses/composites of trypanotolerant and trypanosusceptible cattle were the focus of this analysis. STRUCTURE results indicate that the information acquired from farmers about pure Baoule and Zebu breed types is reasonably accurate with 0.87/0.89 and 0.07/0.06 Baoule ancestry proportions for these

TABLE 1 | Proportions of Baoule admixture calculated with STRUCTURE for Baoule and crosses from Southwest and zebu from Southwest and North.

CHR	PM	M	*FST*	HF_{ST}	Baoule (%)	Zebu (%)	Comp (%)
7	5.77–5.98	6	0.18	0.06	0.93	0.39	0.55
7	21.10–21.30	6	0.19	0.15	0.71	0.28	0.48
7	59.30–59.60	7	0.03	0.01	0.49	0.49	0.57
8	55.20–55.60	6	0.16	0.10	0.79	0.21	0.42
16	23.10–23.40	4	0.27	0.26	0.78	0.24	0.46
17	10.60–10.80	6	0.19	0.41	0.89	0.23	0.5
18	12.40–12.60	7	0.13	0.13	0.64	0.34	0.52
18	17.10–17.30	6	0.23	0.18	0.63	0.08	0.23
19	19.70–19.90	6	0.01	0.02	0.47	0.56	0.57
19	27.80–28.00	7	0.03	0.01	0.59	0.40	0.57
20	19.80–20.00	5	0.28	0.17	0.78	0.27	0.35
20	21.09–22.20	4	0.13	0.21	0.68	0.34	0.41
21	11.60–11.80	8	0.17	0.18	0.76	0.22	0.38
21	20.40–20.60	6	0.23	0.18	0.86	0.26	0.54
21	37.90–38.20	7	0.10	0.11	0.53	0.55	0.45
21	42.80–43.00	7	0.12	0.06	0.68	0.26	0.43
21	56.20–56.50	8	0.00	0.01	0.49	0.51	0.52
21	57.40–57.80	7	0.11	0.08	0.73	0.26	0.47
22	20.60–20.90	5	0.20	0.51	0.82	0.17	0.41
22	51.20–51.40	2	0.23	0.33	0.69	0.33	0.55
23	28.80–29.00	4	0.06	0.17	0.51	0.5	0.5
25	15.40–15.70	7	0.11	0.06	0.66	0.27	0.42
26	23.20–23.40	4	0.34	0.15	0.84	0.22	0.41
All SNPs		135	0.14		0.89	0.06	0.37
Microsatellites		25	0.09		0.87	0.07	0.31

CHR, Chromosome where candidate region is positioned; PM, Position for a corresponding region in Megabases; M, number of markers used in analysis; F_{ST}, fixation index calculated based on average of SNPs; HF_{ST}, fixation index calculated for haplotypes; Baoule%, admixture proportion of Baoule in Baoule; Zebu%, admixture proportion of Baoule in Zebu animals; Comp%, admixture proportion of Baoule in composites.

two breeds based on Microsatellite/SNP markers (see **Table 1**). Average Baoule admixture in background genomes (as assessed by microsatellite markers) of Baoule-Zebu composites was 0.31, which was somewhat, but not significantly ($p = 0.15$ based on a t-test), smaller than the average Baoule admixture in the AAT candidate regions (assessed by SNP markers), 0.37.

Admixture proportions were also determined for each genomic region potentially implicated in AAT tolerance (**Table 1**). Regions with highest Baoule ancestry proportions in the composites were found on chromosome (CHR) 7 [5.77–5.98, 59.30–59.60 Mega bases (Mb)] and on CHR 22 (51.20–51.40 Mb). In some candidate regions, STRUCTURE was not able to cluster individuals, to separate crosses from pure breeds, based on low F_{ST} values (**Table 1**). In order to avoid regions with low F_{ST} values we further considered only the 9 regions where the average F_{ST} of SNPs by SNP or F_{ST} of haplotypes (both provided in **Table 1**) was more than 0.20 (both provided in **Table 1**). Proportions of individuals with admixture estimates of over 0.60 Baoule ancestry in composites for the nine candidate regions were (in descending order): CHR 22 with 62 and 50% (51.20–51.40 and 20.60–20.90 Mb), CHR 21 (11.60–11.80 Mb) 32%, CHR 26 (23.20–23.40 Mb) 32% CHR 16 (23.10–23.40 Mb) 29%, CHR 20 (19.80–20.00 Mb) 24% and CHR 18 (12.40–17.30 Mb) 18%. Average number of haplotypes per region was 12.59 for Baoule, 11.91 for Zebu and 10.64 for composites. Numbers of haplotypes per breed and genomic region are provided in **Table 2**. Haplotypes with more than 10% frequency in at least one of the breeds are given in **Table 3** for the nine candidate regions with $F_{ST} \geq 0.20$. Reconstructed haplotypes showed differentiation of the Zebu and Baoule individuals. For the region on CHR 21 (20.40–20.60 Mb), the total frequency of two haplotypes (out of 22) was 83.34% for Baoule, in composites the frequency of these haplotypes was 51.52% (they had 13 haplotypes in total) whereas their frequency (16 haplotypes) in Zebu animals was 22.03% (**Table 2**). Regions on chromosomes 22 (51.20–51.40 and 20.60–20.90 Mb) and 16 (23.10–23.40 Mb) had 45.59% (out of 98.53%), 32.35% (out of 98.53%) and 34.33% (out of 89.55%) of haplotypes predominant in Baoule. Pairwise chi-squared tests indicated significant differences in haplotype composition for all pairs of breeds (Baoule–Zebu, Baoule–composites, Zebu–composites), in all regions, except for regions on chromosome 20 (19.80–20.00 and 21.90–22.20 Mb), where haplotypes of composites were not significantly different from Zebu ($P = 0.0896, P = 0.1539$). Differences were insignificant after applying Bonferroni correction in the region of Chr 18 (12.40–12.60 Mb) for Zebu and composites and region from CHR 22 (51.20–51.40 Mb) of Baoule and composites. F_{ST} calculated for haplotypes was greatest for CHR 22 (20.60–20.90 Mb), with a value of 0.51.

Genes found in candidate genomic regions studied as recovered from Ensembl (www.ensembl.org) are provided in Supplementary Table 1. Their potential relevance to trypanotolerance based on information from other studies is discussed below.

Discussion

Baoule admixture across the genome, based on a sample of 25 microsatellite markers, of the Baoule-Zebu composites was 0.31, compared to the average Baoule admixture based on SNPs in AAT candidate regions (0.37), see **Table 1**. This difference was not significant, though ($P = 0.15$). Admixture was measured with two distinct types of markers, justification for the process can be found in a study of Schopen et al. (2008), showing that the information content of one microsatellite corresponds to an equivalent to that of about three SNPs in cattle. Similar results were provided by Gärke et al. (2012) when analyzing population differentiation of chicken breeds. The average number of haplotypes per region was 12.59 for Baoule, 11.91 for Zebu and 10.64 for composites. We found slightly more haplotypes in our candidate regions for Baoule (277) compared to Zebu (262), the lower number of haplotypes in composites is most likely due to the smaller sample size. The results are in contrast to Murray et al. (1984) who found higher diversity in *B. indicus* compared to *B. taurus*. It is known that the West African *B. taurus* populations contain a degree of *B. indicus* admixture (Alvarez et al., 2014), but the proportion is small in Baoule (Hanotte et al., 2003; Soudré et al., 2013).

Due to the very low F_{ST} in some candidate regions, STRUCTURE was not able to separate pure breeds in those regions. Overall F_{ST} calculated with SNPs from candidate regions ($F_{ST} = 0.14$) matches that from other studies, see Dayo et al. (2009), while F_{ST} calculated with microsatellites ($F_{ST} = 0.09$) was lower. When looking at F_{ST} for single SNPs, the highest value was 0.70

TABLE 2 | Number of reconstructed haplotypes from SNPs in candidate regions.

CHR	PM	N	B	Z	Comp
7	5.77-5.98	2	4	4	4
7	21.10-21.30	3	7	5	6
7	59.30-59.60	4	14	15	12
8	55.20-55.60	6	22	17	18
16	23.10-23.40	3	8	7	7
17	10.60-10.80	5	8	6	6
18	12.40-12.60	4	12	11	11
18	17.10-17.30	5	15	15	11
19	19.70-19.90	6	12	11	9
19	27.80-28.00	7	15	21	12
20	19.80-20.00	4	10	9	10
20	21.09-22.20	4	6	8	6
21	11.60-11.80	6	22	16	13
21	20.40-20.60	6	7	10	10
21	37.90-38.20	5	10	17	12
21	42.80-43.00	5	13	13	14
21	56.20-56.50	7	20	20	15
21	57.40-57.80	7	29	19	19
22	20.60-20.90	4	4	4	3
22	51.20-51.40	2	2	3	3
25	28.80-29.00	6	25	19	21
26	15.40-15.70	4	12	12	12
Total		105	277	262	234

CHR, Chromosome where candidate region is positioned; PM, Position for a corresponding region in Megabases; N, number of SNPs in candidate region used to reconstruct haplotype; B, number of haplotypes found in pure Baoule animals; Z, number of haplotypes found in pure Zebu animals; Comp, number of haplotypes found in composites.

TABLE 3 | Most frequent haplotypes.

		Baoule		Zebu		Composite	
CHR	**H**	**N**	**% f**	**N**	**% f**	**N**	**% f**
16 (23.10–23.40 Mb)							
	000	68	37.78	1	0.57	10	14.93
	001	16	8.89	2	1.14	3	4.48
	011	59	32.78	11	6.25	13	19.40
	110	15	8.33	100	56.82	25	37.31
	111	6	3.33	41	23.30	12	17.91
	00110	7	3.89	108	60.00	19	27.94
17 (10.60–10.80 Mb)							
	11110	150	83.33	59	32.78	38	55.88
	01001	10	6.45	28	16.00	8	12.70
18 (17.10-17.30 Mb)							
	01101	7	4.52	80	45.71	18	28.57
	11001	52	33.55	15	8.57	13	20.63
	11101	6	3.87	9	5.14	7	11.11
	11111	52	33.55	3	1.71	5	7.94
	0001	10	5.92	13	7.34	2	3.13
20 (19.80–20.00 Mb)							
	0011	93	55.03	19	10.73	13	20.31
	0111	13	7.69	1	0.56	1	1.56
	1001	3	1.78	32	18.08	12	18.75
	1011	24	14.20	27	15.25	3	4.69
	1101	5	2.96	54	30.51	18	28.13
	1111	6	3.55	22	12.43	10	15.63
	0111	27	15.00	2	1.11	1	1.47
20 (21.09–22.20 Mb)							
	1011	10	5.56	74	41.11	20	29.41
	1111	137	76.11	72	40.00	38	55.88
	010011	6	3.41	52	29.38	14	21.21
21 (20.40–20.60 Mb)							
	011111	75	42.61	9	5.08	11	16.67
	110011	8	4.55	71	40.11	7	10.61
	111111	81	46.02	30	16.95	23	34.85
	0101	21	11.67	1	0.56	1	1.47
22 (20.60–20.90 Mb)							
	1101	118	65.56	17	9.44	22	32.35
	1111	40	22.22	159	88.33	44	64.71
	01	61	33.89	136	75.56	36	52.94
22 (51.20–51.40 Mb)							
	11	119	66.11	30	16.67	31	45.59
	0000	10	5.92	60	40.00	18	27.69
26 (23.80–29.00 Mb)							
	0001	3	1.78	17	11.33	6	9.23
	0010	2	1.18	25	16.67	6	9.23
	1101	64	37.87	4	2.67	12	18.46
	1111	42	24.85	13	8.67	8	12.31

CHR, Chromosome and region in Mb of candidate regions H- Reconstructed haplotype; N, Number of haplotypes; %f, frequency of the haplotypes; In table are shown haplotypes with >10% frequency in at least one of the studied populations.

for a SNP on CHR 26 while very low F_{ST} values were found in almost every region (results not shown). F_{ST} values below 0.05 indicate low differentiation, 0.20 moderate to strong and values above 0.65 indicate extreme differentiation (Barreiro et al., 2008). Therefore we concentrated on candidate regions which showed $F_{ST} > 0.20$. Using information from sex linked markers Soudre (2011) found a relative age of admixture of 69 ± 43 years from 2007 data for the crosses analyzed in this study. This is consistent with the findings of Grace (2006) who described the history of introduction of Zebu and crosses for draft power in the Kénédougou region in Burkina Faso. The time range and continuous usage of pure Zebu and Baule animals in the crossbred population are responsible for the a large spread of admixture levels in crosses. A region on CHR 22 (51.20–51.40 Mb) showed greatest admixture deviation in favor of Baoule when compared to the overall genome admixture. CHR 22 has been previously identified to have regions responsible for trypanotolerance (Hanotte et al., 2003; Gautier et al., 2009). Gautier et al. (2009) observed that a region of CHR 22 (region between 43.79–53.04 Mb), overlapping with our region contained a high proportion of SNPs under balancing selection. This result might be related to the maintenance of several haplotypes containing variants under positive selection within different populations. Alternatively, the fixation of the same variant in some populations could also lead to such a trend due to the low level of linkage disequilibrium across populations (Gautier et al., 2007, 2009). Genes involved in trypanotolerance on CHR 22 (between 51.20 and 51.40 Mb) such as MON1A, MST1R, UBA7, FAM212, CAMKV, TRAIP, CDHR4, IP6K1, RNF123, APEH, and MST1 (Supplementary Table 1) were described by O'Gorman et al. (2009). These authors found that trypanotolerant N'Dama cattle displayed a rapid and distinct transcriptional response to infection, with a 10-fold higher number of genes differentially expressed at day 14 post-infection compared to susceptible Boran cattle. Their analyses identified coordinated temporal gene expression changes for both breeds in response to trypanosome infection. Three other protein coding regions found on this part of CHR 22 (ENSBTAG00000040083, CDHR4 and AMIGO) did not show differential expression in their analysis.

One region, CHR 18 (17.10–17.30 Mb) showed significantly lower Baoule admixture than average admixture measured by microsatellites and SNPs. In a mapping study with experimental crosses of trypanotolerant N'Dama and trypanosusceptible Boran cattle, Hanotte et al. (2003) had found that in some instances the trypanosusceptible breed carried trypanotolerant alleles. This region contained one coding gene not previously described in trypanotolerance studies. Noyes et al. (2010) described a QTL on CHR 16 as region responsible in tolerant breeds to cope with anemia. On the region of CHR 16, MIA3, BROX, and AUH genes were found. Three protein coding sequences are present in the candidate region of CHR 21 (20.40–20.60 Mb): RLBP1, FANCI, and POLG, with RLBP1 and POLG showing differential expression. All four genes in the candidate region of chromosome 26 (HPS6, LDB1, NOLC1, and GBF1) were differentially expressed in the study of O'Gorman et al. (2009).

Conclusions

In this study admixture in genomic regions potentially related to trypanotolerance was compared with admixture in the background genome. A non-significant trend of higher proportions of Baoule admixture in the candidate regions was found and a majority of regions (5 of 6) with admixture levels significantly different from background admixture indicated high levels of Baoule ancestry.

In this study, the discovery of trypanotolerance candidate regions was performed via a selection signature approach based on differences of allelic frequencies of trypanotolerant African taurine breeds versus other breeds around the world, using data from the Bovine HapMap consortium. Targeted SNP genotyping in candidate regions was the method of choice. Given the reduction of cost for high density SNP chip genotyping, part of the samples used in this study are now being genotyped with the commerical chip of Illumina Inc.,covering almost 800,000 SNPs. A large number of markers will also allow estimation of individual age of admixture and therefore the number of generations of natural selection acting on the composites. Targeted resequencing approaches of interesting candidate regions that can identify both common and exceedingly rare causal variants could potentially give more insight into trypanotolerance mechanisms.

Silbermayr et al. (2013) developed a novel qPCR assay for indication of infection status of animals with the three trypanosome species involved in AAT (*T. vivax, T. congolense,* and *T. brucei*) from blood samples of most of the animals involved in this study. Zebus were twice as often infected (21.74%) compared to Baoule (9.70%) and composites (9.57%). Phenotypic measures oftrypanosomosis by routine checking of infection status will help to identify best composites (Soudré et al., 2013). Information about best levels of admixture in composites is a premise of more effective and sustainable use of trypanotolerant types of cattle.

Author Contributions

JS conceived the study, with the support of OH. ASo collected samples and background information, as suggested by MW, GB, and MM provided genotyping facilities and support. The study performed to find trypanosoma tolerance region was performed by PB and ASt. Genotyping of bovine microsatellites was performed by ASo and SM, SNPs were genotyped by JB and ASo while KS genotyped parasites. Admixture and F_{ST} analysis was performed by ASm who also drafted the manuscript. Results were interpreted by all authors, PB, JS, GM, and ASm provided the biggest contributions in manuscript revision. All authors read and approved the final manuscript. ASm – Anamarija Smetko; ASo – Albert Soudre; ASt – Alessandra Stella. The views expressed in this publication are those of PB and do not necessarily reflect the views or policies of FAO.

Acknowledgments

We gratefully acknowledge the support of Dr. Delia Grace (ILRI) and Dr. Issa Sidibe (CIRDES, Burkina Faso) before and during the sampling work. The authors are grateful to the Austrian Exchange Service for the grant and for funding the field work, the International Livestock Research Institute, the University of Natural Resources and Life Sciences in Austria, and the Polytechnic University of Bobo-Dioulasso in Burkina Faso for co-funding the field work. We also thank the regional directors and collaborators of the study areas and the farmers for their active collaboration during the surveys.

Supplementary Material

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fgene.2015.00137/abstract

Animal_breed_location_GPS_tryps.xlsx: Animal ID, Animal breed, Region, Village, Latitude, Longitude, Trypanosoma species present in blood sample of this animal.

SNP.ped: SNP data of 214 animals included in the study, 0 represents missing allele. First column is family ID column, followed by animal ID, sire ID, dam ID, sex and phenotype. Phenotype is denoted with -9 as missing and after this columns alleles for SNPs follow.

SNP.map: column with chromosome number, SNP name, Morgan position and bp positions of SNPs included in SNP.ped file.

MIC.ped: Row names are microsatellite names, first column is animal ID, followed by population code, flag code and location. After those columns microsatellites follow.

FAO-marker.xlsx: First column is name of microsatellite, second is chromosome, Genetic Map (MARC) and Sequence Map (STS).

References

Agyemang, K. (2005). "Trypanotolerant livestock in the context of trypanosomiasis intervention strategies," in *PAAT Technical and Scientific Series 7* (Rome: Food and Agriculture Organisation of the United Nations).

Alexander, D. H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19, 1655–1664. doi: 10.1101/gr.094052.109

Alvarez, I., Traore, A., Fernandez, I., Cuervo, M., Lecomte, T., Soudre, A., et al. (2014). Assesing introgression of Sahelian zebu genes into native Bos taurus breeds in Burkina Faso. *Mol. Biol. Rep.* 41, 3745–3754. doi: 10.1007/s11033-014-3239-x

Barreiro, L. B., Laval, G., Quach, H., Patin, E., and Quintana-Murci, L. (2008). Natural selection has driven population differentiation in modern humans. *Nat. Genet.* 40, 340–345. doi: 10.1038/ng.78

Bovine HapMap Consortium. (2009). Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Science* 324, 528–532. doi: 10.1126/science.1167936

Bradley, D. G., MacHugh, D. E., Cunningham, P., and Loftus, R. T. (1996). Mitochondrial diversity and the origins of African and European cattle. *Proc. Natl. Acad. Sci. U.S.A.* 93, 5131–5135. doi: 10.1073/pnas.93.10.5131

Courtin, D., Argiro, L., Jamonneau, V., N'dri, L., N'guessan, P., Abel, L., et al. (2006). Interest of tumor necrosis factor-alpha-308G/A and interleukin-10-592C/A polymorphisms in human African trypanosomosis. *IGE* 6, 123–129. doi: 10.1016/j.meegid.2005.03.002

Courtin, D., Berthier, D., Thevenon, S., Dayo, G. K., Garcia, A., and Bucheton, B. (2008). Host genetics in African trypanosomiasis. *IGE* 8, 229–236. doi: 10.1016/j.meegid.2008.02.007

Courtin, D., Milet, J., Jamonneau, V., Yeminanga, C. S., Kumeso, V. K., Bilengue, C. M., et al. (2007). Association between human African trypanosomiasis and the IL6 gene in a Congolence population. *IGE* 7, 60–68. doi: 10.1016/j.meegid.2006.04.001

Dayo, G. K., Thevenon, S., Berthier, D., Moazami-Goudrazi, K., Denis, C., Cuny, G., et al. (2009). Detection of selection signatures within candidate regions underlying trypanotolerance in outbred cattle populations. *Mol. Ecol.* 18, 1801–1813. doi: 10.1111/j.1365-294X.2009.04141.x

Decker, J. E., McKay, S. D., Rolf, M. M., Kim, J., Alcalá, A. M., Sonstegard, T. S., et al. (2014). Worldwide patterns of ancestry, divergence, and admixture in domesticated cattle. *PLoS Genet.* 10:e1004254. doi: 10.1371/journal.pgen.1004254

Epstein, H. (1971). *The Origin of the Domestic Animals of Africa,* Vol I. New York, NY: Africana Publishing Corporation 1971.

FAO, (2011). *Molecular Genetic Characterization of Animal Genetic Resources.* Rome: FAO Animal Production and Health Guidelines, 9.

Freeman, A. R., Meghen, C. M., Machugh, D. E., Loftus, R. T., Achukwi, M. D., Bado, A., et al. (2004). Admixture and diversity in West African cattle populations. *Mol. Ecol.* 13, 3477–3487. doi: 10.1111/j.1365-294X.2004.02311.x

Gärke, C., Ytournel, F., Bed'hom, B., Gut, I., Lathrop, M., Weigend, S., et al. (2012). Comparison of SNPs and microsatellites for assessing the genetic structure of chicken populations. *Anim. Genet.* 43, 419–428. doi: 10.1111/j.1365-2052.2011.02284.x

Gautier, M., Faraut, T., Moazami- Goudarzi, K., Navratil, V., Foglio, M., Grohs, C., et al. (2007). Genetic and haplotypic structure in 14 European and African cattle breeds. *Genetics* 177, 1059–1070. doi: 10.1534/genetics.107.075804

Gautier, M., Flori, L., Riebler, A., Jaffrezic, F., Laloe, D., Gut, I., et al. (2009). A whole genome Bayesian scan for adaptive genetic divergence in West African cattle. *Bio. Med. Central Genomics* 10, 550. doi: 10.1186/1471-2164-10-550

Grace, D. (2006). *Epidemiology and Control of Cattle Trypanosomiasis in Villages Under Risk of Trypanocide Resistance in West Africa*. Doctoral Thesis, Veterinary university Berlin, Berlin.

Hanotte, O., Bradley, D. G., Ochieng, J. W., Verjee, Y., Hill, E. W., and Rege, J. E. O. (2002). African pastoralism: genetic imprints of origins and migrations. *Science* 296, 336–339. doi: 10.1126/science.1069878

Hanotte, O., Ronin, Y., Agaba, M., Nilsson, P., Gelhaus, A., Horstmann, R., et al. (2003). Mapping of quantitative trait loci controlling trypanotolerance in a cross of tolerant West African N'Dama and susceptible East African Boran cattle. *Proc. Natl. Acad. Sci. U.S.A.* 100, 7443–7448. doi: 10.1073/pnas.1232392100

Hickey, J. M., Kinghorn, B. P., Tier, B., Wilson, J. F., Dunstan, N., and van der Werf, J. H. J. (2011). A combined long-range phasing and long haplotype imputation method to impute phase for SNP genotypes. *GSE* 43, 12. doi: 10.1186/1297-9686-43-12

Hubisz, M., Falush, D., Stephens, M., and Pritchard, J. K. (2009). Inferring weak population structure with the assistance of sample group information. *Mol. Ecol. Resour.* 9, 1322–1332. doi: 10.1111/j.1755-0998.2009.02591.x

Loftus, R. T., MacHugh, D. E., Bradley, D. G., Sharp, P. M., and Cunningham, E. P. (1994). Evidence for two independent domestications of cattle. *Proc. Natl. Acad. Sci. U.S.A.* 91, 119–144. doi: 10.1073/pnas.91.7.2757

MacDonald, K. C., and MacDonald, R. H. (2002). "The origins and development of domesticated animals in Arid West Africa." in *The Origins and Development of African Livestock: Archaeology, Genetics, Linguistics and Ethnography,* eds R. Blench and K. MacDonald (London: UCL Press), 127–162.

MacHugh, D. E., Shriver, M. D., Loftus, R. T., Cunningham, P., and Bradley, D. G. (1997). Microsatellite DNA variation and the evolution, domestication and phylogeography of taurine and zebu cattle (*Bos taurus* and *Bos indicus*). *Genetics* 146, 1071–1086.

Murray, M., Trail, J. C., Davis, C. E., and Black, S. J. (1984). Genetic resistance to African Trypanosomiasis. *J. Infect. Dis.* 149, 311–319. doi: 10.1093/infdis/149.3.311

Naessens, J., Teale, A. J., and Sileghem, M. (2002). Identification of mechanisms of natural resistance to African trypanosomiasis in cattle. *Vet. Immunol. Immunophatol.* 87, 187–194. doi: 10.1016/S0165-2427(02)00070-3

Noyes, H., Brass, A., Obara, I., Anderson, S., Archibald, A. L., Bradely, D. G., et al. (2010). Genetic and expression analysis of cattle identifies candidate genes in pathways responding to Trypanosoma congolense infection. *Proc. Natl. Acad. Sci. U.S.A.* 108, 9304–9309. doi: 10.1073/pnas.1013486108

O'Gorman, G. M., Park, S. D. E., Hill, E. W., Meade, K. G., Coussens, P. M., Agaba, M., et al. (2009). Transcriptional profiling of cattle infected with Trypanosoma congolense higlights gene expression sign underlying trypanotolerance and trypano susceptybility. *BMC Genomics* 10:207. doi: 10.1186/1471-2164-10-207

Oleksyk, T. K., Smith, M. W., and O'Brien, S. J. (2010). Genome wide scans for footprints of natural selection. *Philos. Trans. R. Soc.* 365, 185–205. doi: 10.1098/rstb.2009.0219

Pritchard, J. K., Wen, X., and Falush, D. (2010). *Documentation for Structure Software 2010, Version 2.3*. Available online at: http://pritch.bsd.uchicago.edu/structure_software/release_versions/v2.3.3/structure_doc.pdf

Pritchard, J., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., et al. (2007). PLINK: a toolset for whole-genome association and population-based linkage analysis. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795

SAS Institute Inc, (2009). *SAS/STAT® User's Guide, Version 9.2.* Cary, NC.

Schopen, G. C. B., Bovenhuis, H., Visker, M. H. P. W., and van Arendonk, J. A. M. (2008). Comparison of information content for microsatellites and SNPs in poultry and cattle. *Anim. Genet.* 39, 451–453. doi: 10.1111/j.1365-2052.2008.01736.x

Silbermayr, K., Li, F., Soudré, A., Müller, S., and Sölkner, J. (2013). A novel qPCR assay for the detection of African Animal Trypanosomosis in trypanotolerant and trypanosusceptible cattle breeds. *PLoS Negl. Trop. Dis.* 7:e2345. doi: 10.1371/journal.pntd.0002345

Soudre, A. (2011). *Trypanosomosis, Genetic Diversity and Admixture in Cattle Breeds of Burkina Faso,* Doctoral Thesis, University of Natural Resources and Life Sciences, Vienna.

Soudré, A., Ouédraogo-Koné, S., Wurzinger, M., Müller, S., Hanotte, O., Ouédraogo, A. G., et al. (2013). Trypanosomosis: a priority disease in tsetse-challenged areas of Burkina Faso. *Trop. Anim. Health Prod.* 45, 497–503. doi: 10.1007/s11250-012-0248-4

Stella, A., Hanotte, O., and Boettcher, P. J. (2010). "Selection signatures in trypanotolerant African cattle breeds," in *Plant and Animal Genomes XVIII Conference* (San Diego, CA), 549.

Weir, B. S., and Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution* 38, 1358–1370. doi: 10.2307/2408641

Weir, B. S., and Hill, W. G. (2002). Estimating F-statistics. *Ann. Rev. Genet.* 36, 721–750. doi: 10.1146/annurev.genet.36.050802.093940

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Smetko, Soudre, Silbermayr, Müller, Brem, Hanotte, Boettcher, Stella, Mészáros, Wurzinger, Curik, Müller, Burgstaller and Sölkner. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in GENETICS

ORIGINAL RESEARCH ARTICLE
published: 19 December 2014
doi: 10.3389/fgene.2014.00443

Genomic adaptation of admixed dairy cattle in East Africa

Eui-Soo Kim and Max F. Rothschild*

Department of Animal Science, Iowa State University, Ames, IA, USA

Edited by:
Johann Sölkner, University of Natural Resources and Life Sciences Vienna, Austria

Reviewed by:
Jeffrey O'Connell, University of Maryland School of Medicine, USA
Steve Bishop, University of Edinburgh, UK

***Correspondence:**
Max F. Rothschild, Center for Sustainable Rural Livelihoods, Iowa State University, 2255 Kildee Hall, Ames, IA, USA
e-mail: mfrothsc@iastate.edu

Dairy cattle in East Africa imported from the U.S. and Europe have been adapted to new environments. In small local farms, cattle have generally been maintained by crossbreeding that could increase survivability under a severe environment. Eventually, genomic ancestry of a specific breed will be nearly fixed in genomic regions of local breeds or crossbreds when it is advantageous for survival or production in harsh environments. To examine this situation, 25 Friesians and 162 local cattle produced by crossbreeding of dairy breeds in Kenya were sampled and genotyped using 50K SNPs. Using principal component analysis (PCA), the admixed local cattle were found to consist of several imported breeds, including Guernsey, Norwegian Red, and Holstein. To infer the influence of parental breeds on genomic regions, local ancestry mapping was performed based on the similarity of haplotypes. As a consequence, it appears that no genomic region has been under the complete influence of a specific parental breed. Nonetheless, the ancestry of Holstein-Friesians was substantial in most genomic regions (>80%). Furthermore, we examined the frequency of the most common haplotypes from parental breeds that have changed substantially in Kenyan crossbreds during admixture. The frequency of these haplotypes from parental breeds, which were likely to be selected in temperate regions, has deviated considerably from expected frequency in 11 genomic regions. Additionally, extended haplotype homozygosity (EHH) based methods were applied to identify the regions responding to recent selection in crossbreds, called candidate regions, resulting in seven regions that appeared to be affected by Holstein-Friesians. However, some signatures of selection were less dependent on Holsteins-Friesians, suggesting evidence of adaptation in East Africa. The analysis of local ancestry is a useful approach to understand the detailed genomic structure and may reveal regions of the genome required for specialized adaptation when combined with methods for searching for the recent changes of haplotype frequency in an admixed population.

Keywords: admixture mapping, selection signatures, adaptation to new environment, population structure, gene flow, crossbred cattle

INTRODUCTION

Dairy cattle have been successfully improved to optimize their performance under favorable environments. During the recent decades, Western or exotic dairy cattle have been imported rapidly to diverse geographic regions using new technologies such as artificial insemination (AI). However, purebreds have rarely adapted well to new environments in a short period of time and usually not without significant interventions for health. Holstein-Friesians, considered by many to be the most productive dairy breed in temperate regions, have been imported to tropical or subtropical regions in an attempt to improve milk production, but in many cases they have shown poor performance in comparison with that obtained in the temperate climate (Bohmanova et al., 2007).

In East Africa, European cattle have been introduced since the early twentieth century, creating local cattle herds under strong natural selection by disease and environmental factors. According to a previous study, mostly European ancestry remains in a large herd Kenyan cattle population sample but with little native germplasm (Gorbach et al., 2010) and the Kenyan cattle that have been maintained in local smaller farms were mostly an admixed population of imported breeds, including Holstein-Friesian and Guernsey. Genetic variation between breeds for most quantitative traits represents opportunities to combine breeds to improve productivity (Van Vleck et al., 1986). Despite a presumed advantage of crossbreeding in the F1 generation profit from crosses between Holsteins and other dairy breeds in the United States is not as high as that of Holsteins (VanRaden and Sanders, 2003). Nevertheless, mating between cattle breeds with diverse genetic backgrounds that may increase the adaptability of cattle in other geographic regions.

In Africa, crossbreds resulting from a limited gene pool of exotic cattle imported from more productive countries and strong natural selection for a resistance to infectious disease or other environmental stresses possibly may display a footprint of selection. In particular, animals bred over time in local farms are more likely to be adapted to a new environment, allowing us to obtain clues to reveal the genomic region involved in adaptation. To identify genomic regions corresponding to natural selection, new approaches have been suggested using large scale genetic

markers (Nielsen et al., 2007). In Holstein cattle, signatures of artificial selection have been identified by examining the decay of extended haplotype homozygosity (EHH) (Qanbari et al., 2010; Elferink et al., 2012; Glick et al., 2012) or F_{ST} (Barendse et al., 2009; Flori et al., 2009) with other cattle breeds. Common signatures of selection have been reported in Holstein-Friesians and those are expected to be detected in addition to the unique signatures of selection in cattle sampled from East Africa. Moreover, admixture mapping (Winkler et al., 2010) was applied to infer the background of genomic regions considering the admixed population structure in Kenyan cattle. Since animals in local farms have been maintained by crossbreeding with mostly European or U.S. dairy cattle breeds, inferring local ancestry would help to reveal the genetic background contributing to adaptation in East Africa. This approach could be an attractive complementary analysis to signatures of selection for the analysis of an admixed population. Admixture mapping may also help to reveal favorable parental populations in a genomic region as a result of adaptation.

Large variation in production ability of exotic cattle has been reported in non-temperate climate regions (O'Neil et al., 2010). There is an opportunity to improve livestock when the genes involved in desirable phenotypes are identified in a specific environment. In developing countries, sparse recording of livestock performance and limited use of such records have been obstacles to improve the genetic ability of animals. Additionally, economic traits, including resistance to infectious disease, improved fertility, heat tolerance, or more basic information have not been recorded in many regions due to the limited availability of time, and effort and overall cost. Thus, identification of genomic regions and genes adapted to a specific condition may help breeders to understand the genetic characteristics of population and to establish a plan to improve cattle under severe conditions.

MATERIALS AND METHODS

GENOTYPES AND POPULATION

Based on a previous study, we consider the following animals a crossbred group: 20 bulls that were have been heavily used in breeding programs for genetic improvement of the dairy cattle population in Kenya, and 142 cows and offspring sampled from many small farms in Kenya (crossbreds). Phenotypes of individual bulls and cows were not available. There were 25 animals across four generations from a large Kenyan ranch that were known to be closely related to Holsteins and Friesians (Friesians). DNA was extracted from blood samples and genotyped using the Illumina 50K bovine SNP chip (San Diego, CA, USA). Markers with minor allele frequency >0.01 within any breed were included in the analysis. We used Plink (Purcell et al., 2007) to control the quality of data and the position of each SNP on a bovine chromosome (BTA) was decided using genome assembly UMD 3.1. In order to assess the genetic background of Kenyan cattle, SNP genotypes of Holstein ($n = 62$), Guernsey ($n = 21$), and Norwegian Red ($n = 21$) cattle were obtained from the HapMap project (Bovine HapMap Consortium, 2009). Norwegian Red cattle share the many common ancestors with Ayrshire that have been imported to Kenya.

POPULATION ADMIXTURE

Principal component analysis (PCA) and discriminant analysis of principal components (DAPC) were used to examine the admixture in the Kenyan cattle population. DAPC is an approach that optimizes the separation of individuals into predefined groups using a discriminant function of principal component (Jombart et al., 2010). Based on DAPC, membership probability was obtained to present the overall genetic background of an individual. For this analysis, Western breeds, namely Holstein, Guernsey, and Norwegian Red cattle as well as two Kenyan cattle populations were included. Differentiation of populations was also measured by estimation of single marker F_{ST} (Wright, 1951) across the genome. The software package *adegenet* in R was used to analyze population admixture and F_{ST} (Jombart and Ahmed, 2011).

INFERENCE OF LOCAL ANCESTRY

Although ancestries in the Kenyan dairy cattle of admixed individuals appear complicated, the majority of their ancestors were known to originate from a limited number of imported breeds, particularly Holstein-Friesian and Guernsey (Gorbach et al., 2010). To further understand the observed admixture, we estimated local ancestries at each locus in Kenyan crossbred cattle comparing with animals sampled from their parental breeds. The SNP allele frequency of haplotypes obtained from three European/U.S. dairy breeds was applied to infer the ratio of the parental breeds in a specific genomic regions using LAMP-LD (Baran et al., 2012). This method applied hidden Markov models (HMM) to trace the origins of an admixed population based on the haplotype patterns trained using parental breeds. The genome was divided into non-overlapping 30-SNP window to train the HMM of the ancestral population. The LAMP-LD classifies the origin of a haplotype in crossbreds into at most two ancestral groups. Thus, we could obtain the mean and minimum probability of local ancestry. When one haplotype originated from only one parental breed, it was regarded as true local ancestry to calculate minimum probability, while multiple ancestries were allowed to calculate mean probability of local ancestry. For local ancestry mapping, haplotype phase was decided using Beagle package (Browning and Browning, 2007).

MIGRATION OF HAPLOTYPES FROM PARENTAL BREEDS TO KENYAN CROSSBREDS

In addition to local admixture analysis, frequencies of haplotypes from animals in parental breed were examined in crossbred cattle. We suggest this method to test whether the frequency of a haplotype from Holstein, Ayrshire, or Guernsey cattle has changed substantially in Kenyan crossbreds compared to the frequency of the haplotype in a parental breed. The most frequent haplotypes, probably a signature of the influential bulls, were compared between animals in a breed and crossbreds to assess the migration of haplotypes under selection using a 30-SNP window and starting with a new window every 15-SNPs. Mean and standard deviation of the length of 30-SNP haplotype was 1.48 and 0.38 (Mb), respectively. Next, the difference in the frequency of a common haplotype which originated from a parental breed was calculated

using the equation, $freq(D) = p(B) \times freq(K|B) - freq(B)$, where $freq(B)$ is the frequency of the most common haplotype of a breed, $p(B)$ is the ratio of an ancestry in the region encompassing the haplotype, and $freq(K|B)$ is the frequency of haplotype that originates from a parental breed in Kenyan crossbreds. In order to calculate the expected frequency of the haplotype from a parental breed in crossbreds, the ratio of local ancestry, $p(B)$, was obtained from the results of LAMP-LD. Then, a standardized score was computed by $\frac{\{freq(D) - mean(freq(D))\}}{stdev(freq(D))}$, where $mean(freq(D))$ and $stdev(freq(D))$ are the mean and standard deviation of $freq(D)$, respectively. This score accounts for a migration of the most frequent haplotype from a parental breed to Kenyan crossbred cattle. To calculate scores, the most frequent haplotypes that were also found in Kenyan crossbreds were included. Perl and R scripts were used to assess the flow of haplotypes.

SIGNATURES OF SELECTION

The evidence for positive selection was determined by calculating the value of the standardized integrated EHH (iHS) that measures the relative decay of EHH of the ancestral and derived core allele (Voight et al., 2006). This test was applied to detect selection signatures in Kenyan crossbreds, and presumed parental breeds. Similarly, a comparative EHH score, Rsb (Tang et al., 2007), was calculated to compare the relative decay of EHH for each marker between populations. Using Rsb, we compared the EHH of Kenyan bulls derived from Holstein-Friesians and Holsteins in the United States. The *rehh* R package was used to compute the values of iHS and Rsb with default parameters (Gautier and Vitalis, 2012).

ANNOTATION

Genes in the regions determined by local ancestry mapping and selection signatures were considered to be from candidate regions and retrieved from Biomart in Ensembl (EMBL-EBI). Using Enrichr (Chen et al., 2013) and WikiPathways (Kelder et al., 2009), biological function of genes was annotated.

RESULTS

ADMIXTURE IN KENYAN CATTLE AND IMPORTED BREEDS

The PCA and DAPC provided evidence that Friesian type cattle from the large ranch were highly correlated with Holstein cattle, whereas the animals sampled from smaller farms were admixed populations of Holstein-Friesian, Norwegian Red (or Ayrshire), and Guernsey cattle. Using DAPC with five predefined breeds, crossbreds appeared as an admixed population with Holstein and Kenyan Friesian, but relatively unrelated to Guernsey or Norwegian Red (**Figure 1A**). Three breeds, including Holsteins, Kenyan Friesians, and Kenyan crossbreds were separated by the second linear determinant (LD2), and the first linear determinant (LD1) reflected the relatively lower influence of Guernsey and Norwegian Reds in crossbred cattle. PCA plots showed similar clustering of sampled breeds except Norwegian Red that were plotted in the middle of a broad cluster of crossbreds (**Figure 1B**), which was in an agreement with history of Holstein-Friesians in Kenya.

LOCAL ANCESTRY IN KENYAN CROSSBRED CATTLE

In order to clarify the genome-wide pattern of admixture, local ancestry of cattle was inferred using the haplotype information from dairy breeds related to Kenyan crossbreds. Overall, more than half of haplotypes were inferred to be shared with those of Holsteins (0.53 ± 0.097) under the assumption of admixture with three ancestral dairy cattle breeds (Holstein, Guernsey, and Norwegian Red). When comparing the other two breeds, the genetic background of Norwegian Red cattle (0.32 ± 0.096) was found to be more influential than the ancestry from Guernsey (0.15 ± 0.078) in the Kenyan admixed population. Local ancestry of most regions was not solely dominated by an ancestral breed, whereas the ancestry of Holstein-Friesian was substantial when the influence of a breed was relatively high (>0.75), particularly in regions on BTA 10, 14, 15, or 27 (**Figure 2**). The probability of ancestry was also calculated based on the haplotypes that originated exclusively from an ancestral breed (Figure

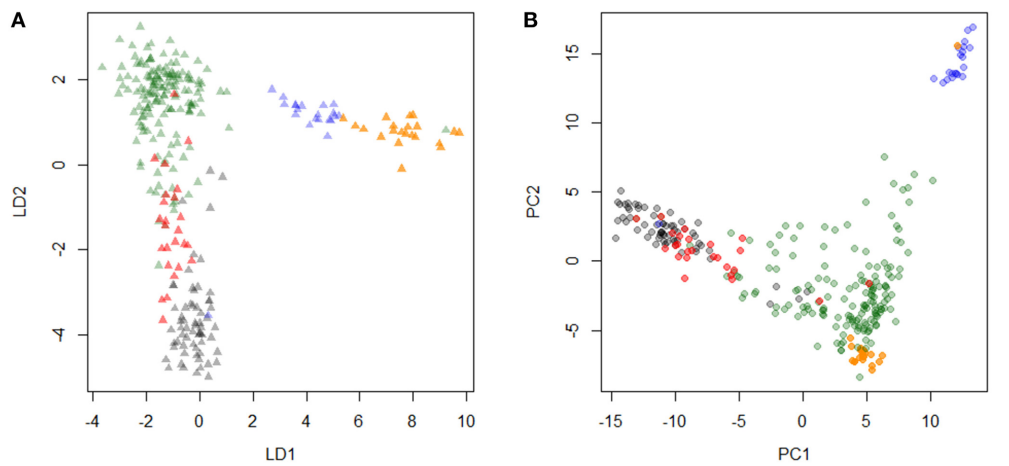


FIGURE 1 | Principal component analysis of Kenyan and other dairy cattle breeds. Discriminant analysis of principal component (DAPC) and principal component analysis (PCA) are plotted (**A,B**, respectively). Each color represents black, Holstein; red, Kenyan Friesian; green, Kenyan crossbreds; blue, Guernsey; orange, Norwegian Red.

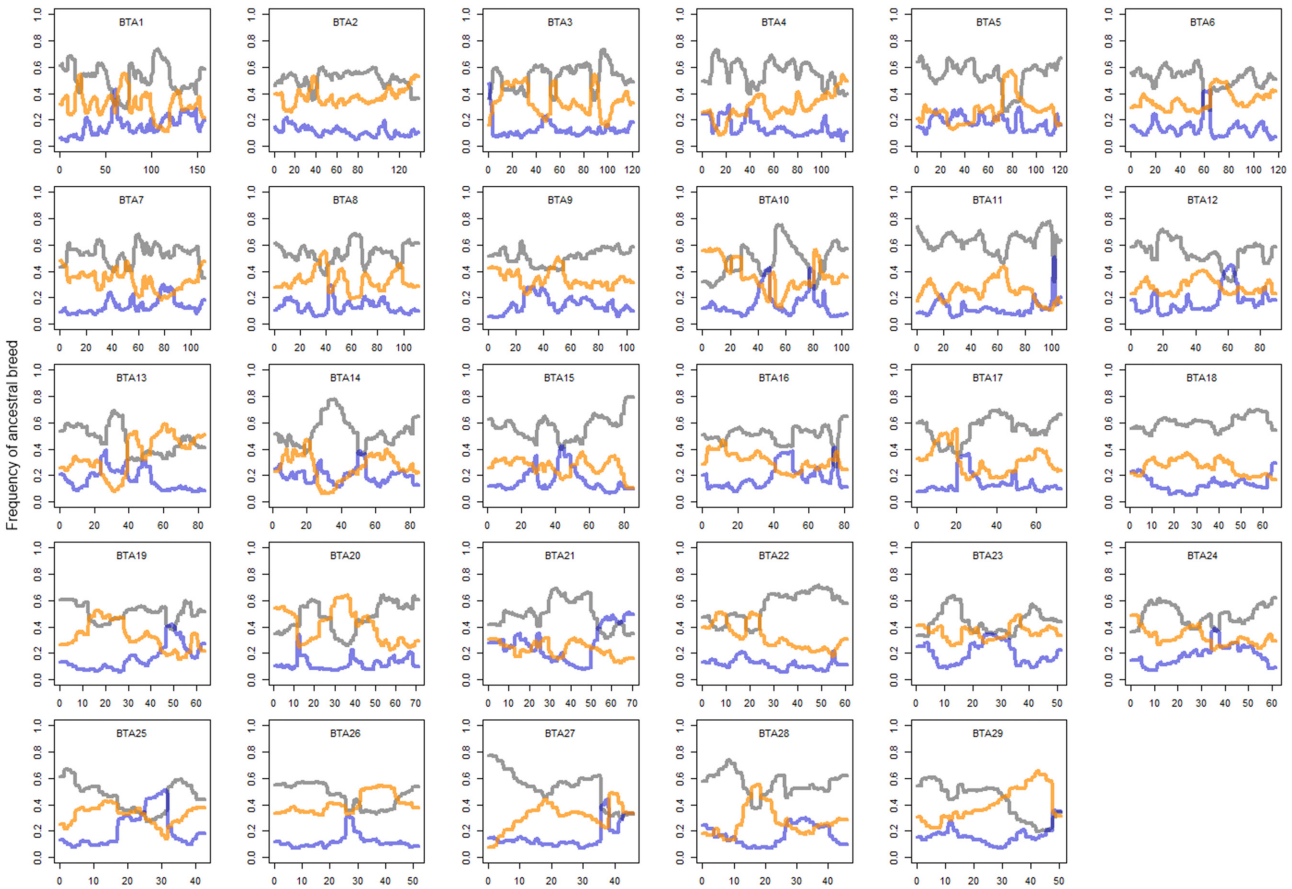


FIGURE 2 | Local ancestry of crossbreds. Gray, orange, and blue lines represent Holstein, Norwegian Red, and Guernsey, respectively. Mean local ancestry of SNP is plotted on genomic position of each chromosome. y axis is percentage ancestral breed and x axis genomic position.

S1). Interestingly, we found Norwegian Reds shared a considerable number of common ancestors with Holsteins, but the majority of haplotypes that derived from only Holstein was great compared to other breeds (Figure S1), representing the strong impact of Holstein-Friesians in Kenyan crossbreds. Nonetheless, selection signatures were not dependent on the ratio of Holstein background in crossbreds. When examining the local ancestry, we found that the ratio of Holstein background was relatively low (<0.5) in the regions on BTA 4 (40 Mb), 7 (45 Mb), 10 (85 Mb), 13 (45 Mb), 20 (20–40 Mb), and 26 (25 Mb).

HAPLOTYPES FLOW FROM HOLSTEINS TO KENYAN CROSSBREDS

The analysis of local admixture helps to infer the overall genetic background of cattle of the region. However, selection signatures are mainly detected by long haplotypes shared by individuals. To reveal the origin of haplotypes, we compared the most frequent haplotypes of crossbreds with their parental breeds (**Figure 2**; Figure S2). As shown above, a considerable amount of consensus haplotypes was found across the genome in all Kenyan dairy cattle (**Figure 3**), in particular those haplotypes with high frequency (>0.3). Approximately, 26% of the most frequent haplotypes originated from Holsteins were found in Kenyan crossbreds, whereas less than 10% of the most frequent haplotypes of Guernsey (8%) or Norwegian Red (6%) were identical to any haplotype in Kenyan crossbred cattle. In most genomic regions, frequency of the haplotype in Kenyan crossbreds was dependent on the frequency of haplotypes within their ancestral breed (**Figure 3**). Despite the fact that the most frequent haplotypes on BTA 10 and 13 remained, these appeared to be unfavorable haplotypes when considering the difference of the expected and observed frequencies in Kenyan crossbreds (**Figure 3**, **Table 1**). The frequency of the most common haplotype in crossbreds was ∼0.3 at 25 Mb on BTA 26, which originated from Holstein-Friesians. In a region on BTA 16, the difference of haplotype frequency between Holsteins and crossbreds was the highest (0.07 vs. 0.34), whereas an identical haplotype was found in all imported breeds in the region from 23 to 24 Mb, resulting in the high levels of the common haplotypes in Kenyan crossbreds (**Figure 4**).

The most frequent haplotypes on BTA 4, 5, 9, 10, 13, and 26 in Holsteins (frequency > 0.3) were shared with Kenyan crossbreds (**Table 1**), but most these haplotypes were less frequent than expected in crossbreds (**Figure 4**). Specifically, frequency of the most common haplotypes on BTA 4, 10, 13, 20, and 26 were higher than 0.4 in Holsteins using 50-SNP windows, which were probably inherited from influential ancestors. However, most these haplotypes appear to harbor unfavorable alleles in Kenyan

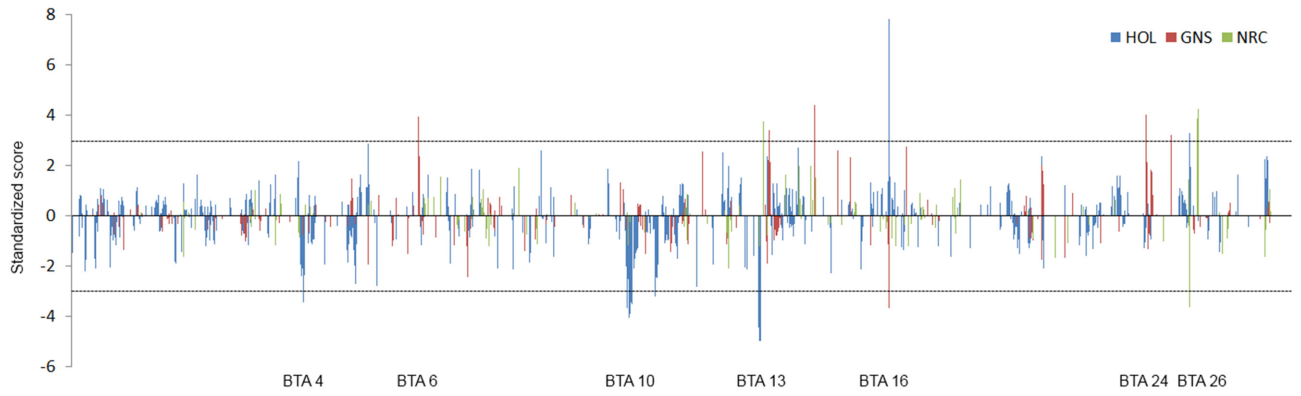


FIGURE 3 | Migration of the most frequent haplotypes from parental breeds to Kenyan crossbreds. The standardized score of haplotype migration of Holstein (blue), Guernsey (red), and Norwegian Red (green) are shown across the genome. Positive score represents the higher frequency than expected frequency of the most frequent haplotype that is originated from Holsteins. Negative value represents the opposite case of positive score. Number on x axis shows chromosome.

Table 1 | Comparisons of the most frequent consensus haplotype in parental breeds and Kenyan crossbreds.

Parental breed[a]	BTA	Region (Mb)	Frequency in parental breed[b]	Expected frequency[c]	Observed frequency[b]	Frequency change[d]	Standardized score[d]
HOL	4	49.22–57.03	0.43	0.11	0.29	−0.18	−3.11
GNS	6	62.96–64.29	0.50	0.21	0.09	0.12	3.94
HOL	10	51.97–66.12	0.34	0.25	0.05	−0.20	−4.08
HOL	13	28.92–34.03	0.44	0.29	0.05	−0.24	−5.00
GNS	13	49.86–51.78	0.67	0.19	0.09	0.10	3.40
HOL	16	43.30–45.76	0.16	0.07	0.34	0.27	7.83
GNS	16	44.05–45.76	0.52	0.19	0.34	−0.15	−3.66
GNS	24	34.81–39.06	0.48	0.18	0.06	0.12	4.03
HOL	26	22.31–24.91	0.52	0.25	0.34	0.09	3.13
NRC	26	22.31–24.91	0.45	0.18	0.33	−0.15	−3.62
NRC	26	37.33–39.74	0.38	0.21	0.05	0.16	4.26

[a] *HOL, GNS, and NRC stand for Holstein, Guernsey, and Norwegian Red cattle, respectively.*

[b] *Haplotype frequency > 0.3 in a parental breed or Kenyan crossbreds is shown.*

[c] *Expected and observed haplotypes from a parental breed in Kenyan crossbreds.*

[d] *Maximum change of haplotype frequency from a parental breed to crossbreds.*

crossbreds since the haplotype from Holstein almost disappeared in crossbreds except for a haplotype on BTA 26. The most frequent haplotypes in the candidate region of crossbreds (**Table 1**) were found to derive from Holstein-Friesians. As expected, the frequency of the most common haplotypes between Holsteins and Kenyan Friesians was correlated ($r = 0.55$). We observed four noticeable differences (frequency > 0.1) between Holsteins and Kenyan Friesians on BTA 10 (50–60 Mb), 13 (30 Mb), 5 (25, 100 Mb), and 22 (45 Mb). In particular, the intervals on BTA 10 and 13 overlap the regions that encompass the most frequent haplotypes inherited from Holstein ancestors in crossbreds (**Table 1**).

SIGNATURES OF SELECTION IN CROSSBREDS, HOLSTEINS, AND KENYAN FRIESIANS

We searched for signatures of selection in Kenyan crossbreds, which could support that a given region(s) was probably involved in adaptation. As a consequence, candidate regions can be defined by density of high $|iHS| > 3$, including the regions on BTA 4, 7, 10, 11, 13, 20, and 26 (**Figure 5**; **Table 2**). The signatures of selection on BTA 4, 10, and 20 agreed with those in Kenyan Friesians and Holsteins, whereas the most significant signals on BTA 13 and 26 were detected only in crossbreds. Next, the relatedness of crossbreds and Holstein-Friesians was assessed using Rsb, which may suggest the evidence of common candidate regions in Holsteins and Friesians as well as new candidate regions within crossbreds. However, most significant values of Rsb (Holsteins/Kenyan Friesians) higher than the absolute level of three were positive values, reflecting higher levels of haplotype homozygosity in U.S. Holsteins compared to the EHH of Kenyan Friesians. The F_{ST} between Kenyan crossbreds and Friesians was 0.029, which was lower than the value (0.069) between Holsteins and crossbreds.

Additionally, Kenyan Friesians were compared with Holsteins. To assess the similarity between Holstein and Kenyan Friesian cattle, firstly, we measured a differentiation of Holsteins and

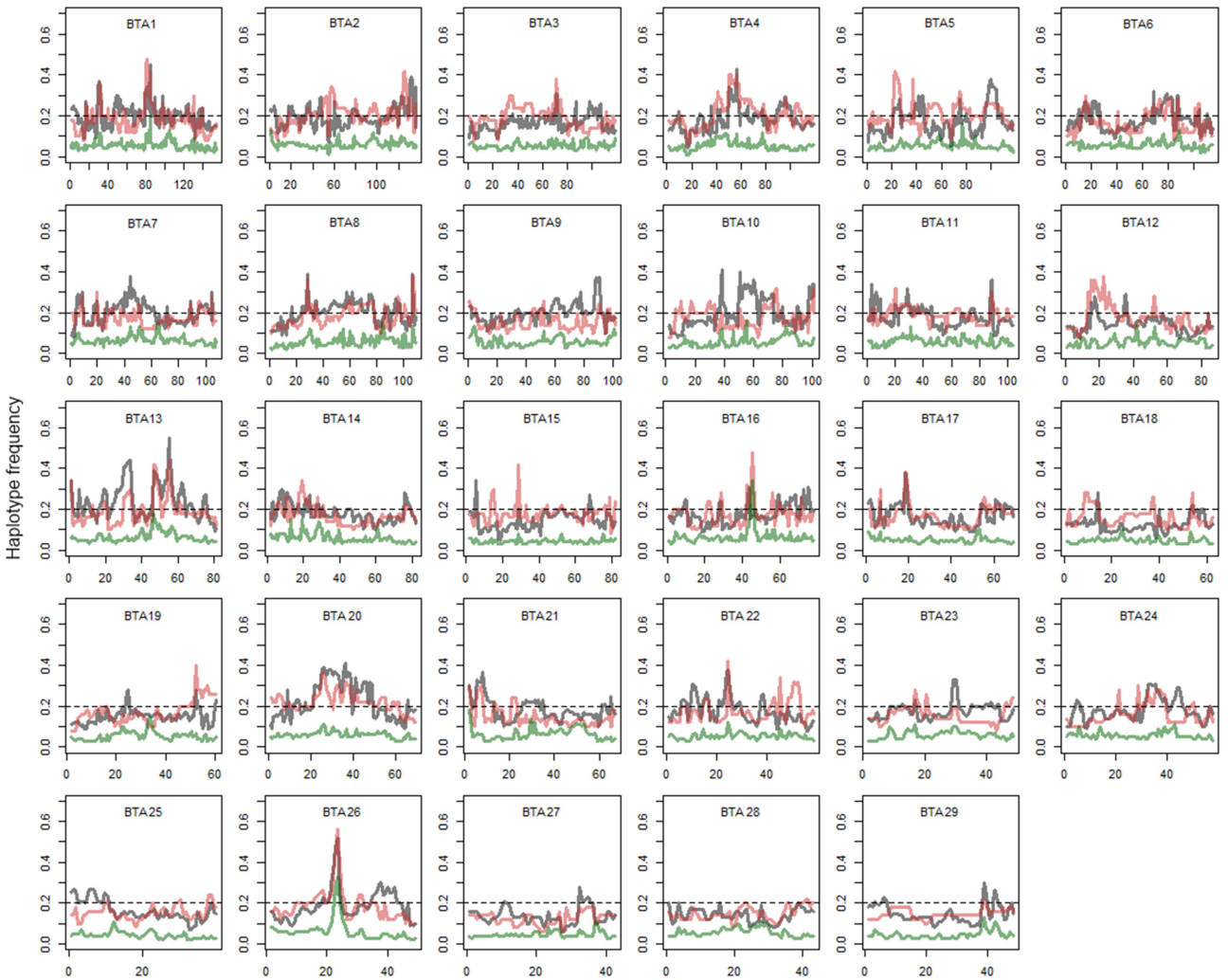


FIGURE 4 | The most frequent haplotypes in Holstein, Kenyan Friesian, and crossbreds. Gray, red, and green lines represent Holstein, Kenyan Friesian, and Kenyan crossbreds, respectively. Gray line, Holstein; Red line, Kenyan Friesian; Green line, Kenyan Crossbred. Frequency of the most common haplotype in each breed is shown. y axis is frequency of most frequent haplotype and x axis shows genomic position (Mb).

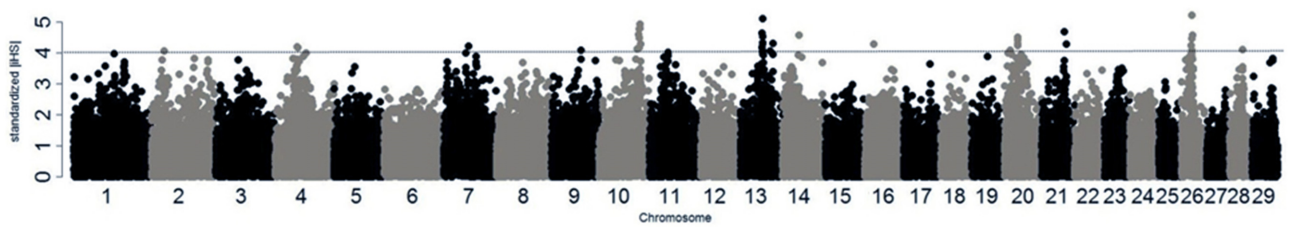


FIGURE 5 | iHS in Kenyan crossbreds. |iHS|at each SNP loci is plotted across the genome. Dotted line indicates |iHS|= 4.

Friesians using F_{ST}. The mean F_{ST} was relatively low (0.012) and only 0.2% of F_{ST} exceeded 0.2 across the genome (Figure S3), which was lower than the mean F_{ST} of Kenyan Friesian and Guernsey (0.078) or Kenyan Friesian and Norwegian Red (0.028). The analysis of EHH revealed several differentially selected regions (|Rsb|> 3) on BTA 1, 3, 5, 9, and 23 that were not found from the results of F_{ST}, which may reflect the history recent selection for a few decades compared to F_{ST}. However, most regions did not differ substantially across the genome. Then, we carried out the analysis of iHS for each group, resulting in 98 significant signals (|iHS|> 3) in Kenyan Friesians (Figure S4), and only 24 significant loci were detected in Holsteins. Using the iHS,

Table 2 | Signatures of selection in Kenyan crossbreds.

BTA	Max (iHS)	Position (Mb)	Region (Mb)[a]	Gene[b]
4	4.21	48.72	37.87–67.08	*MDFIC*
7	4.58	45.01	44.58–49.19	*TRPC6*
10	5.01	87.19	81.45–88.37	*TSHR*
11	4.16	31.43	28.23–32.96	*FBXO11*
13	5.58	47.37	45.07–49.14	*LARP5*
20	4.06	25.5	22.49–39.06	*HSBP3*
26	5.35	23.41	20.10–25.67	*SORCS3*

[a] *Region defined by at least 3 continuous significant (iHS > 3) per 1 Mb.*
[b] *Gene located at the loci (or nearest loci) with maximum iHS.*

signatures of selection of Kenyan Friesians and Holsteins showed low levels of correlation ($r = 0.13$).

GENES INVOLVED IN ADAPTATION

We selected candidate regions that were minimally related to signatures of selection in Holstein-Friesians to find gene(s) that were possibly involved in adaptation. The candidate regions with unique iHS on BTA 11 was chosen first, and then, four regions on BTA 14, 15, and 27 were selected for the excessive ancestry of Holsteins (>0.75). Additionally, the potentially advantageous haplotypes on BTA 16 and 26 in crossbreds were included for this step. The gene *PIK3CD* on BTA 16 is involved in IL-2 signaling and Toll-like receptor (TLR) signaling pathway with *MTOR* (BTA 16) and *NFKB2* (BTA 26), respectively. *IL-2* is a multifunctional cytokine with pleiotropic effects on T cells, B cells, and natural killer cells, and the TLR signaling pathway detects microbial pathogens and is involved in generating innate immune responses. On BTA 11, the genes *LHCGR* and *FSHR* encode receptors of luteinizing hormone/choriogeonadotropin and follicle stimulating hormone that play an important role in the reproductive development process. The gene *CYP17A1* (BTA 26) is also related to the reproductive and gonad development. The candidate region on BTA 14 and 27 was defined at loci with the maximum ratio of Holstein ancestry in flanking region (2 Mb). *NCOA2* (35 Mb) on BTA 14 is related to the regulation of transcription and RNA metabolic process with several genes located on different chromosomal regions, including *RHOQ* (BTA 11), *CTNNBIP1* (BTA 16), *PEX14* (BTA 16), *NFKB2* (BTA 26), *SUFU* (BTA 26), and *LDB1* (BTA 26).

Although haplotypes that originated from Guernsey or Norwegian Reds were not commonly found in Kenyan crossbreds, the analysis of local ancestry revealed the substantial influence of Norwegian Reds in some regions. Haplotypes from Norwegian Red were the most frequent (>0.5) at three regions (>5 Mb) on BTA 5, 20, and 29 (Figure S3) in crossbred cattle, which implies the advantage of Norwegian Red (Ayrshire) background. Among genes located in the regions, *OSMR* on BTA 20 is a notable gene that participates in regulation of inflammatory response, as well as response to cytokine stimulus and cytokine-cytokine receptor interaction with *LIFR* (BTA 20).

DISCUSSION

In East Africa, there has also been a routine use of Holstein-Friesian semen, whereas the larger footprints of selection in Kenya would be affected by a substantial amount of British Friesian introduction compared to recent Holstein. In Kenya, most of the AI is from the government AI Centre, which uses bulls that have been maintained in the country for many years. A previous study found that the Kenyan admixed cattle population consisted of 30–40% Holstein-Friesian and 60–70% Guernsey (Gorbach et al., 2010), but the crossbreds were not likely to come from only two breeds when one considers more European dairy breeds, including Ayrshire, Guernsey, and Jersey, have been widely used for AI. Furthermore, the ancestry of native African or *bos indicus* are likely to be small in Kenyan dairy cattle.

The Kenyan crossbreds in small farms are expected to be more adapted to local environments, although crossbreeding is not a favorable mating system for dairy cattle in Western countries. Interestingly, we could identify several regions under selection using iHS in Kenyan crossbreds. To interpret this, it is noted that dairy cattle in East Africa have been adapted through crossbreeding of imported breeds and probably some indigenous *bos taurus* cattle that cannot be inferred in this study. The sires used for AI varied in their ancestry with anywhere from 30 to 98% Holstein-Friesians (Gorbach et al., 2010), and the contemporary Friesian herd comprises 45% of the national dairy herd in Kenya. The haplotype based methods tend to be sensitive to the influential common ancestors, implying the effect of an ancestral breed. Thus, we assumed selection signatures in Kenyan crossbreds were greatly dependent on the Holstein genetics. Several candidate regions were detected on BTA 7, 13, 20, and 26 in crossbreds and which overlapped for the same regions with high levels of runs of homozygosity (ROH > 0.16) in the North American Holsteins (Kim et al., 2013). However, we could not clarify the similarity of signature of selection between Kenyan cattle and Western dairy breeds only by comparisons of statistical scores based on the decay of EHH.

Identifying the distinct ancestry of genomic segments has a wide range of applications from disease mapping to learning about history (Sankararaman et al., 2008), and allow us to reveal the favorable local genetic background from a specific breed in an admixed livestock population. The analysis of genomic sequences revealed selection for Asian genes that introgressed into European pigs, demonstrating the effect of artificial selection to improve the fertility in Europe during since nineteenth centuries (Bosse et al., 2014). In Kenyan cattle, migration of haplotypes from Western countries was assessed, while time to the first generation of imported parental breeds may be insufficient to identify adaptation under natural selection. Consequently, adaptation of an allele with low initial frequency is unlikely to be clarified by natural or artificial selection during the last few decades.

Estimates of the most frequent haplotype in crossbreds are fairly similar to half of the corresponding haplotypes in Holstein-Friesians. Norwegian Reds have been selected using a multiple breeding objective, with increasing emphasis on functional traits like health and fertility (Mason, 1988). During the twentieth century, U.S. Holstein and Norwegian Red cattle shared some common ancestors because germplasm of Norwegian Red has been exported to the United States for crossbreeding with Holstein cattle (Heins et al., 2006). PCA also supported the history of this breed. On the contrary, the exact percentage

of ancestry is debatable. Among all haplotypes, 17.6% belong to Holstein-Friesian ancestry exclusively, and 1.6 and 0.9% attributed to Norwegian Reds and Guernsey, respectively (Figure S1). Conversely, nearly 80% of haplotypes could belong to any ancestral group. Nonetheless, overall local ancestry information allowed us to reveal possible regions that correspond to adaptation. The excessive ancestry of Holstein (>75%) was inferred at 30 Mb on BTA 14, 80 Mb on BTA15, and 1 Mb on BTA 27 (Figure S4). When comparing to signatures of selection, these regions did not correspond to the selection signatures in Holsteins or crossbred cattle, which implies a potential adaptation of Holstein backgrounds that were not involved in objectives of breeding program in Western countries. Some regions under influence of the same percentage of ancestry, particularly, a region encompassing the MHC from 20 to 30 Mb on BTA 23, may reflect the advantage of high diversity (**Figure 4**) i.e., high diversity in the MHC derived from diverse ancestors is likely to be beneficial in unfavorable and changing environments.

The ratio of Holstein background could be overestimated in crossbreds due to a limited number of ancestral populations used for HMM. Nevertheless, the measurement of a specific haplotype flow allowed us to infer the remaining ancestry in a genomic region. A comparison revealed that the most frequent haplotype on BTA 26 in crossbreds was found in Holstein-Friesian and Norwegian Red cattle. This region on BTA 26 was also detected using iHS in Kenyan crossbreds, but no sizable signature of selection has been reported in the same region in previous studies of Holstein-Friesians. Clearly, the most common haplotype in this candidate region originated from Holstein-Friesian or Norwegian Red ancestry, which may result in significant standardized score in this region. On a wide region of BTA 26, significant associations of unsaturated fatty acid were detected in Dutch Holsteins (Bouwman et al., 2012), overlapping the candidate region on BTA 26 in Kenyan crossbreds. An obvious change in haplotype frequency was found on BTA 16 where all breeds shared the common haplotype in a very narrow region at 42 Mb. However, the similarity in flanking region was the highest when compared with Kenyan Friesians. Thus, this region may be considered either as resulting from adaptation or differential selection between Holsteins and Friesians.

The results from PCA and F_{ST} supported the assumption that Holsteins did not considerably differ from Friesians. Nevertheless, we could not identify common selection signatures from the results of iHS in Holsteins and Kenyan Friesians. A clue to elucidate common selection signatures was found when surveying haplotypes. Indeed, the pattern of the most frequent haplotype resembles greatly that found in Holsteins and Friesians, except two regions on BTA 10 and 13. The selection signature on BTA 10 was reported in German, Dutch, and U.S. Holsteins (Qanbari et al., 2010; Elferink et al., 2012; Kim et al., 2013), but not in Israeli Friesians (Glick et al., 2012). The frequency of the most common haplotype at 30 Mb on BTA 13 has increased during the last few decades in the U.S. Holsteins, and was associated with milk yield (Kim et al., 2013). Furthermore, one of the largest genomic differences between U.S. Holsteins and Kenyan Friesians were also these two regions on BTA 10 and 13, which may account for the differences between U.S. Holsteins and Kenyan crossbreds. It is also worth mentioning the haplotype migration of BTA 4 from Holsteins to Kenyan crossbreds. On BTA 20, haplotypes inherited from Norwegian Red (Ayrshire) appear to be more frequent than those of Holstein-Friesians in crossbreds. Most previous studies reported strong evidence of selection signatures on BTA 20 in Holstein-Friesians, which agrees with common frequent haplotypes in Holsteins and Friesian in our study. These findings may explain some genetic reasons for the poor production abilities of Kenyan dairy cattle.

We assumed that excessive local ancestry of a specific breed could be an evidence of adaptation. However, in principle, the benefits of crossbreeding are mainly attributed to increased levels of heterozygosity in the F1 generation. Conversely, excessive alleles from only one ancestral breed may not be a desirable inheritance mode in crossbreds unless they confer some advantage. We defined cattle in local farms as Kenyan crossbreds, but their population structure is more complex than crossbreds in Western countries. More importantly, Kenyan crossbreds with imported cattle have been exposed to infectious diseases and other conditions under poor environments. Thus, we emphasized on the contribution of a breed to crossbreds in a local segment of genome using admixture mapping (Shriner, 2013).

Despite existence of common haplotypes, some alleles probably associated with high milk yield were not fully transferred from Holsteins and selected for in crossbreds in local farms. This is not surprising because crossbreeding may pursue higher survivability rather than direct improvement of ability to produce milk yield. However, it is unclear which regions were involved in adaptation to the specific environmental conditions in Kenya without records. One earlier original aim of the Kenyan national AI has been to control the spread of infectious disease among cattle (Duncanson, 1977), but there was no long-term selection program for the resistance to a specific disease. Heat tolerance has not been an objective of dairy breeding in Western countries. Therefore, we regard unique selection signatures or exceptional influence of an ancestral breed in crossbreds as evidence of possible adaptation to East Africa.

For worldwide cattle production in the twenty-first century, it will be necessary to explore the adaptation of cattle in unfavorable environments, where there has been selection for alleles at many loci offering specific environmental adaptation (O'Neil et al., 2010). To overcome low survivability and poor productivity, crossbreeding has been widely applied in East Africa. However, the genomic features underlying the profits of crossbreeding have not been investigated with a view of the entire genome. While three or more breeds are expected to have been used in Kenyan crossbred cattle, the composition of long extended haplotypes has been strongly depended on a few influential bulls that were mostly Holstein-Friesian cattle. Those common haplotypes were probably selected for economic traits rather than survivability in Western countries. Nonetheless, some regions with the excessive ancestry of a specific breed appear to be unrelated to the recent signatures of selection in their parental breed. In the Kenyan crossbreds, Holstein-Friesians background is usually expected to provide the ability of higher milk yield, whereas other breeds may support health or fertility. However, we suggest from the results that some local ancestry of Holstein-Friesians may be

advantageous to adaptation to a new environment. Although local adaptation and selection signatures have been identified in Kenyan cattle, these need to be allied to industry efforts to characterize the different aspects of performance in new environments (Rothschild and Plastow, 2014). To better understand adaptation, a genome-wide analysis of local ancestry is required in an admixed population. This type of analysis may enable researchers a clearer view of the details in genetic background that may contribute to survivability. Thus, our results in Kenyan admixed cattle may provide useful information for objective dairy breeding in both temperate and non-temperate climate regions.

ACKNOWLEDGMENTS

Funding was provided by the Ensminger Endowment, State of Iowa and Hatch funding. Previous support the International Livestock Research Institute for data collection is appreciated. Previous contributions to the earlier data collection phase of this project by J. Reecy, S. Kent are appreciated. Comments by T. Sonstegard are appreciated.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://www.frontiersin.org/journal/10.3389/fgene.2014.00443/abstract

REFERENCES

Baran, Y., Pasaniuc, B., Sankararaman, S., Torgerson, D. G., Gignoux, C., Eng, C., et al. (2012). Fast and accurate inference of local ancestry in Latino populations. *Bioinformatics* 28, 1359–1367. doi: 10.1093/bioinformatics/bts144

Barendse, W., Harrison, B. E., Bunch, R. J., Thomas, M. B., and Turner, L. B. (2009). Genome wide signature of positive selection: the comparison of independent samples and the identification of regions associated to traits. *BMC Genomics* 10:178. doi: 10.1186/1471-2164-10-178

Bohmanova, J., Misztal, I., and Cole, J. B. (2007). Temperature-humidity indices as indicators of milk production losses due to heat stress. *J. Dairy Sci.* 90, 1947–1956. doi: 10.3168/jds.2006-513

Bosse, M., Megens, H.-J., Frantz, L., Madsen, O., Larson, G., Paudel, Y., et al. (2014). Genomic analysis reveals selection for Asian genes in European pigs following human-mediated introgression. *Nat. Commun.* 5, 4392. doi: 10.1038/ncomms5392

Bouwman, A. C., Visker, H. M., van Arendonk, J. A., and Bovenhuis, H. (2012). Genomic regions associated with bovine milk fatty acids in both summer and winter milk samples. *BMC Genet.* 13:93. doi: 10.1186/1471-2156-13-93

Bovine HapMap Consortium. (2009). Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Science* 324, 528–532. doi: 10.1126/science.1167936

Browning, S. R., and Browning, B. L. (2007). Rapid and accurate haplotype phasing and missing data inference for whole genome association studies using localized haplotype clustering. *Am. J. Hum. Genet.* 81, 1084–1097. doi: 10.1086/521987

Chen, E. Y., Tan, C. M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G. V., et al. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* 14:128. doi: 10.1186/1471-2105-14-128

Duncanson, G. R. (1977). *The Kenya National Artificial Insemination Service. Animal Breeding: Selected Articles from the World Animal Review.* Rome: FAO.

Elferink, M. G., Bovenhuis, H., Veerkamp, R. F., Coffey, M. P., Wall, E., McParland, S., et al. (2012). "Signatures of selection in Holstein Friesian dairy cattle," in *European Federation of Animal Science, 63rd Annual Meeting* (Bratislava).

Flori, L., Fritz, S., Jaffrézic, F., Boussaha, M., Gut, I., Heath, S., et al. (2009). The genome response to artificial selection: a case study in dairy cattle. *PLoS ONE* 4:e6595. doi: 10.1371/journal.pone.0006595

Gautier, M., and Vitalis, R. (2012). rehh: an R package to detect footprints of selection in genome-wide SNP data from haplotype structure. *Bioinformatics* 28, 1176–1177. doi: 10.1093/bioinformatics/bts115

Glick, G., Shirak, A., Uliel, S., Zeron, Y., Ezra, E., Seroussi, E., et al. (2012). Signatures of contemporary selection in the Israeli Holstein dairy cattle. *Anim. Genet.* 43(Suppl. 1), 45–55. doi: 10.1111/j.1365-2052.2012.02348.x

Gorbach, D. M., Makgahlela, M. L., Reecy, J. M., Kemp, S. J., Baltenweck, I., Ouma, R., et al. (2010). Use of SNP genotyping to determine pedigree and breed composition of dairy cattle in Kenya. *J. Anim. Breed. Genet.* 127, 348–351. doi: 10.1111/j.1439-0388.2010.00864.x

Heins, B. J., Hansen, L. B., and Seykora, A. J. (2006). Production of pure Holsteins versus crossbreds of Holstein with Normande, Montbeliarde, and Scandinavian Red. *J. Dairy Sci.* 89, 2799–2804. doi: 10.3168/jds.S0022-0302(06)72356-6

Jombart, T., and Ahmed, I. (2011). adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* 27, 3070–3071. doi: 10.1093/bioinformatics/btr521

Jombart, T., Devillard, S., and Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet.* 11:94. doi: 10.1186/1471-2156-11-94

Kelder, T., Pico, A. R., Hanspers, K., van Iersel, M. P., Evelo, C., and Conklin, B. R. (2009). Mining biological pathways using wikipathways web services. *PLoS ONE* 4:e6447. doi: 10.1371/journal.pone.0006447

Kim, E. S., Cole, J. B., Huson, H., Wiggans, G. R., van Tassell, C. P., Crooker, B. A., et al. (2013). Effect of artificial selection on runs of homozygosity in U.S. Holstein cattle. *PLoS ONE* 8:e80813. doi: 10.1371/journal.pone.0080813

Mason, I. L. (1988). *World Dictionary of Livestock Breeds, Types and Varieties*, *3rd Edn.* Wallingford, UK: C.A.B International.

Nielsen, R., Hellmann, I., Hubisz, M., Bustamante, C., and Clark, A. G. (2007). Recent and ongoing selection in the human genome. *Nat. Rev. Genet.* 8, 857–868. doi: 10.1038/nrg2187

O'Neil, C. J., Swain, D. L., and Kadarmideen, H. N. (2010). Evolutionary process of Bos taurus cattle in favourable versus unfavorable environments and its implications for genetic selection. *Evol. App.* 3, 422–433. doi: 10.1111/j.1752-4571.2010.00151.x

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795

Qanbari, S., Pimentel, E. C., Tetens, J., Thomas, L., FerThaller, G., Lichtner, P., et al. (2010). A genome-wide scan for signatures of recent selection in Holstein cattle. *Anim. Genet.* 41, 377–389. doi: 10.1111/j.1365-2052.2009.02016.x

Rothschild, M. F., and Plastow, G. S. (2014). Applications of genomics to improve livestock in the developing world. *Livest. Sci.* 166, 76–83. doi: 10.1016/j.livsci.2014.03.020i

Sankararaman, S., Sridhar, S., Kimmel, G., and Halperin, E. (2008). Estimating local ancestry in admixed populations. *Am. J. Hum. Genet.* 82, 290–303. doi: 10.1016/j.ajhg.2007.09.022

Shriner, D. (2013). Overview of admixture mapping. *Curr. Protoc. Hum. Genet.* 76, 1.23.1–1.23.8. doi: 10.1002/0471142905.hg0123s76

Tang, K., Thornton, K. R., and Stoneking, M. A. (2007). New approach for using genome scans to detect recent positive selection in the human genome. *PLoS Biol.* 5:e171. doi: 10.1371/journal.pbio.0050171

VanRaden, P. M., and Sanders, A. H. (2003). Economic merit of crossbred and purebred US dairy cattle. *J. Dairy Sci.* 86, 1036–1044. doi: 10.3168/jds.S0022-0302(03)73687-X

Van Vleck, L. D., Pollak, E. J., and Oltenacu, E. A. B. (1986). *Genetics for the Animal Sciences*. New York, NY: W.H. Freeman and Co.

Voight, B. F., Kudaravalli, S., Wen, X., and Pritchard, J. K. (2006). A map of recent positive selection in the human genome. *PLoS Biol.* 4:e72. doi: 10.1371/journal.pbio.0040072

Winkler, C. A., Nelson, G. W., and Smith, M. W. (2010). Admixture mapping comes of age. *Annu. Rev. Genomics Hum. Genet.* 11, 65–89. doi: 10.1146/annurev-genom-082509-141523

Wright, S. (1951). The genetical structure of populations. *Ann. Eugen.* 15, 323–354. doi: 10.1111/j.1469-1809.1949.tb02451.x

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 06 August 2014; accepted: 03 December 2014; published online: 19 December 2014.
*Citation: Kim E-S and Rothschild MF (2014) Genomic adaptation of admixed dairy cattle in East Africa. Front. Genet. **5**:443. doi: 10.3389/fgene.2014.00443*

This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics.
Copyright © 2014 Kim and Rothschild. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in GENETICS

ORIGINAL RESEARCH ARTICLE
published: 03 February 2015
doi: 10.3389/fgene.2015.00013

Population genetic structure, linkage disequilibrium and effective population size of conserved and extensively raised village chicken populations of Southern Africa

Khulekani S. Khanyile[1,2], Edgar F. Dzomba[2] and Farai C. Muchadeyi[1]*

[1] *Biotechnology Platform, Agricultural Research Council, Pretoria, South Africa*
[2] *Discipline of Genetics, School of Life Sciences, University of KwaZulu-Natal, Pietermaritzburg, South Africa*

Edited by:
Paolo Ajmone Marsan, Università Cattolica del S. Cuore, Italy

Reviewed by:
Ikhide G. Imumorin, Cornell University, USA
Paolo Ajmone Marsan, Università Cattolica del S. Cuore, Italy
Joanna Szyda, Wroclaw University of Life Sciences, Poland

***Correspondence:**
Farai C. Muchadeyi, Biotechnology Platform, Agricultural Research Council, 100 Old Soutpan Road, Pretoria 0110, South Africa
e-mail: muchadeyif@arc.agric.za

Extensively raised village chickens are considered a valuable source of biodiversity, with genetic variability developed over thousands of years that ought to be characterized and utilized. Surveys that can reveal a population's genetic structure and provide an insight into its demographic history will give valuable information that can be used to manage and conserve important indigenous animal genetic resources. This study reports population diversity and structure, linkage disequilibrium and effective population sizes of Southern African village chickens and conservation flocks from South Africa. DNA samples from 312 chickens from South African village and conservation flocks ($n = 146$), Malawi ($n = 30$) and Zimbabwe ($n = 136$) were genotyped using the Illumina iSelect chicken SNP60K BeadChip. Population genetic structure analysis distinguished the four conservation flocks from the village chicken populations. Of the four flocks, the Ovambo clustered closer to the village chickens particularly those sampled from South Africa. Clustering of the village chickens followed a geographic gradient whereby South African chickens were closer to those from Zimbabwe than to chickens from Malawi. Different conservation flocks seemed to have maintained different components of the ancestral genomes with a higher proportion of village chicken diversity found in the Ovambo population. Overall population LD averaged over chromosomes ranged from 0.03 ± 0.07 to 0.58 ± 0.41 and averaged 0.15 ± 0.16. Higher LD, ranging from 0.29 to 0.36, was observed between SNP markers that were less than 10 kb apart in the conservation flocks. LD in the conservation flocks steadily decreased to 0.15 (PK) and 0.24 (VD) at SNP marker interval of 500 kb. Genomewide LD decay in the village chickens from Malawi, Zimbabwe and South Africa followed a similar trend as the conservation flocks although the mean LD values for the investigated SNP intervals were lower. The results suggest low effective population sizes particularly in the conservation flocks. The utility and limitations of the iselect chicken SNP60K in village chicken populations is discussed.

Keywords: genetic diversity, village chickens, SNPs, linkage disequilibrium, effective population size

INTRODUCTION

Extensively raised village chickens are considered a valuable source of biodiversity, with genetic variability developed over thousands of years, that could be useful in future for improvement in response to climate change and consumer demands (Delany, 2004). This diversity ought to be characterized, conserved and manipulated to suit production systems such as free-range organic farming. Surveys that can reveal the effective population sizes, inbreeding levels, the effects of natural and artificial selection, as well as population bottleneck events that shaped these populations' current genetic structures will provide valuable information that can be used to manage and conserve these indigenous animal genetic resources. Previous diversity studies (Muchadeyi et al., 2007; Mtileni et al., 2010, 2011b) used microsatellite markers that were of sparse density and could not be used to extensively estimate the population demographic parameters. However, in the presence of dense marker sets, advanced statistical genomics methods can now be used to build an understanding of population genetic and demographic parameters in the absence of pedigree records.

Linkage disequilibrium (LD), defined as non-random association of alleles at two or more loci (Hedrick, 2004; Qanbari et al., 2010) is a useful tool in genetics and evolutionary biology. Its patterns are useful in understanding the levels of inbreeding (García-Gámez et al., 2012) the genetic background of animal populations (Porto-Neto et al., 2014) and assists in the fine mapping of genes and quantitative trait loci (QTL) of economically important traits (Wragg et al., 2012). The decay and extent of LD at a pair-wise distance can be used to determine the evolutionary history of populations (Andreescu et al., 2007; Lu et al., 2012; Wragg et al., 2012). LD will therefore be of use particularly in extensively raised chicken populations in smallholder

farming systems where it can be used to calculate population genetic parameters in the absence of pedigree data.

The advent of whole genome sequencing and high density SNP genotyping technologies has resulted in increased marker density and facilitated estimation of LD in a number of domesticated animals including chickens. The completion of the first draft of the chicken genome (Hillier et al., 2004) made it possible for the development of high density markers (Groenen et al., 2011; Kranis et al., 2013). The Illumina iSelect chicken SNP60K BeadChip consists of a panel of 57,636 SNPs (Groenen et al., 2011) that have found utility in population genetic studies and LD analysis in various commercial (Qanbari et al., 2010) and traditional chicken populations (Wragg et al., 2012) as well as in other analyses such as mapping of Mendelian traits (Wragg et al., 2012) and in copy number variation screening (Jia et al., 2012).

This study sought to investigate the underlying population diversity and structure, and the extent and decay of LD in extensively raised village chicken populations of Southern Africa using samples obtained from South Africa (SA), Malawi (Mal), and Zimbabwe (Zim). These chicken populations are raised by smallholder communal farmers under village chicken farming systems characterized by low input management, uncontrolled mating systems and intermixing of flocks within and between villages (Muchadeyi et al., 2007). Population genetic structure of these chickens could be a function of small flock sizes, inbreeding (since farmers retain breeding stock from within flocks over a number of generations), as well natural selection from disease outbreaks, extreme weather conditions and poor quality feed. The objectives of the study were therefore to (i) investigate the population structure and diversity (ii) investigate the extent and decay of LD and (iii) estimate LD-based effective population sizes of extensively raised chickens from Southern Africa (Zimbabwe, Malawi, and South Africa) and provide baseline information for their management and conservation purposes.

MATERIALS AND METHODS

CHICKENS POPULATIONS, BLOOD COLLECTION AND SNP GENOTYPING

A total of 312 village chickens were randomly sampled from South Africa, Malawi, and Zimbabwe. South African village chickens were represented by chickens from Limpopo ($n = 15$), Eastern Cape ($n = 26$) and Northern Cape ($n = 35$) provinces, and four conserved flocks of Venda (VD, $n = 20$), Naked Neck (NN, $n = 20$), Potchefstroom Koekoek (PK, $n = 20$) and Ovambo (OV, $n = 10$) that are kept at the Agriculture Research Council Poultry Breeding Resource at Irene in Pretoria. Detailed sampling of these populations was described by Mtileni et al. (2011b). A total of 135 village chickens were sampled from three Zimbabwean agro-ecological zones (AEZ) of AEZ1 ($n = 92$), AEZ3 ($n = 34$), and AEZ5 ($n = 10$). The detailed sampling of Zimbabwe chicken populations is described by Muchadeyi et al. (2007). The sampling locations for both the conservation flocks and field populations of South Africa and Zimbabwe are indicated in **Figure 1**. Thirty chickens sampled from one region of central Malawi (**Figure 1**) were also used in the study. Basically the study selected individuals, households, villages, and regions to obtain genetically unrelated individuals representing a wide geographical location. The distances between villages within a district ranged from 20 to 40 km, and 100 to 500 km between districts within a province and over 1000 km between provinces. The number of individuals varied from 2 to 10 per village depending on per household chicken density in each village. All the village chickens used in this study were not selected for any commercial production traits and were raised by communal farmers under a scavenging system of production.

Blood samples had been collected on FTA Micro Cards (Whatman Bio Science, UK) described in the previous studies (Muchadeyi et al., 2007; Mtileni et al., 2011b). DNA was extracted from these FTA cards using a modified Qiagen® DNeasy Blood and Tissue protocol. DNA quality was checked on a 1% agarose gel where bright sharp bands where observed indicating an intact DNA (no degradation) and DNA concentration of 50 ng/μl for each sample was used for genotyping.

SNP GENOTYPES AND DATA PREPARATION

SNP genotyping was done using the Illumina chicken iSelect SNP60 Bead chip using the Infinium assay compatible with the Illumina HiScan SQ genotsyping platform at the Agricultural Research Council-Biotechnology Platform in South Africa. This Inifinium whole genome genotyping assay is designed to interrogate a large number of SNPs at unlimited levels of loci multiplexing (www.illumina.com). SNP calling was done using Illumina Genome Studio v2.0. The genotype input file was converted into a PLINK (v1.07) (Purcell et al., 2007) input file using a plug-in compatible with the Genome Studio program. SNP quality control was done in a number of stages depending on the downstream analysis.

BASIC POPULATION GENETIC PARAMETERS

A single data set consisting of all seven populations was filtered for SNPs that were monomorphic or had minor allele frequency (MAF) $\leq$0.02 and this resulted in a total sample of 311 chickens across the seven populations. There were 54,115 SNPs available to estimate observed and expected heterozygosity indices (H_O and H_E) as well as the inbreeding co-efficient of each population using PLINK (v1.07) software (Purcell et al., 2007). The inbreeding coefficients of the populations were tested for deviation from zero using paired *t*-tests of the Proc *t*-Test in Statistical Analysis System (SAS, 2011). PLINK (v1.07) software was also used to measure minor allele frequency distribution per population using the comprehensive data set before pruning for MAF. Bins were set for minor allele frequencies of 0–0.05, 0.05–0.1, 0.1–0.2, 0.2–0.3, 0.3–0.4, and 0.4–0.5. The proportion of SNPs per bin was calculated by dividing the number of markers per bin by the total number of markers included in the MAF estimation.

POPULATION STRUCTURE

A comprehensive SNP data set with all seven populations was filtered to remove SNPs that were either on sex chromosomes or had their positions unmapped. Markers with missing data >5%; that were monomorphic or had a MAF $\leq$2% were removed. Individuals with missing genotypes of more than 5% were also dropped. Closely related individuals, as inferred by a kinship estimate $\geq$0.45, were filtered out of the data set together with SNPs in

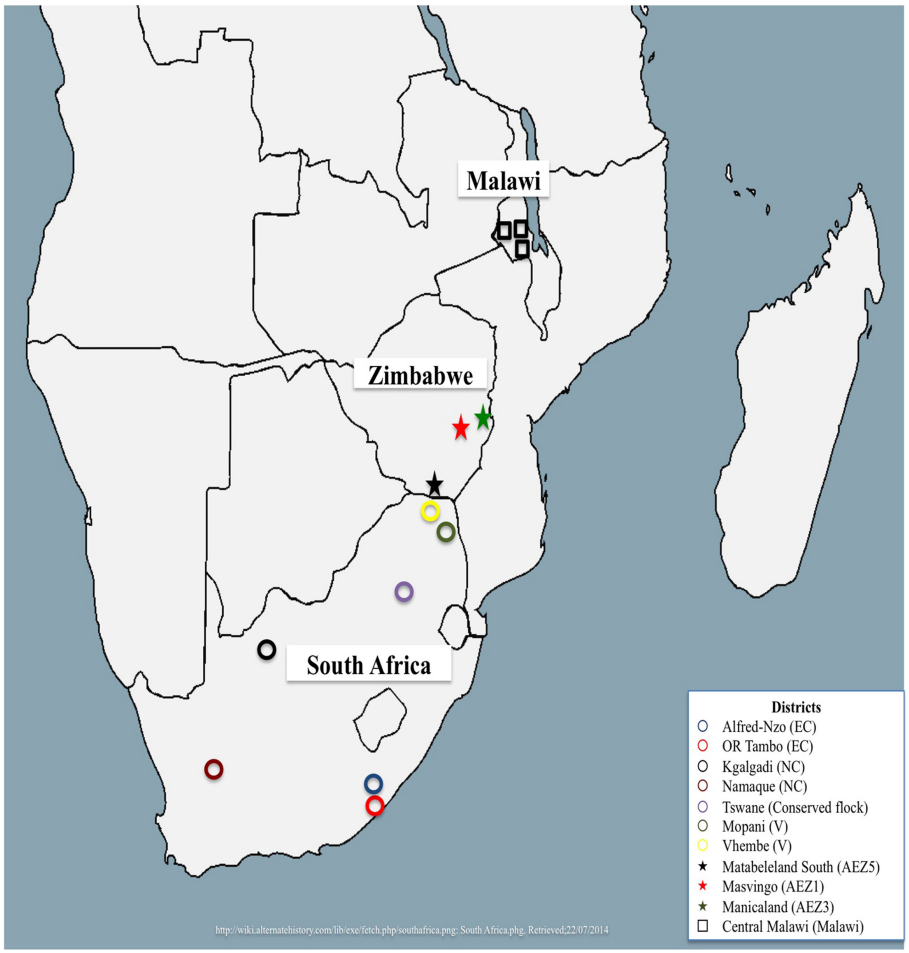


FIGURE 1 | Map showing sampled regions from Malawi, South Africa, and Zimbabwe.

high linkage disequilibrium at a threshold of LD ≥0.2. As a result 29,942 SNPs from 266 village chickens were available for analyses.

A principal component analysis (PCA) was then performed to illustrate the relationship among the extensively raised chicken populations using the Golden Helix SNP Variation Suit (SVS) version 8.1 (Golden Helix Inc., 2014).

In addition, ADMIXTURE 1.23 software (Alexander et al., 2009) was used to infer the most propable number of ancestral populations based on the SNP genotype data. Prior information on breed of origin was not used in the determination of the distinct genetic populations or in assigning individuals to populations. Admixture was run from $K = 2$ to $K = 8$ and the optimal number of clusters (K-value) was determined as that which had the lowest cross validation error (CV-error).

LINKAGE DISEQUILIBRIUM

SNP data for the individual populations were quality-controlled in order to remove SNPs (i) on sex chromosomes or those there were not mapped, (ii) with MAF ≤5%, (iii) those that deviated from Hardy-Weinberg equilibrium (HWE) ($P \leq 0.001$), (iv) with missing genotypes (>5%) as well as for individual chickens with missing genotypes (>5%) and high kinship (IBD ≥0.45) using PLINK (v1.07) (Purcell et al., 2007). After filtering, 46,973, 48,359, 48,482 as well as 38,976, 42,858, 44,920, and 44,403 SNPs on 28 autosomal chromosomes were available for the Malawian ($n = 29$), Zimbabwean ($n = 121$) and South African field populations ($n = 62$) and conservation flocks of NN ($n = 15$), OV ($n = 9$), VD ($n = 12$), and PK ($n = 18$), respectively. The level of identity by descent in the resultant data sets were 0.022, 0.038, 0.060, and 0.072 for the PK, OV, NN, and VD conservation flocks and 0.002, 0.006, and 0.004 for the South African, Malawian, and Zimbabwean village chickens, respectively. These individual population data sets were used for the estimation of linkage disequilibrium and associated estimates.

A pair-wise r^2 estimation was used to measure LD between pairs of SNPs within a chromosome and population using PLINK (v1.07) program (Purcell et al., 2007) for SNPs on autosomal chromosomes 1–28 that had passed the quality control as described above. The r^2 measure, which is defined as the squared correlation coefficient of alleles at two loci was chosen because it is independent of allele frequency (Lu et al., 2012). Briefly, its calculation, considers two loci, A and B, each locus having two alleles

(denoted A_1, A_2; B_1, B_2, respectively) (Qanbari et al., 2010). The frequencies of the haplotypes will then be denoted as f_{11}, f_{12}, f_{21}, and f_{22} for haplotypes A_1B_1, A_1B_2, A_2B_1, and A_2B_2, respectively and as f_{A1}, f_{A2} f_{B1}, and f_{B2} for A_1, A_2, B_1, and B_2, respectively. From this, r^2 was then be calculated as:

$$r^2 = \frac{\left(f_{11}f_{22} - f_{12}f_{21}\right)^2}{fA_1fA_2fB_1fB_2}.$$

By default, PLINK only reports r^2-values above 0.2 and to allow reporting of all r^2-values observed in the populations, the *–r2 –window-ld 0* option was used. An additional option, *–r2 –window-snp 5000 –kb 10000*, allowed for estimation of r^2 for SNP marker pairs separated by at most 5000 SNPs and within a 10 MB SNP interval.

An Analysis of Variance (ANOVA) was conducted using the Generalized Linear Model procedure (Proc GLM) in the SAS (2011) to determine the effects of chromosome, population, the interaction of chromosome-by-population, and SNP marker interval (bp) on LD using the following model:

$$r^2ij = \mu + \mathrm{Pop}_i + \mathrm{Gga}_j + (\mathrm{Pop} \times \mathrm{Gga})_{ij} + b\mathrm{SNP}_{\mathrm{int}} + e_{ik},$$

where: r^2_{ij} was the pairwise LD; μ was the overall population mean and Pop_i was the effect of the ith chicken population from Malawi, Zimbabwe or South Africa; Gga_j was the effect of the jth chromosome 1–28; and $\mathrm{SNP}_{\mathrm{int}}$ represented the effects of SNP interval which were defined as the distance between markers (number of base pairs) and fitted as a covariate with regression coefficient b. The F-test from the ANOVA analysis was used to determine the significance of factors included in the model at $P \leq 0.05$. Linkage disequilibrium decay was estimated genomewide for all subpopulations. Sliding window bins for LD decay were set at 10, 20, 40, 60, 100, 200, 500, 1000, 2000, and 5000 kb for chromosomes 1–28. An additional analysis of the macro-chromosomes 1–5 was done with bins up to 10,000 kb.

TRENDS IN EFFECTIVE POPULATION SIZE

The relationship between *Ne*, recombination frequency and expected LD (r^2) was determined using the following equation from Corbin et al. (2012);

$$E\left[r^2_{adj}\right] = (\alpha + 4N_e c)^{-1}$$

where $\alpha = 1$ when assuming no mutations and 2 if mutation was considered, $r^2_{adj} = r^2 - \frac{1}{2n}$, c was the recombination rate, and n was the chromosomal sample size. The effective population size *Ne*, as $\frac{1}{2c}$ generations, was estimated from the adjusted r^2_{adj} values related to a given genetic distance d in Morgans, assuming, $c = d$ (Qanbari et al., 2010).

For each pair of SNPs on each chromosome, recombination rate was estimated by converting physical marker interval length x_i(MB) to the corresponding genetic length c_i using the formula: $c_i = \bar{o}_i x_i$, where $\bar{o}_i$ is the average ratio of Morgans per kilo base pair on chromosome i, which was taken from the physical lengths of the chicken genome v74 (Ensembl, 2013). The genetic length of chromosomes was adopted from Hillier et al. (2004). The r^2-values range between 0 and 1, whereby a zero value indicates uncorrelated SNPs while a value of one reflects SNPs that are perfectly correlated (Qanbari et al., 2010).

The trends in effective population sizes for each of the defined subpopulations were then estimated by setting bins at 10, 20, 40, 60, 100, 200, 500, 1000, 2000, and 5000 kb. The bins were designed to cover the genome in tens, hundreds, thousands, and hundred thousand base pairs.

RESULTS

SNP MARKER CHARACTERISTICS

Minor allele frequency averaged 0.29 (**Table 1**) and over 8.5% of the SNPs on the Illumina iSelect chicken SNP60K panel had a MAF of less than 0.05 (Supplementary Figure 1). An analysis of the distribution of MAF across all populations showed that over 10% of the markers were within the 0–10% MAF threshold. Of the 57,636 SNPs on the panel, 29,942 were used for the determination of population structure and diversity whilst a range of 38,976 (in the Ovambo) to 48,482 (in Zimbabwe) were used for estimating LD in the different populations (**Table 1**). Majority of the SNPs excluded were either monomorphic or had minor allele frequencies ≤0.02 and were therefore considered not informative for the populations. Over 1000 SNPs had missing genotypes amongst the seven populations. SNPs located on unknown chromosomes, linkage groups, and sex chromosomes were also excluded from further analysis. The proportion of SNPs used for further analysis was 51% for the whole population for estimation of population structure and was over 80% for the village chickens from Malawi, South Africa and Zimbabwe and ranged from 67 to 77% in the conservation flocks for the estimation of LD.

BASIC POPULATION GENETIC PARAMETERS

Observed heterozygosity values averaged 0.62 ± 0.003 across all seven populations. Overall, H_0 in all populations was lower than expected (0.67 ± 0.048) and the populations were therefore significantly inbred ($P \leq 0.05$) inbred. Heterozygosity estimates and inbreeding coefficients were high in the conservation flocks compared to the village flocks. Of the conservation flocks, Ovambo chickens had the lowest levels of inbreeding.

POPULATION STRUCTURE USING PCA AND ADMIXTURE ANALYSIS

Results of the first principal component showed the conserved Venda, Ovambo, Naked Neck and Potchefstroom Koekoek chickens from South Africa grouped into four distinct clusters separated from the village chicken populations. Of the four flocks, the Ovambo clustered closer to the village chickens particularly those sampled from South Africa. Clustering of the village chickens followed a geographic gradient whereby South African chickens were closer to those from Zimbabwe than to chickens from Malawi. The chickens from Malawi clustered together with some chickens from Zimbabwe (**Figure 2**).

The optimal K-value for admixture was $K = 6$ (Supplementary Figure 2) corresponding to the conserved (i) Naked Neck, (ii) Potchefstroom Koekoek (iii) Venda (iv) Ovambo and (v) the village chickens from Malawi and (vi)

Table 1 | SNP distribution after quality control and the minor allele frequency (MAF), observed (H_O), expected (H_E) heterozygosities and inbreeding coefficient (F) of Malawi, South African field (SAField), Zimbabwean chicken populations as well as the Naked Neck (NN), Potchefstroom Koekoek (PK), Ovambo (OV) and Venda (VD) conservation flocks from South Africa.

	Malawi	Zimbabwe	SAField	NN	PK	OV	VD
SNPs							
Total	57,636	57,636	57,636	57,636	57,636	57,636	57,636
Sex chromosome	3202	3202	3202	3202	3202	3202	3202
Unmapped	172	172	172	172	172	172	172
Monomorphic	1992	1878	1501	4488	2341	4702	2975
MAF ≤0.02	3706	2409	2047	9648	5191	4317	5204
HWE ($P \geq 0.001$)	207	278	193	102	109	101	77
Missing Gen > 0.05	1384	1338	2039	1048	2218	2284	1086
PARAMETERS							
H_O	0.62 ± 0.00	0.62 ± 0.03	0.62 ± 0.00	0.62 ± 0.00	0.62 ± 0.00	0.62 ± 0.00	0.62 ± 0.00
H_E	0.68 ± 0.02	0.66 ± 0.04	0.65 ± 0.05	0.70 ± 0.14	0.68 ± 0.06	0.65 ± 0.01	0.71 ± 0.00
Mean MAF ±SD	0.27 ± 0.14	0.28 ± 0.13	0.29 ± 0.13	0.26 ± 0.14	0.26 ± 0.14	0.28 ± 0.13	0.24 ± 0.14
F*	0.15 ± 0.07	0.12 ± 0.12	0.07 ± 0.14	0.21 ± 0.03	0.14 ± 0.15	0.07 ± 0.04	0.24 ± 0.18

**Inbreeding coefficients of all populations were significantly > 0 at P ≤ 0.05.*

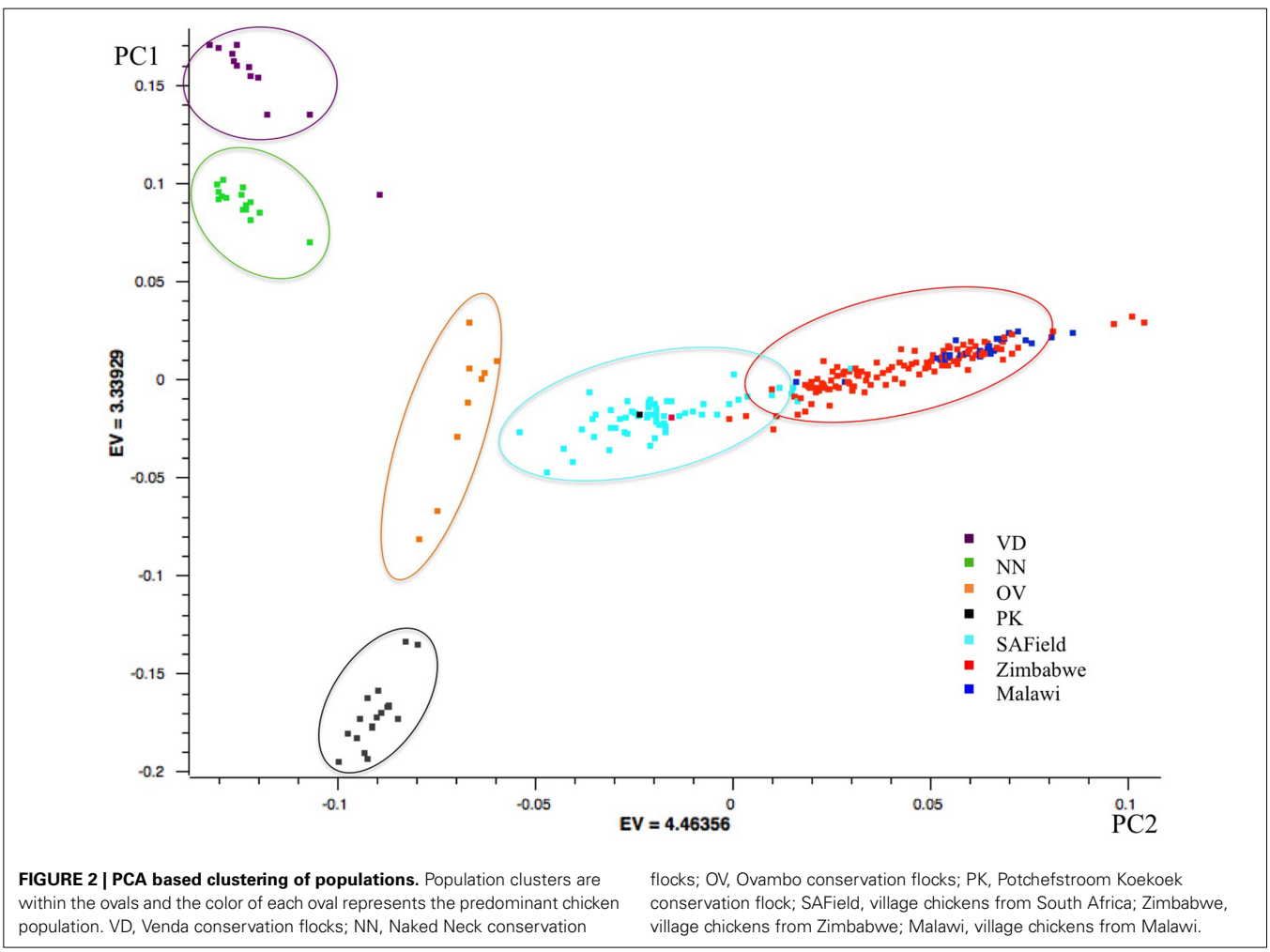


FIGURE 2 | PCA based clustering of populations. Population clusters are within the ovals and the color of each oval represents the predominant chicken population. VD, Venda conservation flocks; NN, Naked Neck conservation flocks; OV, Ovambo conservation flocks; PK, Potchefstroom Koekoek conservation flock; SAField, village chickens from South Africa; Zimbabwe, village chickens from Zimbabwe; Malawi, village chickens from Malawi.

village chicken from South Africa and Zimbabwe that had slight variations in allele frequencies (**Figure 3**). Similar to PCA results, the Ovambo chickens clustered separately but with greater diversity and some similarity to the village chicken populations (**Figure 3**). The clustering also revealed conservation of some village chickens' ancestral genomic components in the Naked Neck and Potchefstroom Koekoek flocks. The Ovambo had diverse genomic elements with some that were concentrated in the Potchefstroom Koekoek, Naked Neck and Venda conservation flocks and others that were found in the village chicken populations from the three countries.

LD ESTIMATES AND THE EFFECTS OF CHROMOSOME, SNP INTERVALS AND BREED

For each chromosome, total length, number of SNPs and average SNP interval are shown in Supplementary Table 1. **Table 2** summarizes r^2 values for the 28 autosomal chromosomes for the four conservation flocks and three village chicken populations of Southern Africa. The SNP interval was not consistent across the genome, ranging from a distance of 0.01 to 0.1 Mb. Macro-chromosomes showed the highest marker distance followed by intermediate chromosomes. SNP intervals were shorter for the micro-chromosomes (Supplementary Table 1). Number of SNPs per chromosome varied with chromosome size between the macro (chromosomes 1–5) that had the highest number of SNPs (976 to 3443) and micro (chromosome 16–28) with fewer numbers of SNPs (5 to 832) per chromosome.

Overall population LD averaged over chromosomes ranged from 0.03 ± 0.07 to 0.58 ± 0.41 and averaged 0.15 ± 0.16 (**Table 2**). The *F*-test results from the analysis of variance showed that pairwise LD varied significantly ($P < 0.001$) among chromosomes, populations and their interaction as well as with SNP marker interval (**Table 3**). Chromosome 16 had high LD in all the conservation flocks, and in Malawi and South Africa, whilst chromosome 25 had low LD across all populations. Chickens from Malawi had higher LD compared to those from South Africa and Zimbabwe. The conservation flocks had significantly higher LD compared to the village flocks. The Naked Neck and Venda conservation flocks had higher LD across chromosomes compared to Potchefstroom Koekoek and Ovambo chickens. The highest LD (0.58 ± 0.41) was observed in the Venda conservation flocks.

Linkage disequilibrium depended on SNP distance. Plots of the rate of LD decay over marker distance over all 28 autosomes and for macro-chromosomes 1–5 are given in **Figures 4A,B**, respectively. Higher LD, ranging from 0.29 to 0.36, was observed between SNP markers that were less than 10 kb apart in the conservation flocks. Within this window, LD was highest in the Venda ($LD = 0.36$) followed by Naked Neck (>0.33) and least in Ovambo and Potchefstroom Koekoek (0.29). LD in the conservation flocks steadily decreased to 0.15 (PK) and 0.24 (VD) at SNP marker interval of 500 kbp. A sudden increase in LD was observed at 500 kbp SNP interval. Genomewide LD decay in the village chickens from Malawi, Zimbabwe and

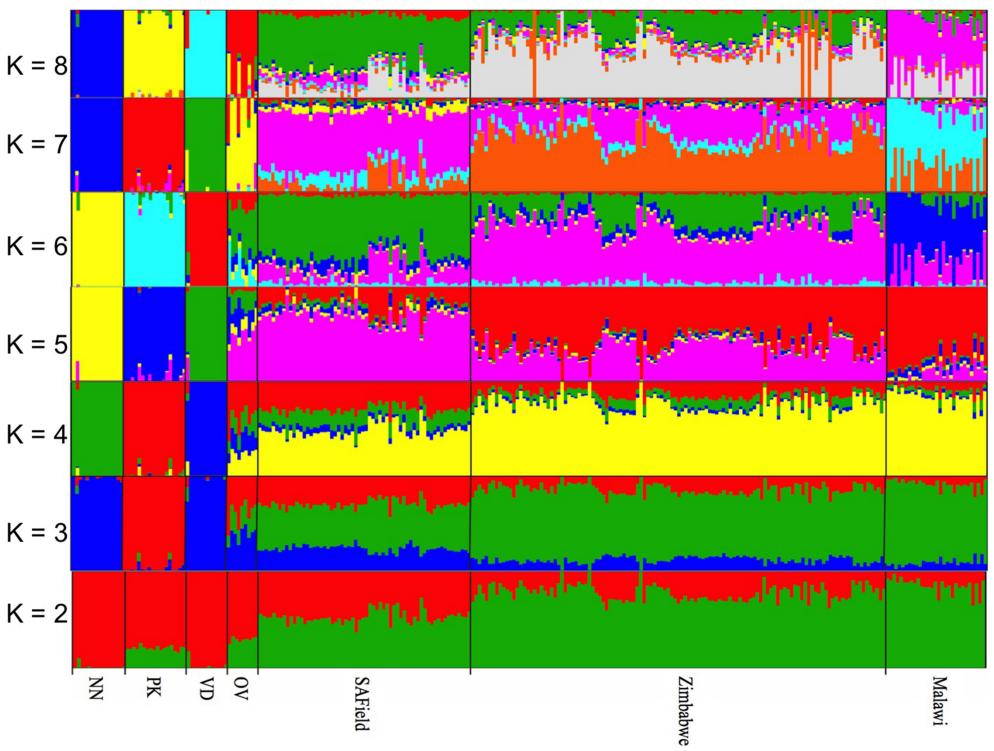


FIGURE 3 | Population substructuring using ADMIXTURE. VD, Venda conservation flocks; NN, Naked Neck conservation flocks; OV, Ovambo conservation flocks; PK, Potchefstroom Koekoek conservation flock; SAField, village chickens from South Africa; Zimbabwe, village chickens from Zimbabwe; Malawi, village chickens from Malawi.

Table 2 | Linkage disequilibrium of Malawi, Zimbabwe, South African Field (SAField) village chickens and the Naked Neck (NN), Venda (VD), Potchefstroom Koekoek (PK) and Ovambo (OV) conservation flocks.

	Linkage Disequilibrium (r^2)						
	Malawi	**Zimbabwe**	**SAField**	**NN**	**VD**	**PK**	**OV**
1	0.09 ± 0.14	0.05 ± 0.10	0.05 ± 0.10	0.23 ± 0.26	0.26 ± 0.28	0.17 ± 0.21	0.21 ± 0.23
2	0.9 ± 0.14	0.05 ± 0.10	0.06 ± 0.11	0.22 ± 0.25	0.27 ± 0.29	0.17 ± 0.21	0.21 ± 0.24
3	0.09 ± 0.14	0.05 ± 0.09	0.05 ± 0.09	0.24 ± 0.26	0.26 ± 0.28	0.17 ± 0.21	0.20 ± 0.22
4	0.09 ± 0.14	0.05 ± 0.10	0.06 ± 0.10	0.23 ± 0.26	0.26 ± 0.29	0.16 ± 0.20	0.21 ± 0.23
5	0.08 ± 0.13	0.05 ± 0.09	0.06 ± 0.10	0.23 ± 0.27	0.24 ± 0.26	0.16 ± 0.20	0.21 ± 0.22
6	0.08 ± 0.12	0.04 ± 0.08	0.05 ± 0.08	0.24 ± 0.26	0.25 ± 0.26	0.17 ± 0.21	0.22 ± 0.23
7	0.09 ± 0.14	0.05 ± 0.10	0.05 ± 0.10	0.23 ± 0.26	0.26 ± 0.29	0.16 ± 0.20	0.20 ± 0.22
8	0.11 ± 0.17	0.07 ± 0.15	0.07 ± 0.15	0.23 ± 0.26	0.28 ± 0.29	0.21 ± 0.26	0.22 ± 0.25
9	0.08 ± 0.12	0.04 ± 0.08	0.05 ± 0.08	0.19 ± 0.22	0.22 ± 0.26	0.16 ± 0.20	0.20 ± 0.22
10	0.08 ± 0.12	0.04 ± 0.09	0.05 ± 0.09	0.21 ± 0.24	0.28 ± 0.31	0.18 ± 0.22	0.20 ± 0.20
11	0.9 ± 0.14	0.06 ± 0.11	0.06 ± 0.12	0.28 ± 0.29	0.22 ± 0.33	0.23 ± 0.26	0.20 ± 0.23
12	0.09 ± 0.12	0.05 ± 0.09	0.05 ± 0.09	0.24 ± 0.27	0.33 ± 0.33	0.17 ± 0.21	0.22 ± 0.23
13	0.09 ± 0.13	0.05 ± 0.09	0.05 ± 0.09	0.23 ± 0.26	0.26 ± 0.28	0.18 ± 0.22	0.22 ± 0.23
14	0.10 ± 0.15	0.05 ± 0.10	0.05 ± 0.10	0.22 ± 0.25	0.24 ± 0.28	0.17 ± 0.21	0.21 ± 0.23
15	0.10 ± 0.15	0.07 ± 0.12	0.07 ± 0.12	0.25 ± 0.27	0.26 ± 0.28	0.18 ± 0.23	0.23 ± 0.24
16	0.10 ± 0.15	0.04 ± 0.08	0.12 ± 0.14	0.27 ± 0.28	0.58 ± 0.41	0.23 ± 0.28	0.25 ± 0.27
17	0.09 ± 0.14	0.05 ± 0.11	0.06 ± 0.11	0.24 ± 0.27	0.24 ± 0.27	0.18 ± 0.22	0.22 ± 0.24
18	0.08 ± 0.11	0.04 ± 0.08	0.04 ± 0.07	0.23 ± 0.26	0.22 ± 0.26	0.17 ± 0.21	0.20 ± 0.22
19	0.08 ± 0.11	0.04 ± 0.09	0.04 ± 0.09	0.21 ± 0.23	0.22 ± 0.25	0.17 ± 0.22	0.19 ± 0.21
20	0.09 ± 0.14	0.05 ± 0.10	0.06 ± 0.11	0.22 ± 0.26	0.32 ± 0.31	0.21 ± 0.25	0.19 ± 0.21
21	0.09 ± 0.13	0.05 ± 0.09	0.05 ± 0.09	0.21 ± 0.25	0.30 ± 0.30	0.18 ± 0.22	0.19 ± 0.21
22	0.08 ± 0.13	0.05 ± 0.09	0.05 ± 0.11	0.21 ± 0.25	0.25 ± 0.29	0.16 ± 0.21	0.19 ± 0.22
23	0.08 ± 0.11	0.03 ± 0.07	0.04 ± 0.08	0.27 ± 0.29	0.23 ± 0.27	0.16 ± 0.21	0.19 ± 0.22
24	0.08 ± 0.11	0.04 ± 0.09	0.04 ± 0.09	0.20 ± 0.21	0.24 ± 0.27	0.14 ± 0.18	0.21 ± 0.22
25	0.07 ± 0.14	0.03 ± 0.07	0.04 ± 0.07	0.17 ± 0.21	0.17 ± 0.22	0.15 ± 0.20	0.19 ± 0.21
26	0.07 ± 0.11	0.04 ± 0.07	0.04 ± 0.07	0.23 ± 0.26	0.20 ± 0.23	0.17 ± 0.21	0.18 ± 0.19
27	0.09 ± 0.14	0.04 ± 0.10	0.05 ± 0.09	0.22 ± 0.25	0.21 ± 0.26	0.17 ± 0.21	0.21 ± 0.23
28	0.09 ± 0.13	0.05 ± 0.10	0.05 ± 0.10	0.19 ± 0.22	0.25 ± 0.29	0.16 ± 0.21	0.18 ± 0.21

Table 3 | The effects of population, chromosome and SNP marker interval on linkage disequilibrium (r^2).

Factor	**DF**	**SS**	**MS**	***F*-value**	***P*-value**
Population	6	4363.4	727.23	1,892,555	***
Chromosome	27	629.76	23.32	606.98	***
SNP Distance	1	954.77	954.77	24,846.4	***
Population x chromosome	162	734.08	4.53	117.92	***

*** $p < 0.0001$.

South Africa followed a similar trend as the conservation flocks but the mean LD values at different SNP intervals were lower (**Figure 4A**).

An additional analysis of LD decay of the macro-chromosomes 1–5 was performed for all the populations and results are illustrated in **Figure 4B**. LD for these macro-chromosomes was high (>0.3) for the conservation flocks for markers within 10 kb intervals and steadily decreased to values lower than 0.15 beyond 8 Mb SNP intervals. Lower LDs of 0.2 were observed in the village chickens at 10 kb SNP interval and then decreased to less than 0.05 beyond 8 Mb.

The trends observed for LD decay per chromosome per population were similar to those for the overall genomewide population (Supplementary Figure 3). However, distinctly high LD was observed for chromosome 16 particularly for the Venda Conservation flock. The sudden increase in LD at SNP intervals higher than 500 kb was observed in some micro-chromosomes (7–11, 13, 14, 20, 23, 24, 26–28) (Supplementary Figure 2). This trend was observed only in certain conservation flocks on chromosomes 9 (VD; OV; NN), 10 (PK; VD; OV; NN), 13 and 14 (VD), and 24 (PK).

EFFECTIVE POPULATION SIZE OVER THE PAST GENERATIONS

Figures 5A,B are plots of the estimated effective population size (N_e) at t generations ago for the village and conservation flocks, respectively. The adjusted LD based estimates of N_e indicated low effective population size of 49–57 in the village chickens and of 31–50 in the conservation flocks 97 generations ago. The graphs also illustrate a steady decrease in effective population size from over 8500 to below 60 within 9000 generations for the village chickens (**Figure 5A**) and from 6000 to below 50, during the same time frame, for the conservation flocks (**Figure 5B**).

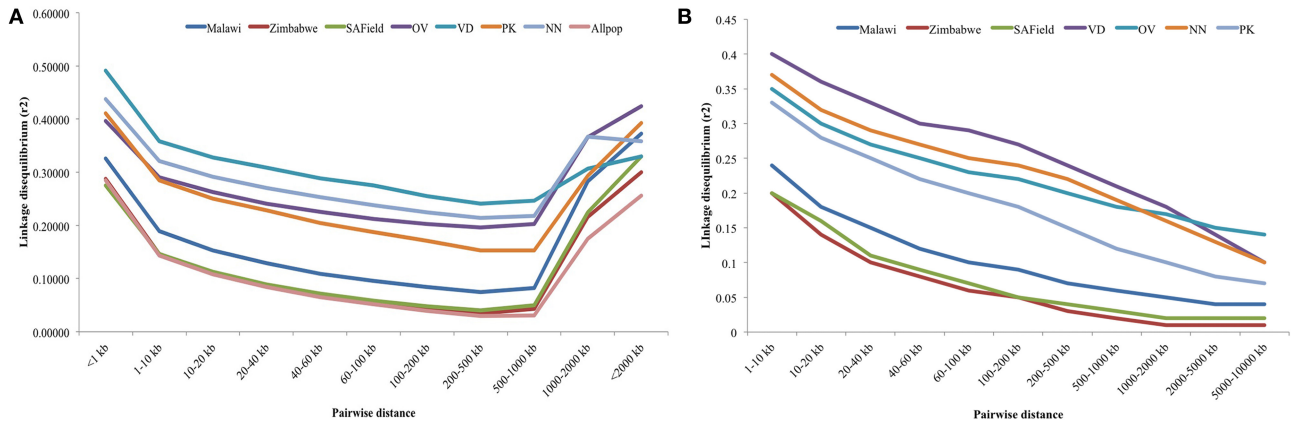


FIGURE 4 | (A) Average LD decay with increased physical distance between SNPs for chromosomes 1–28. VD, Venda conservation flocks; NN, Naked Neck conservation flocks; OV, Ovambo conservation flocks; PK, Potchefstroom Koekoek conservation flock; SAField, village chickens from South Africa; Zimbabwe, village chickens from Zimbabwe; Malawi, village chickens from Malawi. **(B)** Average LD decay with increased physical distance between SNPs for macro-chromosomes 1–5. VD, Venda conservation flocks; NN, Naked Neck conservation flocks; OV, Ovambo conservation flocks; PK, Potchefstroom Koekoek conservation flock; SAField, village chickens from South Africa; Zimbabwe, village chickens from Zimbabwe; Malawi, village chickens from Malawi.

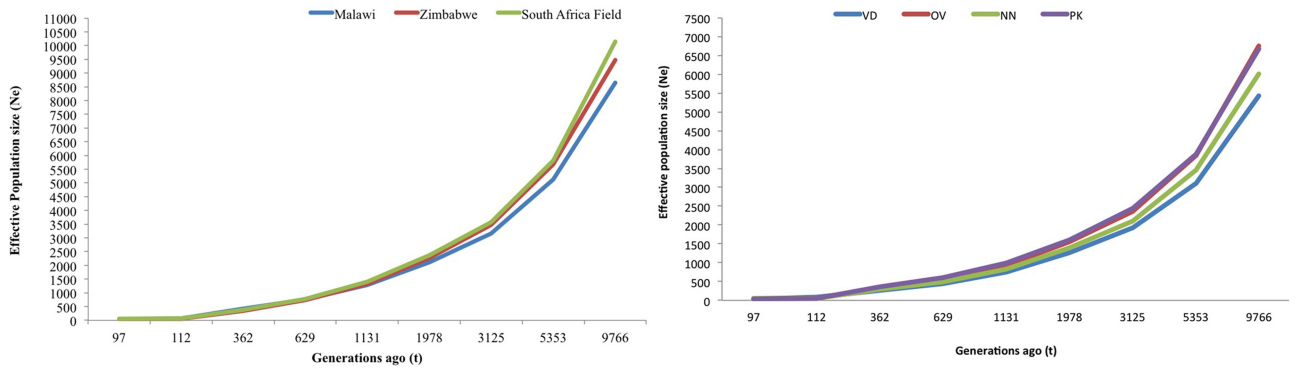


FIGURE 5 | (A) Trends in effective population size of village chicken flocks. **(B)** Trends in effective population size of conservation flocks. VD, Venda conservation flocks; NN, Naked Neck conservation flocks; OV, Ovambo conservation flocks; PK, Potchefstroom Koekoek conservation flock.

DISCUSSION

Village chicken populations in sub-Saharan Africa have not been well studied to estimate genetic and demographic parameters that are shaping their genetic structure. Previous studies have suggested that village chickens are a valuable genetic reservoir, particularly for smallholder resource-limited farmers, due to their ability to thrive in diverse geographical environments characterized by extreme climatic conditions (Hall and Bradley, 1995). Random mating and absence of pedigree data make it difficult to estimate the effective population size and other key population genetic parameters in these populations. In addition, absence of record keeping and organization hinder prospects of conducting genetic improvement programs required to improve phenotypes. Previous genetic diversity studies of village chickens utilized microsatellite markers (Muchadeyi et al., 2007; Mtileni et al., 2011b) as well as mitochondrial DNA (mtDNA) (Mobegi et al., 2006; Adebambo and Consortium, 2009; Mwacharo et al., 2011; Wani et al., 2014). Microsatellite markers are very informative but too sparse over the genome to provide accurate estimates of population genetic parameters. MtDNA sequences are informative for investigating species domestication and migration but they lack genome-wide coverage (Godinho et al., 2008). High density SNP chips have been successfully used in recent studies to characterize LD (Megens et al., 2009; Qanbari et al., 2010) and Mendelian traits and screen for other genetic variants in commercial layers (Qanbari et al., 2010) and in traditional chicken populations raised under production systems similar to those of our village chickens (Wragg et al., 2012). However, there is no information on the utility of this panel of markers for village chickens from the Southern African region. This study therefore used genome-wide SNP data to estimate population structure and diversity, linkage disequilibrium and population demographic history of extensively raised chicken populations of Southern Africa.

Only 51% of the SNPs on the iselect chicken SNP60K panel were used for the estimation of population structure and diversity. The remaining SNPs were excluded because they had MAF below the set threshold (MAF $\leq$0.02) or were in linkage disequilibrium. Conversely, over 80% of the SNPs of the panel were used for LD analysis. The number of monomorphic markers observed in this study was about 5 fold lower than those reported in Qanbari et al. (2010) in commercial layers, and 2–3 fold higher than that reported by Wragg et al. (2012) in traditional village chickens from Ethiopia, Kenya, and Chile. Even if variations in the number of monomorphic markers can be partially explained by the different number of animals sampled, it seems clear that village chickens hold a higher diversity than commercial populations and that the 60K SNP chip has some utility in genomics studies of non-descript chicken populations in spite of the ascertainment bias embedded in its design.

Population structure analysis grouped the conservation flocks into four distinct clusters that were different from the village chickens sampled from the communal farming areas of South Africa, Zimbabwe, and Malawi (**Figures 2**, **3**). This study and that of Mtileni et al. (2011b), which was based on microsatellite markers, showed that the conservation flocks have diverged from their founder village chicken populations. The levels of population divergence showed the Venda and Naked Neck were more distant to the South African village chickens than the Potchefstroom Koekoek and Ovambo chickens. Variations in levels of population divergence could have originated from different founder effects and reduced population sizes in these conservation flocks. The clustering of populations from both the PCA and ADMIXTURE indicated low levels of within population diversity of the Venda and Naked Neck conservation flocks and higher divergences of these populations from the Ovambo, Potchefstroom Koekoek and village flocks. This observation was also supported by the relatively higher heterozygosity deficiency and inbreeding coefficients of the Venda and Naked Neck conservation flocks (**Table 1**).

Conservation chickens were sampled from closed populations, kept at the Agricultural Research Council in South Africa, ranging from 100 to 150 chickens/flock (van Marle-Köster et al., 2008; Mtileni et al., 2011a) that have a narrow genetic base. The flocks were established from chickens sampled from villages in South Africa. The Venda chicken flocks were established from a few individuals based on rare plumage color variants amongst other diverse phenotypes in the Limpopo province of South Africa (van Marle-Köster et al., 2008). The Naked Neck, as their name implies, were similarly sub-sampled from a heterogenous pool of village chickens in the Eastern Cape provinces of South Africa for this phenotype caused by a single gene that is dominantly expressed, and is considered to be one of only a few distinct phenotypes observed in most village chickens in South Africa and other developing countries. On the other hand, the Ovambo chickens were established from a representative sample of village chickens found in the Ovambo regions bordering South Africa and Namibia (van Marle-Köster et al., 2008). The actual numbers of individual chickens used to establish the Naked-Neck, Venda, and Ovambo populations are not known. The Potchefstroom Koekoek was established by crossing a number of lines of White Leghorn females and Black Australorp males. The Barred Plymouth Rock was later introduced to the breeding program (Viljoen, 1986) giving this flock a relatively a broader founder population compared to the other flocks.

Whilst it is evident that the conservation flocks diverged from the village chicken populations they were founded from, results from ADMIXTURE indicated that the Potchefstroom Koekoek and Naked Neck have retained single and unique ancestral genomic components from the founder flocks (**Figure 3**) and could be used to conserve part of the genetic diversity found in the village chickens. In contrast, the Venda conservation flock has evolved into a population with a completely different genomic composition to that of the village chickens. Of the four flocks, the Ovambo chickens appear to have maintained much of the village chicken genetic diversity and could therefore be a good and more representative conservation flock.

The village chicken samples were obtained from multiple agro-ecological zones within a country except for the chickens of Malawi that were obtained from a single agro-ecological zone. The clustering of the village chickens followed a geographical gradient in which the South African chickens were least related to the Malawian chickens. Higher levels of divergence between village chickens from Malawi and South Africa could be explained by geographical distance from each other, lack of gene flow between the two countries and isolated evolution occurring in these populations. Village chickens of South Africa and Zimbabwe had more within population diversity, as indicated by their wide spread clusters, than the conservation flocks. This could be due to a combination of founder effects in the conservation flocks as well as gene flow between the two countries.

Linkage disequilibrium was calculated using 28 of the 38 chicken autosomal chromosomes that were represented on the Illumina iSelect SNP60K bead chip. The 10 autosomes not used for this analysis are micro chromosomes that were not included in the design of the 60K bead chip as they were not yet covered by the genome build *Gallus gallus* v2.1 (Groenen et al., 2011). SNPs on linkage groups and sex chromosomes as well as those of unknown marker positions were excluded from the analysis. Most SNPs were pruned due to monomorphism and minor allele frequency. A threshold of MAF $\leq$0.05 was used prior to LD analysis in this and other studies (Qanbari et al., 2010; Wragg et al., 2012) which, according to Corbin et al. (2010), can increase accuracy on LD measures when sample size is large. It was observed by Corbin et al. (2010) and Corbin et al. (2012) that pruning MAF of more than 0.1 can lead to ascertainment bias on the measures of effective population size particularly in small to moderate sample sizes.

The overall LD values between populations showed significant differences between populations with higher LD observed in the conservation flocks and low LD in the village chicken populations kept by smallholder farmers. Variation in LD between the conservation flocks and village chicken populations could be an indication of different population histories and the influences of different evolutionary mechanisms in terms of bottleneck effect, genetic drift, selection and mutations in different population categories. The least diverse (**Table 1**) and highly divergent (**Figures 2**, **3**) Venda flock was also observed to have high LD compared to the other populations which implies low effective

population size and diversity of this population as suggested by the population structure based methods. LD was consistently high on chromosome 16 and was low on chromosome 25 for both village and conservation flocks. The chromosomal difference in LD supports observations by Andreescu et al. (2007), Megens et al. (2009), and Qanbari et al. (2010). However, studies by Andreescu et al. (2007) and Megens et al. (2009) focused on selected genomic regions and selected chromosomes. Findings from the current study and studies by Megens et al. (2009) indicate that evolutionary forces affecting LD act differently on different chromosomes within populations. Natural selection could be a major factor in village chicken populations that are raised under extensive systems characterized by low production levels and minimal human selection pressures (Mtileni et al., 2010).

The current study also indicated a significant LD decay with increased marker intervals, which generally is a function of increased recombination events with increased genetic distance (Megens et al., 2009). The high GC content and high density of genes on micro-chromosomes compared to macro-chromosomes is also associated with high recombination events, which results in lower LD (Megens et al., 2009) and the current results agreed with the expected trends.

Over and above the expected trends in LD decay with increased marker distances, LD was moderately high and remained well above 0.2 at marker distances of up to 500 kb when using genomewide SNP data and upto 1000 kb for macrochromosomes 1–5 in the conservation flocks (**Figures 4A,B**). On the other hand, LD decayed to relatively lower values below 0.1 in the village chicken populations. The relatively high average LD that starts at very short marker distance of 10 kb and is persistent over long distances could be a reflection of low effective population size in the conservation vs. village chicken populations which will be in agreement with results on population structure (**Figures 2**, **3**) and other population diversity analysis (**Table 1**).

Analysis of trends in effective population size from LD values suggested low effective population sizes particularly in the conservation flocks. Results showed a decrease in genetic variation over time in both conservation and village chicken flocks which could be due to poor management, inbreeding as a result of population sub-structuring within villages or population bottlenecks that could have been experienced during the development of these populations (**Figures 5A,B**). The overlapping generations in smallholder farming systems promote mating of closely related chickens thereby increasing inbreeding levels (which were high in conservation flocks, **Table 1**). On the other hand, although village farmers are known to keep small flocks ranging from 1 to 20 chicken per household significant levels of cock sharing is expected within villages which could actually result in higher effective population sizes. Results from microsatellite analyses (Muchadeyi et al., 2007; Mtileni et al., 2011b) have suggested a high level of population diversity within village chicken populations.

Overall, the study demonstrated the utility of the Illumina chicken iselect SNP 60K panel in extensively raised and conservation flocks with limitations due to high proportion of monomorphic and less polymorphic SNPs. Only a subset of independent SNPs could be used for population structure analysis. The study observed population divergence resulting in clear population boundaries between the conservation and the village flocks. High levels of population diversity were observed in the village chickens as well as the Ovambo conservation flock. A relatively high LD that persisted over longer SNP intervals was observed in the South African conservation flocks and not the village chicken populations. This LD pattern seems to be consistent with low effective population sizes and loss of diversity in conservation populations which could be an effect of small size of the founder populations and them being raised as closed populations prone to the effects of inbreeding and genetic drift.

AUTHOR CONTRIBUTIONS

Khulekani S. Khanyile carried out the laboratory analyses, statistical analyses, and interpretation of the data and drafted the manuscript. Farai C. Muchadeyi and Edgar F. Dzomba assisted with the acquisition of funding, designing and execution of the experiment and revised the manuscript critically for important intellectual content.

ACKNOWLEDGMENTS

Genotyping of samples was funded by the Agricultural Research Council-Biotechnology Platform (ARC-BTP) while Mr. K. S. Khanyile held fellowships from the ARC-Professional Development Program, National Research Foundation of South Africa and the University of KwaZulu-Natal postgraduate program.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://www.frontiersin.org/journal/10.3389/fgene.2015.00013/abstract

REFERENCES

Adebambo, A. O., and Consortium, C. D. (2009). Mitochondrial DNA D-loop analysis of South Western Nigerian chicken. *Arch. Zootec.* 58, 637–643. doi: 10.4321/S0004-05922009000400001

Alexander, D. H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19, 1655–1664. doi: 10.1101/gr.094052.109

Andreescu, C., Avendano, S., Brown, S. R., Hassen, A., Lamont, S. J., and Dekkers, J. C. (2007). Linkage disequilibrium in related breeding lines of chickens. *Genetics* 177, 2161–2169. doi: 10.1534/genetics.107.082206

Corbin, L., Blott, S., Swinburne, J., Vaudin, M., Bishop, S., and Woolliams, J. (2010). Linkage disequilibrium and historical effective population size in the Thoroughbred horse. *Anim. Genet.* 41, 8–15. doi: 10.1111/j.1365-2052.2010.02092.x

Corbin, L., Liu, A., Bishop, S., and Woolliams, J. (2012). Estimation of historical effective population size using linkage disequilibria with marker data. *J. Anim. Breed. Genet.* 129, 257–270. doi: 10.1111/j.1439-0388.2012.01003.x

Delany, M. E. (2004). Genetic variants for chick biology research: from breeds to mutants. *Mech. Dev.* 121, 1169–1177. doi: 10.1016/j.mod.2004.05.018

Ensembl. (2013). *Chicken Genome [Online]*. Available online at: http://www.ensembl.org/Gallus_gallus/Location/Genome (Accessed October 3, 2013).

García-Gámez, E., Sahana, G., Gutiérrez-Gil, B., and Arranz, J. (2012). Linkage disequilibrium and inbreeding estimation in Spanish Churra sheep. *BMC Genet.* 13:43 doi: 10.1186/1471-2156-13-43

Godinho, R., Crespo, E. G., and Ferrand, N. (2008). The limits of mtDNA phylogeography: complex patterns of population history in a highly structured Iberian lizard are only revealed by the use of nuclear markers. *Mol. Ecol.* 17, 4670–4683. doi: 10.1111/j.1365-294X.2008.03929.x

Golden Helix Inc. (2014). *SNP and Variation Suite (Version 8.1) [Software](2014)*. Bozeman, MT. Available online at: http://www.goldenhelix.com.

Groenen, M., Megens, H. J., Zare, Y., Warren, W., Hillier, L., Crooijmans, R., et al. (2011). The development and characterization of a 60K SNP chip for chicken. *BMC Genomics* 12:274. doi: 10.1186/1471-2164-12-274

Hall, S. J., and Bradley, D. G. (1995). Conserving livestock breed biodiversity. *Trends Ecol. Evol.* 10, 267–270. doi: 10.1016/0169-5347(95)90005-5

Hedrick, P. W. (2004). *Genetics of Populations.* Canada: Jones and Bartlett.

Hillier, L. W., Miller, W., Birney, E., Warren, W., Hardison, R. C., Ponting, C. P., et al. (2004). Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432, 695–716. doi: 10.1038/nature03154

Jia, X., Chen, S., Zhou, H., Li, W., and Yang, N. (2012). Copy number variations identified in the chicken using a 60K SNP BeadChip. *Anim. Genet.* 44, 276–284. doi: 10.1111/age.12009

Kranis, A., Gheyas, A. A., Boschiero, C., Turner, F., Yu, L., Smith, S., et al. (2013). Development of a high density 600K SNP genotyping array for chicken. *BMC Genomics* 14:59. doi: 10.1186/1471-2164-14-59

Lu, D., Sargolzaei, M., Kelly, M., Changxi, L., Vander Voort, G., Wang, Z., et al. (2012). Linkage disequilibrium in Angus, Charolais, and Crossbred beef cattle. *Front. Genet.* 3:10. doi: 10.3389/fgene.2012.00152

Megens, H. J., Crooijmans, R. P., Bastiaansen, J. W., Kerstens, H. H., Coster, A., Jalving, R., et al. (2009). Comparison of linkage disequilibrium and haplotype diversity on macro-and microchromosomes in chicken. *BMC Genet.* 10:86. doi: 10.1186/1471-2156-10-86

Mobegi, A., Rege, J., Nyamu, A., and Sendalo, D. (2006). "Mitochondrial DNA D-loop sequences reveal the genetic diversity of African chicken," in *The Role of Biotechnology in Animal Agriculture to Address Poverty in Africa: Opportunities and Challenges. Proceedings of the 4th All African Conference on Animal Agriculture and the 31st Annual Meeting of the Tanzania Society for Animal Production (TSAP),* (Arusha: All Africa Society for Animal Production AASAP), 293–298.

Mtileni, B., Muchadeyi, F., Maiwashe, A., Chimonyo, M., Groeneveld, E., Weigend, S., et al. (2011a). Diversity and origin of South African chickens. *Poult. Sci.* 90, 2189–2194. doi: 10.3382/ps.2011-01505

Mtileni, B., Muchadeyi, F., Maiwashe, A., Groeneveld, E., Groeneveld, L., Dzama, K., et al. (2010). "Domestication of South African chicken genetic resources," in *Proceedings of the 9th World Congress on Genetics Applied to Livestock Production* (Liepzig).

Mtileni, B., Muchadeyi, F., Maiwashe, A., Groeneveld, E., Groeneveld, L., Dzama, K., et al. (2011b). Genetic diversity and conservation of South African indigenous chicken populations. *J. Anim. Breed. Genet.* 128, 209–218. doi: 10.1111/j.1439-0388.2010.00891.x

Muchadeyi, F., Eding, H., Wollny, C., Groeneveld, E., Makuza, S., Shamseldin, R., et al. (2007). Absence of population substructuring in Zimbabwe chicken ecotypes inferred using microsatellite analysis. *Anim. Genet.* 38, 332–339. doi: 10.1111/j.1365-2052.2007.01606.x

Mwacharo, J., Bjørnstad, G., Mobegi, V., Nomura, K., Hanada, H., Amano, T., et al. (2011). Mitochondrial DNA reveals multiple introductions of domestic chicken in East Africa. *Mol. Phylogenet. Evol.* 58, 374–382. doi: 10.1016/j.ympev.2010.11.027

Porto-Neto, L. R., Kijas, J. W., and Reverter, A. (2014). The extent of linkage disequilibrium in beef cattle breeds using high-density SNP genotypes. *Genet. Sel. Evol.* 46:22. doi: 10.1186/1297-9686-46-22

Purcell, S. N. B., Todd-Brown, K., Thomas, L., Ferreira, M., Bender, D., Maller, J., et al. (2007). *PLINK: A Toolset for Whole-Genome Association and Population-Basedlinkage Analysis [Online]*. Available online at: http://pngu.mgh.harvard.edu/purcell/plink/ (Accessed April 2013).

Qanbari, S., Hansen, M., Weigend, S., Preisinger, R., and Simianer, H. (2010). Linkage disequilibrium reveals different demographic history in egg laying chickens. *BMC Genet.* 11:103. doi: 10.1186/1471-2156-11-103

Statistical Analysis System. (2011). *SAS/STAT User's Guide: Release 8.1 Edition*. Cary, NC:SAS Institute Inc.

van Marle-Köster, E., Hefer, C. A., Nel, L. H., and Groenen, M. A. M. (2008). Genetic diversity and population structure of locally adapted South African chicken lines: implications for conservation. *South African J. Anim. Sci.* 38, 271–281.

Viljoen, W. C. J. (1986). "Hoenderrasse – pluimvee," in *Boerdery in Suid Afrika, Pamphlet A2 / 1979* (Pretoria: Department of Agriculture and Water Supply).

Wani, C. E., Yousief, I. A., Ibrahim, M. E., and Musa, H. H. (2014). Molecular characterization of sudanese and Southern Sudanese Chicken Breeds using mtDNA D-Loop. *Genet. Res. Int*. 2014:928420. doi: 10.1155/2014/928420

Wragg, D., Mwacharo, J., Alcalde, J., Hocking, P., and Hanotte, O. (2012). Analysis of genome-wide structure, diversity and fine mapping of Mendelian traits in traditional and village chickens. *Heredity* 109, 6–18. doi: 10.1038/hdy.2012.9

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 03 April 2014; accepted: 11 January 2015; published online: 03 February 2015.

Citation: Khanyile KS, Dzomba EF and Muchadeyi FC (2015) Population genetic structure, linkage disequilibrium and effective population size of conserved and extensively raised village chicken populations of Southern Africa. Front. Genet. ***6****:13. doi: 10.3389/fgene.2015.00013*

This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics.

Copyright © 2015 Khanyile, Dzomba and Muchadeyi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in GENETICS

ORIGINAL RESEARCH ARTICLE
published: 22 September 2014
doi: 10.3389/fgene.2014.00333

Genetic diversity and population structure among six cattle breeds in South Africa using a whole genome SNP panel

Sithembile O. Makina[1,2]*, Farai C. Muchadeyi[3], Este van Marle-Köster[2], Michael D. MacNeil[1,4,5] and Azwihangwisi Maiwashe[1,4]

[1] Agricultural Research Council-Animal Production Institute, Irene, South Africa
[2] Department of Animal and Wildlife Sciences, University of Pretoria, Hatfield, South Africa
[3] Agricultural Research Council-Biotechnology Platform, Onderstepoort, South Africa
[4] Department of Animal, Wildlife and Grassland Sciences, University of Free State, Bloemfontein, South Africa
[5] Delta G, Miles City, MT, USA

Edited by:
Johann Sölkner, BOKU -University of Natural Resources and Life Sciences Vienna, Austria

Reviewed by:
Rodolfo Juan Carlos Cantet, Universidad de Buenos Aires, Argentina
Kwan-Suk Kim, Chungbuk National University, South Korea

***Correspondence:**
Sithembile O. Makina, Agricultural Research Council-Animal Production Institute, Private Bag X 2, Irene 0062, South Africa
e-mail: qwabes@arc.agric.za

Information about genetic diversity and population structure among cattle breeds is essential for genetic improvement, understanding of environmental adaptation as well as utilization and conservation of cattle breeds. This study investigated genetic diversity and the population structure among six cattle breeds in South African (SA) including Afrikaner ($n = 44$), Nguni ($n = 54$), Drakensberger ($n = 47$), Bonsmara ($n = 44$), Angus ($n = 31$), and Holstein ($n = 29$). Genetic diversity within cattle breeds was analyzed using three measures of genetic diversity namely allelic richness (A_R), expected heterozygosity (H_e) and inbreeding coefficient (f). Genetic distances between breed pairs were evaluated using Nei's genetic distance. Population structure was assessed using model-based clustering (ADMIXTURE). Results of this study revealed that the allelic richness ranged from 1.88 (Afrikaner) to 1.73 (Nguni). Afrikaner cattle had the lowest level of genetic diversity ($H_e = 0.24$) and the Drakensberger cattle ($H_e = 0.30$) had the highest level of genetic variation among indigenous and locally-developed cattle breeds. The level of inbreeding was lower across the studied cattle breeds. As expected the average genetic distance was the greatest between indigenous cattle breeds and *Bos taurus* cattle breeds but the lowest among indigenous and locally-developed breeds. Model-based clustering revealed some level of admixture among indigenous and locally-developed breeds and supported the clustering of the breeds according to their history of origin. The results of this study provided useful insight regarding genetic structure of SA cattle breeds.

Keywords: South Africa, cattle breeds, genetic resources, genetic diversity, population structure

BACKGROUND

African cattle breeds can be divided into two major categories, namely Taurine cattle (*Bos taurus*) and Indicine cattle (*Bos indicus*). *Bos indicus* is subdivided into zebu proper and zebu crossbred-types and is phenotypically identifiable by the presence of a substantial cerciothoracic hump (Rege, 1999). The position of the hump on the animal's back is used to classify the zebu proper and zebu crossbred types into cervico thoracic-humped and thoracic-humped stocks (Epstein, 1971). Cervico-thoracic-humped cattle occur in or are derived from, contact areas of thoracic-humped Zebu and humpless cattle. In crossbreds of humped and thoracic-humped Zebu cattle, the hump is usually cervico-thoracic and these cattle are referred to as Sanga. However, the Sanga is nowadays considered a separate group of cattle. Thus, African cattle can be classified into four different groups distinguished namely *B. taurus*, *B. indicus*, *Sanga*, and *Sanga' zebu types* (Rege, 1999). Afrikaner and Nguni cattle are classified under the Sanga group and indigenous to South Africa. Drakensberger and Bonsmara cattle are also classified under Sanga types, however, the origin of the Drakensberger cattle is unclear with a history dating back to the early settlers in the late 1700's (Scholtz et al., 2010). The Bonsmara cattle was developed at Mara and Messina Research Station from 1937 to 1963 using Milk Short Horn, Hereford, and Afrikaner cattle with the aim to produce a locally adapted beef breed (Bonsma, 1980). Angus and Holstein belong to *Bos taurus* group and these originate from British and Europe, respectively.

The Afrikaner is one of the oldest breeds with a medium–frame, yellow to red colored with lateral horns with a typical twist. It has exceptional good quality meat and is the ideal minimum care and maximum profit breed (Strydom et al., 2000). Nguni cattle are characterized by their multi-colored coats, which can present many different patterns (white, brown, golden yellow, black, dappled, or spotty), but their noses are always black-tipped and they present a variety of horn shapes. This small framed breed has been kept in rural areas for centuries and often used as dam lines in crossbreeding systems (Scholtz et al., 2011). Drakensberger is a medium to large frame breed and has a black smooth coat. A study by Strydom (2008) has shown that the Drakensberger compare well to British and Europe breeds with regard to meat quality. Bonsmara is medium to large framed, smooth coated with heat and tick tolerance and current the breed

with the largest number of registered females in South Africa (Muchenje et al., 2008).

Bos indicus are known to be adapted to the sub-tropical areas in Africa and have a higher tolerance to various diseases (Muchenje et al., 2008; Marufu et al., 2011). These breeds are also suited to low input systems with lower maintenance and management requirements. In a changing South African environment breeds such as the Afrikaner, Nguni, Drakensberger, and Bonsmara holds potential. Despite their large numbers and not endangered, breeds genetic diversity information is essential for control of inbreeding and effective utilization of breed specific characteristics. The adaptive traits are of importance and there is worldwide a drive for effective management of indigenous genetic resources as they could be most valuable in selection and breeding programs in times of biological stress such as famine, drought, or disease epidemics (FAO, 2010). In order to effectively manage these cattle breeds comprehensive knowledge of their characteristics is required. These include population size and structure as well as knowledge of within and between breeds' divergence (Boettcher et al., 2010; Groeneveld et al., 2010). In South Africa a number of studies have focused on the characterization of small stock such as goats: Visser et al. (2004); sheep: Soma et al. (2012), Qwabe et al. (2012). Limited studies have focused on the genetic characterization of South African cattle breeds and this thus emphasized the need for a genetic characterization of these breeds as genetic resources.

Worldwide genetic markers have been used to assess the genetic variation among many cattle breeds relative to their area of origin (Blott et al., 1998; Hanotte et al., 2002; Gautier et al., 2007; Edea et al., 2013). Results have shown that genetic diversity of breeds is directly linked to their areas of origin, indicating that breeds which have diverged more recently were generally closer together geographically. These studies have also demonstrated larger differences between taurine and indicine breeds due to a greater time since their divergence (McKay et al., 2008; Edea et al., 2013). In addition, significant differences were reported between beef and dairy cattle compared to within beef or dairy; this was attributed to different selection pressure across these contemporary groups (Hayes et al., 2003).

This study therefore investigated genetic diversity and population structure within and between six cattle breeds in South African including Afrikaner, Nguni, Drakensberger, Bonsmara, Angus, and Holstein using genome wide single nucleotide polymorphism (SNP) generated from the Illumina Bovine SNP50BeadChip.

MATERIALS AND METHODS

ANIMAL RESOURCES

A total of 249 animals including three indigenous breeds (Afrikaner = 44), (Nguni = 54), (Drakensberger = 47), one composite (locally-developed) (Bonsmara = 44), and two *Bos taurus* (Angus = 31) and (Holstein = 29) cattle breeds were included in this study. Breeders and Research Stations which keep pure breeds of the populations included in this study were identified and requested to provide animals for blood sampling. All animal handling and sample collection were done according to the regulations of the Animal Ethics Committee of the University of Pretoria (E087-12). To maximize the genetic diversity within each sampled population, pedigree data were used to select against full and half sib animals. **Figure 1** show the map of South Africa indicating the location of farms and research station where populations under study were sampled. The sampling of these animals included collection of 10 ml whole blood using EDTA VACUETTE® tubes. Holstein (48) semen samples were obtained with permission from an artificial insemination company (Taurus, South Africa). However, to maximize the genetic diversity within Holstein samples, identity by descent analysis was performed using data generated from the Bovine SNP50 BeadChip to select the least related bulls. In which a total of 29 least related bulls were selected for the purpose of this study.

GENOTYPING AND QUALITY CONTROL

Genomic DNA was extracted at the ARC-Biotechnology Platform from whole blood and semen samples using the Qiagen DNeasy extraction kit (Qiagen, South Africa) according to the manufacturer's protocol. The protocol was adapted for the semen samples where Dithiothreitol (DTT) was added with proteinase K in the first step. Genomic DNA for all samples was quantified using a Qubit® 2.0 Fluorometer and the Nanodrop Spectrophotometer (Nanodrop ND-1000). In addition, gel electrophoresis was performed to quantify the DNA.

Genotyping was conducted at the ARC-Biotechnology Platform with the Illumina BovineSNP50 BeadChip v2 which features 54,609 SNP probes distributed across the whole bovine genome with an average spacing of 49.9 kb (Matukumalli et al., 2009). Approximately 12 μL of DNA loaded in each well of a BeadChip of genomic DNA was used to genotype each sample. Samples were processed according to the Illumina Infinium–II assay protocol (Illumina, Inc. San Diego, CA, 92122, USA). Quality control criteria were performed across six cattle breeds to remove from further analysis any SNPs with less than 95% call rate, SNPs with less than 0.02 MAF and samples with more than 10% missing genotypes (Purcell et al., 2007). This left about 46,236 SNPs across the breeds. Furthermore, SNPs that were in high LD were pruned using the following parameter; –indep 50 5 2 in plink (Purcell et al., 2007); this left about 21,290 SNPs for further analysis. Pruning of SNPs that are in high LD have been shown to counter the effect of ascertainments bias and to generate meaningful comparison between breeds (Kijas et al., 2009).

ESTIMATES OF WITHIN BREED GENETIC DIVERSITY

Three measures of genetic variability were used to compare the levels of heterogeneity within the cattle breeds (allelic richness, expected heterozygosity, and inbreeding coefficient). Allelic richness (A_R) was determine within each population using ADZE v 1.07 (Szpiech et al., 2008), while expected heterozygosity (H_e) and Inbreeding coefficient (f) was calculated using Plink v1.07 (Purcell et al., 2007) under the default setting.

ANALYSES OF MOLECULAR VARIANCE (AMOVA) AND POPULATION DIFFERENTIATION

Analyses of molecular variance to determine the partition of genetic diversity was first performed among indigenous and

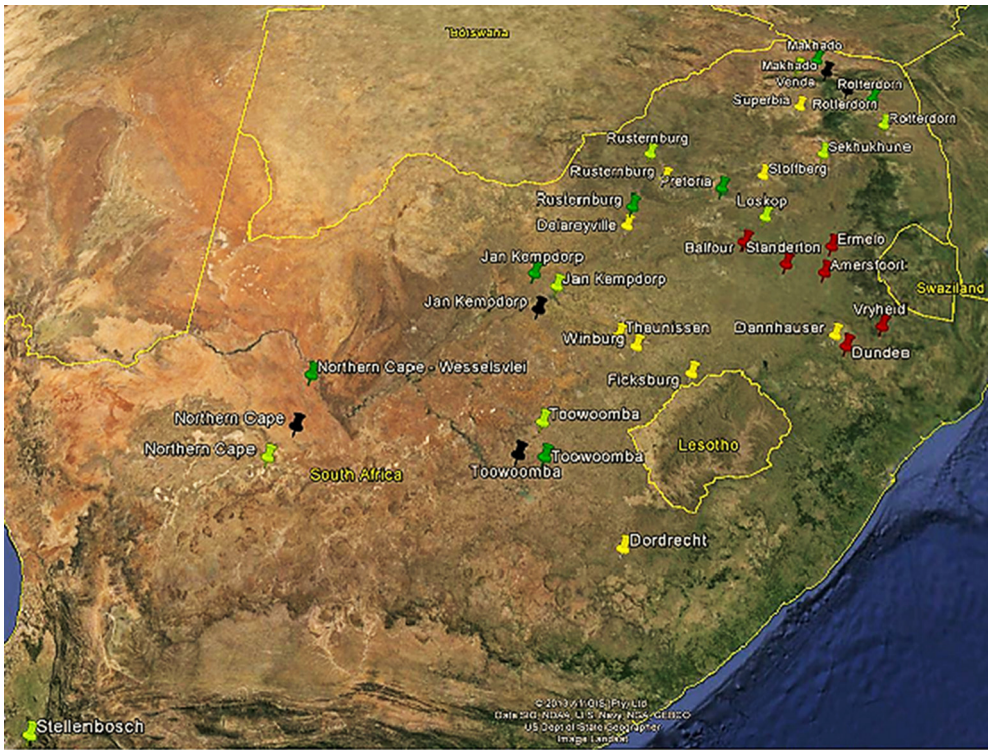


FIGURE 1 | Geographic origin of five cattle breeds in South Africa sampled in the current study. Afrikaner (yellow) (44), Nguni (light green) (56), Drakensbureger (red) (47), Bonsmara (dark green), and Angus (black) (31).

locally-developed cattle breeds and then amongst all six cattle breeds with the program ARLEQUIN 3.1 version (Excoffier et al., 2005).

Populations differentiation was evaluated using pairwise F_{ST} estimates according to Weir and Cockerham (1984) using Golden Helix SNP Variation Suite (SVS) Version 8.1(Golden Helix Inc., Bozeman, MT, 2012).

ALLELE SHARING AND GENETIC DISTANCE

Genetic distance between all pairwise combination of individuals (D) was estimated as one minus the average proportion of allele shared (Purcell et al., 2007) where the average proportion of allele shared was calculated as Dst using Plink v1.07 (Purcell et al., 2007) as:

$$\text{Dst} = \frac{\text{IBS2} + 0.5^{*}\text{IBS1}}{\text{N}}$$

Where IBS1 and IBS2 are the number of loci which are shared either 1 or 2 alleles identical by state (IBS), respectively, and N is the number of loci tested.

Pairwise genetic distance among cattle breeds was estimated based on Nei's (1987) unbiased genetic distance using Phylip v 3.695 genetic software (Felsenstein, 1989), in which a Neighbor-joining (NJ) relationship tree was then constructed using DrawTree application within Phylip v 3.695 software (Felsenstein, 1989).

STRUCTURE ANALYSIS

To investigate the population structure of the studied cattle breeds, ADMIXTURE 1.2.3 Software (Alexander et al., 2009) was used. In order to infer the true number of genetic populations (clusters or K) between the six cattle breeds. Prior population information was ignored before testing and identifying distinct genetic populations, and assigning individuals to populations. ADMIXTURE uses cross validation (CV) procedure to estimate most preferable K. Most preferable K exhibit a low cross-validation error compared to other K-values. In the current study CV error estimates were plotted (**Figure 2**) for comparison of K and $K = 6$ exhibited low cross validation error values thus $K = 6$ was taken as the most probable number of inferred populations.

RESULTS

SNP POLYMORPHISM AND WITHIN BREED GENETIC DIVERSITY

Parameter for SNP validation that included the level of polymorphism, minor allele frequency (MAF) and deviation from Hardy Weinberg equilibrium (HWE) for all six cattle breeds in this study were previously reported (Makina et al., submitted). In summary, examination across breeds revealed that about 56% of SNPs were polymorphic in all breeds and the distribution of MAF showed that nearly half of the SNPs (41%) showed a higher degree of polymorphism (MAF $\geq$ 0.05) across the breeds. With regard to deviation from HWE only between 5 and 6% of SNP were shown to deviate from HWE ($P \leq 0.05$) across the six breeds.

Table 1 presents three measures of within breed diversity across the breeds: Afrikaner cattle had the highest number alleles per locus ($A_R = 1.88$) while the Nguni cattle had the lowest number of alleles per locus ($A_R = 1.73$). However, the Afrikaner cattle was observed to have the lowest level of expected heterozygosity ($H_e = 0.24$) in this study. Among indigenous and locally-developed breeds the Drakensberg cattle ($H_e = 0.30$) had the highest level of genetic diversity. Looking across all six breeds Angus and Holstein cattle had the highest level of gene diversity ($H_e = 0.31$). The level of inbreeding was low across the breeds in this study ranging from 0.004 (Afrikaner) to −0.002 (Drakensberger).

ANALYSES OF MOLECULAR VARIANCE AND POPULATION DIFFERENTIATION

Analysis of Molecular Variance illustrated that within breed genetic variation accounted for 90% among indigenous and locally-developed breeds. On the other hand when indigenous and locally-developed breeds were grouped together with *Bos taurus* cattle 92% of genetic diversity occurred within breeds while only 8% occurred between the breeds (**Table 2**).

Populations differentiation estimates showed that F_{ST} varied from 0.043 (Nguni-Drakensberger) to 0.081 (Afrikaner-Drakensberger) among indigenous and locally-developed breeds and from 0.078 (Drakensberger-Angus) to 0.159 (Afrikaner-Holstein) across all six breeds (**Table 3**).

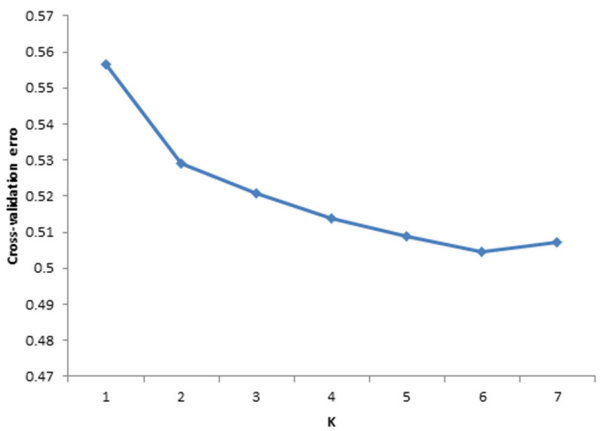


FIGURE 2 | Cross validation plot for six cattle breeds in South Africa. Based on cross validation error the plot indicated that $k = 6$ is optimal for data set.

Table 1 | Sample size and genetic diversity within six cattle breeds in South Africa.

Breed	Code	*n*	A_R (*SD*)	H_e (*SD*)	F_{is}
Afrikaner	AFR	42	1.88 (0.12)	0.24 (0.18)	0.004
Nguni	NGU	54	1.73 (0.11)	0.28 (0.17)	0.005
Drakensberger	DRA	47	1.85 (0.12)	0.30 (0.17)	−0.002
Bonsmara	BON	44	1.84 (0.11)	0.29 (016)	−0.017
Angus	ANG	31	1.80 (0.13)	0.31 (0.16)	−0.012
Holstein	HOL	29	1.81 (0.13)	0.31 (0.18)	−0.026

GENETIC DISTANCE WITHIN AND BETWEEN CATTLE BREEDS

The average genetic distance between individuals drawn from the same breeds was 0.20 ± 0.01 within the Afrikaner cattle, 0.23 ± 0.01 within the Nguni, 0.25 ± 0.01 with the Drakensberger, 0.24 ± 0.01 within the Bonsmara, 0.25 ± 0.02 within the Angus and Holstein 0.25 ± 0.01. The average genetic distance between individuals drawn from different breeds ranged from 0.23 ± 0.005 (Afrikaner-Nguni) to 0.29 ± 0.004 (Angus and Holstein).

Topological relationships between breeds, from Neighbor-Joining tree clearly separated *Bos taurus* breeds (Angus and Holstein) from indigenous and locally-developed cattle breeds (Afrikaner, Nguni, Drakensberger, and Bonsmara) (**Figure 3**). Three main groups were separated: the group formed by Nguni, Drakensberger, and Bonsmara, the group formed by Afrikaner cattle and the group formed by the *Bos taurus* breeds (Angus and Holstein).

Table 2 | Analysis of Molecular Variance among six cattle breeds in South Africa.

Data set	Variance component (%)		
	Among groups	Among populations within group	Within populations
All six cattle breeds	7.80	0.70	91.45
Indigenous and local developed breeds	7.80	1.40	90.80

Table 3 | Wright fixation index (F_{ST}) pair-wise among six cattle breeds in South African.

	Afrikaner	Nguni	Drakensberger	Bonsmara	Angus	Holstein
Afrikaner						
Nguni	0.064					
Drakensberger	0.080	0.044				
Bonsmara	0.071	0.044	0.043			
Angus	0.151	0.108	0.078	0.083		
Holstein	0.159	0.114	0.084	0.099	0.098	

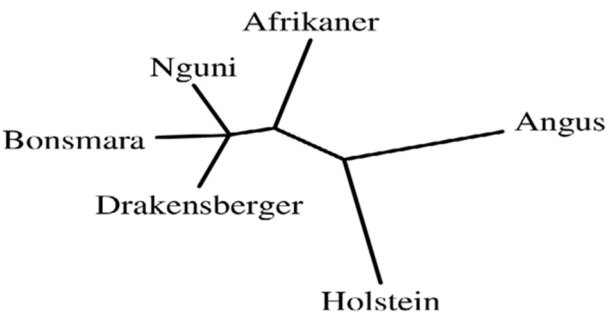


FIGURE 3 | Genetic distances between six cattle breeds in South Africa: Neighbor-joining relationship tree of tested cattle breeds.

POPULATION STRUCTURE ANALYSIS BETWEEN SIX CATTLE BREEDS IN SOUTH AFRICA

The proportions of individuals in each of the breeds in the six most likely clusters inferred by the ADMIXTURE are presented in **Table 4** and this corresponded to the six different breeds included in the study. This revealed that 94% of Afrikaner breed were assigned to cluster one, 84% of Nguni were assigned to cluster two with 8% of its genome assigned to cluster one, 81% of Drakensberger were assigned to cluster three with 5% of its genome assigned to clusters two, four, and five, 89% of Bonsmara were assigned to cluster four with 3% of its genome assigned to cluster two, 93% of Angus were assigned to cluster five and 97% of Holstein were assigned to cluster six. The results presented in **Figure 4** ($k = 6$) demonstrated that among the SA indigenous and locally-developed breeds (Afrikaner, Nguni, Drakensberger, and Bonsmara), the Afrikaner population had the least level of admixture while the Drakensberger had the most level of admixture. The Nguni cattle showed some signals of admixture with Afrikaner breed while the Drakensberger cattle revealed some signals of admixture with Nguni, Bonsmara, and Angus. Bonsmara cattle shared more genetic links with the Nguni cattle than with other indigenous breeds. When comparing all six breeds Afrikaner, Angus, and Holstein populations showed the lowest level of admixture in the current study.

DISCUSSION

Information about genetic diversity and population structure among cattle breeds is essential for genetic improvement, understanding of environmental adaptation as well as utilization and conservation of cattle breeds (Groeneveld et al., 2010). This study investigated the genetic diversity and population structure among six cattle breeds in South Africa. Among indigenous and locally-developed breeds; Drakensberger cattle demonstrated the highest level of genetic variability ($H_e = 0.30$) while the Afrikaner demonstrated the lowest level of genetic diversity. The lower level of genetic variability observed within the Afrikaner cattle could be due to the present of strong selection and use of elite sires which is common among stud and commercial herds and small effective population size. This lower level should be noted in Afrikaner and step toward increasing diversity should be prioritized. This could include exchange of bulls from the different genetic pools. The negative correlation observed between allelic richness and expected heterozygosity in the Afrikaner cattle could be attributed to the processes that differential affect these two measures of diversity, such as bottleneck, selection and increased gene flow between populations within the Afrikaner (Comps et al., 2001).

Angus and Holstein cattle ($H_e = 0.31$) demonstrated the highest level of genetic variability compare to all other breeds. The highest genetic diversity observed in *Bos taurus* breeds were in agreement with the results of Lin et al. (2010) who reported highest genetic variability within *Bos taurus* compared to *Bos indicus* and also to Edea et al. (2013) who reported more genetic diversity in Hanwoo ($H_e = 0.41$) breed than in Ethiopia cattle breeds (between $H_e = 0.37$–0.38) based on SNP data. Heterozygosity values observed in this study were comparable to the previously reported heterozygosity among African ($H_e = 0.25$) and European ($H_e = 0.30$) cattle breeds using SNPs (Gautier et al., 2007). The levels of inbreeding observed in this study were lower across the breeds. However, it should be noted that this may not indicate the real status of inbreeding within these cattle breeds as allele frequencies may be poor estimate of inbreeding. Assessment of the inbreeding level should be done every 5 years to determine any unfavorable change in inbreeding levels, so that appropriate steps could be taken to prevent increases in inbreeding.

Analysis of molecular variance among indigenous and locally-developed breed revealed that about 90% of the genetic variation occurred within the populations. This was lower than the within-population genetic variation (99%) observed among Ethiopia populations by Edea et al. (2013). Combining all six breeds showed that 92% of total variation was within populations. This was higher than 81% observed among Ethiopia and Hanwoo cattle populations.

As expected genetic differentiation (F_{ST}) among the indigenous and locally-developed breeds was lower than African-*Bos taurus* pairs, ranging from 4 to 8%. This was lower than 12% observed among West African cattle breeds by Gautier et al. (2007), but higher than 1% reported among Ethiopian cattle breeds (Edea et al., 2013). Among indigenous and locally-developed and *Bos taurus* cattle breeds genetic differentiation ranged between 8 and 15%; this was comparable to 15% reported between African and European breeds by Gautier et al. (2007) and 17% reported by Edea et al. (2013) among Ethiopia and Hanwoo cattle populations.

The average genetic distance between pairs of animals drawn from the same breeds ranged from 0.20 (Afrikaner) to 0.25 (Angus and Holstein). Average genetic distance between pairs of animal (0.21) was previously reported within 19 cattle breeds (Bovine HapMap Consortium, 2009). As expected average genetic distance between individuals drawn from different breed was

Table 4 | Proportion of membership of the analyzed South African cattle breeds in each of the six clusters inferred in the ADMIXTURE program.

Predefined populations	Inferred clusters						
	1	2	3	4	5	6	*n*
Afrikaner	**0.938**	0.036	0.011	0.006	0.005	0.011	42
Nguni	0.083	**0.838**	0.032	0.032	0.007	0.009	54
Drakensberger	0.032	0.048	**0.806**	0.040	0.045	0.028	47
Bonsmara	0.005	0.030	0.013	**0.887**	0.017	0.006	44
Angus	0.003	0.012	0.005	0.034	**0.932**	0.015	31
Holstein	0.000	0.000	0.008	0.005	0.018	**0.969**	29

Bold indicate inferred cluster.

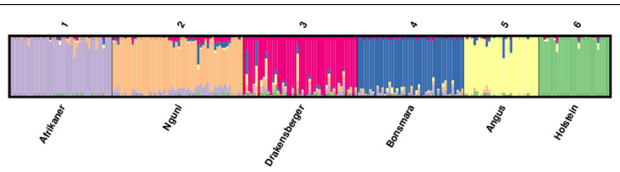


FIGURE 4 | ADMIXTURE clustering of six cattle breeds in South Africa.

higher than those drawn from within breeds, ranging from 0.23 (Nguni-Afrikaner) to 0.29 (Angus-Holstein).

Phylogenic analyses confirmed the closer relationship among indigenous and locally-developed breeds and clearly separated indigenous and locally-developed breeds from *Bos taurus* breeds; this was in agreement with the great divergence between African and European/British breeds observed by Gautier et al. (2007). It will be interesting to expand this breed level analysis in subsequent studies through the inclusion of all SA cattle breeds to better understand genetic relationship among SA cattle breeds.

Population structure analysis revealed some signals of admixture and genetic relationship between Afrikaner, Nguni and Drakensberger and Bonsmara. Nguni cattle shared some genetic links with the Afrikaner cattle, with about 8% of its genome derived from the Afrikaner cattle. This may reflect co-ancestry regarding the origin of these breeds as both these came from the same migration route into the Southern Africa (Scholtz et al., 2011). On the other hand, the Bonsmara cattle shared some genetic links with the Nguni cattle (3%) but only limited genetic links with Afrikaner cattle (0.5%); which was unexpected since the Bonsmara cattle was developed through crossbreeding of Afrikaner cattle with exotic breeds such as Hereford and Milk Shorthorn during the early sixties (Bonsma, 1980). However, it should be noted that when Afrikaner and Nguni cattle were brought to the Southern Africa by the Khoi-Khoi people, Afrikaner cattle migrated along the western side of Southern Africa whilst the Nguni cattle migrated along the eastern side of Southern African (Scholtz et al., 2011), and the Bonsmara cattle was developed in the eastern part of South Africa which predominantly consisted of the Nguni cattle. The observed low relationship between Bonsmara and Afrikaner may also be attributed to genetic drift or small sample size. The Drakensberger cattle was the most admixtured breed in this study with about 5% of its genome derived from the Nguni, Bonsmara and Angus and 3% from Afrikaner and Holstein; this was in agreement with the history of this breed which is believed to have unclear origin (Scholtz, 2010). Afrikaner cattle was the least admixed breed in this study, this was in agreement with the history of this breed as it was the first indigenous South African breed to form a breed society in 1912, thus this breeds may have been closed within the breeding society where only registered animals are allowed within the society. Limited genetic component was shared between indigenous *Bos taurus* breeds, this indicated distinct genetic resources in South African which should be utilization and conservation separately.

In general phylogenetic and population structure analysis revealed distinctiveness among South African (indigenous and locally-developed cattle breeds) and *Bos taurus* cattle breeds which is in agreement with their separate domestication and great time divergence (McKay et al., 2008). The presence of some admixture among South African cattle breeds was in accordance with previous results of genetic diversity studies among cattle breeds that are generally closer together geographically (McKay et al., 2008; Edea et al., 2013). This indicated that the genetic diversity of breeds is directly linked to the areas of origin, suggesting that breeds which have diverged more recently have a generally closer relationship than breeds which diverged long time ago (Maudet et al., 2002).

CONCLUSION

This study revealed low to moderate genetic diversity within six cattle breeds in South Africa and showed a closer relationship among indigenous and locally-developed cattle breeds. Clear genetic divergence between South African (indigenous and locally-developed cattle breeds) and *Bos taurus* cattle breeds was observed which suggested distinct genetic resource in South Africa cattle breeds that should be proper utilization and conservation in order to cope with unpredictable future environments. Information generated from this study forms the basis for future management of these cattle breeds.

AUTHOR CONTRIBUTIONS

Sithembile O. Makina collected the genetic materials, carried out the laboratory analyses, statistical analyses, interpretation of the data and drafted the manuscript. Azwihangwisi Maiwashe and Farai C. Muchadeyi assisted with the acquisition of funding. All authors participated in the design and coordination of the study. Azwihangwisi Maiwashe, Farai C. Muchadeyi, Este van Marle-Köster and Michael D. MacNeil revised the manuscript critically for important intellectual content. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

The authors would like to thank the breeders and research institutions for provision of animals for blood samples. Provision of semen on Holstein bulls by Taurus Co-operative is also acknowledged. ARC-Biotechnology Platform is acknowledged for availing their laboratory resources for genotyping of samples. Financial support from the ARC is greatly appreciated.

REFERENCES

Alexander, D. H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19, 1655–1664. doi: 10.1101/gr.094052.109

Blott, S. C., Williams, J. L., and Haley, C. S. (1998). Genetic relationship among European cattle breeds. *Anim. Genet.* 29, 273–282. doi: 10.1046/j.1365-2052.1998.00327.x

Boettcher, P. J., Tixier-Boichard, M., Toro, M. A., Simianer, H., Eding, H., Gandini, G., et al. (2010). Objectives, criteria, and methods for using molecular genetic data in priority setting for conservation of animal genetic resources. *Anim. Genet.* 41, 64–77. doi: 10.1111/j.1365-2052.2010.02050.x

Bonsma, J. C. (1980). "Cross-breeding, breed creation and the genesis of the Bonsmara," in *Livestock Production: A Global Approach*, ed J. C. Bonsma (Cape Town: Tafelberg Publishers Ltd.), 126–136.

Comps, B., Gömöry, D., Letouzey, J., Thiébaut, B., and Petit, R. J. (2001). Diverging trends between heterozygosity and allelic richness during postglacial colonization in the European beech. *Genetics* 157, 389–397. doi: 10.1126/science.1167936

Edea, Z., Dadi, H., Kim, S., Dessie, T., Lee, T., Kim, H., et al. (2013). Genetic diversity, population structure and relationships in indigenous cattle populations of Ethiopia and Korean Hanwoo breeds using SNP markers. *Front. Genet.* 4:35. doi: 10.3389/fgene.2013.00035

Epstein, H. (1971). *The Origin of the Domestic Animals of Africa,* Vol. 1 *Dog, Cattle, Buffalo*. Revised by I. L. Mason. New York, London, Munich: Africana Publishing Corporation.

Excoffier, L., Laval, G., and Schneider, S. (2005). *Arlequin Version 3.0: An integrated Software Package for Population Genetics Data Analysis*. Switzerland: Computational and Molecular Population Genetic Laboratory (CMPG), Institute of Zoology, University of Berne.

FAO. (2010). *Progress Report of the FAO on Selected Activities Related to Agricultural Biodiversity.* Avilable online at: http://www.fao.org/docrep/meeting/018/k5700e.pdf

Felsenstein, J. (1989). PHYLIP – Phylogeny Inference Package (version 3.6). *Cladistics* 5, 164–166.

Gautier, M., Faraut, T., Moazami-Goudarzi, K., Navratil, V., Foglio, M., Grohs, C., et al. (2007). Genetic and haplotypic structure in 14 European and African cattle breeds. *Genet. Soc. Am.* 177, 1059–1070. doi: 10.1534/genetics.107.075804

Golden Helix Inc. (2012). *SNP & Variation Suite Manual*, *Version 8.1.* Available online at: http://www.goldenhelix.com

Groeneveld, L. F., Lenstra, J. A., Eding, H., Toro, M. A., Scherf, B., Pilling, D., et al. (2010). Genetic diversity in farm animals – a review. *Anim. Genet.* 41, 6–31. doi: 10.1111/j.1365-2052.2010.02038.x

Hanotte, O., Bradley, D. G., Ochieng, J. W., Verjee,Y., Hill, E. W., and Rege, J. E. O. (2002). African pastoralism: genetic imprints of origins and migrations. *Science* 296, 336–339. doi: 10.1126/science.1069878

Hayes, B. J., Visscher, P. M., McPartlan, H. C., and Goddard, M. E. (2003). Novel multilocus measure of linkage disequilibrium to estimate past effective population size. *Genome Res.* 13, 635–643. doi: 10.1101/gr.387103

Kijas, J. W., Townley, D., Dalrymple, B. P., Heaton, M. P., Maddox, J. F., McGrath, A., et al. (2009). A genome wide survey of SNP variation reveals the genetic structure of sheep breeds. *PLoS ONE* 4:e4668. doi: 10.1371/journal.pone.0004668

Lin, B. Z., Sasazaki, S., and Mannen, H. (2010). Genetic diversity and structure in *Bos taurus* and *Bos indicus* populations analyzed by SNP markers. *Anim. Sci. J.* 81, 281–289. doi: 10.1111/j.1740-0929.2010.00744.x

Marufu, M. C., Qokweni, L., Chimonyo, M., and Dzama, K. (2011). Relationships between tick counts and coat characteristics in Nguni and Bonsmara cattle reared on semiarid rangelands inSouth Africa. *Ticks Tick Borne Dis.* 2, 172–177. doi: 10.1016/j.ttbdis.2011.07.001

Matukumalli, L. K., Lawley, C. T., Schnabel, R. D., Taylor, J. F., Allan, M. F., Heaton, M. P., et al. (2009). Development and characterization of a high density SNP genotyping assay for cattle. *PLoS ONE* 4:e5350. doi: 10.1371/journal.pone.0005350

Maudet, C., Luikart, G., and Taberlet, P. (2002). Genetic diversity and assignment tests among seven French cattle breeds based on microsatellite DNA analysis. *J. Anim. Sci*. 80, 942–950.

McKay, S. D., Schnabel, R. D., Murdoch, B. M., Matukumalli, L. K., Aerts, J., Coppieters, W., et al. (2008). An assessment of population structure in eight breeds of cattle using a whole genome SNP panel. *BMC Genet.* 9:37. doi: 10.1186/1471-2156-9-37

Muchenje, V., Dzama, K., Chimonyo, M., Raats, J. G., and Strydom, P. E. (2008). Tick susceptibility and its effects on growth performance and carcass characteristics of Nguni, Bonsmara and Angus steers raised on natural pasture. *Animal* 2, 298–304. doi: 10.1017/S1751731107001036

Nei, M. (1987). *Molecular Evolutionary Genetics*. New York, NY: Columbia University Press.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795

Qwabe, S. O., van Marle-Köster, E., and Visser, C. (2012). Genetic diversity and population structure of the endangered Namaqua Afrikaner sheep. *Trop. Anim. Health Prod.* 45, 511–516. doi: 10.1007/s11250-012-0250-x

Rege, J. E. O. (1999). The state of African cattle genetic resources. I. Classification framework and identification of threatened and extinct breeds. *Anim. Gen. Res. Info. Bull*. 25, 1–25.

Scholtz, M. M. (2010). *Beef Breeding in South Africa*, 2nd Edn. Pretoria: ARC Pretoria.

Scholtz, M., Bosman, D. J., Erasmus, G. J., and Maiwashe, A. (2010). "Selection as the base of improvement in beef cattle," in *Beef Breeding in South Africa*, 2nd Edn., ed M. Scholtz (Pretoria: Agricultural Research Council), 2–10.

Scholtz, M. M., Gertenbach, W., and Hallowell, G. (2011). *The Nguni Breed of Cattle: Past, Present and Future*. Pretoria: Thee Nguni Cattle Breeders' Society.

Soma, P., Kotze, A., Grobler, J. P., and van Wyk, J. B. (2012). South African sheep breeds: population genetic structure and conservation implications. *Small Rumin. Res.* 103, 528–532. doi: 10.1016/j.smallrumres.2011.09.041

Strydom, P. E. (2008). Do indigenous Southern African cattle breeds have the right genetics for commercial production of quality meat? *Meat Sci.* 80, 86–93. doi: 10.1016/j.meatsci.2008.04.017

Strydom, P. E., Naude, R. T., Smith, M. F., Scholtz, M. M., and van Wyk, J. B. (2000). Characterization of indigenous African cattle breeds in relation to carcass characteristics. *J. Anim. Sci.* 70, 241–252.

Szpiech, Z. A., Jackobson, N. A., and Rosenberg, N. A. (2008). A rerafaction approach for counting allele private to combinations of populatons. *Bioinformatics* 24, 2498–2540. doi: 10.1093/bioinformatics/btn478

The Bovine HapMap Consortium. (2009). Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Science* 324, 528–532. doi: 10.1126/science.1167936

Visser, C., Hefer, C. A., van Marle-Köster, E., and Kotze, A. (2004). Genetic variation of three commercial and three indigenous goat populations in South Africa. *J. Anim. Sci.* 34, 24–27.

Weir, B. S., and Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution* 38, 1358–1370. doi: 10.2307/2408641

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 23 May 2014; accepted: 04 September 2014; published online: 22 September 2014.

*Citation: Makina SO, Muchadeyi FC, van Marle-Köster E, MacNeil MD and Maiwashe A (2014) Genetic diversity and population structure among six cattle breeds in South Africa using a whole genome SNP panel. Front. Genet. **5**:333. doi: 10.3389/fgene.2014.00333*

This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics.

Copyright © 2014 Makina, Muchadeyi, van Marle-Köster, MacNeil and Maiwashe. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

THE ROLE OF SOCIAL SCIENCE IN THE MANAGEMENT OF FAnGR

frontiers in GENETICS

PERSPECTIVE ARTICLE
published: 11 February 2015
doi: 10.3389/fgene.2015.00023

Comparing decision-support systems in adopting sustainable intensification criteria

Bouda Vosough Ahmadi[1]*, Dominic Moran[1], Andrew P. Barnes[1] and Philippe V. Baret[2]

[1] Land Economy, Environment and Society Research Group, Scotland's Rural College, Edinburgh, UK
[2] Agronomy, Agroecology, Earth and Life Institute, Université catholique de Louvain, Louvain-la-Neuve, Belgium

Edited by:
Stéphane Joost, Swiss Federal Institute of Technology in Lausanne, Switzerland

Reviewed by:
Maria Wurzinger, BOKU-University of Natural Resources and Life Sciences, Austria
Julie Labatut, National Institute of Agronomic Research, France

***Correspondence:**
Bouda Vosough Ahmadi, Land Economy, Environment and Society Research Group, Scotland's Rural College, King's Buildings, West Mains Road, Edinburgh EH9 3JG, UK
e-mail: bouda.v.ahmadi@sruc.ac.uk

Sustainable intensification (SI) is a multifaceted concept incorporating the ambition to increase or maintain the current level of agricultural yields while reduce negative ecological and environmental impacts. Decision-support systems (DSS) that use integrated analytical methods are often used to support decision making processes in agriculture. However, DSS often consist of set of values, objectives, and assumptions that may be inconsistent or in conflict with merits and objectives of SI. These potential conflicts will have consequences for adoption and up-take of agricultural research, technologies and related policies and regulations such as genetic technology in pursuit of SI. This perspective paper aimed at comparing a number of frequently used socio-economic DSS with respect to their capacity in incorporating various dimensions of SI, and discussing their application to analyzing farm animal genetic resources (FAnGR) policies. The case of FAnGR policies was chosen because of its great potential in delivering merits of SI. It was concluded that flexible DSS, with great integration capacity with various natural and social sciences, are needed to provide guidance on feasibility, practicality, and policy implementation for SI.

Keywords: decision-support systems (DSS), sustainable intensification, farm animal genetic resources, models, social science

INTRODUCTION

The growing human population, and growing global demand for food, are major challenges that will need to be addressed in a world with a potentially dramatically changing climate, and with diminishing natural resources such as farm animal genetic resources (FAnGR; Tilman et al., 2011; Tscharntke et al., 2012). These challenges may require a re-appraisal of the capacity to increase food production, especially livestock products without damaging the important environment and ecosystem services they provide. This is referred to as sustainable intensification (SI) approach by some researchers and politicians (Pretty et al., 2011; Tilman et al., 2011; Garnett et al., 2013) but contested by others (Reed, 2012; Kuyper and Struik, 2014; Loos et al., 2014). SI is a multifaceted concept incorporating the ambition to increase or maintain the current level of agricultural yields while reducing negative ecological and environmental impacts by using a broad range of production methods and technologies and by altering consumption patterns. Four key criteria of SI listed by Godfray and Garnett (2014) include: (i) increase or maintain yield, (ii) reduce or maintain land use, (iii) reduce negative environmental and ecological externalities, and (iv) consider/use all forms of agriculture without prejudice. To achieve the objectives of SI by implementing these four main criteria, agricultural, ecological and environmental policies and regulations need to be adjusted accordingly. Policy decisions are often supported and informed by the results of scientific and socio-economic research. Decision-support systems (DSS) are considered as set of scientific and analytical tools and approaches that are used in interpreting research results into policy relevant outcomes.

Decision-support systems can be used to assist agricultural systems' players and policy makers to achieve objectives of SI by incorporating these four criteria and their subsets in such analytical frameworks. DSS are often used at farm level in informing farmers to improve plans and decisions. They are also used to assist policy makers to evaluate and *ex ante* assessment of future policies. Results of DSS provide an optimum plan of action that can be applied to enterprise, farm, regional, national, or global levels (Geels and Schot, 2007). At farm level, in addition to bio-physical variations of farms (i.e., technical characteristics), goals, and perception of farmers about their farming and agri-ecological systems, as well as their risk attitudes (i.e., social characteristics) vary considerably. Traditionally bio-physical and technical characteristics including technical coefficients, representing specific production functions, are included in certain DSS in form of constraints and activity requirements. However, inclusions of social characteristics and agri-ecological/environmental externalities of farming practices, farmers' perceptions, behaviors and attitudes in these frameworks are proved to be challenging and less comprehensive (Vanwindekens et al., 2013).

Another challenge in developing and using DSS relevant to policy analysis is inclusion of public and private goods characteristics. Agriculture is inherently multifunctional and often includes both private and public good such as producing food,

fiber, etc., with having a profound impact on economies, ecosystems, and environment (Pretty et al., 2001). Farming practices are considered as business activities that generate products and income for farmers (private good) but at the same time could generate positive (e.g., ecosystem services) and negative externalities affecting environmental and ecological systems (public good). Estimating financial performance of farming practices is relatively easy and is routinely done at farm or sector levels using budgeting techniques. However, incorporating agri-ecological/environmental costs and benefits of farming practices (i.e., public good element) is challenging and require getting support from other methodologies. For example the total economic value (TEV) approach is often used to capture these costs and benefits. Direct use value, indirect use value, option value, bequest value, and existence value are components of TEV (Pearce and Moran, 1994). Some of these components also provide a mixture of private and public goods. Other approaches that could support and inform DSS in assessing agri-ecological costs and benefits are: empirical approaches, willingness to pay, contingent valuation, hedonic pricing, and use of experimental data (Randall, 2002). The capability and capacity of DSS to adopt these approaches and capture public/private values vary substantially and therefore both developer and end-user of the results of DSS need to be aware of these differences. In addition, reducing negative ecological and environmental externalities is an important criterion of SI. To go beyond this, "positive" externalities such as ecosystem services could be integrated in DSS.

To incorporate SI's criteria including social, ecological, and environmental externalities in DSS that ultimately enhance agricultural policies, greater integration of social and technical aspects of farming practices is needed. A wide range of DSS have been developed and applied to different production and agricultural system (Janssen and van Ittersum, 2007). The objectives of this paper are to revisit and compare the capacity of six widely used DSS in adopting the four key SI criteria, agro-ecological/environmental externalities and socio-technical aspects of farming practices, and to discuss the application of DSS to analyzing policies related to conservation of FAnGR.

REVIEW OF DSS

Agricultural systems and practices are studied using both sociological (anthropological) science methods and technical/engineering sciences. DSS applied to agricultural systems often use statistical and mathematical modeling techniques and are classified based on their purpose, methodology, and assumptions. On this basis, DSS are classified under four main categories: empirical, mechanistic, positive, and normative (Hazel and Norton, 1986). Empirical models are built using observed data and aiming to discover relationships that were not expected *ex ante*. Mechanistic models are built on existing scientific theory and knowledge and are mainly used for *ex ante* scenario analysis (Janssen and van Ittersum, 2007). DSS could be developed using either positive or normative approach. Positive approach tries to mimic the actual behavior of the farmers or managers whereas normative approach tries to find optimum solution for a given system.

COMPARISONS

Six DSS approaches namely: structural equation modeling (SEM), linear- and non-linear programming (LP and NLP), positive mathematical programming (PMP), multiple criteria decision making (MCDM), cognitive mapping (CM), and dynamic programming (DP) were selected for comparison in this study. **Table 1** summarizes the characteristics of the mentioned DSS approaches. Among the selected DSS approaches, SEM and CM are considered as empirical methods that are mainly used in *ex post* analysis aiming at revealing relationships in observed data that will be used to predict outcomes in future. Technical aspects of farming practices could be included to some extend in these methods but less than mechanistic models. Both SEM and CM are strong in looking at social aspects of socio-ecological farming systems including farmers, behavior, perceptions, and goals. These social attributes could be related to ecological and environmental issues and therefore could provide useful insight. Considering mentioned characteristics the potential of these methods in assisting with SI merits are judged to be moderate to high.

LP/NLP, PMP, MCDM, and DP are considered as mechanical approaches that are built based on theory and knowledge to find solutions for management problems in relation to farming systems by running the models under different scenarios and policy assumptions. LP/NLP and PMP have good capacity in incorporating technical aspects of the systems (including economics, production, environment, and ecology/biodiversity; Cypris, 2000; Vosough Ahmadi et al., 2011, 2015; Stott et al., 2012). However, they have fairly limited capacity in covering social and behavioral characteristics of the farmers. MCDM approaches are less sophisticated in terms of the level of technical details of the systems but could potentially include different goals of various stakeholders or certain view points (i.e., goals). Social and technical aspects of environmental and ecological issues could be covered to some extend by these approaches. DP is an example of DSS approaches that could assist farm managers with decisions within a short time (Kennedy, 1986). They are capable to incorporate risk and stochastic events but relatively low complexity of the system could be built in these models (Stott, 1994). They are not usually capable of inclusion of social aspects and not very strong in incorporating environmental and biodiversity elements. In terms of capability of inclusion of SI goals and merits, in our judgment LP/NLP and CM are considered as approaches with a very high capacity. After these methods, MCDM are considered as highly capable in incorporating and helping with SI concept. SEM and DP are considered as moderate in terms of their capability of adopting SI criteria.

INTEGRATING SI CRITERIA IN DSS

In the following lines the application and usefulness of the mentioned DSS approaches in relation to SI's four criteria definition by Godfray and Garnett (2014), is discussed.

(S1) Increasing/maintaining yield (intensification aspect): This criterion is related to utilizing technologies and also to some extent to improving management of crops and livestock (e.g., controlling diseases) that leads to higher yield. DSS could help with informing decision makers with control of diseases, and short

Table 1 | Comparisons of selected DSS including characteristics and their capacity of integrating SI criteria and socioeconomic aspects.

Method	Descriptive						Capacity of integrating*						
							Sustainable intensification criteria**				Positive externalities (e.g., ecosystem services)	Social aspects (goals)	Attitude (e.g., risk)
	Type	Level	Time aspect	Model inputs	Objective and assumptions	Technical	S1	S2	S3	S4			
SEM	Empirical	Farm	Retrospect (*ex post*)	Independent and explanatory variables	Explore relationships in observed data	No	Yes	No	Yes	Yes	Yes (M)	Yes	Yes
LP and NLP	Mechanical/ normative or positive	Animal/ crop/ farm	Season/year/ static/dynamic/ (*ex ante*)	Activities, constraints, coefficients	Maximizing or minimizing objectives. Linear/non-linear relationships	Yes	Yes (L)	Yes	Yes	Yes	Yes (VH)	Ye (L)	Ye (L)
PMP	Mechanical/ positive	Animal/ crop/farm	Season/year/ static/dynamic/ (*ex ante*)	Activities, constraints, coefficients	Mimic actual farm practice. Linear/non-linear relationships	Yes (L)	Yes (L)	Yes	Yes	Yes	Yes (H)	No	Yes (L)
MCDM	Mechanical	Farm/ region/ country	Season/year/ static/dynamic/ (*ex ante*)	Goals as objectives, constraints	One goal optimized, constraints on others get tighter	Yes (L)	Yes	Yes	Yes	Yes	Yes (H)	Yes (H)	Yes (H)
CM	Empirical	Individual (farm)/ social	Year (*ex post*)	Quantitative, operations, drivers, constraints	Incorporate social aspect in DSS, farmer	Yes (L)	Yes	Yes (L)	Yes (L)	Yes (L)	Yes (VH)	Ye (VH)	Yes (VH)
DP	Mechanical/ normative	Animal/ crop/farm	Year/dynamic (*ex ante*)	Pay-off values, transition probabilities	Profit maximizing	Yes (L)	No	Yes	Yes (VL)	Yes (VL)	Yes (M)	No	Yes (L)

**Capacity of integrating was specified by "yes" or "no". The following scale was used for the "yes" category: VL, very limited; L, limited; M, moderate; H, high; VH, very high.*

***S1, increasing/maintaining yield; S2, using less land or maintaining current land usage; S3, less environmental and ecological damage, more biodiversity and ecosystem services; S4, utilizing all types of technology without prejudice.*

SEM, structural equation modeling; LP, linear programming; NLP, non-linear programming; PMP, positive mathematical programming; MCDM, multiple criteria decision making; CM, cognitive mapping; DP, dynamic programming.

and long term optimum management for example in relation to keeping/replacement of animals or crop rotation etc.

(S2) Using less land or maintaining current land usage (intensification aspect): DSS and in particular mechanistic models could provide insight on the impact of reducing available land on production and could suggest alternative solutions if technology allows.

(S3) Less environmental and ecological damage, more biodiversity and ecosystem services (sustainability aspect): This condition could potentially be included at both technical and social levels in DSS models. However, in majority of the available models environmental and ecological aspects have been added as constraints to the systems whereas it could be considered as objective of the farming in these models.

(S4) Utilizing all types of technology without prejudice (both intensification and sustainability elements): There is an on-going debate about this criterion of SI (Loos et al., 2014). In DSS approaches such as CM and MCDM, the perceptions and goals of farmers with respect to using particular technologies to improve/increase yield or to protect environment and biodiversity could be analyzed and included in the models. In this case individual and social believes/perceptions of farmers that are added to model will assist policy makers to come up with effective policies.

All the four mentioned SI criteria could be considered as objectives and opportunity or could be as constraints of the agricultural systems in DSS models. Similarly they are influenced by short and long term goals, and perceptions and behavior of farmers. These criteria are also directly related to technological advances that help with increase/maintain yield but lowering negative externalities and also by increasing efficiency.

APPLICATION TO FAnGR POLICIES

In the context of FAnGR conservation and biodiversity policies, the issue of allocation of limited preserving genetic diversity budget in determining actual conservation priorities among endangered species has been included in a number of theoretical and operational DSS by a number of authors (Weitzman, 1998; Naidoo and Iwamura, 2007). In most of these DSS, objective was to preserve maximum diversity given the limited financial, technological, and perhaps logistical resources. Probability of extinction has been a core element of conservation DSS modeling. In addition, discounting future benefits and costs as a basis for economic justification of conservation policies has been taken into account. More recent applications of DSS to FAnGR context showed that supply chain management, cooperation, management of common goods in relation to biological resources and data management are important elements that need to be considered in developing and using DSS (Labatut et al., 2010, 2012). A number of other areas for consideration are: the goals of conservation, intrinsic value of breeds, public and private good elements of FAnGR, the impact of genetic technologies on society and power in the breeding systems. Also the impact of demand of agricultural product and services on commodity market prices at farm level that are not usually explicitly included in DSS models must be incorporated in the models subject to data availability.

Means to SI in livestock production or in other words means to improve sustainability and productivity of farm animals need to be sought through breeding, genetic engineering, nutrition, health, and welfare. For example new phenotypes linked to sustainable animal productivity could be developed and integrated into breeding schemes. SI's merits could also be achieved through economically justified conservation of FAnGR that depends on the increased adaptive capacity in response to change that such preservation in a genome resource bank offers beyond that of alternatives. The technological preservation of FAnGR does not only require economic and scientific input, to direct optimal decision making, but social science methods to reflect the historical, cultural, and social aspects of genetic resources at farm level and beyond. DSS therefore play a crucial role in integrating both technical and social aspects of farming practices and provide an improved policy and practical guidance to tackle major global challenges ahead.

ACKNOWLEDGMENTS

The authors acknowledge the European Science Foundation and the Scottish Government for funding this work. This research was undertaken within the European Science Foundation exchange grant number 4576 and the Scottish Government Rural Affairs and the Environment Portfolio Strategic Research Programme 2011–2016, Theme 4: Economic Adaptation.

REFERENCES

Cypris, C. (2000). *Positive Mathematical Programming (PMP) in the Agrarian Sector Model RAUMIS*. Bonn: FAA.

Garnett, T., Appleby, M. C., Balmford, A., Bateman, I. J., Benton, T. G., Bloomer, P., et al. (2013). Sustainable intensification in agriculture: premises and policies. *Science* 341, 33–34. doi: 10.1126/science.1234485

Geels, F. W., and Schot, J. W. (2007). Typology of sociotechnical transition pathways. *Res. Policy* 36, 399–417. doi: 10.1016/j.respol.2007.01.003

Godfray, H. C. J., and Garnett, T. (2014). Food security and sustainable intensification. *Philos. Trans. R. Soc. B Biol. Sci.* 369, 20120273. doi: 10.1098/rstb.2012.0273

Hazel, P. B., and Norton, R. D. (1986). *Mathematical Programming for Economic Analysis in Agriculture*. New York, NY: Macmillan.

Janssen, S. J. C., and van Ittersum, M. K. (2007). Assessing farm innovations and responses to policies: a review of bio-economic farm models. *Agric. Syst.* 94, 622–636. doi: 10.1016/j.agsy.2007.03.001

Kennedy, J.O.S. (1986). *Dynamic Programming: Application to Agriculture and Natural Resources*. London: Elsevier.

Kuyper, T. W., and Struik, P. C. (2014). Epilogue: global food security, rhetoric, and the sustainable intensification debate. *Curr. Opin. Environ. Sustain.* 8, 71–79. doi: 10.1016/j.cosust.2014.09.004

Labatut, J., Aggeri, F., and Girard, N. (2012). Discipline and change: how technologies and organizational routines interact in new practice creation. *Organ. Stud.* 33, 39–69. doi: 10.1177/0170840611430589

Labatut, J., Girard, N., Astruc, J. M., Barillet, F., Bibe, B., Soulas, C., et al. (2010). "Renewing collaborative design in the management of animal genetic resources," in *Proceedings of the IFSA Symposium*, Vienna.

Loos, J., Abson, D. J., Chappell, M. J., Hanspach, J., Mikulcak, F., Tichit, M., et al. (2014). Putting meaning back into "sustainable intensification." *Front. Ecol. Environ.* 12, 356–361. doi: 10.1890/130157

Naidoo, R., and Iwamura, T. (2007). Global-scale mapping of economic benefits from agricultural lands: implications for conservation priorities. *Biol. Conserv.* 140, 40–49. doi: 10.1016/j.biocon.2007.07.025

Pearce, D. W., and Moran, D. (eds). (1994). *The Economic Value of Biodiversity*. London: Earthscan.

Pretty, J., Brett, C., Gee, D., Hine, R., Mason, C., Morison, J., et al. (2001). Policy challenges and priorities for internalizing the externalities of modern agriculture. *J. Environ. Plan. Manag.* 44, 263–283. doi: 10.1080/09640560123782

Pretty, J., Toulmin, C., and Williams, S. (2011). Sustainable intensification in African agriculture. *Int. J. Agric. Sustain.* 9, 5–24. doi: 10.3763/ijas.2010.0583

Randall, A. (2002). Valuing the outputs of multifunctional agriculture. *Eur. Rev. Agric. Econ.* 29, 289–307. doi: 10.1093/eurrag/29.3.289

Reed, M. (2012). "Contesting 'sustainable intensification' in the UK: the emerging organic discourse," in *Organic Food and Agriculture—New Trends and Developments in the Social Sciences*, ed. M. Reed (Croatia: InTech), 131–144.

Stott, A. W. (1994). The economic advantage of longevity in the dairy cow. *J. Agric. Econ.* 45, 113–122. doi: 10.1111/j.1477-9552.1994.tb00382.x

Stott, A. W., Vosough Ahmadi, B., Dwyer, C. M., Kupiec, B., Morgan-Davies, C., Milne, C. E., et al. (2012). Interactions between profit and welfare on extensive sheep farms. *Anim. Welf.* 21, 57–64. doi: 10.7120/096272812X13345905673683

Tilman, D., Balzer, C., Hill, J., and Befort, B. L. (2011). Global food demand and the sustainable intensification of agriculture. *Proc. Natl. Acad. Sci. U.S.A.* 108, 20260–20264. doi: 10.1073/pnas.1116437108

Tscharntke, T., Clough, Y., Wanger, T. C., Jackson, L., Motzke, I., Perfecto, I., et al. (2012). Global food security, biodiversity conservation and the future of agricultural intensification. *Biol. Conserv.* 151, 53–59. doi: 10.1016/j.biocon.2012.01.068

Vanwindekens, F. M., Stilmant, D., and Baret, P. V. (2013). Development of a broadened cognitive mapping approach for analysing systems of practices in social–ecological systems. *Ecol. Model.* 250, 352–362. doi: 10.1016/j.ecolmodel.2012.11.023

Vosough Ahmadi, B., Shrestha, S., Thomson, S. G., Barnes, A. P., and Stott, A. W. (2015). Impacts of greening measures and flat rate regional payments of the Common Agricultural Policy on Scottish beef and sheep farms. *J. Agric. Sci.* doi: 10.1017/S0021859614001221 [Epub ahead of print].

Vosough Ahmadi, B., Stott, A. W., Baxter, E. M., Lawrence, A. B., and Edwards, S. A. (2011). Animal welfare and economic optimisation of farrowing systems. *Anim. Welf.* 20, 57–67.

Weitzman, M. L. (1998). The Noah's ark problem. *Econometrica* 6, 1279–1298.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 31 October 2014; accepted: 16 January 2015; published online: 11 February 2015.

Citation: Vosough Ahmadi B, Moran D, Barnes AP and Baret PV (2015) Comparing decision-support systems in adopting sustainable intensification criteria. Front. Genet. ***6****:23. doi: 10.3389/fgene.2015.00023*

This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics.

Copyright © 2015 Vosough Ahmadi, Moran, Barnes and Baret. This is an open-access article distributed under the terms of the **Creative Commons Attribution License** *(CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*

frontiers in GENETICS

ORIGINAL RESEARCH ARTICLE
published: 18 December 2014
doi: 10.3389/fgene.2014.00440

Citizens' preferences for the conservation of agricultural genetic resources

Eija Pouta*, Annika Tienhaara and Heini Ahtiainen

MTT Agrifood Research Finland, Helsinki, Finland

Edited by:
Juha Kantanen, MTT Agrifood Research Finland, Finland

Reviewed by:
Jutta Roosen, Technische Universität München, Germany
Ann Bruce, University of Edinburgh, UK

****Correspondence:***
Eija Pouta, MTT Agrifood Research Finland, Latokartanonkaari 9, FI-00790 Helsinki, Finland
e-mail: eija.pouta@mtt.fi

Evaluation of conservation policies for agricultural genetic resources (AgGR) requires information on the use and non-use values of plant varieties and animal breeds, as well as on the preferences for *in situ* and *ex situ* conservation. We conducted a choice experiment to estimate citizens' willingness to pay (WTP) for AgGR conservation programmes in Finland, and used a latent class model to identify heterogeneity in preferences among respondent groups. The findings indicate that citizens have a high interest in the conservation of native breeds and varieties, but also reveal the presence of preference heterogeneity. Five respondent groups could be identified based on latent class modeling: one implying lexicographic preferences, two with reasoned choices, one indicating uncertain support and one with a preference for the current status of conservation. The results emphasize the importance of *in situ* conservation of native cattle breeds and plant varieties in developing conservation policies.

Keywords: native breeds, native varieties, genetic resources, choice experiment, preference heterogeneity, valuation

INTRODUCTION

The intensification of agriculture has led to marked changes in the utilization of agricultural genetic resources (AgGR), and many previously common cultivated plant varieties as well as native animal breeds that are of interest in terms of food and agricultural production have become rare or even endangered (Drucker et al., 2001; FAO, 2007, 2010). In Finland, several native breeds, such as the Eastern and Northern Finncattle, the Kainuu Gray Sheep and the Åland Sheep, are endangered according to the FAO classification (FAO, 2007), and the majority of old Finnish crop varieties as well as the Finnish landrace pig are already extinct.

Decisions on the focus and extent of genetic resource conservation should consider both the costs and benefits of conservation. The full benefits of conserving AgGR are not revealed by markets, as the resources are either not traded in the markets or the price of agricultural products does not completely capture their value (Oldfield, 1989; Brown, 1990; Drucker et al., 2001). These market failures result in an inefficient allocation of resources, i.e., the level of conservation is too low as the full benefits are not considered. Although the importance of economic analyses has been recognized, the literature on the monetary value of genetic resources in agriculture is still relatively limited (e.g., Evenson et al., 1998; Rege and Gibson, 2003; Ahtiainen and Pouta, 2011).

Conservation policies for AgGR in Finland, as in many other European countries, are currently based on international agreements such as the Convention on Biological Diversity (1992) and the Global Plan of Action for Animal Genetic Resources (FAO, 2007). National genetic resource programmes were initiated for plants in the year 2003 and for farm animals in 2005 to strengthen the conservation of genetic resources in Finland. Although there has been some progress in putting the programmes into action, they have not been fully implemented. This may reflect, for example, the lack of political interest in the conservation.

To evaluate conservation policies, there is a need for monetary benefit estimates that encompass both use and non-use values associated with genetic resources. Use values refer to the benefits obtained from current and future use of genetic resources in production and breeding, while non-use values are generated from the knowledge that genetic resources, e.g., certain breeds, exist and are saved for future generations. Stated preference methods, such as the discrete choice experiment (CE) method, are capable of estimating both use and non-use values in monetary terms. A choice experiment is a survey-based method whereby respondents are asked to choose between two or more discrete alternatives that are described with attributes. By varying attribute levels and including a price variable as one of the attributes, respondents' willingness to pay (WTP) for a policy alternative or attribute level is indirectly revealed based on the choices they make (e.g., Hanley et al., 2001). The CE method has been found suitable for valuing genetic resources due to its flexibility and ability to value the different traits that breeds or varieties may have. The CE method can also be used to evaluate the means of conservation *in situ* (live animals and plants) and *ex situ* (as seeds, cryopreserved embryos and other genetic material), and both plant genetic resources (PGR) and animal genetic resources (AnGR).

Previous choice experiments have focused on valuing breeds or varieties and their attributes, especially related to their use in agriculture (Birol et al., 2006; Ouma et al., 2007), and applications focusing on consumer or citizen values for AgGR are rare. Valuation studies on biodiversity have found heterogeneity in consumer preferences, and even identified lexicographic preferences toward conservation (Hanley et al., 1995; Sælensminde,

2006). Lexicographic preferences imply that people are unwilling to accept any trade-offs for changes in environmental goods, such as biodiversity, and may arise when an individual believes that the environment should be protected without regard to the costs. In the context of AgGR, preference heterogeneity has mainly been studied among farmers (e.g., Ouma et al., 2007; Omondi et al., 2008; Roessler et al., 2008), and there have been only few empirical studies of heterogeneity of citizen preferences (Zander et al., 2013) or lexicographic preferences.

In this paper, we present the results of a choice experiment conducted to estimate the benefits of genetic resource conservation programmes in Finland. We tested the effect of *in situ* and *ex situ* conservation on citizens' choices between programmes. We also analyzed whether plant varieties and animal breeds are perceived as equally valuable by citizens. As heterogeneity in the preferences for the conservation of AgGR is likely, we tested for the existence of citizen segments that place different values on the conservation of genetic resources.

We expected that AgGR would be rather unfamiliar to some of the respondents of the valuation survey. However, in valuation surveys, respondents are assumed to make "informed" choices when responding to value elicitation questions (e.g., Blomquist and Whitehead, 1998). To obtain informed choices that produce valid estimates of WTP, surveys need to provide a sufficient amount of neutral information on the environmental good while avoiding information overload. Providing more information on the quality (characteristics and services) of an environmental good can increase the stated WTP, have no effect, or in some cases reduce WTP (Blomquist and Whitehead, 1998).

There is a substantial body of literature on the effects of information and respondent effort in contingent valuation studies (e.g., Cameron and Englin, 1997; Blomquist and Whitehead, 1998; Berrens et al., 2004), and some choice experiment studies have also examined the issue, mainly focusing on respondent effort (Hu et al., 2009; Vista et al., 2009). Hu et al. (2009) used data from a choice experiment concerning genetically modified food to simultaneously model voluntary information access and product choices. They found that information was accessed rather infrequently, and that those who held critical views on GM food accessed information more often. There were interlinkages between information access and choices, but they were complex and varied between individuals. Vista et al. (2009) examined the effect of time spent on attribute information, choice questions or completing the survey, finding no significant effects on parameter estimates.

Here, we were particularly interested in examining how the use of information differs between respondent segments. In the survey, respondents had the opportunity to obtain additional information on genetic resources by accessing a hyperlink to a web page. The Internet survey allowed us to measure whether the respondents accessed the additional information and how much time they used to read it. Offering the opportunity for voluntary access to information instead of using different information treatments for split samples has the advantage of not assuming that respondents read all the information that is provided (Hu et al., 2009). Furthermore, we tested the effects of response certainty and self-perceived carefulness in filling the survey as sources of preference heterogeneity.

The rest of the paper is organized as follows. Section Materials and Methods introduces the data and statistical models used in the analysis. Results are presented in section Results, and section Discussion and Conclusions provides discussion and conclusions.

MATERIALS AND METHODS

DATA COLLECTION

The survey data were collected using an Internet survey during the summer of 2011. The sample was drawn from the Internet panel of a private survey company, Taloustutkimus, which comprises 30,000 respondents who have been recruited to the panel using random sampling to represent the population (Taloustutkimus, 2013). After a pilot survey of 138 people, a random sample of 6200 respondents was selected, of which 2426 completed parts of the survey and 1495 completed the survey entirely. These numbers correspond to response rates of 39 and 24%, respectively. Based on the socio-demographic variables, the data represented the population rather well (**Table 1**).

SURVEY DESIGN

In the first section, the survey introduced the most common Finnish native animal breeds and plant varieties by explaining what landraces are and giving examples. After asking the respondents about their familiarity with PGR and AnGR, all respondents were offered a short piece of information on the conservation of these breeds and varieties. Next, the respondents were given the opportunity to obtain further information by clicking on two hyperlinks, one for PGR and the other for AnGR. Providing voluntary access to additional information made it possible to identify those respondents who accessed the information, and the time spent on the information page was also recorded (Hu et al., 2009). The additional information provided in our survey included motives for conservation, descriptions of the *in situ* and *ex situ* conservation methods and facts about the sustainable use of genetic resources. After several questions concerning perceptions of genetic resources, the survey proceeded to the choice experiment.

Table 1 | Descriptive statistics ($n = 1608$).

	In the data	In the population[a]
Proportion of females, %	48	51
Mean age, years	52	47
Proportion of people with a higher educational level, %	24	23
Proportion of people living in households with a gross income under €40,000, %	43	53
Proportion of people with children (<18 years) in the family, %	35	40
Proportion of people living in South Finland, %	40	41

[a] *Statistics Finland 2010, www.stat.fi.*

The choice experiment was framed by telling respondents that the conservation of native plant varieties and animal breeds is not yet comprehensive in Finland. The survey presented a programme that would conserve the majority of the varieties and breeds on farms and in gene banks. The operation of gene banks would be extended to missing plants and varieties, and conservation on farms would be enhanced by developing the support provided to farmers for conservation activities. Furthermore, those who are using native varieties in gardens were stated to be supported monetarily and by providing guidance.

The survey explained that the conservation programme would be financed with an increase in income tax between the years 2012 and 2021, and that depending on the extent of the programme, the cost to taxpayers would vary, but all taxpayers would participate in financing the programme. The conservation measures (attributes) of the alternative programmes were illustrated to the respondents using a table.

Table 2 presents the attributes together with their descriptions and levels. The first attribute level is always the level specified in the status quo alternative (current state). The attributes included conservation measures of both plant varieties and animal breeds in gene banks and farms. Instead of having a separate attribute for each native breed, only one attribute for breeds in gene banks and one on farms was included to have the same number of attributes for varieties and breeds, and to ease the cognitive burden of the respondents. The native breeds in gene banks attribute had eight levels and native breeds on farms nine levels, including the status quo attribute level.

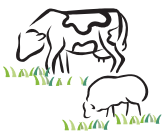
After introducing the attributes, the respondents were presented with six choice tasks. Each choice task included three alternatives: the status quo alternative, described as maintaining the current situation, and two policy alternatives describing an improved level of conservation compared to the current level. Each alternative was described with five conservation attributes, their levels and the cost attribute. The status quo alternative was uniform across choice tasks. An example of a choice task is shown in **Table 3**.

We employed an efficient experimental design to allocate the attribute levels to the choice tasks in the choice experiment survey. Efficient designs aim to generate parameter estimates with standard errors that are as low as possible, and thus produce the maximum information from each choice situation (e.g., Rose and Bliemer, 2009). The generation of efficient designs requires the specification of priors for the parameter estimates. In the pilot survey, we employed zero priors in the design, and used the parameter estimates obtained in the pilot study to construct the final experimental design. In the final study, we employed a Bayesian D-efficient design using Ngene (v. 1.0.2), taking 500 Halton draws for the prior parameter distributions. Bayesian designs take into account the uncertainty related to the parameter priors. Instead of fixed priors, they make use of random priors by specifying a mean and standard deviation for the prior.

Table 2 | Attributes of conservation programmes and their levels.

Attribute	Description	Current state	Levels (unit)
Native food plant varieties in gene banks	Native food plants are stored in a gene bank, either as seeds or plant parts.	The gene bank contains seeds from about 300 landrace varieties. Plants that are added vegetatively (e.g., berry and apple varieties) are missing.	300, 400, 500 (number of plants)
Farms growing native food plants	Farmers and hobby gardeners cultivate native food plants on farms or in gardens.	Seven farms grow seeds of native food plants with agri-environmental support. Other activities than growing seeds are not supported.	7, 500, 1000 (number of farms)
Native ornamental plant varieties mapped and in gene banks	Scientists identify and register native ornamental plants. Varieties are preserved in a gene bank, either as seeds or plant parts.	Only a small proportion of the native ornamental plants are known. Storage in the official gene bank is not provided.	small proportion, about half, majority (proportion of plants)
Native breeds in gene banks	Landrace breeds are kept in a gene bank as gametes and embryos.	The gene bank contains Western, Eastern and Northern Finncattle, as well as Finn-, Åland and Kainuu sheep. Native chicken, goat and horse breeds are missing from the gene bank.	3 cattle breeds and 3 sheep breeds (status quo level), + all combinations of goat, horse and chicken breeds
Native breeds on farms	Native breeds are kept on farms in their natural environment. A breed is considered to be endangered if the number of females is less than 1000.	Farms secure goat, horse and chicken breeds, Finnish sheep and Western Finncattle. Eastern and Northern Finncattle, as well as Åland and Kainuu sheep, are endangered.	1 cattle breed, 1 sheep breed, goat, horse and chicken (status quo level), + all combinations of additional 1-2 cattle and sheep breeds
Cost	Cost for taxpayers, €/year during 2012–2021.	No additional costs.	0, 5, 20, 40, 80, 100, 150, 300 (€)

Table 3 | Example of a choice set.

		Current state	Conservation programme A	Conservation programme B
Native food plant varieties in gene banks		Approximately 300	400	400
Farms growing native food plants		7 farms	2000 farms	1000 farms
Native ornamental plant varieties mapped and in gene banks		Some	Majority	About half
Native breeds in gene banks		3 cattle breeds, 3 sheep breeds	3 cattle breeds, 3 sheep breeds, chicken, goat, horse	3 cattle breeds, 3 sheep breeds, goat
Native breeds on farms		Goat, horse, chicken, 1 cattle breed, 1 sheep breed	Goat, horse, chicken, 3 cattle breeds, 1 sheep breed	Goat, horse, chicken, 2 cattle breeds, 3 sheep breeds
Cost for taxpayers, €/year during 2012–2021	€	€0/year	€80/year	€200/year
I support the alternative		( )	( )	( )

In the design phase, animal breeds in gene banks and on farms were treated as separate attributes, but were later combined to the "Native breeds in gene banks" and the "Native breeds on farms" attributes in the choice tasks presented to the respondents. Bayesian priors were employed for the chicken attribute and the number of cattle breeds on the farm attribute, and fixed priors for all other attributes. We generated 180 choice tasks, blocking them into 30 subsets, which resulted in six choice situations presented for each respondent. The final design had a D-error of 0.002.

STATISTICAL MODELS

The choices between environmental programmes were originally modeled with a conditional logit model (also called a multinomial logit model) (McFadden, 1974). The conditional logit, however, assumes a similar preference structure for all respondents, which implies that they have similar tastes for the attributes of conservation. In this study, we were particularly interested in defining heterogeneous citizen segments, which have a similar preference structure within each segment. One approach that allows this heterogeneity is the latent class model (Boxall and Adamowicz, 2002), which has frequently also been applied in choice experiment models of environmental conservation programmes (e.g., Garrod et al., 2012; Grammatikopoulou et al., 2012). In the latent class model, preferences are assumed to be homogeneous in each segment, but to vary between the segments.

In the modeling, price was treated as a continuous variable and the other attributes were effects-coded, implying that the parameters will sum to zero over the categories of the nominal variable concerned. The status quo attribute levels were thus included in the model, and could obtain either negative or positive coefficients depending on their effect on respondent's utility. Alternative-specific constants (ASC) were included for all alternatives in order to allow systematic choice tendencies not explained with the parameters describing the attributes.

Heterogeneity was statistically included in the latent class model by simultaneously dividing individuals into behavioral groups or latent segments, and estimating a choice model for each of these classes. The estimation was carried out by assuming first one class, then two classes, three classes and so forth. In each step, the explanatory power of the model was assessed to decide on the optimal number of classes. For this purpose, we used the Bayesian information criterion (BIC) and Akaike information criterion (AIC), which are log-likelihood scores with correction factors for the number of observations and the number of parameters. The latent class model also enables the calculation of the WTP for the attributes for each citizen segment.

The relationship between the individual characteristics and the latent classes was examined *a posteriori* of the actual estimation of the latent class model in order to describe the heterogeneous citizen segments. Thus, the segments were formed solely based

on the conservation program choices. The membership in the most probable segment was regressed using a logistic regression to characterize each class compared to the rest of the respondents. The explanatory variables for the class memberships included respondents' socioeconomic characteristics, perceived values and responsibilities, use of provided information, response certainty and self-reported perception of the carefulness of completing the survey. The independent variables in the logistic regression models and their descriptive statistics are presented in **Table 4**.

RESULTS

In 24% of the choice sets, the respondents chose the status quo option, i.e., the current state without any additional program to conserve AgGR. The probability of choosing one of the two alternative conservation programs varied between 46% for the lowest cost level of €5 and 28% for the highest cost level of €300.

Table 5 presents the conditional logit model results for the choice of the conservation programme. As expected, an increase in the programme cost negatively affected the probability of choosing it. Turning to consider the genetic resource attributes, the number of food plants in the gene bank was not statistically significant. All other attributes were significant in determining respondents' choices. A higher number of farms growing native plant varieties increased the choice probability. The larger the number of ornamental plants to be mapped and conserved in gene banks, the more probable it was that the respondent would choose the programme. Conserving native breeds of Finnish goats, horses and chickens in the gene bank all increased the support for the programme. The effect was highest for horse, followed by chicken and goat. The guaranteed existence of cattle breeds on farms had a positive and significant effect on choice. As expected, the effect was greater if the number of conserved cattle breeds was three instead of two. This was also the case with sheep breeds, although the conservation of two breeds did not have a positive effect on choice compared with the status quo of one conserved breed.

The alternative specific constants (ASC) capture the tendency to choose one of the alternatives which is not explained by

Table 4 | Variables in the logistic regression models.

Characteristic	Description	Mean	Standard deviation	Min	Max
Female	1 if the respondent is female, if male	0.49	0.50	0	1
Year of birth	Respondents year of birth, continuous	1960	15	1931	1992
High income	1 if household income is over €50,000 per year, 0 otherwise	0.45	0.49	0	1
High education	1 if respondents education level is university education, 0 otherwise	0.24	0.46	0	1
Eastern Finnish	1 if respondents lives in Eastern Finland, 0 otherwise	0.11	0.32	0	1
Childhood in city	1 if respondent spent his/her childhood in a city, 0 otherwise	0.41	0.49	0	1
Certainty	Mean of respondent's certainty in the conservation programme choices, on a scale of 10 completely certain—1 not at all certain.	6.85	2.23	1	10
Agri-environmental attitude	Importance of environmental issues in agriculture, mean of nine measures on scales from 1 to 4	3.26	0.44	1	4
Relative importance of preserving AgGR	The importance of preserving native breeds and varieties relative to other environmental protection measures, 1 if both equally important, >1 if preserving native breeds and varieties more important, <1 if other environmental protection measures more important	0.94	0.16	0.36	1.66
Existence value	Factor score based on 8 measures of the importance of existence values, continuous*	0.00	1.00	−4.38	2.39
Use values	Factor score based on 8 measures of the importance of use values, continuous*	0.00	1.00	−3.78	2.62
Citizen responsibility	Factor score based on 9 measures of stakeholder responsibilities in conservation*	0.00	1.00	−3.38	2.30
Consumer responsibility	Factor score based on 9 measures of stakeholder responsibilities in conservation*	0.00	1.00	−5.27	2.01
Farmer responsibility	Factor score based on 9 measures of stakeholder responsibilities in conservation*	0.00	1.00	−3.12	2.88
Familiarity of products	Familiarity of AgGR products, mean of 10 measures on scales from 1 to 3	2.03	0.42	1	3
Info use (animals) > 0.5 min	1 if respondent used more than 30 s for additional information about breeds, 0 otherwise	0.33	0.47	0	1
Info use (plants) > 0.5 min	1 if respondent used more than 30 s for additional information about varieties, 0 otherwise	0.35	0.48	0	1
Hasty response	1 if respondent evaluated his/her response as hasty, 0 if careful	0.05	0.22	0	1

**Detailed description of these variables can be found in Tienhaara et al. (2014).*

Table 5 | Conditional logit (CL) model results.

Variable	Coefficient	Wald *p*-value
ASC1 (SQ)	−0.263***	0.000
ASC2	0.291***	
ASC3	−0.028	
Cost	−0.005***	0.000
300 plants in bank (SQ)	0.002	1.000
400 plants in bank	−0.002	
500 plants in bank	0.000	
7 plants on farms (SQ)	−0.199***	0.000
500 plants on farms	0.075***	
1000 plants on farms	0.124***	
Ornamental plants in bank (SQ)	−0.057**	0.008
Ornamental plants in bank L2	−0.004	
Ornamental plants in bank L3	0.061***	
Goats (SQ)	−0.039***	0.005
Goats in bank	0.039***	
Horses (SQ)	−0.075***	0.000
Horses in bank	0.075***	
Chickens (SQ)	−0.047***	0.001
Chickens in bank	0.047***	
1 cattle breed on farms (SQ)	−0.114***	0.000
2 cattle breeds on farms	0.025	
3 cattle breeds on farms	0.089***	
1 sheep breed on farms (SQ)	0.020	0.027
2 sheep breeds on farms	−0.052***	
3 sheep breeds on farms	0.032	
No. of respondents	1608	
No. of observations	9484	
Correct predictions %	48	
R^2	0.04	

*z-test: *** 99% significance level; ** 95% significance level.*
SQ, attribute level in the status quo alternative.

the attributes. The negative ASC1 (SQ) coefficient showed the reluctance to choose the status quo alternative regardless of the attribute levels in the policy alternatives. Furthermore, the ASC2 and ASC3 coefficients differed unexpectedly in sign and significance. The positive coefficient for ASC2 and negative for ASC3 indicated that the conservation programme that was presented first received more support. This was surprising, as the programmes were not presented in a specific order in the survey. The model predicted 48% of the choices right, clearly exceeding the probability of correct random choices of 33%, leading to a relatively weak goodness of fit.

The homogeneity of preferences was tested in the estimation of the latent class models. Based on the AIC and BIC, the estimation process showed that a model of five citizen clusters provided the best fit of the data. **Table 6** presents the latent class model results with the cluster names, and the logit model for the membership of each cluster is presented in **Table 7**.

The latent class model showed that although preferences for some attributes, such as conserving goat and chicken breeds in gene banks and cattle breeds on farms, did not differ significantly between clusters, there was significant heterogeneity in preferences for most of the attributes. The first class, named as "conservationists," comprised 27% of the respondents. They did not take the personal cost of the conservation programme into account in their decision process, as the coefficient of the cost variable was not significant. Instead, almost all the conservation attributes had significant and positive signs. Contrary to other clusters, most plant-related attributes were significant for conservationists. They also valued the conservation of ornamental plants. **Table 5** also shows that this cluster perceived higher use and existence values from genetic resource conservation than respondents in other segments, and also higher than average certainty in their responses to the choice tasks. This class contained more men than women and considered the conservation not to be a responsibility of farmers. For this cluster we also tested the effect gardening as a hobby, but it did not turn out to be significant. Thus, it seems that these respondents did not support the program because of the possible private good aspect of measures to support native varieties in gardens.

The second cluster, covering 26% of the respondents, was named as "bid-sensitive animal conservers." This group had a higher tendency to choose the improvement programmes compared to the status quo. The coefficient of the bid was significant and the second smallest of all clusters. In this cluster, the emphasis of preferences was on the conservation of animal breeds. The conservation of plant varieties in gene banks was even valued negatively. These respondents perceived more often than average that citizens and consumers should be responsible for the conservation of genetic resources. They also had positive agri-environmental attitudes. Furthermore, the respondents in this cluster used more than the average time to familiarize themselves with the information available in the survey concerning PGR, and they were slightly younger than the average respondent.

A confusing aspect in the third cluster was the large difference between the ASC for the two conservation programmes. This group, comprising 17% of the respondents, had a considerably greater tendency to choose conservation programme A rather than B or the status quo, although this could not be explained by the experimental design and attribute levels. The bid variable followed expectations, but for the other attributes, only plants on farms and the class-independent variables were significant. The logistic regression revealed that members of this cluster were older and had a lower income, and they emphasized the responsibility of citizens in conservation. Geographically, this cluster had more members who lived in Eastern Finland. The respondents in this group were relatively uncertain of their preferences, used the additional information less, and responded, according to their self-evaluation, less carefully than other respondents. As there were random tendencies in their support for a programme (ASC), but they still preferred an increase in several conservation attributes, they were named as "uncertain supporters."

The fourth class, with 17% of respondents, clearly preferred the status quo option, as the ASC for the programme options were negative. The coefficient of the bid variable was not significant. Among these "status quo preferers," the choice was consistent with their negative attitudes, as the relative importance of AgGR was low, as well as the perceived existence and use values. Citizens

Table 6 | Latent class models for conservation programme choice.

	Class 1	Class 2	Class 3	Class 4	Class 5	Overall	
Pseudo R^2	0.131	0.288	0.019	0.015	0.472	0.559	
Class size	0.27	0.26	0.17	0.17	0.13		
Class names	**Conserva-tionists**	**Bid-sensitive animal conservers**	**Uncertain supporters**	**Status quo preferers**	**Bid sensitives**	**Wald *p*-value**	**Wald (=) *p*-value**
Attributes	**Coefficients and significance levels**						
ASC 1 (SQ)	−0.990***	−2.937***	−0.841***	1.668***	−0.554**	0.000	0.000
ASC 2	0.332***	1.499***	1.757***	−0.414**	0.478***		
ASC 3	0.658***	1.438***	−0.916***	−1.254***	0.076		
Cost	0.000	−0.018***	−0.003*	−0.001	−0.041***	0.000	0.000
300 plants in bank (SQ)	−0.162***	0.138**	0.018	0.412**	−0.322***	0.003	0.001
400 plants in bank	0.025	−0.007	0.078	−0.166	0.225*		
500 plants in bank	0.137**	−0.131*	−0.096	−0.245	0.097		
7 plants on farms (SQ)	−0.621***	−0.120*	−0.261**	−0.006	−0.169	0.000	0.000
500 plants on farms	0.125**	0.208***	0.237*	0.003	0.104		
1000 plants on farms	0.496***	−0.088	0.024	0.003	0.065		
Ornamental plants in bank (SQ)	−0.462***	0.015	0.116	−0.004	−0.332**	0.000	0.000
Ornamental plants in bank L2	0.158***	0.002	0.023	−0.053	0.16		
Ornamental plants in bank L3	0.304***	−0.017	−0.139	0.057	0.172		
Goats (SQ)	−0.063***	−0.063***	−0.063***	−0.063***	−0.063***	0.001	C.i.
Goats in bank	0.063***	0.063***	0.063***	0.063***	0.063***		
Horses (SQ)	−0.152***	−0.128***	−0.075	0.447***	−0.256***	0.000	0.000
Horses in bank	0.152***	0.128***	0.075	−0.447***	0.256***		
Chickens (SQ)	−0.062***	−0.062***	−0.062***	−0.062***	−0.062***	0.001	C.i.
Chickens in bank	0.062***	0.062***	0.062***	0.062***	0.062***		
1 cattle breed on farms (SQ)	−0.144***	−0.144***	−0.144***	−0.144***	−0.144***	0.000	C.i.
2 cattle breeds on farms	0.034	0.034	0.034	0.034	0.034		
3 cattle breeds on farms	0.110***	0.110***	0.110***	0.110***	0.110***		
1 Sheep breed on farms (SQ)	−0.213***	0.046	−0.036	0.581***	−0.245**	0.000	0.001
2 Sheep breeds on farms	0.056	−0.04	−0.156	−0.282	0.116		
3 Sheep breeds on farms	0.157***	−0.007	0.192	−0.300	0.128		
No. of respondents	1608						
No. of observations	9484						
Correct predictions %	85						

*z-test: *** 99% significance level; ** 95% significance level; * 90% significance level.*

SQ, attribute level in the status quo alternative.

C.i., class independent.

and consumers were less frequently seen as those responsible for conservation; instead, it was perceived as a responsibility of the farmers. This class was characterized by an older age, lower educational level and growing up on a farm.

The fifth class of respondents (13%), named as "bid sensitives," were the most sensitive to the cost of the programme of all groups. Nevertheless, the ASC revealed that they were interested in conservation, and almost all conservation attributes had significant coefficients. Among these respondents, particularly the *ex situ* conservation of Finnhorse positively affected their choices. In this class, the conservation of genetic resources was not seen as a responsibility of citizens or consumers. The logit model for this group showed that they evaluated themselves as careful respondents but felt somewhat uncertain of their choices. They were younger than average and less familiar with products from traditional breeds and varieties.

WTP for different attributes was calculated based on the conditional logit model and the latent class model for those classes for which the cost coefficient was significant (**Table 8**). WTPs based on the conditional logit model indicated that plants on farms, cattle breeds and horses were most highly valued. In general, there was substantial variation in WTPs between the classes. In class 3, WTPs were higher due to the low importance of the cost attribute.

DISCUSSION AND CONCLUSIONS

The results of a choice experiment concerning agricultural genetic resource policies showed that citizens are interested in

Table 7 | Logistic regression models profiling consumer classes.

Class	Class 1	Class 2	Class 3	Class 4	Class 5
Variable	**Coefficients and significance levels**				
Constant	−2.76***	−43.31***	48.77***	39.90**	−29.46**
Female	−0.46***				
Year of birth		0.02***	−0.02***	−0.02**	0.02*
High income			−0.39**		
High education				−0.72***	
Eastern Finnish			0.40*		
Childhood in city				−0.68**	
Certainty	0.12***		−0.09**		−0.08**
Agri-environmental attitude	0.37*	0.43**			
Relative importance of AgGR		−1.482***	1.412**	−1.82**	
Existence values	0.32***			−0.50***	
Use values	0.38***			−0.39***	
Citizen responsibility		0.29***	0.21**	−1.06***	−0.43***
Consumer responsibility		0.17**		−0.31**	−0.38***
Farmer responsibility	−0.16**			0.27**	
Familiarity of products					−0.48**
Info use (animals) > 0.5 min				−0.39*	
Info use (plants) > 0.5 min		0.54***	−0.47***		
Hasty response			0.70*		−1.08**
N	1088	1201	1098	1077	1199
Nagelkerke R^2	0.103	0.083	0.071	0.397	0.104
Chi-squared	81.99	71.44	46.48	252.37	68.25
p-Value	0.000	0.000	0.000	0.000	0.000
Correctly classified (cut 0.5)	69.6	71.6	83.9	90.4	86.8

*Variables are significant at the *** 99% level, ** 95% level, * 90% level.*

Table 8 | Annual willingness to pay (in 2009 €) for attributes.

	Conditional logit model	Latent class model, Class 2	Latent class model, Class 3	Latent class model, Class 5
Plants in bank (400)	–	–	–	13
Plants in bank (500)	–	-15	–	–
Plants on farms (500)	60	19	7	–
Plants on farms (1000)	70	–	–	–
Ornamental plants (majority) inventoried and in bank	14	–	–	–
Goats in bank	17	7	105	3
Horses in bank	33	15	–	12
Chickens in bank	20	7	104	3
3 cattle breeds on farms	44	14	211	6
2 sheep breeds on farms	−15	–	–	–

–, Indicates that the estimate is missing due to the non-significance of the cost coefficient.

the conservation of native breeds and varieties in agriculture. However, there was considerable variation in preferences between citizen segments. Of the five identified groups, two groups covering over half of the respondents had a high interest in the conservation of native breeds and varieties. Respondents in one of the segments clearly preferred the current state of conservation to additional conservation efforts, while one group had a favorable attitude toward conservation if the expenses were on a low level, and respondents in one segment were supportive but wavering in their preferences. The respondent groups were identified based on their preferences for conservation, and they also differed with respect to the use of additional information, their response carefulness and the certainty of the stated WTP.

Similar to previous studies of consumer preferences on biodiversity (e.g., Hanley et al., 1995), we also found lexicographic preferences for conserving AgGR. Those were expressed by the largest group of respondents (27%), as their interest in conservation was high regardless of the costs. Lexicographic choices can occur as a result of simplification if the respondent finds the choice task too difficult to handle or as a result of actual lexicographic preferences (Sælensminde, 2006). In our case, it is difficult to determine whether respondents exhibited lexicographic preferences because they wanted to simplify the choice tasks or because the differences in the attribute levels were large. Respondents in the group which exhibited lexicographic preferences were more certain about their preferences, which supports the phenomenon of actual lexicographic preferences as the reason for their choices. In addition, their positive perceptions concerning the existence and use values of genetic resources support the observation of actual lexicographic preferences.

Due to the preference structures, WTP estimates were only obtained for three respondent groups and some of the attributes. In those groups where the cost variable was significant and meaningful WTP estimates could thus be estimated, the marginal WTPs were considerably lower than the WTPs of the whole sample based on the conditional logit model. This implies that in the whole sample, the results were influenced by the groups that were insensitive to the costs of conservation.

Our results can be compared with those obtained by Zander et al. (2013), who assessed the economic value of conservation programs for two Italian cattle breeds using a choice experiment directed to citizens. Zander et al. (2013) also found preference heterogeneity for most of the attributes of the conservation programs, as well as differences in the sensitivity to the cost attribute. According to their findings, 85% of the respondents supported increased conservation, and the mean WTP was 90€

for conserving each breed. The present results can also be linked to previous results of heterogeneity among farmers using native breeds and varieties. Soini et al. (2012) identified a segment of production-oriented farmers among European cattle breeders that would benefit from increased subsidies for keeping native breeds on farms. If the subsidies were increased to correspond to citizens' WTP, it would help particularly this subsidy-dependent group of farmers.

As the survey was Internet-based, we were able to obtain information on the time used for obtaining additional information about plant and AnGR. These variables, combined with certainty, could partly explain the membership in the latent classes. However, similarly to Hu et al. (2009) and Vista et al. (2009), there were no clear tendencies for the use of information to be associated with a lower or higher WTP. Further research is, however, needed to clarify the associations between preferences, uncertainty and information acquisition in the case of genetic resources.

The results provide implications concerning how to direct the conservation policies for AgGR in Finland. The WTP estimates for the attributes of the conservation programmes indicated that the participants valued particularly *in situ* conservation in the case of PGR, which would also imply the existence of native plant varieties in the landscape. However, a moderate level of this *in situ* conservation would be sufficient, as the highest level increased the WTP only slightly. For the conservation of animal breeds, the results emphasize the importance of *in situ* conservation of cattle breeds. The weak support for the conservation of sheep breeds compared to cattle breeds was understandable, as Finnsheep breeds are less familiar to the public. However, the low, even negative, WTP for the conservation of sheep breeds is in contradiction with the importance of Finnsheep in breeding (e.g., Thomas, 2010). *Ex situ* conservation of those animal breeds that are at present insufficiently protected in gene banks was perceived as important, particularly the conservation of the genetic material of the Finnhorse.

Although the cost-effectiveness of AgGR conservation is case-dependent, some previous studies have recommended *ex situ* conservation in gene banks as a less expensive, less vulnerable and less policy-sensitive method of conservation (Dulloo et al., 2010; Silversides et al., 2012). These cost-effectiveness considerations do not, however, take into account the additional benefits that may be associated with *in situ* conservation, such as the visibility of local breeds and varieties in the landscape or the opportunity to use local breed products. Thus, taking into account citizens' preferences for *in situ* and *ex situ* conservation and using cost-benefit analysis in policy evaluation may shift the priorities of agricultural genetic resource conservation policies.

In this study, the conservation policies were based on equal participation of all citizens, as the policy was financed with taxes. An alternative approach would be to apply market-based incentives, e.g., *payments for environmental services (PES)* for the conservation of genetic resources (McNeely, 2006; Wunder, 2007; Narloch et al., 2011). PES would imply that actors who are major users of the resources are involved in making and adapting rules for conservation markets. For future experiments of PES, our results of the citizen groups that are most interested provide information for identifying the interested parties for the markets of AgGR.

REFERENCES

Ahtiainen, H., and Pouta, E. (2011). The value of genetic resources in agriculture: a meta-analysis assessing existing knowledge and future research needs. *Int. J. Biodivers. Sci. Ecosyst. Serv. Manage.* 7, 27–38. doi: 10.1080/21513732.2011.593557

Berrens, R. P., Bohara, A. K., Jenkins-Smith, H. C., Silva, C. L., and Weimer, D. L. (2004). Information and effort in contingent valuation surveys: application to global climate change using national internet samples. *J. Environ. Econ. Manage.* 47, 331–363. doi: 10.1016/S0095-0696(03)00094-9

Birol, E., Smale, M., and Gyovai, Á. (2006). Using a choice experiment to estimate farmers' valuation of agrobiodiversity on hungarian small farms. *Environ. Resour. Econ.* 34, 439–469. doi: 10.1007/s10640-006-0009-9

Blomquist, G., and Whitehead, J. (1998). Resource quality information and validity of willingness to pay in contingent valuation. *Resour. Energy Econ.* 20, 179–196. doi: 10.1016/S0928-7655(97)00035-3

Boxall, C. P., and Adamowicz, L. W. (2002). Understanding heterogeneous preferences in random utility models: a latent class approach. *Environ. Resour. Econ.* 23, 421–446. doi: 10.1023/A:1021351721619

Brown, G. M. (1990). "Valuation of genetic resources," in *The Preservation and Valuation of Biological Resources,* eds G. H. Orians, G. M. Jr. Brown, W. E. Kunin, and J. E. Swierzbinski (Seattle, WA: University of Washington Press), 203–228.

Cameron, T. A., and Englin, J. (1997). Respondent experience and contingent valuation of environmental goods. *J. Environ. Econ. Manage.* 33, 296–313. doi: 10.1006/jeem.1997.0995

Drucker, A. G., Gomez, V., and Anderson, S. (2001). The economic valuation of farm animal genetic resources: a survey of available methods. *Ecol. Econ.* 36, 1–18. doi: 10.1016/S0921-8009(00)00242-1

Dulloo, M. E., Hunter, D., and Borelli, T. (2010). *Ex situ* and *in situ* conservation of agricultural biodiversity: major advances and research needs. *Notulae Botanicae Horti Agrobotanici Cluj-Napoca* 38, 123–135.

Evenson, R. E., Gollin, D., and Santaniello, V. (1998). "Introduction and overview: agricultural values of plant genetic resources," in *Agricultural Values of Plant Genetic Resources*, eds R. E. Evenson, D. Gollin, and V. Santaniello (Wallingford: CABI Publishing), 1–25.

FAO, (2007). *The State of the World's Animal Genetic Resources for Food and Agriculture*. Rome: Commission on Genetic Resources for Food and Agriculture, FAO.

FAO, (2010). *The Second Report on the State of the World's Plant Genetic Resources for Food and Agriculture.* Available online at: http://www.fao.org/docrep/013/i1500e/i1500e.pdf [Accessed 11 October 2011].

Garrod, G., Ruto, E., Willis, K., and Powe, N. (2012). Heterogeneity of preferences for the benefits of environmental stewardship: a latent-class approach. *Ecol. Econ.* 76, 104–111. doi: 10.1016/j.ecolecon.2012.02.011

Grammatikopoulou, I., Pouta, E., Salmiovirta, M., and Soini, K. (2012). Heterogeneous preferences for agricultural landscape improvements in southern Finland. *Landsc. Urban Plan.* 107, 181–191. doi: 10.1016/j.landurbplan.2012.06.001

Hanley, N., Mourato, S., and Wright, R. E. (2001). Choice modelling approaches: a superior alternative for environmental valuation? *J. Econ. Surv.* 15, 435–462. doi: 10.1111/1467-6419.00145

Hanley, N., Spash, C., and Walker, L. (1995). Problems in valuing the benefits of biodiversity protection. *Environ. Res. Econ.* 5, 249–272. doi: 10.1007/BF00691519

Hu, W., Adamovicz, W. L., and Veerman, M. M. (2009). Consumers' preferences for GM food and voluntary information access: a simultaneous choice analysis. *Can. J. Agri. Econ.* 57, 241–267. doi: 10.1111/j.1744-7976.2009.01150.x

McFadden, D. (1974). "Conditional logit analysis of qualitative choice behaviour," in *Frontiers in Econometrics*, ed P. E. Zarembka (New York, NY: Academic Press), 105–142.

McNeely, J. A. (2006). Using economic instruments to overcome obstacles to *in situ* conservation of biodiversity. *Integr. Zool.* 1, 25–31. doi: 10.1111/j.1749-4877.2006.00009.x

Narloch, U., Drucker, A., and Pascual, U. (2011). Payments for agrobiodiversity conservation services for sustained on-farm utilization of

plant and animal genetic resources. *Ecol. Econ.* 70, 1837–1845. doi: 10.1016/j.ecolecon.2011.05.018

Oldfield, M. L. (1989). *The Value of Conserving Genetic Resources*. Sunderland, MA: Sinauer Associates Inc.

Omondi, I., Baltenweck, I., Drucker, A. G., Obare, G., and Zander, K. K. (2008). Valuing goat genetic resources: a pro-poor growth strategy in the Kenyan semi-arid tropics. *Trop. Anim. Health Prod.* 40, 583–589. doi: 10.1007/s11250-008-9137-2

Ouma, E., Abdulai, A., and Drucker, A. (2007). Measuring heterogeneous preferences for cattle traits among cattle-keeping households in East Africa. *Am. J. Agric. Econ.* 89, 1005–1019. doi: 10.1111/j.1467-8276.2007.01022.x

Rege, J. E. O., and Gibson, J. P. (2003). Animal genetic resources and economic development: issues in relation to economic valuation. *Ecol. Econ.* 45, 319–330. doi: 10.1016/S0921-8009(03)00087-9

Roessler, R., Drucker, A. G., Scarpa, R., Markemann, A., Lemke, U., Thuy, L. T., et al. (2008). Using choice experiments to assess smallholder farmers' preferences for pig breeding traits in different production systems in North-West Vietnam. *Ecol. Econ.* 66, 184–192. doi: 10.1016/j.ecolecon.2007.08.023

Rose, J. M., and Bliemer, M. C. J. (2009). Constructing efficient stated choice experimental designs. *Trans. Rev.* 29, 597–617. doi: 10.1080/01441640902827623

Sælensminde, K. (2006). Causes and consequences of lexicographic choices in stated choice studies. *Ecol. Econ.* 59, 331–340. doi: 10.1016/j.ecolecon.2005.11.001

Silversides, F. G., Purdy, P. H., and Blackburn, H. D. (2012). Comparative costs of programmes to conserve chicken genetic variation based on maintaining living populations or storinf cryopreserved material. *Br. Poult. Sci.* 53, 599–607. doi: 10.1080/00071668.2012.727383

Soini, K., Diaz, C., Gandini, G., de Haas, Y., Lilja, T., Martin-Collado, D., et al. (2012). Developing a typology for local cattle breed farmers in Europe. *J. Anim. Breed. Genet.* 129, 436–447. doi: 10.1111/j.1439-0388.2012.01009.x

Taloustutkimus. (2013). *Internet Panel*. Available online at: http://www.taloustutkimus.fi/in-english/products_services/internet_panel/ [Accessed 28 August 2013].

Thomas, D. L. (2010). Performance and utilization of Northern European short-tailed breeds of sheep and their crosses in North America: a review. *Animal* 4, 1283–1296. doi: 10.1017/S1751731110000856

Tienhaara, A., Ahtiainen, H, and Pouta, E. (2014). "The consumer and citizen roles and motives in valuation of agricultural genetic resources in Finland," in *Paper Presented in the 20th Annual Conference of the European Association of Environmental and Resource Economists* (Toulouse). Available online at: http://www.webmeets.com/EAERE/2013/m/viewpaper.asp?pid=870

Vista, A. B., Rosenberger, R. S., and Collins, A. R. (2009). If you provide, will they read it? Response time effects in a choice experiment. *Can. J. Agri. Econ.* 57, 365–377. doi: 10.1111/j.1744-7976.2009.01156.x

Wunder, S. (2007). The efficiency of payments for environmental services in tropical conservation. *Conserv. Biol.* 21, 48–58. doi: 10.1111/j.1523-1739.2006.00559.x

Zander, K., Signorello, G., De Salvo, M., Gandini, G., and Drucker, A. (2013). Assessing the total economic value of threatened livestock breeds in Italy: Implications for conservation policy. *Ecol. Econ.* 93, 219–229. doi: 10.1016/j.ecolecon.2013.06.002

Conflict of Interest Statement: The Guest Associate Juha Kantanen declares that, despite being affiliated to the same institution as the authors, the review process was handled objectively and no conflict of interest exists. The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 29 September 2014; accepted: 01 December 2014; published online: 18 December 2014.

*Citation: Pouta E, Tienhaara A and Ahtiainen H (2014) Citizens' preferences for the conservation of agricultural genetic resources. Front. Genet. **5**:440. doi: 10.3389/fgene.2014.00440*

This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics.

Copyright © 2014 Pouta, Tienhaara and Ahtiainen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

OPINION
published: 30 April 2015
doi: 10.3389/fgene.2015.00153

Farm animal genetic and genomic resources from an agroecological perspective

Michèle Tixier-Boichard, Etienne Verrier, Xavier Rognon and Tatiana Zerjal*

GABI, INRA, AgroParisTech, Université Paris-Saclay, Paris, France

Keywords: agroecology, ecosystem services, local breeds, genetic variation, genomics, animal breeding

Agroecology, as a scientific approach, relies on a better knowledge of biodiversity at all levels of organization and function, in order to better manage agricultural production systems, from farm scale to landscape. Ecological concepts such as functional redundancy, complementary use of resources, can be applied to farming systems, with the purpose of improving their resilience. Transposing the concepts of agroecology to livestock production has been recently proposed by Dumont et al. (2013). One of the principles proposed for the design of sustainable animal production systems is to enhance diversity within animal production systems in order to strengthen their resilience. Why is it so? An increased biodiversity allows benefiting from complementary aptitudes. For example, in the case of disease resistance, the diversity of hosts will limit the risk of the specialization of a highly pathogenic agent with devastating consequences. It does not mean that diseases will not occur but the spread of infections and the overall impact on animal health should be limited (Springbett et al., 2003).

How the agroecology concepts can be applied to farm animal genetic resources and how the genomics approach may be used to facilitate it?

Biodiversity in livestock production systems may be considered at all scales, from individuals and breeds to species and ecosystems. Several levels may be considered to increase biodiversity in livestock production systems, such as intermixing species within production systems. This paper will focus on within-species biodiversity that goes from local breeds to highly selected populations. At the population scale, one possibility is to increase the number of breeds in use, or to produce new composite populations, as done for the Creole cattle in the French West Indies (Gautier and Naves, 2011). At the local scale (i.e., the farm), the increase of biodiversity may be obtained by intermixing breeds, by using crossbreeding but also by monitoring individual genetic diversity within a group. What could be the drivers for such a trend toward increased biodiversity? Stratified crossbreeding between local breeds is likely to increase the level of diversity at the individual level, because of a higher heterozygosity of F1, but may not increase the diversity within a group since F1 animals are likely to inherit the most frequent alleles present in their parental breeds. The main issue is still to maintain genetic diversity within each parental breed. Breeding for an increased production level has led to a decrease of within-breed variability (Danchin-Burge et al., 2012) although measures have been taken to limit this decrease. Actually, another incentive than selection response is needed to trigger an increased use of biodiversity in livestock systems.

We propose to use the conceptual framework of ecosystem services for this purpose. Ecosystem services are benefits that human populations get from natural or managed ecosystems. This framework includes not only the provision of food but also environmental and socio-cultural services. Although the concept of ecosystem services is still evolving and open to debate (Lele et al., 2013), it is applied to value ecosystem services in complex ecosystems incorporating livestock, as shown by Silvestri et al. (2013) in Kenya. The need to quantify various services and their interactions is opening a new research field, which is relevant to characterize genetic

OPEN ACCESS

Edited by:
Michael William Bruford, Cardiff University, UK

Reviewed by:
Stephen Hall, University of Lincoln, UK

****Correspondence:***
Michèle Tixier-Boichard, michele.boichard@jouy.inra.fr

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: *26 January 2015*
Accepted: *05 April 2015*
Published: *30 April 2015*

Citation:
Tixier-Boichard M, Verrier E, Rognon X and Zerjal T (2015) Farm animal genetic and genomic resources from an agroecological perspective. Front. Genet. 6:153. doi: 10.3389/fgene.2015.00153

resources in various production systems. Moreover, the frame of ecosystem services is strongly connected with sustainability, as shown by Broom et al. (2013) for sylvo-pastoral systems. However, it should not be restrained to the extensive systems but should involve all production systems with the genetic resources embedded in them.

The framework of ecosystem services requires the development of a multi-criteria assessment of the added value of an individual or a breed, beyond its contribution to food production, which is generally valued by a market price. Regarding local breeds, the production of high quality food for niche markets has often been considered as a very good opportunity for their preservation, besides the development of commercial breeds (Lambert-Derkimba et al., 2013). In addition to distinctive products, local breeds may also provide environmental services, which are still difficult to assess and value. Generally, such niche markets remain fragile, and the cultural services they provide are difficult to quantify. Environmental services may be positive or negative, local or distant. For example, positive services may involve fire prevention through extensive grazing, maintenance of open landscapes or habitats for wild species, manure production as a fertilizer for crops. On the opposite, negative services may be the increased pollution due to excess of manure, at a local scale, or the increased deforestation for production of soy-bean or maize, at a distant scale. Socio-cultural services may involve rural development, landscape management, ecotourism, which valuation is a real challenge. Raising livestock is an activity embedded into an agroecosystem, combining livestock and plant production benefits from complementarity in the use of resources.

Research is needed to develop methods and indicators to quantify different types of services in order to integrate them into breeding goals. There will be trade-offs between services, as well as between productive and adaptive traits in livestock. Quantitative genetics has tools to combine different traits in a breeding goal, whatever the correlation between traits, and could do the same to combine different services provided that appropriate weighting parameters are defined. Stratified crossbreeding is another approach to combine different aptitudes in F1 animals in order to optimize trade-offs between traits. The possibility to apply this approach to other services than production remains to be investigated.

The identification of traits or functions which are the basis for these services is an important research issue where genomics can play a major role. The genome is an archive of the population past and recent history but, until recently, few markers were available to read this archive. In the last 10 years, whole genome sequencing tools have developed allowing to understand population history and to unravel the genetic determinism of complex traits. Thus, genome sequencing has provided a universal frame for all geneticists working on a species, to share tools and data. At the same time conservation genetics is moving toward conservation genomics (Ouborg et al., 2010). High density SNP chips are powerful tools to monitor genetic diversity at different scales: animals within a herd, herds within a landscape and breeds within a species. In terms of functional diversity, sequencing can be used to reveal the footprints of natural or artificial selection, provided that relevant statistical methods are used to distinguish these footprints from drift effects. For example, landscape genomics may help to identify the genetic basis for some environmental services, such as climatic adaptation (e.g., Flori et al., 2012), tolerance to rough diets, or resistance to pathogens, provided that data are available to document these services. At present, genomic selection has revolutionized the organization of cattle selection and it is likely to be applied in other species, provided that a sufficiently large reference population is established to ensure reliable associations between genotype and phenotype. Such data sets already exist, as shown by Hayes et al. (2009) who studied adaptation to tropical climate of Holstein Friesian in Australia.

Genomics can support an agroecological management of animal genetic resources at three levels.

The first one is the monitoring of genetic diversity at any scale, and more particularly at the herd level: genotyping animals within a herd will allow to calculate the mean genetic distance of any external animal, sire, or dam, to the herd, and to integrate the benefit associated with the introduction of an external animal to the herd diversity. This procedure is similar to those used to minimize inbreeding in a mating plan at a breed level, but it takes place at the herd level, to maximize its diversity. This will require the reduction of costs for genotyping. As an example, the low density (LD) cattle chip is the most used and the cheapest SNP array. Other requirements are that breeders have easy access to the genotyping results, through their breeding organization or technical services, and to guidelines explaining how to use these results to monitor the diversity within herds. The benefit is to maximize genome diversity and the overall genetic resource within the herd. Research is needed to prove that maximizing genetic diversity will indeed improve the tolerance of the herd to climate change and extreme events which are likely to become more frequent, and guarantee a stable performance of the group, which is an important objective for the farmer. Genomic selection and individual dairy cow genotyping are producing large datasets suitable to test such a relationship between genetic diversity and herd resilience to extreme events. This would represent a change in breeding methods, since the best herd will not be an ensemble of the best animals but the best set of diverse animals.

The second one is the association of genomic regions with a range of phenotypes or performance levels. Monitoring these regions may provide a tool to tailor the genetic make-up of a herd with regards to specific traits that need to be combined at the herd level. This approach targets specific aptitudes to be combined, going beyond the first approach which aimed at verifying that the herd harbors a sufficiently high diversity, leaving options for future choices. The main challenge is to define the relevant phenotype to be predicted with molecular markers. Which phenotypes correspond to environmental or socio-cultural services? Research is needed on these issues, in close connection with the farmers, who are willing to better characterize their breed and raise awareness about its value, and with other actors who are benefiting from the service. Genomic regions associated to a desired trait or service should be validated at a large scale so that their effect is likely to be observed in any herd or any herd x environment combination. However, environmental services may depend on the herd location and be difficult to assess

on a large scale because of genotype by environment interactions (GxE). In this case the farmer may choose a specific genetic profile for its herd.

The third one is targeting local breeds, where genomic selection and large scale reference populations are not available. There are two main issues for these breeds: managing their genetic variability and characterizing their specific features. Such breeds harbor original characteristics, either for product quality, or adaptation to specific environments, but the genetic basis for these characteristics is seldom known. If farmers had access to this information, how would they use it? What if adaptive features involved epigenetic changes that cannot be tested by standard SNP genotyping? Research is obviously needed in this regard. The use of such information may require a change of scale in the breed management, to record more traits and share more data between herds. This is more an organizational challenge and a social issue than a genetic issue. Furthermore, the identification of original features in local breeds may trigger the onset of crossbreeding programs to introduce such features in commercial populations, which raises strong issue for benefit sharing and for preserving breed specificities. Implementing genomics for the management of genetic variability appears relatively easy, provided that the genotyping cost is affordable, which is not so obvious for small populations. To share costs, a multi-breed genotyping tools could be an option to explore. With this approach, farmers will get accurate information about drift, population fragmentation or introgression events affecting a breed. Population fragmentation is typically expected when small size herds do not exchange animals, to a point that the genetic difference can be very high among them. Such a situation with a low within-herd variability but a high between herds variability could be favorable at the breed level from the viewpoint of agroecology (diversity and complementarity between herds) but it might question the breed definition and the breed identity. The same issue could appear if introgression events are detected: the breed would not anymore be the "pure" local breed as often described, even though this introgression could contribute to the breed evolution and facilitate its adaptation to future conditions. It is not to be excluded that the genomic characterization will enhance social issues, revealing how farmers are managing their breed.

In conclusion, setting the management of animal genetic resources in an agroecological perspective raises two major issues where research is needed: the possible transposition of ecosystem services to animal breeding and the impact of genomic tools on the breed concept and the management of genetic diversity at different scales.

References

Broom, D. M., Galindo, F. A., and Murgueitio, E. (2013). Sustainable, efficient livestock production with high biodiversity and good welfare for animals. *Proc. Royal Soc. B Biol. Sci.* 280, 2013–2025. doi: 10.1098/rspb.2013.2025

Danchin-Burge, C., Leroy, G., Moureaux, S., and Verrier, E. (2012). Evolution of the genetic variability of 8 French dairy cattle breeds assessed by pedigree analysis. *J. Anim. Breed. Genet.* 129, 206–217. doi: 10.1111/j.1439-0388.2011.00967.x

Dumont, B., Fortun-Lamothe, L., Jouven, M., Thomas, M., and Tichit, M. (2013). Prospects from agroecology and industrial ecology for animal production in the 21st century. *Animal* 7, 1028–1043. doi: 10.1017/S1751731112002418

Flori, L., Gonzatti, M. I., Thevenon, S., Chantal, I., Pinto, J., Berthier, D., et al. (2012). A quasi-exclusive european ancestry in the Senepol tropical cattle breed highlights the importance of the slick locus in tropical adaptation. *PLoS ONE* 7:e36133. doi: 10.1371/journal.pone.0036133

Gautier, M., and Naves, M. (2011). Footprints of selection in the ancestral admixture of a New World Creole cattle breed. *Mol. Ecol.* 20, 3128–3143. doi: 10.1111/j.1365-294X.2011.05163.x

Hayes, B., Bowman, P. J., Chamberlain, A. J., Savin, K., van Tassell, C. P., Sonstegard, T. S., et al. (2009). A validated genome wide association study to breed cattle adapted to an environment altered by climate change. *PLoS ONE* 4:e6676. doi: 10.1371/journal.pone.0006676

Lambert-Derkimba, A., Lauvie, A., and Verrier, E. (2013). How the development of products valorizing local breeds changes breeding goals: examples from French cattle breeds. *Anim. Genet. Resour.* 53, 135–140. doi: 10.1017/S2078633612000768

Lele, S., Springate-Baginski, O., Lakerveld, R., Deb, D., and Dash, P. (2013). Ecosystem services: origins, contributions, pitfalls, and alternatives. *Conserv. Soc.* 11, 343–358. doi: 10.4103/0972-4923.125752

Ouborg, N. J., Pertoldi, C., Loeschke, C., Bijlsma, R., and Hedrick, P. W. (2010). Conservation genetics in transition to conservation genomics. *Trends Genet.* 26, 177–187. doi: 10.1016/j.tig.2010.01.001

Silvestri, S., Zaibet, L., Said, M. Y., and Kifugo, S. C. (2013). Valuing ecosystem services for conservation and development purposes: a case study from Kenya. *Environ. Sci. Policy* 31, 23–33. doi: 10.1016/j.envsci.2013.03.008

Springbett, A. J., MacKenzie, K., Woolliams, J. A., and Bishop, S. C. (2003). The contribution of genetic diversity to the spread of infectious diseases in livestock populations. *Genetics* 165, 1465–1474.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Tixier-Boichard, Verrier, Rognon and Zerjal. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in GENETICS

REVIEW ARTICLE
published: 25 February 2015
doi: 10.3389/fgene.2015.00052

Utilization of farm animal genetic resources in a changing agro-ecological environment in the Nordic countries

Juha Kantanen[1,2]*, Peter Løvendahl[3], Erling Strandberg[4], Emma Eythorsdottir[5], Meng-Hua Li[1,6], Anne Kettunen-Præbel[7], Peer Berg[7] and Theo Meuwissen[8]

[1] Green Technology, Natural Resources Institute Finland, Jokioinen, Finland
[2] Department of Biology, University of Eastern Finland, Kuopio, Finland
[3] Center for Quantitative Genetics and Genomics, Department of Molecular Biology and Genetics, Aarhus University, Tjele, Denmark
[4] Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala, Sweden
[5] Faculty of Land and Animal Resources, Agricultural University of Iceland, Reykjavik, Iceland
[6] Key Laboratory of Animal Ecology and Conservation Biology, Institute of Zoology, Chinese Academy of Sciences, Beijing, China
[7] NordGen – Nordic Genetic Resource Center, Aas, Norway
[8] Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, Aas, Norway

Edited by:
Ino Curik, University of Zagreb, Croatia

Reviewed by:
Shaolin Wang, University of Virginia, USA
Mahdi Saatchi, Iowa State University, USA

***Correspondence:**
Juha Kantanen, Green Technology, Natural Resources Institute Finland, Myllytie 1, FI-31600 Jokioinen, Finland
e-mail: juha.kantanen@luke.fi

Livestock production is the most important component of northern European agriculture and contributes to and will be affected by climate change. Nevertheless, the role of farm animal genetic resources in the adaptation to new agro-ecological conditions and mitigation of animal production's effects on climate change has been inadequately discussed despite there being several important associations between animal genetic resources and climate change issues. The sustainability of animal production systems and future food security require access to a wide diversity of animal genetic resources. There are several genetic questions that should be considered in strategies promoting adaptation to climate change and mitigation of environmental effects of livestock production. For example, it may become important to choose among breeds and even among farm animal species according to their suitability to a future with altered production systems. Some animals with useful phenotypes and genotypes may be more useful than others in the changing environment. Robust animal breeds with the potential to adapt to new agro-ecological conditions and tolerate new diseases will be needed. The key issue in mitigation of harmful greenhouse gas effects induced by livestock production is the reduction of methane (CH_4) emissions from ruminants. There are differences in CH_4 emissions among breeds and among individual animals within breeds that suggest a potential for improvement in the trait through genetic selection. Characterization of breeds and individuals with modern genomic tools should be applied to identify breeds that have genetically adapted to marginal conditions and to get critical information for breeding and conservation programs for farm animal genetic resources. We conclude that phenotyping and genomic technologies and adoption of new breeding approaches, such as genomic selection introgression, will promote breeding for useful characters in livestock species.

Keywords: adaptation, animal genetic resources, climate change, genomics, genomic selection, livestock, methane, mitigation

INTRODUCTION

Studies on impacts of climate change on primary industries in the Nordic countries (Denmark, Finland, Iceland, Norway, and Sweden) have focused mainly on agricultural productivity, land use, political issues, and water resource availability (e.g., Olesen and Bindi, 2002; Ciscar et al., 2011; Hakala et al., 2011; Olesen et al., 2011; Höglind et al., 2013). Different climate change scenarios and adaptation strategies for northern European conditions have also been discussed (e.g., Benestad, 2005; Olesen et al., 2011). In studies and reports, climate change and livestock issues have been only modestly considered even though livestock production is the most important sector in northern European agriculture as measured by the total value of production (e.g., Niemi and Ahlstedt, 2014) and has effects on and is influenced by climate change. Domestic animal genetic resources for food and agriculture in particular have not yet been adequately considered in strategies for adaptation to and mitigation of current global climate changes (McMichael et al., 2007; Hoffman, 2010) and issues on genetic resources are typically focussed on future plant breeding scenarios (e.g., Ceccarelli et al., 2010; Olesen et al., 2011). However, The Global Plan of Action for Animal Genetic Resources (GPA), published by the Commission on Genetic Resources for Food and Agriculture, lists several associations between animal genetic resources and climate change (FAO, 2007). As pointed out by Hoffman (2010) and Pilling and Hoffman (2011), sustainability and robustness of animal production systems and future food

security require accessibility to a wide diversity of animal genetic resources. Animal genetic resources are defined as genetic diversity in domesticated animal species having economic or other socio-cultural values and found among species, among animal breeds within the species and in cryoconserved material (embryos and semen). Genetic diversity refers to differences in allele frequencies and allele combinations among breeds of farm animal species and the spectrum of genetic variation within the breeds.

In GPA, climate change was widely recognized as a major challenge for agriculture and food security (FAO, 2007). GPA is based on achievements and common agreements reached at the International Technical Conference on Animal Genetic Resources held in Interlaken, Switzerland in 2007. It includes four priority areas, providing suggestions and guidelines for characterization, sustainable use and conservation of animal genetic resources and institutional capacity building related to these issues.

Howden et al. (2007) suggested several practical approaches that could advance the potential of livestock production systems to adapt to climate change and reduce greenhouse gas (GHG) emissions. These included altered rotation of pasture, modifications of times of grazing and timing of reproduction, alteration of forage crops, adequate water supplies, use of supplementary feeds and concentrates, and reduced need for winter housing in cold climates. Various political options are also available to regulate livestock production and consumption of products of animal origin, which can lead to a reduction in GHG emissions (Gerber et al., 2010; Garnett, 2011). Political regulation can diminish GHG emissions through taxation and subsidies and by promoting new energy-saving technologies and use of cleaner and renewable fuels, by creating a portfolio of products for particular markets, and by influencing consumer behavior. One useful way to diminish GHG is to reduce meat consumption, particularly in rich countries (Garnett, 2011).

However, there are several genetic and animal breeding questions that should be considered in climate change strategies (Hoffman, 2010; Wall et al., 2010; Bruce, 2013), such as choosing among breeds and even among species suited to changing circumstances (Seo et al., 2010). There may be an increased demand for robust animal breeds with the potential to adapt to changes in environmental conditions and tolerate new livestock diseases (Hoffman, 2010). Characterization of breeds with modern genomic tools can be applied to identify breeds that have genetically adapted to marginal circumstances. The genomic data also provide critical information for conservation programs for farm animal genetic resources. All these genetic issues were examined and discussed in the Nordic Research Network on Animal Genetic Resources in the Adaptation to Climate Change (AnGR-NordicNET[1]). AnGR-NordicNET's aims were to provide material, results and conclusions for a Nordic strategy for the conservation, utilization and investigation of animal genetic resources within adaptation and mitigation issues. AnGR-NordicNET was part of the program "Climate Change Impacts, Adaptation and Mitigation in Nordic Primary Industries," which is a thematic research network program developed by the Nordic Council of Ministers (Barua et al., 2014). In this paper, we review some conclusions of AnGR-NordicNET and the current knowledge of climate change effects on the Nordic agro-ecosystems and livestock production and give recommendations for animal breeding that consider adaptation and mitigation issues. Moreover, we discuss the values of animal genetic resources for future breeding work.

[1]https://sites.google.com/a/nordgen.org/angr/home

CHANGES IN AGROCLIMATIC CONDITIONS IN NORTHERN EUROPE

Current climate change, detected as alterations in atmospheric composition, is mainly caused by human activities, e.g., the burning of fossil fuels, urbanization, shifts in land use, agricultural practices, and livestock production (Meehl et al., 2007). N-fertilizer production and application, on-farm use of fossil fuels, clearing forests and other land to grow feed for animals and graze livestock, manure management, manure emissions and processing, and transporting the end products are examples of activities in livestock production that produce GHG (Gill et al., 2010). The changes in atmospheric composition arise from anthropogenic emissions of, e.g., carbon dioxide (CO_2), methane (CH_4) and nitrous oxide (N_2O) (Karl and Trenberth, 2003). According to FAO's report (Steinfeld et al., 2006), globally 18 per cent of anthropogenic GHG emissions are attributable to cattle, sheep, goats and other domestic ruminant species, camels, horses, pigs and poultry. However, the proportion of GHG coming from livestock production can vary nationally and even regionally depending on the density of livestock populations and severity of impacts of livestock production on the environment (Mitloehner, 2010). A review on GHG emissions from livestock production in the Nordic countries is given elsewhere (Åby et al., 2014), showing national emissions from agriculture generally being lower than the global average. Animal production based on ruminants produces CH_4 and N_2O in the main while that of monogastric species produces N_2O (Wall et al., 2010). There is a risk that livestock-related GHG emissions will increase in the future: the human population will continue to grow and demand for animal products will increase both globally and in the Nordic countries (Delgado, 2003; Åby et al., 2014; Gerland et al., 2014). This will lead to an increase in domestic animal populations and the mitigation and adaptation issues related to livestock production will become increasingly important.

The rise of average annual surface temperature, variation in precipitation events, and the increased occurrence of extreme weather events, such as warm periods, heat waves and heavy rainfall, are examples of climate change (Bernstein et al., 2007), all of which have impacts on agriculture and livestock production. The International Panel on Climate Change (IPCC) presented scenarios on trends in climate variables occurring during the 21st Century if no successful actions are taken to diminish GHG emissions. Different biogeographic zones, which are separated according to climatological, biogeographic and geological factors, are assumed to experience different climatic changes (Benestad, 2005; Bernstein et al., 2007; Peel et al., 2007) and climate is changing in slightly different ways also across the Nordic countries. For example, in Denmark, which belongs mostly to the North Atlantic biogeographic zone, characterized by a mild and humid climate, the annual mean temperature is estimated to increase by +2°C during the 21st Century, leading to drier and hotter summers.

Also precipitation in winter is expected to increase (estimation of 0.5 mm/month per decade). In Norway, Sweden and Finland, which mostly belong to the Boreal or North Alpine biogeographic zones, the annual mean temperature is expected to increase by > 3°C over the course of the 21st Century. These regions are characterized by a marked increase in annual precipitation (close to +1 mm/month per decade), wetter winters and risks for floods. The number of days with snow cover and/or frost will be fewer in the future and the snow conditions will not be as reliable as now (Jylhä et al., 2008). In Iceland, a country in the Arctic biogeographic zone with low temperatures, extreme annual variation in sunlight and short intensive growing seasons, climate change is already documented as affecting the distribution of moisture, resulting in shifts in distribution of plants and wildlife animals. The annual surface temperature is expected to increase by 2–4°C, mainly in winter, and precipitation will be as much as 20% higher in many areas (Bernstein et al., 2007).

These general outcomes of climate change will vary even within the northern European biogeographic regions as revealed by so-termed downscaled regional climate models with a spatial resolution of 50 km or less (Benestad, 2005 and references therein; Bernstein et al., 2007). In regions with complex landscape structures, e.g., typical in Norway, a pronounced local pattern in temperature and rainfall can be detected. In the Nordic region, the strongest warming is estimated for the high mountains in southern Norway, and the interior regions of Finland, Sweden and Norway, which all are important dairy production areas. The strongest trends in precipitation are assumed in the regions of Norway that are characterized by abundant sloping geography.

EFFECTS OF CLIMATE CHANGE ON LIVESTOCK PRODUCTION

In the long run, and currently to varying extents, the climate changes described will have various direct and indirect effects on livestock production (Nardone et al., 2010). Air temperature, humidity, air movement, and precipitation are environmental factors that affect daily weather conditions and directly affect animal welfare with the potential to create heat stress (Robinson, 2001; Nardone et al., 2006). In the Nordic countries, the ruminant farm animal species (mainly small ruminants) graze from spring to late autumn (reindeer remain outside all the time, as do honeybees) and are more subject to the direct effects of climate change than the monogastric species, for which farming is more industrialized. Animals can suffer from occasional heat stress during the summer season even in northern Europe. Ravagnolo et al. (2000) estimated, for example, that when the temperature is +25°C and relative humidity 50%, lactating cows are outside of their optimal ambient temperature zone. When relative humidity increases, the threshold temperature decreases. Several energy-requiring physiological and metabolic functions, such as increased respiration, increased water intake and reduced feed intake, are needed to maintain optimal body temperature. These adaptations, however, lead to lower productivity and fertility (Ravagnolo et al., 2000; De Rensis and Scaramuzzi, 2003; West, 2003; Nardone et al., 2006). There are differences among species, among breeds within species, and among individuals within breeds regarding heat stress tolerance. The ruminants' ability to thermoregulate is typically better than that of the monogastrics (Nardone et al., 2006). In addition, modern highly productive farm animal breeds, which typically show increased metabolic heat production may tolerate extreme climatic conditions less well than moderate and low-output breeds (Nardone et al., 2006; Hoffman, 2010 and references therein). Ravagnolo and Misztal (2002) showed that there is genetic variation among individual (Holstein) cows in their heat stress sensitivity, both with respect to milk yield and fertility, and that high-yielding cows were more prone to decrease their production when heat stressed. The heat stress sensitivity for milk yield was, however, not genetically correlated to that for fertility, and the authors hypothesize that different metabolic and physiological processes are responsible for heat tolerance for these two traits.

An already existing problem associated with climate change that has increasingly unfavorable effects on animal welfare and livestock production is the occurrence and frequency of animal diseases (Gale et al., 2009). For example, the spread of Bluetongue disease virus and Schmallenberg virus is evidently associated with climate change (Guis et al., 2012). Bluetongue disease, which is a viral disease in ruminants transmitted by bloodsucking midges (*Culicoides* spp.), has been found in Denmark, Norway and Sweden, but no cases to date have been detected in Finland and Iceland. Schmallenberg virus has spread in all Nordic countries except Iceland[2]. The trend is that future agroenvironments will be more favorable for several diseases than the present-day environments. Global warming and incidence of extreme meteorological events (droughts and increased rainfall) will create, or may already have created, favorable microenvironments for various viruses, their vector species and fungal and bacterial pathogens. The density of insect vectors may also increase as a result of changes in annual behavioral cycles of migratory birds (Gale et al., 2009 and references therein). Global warming may shift timing when insectivorous birds migrate and nest, leading to the loss of synchrony between nesting and peak food abundance for migratory birds (Both et al., 2004).

One additional challenge is that pathogens typically have specific characters that allow rapid spread. For example, RNA-viruses have a high mutation rate and can adapt to new circumstances quickly (Duffy et al., 2008; Gale et al., 2009 and references therein). New viral vector-borne diseases may not necessarily originate from close geographic regions but may come from distant regions, as the history of Bluetongue disease demonstrates (Guis et al., 2012). In general, *Culicoides* spp. are a major threat to animal welfare by spreading viruses that cause serious diseases (Gale et al., 2009). Midges can transmit pathogens and diseases to livestock species from wild species, as exemplified by Epizootic Hemorrhagic Disease that has spread from wild deer to livestock (Savini et al., 2011). Midges are not the only invertebrates that vector many livestock diseases. Ticks, mosquitoes and lymnaeid snails also transmit extremely harmful diseases to livestock (Scott and Smith, 1994; Randolph, 2009; Caron et al., 2014). Moreover, it is expected that the increased annual temperature, milder winters and higher rainfall will improve developmental success of helminth parasites, such as gastrointestinal nematodes and flukes, which will have more pronounced negative effects on the welfare

[2]http://www.nordrisk.dk/

of grazing cattle and sheep, and livestock production in general (van Dijk et al., 2010).

As a result of climate change, animal feeding strategies in the Nordic countries may need modifying. Climate change will have positive impacts on the "domestic" production of fodder plants in the Nordic countries. Plant growth, yield and the production of crop and pasture species will benefit from increases in atmospheric CO_2 concentration, a warmer climate and a longer growing season. However, these agroclimatic changes will also bring new challenges with the expansion of new weeds, insect pests and plant diseases to the northern European regions and problems with overwintering of perennial fodder plants (Tubiello et al., 2007; Hakala et al., 2011; Olesen et al., 2011; Höglind et al., 2013). More chemical control in plant protection, resistant cultivars and plant rotations will be needed to overcome the negative effects of climate change in plant production (Ceccarelli et al., 2010; Hakala et al., 2011). New varieties that tolerate increased precipitation and annual fodder plants, such as maize (*Zea mays* L.), might be commonly cultivated by the end of 21st Century also in Scandinavia and Finland (Olesen et al., 2011). However, the Nordic cultivation traditions for perennial forage grasses are likely to continue (at least in the northernmost and eastern regions), such as perennial ryegrass (*Lolium perenne* L.) and timothy (*Phleum pratense* L.) because perennial plants that are relatively tolerant of less optimal overwintering conditions will be favored even under the changing agroclimatic conditions (Höglind et al., 2013).

In the Nordic countries, imported fodder, mainly proteins and cereal concentrates, is important, particularly in production based on monogastric species (Åby et al., 2014). Feed trading exists both among EU-countries and non-European countries, e.g., Brazil. It is suggested that global warming and extreme meteorological events will decrease crop yields and agricultural productivity in the southern countries, leading to reduced availability and increased prices of grains for animal feeds in the future (Wheeler and Reynolds, 2013). This calls for improvement of self-sufficiency in fodder production in the Nordic countries for future food-security. Such self-sufficiency can be improved by using more fertilizers, pest control chemicals and other inputs in fodder production, and through plant breeding and changes in land-use. However, deforestation of new land for cultivation of fodder plants may face various restrictions owing to international political agreements and for environmental reasons. This may lead to the utilization of less productive marginal land for fodder production and pastures and could provide possibilities to utilize low-input breeds and support their conservation (Sæther et al., 2006). In recent years the trend has been in the opposite direction and fewer pastures than previously have been used to feed cattle (Åby et al., 2014). The socio-economic approaches and subsidy policies should be developed in order to make the use of low-input breeds in animal production a realistic option for farmers.

It appears that higher yields of fodder and pasture plants will lead to increased profitability of animal production in the Nordic countries (Ciscar et al., 2011). The Nordic livestock production systems, however, have to cope with various challenging circumstances in the future, e.g., to improve self-sufficiency in fodder production, as well as to mitigate harmful environmental effects caused by their production. Livestock production is a substantial source of GHG and there is an urgent need to modify the production systems. Diversity in production systems may increase, which calls for matching the genotypes to each system. The use of farm animal genetic resources and animal breeding play a role in this context in finding solutions to new challenges and making livestock production more environmentally friendly.

CHARACTERIZATION OF ANIMALS' CH_4 PRODUCTION

The key issue associated with negative environmental impacts induced by livestock production and mitigation of GHG effects is the reduction of CH_4 emissions from ruminants, especially from beef and dairy cattle (Martin et al., 2010; Wall et al., 2010). There have been several methods used to measure CH_4 concentrations, such as gas chromatography, mass spectroscopy, and a tunable laser diode technique (Johnson and Johnson, 1995). Currently it is common to use automatic advanced technology based on infrared detectors, either in respiration chambers or with more recently developed methods in feeding stations with automatic milking robots (e.g., Garnsworthy et al., 2012; Lassen et al., 2012, 2014). The use of respiration chambers gives highly accurate measurements, but the capacity is limited to a few animals per week. The feeding station methods are less accurate but have higher capacity, up to 60 cows per week per unit, making them suitable for genetic studies at a pilot scale (Lassen et al., 2014).

Microbial fermentation of feed in the rumen produces short-chain fatty acids, such as acetate, propionate and butyrate, which are used as the animal's energy source. This fermentation results in high levels of enteric CH_4 (Martin et al., 2010). It should be pointed out that manipulation of feeding affecting rumen microbial populations is one of the main approaches to decreasing the levels of CH_4 emissions (Boadi et al., 2004; Hook et al., 2010; Martin et al., 2010). For example, increasing the energy density of the diet decreases CH_4 production per unit of digestible energy consumed (Yates et al., 2000). However, this would mean an increase in cereals and other high energy components in cattle feed rations. This can be considered as unwanted in terms of resource utilization in food production for a growing human population and would also mean that the Nordic countries become more dependent on imported feedstuff. In addition to the manipulation of the animals' diets, selective breeding work is the other principal means used to mitigate GHG emissions (Wall et al., 2010; Bruce, 2013).

Feeding experiments in cattle and sheep indicated that there are variations among individual animals in the production of CH_4 when they are fed the same diets. In addition, as reviewed by Wall et al. (2010), there exists variation in CH_4 emissions among individual cattle and among breeds, suggesting potential for improvement of the trait through genetic selection. However, Martin et al. (2010) were less optimistic; they concluded that repeatability of the successive measurements has been low in experiments and is heavily dependent on diet and physiological stage of the animals.

Characterization of individual animal CH_4 emissions for genetic selection is an urgent matter. The COST-action project METHAGENE focuses on the harmonization of CH_4 measurement techniques and develops approaches for incorporating CH_4

emissions into national breeding strategies. Taking into account that methane is a product of rumen microbial fermentation processes that are directly affected by diet, better understanding of animal genome interaction with own rumen microbiome under various feeding conditions needs to be taken into account in drawing mitigation strategies. These subjects are addressed in a number of national and international research projects (e.g., EU-FP7-project RUMINOMICS; REMRUM in Denmark; Rumen Microbial Genomics Network). Results are expected from these projects in the near future.

CHARACTERIZATION OF ANIMALS' ENVIRONMENTAL ADAPTATION

If an animal population survives, is productive and reproduces in a given environment, we can say that this population comprises suitable, adapted phenotypes for that environment. The adaptations, such as disease and heat resistance, water scarcity tolerance and ability to cope with poor quality feed, are valuable characteristics of a breed and have importance when mitigating and adapting to environmental changes (Hoffman, 2010; Mirkena et al., 2010). Breeds can become adapted to specific environments through natural and artificial selection. "Adaptation traits" are complex and often polygenically controlled (Pritchard et al., 2011).

The interactions between genotypes and environments are typically examined in livestock species using quantitative genetics approaches (Falconer and MacKay, 1996). Dense SNP-markers and next-generation-sequencing (NGS) technology can also be used to search for adaptation patterns and selection footprints in animal genomes that result from long-term natural and artificial selection (Harrison et al., 2012; Guo et al., 2014; Lv et al., 2014).

Genotype by environment interaction (G × E) means that genotypes react differently to environmental changes (Falconer and MacKay, 1996). For example, genotype A can perform better and display superior fitness in high altitude regions than genotype B, while at sea level genotype B is the superior phenotype. Or genotype A performs better in both environments but the difference between the two genotypes is larger in one environment than in another. G × E has been an active research field in animal breeding and quantitative genetics. If there is information available on performance of animals over a wide range of environments, it is possible to use a reaction norm approach for estimating breeding values. The reaction norm can predict the performance of an individual in an environment the animal has not been in (Calus et al., 2002; Kolmodin et al., 2002). Reaction norms have up till now mainly been estimated using traditional quantitative genetics, but there is no theoretical reason why they could not also be estimated using molecular genetic information, which would probably lead to greater accuracy in estimating breeding values of young animals (Silva et al., 2014).

From a genomics point of view, adaptations of animal breeds to environments or diets are typically associated with structural and functional genomic variations (Axelsson et al., 2013; Li et al., 2013; Guo et al., 2014; Lv et al., 2014). Dense whole-genome SNP-chips and NGS applications, such as whole genome and mRNA sequencing, analysis of regulatory (miRNAs) elements, and DNA methylation profiles for epigenetic analysis, can be used to investigate genetic background of adaptations in livestock breeds and species (Bartel, 2004; Pritchard et al., 2011; Feil and Fraga, 2012; Harrison et al., 2012; Jiang et al., 2014; Lee et al., 2014). Pairwise comparisons between closely related taxa (for example breeds originating from different environments) provide a powerful approach to identifying loci that show divergence between populations and which may have been under positive selection (Harrison et al., 2012; Li et al., 2013; Jiang et al., 2014; Lv et al., 2014). There is a body of different robust statistical and bioinformatics methods for detecting selection signatures (e.g., Beaumont and Balding, 2004; Joost et al., 2007; Frichot et al., 2013; Wolf, 2013 and many others) that have been successfully used in genome-wide SNP and genomic sequence studies (e.g., Guo et al., 2014; Lv et al., 2014).

Measures of CH_4 concentrations from ruminants and the characterization of individuals and breeds using modern genomic, biometrical and bioinformatic tools play a pivotal role in the implementation of the strategic priority areas of GPA (FAO, 2007), though there is still a need to document the marginal effect of including CH_4 emissions in breeding schemes selecting for efficiency and productivity. With this new information we will understand better characteristics of farm animal genetic resources and can develop animal breeding and sustainable utilization of genetic resources that will make livestock production more environmental friendly.

BREEDING GOALS CONSIDERING CLIMATE CHANGE

Mitigation through selection refers to breeding animals that have high productivity and efficiency, fertility, good health, robustness and that produce less GHG (Boadi et al., 2004; Wall et al., 2010; Bruce, 2013; Hietala et al., 2014). The breeding goals for adaptation are very similar to those for mitigation: in adaptation to new environmental circumstances and production environments, we consider that fertility, feed conversation rate and particularly health traits, are very important. As pointed out in several previous papers, the improvement in productivity (higher average milk and meat yields etc.) means fewer emissions per product. In addition, fewer animals are needed to meet the demand for animal products (e.g., Boadi et al., 2004; Wall et al., 2010; Bruce, 2013). Improving fertility, on the other hand, means shorter unproductive periods, and improving calving and maternal traits, diminishing emissions by improving survival of offspring. Major production traits such as feed conversion rate, fertility, health and other fitness traits have been shown to have a genetic component, demonstrating that there are possibilities to improve them via selection.

The Nordic breeding programs have typically broad breeding goals and both production and health characters are considered (e.g., Miglior et al., 2005; Åby et al., 2013; Hietala et al., 2014). Traits important for mitigation and adaptation are typically either directly or indirectly considered in the Nordic multitrait breeding schemes, which makes it easier to breed animals that are needed for future livestock production (Åby et al., 2013). Currently, fertility, health and other fitness and functional traits have received more attention in breeding goals than previously (e.g., Hietala et al., 2014). This trend can be considered highly recommendable because, for example, both fertility and health traits of dairy cattle

(and several other farm animal species) have deteriorated, especially in populations where the traits have not been considered in the total breeding values (Lucy, 2001; Miglior et al., 2005). Good fertility, e.g., in dairy cows, is known to correlate negatively with genetic merit for milk production (Rauw et al., 1998; van der Waaij, 2004).

One option for including CH_4 production in a future breeding program is to carry out direct selection on the trait. However, to do this, there is a need for phenotypic recording of direct measurements of the trait in many ruminants in several herds in order to create a reference population to estimate genomic breeding values (Hansen Axelsson et al., 2013, 2015). For this to happen there is a need to develop better and cheaper measurement techniques (Hansen Axelsson et al., 2013). More research is in progress in this field and some of it is supported by the EU-COST project METHAGENE. In the meantime, we can improve the trait indirectly through selection of proxy traits that are correlated with CH_4 emissions per unit of product (e.g., milk yield, fertility, feed efficiency, and longevity of the animals; Capper et al., 2009; Bruce, 2013; Hietala et al., 2014).

Moreover, with the advent of relatively cheap SNP-chips with tens or hundreds of thousands of markers, it is also possible to estimate genomic breeding values for animals that have not themselves, nor their close relatives, lived and produced in the environment where they or their offspring are expected to live. Stated differently, it would be possible to find markers that are associated with performance in conditions that we believe we will have in the Nordic countries in the coming decades if we can genetically evaluate animals that currently live under such conditions elsewhere.

AVAILABLE ANIMAL GENETIC RESOURCES

Currently there are three types of livestock breed available in the Nordic countries for future selection programs: (1) the major commercial breeds, (2) the minor breeds, which are typically native breeds and that are also used in commercial herds but more typically in special production situations, and (3) endangered breeds, which are also native breeds and kept for recreational purposes and rarely for production purposes.

The major breeds dominate production systems and they may possess important within-breed genetic variation to select for adaptation to new agro-ecological conditions and mitigation of harmful effects of animal production on climate change (e.g., Gomes da Silva, 1973). It is very important that these breeds do not run into inbreeding problems, otherwise inferior alternative breeds, if they still exist at that time, will have to be introduced into the production system. Inbreeding problems have to be avoided also for the minor and endangered breeds in order to maintain viability over many future generations. Long-term selection experiments have shown that managed populations can be sustained without significant loss of genetic variation for more than 100 generations when the effective population size is maintained at 100 or more (Hill, 2000). However, the effective population size of major commercial breeds is typically much less than 100 (Kantanen et al., 1999; Taberlet et al., 2008). Optimal contribution theory provides a framework for maximizing response to selection while controlling the effective population size (Meuwissen, 1997). Software for optimum contribution selection exists, but improvements are needed in order to address the different situations that occur in practical breeding schemes.

The native breeds have the longest adaptation history to Nordic environmental and production conditions. These breeds are based on ancient animal populations that spread to northern Europe thousands of years ago when the transition from hunting-fishing-gathering livelihoods to animal farming and cultivation began (Kantanen et al., 2000; Bläuer and Kantanen, 2013; Niemi et al., 2013). Therefore, we argue that the Nordic native breeds, which typically are minor and endangered breeds, may possess structural and functional genomic variations for specific traits, such as disease resistances. For example, the native Finncattle display a high level of polymorphism in the Major-Histocompatibility-Complex system (more specifically at the *BoLA-DRB3* locus) that controls a major part of the immune system (Kostia, 2000). The Nordic native cattle breeds exhibit allelic combinations in the casein loci that have a positive impact on processing properties of milk (Lien et al., 1999). In addition, there are several anecdotes about adaptive characters of native breeds that should be scientifically studied and critically evaluated.

Due to climatic changes, the commercial and widespread breeds may show shortcomings in some traits, such as insufficient resistance to a new disease or tolerance to other environmental stress (Nardone et al., 2006; Hoffman, 2010). The minor and endangered breeds may possess genes that code for specific traits, such as disease resistances, which may become desired by the major breed owners, but for which the major breed does not possess the necessary genetic variation. However, the major breeds can be selected for any desired trait, just as the minor or endangered breeds were once selected for this trait, but it may take many generations to establish the desired trait in the major breed. The major breed could benefit from alleles available in minor and endangered breeds using crossing and genomic introgression, and genomic marker information to introgress favorable alleles, while keeping favorable alleles for production traits in the major breed (Ødegård et al., 2009). If the trait is due to a single or a few genes, such genes can be mapped and be introgressed into the commercial breed (Ødegård et al., 2009). Although often successful in plant breeding, this approach is often not feasible for livestock because most livestock traits are complex, i.e., highly polygenic, and introgression takes ∼5 generations, which in livestock might easily be 10 years or more. Crossbreeding systems can be devised that at least partly convey the desired trait from the rare into the commercial breed. Alternatively, a Genomic Selection Introgression approach can be employed (Ødegård et al., 2009), where genomic selection is applied for a rapid introduction of a new trait in the commercial breed.

In this process, minor breeds that possess the trait represent a much more useful resource than the endangered breeds because when crossed with the major breed, their offspring combine the desired trait with commercial viability (since both parental breeds are commercially viable). Examples of this situation are the use of Nordic red bulls on US-Holstein cows to improve their fertility (considering Nordic Reds as a minor breed at the global cattle breeding scale), the use of Chinese Meishan pigs to improve

fertility in some European pig breeding programs and crossing less fertile sheep breeds with the highly prolific Finnsheep.

In general, minor and endangered breeds represent a valuable resource for commercial breeding schemes to increase the rate at which desirable traits can be established in major breeds. Thus, to address future unforeseen production challenges such as climatic changes, which require new, desired traits in major breeds, it is important to maintain a large number of minor breeds that are improved for specialized production environments, and, to a lesser extent, in endangered breeds. In all Nordic countries there are national strategies to conserve both *in vivo* and *in vitro* native breeds and their genetic resources. However, these strategies should be revised, e.g., by considering the geographic distribution of rare breeds within the countries and strengthening cryopreservation of genetic materials. Several native breeds exist in relatively restricted local areas and in the outbreaks of serious animal diseases the whole breed or most of it can be lost. Nordic breeds have been previously analyzed for neutral genetic markers (e.g., Tapio et al., 2006, 2010; Li et al., 2007; Kantanen et al., 2009), but more characterization of conservation values and adaptations is needed in order to promote efficient use of genetic resources in the future. In the AnGr-NordicNET project, a new measure of valuing breeds for conservation, termed "adaptivity coverage" has been developed (Wellman et al., 2014). This quantifies how well a set of breeds could be adapted to wide range of environments within a limited timespan. In this quantification, adaptivity coverage considers both neutral and non-neutral genetic variation.

CONCLUSION

The Nordic multitrait breeding programs for several animal breeds consider directly or indirectly traits that are important for mitigation of environmental effects of livestock production or advance animals' adaptation to new agroecological conditions. The important traits in this context are, for example, productivity in general, fertility, feed conversation rate and health. However, fertility, health and other fitness traits should receive more weight and value in animal breeding to strengthen adaptation potential. Moreover, the breeding programs should maintain high effective population sizes in order to keep high genetic variation in major and minor breeds. Including CH_4 production of ruminants as a trait in breeding programs needs more research and the development of better and cheaper CH_4 measurement techniques. In the future genomic selection and genomic selection introgression approaches may play pivotal roles, particularly in "adaptation breeding." Valuable alleles in terms of adaptation to climate change can be introduced into major breeds from conserved native breeds through genomic selection introgression breeding. Therefore, *in vivo* and *in vitro* conservation of minor and endangered breeds, which are typically native breeds, should be strengthened and their adaptation traits investigated using modern genomic and bioinformatics tools.

AUTHOR CONTRIBUTIONS

All authors have designed the review paper, drafted the manuscript and revised it critically. All authors have approved the version to be published.

ACKNOWLEDGMENTS

The funding given by Nordic Council of Ministers, NordForsk and NordGen—Nordic Genetic Resource Centre is greatly acknowledged.

REFERENCES

Åby, B. A., Aass, L., Sehested, E., and Vangen, O. (2013). Effect of incorporating greenhouse gas emission costs into economic values of traits for intensive and extensive beef cattle breeds. *Livest. Sci.* 158, 1–11. doi: 10.1016/j.livsci.2013.09.009

Åby, B. A., Kantanen, J., and Meuwissen, T. (2014). Future trends in livestock production in the Nordic countries with a changing climate and human population growth. *Acta Agric. Scand. A Anim.* 64, 73–79. doi: 10.1080/09064702.2014.950321

Axelsson, E., Ratnakumar, A., Arendt, M.-L., Maqbool, K., Webster, M. T., Perloski, M., et al. (2013). The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature* 495, 360–365. doi: 10.1038/nature11837

Bartel, D. P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116, 281–297. doi: 10.1016/S0092-8674(04)00045-5

Barua, S. K., Berg, P., Bruvoll, A., Cederberg, C., Drinkwater, K. F., Eide, A., et al. (2014). *Climate Change and Primary Industries. Impacts, Adaptation and Mitigation in the Nordic Countries* (TemaNord 2014: 552). Copenhagen: Nordic Council of Ministers. doi: 10.6027/TN2014-552

Beaumont, M. A., and Balding, D. J. (2004). Identifying adaptive genetic divergence among populations from genome scans. *Mol. Ecol.* 13, 969–980. doi: 10.1111/j.1365-294X.2004.02125.x

Benestad, R. E. (2005). Climate change scenarios for northern Europe from multimodel IPCC AR4 climate simulations. *Geophys. Res. Lett.* 32, L17704. doi: 10.1029/2005GL023401

Bernstein, L., Bosch, P., Canziani, O., Chen, Z., Christ, R., Davidson, O., et al. (2007). *IPCC, 2007: Climate Change 2007: Synthesis Report. Contribution of Working Groups I, II, and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Geneva: Intergovernmental Panel on Climate Change, 52 p. + Appendix.

Bläuer, A., and Kantanen, J. (2013). Transition from hunting to animal husbandry in Southern, Western and Eastern Finland: new dated osteological evidence. *J. Archaeol. Sci.* 40, 1646–1666. doi: 10.1016/j.jas.2012.10.033

Boadi, D., Benchaar, C., Chiquette, J., and Massé, D. (2004). Mitigation strategies to reduce enteric methane emissions from dairy cows: update review. *Can. J. Anim. Sci.* 84, 319–335. doi: 10.4141/A03-109

Both, C., Artemyev, A. V., Blaauw, B., Cowie, R., J., Dekhuijzen, A. J., Eeva, T., et al. (2004). Large-scale geographical variation confirms that climate change causes birds to lay earlier. *Proc. R. Soc. Lond. B* 271, 1657–1662. doi: 10.1098/rspb.2004.2770

Bruce, A. (2013). The lore of low methane livestock: co-producing technology and animals for reduced climate change impact. *Life Sci. Soc. Policy* 9, 10. doi: 10.1186/2195-7819-9-10

Calus, M., Groen, A. F., and De Jong, G. (2002). Genotype × environment interaction for protein yield in Dutch dairy cattle as quantified by different models. *J. Dairy Sci.* 85, 3115–3123. doi: 10.3168/jds.S0022-0302(02)74399-3

Capper, J. L., Cady, R. A., and Bauman, D. E. (2009). The environmental impact of dairy production: 1944 compared with 2007. *J. Anim. Sci.* 87, 2160–2167. doi: 10.2527/jas.2009-1781

Caron, Y., Martens, K., Lempereur, L., Saegerman, C., and Losson, B. (2014). New insight in lymnaeid snails (Mollusca, Gastropoda) as intermediate hosts of *Fasciola hepatica* (Trematoda, Digenea) in Belgium and Luxembourg. *Parasit. Vectors* 7:66. doi: 10.1186/1756-3305-7-66

Ceccarelli, S., Grando, S., Maatougui, M., Michael, M., Slash, M., Haghparast, R., et al. (2010). Plant breeding and climate change. *J. Agric. Sci.* 148, 627–637. doi: 10.1017/S0021859610000651

Ciscar, J.-C., Iglesias, A., Feyen, L., Szabó, L., Van Regemorter, D., Amelung B., et al. (2011). Physical and economic consequences of climate change in Europe. *PNAS* 108, 2678–2683. doi: 10.1073/pnas.1011612108

De Rensis, F., and Scaramuzzi, R. J. (2003). Heat stress and seasonal effects on reproduction in the dairy cow—a review. *Theriogenology* 60, 1139–1151. doi: 10.1016/S0093-691X(03)00126-2

Delgado, C. L. (2003). Rising consumption of meat and milk in developing countries has created a new food revolution. *J. Nutr.* 133, 3907S–3910S.

Duffy, S., Shackelton, L. A., and Holmes, E. C. (2008). Rates of evolutionary change in viruses: patterns and determinants. *Nat. Rev. Genet.* 9, 267–276. doi: 10.1038/nrg2323

Falconer, D. S., and MacKay, T. F. C. (1996). *Introduction to Quantitative Genetics*. Harlow: Longman Group Ltda.

FAO. (2007). *Global Plan of Action for Animal Genetic Resources and Interlaken Declaration*. Rome: Food and Agriculture Organization of the United Nations.

Feil, R., and Fraga, M. F. (2012). Epigenetics and the environment: emerging patterns and implications. *Nat. Rev. Genet.* 13, 97–109. doi: 10.1038/nrg3142

Frichot, E., Schoville, S. D., Bouchard, G., and Francois, O. (2013). Testing for associations between loci and environmental gradients using latent factor mixed models. *Mol. Biol. Evol.* 30, 1687–1699. doi: 10.1093/molbev/mst063

Gale, P., Drew, T., Phipps L. P., David, G., and Wooldridge, M. (2009). The effect of climate change on the occurrence and prevalence of livestock diseases in Great Britain: a review. *J. Appl. Microbiol.* 106, 1409–1423. doi: 10.1111/j.1365-2672.2008.04036.x

Garnett, T. (2011). Where are the best opportunities for reducing greenhouse gas emissions in the food system (including the food chain)? *Food Policy* 36, S23–S32. doi: 10.1016/j.foodpol.2010.10.010

Garnsworthy, P. C., Craigon, J., Hernandez-Medrano, J. H., and Saunders, N. (2012). Variation among individual dairy cows in methane measurements made on farm during milking. *J. Dairy Sci.* 95, 3181–3189. doi: 10.3168/jds.2011-4606

Gerber, P., Key, N., Portet, F., and Steinfeld, H. (2010). Policy options in addressing livestock's contribution to climate change. *Animal* 4, 393–406. doi: 10.1017/S1751731110000133

Gerland, P., Raftery, A. E., Ševčíková, H., Li, N., Gu, D., and Spoorenberg, T. (2014). World population stabilization unlikely this century. *Science* 346, 234–237. doi: 10.1126/science.1257469

Gill, M., Smith, P., and Wilkinson, J. M. (2010). Mitigating climate change: the role of domestic livestock. *Animal* 4, 323–333. doi: 10.1017/S1751731109004662

Gomes da Silva, R. (1973). Improving tropical beef cattle by simultaneous selection for weight and heat tolerance. Heritabilities and correlations of the trait. *J. Anim. Sci.* 37, 637–642.

Guis, H., Caminade, C., Calvete, C., Morse, A. P., Tran, A., and Baylis, M. (2012). Modelling the effects of past and future climate on the risk of bluetongue emergence in Europe. *J. R. Soc. Interface* 9, 339–350. doi: 10.1098/rsif.2011.0255

Guo, X., Wang, Z., Li, N., Qiu, F., Xu, Z., Yan, D., et al. (2014). Whole genome sequencing of six dog breeds from continuous altitudes reveals adaptation to high-altitude hypoxia. *Gen. Res.* 24, 1308–1315. doi: 10.1101/gr.171876.113

Hakala, K., Hannukkala, A. O., Huusela-Veistola, E., Jalli, M., and Peltonen-Sainio, P. (2011). Pests and diseases in a changing climate: a major challenge for Finnish crop production. *Agric. Food Sci.* 20, 3–14. doi: 10.2137/145960611795163042

Hansen Axelsson, H., Fikse, W. F., Kargo, M., Sørensen, A. C., Johansson, K., and Rydhmer, L. (2013). Genomic selection using indicator traits to reduce the environmental impact of milk production. *J. Dairy Sci.* 96, 5306–5314. doi: 10.3168/jds.2012-6041

Hansen Axelsson, H., Thomasen, J. R., Sørensen, A. C., Rydhmer, L., Kargo, M., Johansson, K., et al. (2015). Breakeven prices for recording of indicator traits to reduce the environmental impact of milk production. *J. Anim. Breed. Genet.* 132, 30–41. doi: 10.1111/jbg.12106

Harrison, P. W., Wright, A. E., and Mank, J. E. (2012). The evolution of gene expression and the transcriptome-phenotype relationships. *Semin. Cell Dev. Biol.* 23, 222–229. doi: 10.1016/j.semcdb.2011.12.004

Hietala, P., Wolfová, M., Wolf, J., Kantanen, J., and Juga, J. (2014). Economic values of production and functional traits, including residual feed intake, in Finnish milk production. *J. Dairy Sci.* 97, 1092–1106. doi: 10.3168/jds.2013-7085

Hill, W. G. (2000). Maintenance of quantitative genetic variation in animal breeding programmes. *Livest. Prod. Sci.* 63, 99–109. doi: 10.1016/S0301-6226(99)00115-3

Hoffman, I. (2010). Climate change and the characterization, breeding and conservation of animal genetic resources. *Anim. Genet.* 41(Suppl. 1), 32–46. doi: 10.1111/j.1365-2052.2010.02043.x

Höglind, M., Thorsen, S. M., and Semenov, M. A. (2013). Assessing uncertainties in impact of climate change on crass production in Northern Europe using ensembles of global climate models. *Agric. Forest Meteorol.* 170, 103–113. doi: 10.1016/j.agrformet.2012.02.010

Hook, S. E., Wright, A. G., and McBride, B. W. (2010). Methanogens: methane producers of the rumen and mitigation strategies. *Archaea* 2010, 945785. doi: 10.1155/2010/945785

Howden, S. M., Soussana, J.-F., Tubiello, F. N., Chhetri, N., Dunlop, M., and Meinke, H. (2007). Adapting agriculture to climate change. *PNAS* 104, 19691–19696. doi: 10.1073/pnas.0701890104

Jiang, Y., Xie, M., Chen, W., Talbot, R., Maddox, J., Faraut, T., et al. (2014). The sheep genome illuminates biology of the rumen and lipid metabolism. *Science* 344, 1168–1173. doi: 10.1126/science.1252806

Johnson, K. A., and Johnson, D. E. (1995). Methane emissions from cattle. *J. Anim. Sci.* 73, 2483–2492.

Joost, S., Bonin, M. W., Bruford, M., Després, L., Conord, C., Erhardt, G., et al. (2007). A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. *Mol. Ecol.* 16, 3955–3969. doi: 10.1111/j.1365-294X.2007.03442.x

Jylhä, K., Fronzek, S., Tuomenvirta, H., Carter, T. R., and Ruosteenoja, K. (2008). Changes in frost, snow and Baltic Sea ice by the end of the twenty-first century based on climate model projections for Europe. *Clim. Change* 86, 441–462. doi: 10.1007/s10584-007-9310-z

Kantanen, J., Edwards, C. J., Bradley, D. G., Viinalass, H., Thessler, S., Ivanova, Z., et al. (2009). Maternal and paternal genealogy of Eurasian taurine cattle (*Bos taurus*). *Heredity* 103, 404–415. doi: 10.1038/hdy.2009.68

Kantanen, J., Olsaker, I., Adalsteinsson, S., Sandberg, K., Eythorsdottir, E., Pirhonen, K., et al. (1999). Temporal changes in genetic variation of North European cattle breeds. *Anim. Genet.* 30, 16–27. doi: 10.1046/j.1365-2052.1999.00379.x

Kantanen, J., Olsaker, I., Holm, L.-E., Lien, S., Vilkki, J., Brusgaard, K., et al. (2000). Genetic diversity and population structure of 20 North European cattle breeds. *J. Hered.* 91, 446–457. doi: 10.1093/jhered/91.6.446

Karl, T. R., and Trenberth, K. E. (2003). Modern global climate change. *Science* 302, 1719–1723. doi: 10.1126/science.1090228

Kolmodin, R., Strandberg, E., Madsen, P., Jensen, J., and Jorjani, H. (2002). Genotype by environment interaction in Nordic dairy cattle studied by use of reaction norms. *Acta Agric. Scand. A Anim.* 52, 11–24. doi: 10.1080/09064700252806380

Kostia, S. (2000). *Genomic Evolution and Diversity in Artiodactyla*. Ph.D. thesis, Department of Biosciences, Division of Genetics, University of Helsinki, Helsinki. 41.

Lassen, J., Garnsworthy, P. C., Chagunda, M. G. G., Negussie, E., Løvendahl, P., and de Haas, Y. (2014). "Progress with genetic selection for low methane traits in dairy cows," in *Proceedings, 10th World Congress of Genetics Applied to Livestock Production*, Vancouver. Available at: https://asas.org/docs/default-source/wcgalp-proceedings-oral/036_paper_9384_manuscript_600_0.pdf?sfvrsn=2

Lassen, J., Løvendahl, P., and Madsen, J. (2012). Accuracy of noninvasive breath methane measurements using Fourier transform infrared methods on individual cows. *J. Dairy Sci.* 95, 890–898. doi: 10.3168/jds.2011-4544

Lee, J.-R., Hong, C. P., Moon, J.-W., Jung, Y.-D., Kim, D.-S., Kim, T.-H., et al. (2014). Genome-wide analysis of DNA methylation patterns in horse. *BMC Genom.* 15:598. doi: 10.1186/1471-2164-15-598

Li, M.-H., Tapio, I., Vilkki, J., Ivanova, Z., Kiselyova, T., Marzanov, N., et al. (2007). Genetic structure of cattle populations (*Bos taurus*) in northern Eurasia and the neighboring Near Eastern regions: implications for breeding strategies and conservation. *Mol. Ecol.* 16, 3839–3853. doi: 10.1111/j.1365-294X.2007.03437.x

Li, M., Tian, S., Jin, L., Zhou, G., Li, Y., Zhang, Y., et al. (2013). Genomic analyses identify distinct patterns of selection in domesticated pigs and Tibetan wild boars. *Science* 45, 1431–1439. doi: 10.1038/ng.2811

Lien, S., Kantanen, J., Olsaker, I., Holm, L.-E., Eythorsdottir, E., Sandberg, K., et al. (1999). Comparison of milk protein allele frequencies in Nordic cattle breeds. *Anim. Genet.* 30, 85–91. doi: 10.1046/j.1365-2052.1999.00434.x

Lucy, M. C. (2001). Reproductive loss in high-producing dairy cattle: where will it end? *J. Dairy Sci.* 84, 1277–1293. doi: 10.3168/jds.S0022-0302(01)70158-0

Lv, F.-H., Agha, S., Kantanen, J., Colli, L., Stucki, S., Kijas, J. W., et al. (2014). Adaptations to climate-mediated selective pressures in sheep. *Mol. Biol. Evol.* 31, 3324–3343. doi: 10.1093/molbev/msu264

Martin, C., Morgavi, D. P., and Doreau, M. (2010). Methane mitigation in ruminants: from microbe to the farm scale. *Animal* 4, 351–365. doi: 10.1017/S1751731109990620

McMichael, A. J., Powles, J. W., Butter, C. D., and Uauy, R. (2007). Food, livestock production, energy, climate change, and health. *Lancet* 370, 1253–1263. doi: 10.1016/S0140-6736(07)61256-2

Meehl, G. A., Stocker, T. F., Collins, W. D., Friedlingstein, P., Gaye, A. T., Gregory, J. M., et al. (2007). "Global climate projections," in *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment*

Report of the Intergovernmental Panel on Climate Change, eds S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, et al. (Cambridge: Cambridge University Press), 747–845.

Meuwissen, T. (1997). Maximizing the response of selection with a predefined rate of inbreeding. *J. Anim. Sci.* 75, 934–940.

Miglior, F., Muir, B. L., and Van Doormaal, B. J. (2005). Selection indices in holstein cattle of various countries. *J. Dairy Sci.* 88, 1255–1263. doi: 10.3168/jds.S0022-0302(05)72792-2

Mirkena, T., Duguma, G., Haile, A., Tibbo, M., Okeyo, A. M., Wurzinger, M., et al. (2010). Genetics of adaptation in domestic farm animals: a review. *Livest. Sci.* 132, 1–12. doi: 10.1016/j.livsci.2010.05.003

Mitloehner, F. M. (2010). Is the rising demand for animal protein fuelling climate change? *J. Anim. Breed. Genet.* 127, 421–422. doi: 10.1111/j.1439-0388.2010.00909.x

Nardone, A., Ronchi, B., Lacetera, N., and Bernabucci, U. (2006). Climatic effects on productive traits in livestock. *Vet. Res. Commun.* 30(Suppl. 1), 75–81. doi: 10.1007/s11259-006-0016-x

Nardone, A., Ronchi, B., Lacetera, N., Ranieri, M. S., and Bernabucci, U. (2010). Effects of climate changes on animal production and sustainability of livestock systems. *Livest. Sci.* 130, 57–69. doi: 10.1016/j.livsci.2010.02.011

Niemi, J., and Ahlstedt, J. (eds). (2014). *Finnish Agriculture and Rural Industries 2014*. Jokioinen: MTT Economic Research 115a.

Niemi, M., Bläuer, A., Iso-Touru, T., Nyström, V., Harjula, J., Taavitsainen, J.-P., et al. (2013). Mitochondrial DNA and Y-chromosomal diversity in ancient populations of domestic sheep (*Ovis aries*) in Finland: comparison with contemporary sheep breeds. *Genet. Sel. Evol.* 45, 2. doi: 10.1186/1297-9686-45-2

Ødegård, J., Sonesson, A. K., Yazdi, M. H., and Meuwissen, T. H. E. (2009). Introgression of a major QTL from an inferior into a superior population using genomic selection. *Genet. Sel. Evol.* 41, 38. doi: 10.1186/1297-9686-41-38

Olesen, J. E., and Bindi, M. (2002). Consequences of climate change for European agricultural productivity, land use and policy. *Eur. J. Agron.* 16, 239–262. doi: 10.1016/S1161-0301(02)00004-7

Olesen, J. E., Trnka, M., Kersebaum, K. C., Skjelvåg, A. O., Sequin, B., Peltonen-Sainio, P., et al. (2011). Impacts and adaptation of European crop production systems to climate change. *Eur. J. Agron.* 34, 96–112. doi: 10.1016/j.eja.2010.11.003

Peel, M. C., Finlayson, B. L., and McMahon, T. A. (2007). Updated world map of the Köppen-Geiger climate classification. *Hydrol. Earth Syst. Sci.* 11, 1633–1644. doi: 10.5194/hess-11-1633-2007

Pilling, D., and Hoffman, I. (2011). *Climate Change and Animal Genetic Resources for Food and Agriculture: State of Knowledge, Risks and Opportunities*. Background Study Paper No. 53. Rome: Commission on Genetic Resources for Food and Agriculture.

Pritchard, J. K., Pickrell, J. K., and Coop, G. (2011). The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Curr. Biol.* 20, R208–R215. doi: 10.1016/j.cub.2009.11.055

Randolph, S. E. (2009). Tick-borne disease systems emerge from the shadows: the beauty lies in molecular detail, the message in epidemiology. *Parasitology* 136, 1403–1413. doi: 10.1017/S0031182009005782

Rauw, W. M., Kanis, E., Noordhuizen-Stassen, E. N., and Grommers, F. J. (1998). Undesirable side effects of selection for high production efficiency in farm animals: a review. *Livest. Prod. Sci.* 56, 15–33. doi: 10.1016/S0301-6226(98)00147-X

Ravagnolo, O., and Misztal, I. (2002). Effect of heat stress on nonreturn rate in Holstein cows: genetic analyses. *J. Dairy Sci.* 85, 3092–3100. doi: 10.3168/jds.S0022-0302(02)74396-8

Ravagnolo, O., Misztal, I., and Hoogenboom, G. (2000). Genetic component of heat stress in dairy cattle, development of heat index function. *J. Dairy Sci.* 83, 2120–2125. doi: 10.3168/jds.S0022-0302(00)75094-6

Robinson, P. J. (2001). On the Definition of a Heat Wave. *J. Appl. Meteorol. Clim.* 40, 762–775. doi: 10.1175/1520-0450(2001)040<0762:OTDOAH>2.0.CO;2

Savini, G., Afonso, A., Mellor, P., Aradaib, I., Yadin, H., and Sanaa, M. (2011). Epizootic hemorrhagic disease. *Res. Vet. Sci.* 91, 1–17. doi: 10.1016/j.rvsc.2011.05.004

Seo, S. N., McCarl, B. A., and Mendelsohn, R. (2010). From beef to sheep under global warming? An analysis of adaptation by livestock species choice in South America. *Ecol. Econ.* 69, 2486–2494. doi: 10.1016/j.ecolecon.2010.07.025

Scott, M. E., and Smith, G. (1994). *Parasitic and Infectious Diseases: Epidemiology and Ecology*. San Diego: Academic Press.

Silva, F. F., Mulder, H. A., Knol, E. F., Lopes, M. S., Guimarães, S. E. F., Lopes, P. S., et al. (2014). Sire evaluation for total number born in pigs using a genomic reaction norms approach. *J. Anim. Sci.* 92, 3825–3834. doi: 10.2527/jas.2013-6486

Steinfeld, H., Gerber, P., Wassenaar, T., Castel, V., Rosales, M., and de Haan, C. (2006). *Livestock's Long Shadow*. Rome: The Food and Agricultural Organization of the United Nations (FAO).

Sæther, N. H., Sickel, H., Norderhaug, A., Sickel, M., and Vangen, O. (2006). Plant and vegetation preferences for a high and a moderate yielding Norwegian dairy cattle breed grazing semi-natural mountain pastures. *Anim. Res.* 55, 363–387. doi: 10.1051/animres:2006033

Taberlet, P., Valentini, A., Rezaei, H. R., Naderi, S., Pompanon, S., Negrini, R., et al. (2008). Are cattle, sheep, and goats endangered species? *Mol. Ecol.* 17, 275–284. doi: 10.1111/j.1365-294X.2007.03475.x

Tapio, I., Värv, S., Bennewitz, J., Maleviciute, J., Fimland, E., Grislis, Z., et al. (2006). Prioritization for conservation of northern European cattle breeds based on analysis of microsatellite data. *Conserv. Biol.* 20, 1768–1779. doi: 10.1111/j.1523-1739.2006.00488.x

Tapio, M., Ozerov, M., Tapio, I., Toro, M. A., Marzanov, N., Ćinkulov, M., et al. (2010). Microsatellite-based genetic diversity and population structure of domestic sheep in northern Eurasia. *BMC Genet.* 11:76. doi: 10.1186/1471-2156-11-76

Tubiello, F. N., Soussana, J.-F., and Howden, S. M. (2007). Crop and pasture response to climate change. *PNAS* 104, 19686–19690. doi: 10.1073/pnas.0701728104

van der Waaij, E. H. (2004). A resource allocation model describing consequences of artificial selection under metabolic stress. *J. Anim. Sci.* 82, 973–981.

van Dijk, J., Sargison, N. D., Kenyon, F., and Skuce, P. J. (2010). Climate change and infectious disease helminthological challenges to farmed ruminants in temperate regions. *Animal* 4, 377–392. doi: 10.1017/S1751731109990991

Wall, E., Simm, G., and Moran, D. (2010). Developing breeding schemes to assist mitigation of greenhouse gas emissions. *Animal* 4, 366–376. doi: 10.1017/S175173110999070X

Wellman, R., Bennewitz, J., and Meuwissen, T. (2014). A unified approach to characterize and conserve adaptive and neutral genetic diversity in subdivided populations. *Genet. Res.* 96:e16. doi: 10.1017/S0016672314000196

West, J. W. (2003). Effects of heat-stress on production in dairy cattle. *J. Dairy Sci.* 86, 2131–2144. doi: 10.3168/jds.S0022-0302(03)73803-X

Wheeler, T., and Reynolds, C. (2013). Predicting the risks from climate change to forage and crop production for animal feed. *Anim. Front.* 3:36–41. doi: 10.2527/af.2013-0006

Wolf, J. B. (2013). Principles of transcriptome analysis and gene expression quantification: an RNA-seq tutorial. *Mol. Ecol. Res.* 13, 559–572. doi: 10.1111/1755-0998.12109

Yates, C. M., Cammell, S. B., France, J., and Beever, D. E. (2000). Predictions of methane emissions from dairy cows using multiple regression analysis. *Proc. Br. Soc. Anim. Sci.* 2000:94.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 31 October 2014; accepted: 05 February 2015; published online: 25 February 2015.

*Citation: Kantanen J, Løvendahl P, Strandberg E, Eythorsdottir E, Li M-H, Kettunen-Præbel A, Berg P and Meuwissen T (2015) Utilization of farm animal genetic resources in a changing agro-ecological environment in the Nordic countries. Front. Genet. **6**:52. doi: 10.3389/fgene.2015.00052*

This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics.

Copyright © 2015 Kantanen, Løvendahl, Strandberg, Eythorsdottir, Li, Kettunen-Præbel, Berg and Meuwissen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

DEMOGRAPHIC EVENTS AND DIVERSITY IN CATTLE

frontiers in
GENETICS

REVIEW ARTICLE
published: 10 February 2015
doi: 10.3389/fgene.2015.00036

Genomic data as the "hitchhiker's guide" to cattle adaptation: tracking the milestones of past selection in the bovine genome

Yuri T. Utsunomiya[1], Ana M. Pérez O'Brien[2], Tad S. Sonstegard[3], Johann Sölkner[2] and José F. Garcia[1,4]*

[1] *Departamento de Medicina Veterinária Preventiva e Reprodução Animal, Faculdade de Ciências Agrárias e Veterinárias, Universidade Estadual Paulista (UNESP), Jaboticabal, São Paulo, Brazil*
[2] *Division of Livestock Sciences, Department of Sustainable Agricultural Systems, University of Natural Resources and Life Sciences (BOKU), Vienna, Austria*
[3] *Animal Genomics and Improvement Laboratory, Agricultural Research Service, United States Department of Agriculture, Beltsville, MA, USA*
[4] *Laboratório de Bioquímica e Biologia Molecular Animal, Departamento de Apoio, Saúde e Produção Animal, Faculdade de Medicina Veterinária de Araçatuba, Universidade Estadual Paulista (UNESP), Araçatuba, São Paulo, Brazil*

Edited by:
Stéphane Joost, École Polytechnique Fédérale de Lausanne, Switzerland

Reviewed by:
Kwan-Suk Kim, Chungbuk National University, Korea (South)
Francois Pompanon, University Grenoble Alpes, France

Correspondence:
José F. Garcia, Laboratório de Bioquímica e Biologia Molecular Animal, Departamento de Apoio, Saúde e Produção Animal, Faculdade de Medicina Veterinária de Araçatuba, Universidade Estadual Paulista (UNESP), Araçatuba, Rua Clóvis Pestana 793, Araçatuba, São Paulo 16050680, Brazil
e-mail: jfgarcia@fmva.unesp.br

The bovine species have witnessed and played a major role in the drastic socio-economical changes that shaped our culture over the last 10,000 years. During this journey, cattle "hitchhiked" on human development and colonized the world, facing strong selective pressures such as dramatic environmental changes and disease challenge. Consequently, hundreds of specialized cattle breeds emerged and spread around the globe, making up a rich spectrum of genomic resources. Their DNA still carry the scars left from adapting to this wide range of conditions, and we are now empowered with data and analytical tools to track the milestones of past selection in their genomes. In this review paper, we provide a summary of the reconstructed demographic events that shaped cattle diversity, offer a critical synthesis of popular methodologies applied to the search for signatures of selection (SS) in genomic data, and give examples of recent SS studies in cattle. Then, we outline the potential and challenges of the application of SS analysis in cattle, and discuss the future directions in this field.

Keywords: positive selection, *Bos taurus*, *Bos indicus*, adaptation, single nucleotide polymorphism

INTRODUCTION

The ability of domestic cattle to convert low-quality forage into meat, milk, and draft power is of direct importance to the livelihood of the human species. This ability is tightly linked to the adaptation of indigenous and exotic cattle to the diverse environments found around the world, which may result from complex—mostly untold—stories of migration, expansion, exposure to diseases, admixture, climate changes, and selective pressures (Ajmone-Marsan, 2010). These past events have shaped the genetic diversity of domestic cattle throughout history, and their present genomes may shelter tractable signatures of these phenomena.

Footprints of selection, such as specific patterns of change in allele frequencies, diversity loss, and haplotype structure, are currently detectable from single nucleotide polymorphism (SNP) data by well-established methodologies (Sabeti et al., 2006; Oleksyk et al., 2010), and can unravel past responses of the cattle genome to natural and human-driven selection, as well as evidences of loci and variants underlying adaptive and economically important traits. Detecting these selection signatures (SS) may not only help to shed some light on the key adaptive events that have generated the enormous phenotypic variation observed between cattle breeds today, but can also be of biotechnological relevance.

In recent years, SS studies are becoming increasingly popular because they offer a complementary strategy to genome-wide association (GWA) studies on mapping variants impacting traits of interest, helping to link phenotypes to gene function. In a typical GWA analysis, one starts from a phenotype and scan genotypes to identify underlying large and moderate effect variants (Bush and Moore, 2012). Generally, SS studies take the opposite direction: one starts with an evidence of selection in samples sharing geographical proximity, environmental factors or a common phenotype, and attempts to find selected mutations (Sabeti et al., 2006).

The major motivation for SS studies is that this type of approach can pinpoint chromosomal segments sheltering large effect mutations even if they no longer segregate in a population. In such cases, these variants cannot be detected by classical quantitative genetics methods unless linkage experiments are designed using crosses (Ramey et al., 2013). Another appealing feature of SS studies is that they typically require over 10 fold smaller sample sizes in comparison to GWA studies. Moreover, SS can reveal signals on genes controlling traits that are difficult, expensive or

even impossible to measure on a large population (i.e., disease resistance).

Although identifying SS is of paramount importance to uncover variants responsible for adaptive traits, its application to cattle data must be carefully interpreted as important demographic events such as severe population bottlenecks, genetic drift, and admixture, as well as confounding effects derived from the development of SNP panels, may give rise to false signals. Furthermore, SS have serious limitations in targeting specific traits, so assigning signals to phenotypes is a non-trivial problem.

Here, we provide a brief review of the demographic history of cattle as it is known, present a critical review of some of the main methodologies available for the detection of putative loci under selection, and provide examples of recently published results in cattle. Then, we outline the potential and the challenges of the application of these methods to cattle data, and discuss the future directions in this field.

CATTLE DEMOGRAPHIC HISTORY

DIFFERENTIATION AND DOMESTICATION

The humpless taurine (*Bos taurus*) and the humped indicine or zebu cattle (*Bos indicus*) descend from the wild ox or auroch (*Bos primigenius*), which has been extinct since 1627 (Mona et al., 2010). The two populations of wild aurochsen that formed the ancestral pool for these interfertile cattle species may have diverged some 280,000 years ago (Murray et al., 2010), and were subjected to many demographic events before being domesticated by our species, including severe bottlenecks, admixture and natural selection.

Although the domestication of taurine and zebu cattle is still an open question and an active field of investigation, evidences collected over 20 years of research on molecular genetic diversity of cattle (see Groeneveld et al., 2010, for further review), combined with historic and archaeological data, support that these species were independently domesticated in at least two episodes (Loftus et al., 1994; Bruford et al., 2003): some 10,000 years ago, taurine animals were captured from the wild in the Fertile Crescent (modern-day countries of Israel, Jordan, Lebanon, Cyprus and Syria, and parts from Egypt, Turkey, Iraq, Iran, and Kuwait); 1500 years later, zebu cattle were domesticated in the Indus valley (present Northeastern Afghanistan, Pakistan, and Northwestern India).

EXPANSION

Taurine cattle have an almost cosmopolitan distribution today. From the domestication centre in Southwestern Asia, they followed human migrations and slowly expanded over Asia and Europe. An independent domestication episode of taurine cattle in Northeastern Africa is disputed, but as molecular data are not conclusive, the divergence of African and Eurasian taurine populations from a common ancestor domesticated in the Fertile Crescent is well supported (Hanotte et al., 2002; Ajmone-Marsan, 2010; Decker et al., 2014). Other isolated micro-events of domestication are also not discarded, not even interbreeding of domestic animals with wild aurochsen (Bonfiglio et al., 2010). Taurine cattle reached the New World by European importations after the discovery of America in the late fifteenth century, and their descendants living today are broadly referred as Creole cattle.

Zebu cattle also spread around the world accompanying human migrations, but became more endemic in tropical and subtropical regions due to their adaptability to these environments. Zebu cattle were probably introduced in the African continent in the seventh century by imports of *B. indicus* sires during Arabian invasions (Bradley et al., 1996). Indian zebu cattle were only introduced in Central and South America in the early twentieth century, and started to be massively imported to the American continent after 1950.

FORMATION OF SPECIALIZED BREEDS AND FURTHER CATTLE ADAPTATION

After domestication, farmers started to control cattle mating according to their interest in traits such as size, behavior, and milk production. This "breed the best to the best and hope for the best" strategy exerted a high artificial selective pressure, triggering a severe decline in cattle effective population size. Estimates from a genome-wide linkage disequilibrium analysis using a medium density SNP panel suggested that domestication was responsible for a 50 to 70-fold decline in the effective population size in comparison to the wild ancestors of taurine and zebu cattle (The Bovine HapMap Consortium, 2009).

About 200 years ago, farmers invented the concept of breed and started to limit germplasm exchange in order to standardize cattle populations based on morphology and performance, giving rise to over 1200 cataloged cattle breeds today (Taberlet et al., 2008; see also: http://www.ansi.okstate.edu/breeds/cattle/). In spite of this apparent rich source of genetic resources, over 200 additional breeds are already extinct, and many others are at risk (FAO, 2007).

Due to their high productive performance for milk and meat traits, some cattle breeds were adopted worldwide, such as the Holstein-Friesian, which is present in 128 countries (FAO, 2007). However, most of these specialized groups of genetically distinct animals are local, and even though they are not under the spotlights, they exhibit high adaption to their environments. Therefore, small local breeds should be considered as primary targets of SS studies, as several of them are endangered and their genomes shelter footprints of adaptation.

IMPLICATIONS TO POSITIVE SELECTION MAPPING

In theory, neutrally evolving alleles form the majority of the genome variation within and between species, and events such as population contraction, expansion, migration, isolation and admixture are responsible for random drift of such alleles. While the demographic history of a population determines its neutral allele frequency spectrum, alleles that impact fitness, survival, reproduction, or traits of human interest are subjected to natural or artificial selection, such that their frequencies deviate from the distribution of neutral alleles. Therefore, mapping loci under selection implicates searching for outlier alleles that substantially differ from the genome background.

The idea of outlier loci is dependent upon knowledge about the frequency distribution of neutrally evolving alleles. As different populations have distinct demographic histories, their allele frequency spectrum also differ. Consequently, statistical methods designed to map loci under selection must be calibrated to either a demographic model for the population under study or

to the empirical distribution of scores across the genome, assuming most of the genome is evolving under neutrality. As most of the events that shaped cattle diversity at the species and breed level are still obscure, methods that are robust across different demographic scenarios that take advantage of genome-wide scores to detect outlier loci are more appropriate to analyze cattle data. Nevertheless, one can use the demographic model for cattle differentiation, domestication and expansion provided in this section and combine with specific models for specialized cattle breeds to simulate the distribution under neutrality and compare with empirical data.

METHODS FOR DETECTING POSITIVE SELECTION IN THE CATTLE GENOME

As proposed by Darwin and Wallace (1858), positive selection is the phenomenon where phenotypes that increase the likelihood of survival and reproduction (i.e., that increase an individual's fitness) become more prevalent in populations over time. In the context of the genome sequence, if a specific allele confers advantage, its carrier is more likely to thrive and leave more offspring than non-carriers, causing the haplotype containing that beneficial allele to spread quickly and increase in frequency in the population (Sabeti et al., 2002, 2006).

The majority of the genetic variation found within and between populations is deemed to have little or no effect on fitness, so that haplotype frequencies follow Hardy-Weinberg expectations (Kimura, 1968; Hellmann et al., 2003). Therefore, positive selection leaves distinctive tractable patterns of genetic variation that differ from the neutrally evolving background DNA sequence. These patterns are broadly referred as "signatures" or "footprints" of selection (Oleksyk et al., 2010).

In this section, we describe classes of signatures in terms of different signals. In each class, whenever convenient, we describe the basis of some popularly adopted methods to highlight their strengths and weaknesses for cattle studies. For a broader overview, we suggest the reader to consult other previously published reviews on the topic (Sabeti et al., 2006; Oleksyk et al., 2010; Vitti et al., 2013; Qanbari and Simianer, 2014). Additionally, we provide examples of studies reporting putative loci under selection in different cattle breeds.

LOCAL GENETIC DIVERSITY DEPRESSION

At each generation, recombination shuffles and breaks haplotypes down, producing linkage equilibrium. When selection increases the frequency of an advantageous allele, neighbor variants "hitchhike" and rise in frequency together so quickly that recombination does not prevent linkage disequilibrium, causing an entire chromosome segment to lose diversity. Therefore, positive selection can be probed by searching for chromosome regions where heterozygosity is much lower than expected under neutrality (Oleksyk et al., 2008).

Ramey et al. (2013) scanned Illumina® BovineSNP50 (50 k) genotypes of over 6000 animals from 13 taurine and one zebu breed using an *ad-hoc* method to identify selective sweeps reaching fixation. Briefly, they considered a candidate locus under selection if at least five contiguous SNPs presenting a minor allele frequency (MAF) > 0.01 spanned a chromosome segment of 200 kb or more. Although the strategy was successful to find validated selective sweeps, such as the *POLL* locus that controls horn development (Georges et al., 1993), they showed that the SNP ascertainment bias of the 50 k assay incurred in false positives in breeds that are genetically distant from the SNP discovery breeds. This is not unexpected, as the influence of the 50 k bias on population genetics parameters, such as heterozygosity, genetic structure and differentiation, is well documented (Decker et al., 2009). Therefore, methods relying on heterozygosity are largely prone to type I errors generated from the SNP discovery process. Ideally, genotype data should be generated from less biased commercial arrays, such as the Illumina® BovineHD (HD), or from re-sequencing efforts.

Another approach to search for deficit of heterozygosity is the identification of islands of runs of homozygosity (ROH). As long stretches of consecutive homozygous genotypes indicate identical-by-descent haplotypes (Gibson et al., 2006; Lencz et al., 2007), ROH have been recently used to characterize genome-wide inbreeding in cattle (Purfield et al., 2012; Ferenčaković et al., 2013; Kim et al., 2013). However, as recombination is random, the distribution of ROH across samples is expected to be highly heterogeneous under neutrality, so genomic hotspots of ROH can be a signal of selective sweep (Curik et al., 2014). Interestingly, the length of the run is negatively correlated with the number of generations back to the selective pressure or inbreeding event (Howrigan et al., 2011), so ROH size can be used to date the age of the selective sweep.

Focusing on the identification of ROH islands produced by recent artificial selection in U.S. Holstein cattle, Kim et al. (2013) proposed a new statistic, namely F_{RL}, that compares local homozygosity between a population under artificial selection and a control population. Briefly, for each SNP, animals are scored as 1 if the locus is encompassed by a ROH or 0 otherwise. Then, F_L is computed as the proportion of animals with scores equal to 1. Finally, F_{RL} is obtained as:

$$F_{RL} = \ln\left(\frac{F_{L(selected)}}{F_{L(control)}}\right)$$

All scores are standardized (i.e., subtracting each value by the average score and dividing by the standard deviation) to produce a distribution with mean zero and variance one. Extreme positive values of F_{RL} represent changes in allele frequency in the artificially selected population in comparison to the control group, and therefore reflect homozygosity attributable to recent inbreeding, drift or selection at the analyzed locus. As for the approach adopted by Ramey et al. (2013), SNP density and ascertainment bias are important confounders that can produce false positive ROH (Ferenčaković et al., 2013), so commercial SNP assays must be used with caution here.

Methods that search for local absence of heterozygosity are most powerful to detect haplotypes that hitchhiked to fixation. However, allowing for some heterozygosity can be of help to detect partial sweeps, i.e., haplotypes under ongoing selection that have not reached fixation yet. For a given bi-allelic site, let n_{REF} be the number of observed reference alleles, n_{ALT} the number of observed alternative alleles, and p and q their respective

frequencies. The expected frequency of heterozygote genotypes under Hardy-Weinberg Equilibrium is given by:

$$H = 2pq = \frac{2n_{REF}n_{ALT}}{(n_{REF}+n_{ALT})^2}$$

This expectation can be easily extended for a chromosomal segment containing multiple bi-allelic sites as:

$$H = \frac{2\Sigma n_{REF}\Sigma n_{ALT}}{(\Sigma n_{REF} + \Sigma n_{ALT})^2}$$

Rubin et al. (2010) proposed running sliding windows across the genome, calculating H for these windows and then standardizing the obtained values. The method named as ZH_p produces standard deviation scores, and extremely negative values represent chromosome windows where the regional diversity is substantially lower than the average genome diversity. Although the method has been successfully used to map candidate variants under selection in chickens (Rubin et al., 2010), dogs (Axelsson et al., 2013) and pigs (Rubin et al., 2012), no studies applying ZH_p to cattle data have been published to date.

CHANGE IN THE ALLELE FREQUENCY SPECTRUM

After a complete selective sweep (i.e., the haplotype containing the selected variant reaches fixation), new mutations slowly restore local diversity in the course of many generations. As mutations are generally rare and may take a large number of generations to drift to high frequency under neutral evolution, the local heterozygosity depression signal is deemed to remain robust for several generations after the selective pressure has occurred.

Newly acquired mutations or derived alleles (i.e., variants that differ from the original or ancestral allele) occur in lower frequency in comparison to ancestral alleles under neutrality, but when they arise within a selective sweep they will hitchhike to high frequency quickly in the selected population. Therefore, another class of signals that emerge after the depression in local diversity is the enrichment for moderate to high frequency derived alleles. Several methods have been proposed in this category, including Tajima's (1983), Fay and Wu's (2000), and $\triangle DAF$ (Grossman et al., 2010).

The limitation of the use of these methods in cattle data is that inference of ancestral alleles should be preferably performed by comparison of domesticated cattle genomes with wild type genomes. As these are not available, ancestral allele information for cattle SNP assays were derived from genotypes of outgroup species, such as Gaur (*Bos gaurus*), Buffalo (*Bubalus bubalis*) and Yak (*Bos grunniens*), which are assumed to descend from a common founder *Bovinae* species (Matukumalli et al., 2009; Utsunomiya et al., 2013). However, as not all SNP probes hybridize against the genomes of these outgroup species, ancestral allele information for cattle is incomplete. Future availability of genome assemblies for these outgroup species may help to better infer ancestral status for common SNPs.

One way to bypass the limitation of ancestral allele information is by using samples from several genetically distinct populations to estimate average allele frequencies that could represent the spectrum in the common ancestral population. In this approach, instead of searching for enrichment of high frequency derived or rare alleles, one may look for a shift in the allele frequency spectrum in one population in comparison to the average across populations. Stella et al. (2010) successfully applied the negative composite log-likelihood (*CLL*) approach, an extension of the composite likelihood ratio (*CLR*) test (Kim and Stephan, 2002; Nielsen et al., 2005) to 13 taurine, 4 zebu and 2 synthetic breeds, and reported several candidate loci under selection. These included *KIT* (mast/stem cell growth factor receptor gene), responsible for the piebald and color sidedness phenotype (Durkin et al., 2012), and *MC1R* (melanocortin 1 receptor gene), incriminated in the black/red coat color in Holstein and Angus (Klungland et al., 1995). These loci were further confirmed using whole genome sequence data of German Fleckvieh (Qanbari et al., 2014).

Briefly, *CLL* is computed for each SNP window as:

$$CLL = -\sum_{i=1}^{k} \ln\left[1 - Pr\left(d < |p_i - \overline{p}_i| \mid \mu_i\right)\right]$$

Where, relative to SNP i, d is any random value from the theoretical distribution of allele frequencies with mean $\mu_i = \overline{p}_i$, $\overline{p}_i$ is the reference allele frequency averaged across populations, and p_i is the allele frequency in the population being investigated. The theoretical distribution of allele frequencies can be modeled as a binomial or a normal approximation to the binomial distribution. As calculations are very straightforward and only require a dataset with multiple breeds instead of ancestral allele information, it is general enough for cattle SNP data.

Another extension of the *CLR* statistic was proposed by Chen et al. (2010), namely *XPCLR* (cross-population composite likelihood ratio test), which attempts to detect a shift in allele frequency in a target population in respect to a reference population. Lee et al. (2014) applied *XPCLR* to Holstein and Hanwoo whole genome sequence data and reported that the chromosome segment encompassing the kappa-casein gene (*CSN3*) exhibited high *XPCLR* scores in Holstein cattle.

LONG-RANGE HAPLOTYPES

The concept of Extended Haplotype Homozygosity (*EHH*), first introduced by Sabeti et al. (2002), attempts to identify haplotypes that increased so rapidly in frequency that recombination could not substantially break them down, so linkage disequilibrium presents a long-range persistency. Briefly, consider N chromosomes in a sample, and G unique haplotypes extending from a core SNP site to an upstream or downstream position x, with each group g having n_g observations. The *EHH* score for the entire sample is calculated as:

$$EHH = \frac{\sum_{g=1}^{G}\binom{n_g}{2}}{\binom{N}{2}}$$

This score serves as a proxy for the probability of identity-by-descent of haplotypes within the chromosome segment being investigated. Generally, *EHH* is calculated at varying distances from the core SNP position, so that the decay of *EHH* as a

function of physical distance can be assessed to determine the extension of the haplotype homozygosity. From this seminal concept, a family of statistical methods was developed in order to scan entire genomes in the search for evidence of selection.

Voight et al. (2006) proposed to measure how rapidly *EHH* decays from a core SNP site by calculating the area under the *EHH* curve,

$$iHH = \int_a^b EHH(x)dx$$

where *iHH* represents the definite integral of *EHH* evaluated over the domain of the chromosome segment delimited by upstream position *a* and downstream position *b* where *EHH* decays to some arbitrary small value (originally 0.05). As the area under the curve is not tractable analytically, a trapezoid quadrature with non-uniform grid can be adopted as a deterministic approximation:

$$iHH \sim \sum_{k=1}^{K} (x_{k+1} - x_k) \frac{(EHH_{k+1} + EHH_k)}{2}$$

A within population score, namely Integrated Haplotype Score (*iHS*), for a given site *i*, was introduced by Voight et al. (2006) as the log-ratio between the integrated *EHH* for the haplotypes containing the ancestral allele (iHH_A) and the derived allele (iHH_D):

$$iHS_i = \ln\left(\frac{iHH_{A,i}}{iHH_{D,i}}\right)$$

These scores are then standardized to have mean zero and variance one.

Extremely negative standardized *iHS* values have been of particular interest in human genetics, as they represent a recently acquired mutation that increased very rapidly in frequency (i.e., there is a partial sweep due to ongoing selection) or a haplotype that hitchhiked to fixation and then became enriched for derived alleles (Voight et al., 2006). However, a sweep can also produce large positive *iHS* values at nearby SNPs if ancestral alleles hitchhike with the selected site, so the chromosome region surrounding the selected variant typically exhibits a cluster of extreme positive and negative *iHS* values. Furthermore, in the context of cattle data, artificial selection and domestication probably favored "beneficial" alleles in the sense of human interest, regardless if it is ancestral or derived. Therefore, both positive and negative values should be investigated in cattle data. This implicates that the absolute value of standardized *iHS* scores should be preferred over the signed values, or, equivalently, that a two tailed hypothesis test should be assumed. As only partial ancestral allele information is available for cattle SNP assays, and the search for footprints of selection by *iHS* in cattle should disregard the direction of the sweep, a more appropriate generalized version of *iHS* can be postulated as the log-ratio between the integrated *EHH* for an arbitrary reference allele (iHH_{REF}) and for the alternative allele (iHH_{ALT}).

One of the limitations of this method is that if a given marker presents a nearly or completely fixed allele in the population being analyzed, this allele will have no integral to be calculated or an integral close to zero, so the log-ratio will result in a positive or negative infinite value. In this scenario, the calculation of *iHS* must be conditioned by $iHH_{REF} > 0$ and $iHH_{ALT} > 0$, which indirectly leads to a minor allele frequency (MAF) constraint. This limitation renders *iHS* underpowered to detect very recent nearly fixed selective sweeps, which are of primary interest in the cattle community. However, as discussed earlier, a crucial point to be considered is that contiguous chromosomal segments containing SNPs with MAF = 0 can also result from SNP chip ascertainment bias, which may produce false positive signals.

Tang et al. (2007) and Sabeti et al. (2007) have independently developed equivalent methods, *Rsb* and *XPEHH*, respectively, which attempt to compare long-range haplotypes between populations in order to increase the power of selective sweep detection. The most crucial improvement is that, for each population being analyzed *iHH* is calculated for the entire sample, instead of being partitioned between derived and ancestral alleles. This eliminates the MAF constraint and recovers the power to detect sweeps reaching fixation. The comparison with a population where the selective sweep may not have occurred adds extra power to the method. Calculations are performed as follows:

$$XPEHH_i = Rsb_i = \ln\left(\frac{iHH_{pop1,i}}{iHH_{pop2,i}}\right)$$

Where, relative to SNP *i*, $iHH_{pop1,i}$ is the integrated *EHH* in the first population and $iHH_{pop2,i}$ is the integrated *EHH* in the second population. Scores are also standardized to produce a distribution of standard deviates. Positive values indicate selective sweeps in the population used in the numerator, while negative values indicate selection in the population used in the denominator. Here, it is easy to keep track of the signals by using one-tailed hypothesis tests.

Studies applying *EHH*-based methods to cattle data are numerous (for instance, Hayes et al., 2008; Gautier and Naves, 2011; Qanbari et al., 2011, 2014; Flori et al., 2012; Utsunomiya et al., 2013; Huson et al., 2014). The reported loci are deemed to be genome responses to a variety of different selective pressures, such as milk and meat production, coat color, heat stress, and reproductive performance. Among these, one particularly interesting selective sweep, most likely related to adaptation to heat stress, has been reported in Creole cattle, including Senepol, Carora, Romosinuano, and cross-bred lineages (Flori et al., 2012; Huson et al., 2014). These cattle breeds present the slick hair coat phenotype, a dominant trait associated to heat tolerance in tropically adapted cattle that descend from Spanish cattle introduced to the New World. The chromosome segment containing the selective sweep ranges from 37.5 to 39.6 Mb on chromosome 20, with a variable peaking position (39.5 or 37.7 Mb) depending on the SNP panel (BovineSNP50 or BovineHD) and dataset analyzed (Flori et al., 2012; Huson et al., 2014). The disputed positional candidate genes are the retinoic acid induced 14 (*RAI14* or *NORPEG*), prolactin receptor (*PRLR*), and S-phase kinase-associated protein 2 (*SKP2*). A strong candidate mutation has been recently proposed for *PRLR*, a single base deletion in exon 10 (ss1067289408) predicted to cause a frameshift that introduces

a premature stop codon (p.Leu462*) and consequent loss of 120 C-terminal amino acids from the long isoform of the prolactin receptor (Littlejohn et al., 2014).

POPULATION DIFFERENTIATION

Following the same principle as in *Rsb*/*XPEHH*, *XPCLR* and F_{RL}, although positive selection may act across populations sharing geographical proximity, environmental factors or a common phenotype, outgroup populations may not share the same selective pressure. Therefore, changes in allele frequency promoted by selection in one group will not be detectable in the other, and large differences in allele frequencies between populations will be observed.

The fixation index F_{ST} (Wright, 1950; Weir and Cockerham, 1984) and its abundant estimators is the gold standard method for detecting highly differentiated loci between populations. In essence, given the average allele frequency $\overline{p}$ across subpopulations, F_{ST} is simply the ratio between the variance in the allele frequency in different subpopulations $\sigma_S^2 = \sum_{j=1}^{k} \left(p_j - \overline{p}\right)^2$ and the variance in the total population $\sigma_T^2 = \overline{p}(1-\overline{p})$. In pair-wise comparisons, calculations simplify to:

$$F_{ST} = \frac{(p_1 - p_2)^2}{\left(p_1 + p_2\right)\left(2 - p_1 - p_2\right)}$$

Scores can be averaged across SNP windows or smoothed against genomic positions using a local variable bandwidth kernel estimator (Porto-Neto et al., 2013).

Flori et al. (2009) applied F_{ST} to three French dairy cattle breeds (Holstein, Normande, and Montbéliarde), and found that some of the putative loci under selection in that study overlapped genes that strongly affect milk production traits, such as the growth hormone receptor (*GHR*), and coat color, for instance *MC1R*. Porto-Neto et al. (2013) generated a comprehensive map of divergent loci between taurine and zebu cattle using over 777,000 SNPs and 13 cattle breeds, and reported that the highest scoring locus in the F_{ST} analysis maps to chromosome 7:47.2–53.7 Mb, which shelters a cluster of immune-related genes and *SPOCK1*, a gene previously implicated in puberty (Fortes et al., 2010).

Extensions of F_{ST} led to the development of the *FLK* test (Lewontin and Krakauer, 1973; Bonhomme et al., 2010; Fariello et al., 2013), which not only account for effective population size and hierarchical structure among populations, but also have known distributional properties under neutrality. Briefly, let q_i be the vector of reference allele frequencies at marker i for the populations being compared, and q_0 the ancestral allele frequency for the same allele. The *FLK* method relies on the linear model:

$$q_i = 1_n q_0 + e$$

where the residual term e is assumed $\mathcal{N}(0, V_i)$, and V_i is the expected variance-covariance matrix for vector q_i. This matrix is modeled as:

$$V_i = \mathcal{F} q_0 (1 - q_0)$$

where $\mathcal{F}$ is a kinship matrix. The diagonal elements of $\mathcal{F}$ represent the expected inbreeding coefficients in each respective population, and off-diagonal elements represent the amount of drift accumulated on the different branches of the population tree. A weighted least squares estimate of q_0 is then obtained as:

$$\hat{q}_0 = \left(1_n' \mathcal{F}^{-1} 1_n\right)^{-1} 1_n' \mathcal{F}^{-1} q_i$$

The *FLK* score is a measure of goodness-of-fit of this model, and is calculated as the deviance (i.e., residual sum of squares):

$$FLK_i = \left(q_i - \hat{q}_0 1_n\right)' V_i^{-1} \left(q_i - \hat{q}_0 1_n\right)$$

Under the assumption of a star-like tree-pure drift model (i.e., the populations being compared evolved in parallel from a single ancestral population, with no mutations or admixture), matrix $\mathcal{F}$ can be simplified to $\mathcal{F} = I_n \overline{F_{ST}}$, where I_n is an identity matrix and $\overline{F_{ST}}$ is the average F_{ST} over all SNP loci. In this case, the average allele frequency across populations is an unbiased estimator of q_0, and the deviance is simplified to $F_{ST}(n-1)/\overline{F_{ST}}$, which gives a test statistic that is linearly correlated with F_{ST}. As discussed later, while F_{ST} has no known theoretical distribution under the neutral model, *FLK* can be modeled as a chi-squared distribution. Moreover, *FLK* outperformed F_{ST} in simulations, especially when scores were computed based on haplotypes (*hapFLK*) instead of single markers (Fariello et al., 2013). This method is a suitable alternative to F_{ST} for cattle data.

DISTRIBUTIONS UNDER THE NULL HYPOSTHESIS AND *p*-VALUES

Although different in formulation and assumptions, all methods presented so far attempt to address the same null and alternative hypothesis: H_0 = locus is neutral; H_1 = locus is not neutral. Under the hypothesis of neutrality, *EHH*-based methods are standard normal deviates (Voight et al., 2006; Sabeti et al., 2007; Tang et al., 2007). Hence, the probability that SNP i with *Rsb* or *XPEHH* score x_i is neutral in the population used as numerator can be approximated by an upper tail p-value derived from the normal cumulative density function (CDF) Φ:

$$Pr\left(Rsb, XPEHH > x_i \mid neutral\right) = 1 - \Phi\left(x_i\right)$$

Likewise, the probability that SNP i with *iHS* score x_i is neutral, given that both reference and alternative alleles are of interest, can be approximated by a two-tailed p-value:

$$\begin{aligned} Pr\left(|iHS| > |x_i| \;\middle|\; neutral\right) &= 1 - \left|\Phi\left(x_i\right) - \Phi\left(-x_i\right)\right| \\ &= 1 - 2\left|\Phi\left(x_i\right) - 0.5\right| \end{aligned}$$

Recently, concerns have emerged on the interpretation of p-values in signatures of selection analyses (Simianer et al., 2014). It is argued that, at least for the cases of *EHH*-based methods, scores standardized using genome-wide data are not test statistics in the classical sense but only deviates from an average (Voight et al., 2006), so p-values would only represent quantiles from an empirical distribution, rather than formal significance values. This implies that the probability of obtaining an arbitrary score given the locus is neutral cannot be exactly computed once the

underlying true probability distribution may vary according to different demographic scenarios. The major caveat here is, for any given test discussed, the proposed asymptotic distribution under the hypothesis of neutrality is still largely based on coalescent simulations with demographic models calibrated for human data. However, if the majority of the genome is under neutrality and the null distribution is robust to a wide range of demographic models, one may expect that genome-wide distribution of scores in cattle data should mimic the simulations performed for the human model, serving as a control (Gianola et al., 2010). This has been observed in SS studies in cattle (Gautier and Naves, 2011; Flori et al., 2012; Utsunomiya et al., 2013), and therefore these theoretical distributions under neutrality are suitable for cattle data.

In the case of F_{ST}, although approximate distributions are available (e.g., exponential or beta), such approximations are sub-optimal. One advantage of *FLK* over F_{ST} is that scores are asymptotically distributed as χ^2_ν under the null hypothesis, where degrees of freedom ν can be equivalently calculated as *number of subpopulations* − 1 or as the average *FLK* across all loci. Upper tail *p*-values can be computed as:

$$Pr\,(FLK > x_i \mid neutral) = 1 - F_\nu(x_i)$$

where F_ν is the χ^2 CDF with ν degrees of freedom.

Although F_{ST} and *CLL* scores have unknown distributions under neutrality, both rely on population comparisons. The use of datasets consisting of multiple populations allows for permutation tests via random sampling of individuals or random sorting of population labels in order to compute an empirical null distribution. Nevertheless, permutation tests are computationally intensive and may be impracticable in re-sequencing data.

The *ZHp* method suffers from the same problem, with the additional limitation of not allowing data permutations when a single population is surveyed. Although a truncated normal distribution could be suggested to approximate its null distribution, this has not been properly assessed in practice. Another challenge is deriving significance values when scores are averaged across SNP windows. Following the assumption that the majority of the loci (single markers or SNP windows for that matter) are under neutrality, an empirical CDF derived from genome-wide scores should converge to the underlying true CDF, so probability values could be empirically obtained from a step function:

$$\begin{aligned} Pr\,(Test > x_i \mid neutral) &= 1 - ECDF\,(x_i) \\ &= 1 - \frac{\left(number\ of\ scores\ \leq\ x_i\right)}{number\ of\ observed\ scores} \end{aligned}$$

The probability that SNP i with score x_i in a given test is selected is not as trivial to approximate. While the distributions under the hypothesis of neutrality seems to be robust across a wide range of demographic models, the distribution under the hypothesis of selection may vary widely depending upon the demographic history that shaped the allele frequency spectrum. Therefore, there is no unique theoretical distribution to represent selected variants, and coalescent simulations using well calibrated demographic models are required in order to generate empirical distributions. This has been successfully done for human data (Grossman et al., 2010), but is highly challenging in cattle as most of the demographic history of the bovine species is yet to be uncovered. Nevertheless, promising methods to infer demographic scenarios from the data, including estimates of population sizes and population separations over time, are now emerging (Schiffels and Durbin, 2014), which could be useful to elucidate cattle history and customize coalescent simulations based on empirical data.

AVAILABLE SOFTWARE AND ANALYSIS BEST PRACTICES

Table 1 summarizes all the software mentioned below. The first step to be considered before starting a SS analysis is filtering out poor quality data and markers and samples that are not informative or that may confound the analysis. Discussing the particularities of measures of quality control is beyond the scope of the present article, but some of the best practices in SS studies can emerge from common sense. First, metrics such as Hardy-Weinberg Equilibrium (HWE) and MAF, highly used in GWA studies, should be applied with caution. Elimination of markers with extreme deviations from HWE expectations may counteract the whole SS enterprise, as in this type of study outlier loci are being sought. Likewise, MAF controls may cause signals that are reaching fixation to be completely lost. A situation where HWE and MAF thresholds can be benign is when only markers with extreme excess of heterozygotes are eliminated, and MAF constraints are applied to the pooled allele frequencies across all populations in the study. We suggest *PLINK* v1.07 (Purcell et al., 2007) or *PLINK* v1.90 for this first data screening. Other usual filters such as GenCall and GenTrain scores and call rates should follow the same guidelines as in GWA studies.

Another important issue is cryptic relatedness. Eliminating samples of closely related animals is of paramount importance to reduce false positive signals. We have previously reported an algorithm to find the maximum independent set based on identity-by-descent, i.e., maximize the number of samples while eliminating first and second degree relationships using SNP data (Utsunomiya et al., 2013). Other software such as *PLINK* v1.90 and *GCTA* v1.24.2 (Yang et al., 2011) also provide means to find the optimal set of independent individuals.

The next key step is producing high quality phased data. There are several methods and implementations for phasing, such as *fastPHASE* v1.2 (Scheet and Stephens, 2006), *Beagle* v3.3 or later (Browning and Browning, 2008), and *SHAPEIT2* (O'Connell et al., 2014). Although *fastPHASE* exhibits high phasing accuracy, it is orders of magnitude more computationally intensive than *Beagle* or *SHAPEIT2*. It is important to notice that phasing is not a straightforward procedure, and is highly prone to errors. Consequently, results from haplotype-based methods should be assessed with caution. The effect of haplotype errors on SS results remains underexplored and at some extent neglected.

Regarding the SS analysis *per se*, *EHH*-based methods can be computed using *Sweep*, *selscan* (Szpiech and Hernandez, 2014) or the *R* package *rehh* (Gautier and Vitalis, 2012). F_{RL} has a dedicated software, as well as *CLR* (available at: and *XPCLR*. For ROH-based tests, runs can be computed using either *PLINK* or *SNP & Variation Suite v7.6.8* or later, and F_{RL}can be easily computed with home grown scripts. In the cases of F_{ST}, *CLL*, ZH_p and

Table 1 | Available software for signatures of selection analysis.

Software	Analysis	Availability	References
PLINK v1.07 and v1.90beta	Quality control, data management, cryptic relatedness check, and computation of runs of homozygosity and FST	http://pngu.mgh.harvard.edu/~purcell/plink/ https://www.cog-genomics.org/plink2	Purcell et al., 2007
SNP & Variation Suite v7.6.8	Quality control, data management, cryptic relatedness check, and computation of runs of homozygosity and FST	http://www.goldenhelix.com	Golden Helix, Bozeman, MT, USA
GCTA v1.24.2	Quality control, data management, and cryptic relatedness check	http://www.complextraitgenomics.com/software/gcta/	Yang et al., 2011
fastPHASE v1.2	Phasing and imputation of missing genotypes	http://stephenslab.uchicago.edu/software.html#fastphase)	Scheet and Stephens, 2006
Beagle v3.3 and v4.0	Phasing and imputation of missing genotypes	http://faculty.washington.edu/browning/beagle/beagle.html	Browning and Browning, 2008
SHAPEIT2	Phasing and imputation of missing genotypes	https://mathgen.stats.ox.ac.uk/genetics_software/shapeit/shapeit.html	O'Connell et al., 2014
Sweep	EHH-based methods	http://www.broadinstitute.org/mpg/sweep/	Sabeti et al., 2002
selscan	EHH-based methods	https://github.com/szpiech/selscan	Szpiech and Hernandez, 2014
rehh	EHH-based methods	http://cran.r-project.org/web/packages/rehh/index.html	Gautier and Vitalis, 2012
FLK	Computation of the FLK statistics	https://qgsp.jouy.inra.fr/index.php?option=com_content&view=article&id=50&Itemid=55	Bonhomme et al., 2010; Fariello et al., 2013
SweepFinder	Computation of the CLR statistics	http://people.binf.ku.dk/rasmus/webpage/sf.html	Nielsen et al., 2005
XP-CLR v1.0	Computation of the XPCLR statistics	http://genetics.med.harvard.edu/reich/Reich_Lab/Software.html	Chen et al., 2010
cosi and cosi2	Coalescent simulations	http://www.broadinstitute.org/~sfs/cosi/ http://www.broadinstitute.org/~ilya/cosi2/	Schaffner et al., 2005; Shlyakhter et al., 2014
MSMS	Coalescent simulations	http://www.mabs.at/ewing/msms/index.shtml	Ewing and Hermisson, 2010

the method reported by Ramey et al. (2013), allele frequencies and genotype counts can be obtained from either *PLINK* or *SNP & Variation Suite*, and calculation of scores can be easily implemented in *R* (available at: http://www.r-project.org/) or other languages. Simulations of population genetics datasets under specific demographic models, including neutral and selected loci, can be performed using coalescent simulators such *cosi* (Schaffner et al., 2005), *cosi2* (Shlyakhter et al., 2014), or *MSMS* (Ewing and Hermisson, 2010).

COMBINING SELECTION SIGNALS

The available methodologies to detect positive selection differ substantially from each other in terms of the pattern of genetic variation encrypting a "signal." However, all of them have a shared objective: to identify loci that have undergone positive selection. Indeed, at least for recent selective pressures (up to a few thousand generations back), a selected variant is expected to be in a chromosome segment where there has been loss of diversity, enrichment for derived or rare alleles, population differentiation, and highly frequent long-range haplotypes. Therefore, collecting evidence across different methodologies targeting distinct classes of signals may help in identifying loci under positive selection. This section explores the statistical properties and limitations of some of the available methods designed to combine different methods for signatures of selection.

COMPOSITE OF MULTIPLE SIGNALS (CMS)

In the original implementation of the method (Grossman et al., 2010), CMS is a local test designed to narrow down signals detected from s distinct genome-wide scans, and is defined as the approximate joint posterior probability that a given variant is selected:

$$CMS_{local} = \Pr\left(selected \mid x_{ij}\right) = \prod_{j=1}^{s} \frac{\Pr\left(x_{ij}|selected\right)\Pr\left(selected\right)}{\Pr\left(x_{ij} \mid selected\right)\Pr\left(selected\right) + \Pr(x_{ij}|neutral)\Pr(neutral)}$$

The genome-wide extension (Grossman et al., 2013) focuses on the product of the Bayes Factors for each one of the tests to be combined. For each test, BF is computed as the ratio between the posterior and prior odds:

$$\mathrm{CMS_{GW}} = \prod_{j=1}^{s} \mathrm{BF_{ij}} = \prod_{j=1}^{s} \frac{\Pr(x_{ij}|selected)\Pr(selected)}{\Pr(x_{ij}|neutral)\Pr(neutral)}$$

In the absence of prior information on the number of loci under selection across the genome, CMS scores simplify to:

$$\mathrm{CMS_{GW}} = \prod_{j=1}^{s} \frac{\Pr(x_{ij}|selected)}{\Pr(x_{ij}|neutral)}$$

A challenging aspect of the implementation of these methods is computing $\Pr(x_{ij}|selected)$, which requires simulations under clear demographic assumptions. In contrast, as discussed earlier, many methods designed to detect markers under positive selection allow for approximating $\Pr(x_{ij}|neutral)$ from asymptotic theoretical or empirical distributions. Therefore, composite tests considering only the distributions under neutrality are more appropriate for cattle data.

The original CMS score can be modified to take advantage of the assumed distributions under neutrality in order to relax demographic assumptions and avoid expensive simulations. The most essential modification involves reformulating the problem of detecting markers departing from neutrality. Instead of considering the assessment of whether a marker has been selected or not, one can look for support from the data against the null model, i.e., that the marker does not fit well to the neutral model. First, let the new CMS score be the approximate joint posterior probability of a given variant not being neutral:

$$\begin{aligned}\mathrm{CMS_{new}} &= \Pr\left(not\ neutral|x_{ij}\right) \\ &= \prod_{j=1}^{s} \frac{\Pr(x_{ij}|not\ neutral)\Pr(not\ neutral)}{\Pr\left(x_{ij}|neutral\right)\Pr(neutral) + \Pr\left(x_{ij}|not\ neutral\right)\Pr(not\ neutral)}\end{aligned}$$

Here, $\Pr(x_{ij}|neutral)$ is computed directly from its theoretical distribution, and $\Pr\left(x_{ij} \mid not\ neutral\right)$ is computed as $1 - \Pr(x_{ij}|neutral)$. Also, it can be assumed that the prior $\Pr(not\ neutral) = 1 - \Pr(neutral)$. Therefore, the new CMS score can be re-written as:

$$\mathrm{CMS_{new}} = \prod_{j=1}^{s} \frac{\left[1 - \Pr\left(x_{ij}|neutral\right)\right]\left[1 - \Pr(neutral)\right]}{\Pr\left(x_{ij}|neutral\right)\Pr(neutral) + \left[1 - \Pr\left(x_{ij}|neutral\right)\right]\left[1 - \Pr\left(neutral\right)\right]}$$

Likewise, the genome-wide modified CMS scores can be reformulated as:

$$\mathrm{CMS_{GW-new}} = \prod_{j=1}^{s} \frac{\Pr(x_{ij}|not\ neutral)}{\Pr(x_{ij}|neutral)} = \prod_{j=1}^{s} \frac{1 - \Pr(x_{ij}|neutral)}{\Pr(x_{ij}|neutral)}$$

It is important to note that this modification does not allow for the same interpretation as the original CMS method: the composite likelihood does not indicate selection, but rather, that a marker does not fit well the neutral model.

META ANALYSIS OF SELECTION SIGNALS (META-SS)

Following the ideas expanded from the landmark publication of Grossman et al. (2010), given the probabilities under the null hypothesis for each test, our interest is to identify loci presenting consistent rejection of the neutral model across the different tests. For any given statistic, *p*-values are uniformly distributed in the interval between 0 and 1 under the null hypothesis. This property makes possible to use an inverse CDF, such as the Gaussian density, to produce scores for each test derived from a single theoretical distribution. Therefore, for each SNP i and test j, a new score can be computed as $Z_{ij} = \Phi^{-1}(1 - P_{ij})$, where P_{ij} is the *p*-value. These Z-transformed *p*-values can be then averaged and standardized to produce a composite score.

We have previously described $meta - SS$ (Utsunomiya et al., 2013), an abstraction of the Stouffer Z-transformation for combining different selection signals using the aforementioned framework. As the Stouffer method assumes the tests are uncorrelated under the shared null hypothesis and the use of pair-wise comparisons produce correlated scores, a weighted average was originally proposed to penalize dependent tests:

$$meta - SS = \frac{\sum_{j=1}^{s} \omega_j Z_j}{\sqrt{\sum_{j=1}^{k} \omega_j^2}}$$

where ω_j is the weight for test j. In this setting, a uniform penalization can be applied to control for the inflation of correlated tests. As this penalization does not incorporate the strength of correlations among tests, the $meta - SS$ test can be modified to explicitly account for the magnitude of correlations between scores. Considering all scores are equally weighted, the corrected composite score can be computed as:

$$meta - SS = \frac{\sum_{j=1}^{s} Z_j}{\sqrt{k + 2R}}$$

where R is the sum of all pairwise Pearson's product-moment correlations. Under the hypothesis of neutrality, these composite scores are distributed as $N(0, 1)$, so the higher is the Z-transformed value, the worse the marker fits the neutral model. Upper tail *p*-values can then be obtained from the standard normal CDF.

Obvious limitations from $meta - SS$ and CMS is the inability to incorporate statistics for which *p*-values cannot be derived. Randhawa et al. (2014) proposed Composite Selection Signals (CSS), a nonparametric interpretation of $meta - SS$, where fractional ranking is used instead of *p*-values to combine different tests. Briefly, the vector of test statistics for method j is first sorted and then ranked, taking values $1, \ldots, k$. Next, the vector of ranks is re-scaled by dividing all elements by $k + 1$, thus producing a variable ranging from 0 to 1. These re-scaled ranks are treated as they were *p*-values for the test statistics, and then combined as in the $meta - SS$ approach. This strategy is equivalent to computing probabilities from an empirical CDF using a step function, as discussed earlier, which has an appealing feature: as fractional ranking can be generated for any particular test, signature blending is made feasible even if the theoretical distributions are unknown or if scores have been averaged in chromosome windows. However, a caveat is that the magnitude of the actual test statistics may be lost, so one may expect loss of power compared to the use of theoretical or simulated distributions.

PRINCIPAL COMPONENTS ANALYSIS

Simianer et al. (2014) proposed combining different tests by applying an eigendecomposition of the correlation matrix of the scores. The attractive feature of this method is the possibility of using standardized scores instead of approximate probabilities. However, as each principal component has heterogeneous loadings from each test, deriving a single synthetic score that summarizes all different tests remains a challenge in this framework.

CHALLENGES AND FUTURE DIRECTIONS

In theory, genome-wide genotypes are a vast source of information that can be explored in the search for large effect mutations that underwent selection. However, the existing data and methods still suffer from power issues and confounding effects that can give rise to false positive and false negative signals.

Although simulations suggest that only marginal gains in power are obtained when the sample size is increased from tens to hundreds of unrelated samples, marker density and allele frequency spectrum seems to impact power dramatically (Lappalainen et al., 2010; Simianer et al., 2014). Genotypes derived from commercial SNP arrays have two important limitations in this context: (1) incomplete genome coverage by markers; and (2) ascertainment bias. The search for SS must be preferentially performed using high density SNP panels, although optimal average intermarker distances to detect a sweep may vary depending on the effective population size, extent of linkage disequilibrium and the nature of the signal. Regarding ascertainment bias, commercial SNP arrays are suitable for cattle populations that are closely related to the breeds used in the SNP discovery process, but there is no guarantee they will be informative in genetically distant populations. Indeed, with a few exceptions, little congruence has been reported between candidate selected loci identified using whole genome sequence and different commercial genotyping platforms in African humans not included in the HapMap data (Lachance and Tishkoff, 2013). Altogether, these arguments suggest that re-sequencing data is the optimal choice in SS studies in cattle. At some extent, the HD assay is appropriate, as it has a high-density coverage of the genome with SNPs that are less biased than competing panels.

Another important source of confounding comes from the methods available to detect SS. First, all methods assume that individuals have no recent relationships in their pedigrees, a condition that is hardly true and generally ignored. It is essential to filter the data for cryptic relationships and assure to include only samples that are unrelated for at least two generations. Second, most of the methods rely on haplotypes and SNP coordinates, so further improvement of phasing strategies and of the bovine reference genome assembly is crucial to assure high quality results. Third, variants can depart from neutrality not only due to positive selection, but also as consequence of demographic events such as bottlenecks, genetic drift and admixture. Distinguishing loci under selection from neutrally evolving loci remains a major challenge in the field, and will require refinement of existing methods and development of new tests. Nevertheless, combining signals across different methods seems to be a promising approach to mitigate the individual methodological limitations. Also, when available, the concomitant analysis of environmental data (e.g., temperature, humidity, precipitation, disease prevalence, etc.) may be of great help in distinguishing true positives and accelerating the link between signal and phenotype (Lv et al., 2014).

Well-planned study designs will be crucial to exploit the full potential of SS in the detection of large effect mutations favored by selection. The identification of common adaptive phenotypes, together with geographical information data, should be an important player in sampling and decisions of population comparisons. Cattle breeds that are not highly productive but that exhibit genetic local adaptation should be considered as priority targets, as their environmental fitness was probably forged by hundreds of years of natural and artificial selection. In the context of artificial selection for complex traits, as large cattle pedigree cohorts for genomic selection become available, it will be soon possible to actually assess rapid changes in allele frequency using historical data, rather than present date data only. First demonstrations of such ideas were presented by Decker et al. (2012), and are likely to be incorporated as routine monitoring tools of genomic resources in breeding programs in the future.

Recently, comparing candidate loci across GWA studies has been facilitated in cattle with the advent of the Animal QTLdb (Hu and Reecy, 2007). Similarly, results from a SS scan on the human 1000 Genomes data reported by Grossman et al. (2013) were made available through the CMS viewer tool (available at: http://www.broadinstitute.org/mpg/cmsviewer/). Pybus et al. (2014) have also launched a comprehensive database of SS in the 1000 Genomes data (available at: http://hsb.upf.edu/). The research community would highly benefit from the development of a SS database for livestock species, which would not only facilitate cross-referencing, but would also help researchers willing to dig deep into the functional meaning of the signals to select promising candidates emerging from multiple preexisting studies.

Finally, similarly to the recent developments in human SS (Kamberov et al., 2013), unraveling the functional relevance of the putative selected variants will demand interdisciplinary reasoning, compilation of a wide range of data types (e.g., transcriptomic, proteomic), and assemblage of an arsenal of *post-hoc* assays, such as genomic editing, culture, phenotyping and challenge of specific cell lines, production of knock-out models, and generation of cross-bred lines for confirmatory segregation analyses.

ACKNOWLEDGMENTS

The first author is supported by São Paulo Research Foundation (FAPESP) - process 2014/01095-8. Mention of trade name proprietary product or specified equipment in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the authors or their respective institutions.

REFERENCES

Ajmone-Marsan, P. (2010). On the origin of cattle: how aurochs became cattle and colonized the world. *Evol. Anthropol. Issues News Rev.* 19, 148–157. doi: 10.1002/evan.20267

Axelsson, E., Ratnakumar, A., Arendt, M.-L., Maqbool, K., Webster, M. T., Perloski, M., et al. (2013). The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature* 495, 360–364. doi: 10.1038/nature 11837

Bonfiglio, S., Achilli, A., Olivieri, A., Negrini, R., Colli, L., Liotta, L., et al. (2010). The enigmatic origin of bovine mtDNA haplogroup R: Sporadic interbreeding or an independent event of Bos primigenius domestication in Italy? *PLoS ONE* 5:e15760. doi: 10.1371/journal.pone.0015760

Bonhomme, M., Chevalet, C., Servin, B., Boitard, S., Abdallah, J., Blott, S., et al. (2010). Detecting selection in population trees: the lewontin and krakauer test extended. *Genetics* 186, 241–262. doi: 10.1534/genetics.110.117275

Bradley, D. G., MacHugh, D. E., Cunningham, P., and Loftus, R. T. (1996). Mitochondrial diversity and the origins of African and European cattle. *Proc. Natl. Acad. Sci. U.S.A.* 93, 5131–5135. doi: 10.1073/pnas.93.10.5131

Browning, B. L., and Browning, S. R. (2008). A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am. J. Hum. Genet.* 84, 210–223. doi: 10.1016/j.ajhg.2009.01.005

Bruford, M. W., Bradley, D. G., and Luikart, G. (2003). DNA markers reveal the complexity of livestock domestication. *Nat. Rev. Genet.* 4, 900–910. doi: 10.1038/nrg1203

Bush, W. S., and Moore, J. H. (2012). Chapter 11: genome-wide association studies. *PLoS Comput. Biol.* 8:e1002822. doi: 10.1371/journal.pcbi.1002822

Chen, H., Patterson, N., and Reich, D. (2010). Population differentiation as a test for selective sweeps. *Genome Res* 20, 393–402. doi: 10.1101/gr.100545.109

Curik, I., Ferenčaković, M., and Sölkner, J. (2014). Inbreeding and runs of homozygosity: a possible solution to an old problem. *Livest. Sci.* 166, 26–34. doi: 10.1101/gr.100545.109

Darwin, C., and Wallace, A. (1858). On the tendency of species to form varieties; and on the perpetuation of varieties and species by natural means of selection. *J. Proc. Linn. Soc. Lond. Zool.* 3, 45–62. doi: 10.1111/j.1096-3642.1858.tb02500.x

Decker, J. E., McKay, S. D., Rolf, M. M., Kim, J. W., Molina Alcalá, A., Sonstegard, T. S., et al. (2014). Worldwide patterns of ancestry, divergence, and admixture in domesticated cattle. *PLoS Genet.* 10:e1004254. doi: 10.1371/journal.pgen.1004254

Decker, J. E., Pires, J. C., Conant, G. C., McKay, S. D., Heaton, M. P., Chen, K., et al. (2009). Resolving the evolution of extant and extinct ruminants with high-throughput phylogenomics. *Proc. Natl. Acad. Sci. U.S.A.* 106, 18644–18649. doi: 10.1073/pnas.0904691106

Decker, J. E., Vasco, D. A., McKay, S. D., McClure, M. C., Rolf, M. M., Kim, J., et al. (2012). A novel analytical method, Birth Date Selection Mapping, detects response of the Angus (*Bos taurus*) genome to selection on complex traits. *BMC Genomics* 13:606. doi: 10.1186/1471-2164-13-606

Durkin, K., Coppieters, W., Drögemüller, C., Ahariz, N., Cambisano, N., Druet, T., et al. (2012). Serial translocation by means of circular intermediates underlies colour sidedness in cattle. *Nature* 482, 81–84. doi: 10.1038/nature 10757

Ewing, G., and Hermisson, J. (2010). MSMS: a coalescent simulation program including recombination, demographic structure and selection at a single locus. *Bioinformatics* 26, 2064–2065. doi: 10.1093/bioinformatics/btq322

FAO. (2007). *The State of the World's Animal Genetic Resources for Food and Agriculture*. Available online at: http://www.fao.org/docrep/010/a1250e/a1250e00.htm.

Fariello, M. I., Boitard, S., Naya, H., SanCristobal, M., and Servin, B. (2013). Detecting signatures of selection through haplotype differentiation among hierarchically structured populations. *Genetics* 193, 929–941. doi: 10.1534/genetics.112.147231

Fay, J. C., and Wu, C. I. (2000). Hitchhiking under positive Darwinian selection. *Genetics* 155, 1405–1413.

Ferenčaković, M., Sölkner, J., and Curik, I. (2013). Estimating autozygosity from high-throughput information: effects of SNP density and genotyping errors. *Genet. Sel. Evol.* 45:42. doi: 10.1186/1297-9686-45-42

Flori, L., Fritz, S., Jaffrézic, F., Boussaha, M., Gut, I., Heath, S., et al. (2009). The genome response to artificial selection: a case study in dairy cattle. *PLoS ONE* 4:e6595. doi: 10.1371/journal.pone.0006595

Flori, L., Gonzatti, M. I., Thevenon, S., Chantal, I., Pinto, J., Berthier, D., et al. (2012). A quasi-exclusive European ancestry in the Senepol tropical cattle breed highlights the importance of the slick locus in tropical adaptation. *PLoS ONE* 7:e36133. doi: 10.1371/journal.pone.0036133

Fortes, M. R. S., Reverter, A., Zhang, Y., Collis, E., Nagaraj, S. H., Jonsson, N. N., et al. (2010). Association weight matrix for the genetic dissection of puberty in beef cattle. *Proc. Natl. Acad. Sci. U.S.A.* 107, 13642–13647. doi: 10.1073/pnas.1002044107

Gautier, M., and Naves, M. (2011). Footprints of selection in the ancestral admixture of a New World Creole cattle breed. *Mol. Ecol.* 20, 3128–3143. doi: 10.1111/j.1365-294X.2011.05163.x

Gautier, M., and Vitalis, R. (2012). Rehh An R package to detect footprints of selection in genome-wide SNP data from haplotype structure. *Bioinformatics* 28, 1176–1177. doi: 10.1093/bioinformatics/bts115

Georges, M., Drinkwater, R., King, T., Mishra, A., Moore, S. S., Nielsen, D., et al. (1993). Microsatellite mapping of a gene affecting horn development in *Bos taurus*. *Nat. Genet.* 4, 206–210. doi: 10.1038/ng0693-206

Gianola, D., Simianer, H., and Qanbari, S. (2010). A two-step method for detecting selection signatures using genetic markers. *Genet. Res. (Camb.)* 92, 141–155. doi: 10.1017/S0016672310000121

Gibson, J., Morton, N. E., and Collins, A. (2006). Extended tracts of homozygosity in outbred human populations. *Hum. Mol. Genet.* 15, 789–795. doi: 10.1093/hmg/ddi493

Groeneveld, L. F., Lenstra, J. A., Eding, H., Toro, M. A., Scherf, B., Pilling, D., et al. (2010). Genetic diversity in farm animals–a review. *Anim. Genet.* 41(Suppl. 1), 6–31. doi: 10.1111/j.1365-2052.2010.02038.x

Grossman, S. R., Andersen, K. G., Shlyakhter, I., Tabrizi, S., Winnicki, S., Yen, A., et al. (2013). Identifying recent adaptations in large-scale genomic data. *Cell* 152, 703–713. doi: 10.1016/j.cell.2013.01.035

Grossman, S. R., Shylakhter, I., Karlsson, E. K., Byrne, E. H., Morales, S., Frieden, G., et al. (2010). A composite of multiple signals distinguishes causal variants in regions of positive selection. *Science* 327, 883–886. doi: 10.1126/science.1183863

Hanotte, O., Bradley, D. G., Ochieng, J. W., Verjee, Y., Hill, E. W., and Rege, J. E. O. (2002). African pastoralism: genetic imprints of origins and migrations. *Science* 296, 336–339. doi: 10.1126/science.1069878

Hayes, B. J., Lien, S., Nilsen, H., Olsen, H. G., Berg, P., Maceachern, S., et al. (2008). The origin of selection signatures on bovine chromosome 6. *Anim. Genet.* 39, 105–111. doi: 10.1111/j.1365-2052.2007.01683.x

Hellmann, I., Ebersberger, I., Ptak, S. E., Pääbo, S., and Przeworski, M. (2003). A neutral explanation for the correlation of diversity with recombination rates in humans. *Am. J. Hum. Genet.* 72, 1527–1535. doi: 10.1086/375657

Howrigan, D. P., Simonson, M. A., and Keller, M. C. (2011). Detecting autozygosity through runs of homozygosity: A comparison of three autozygosity detection algorithms. *BMC Genomics* 12:460. doi: 10.1186/1471-2164-12-460

Hu, Z. L., and Reecy, J. M. (2007). Animal QTLdb: beyond a repository - A public platform for QTL comparisons and integration with diverse types of structural genomic information. *Mamm. Genome* 18, 1–4. doi: 10.1007/s00335-006-0105-8

Huson, H. J., Kim, E. S., Godfrey, R. W., Olson, T. A., McClure, M. C., Chase, C. C., et al. (2014). Genome-wide association study and ancestral origins of the slick-hair coat in tropically adapted cattle. *Front. Genet.* 5:101. doi: 10.3389/fgene.2014.00101

Kamberov, Y. G., Wang, S., Tan, J., Gerbault, P., Wark, A., Tan, L., et al. (2013). Modeling recent human evolution in mice by expression of a selected EDAR variant. *Cell* 152, 691–702. doi: 10.1016/j.cell.2013.01.016

Kim, E. S., Cole, J. B., Huson, H., Wiggans, G. R., Van Tassel, C. P., Crooker, B. A., et al. (2013). Effect of artificial selection on runs of homozygosity in U.S. *Holstein cattle. PLoS ONE* 8:e80813. doi: 10.1371/journal.pone.0080813

Kim, Y., and Stephan, W. (2002). Detecting a local signature of genetic hitchhiking along a recombining chromosome. *Genetics* 160, 765–777.

Kimura, M. (1968). Evolutionary rate at the molecular level. *Nature* 217, 624–626. doi: 10.1038/217624a0

Klungland, H., Vage, D. I., Gomez-Raya, L., Adalsteinsson, S., and Lien, S. (1995). The role of melanocyte-stimulating hormone (MSH) receptor in bovine coat color determination. *Mamm. Genome* 6, 636–639. doi: 10.1007/BF00352371

Lachance, J., and Tishkoff, S. A. (2013). SNP ascertainment bias in population genetic analyses: Why it is important, and how to correct it. *Bioessays* 35, 780–786. doi: 10.1002/bies.201300014

Lappalainen, T., Salmela, E., Andersen, P. M., Dahlman-Wright, K., Sistonen, P., Savontaus, M., et al. (2010). Genomic landscape of positive natural selection

in Northern European populations.*Eur. J. Hum. Genet.* 18, 471–478. doi: 10.1038/ejhg.2009.184

Lee, H. J., Kim, J., Lee, T., Son, J. K., Yoon, H. B., Baek, K. S., et al. (2014). Deciphering the genetic blueprint behind Holstein milk proteins and production.*Genome Biol. Evol.* 6, 1366–1374. doi: 10.1093/gbe/evu102

Lencz, T., Lambert, C., DeRosse, P., Burdick, K. E., Morgan, T. V., Kane, J. M., et al. (2007). Runs of homozygosity reveal highly penetrant recessive loci in schizophrenia. *Proc. Natl. Acad. Sci. U.S.A.* 104, 19942–19947. doi: 10.1073/pnas.0710021104

Lewontin, R. C., and Krakauer, J. (1973). Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics* 74, 175–195.

Littlejohn, M. D., Henty, K. M., Tiplady, K., Johnson, T., Harland, C., Lopdell, T., et al. (2014). Functionally reciprocal mutations of the prolactin signalling pathway define hairy and slick cattle. *Nat. Commun.* 5, 5861. doi: 10.1038/ncomms6861

Loftus, R. T., MacHugh, D. E., Bradley, D. G., Sharp, P. M., and Cunningham, P. (1994). Evidence for two independent domestications of cattle. *Proc. Natl. Acad. Sci. U.S.A.* 91, 2757–2761. doi: 10.1073/pnas.91.7.2757

Lv, F., Agha, S., Kantanen, J., Colli, L., Stucki, S., and Kijas, J. W. (2014). Adaptations to climate-mediated selective pressures in sheep. *Mol. Biol. Evol.* 31, 3324–3343. doi: 10.1093/molbev/msu264

Matukumalli, L. K., Lawley, C. T., Schnabel, R. D., Taylor, J. F., Allan, M. F., Heaton, M. P., et al. (2009). Development and characterization of a high density snp genotyping assay for cattle. *PLoS ONE* 4:e5350. doi: 10.1371/journal.pone.0005350

Mona, S., Catalano, G., Lari, M., Larson, G., Boscato, P., Casoli, A., et al. (2010). Population dynamic of the extinct European aurochs: genetic evidence of a north-south differentiation pattern and no evidence of post-glacial expansion. *BMC Evol. Biol.* 10:83. doi: 10.1186/1471-2148-10-83

Murray, C., Huerta-Sanchez, E., Casey, F., and Bradley, D. G. (2010). Cattle demographic history modelled from autosomal sequence variation. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 365, 2531–2539. doi: 10.1098/rstb.2010.0103

Nielsen, R., Williamson, S., Kim, Y., Hubisz, M. J., Clark, A. G., and Bustamante, C. (2005). Genomic scans for selective sweeps using SNP data. *Genome Res.* 15, 1566–1575. doi: 10.1101/gr.4252305

O'Connell, J., Gurdasani, D., Delaneau, O., Pirastu, N., Ulivi, S., Cocca, M., et al. (2014). A general approach for haplotype phasing across the full spectrum of relatedness. *PLoS Genet.* 10:e1004234. doi: 10.1371/journal.pgen.1004234

Oleksyk, T. K., Smith, M. W., and O'Brien, S. J. (2010). Genome-wide scans for footprints of natural selection. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 365, 185–205. doi: 10.1098/rstb.2009.0219

Oleksyk, T. K., Zhao, K., De La Vega, F. M., Gilbert, D. A., O'Brien, S. J., and Smith, M. W. (2008). Identifying selected regions from heterozygosity and divergence using a light-coverage genomic dataset from two human populations. *PLoS One* 3:e1712. doi: 10.1371/journal.pone.0001712

Porto-Neto, L. R., Sonstegard, T. S., Liu, G. E., Bickhart, D. M., Da Silva, M. V. B., Machado, M., et al. (2013). Genomic divergence of zebu and taurine cattle identified through high-density SNP genotyping. *BMC Genomics* 14:876. doi: 10.1186/1471-2164-14-876

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795

Purfield, D. C., Berry, D. P., McParland, S., and Bradley, D. G. (2012). Runs of homozygosity and population history in cattle. *BMC Genet.* 13:70. doi: 10.1186/1471-2156-13-70

Pybus, M., Dall'Olio, G. M., Luisi, P., Uzkudun, M., Carreño-Torres, A., Pavlidis, P., et al. (2014). 1000 Genomes Selection Browser 1.0: a genome browser dedicated to signatures of natural selection in modern humans. *Nucleic Acids Res.* 42, D903–D909. doi: 10.1093/nar/gkt1188

Qanbari, S., Gianola, D., Hayes, B., Schenkel, F., Miller, S., Moore, S., et al. (2011). Application of site and haplotype-frequency based approaches for detecting selection signatures in cattle. *BMC Genomics* 12:318. doi: 10.1186/1471-2164-12-318

Qanbari, S., Pausch, H., Jansen, S., Somel, M., Strom, T. M., Fries, R., et al. (2014). Classic selective sweeps revealed by massive sequencing in cattle. *PLoS Genet.* 10:e1004148. doi: 10.1371/journal.pgen.1004148

Qanbari, S., and Simianer, H. (2014). Mapping signatures of positive selection in the genome of livestock. *Livest. Sci.* 166, 133–143. doi: 10.1016/j.livsci.2014.05.003

Ramey, H. R., Decker, J. E., McKay, S. D., Rolf, M. M., Schnabel, R. D., and Taylor, J. F. (2013). Detection of selective sweeps in cattle using genome-wide SNP data. *BMC Genomics* 14:382. doi: 10.1186/1471-2164-14-382

Randhawa, I. A. S., Khatkar, M. S., Thomson, P. C., and Raadsma, H. W. (2014). Composite selection signals can localize the trait specific genomic regions in multi-breed populations of cattle and sheep. *BMC Genet.* 15:34. doi: 10.1186/1471-2156-15-34

Rubin, C. J., Megens, H. J., Martinez Barrio, A., Maqbool, K., Sayyab, S., Schwochow, D., et al. (2012). Strong signatures of selection in the domestic pig genome. *Proc. Natl. Acad. Sci. U.S.A.* 109, 19529–19536. doi: 10.1073/pnas.1217149109

Rubin, C.-J., Zody, M. C., Eriksson, J., Meadows, J. R. S., Sherwood, E., Webster, M. T., et al. (2010). Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature* 464, 587–591. doi: 10.1038/nature08832

Sabeti, P. C., Reich, D. E., Higgins, J. M., Levine, H. Z. P., Richter, D. J., Schaffner, S. F., et al. (2002). Detecting recent positive selection in the human genome from haplotype structure. *Nature* 419, 832–837. doi: 10.1038/nature01140

Sabeti, P. C., Schaffner, S. F., Fry, B., Lohmueller, J., Varilly, P., Shamovsky, O., et al. (2006). Positive natural selection in the human lineage. *Science* 312, 1614–1620. doi: 10.1126/science.1124309

Sabeti, P. C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., et al. (2007). Genome-wide detection and characterization of positive selection in human populations. *Nature* 449, 913–918. doi: 10.1038/nature06250

Schaffner, S. F., Foo, C., Gabriel, S., Reich, D., Daly, M. J., and Altshuler, D. (2005). Calibrating a coalescent simulation of human genome sequence variation. *Genome Res.* 15, 1576–1583. doi: 10.1101/gr.3709305

Scheet, P., and Stephens, M. (2006). A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *Am. J. Hum. Genet.* 78, 629–644. doi: 10.1086/502802

Schiffels, S., and Durbin, R. (2014). Inferring human population size and separation history from multiple genome sequences. *Nat. Genet.* 46, 919–925 doi: 10.1038/ng.3015

Shlyakhter, I., Sabeti, P. C., and Schaffner, S. F. (2014). Cosi2: an efficient simulator of exact and approximate coalescent with selection. *Bioinformatics* 30, 3427–3429. doi: 10.1093/bioinformatics/btu562

Simianer, H., Ma, Y., and Qanbari, S. (2014). "Statistical problems in livestock population genomics," in *Proceedings, 10th World Congress of Genetics Applied to Livestock Production* (Vancouver, BC).

Stella, A., Ajmone-Marsan, P., Lazzari, B., and Boettcher, P. (2010). Identification of selection signatures in cattle breeds selected for dairy production.*Genetics* 185, 1451–1461. doi: 10.1534/genetics.110.116111

Szpiech, Z. A., and Hernandez, R. D. (2014). Selscan: an efficient multithreaded program to perform ehh-based scans for positive selection. *Mol. Biol. Evol.* 31, 2824–2827. doi: 10.1093/molbev/msu211

Taberlet, P., Valentini, A., Rezaei, H. R., Naderi, S., Pompanon, F., Negrini, R., et al. (2008). Are cattle, sheep, and goats endangered species? *Mol. Ecol.* 17, 275–284. doi: 10.1111/j.1365-294X.2007.03475.x

Tajima, F. (1983). Evolutionary relationship of DNA sequences in finite populations. *Genetics* 105, 437–460.

Tang, K., Thornton, K. R., and Stoneking, M. (2007). A new approach for using genome scans to detect recent positive selection in the human genome. *PLoS Biol.* 5, 1587–1602. doi: 10.1371/journal.pbio.0050171

The Bovine HapMap Consortium. (2009). Genome-wide survey of snp variation uncovers the genetic structure of cattle breeds. *Science* 324, 528 –532. doi: 10.1126/science.1167936

Utsunomiya, Y. T., Pérez O'Brien, A. M., Sonstegard, T. S., Van Tassell, C. P., do Carmo, A. S., Mészáros, G., et al. (2013). Detecting loci under recent positive selection in dairy and beef cattle by combining different genome-wide scan methods. *PLoS ONE* 8:e64280. doi: 10.1371/journal.pone.0064280

Vitti, J. J., Grossman, S. R., and Sabeti, P. C. (2013). Detecting natural selection in genomic data. *Annu. Rev. Genet.* 47, 97–120. doi: 10.1146/annurev-genet-111212-133526

Voight, B. F., Kudaravalli, S., Wen, X., and Pritchard, J. K. (2006). A map of recent positive selection in the human genome. *PLoS Biol.* 4:72. doi: 10.1371/journal.pbio.0040072

Weir, B. S., and Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution* 38, 1358–1370.

Wright, S. (1950). Genetical structure of populations. *Nature* 166, 247–249. doi: 10.1038/166247a0

Yang, J., Lee, S. H., Goddard, M. E., and Visscher, P. M. (2011). GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* 88, 76–82. doi: 10.1016/j.ajhg.2010.11.011

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 31 October 2014; accepted: 26 January 2015; published online: 10 February 2015.

Citation: Utsunomiya YT, Pérez O'Brien AM, Sonstegard TS, Sölkner J and Garcia JF (2015) Genomic data as the "hitchhiker's guide" to cattle adaptation: tracking the milestones of past selection in the bovine genome. Front. Genet. ***6****:36. doi: 10.3389/fgene.2015.00036*

This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics.

Copyright © 2015 Utsunomiya, Pérez O'Brien, Sonstegard, Sölkner and Garcia. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 02 June 2015
doi: 10.3389/fgene.2015.00191

Revisiting demographic processes in cattle with genome-wide population genetic analysis

Pablo Orozco-terWengel[1]*, *Mario Barbato*[1], *Ezequiel Nicolazzi*[2], *Filippo Biscarini*[2], *Marco Milanesi*[3], *Wyn Davies*[4], *Don Williams*[5], *Alessandra Stella*[2], *Paolo Ajmone-Marsan*[3] and *Michael W. Bruford*[1]

[1] School of Biosciences, Cardiff University, Cardiff, UK, [2] Parco Tecnologico Padano, Lodi, Italy, [3] Faculty of Agriculture, Università Cattolica del Sacro Cuore, Piacenza, Italy, [4] Dinefwr, National Trust, Llandeilo, UK, [5] Independent Researcher, Swansea, UK

OPEN ACCESS

Edited by:
Peter Dovc,
University of Ljubljana, Slovenia

Reviewed by:
Ino Curik,
University of Zagreb, Croatia
Lingyang Xu,
University of Maryland, USA

***Correspondence:**
Pablo Orozco-terWengel,
School of Biosciences, Cardiff University, Sir Martin Evans Building, Museum Avenue, Cardiff CF10 3AX, UK
orozco-terwengelpa@cardiff.ac.uk

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: 12 February 2015
Paper pending published: 26 March 2015
Accepted: 12 May 2015
Published: 02 June 2015

Citation:
Orozco-terWengel P, Barbato M, Nicolazzi E, Biscarini F, Milanesi M, Davies W, Williams D, Stella A, Ajmone-Marsan P and Bruford MW (2015) Revisiting demographic processes in cattle with genome-wide population genetic analysis. Front. Genet. 6:191. doi: 10.3389/fgene.2015.00191

The domestication of the aurochs took place approximately 10,000 years ago giving rise to the two main types of domestic cattle known today, taurine (*Bos taurus*) domesticated somewhere on or near the Fertile Crescent, and indicine (*Bos indicus*) domesticated in the Indus Valley. However, although cattle have historically played a prominent role in human society the exact origin of many extant breeds is not well known. Here we used a combination of medium and high-density Illumina Bovine SNP arrays (i.e., ∼54,000 and ∼770,000 SNPs, respectively), genotyped for over 1300 animals representing 56 cattle breeds, to describe the relationships among major European cattle breeds and detect patterns of admixture among them. Our results suggest modern cross-breeding and ancient hybridisation events have both played an important role, including with animals of indicine origin. We use these data to identify signatures of selection reflecting both domestication (hypothesized to produce a common signature across breeds) and local adaptation (predicted to exhibit a signature of selection unique to a single breed or group of related breeds with a common history) to uncover additional demographic complexity of modern European cattle.

Keywords: cattle, *Bos taurus*, *Bos indicus*, phylogeography, adaptation, SNP array

Introduction

Modern cattle comprise two species, *Bos taurus* and *Bos indicus,* both derived from the extinct aurochs (*B. primigenius*). The divergence between the two species has been dated ∼250,000 years ago on the basis of mitochondrial DNA haplotypes (Bradley et al., 1996). The two species are interfertile and a number of hybrid breeds have been established (e.g., the American Beefmaster that was established from a Shorthorn × Hereford × Brahman cross). The details of cattle domestication have been a contentious, and arguments in favor of two or three domestication events have been put forward. The hypothesis arguing for two domestication sites suggests that taurine cattle were domesticated in the Fertile Crescent, likely from *B. primigenius*, while indicine cattle were domesticated in the Indus valley, likely from *B. primigenius namadicus* (Loftus et al., 1994; Troy et al., 2001; Lewis et al., 2011). The hypothesis arguing for three domestication sites adds a Northeast African center, that gave rise to the African taurine breeds deriving form *B. primigenius opisthonomous* (Wendorf and Schild, 1994; Payne and Hodges, 1997). Mitochondrial DNA evidence suggests that two main lineages were domesticated, one now typical of taurine

breeds, and the other now typical of indicines (Loftus et al., 1994; Bradley et al., 1996). The domestication process is thought to have originated sometime between 8000 and 10,000 years ago (Ajmone-Marsan et al., 2010; Gautier et al., 2010; Lewis et al., 2011). Several potential routes have been hypothesized to explain the dispersal of taurine cattle from the Fertile crescent, mediated by Neolithic farmers, including alternative routes into Europe, occurring on at least two occasions (a Danubian route and a Mediterranean route Payne and Hodges, 1997; Pellecchia et al., 2007; Ajmone-Marsan et al., 2010; Taberlet et al., 2011). As aurochs only became extinct recently (Machugh et al., 1998; Beja-Pereira et al., 2006), it is possible that both taurine and indicine cattle crossed with their wild ancestor during their expansion out of the domestication centers, although the evidence supporting it is much debated (Troy et al., 2001; Gotherstrom et al., 2005; Beja-Pereira et al., 2006; Bollongino et al., 2008; Perez-Pardal et al., 2010; Edwards et al., 2011).

Several studies have addressed the population structure and phylogenetic relationships between taurine and indicine cattle breeds. Initially, studies were based on allozymes, mitochondrial DNA sequences, and microsatellites (Loftus et al., 1994; Medjugorac et al., 1994; Bradley et al., 1996; Machugh et al., 1998; Kantanen et al., 2000; Hanotte et al., 2002; Gautier et al., 2007) describing patterns of genetic variation in cattle, e.g., the two main lineages and differentiation between taurine and indicine breeds, albeit largely focusing on taurines. Nevertheless, such studies identified a clear differentiation between breeds from the two species, and between African and European taurines (Gautier et al., 2007). Interestingly, this handful of markers separated the African taurine breeds (e.g., Lagune, N'Dama and Somba) living in tsetse fly endemic regions from those that did not co-occur with these flies (e.g., Kuri and Borgou), suggesting that demographic signals among these breeds are strong and possibly related to properties of the environment they occupy. Recently, surveys of genome-wide genetic variation in many breeds have been carried out using SNP arrays (Bovine Hapmap et al., 2009; Mctavish et al., 2013; Decker et al., 2014). These more comprehensive analyses included tens of thousands of SNPs distributed across all chromosomes genotyped in several breeds from around the world. With these data, the differentiation between taurine and indicine breeds has been measured at approximately 10%, with approximately 5% for the divergence between African and non-African taurines (Gautier et al., 2010; Mctavish et al., 2013; Decker et al., 2014). Analyses using this type of large-scale data further identified a split between African longhorn taurines (N'Dama) and shorthorn taurines (e.g., Somba and Lagune—Gautier et al., 2010). Such analyses have also proved powerful enough to identify the mixed indicine/taurine genetic component of recently established hybrid breeds such as Santa Gertrudis and Beefmaster (Gautier et al., 2010; Mctavish et al., 2013).

Cattle are the most common large livestock species in the world; the global population size of which is approximately 1400 million animals (Felius et al., 2011; Taberlet et al., 2011). Among these, 1311 breeds have been recognized, 209 of which are now extinct (Rischkowsky and Pilling, 2007). Since domestication, cattle have spread around the world, with farmers selecting animals with desirable characteristics and establishing many breeds. This process, while not passive, was until recently slow and resulted in animals becoming adapted to the local conditions, e.g., feed types, local weather, and diseases (Wood, 1973; Cobb, 2006; Russell, 2007). The last two centuries saw a dramatic acceleration in the process of artificial/directional selection, largely as a consequence of the development of the breed concept, artificial insemination, and the development of statistical approaches used to estimate breeding values used to choose sires and dams for breeding purpose. This has limited the intercrossing between animals of divergent genetic background in order to increase or maintain breed purity, especially in dairy cattle [in beef cattle, conversely, cross-breeding is often used to exploit also non-additive genetic variation i.e., heterosis (Simm, 1998); still, pure lines for cross-breeding need to be selected and maintained]. Additionally, as a consequence of artificial insemination and marketing policies, few "champion" males with characteristics of interest have been used extensively to fertilize large proportions of females. The above developments have led to: (i) intense within-breed (or line) selection; (ii) importation/exportation of elite breeding stock to different environments/countries (migration); (iii) strong founder effect of "champion/top ranked" males both in breed formation or in already established breeds whose effective population size has thus been shrunken; (iv) larger genetic drift as a consequence of breed isolation. Generally, by trying to keep breeds pure, gene-flow among them has been reduced or even halted (unless done to improve breeds by upgrading Felius et al., 2011), and by using relatively few males to fertilize many females, inbreeding is increased (Taberlet et al., 2011). These approaches are used to manipulate the properties of a breed at the expense of the decrease in the genetic variability (Bulmer, 1971) which may be accompanied by a decrease in genetic health of the breed (e.g., reduction in fertility in Holstein cattle via the increase in frequency of deleterious variants through the process of inbreeding) (Pryce et al., 1997, 2004; Biffani et al., 2005). Overall, this change in farming practices resulted in a change from limited selection throughout much of the domestic history of cattle (∼9800 years) to relatively stronger selection in the last ∼200 years.

The relatively rapid distribution of cattle around the world over the last ∼10,000 years suggests that populations have been potentially exposed to selective pressures deriving from the environmental variables that they had not been exposed previously, e.g., diseases occurring outside of the domestication centers (Hanotte et al., 2002; Felius et al., 2014; Xu et al., 2014). Consequently, while some cattle breeds represent recent introductions to new geographical areas such as the New World or Oceania, other breeds represent long established populations outside of the domestication center. The genetic makeup of such populations is a valuable source to understand the processes driving local adaptation, as the former represent genetic pools currently being shaped by local selective pressures (i.e., may serve as examples of selection in action) (Flori et al., 2012), while the latter represent established populations where the extant genetic variation reflects the outcome of local adaptation (Felius et al., 2014). In particular, long established local breeds are likely to represent the reservoirs of important genetic variation

with adaptive potential that needs to be characterized (Taberlet et al., 2011; Felius et al., 2014) before it possibly disappears as a consequence of the widespread use of a handful of industrial breeds among others (Fao, 2007; Thornton, 2010; Herrero and Thornton, 2013; Perry et al., 2013).

Here, we collated a large-scale genome-wide dataset of SNP polymorphism in cattle breeds from around the world representing published taurine and indicine breeds, and new data for four European taurine breeds, all genotyped for a common set of ∼35,000 SNPs in over 1300 animals. We used these data to characterize the distribution of genetic variation among cattle breeds around the world, and the genealogical relationships between these breeds. The dataset described here represents multiple breeds from various continents (e.g., multiple European and African taurines, as well as Asian and African indicines) providing a unique experimental set up that allows searching for signatures of selection that are consistent across multiple breeds from different continents. This approach enables to separate signatures of selection that may be breed specific from selection signals common for several breeds occurring in the same geographic region. We extend the published dataset with new SNP data to a case study where we compare white cattle breeds of conservation concern in the UK (White Park and Chillingham) with a candidate Roman ancestor, the Chianina, to help resolve the commonly held belief of a Roman origin of the Welsh White Park cattle (Felius et al., 2011).

Materials and Methods

SNP Array Data

SNP array data was obtained for a total of 1346 animals representing 46 cattle breeds, and four species related to cattle (*B. gaurus*—Gaur, *B. javanicus*—Banteng, *B. gruniensis*—Yak, and *Bison bison*—Bison) used as outgroups (**Table 1**). The data comprises a combination of previously published datasets and new data collected for this study. The published data represents breeds published by Gautier et al. (2009); Gautier et al. (2010) and the Bovine HapMap project (Bovine Hapmap et al., 2009) genotyped with Illumina's BovineSNP50 v.1 and v.2 chip assay, respectively (Bovine Hapmap et al., 2009) (**Table 1**). Four new breeds were genotyped with the BovineHD Beach Chip for this project, namely, Welsh While Park (Dinefwr), Chianina, Romagnola and Chillingham. All SNP coordinates were converted to UMD3.1 bovine assembly (RefSeq: GCF_000003055.5). The datasets were merged using PLINK v1.7 (Purcell et al., 2007). The merged dataset was filtered to keep only individuals with 95% of their SNPs called and SNPs with 95% call rate across all samples. The SNPs and individuals that did not pass these filters were discarded. The total SNPs left in the dataset was 36,503 (data is available upon request from the authors). The data for each breed was phased with fastPHASE v1.4 (Scheet and Stephens, 2006).

Population Divergence

An analysis of population structure was carried out using the software Admixture v1.22 (Alexander et al., 2009), which uses a model-based estimation of indivdual ancestry for a range of prior values of K defined by the user. A cross-validation approach is used to determine the most likely number of populations (K) in the data. For each tested value of K, Admixture estimates the proportion of each individual's genotype deriving from each cluster. The values of K tested were in the range between 1 and 60 to accommodate potential population structure within breeds. For this analysis we excluded outgroup samples, breeds represented by fewer than 10 individuals, and SNPs with a minimum allele frequency lower than 0.01, linkage disequilibrium higher than 0.1 using a sliding window approach of 50 SNPs and a step size of 10 SNPs (Alexander et al., 2009). A principal components analysis was carried out using flashpca v1.2 (Abraham and Inouye, 2014) with default settings in order to investigate the ordinal relationships among breeds and individuals. Lastly, a NeighbourNet network was constructed using Reynold's distance among breeds using Splitstree v4.13.1 (Huson and Bryant, 2006), and a dendrogram using the same genetic distance and 100 bootstrap replicates to assess the statistical support of breed; clustering was performed using Phylip v3.69 (Felsenstein, 1993).

Within this framework, the Welsh White Park cattle breed was analyzed along the other cattle breeds to identify their relative position within the dendrogram and PCA explaining the similarities between cattle breeds. In particular, the common belief of a Roman origin of the Welsh White Park was addressed by comparing the position of this breed in relation to the Italian breeds, in particular the large Chianina breed, one of the oldest cattle breeds originating from the region of Valdichiana in central Italy, and Romagnola. Additionally, we estimated the relationships between Welsh White Park and other 17 breeds in the dataset using Treemix v1.12 (Pickrell and Pritchard, 2012) in order to determine the historical relationships between these breeds in terms of splits and migration (mixtures) between breeds. The breeds used for this analysis were Welsh White Park, the Italian Chianina, Romagnola (ROM), and Piedmontese, the British breeds Chillingham, Hereford, Angus, Jersey (JER), as well as five other non-African taurines (Holstein (HOL), Brown Swiss, Charolais, Normande, Vosgienne), the four African taurines Baoule, Lagune, Somba and N'Dama (NDA), and the indicine breed Brahman as outgroup. Treemix was run iteratively for values of the migration parameter between 0 and 12. The f index representing the fraction of the variance in the sample covariance matrix ($\widehat{W}$) accounted for by the model covariance matrix (W) was used to identify the number of modeled migration events that best fitted the data (Pickrell and Pritchard, 2012).

Demographic Estimation

Recent demographic history was measured by the trend in effective population size (Ne) change over the last 8000 years using the software SNeP v1.0 and default settings (Barbato et al., 2015). Loci with missing data were excluded from this analysis, as well as SNPs with a minimum allele frequency (MAF) smaller than 5% (Sved et al., 2008; Corbin et al., 2012). Linkage disequilibrium (LD) was calculated between each pair of SNPs separated by a minimum distance of 5 kbp and maximum distance of 1 Mbp using Hill and Robertson's r^2

TABLE 1 | Breed abbreviations and species identification.

Abb	BN	Tax	*D*	*N*	Ho	*F*
ABO	Abondance	*Bos taurus*	2	22	0.308	0.089
ANG	Angus	*Bos taurus*	1	61	0.305	0.098
AUB	Aubrac	*Bos taurus*	2	22	0.301	0.112
BPN	Bretonne black pied	*Bos taurus*	2	18	0.317	0.064
BRU	French brown swiss	*Bos taurus*	2	18	0.288	0.150
BSW	Brown swiss	*Bos taurus*	1	24	0.280	0.171
CHA	Charolais	*Bos taurus*	1	20	0.314	0.071
CHL	Charolais	*Bos taurus*	2	26	0.320	0.054
GAS	Gascon	*Bos taurus*	2	22	0.305	0.100
GNS	Guernsey	*Bos taurus*	1	21	0.275	0.188
HFD	Hereford	*Bos taurus*	1	31	0.304	0.101
HO2	Holstein	*Bos taurus*	2	31	0.316	0.066
HOf	Holstein	*Bos taurus*	1	30	0.313	0.077
HOL	Holstein	*Bos taurus*	2	64	0.319	0.057
JE2	Jersey	*Bos taurus*	2	28	0.263	0.223
JEf	Jersey	*Bos taurus*	1	21	0.277	0.180
JER	Jersey	*Bos taurus*	2	28	0.263	0.223
LMS	Limousin	*Bos taurus*	1	44	0.309	0.086
MAN	Maine-Anjou (Rouge des Près)	*Bos taurus*	2	16	0.303	0.105
MAR	Maraichine (Parthenaise)	*Bos taurus*	2	19	0.318	0.059
MON	Montbeliarde	*Bos taurus*	2	30	0.299	0.116
NOR	Normande	*Bos taurus*	2	30	0.307	0.094
NRC	Norwegian red cattle	*Bos taurus*	1	21	0.317	0.063
OUL	Oulmès Zaer	*Bos taurus*	2	27	0.288	0.149
PMT	Piedmontese	*Bos taurus*	1	24	0.321	0.053
PRP	French red pied lowland	*Bos taurus*	2	22	0.325	0.039
RGU	Red Angus	*Bos taurus*	1	15	0.305	0.100
RMG	Romagnola	*Bos taurus*	1	24	0.291	0.141
ROM	Romagnola	*Bos taurus*	3	13	0.293	0.134
CHI	Chianina	*Bos taurus*	3	14	0.285	0.158
CIL	Chillingham	*Bos taurus*	3	16	0.026	0.924
W_P	White park	*Bos taurus*	3	15	0.245	0.276
SAL	Salers	*Bos taurus*	2	22	0.285	0.157
TAR	Tarine	*Bos taurus*	2	18	0.300	0.113
VOS	Vosgienne	*Bos taurus*	2	20	0.312	0.078
BAO	Baoule	*Bos taurus*	2	29	0.216	0.362
LAG	Lagune	*Bos taurus*	2	30	0.183	0.460
NDA	N'Dama	*Bos taurus*	1	25	0.209	0.381
ND1	N'Dama	*Bos taurus*	2	14	0.235	0.307
ND2	N'Dama	*Bos taurus*	2	17	0.237	0.299
ND3	N'Dama	*Bos taurus*	2	25	0.210	0.381
SOM	Somba	*Bos taurus*	2	30	0.217	0.359
BRM	Brahman	*Bos indicus*	1	25	0.190	0.440
GIR	Gir	*Bos indicus*	1	24	0.160	0.528
NEL	Nelore	*Bos indicus*	1	21	0.161	0.524
ZBO	Zebu Bororo	*Bos indicus*	2	23	0.240	0.292
ZFU	Zebu Fulani	*Bos indicus*	2	30	0.241	0.289
ZMA	Zebu from Madagascar	*Bos indicus*	2	30	0.194	0.427
BMA	Beefmaster	*hybrid*	1	24	0.328	0.031
SGT	Santa Gertrudis	*hybrid*	1	24	0.313	0.075

(Continued)

TABLE 1 | Continued

Abb	BN	Tax	*D*	*N*	Ho	*F*
BOR	Borgou	*hybrid*	2	30	0.263	0.222
KUR	Kuri	*hybrid*	2	30	0.261	0.229
SHK	Sheko	*hybrid*	1	20	0.250	0.260
SIM	Simmental	*Bos taurus*	1	3	NA	NA
GBV	Gelbvieh	*Bos taurus*	1	3	NA	NA
OBB	North American Bison	*Bos Bison*	2	4	NA	NA
OBJ	Banteng	*Bos javanicus*	1	2	NA	NA
OGR	Gaur	*Bos gaurus*	1	4	NA	NA
OYK	Yak	*Bos grunniens*	1	2	NA	NA

*The three letter code used to identify the breed throughout the manuscript is shown (**Abb**), along with the breed name (**BN**) and the taxonomic status indicating the species name or whether the breed is a B. taurus x B. indicus hybrid (**Tax**), the data set's origin (**D**), namely Bovine Hapmap Consortium (1), Gautier et al., 2010 (2), this study (3), the breed's sample size (**N**), the observed heterozygosity (**Ho**), and the inbreeding coefficient (**F**).*

(Hill and Robertson, 1968). LD values were grouped in 30 distances bins, such that each bin represented approximately the same number of pairwise LD estimates used to estimate Ne (Barbato et al., 2015). Within each bin the mean r^2 was calculated and used to estimate $N_t = 1/(4f(c))\ (1/r^2 - 1)$, where $f(c) = c\ [(1-c/2)/(1-2)^2]$ (Sved, 1971); c is the linkage distance inferred from the physical distance between SNPs assuming 1 $Mb \sim 1$ cM and N_t represents the effective population size estimate at $t = 1/2c$ generations ago.

Signatures of Selection

As the dataset analyzed here comprised *B. taurus* breeds from Europe and Africa, as well as *B. indicus* breeds from Asia and Africa, we sought to identify signatures of selection in a replicated manner (**Figure 1**). Initially we compared four taurine breeds (two European—Gascon and Chianina; and two African—Baoule and N'Dama) with four indicine breeds (two African – Zebu Fulani and Zebu Bororo; and the Asian Gir and Brasilian Nelore; Nelore was recently derived from Indian Ongole cattle, therefore, we broadly refer to these breeds as Asian indicine) to identify signatures of selection between *B. taurus* and *B. indicus*. Within *B. taurus*, we randomly chose three European breeds to compare against three African breeds in order to identify selection signatures specific to breeds in these continents. In the same way, for *B. indicus* we compared three Asian breeds against three African breeds (**Figure 1**). In contrast to previous studies, this experimental design enabled us identifying potential signatures of selection reflecting adaption to the local environment, instead of breed specific signatures that potentially reflect the breed's particular history.

The test of selection used was the cross population extended haplotype homozygosity test (XP-EHH) calculated for each SNP in the dataset as implemented in Selscan v1.0 (Szpiech and Hernandez, 2014). Selscan was used with default settings except for the maximum gap in bp allowed between SNPs, which was set to 400,000 to address the BovineSNP50 SNP density. For the comparisons between taurine and indicine breeds the following pairwise combinations were used: N'Dama (African taurine)/Gir (Asiatic indice), Gascone (European taurine)/Zebu Bororo (African indicine), Baoule (African taurine)/Zebu Fulani

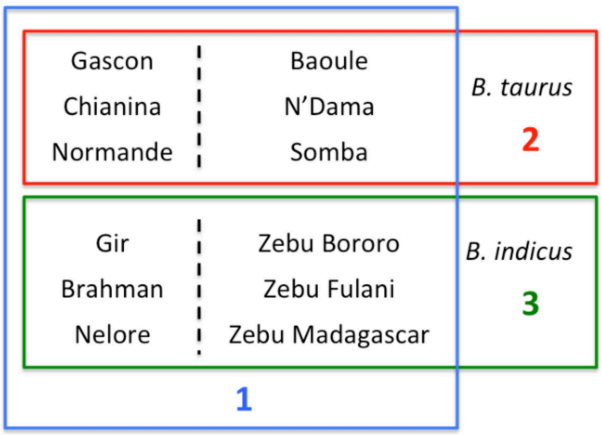


FIGURE 1 | Pairwise comparisons for analysis of selection. Each box corresponds to one of the three types of pairwise comparisons carried out to detect signatures of selection. The blue box (comparison 1) indicates the comparison between four taurine and four indicine breeds used to detect signatures of selection between these breeds. The red box (comparison 2) shows the taurine breeds used in the comparison between African and not African breeds. The green box (comparison 3) shows the indicine breeds used in the comparison between African and not African breeds.

(African indicine) and Chianina (European taurine)/Nelore (Asian indicine). For the within-taurine comparison three pairwise tests were done between a European breed and an African breed, and for the within-indicine comparison three pairwise comparisons were carried out between an Asiatic breed and an African breed (**Figure 1**). For each pairwise comparison, the XP-EHH results where standardized for each chromosome separately, and the 5 and 95% quantiles of the standardized XP-EHH distribution for all markers where used as threshold to identify outlier SNPs, i.e., SNPs with XP-EHH values equal or smaller (more negative) or equal or larger (more positive) than the chosen thresholds. As multiple XP-EHH were carried out for each comparison (e.g., four comparisons between taurine and indicine breeds) this analysis was the equivalent of a replicated test. We took advantage of this design to identify SNPs under selection as those that showed for several pairwise comparisons

an extreme XP-EHH with the same sign (e.g., all pairwise comparisons should an extreme positive XP-EHH).

For the between species comparison we defined the consistent SNPs under selection as those showing the same XP-EHH trend in at least three of the four pairwise comparisons. We did not restrict ourselves to only those SNPs where the four tests showed the same result to avoid excluding SNPs that failed to pass the 5% MAF threshold within a particular breed but which showed a consistent signature of selection across the remaining three comparisons. For the within species comparison of breeds occurring on different continents we identified consistent SNPs under selection as those showing an extreme XP-EHH with the same sign in each of the three comparisons made. The SNPs showing consistent signatures of selection across multiple pairwise comparisons where linked to neighboring genes using a window approach of 50 K base pairs, and a Gene Ontology analysis was carried out on these SNPs using Gorilla (www.cblgorilla.cs.technion.ac.il; Eden et al., 2009).

Results

Of the 36,503 SNPs analyzed in 1345 animals representing taurine and indicine breeds, the observed heterozygosity varied between a minimum of 0.026 (Chillingham) and a maximum of 0.33 (the hybrid breed Beefmaster; **Table 1**). The comparison of expected heterozygosity (unbiased to sample size) between the taurine and indicine breeds was not significant, as well as the comparison between breeds in and out of Africa (all Welch *t*-test *p*-values > 0.2). However, the average observed heterozygosity in taurine breeds was slightly larger than that in indicine breeds [0.28 standard deviation (sd) 0.05 and 0.20 sd 0.04, respectively]. The average expected heterozygosity across breeds in both datasets was similar, i.e., 0.34 sd 0.0001. The inbreeding coefficient (F_{IS}) was significantly lower for taurine breeds (F_{IS}: 0.18 sd 0.16) than for indicine breeds (F_{IS}: 0.42 sd 0.11; Welch *t*-test *p*-value: 0.0011). Within *B. taurus*, the African breeds presented a significantly higher inbreeding coefficient than the non-African taurines (Welch *t*-test *p*-value: 1.54×10^{-7}). Similarly, the non-African taurines also presented significantly lower F_{IS} values than either of the groups of indicine breeds (Welch *t*-test *p*-value: 1.84×10^{-5} and 0.026, respectively). Among the non-African taurines, the Chillingham presented by far the highest inbreeding coefficient of the entire dataset (F_{IS}: 0.92), on average five-fold larger than other taurines, and two-fold larger than the indicine breeds. The Welsh White Park is thought to be related to the Chillingham breed, but presented an observed heterozygosity (0.245) only slightly smaller than the average value observed in the other non-African taurines. The F_{IS} value for Welsh White Park (0.27) was less than a third of that of Chillingham but still almost twice the average across the other non-African taurine breeds (0.14 sd 0.14). With the exception of Chillingham, the four breeds genotyped here presented typical observed heterozygosity and F_{IS} values for European taurine breeds (**Table 1**). Removing from the analysis the breeds with samples sizes smaller than 20 did not change the results (results not shown).

Admixture and Genetic Relationships

Admixture analysis between taurine and indicine breeds was initially run for values of the number of clusters (K) between 1 and 60. The cross validation (CV) statistic used to choose the most suitable number of clusters had its lowest value at $K = 60$. However, the shape of the CV curve suggested that higher K values may present lower CV. Consequently, we carried out runs of Admixture for larger values of K (e.g., 80). However, the CV values continued to reduce the larger the K numbers tested (SM Figure 1). While it may be possible that some family structure within breeds explains this trend, the substantially longer run time needed by the algorithm for such K values combined with the small decrease in CV suggests that clustering solutions larger than K ∼60 potentially represent spurious results (a similar observation was made by Mctavish et al. (2013) who only achieved marginal increases on the likelihood scores using structure in their cattle dataset for values of K beyond 3).

The hierarchical clustering analysis with Admixture showed that the largest difference for $K = 2$ was between taurine and indicine breeds. Conditioning the data to $K = 3$ separated the taurine breeds with an African origin from the non-African, suggesting a large differentiation between taurine breeds in these two groups of samples (**Figure 2**). Contrastingly, for values up to $K = 10$, the African indicine breeds presented a remarkably similar distribution of genetic variation to that of the Asian indicine breeds, which did not form a separate cluster. Nonetheless, the African indicine breeds consistently presented a distinctive proportion of their genotype representing taurine ancestry throughout the analysis (a pattern almost entirely absent in the breeds of Asian origin). The *B. taurus*/*B. indicus* hybrids show a genetic make-up consisting of non-African taurine and *B. indicus* genetic components (Beefmaster and Santa Gertrudis), or African taurine and *B. indicus* (Borgou, Kuri, and Sheko).

The Admixture analysis was complemented with a principal component analysis (**Figure 3**). The first principal component (PC1) explained ∼10% of the total variation in the data and separated taurine form indicine breeds. Besides of PC1, the only other component explaining a substantial part of the variance was PC2 (∼5%), which separates the African taurine from the remaining taurine, as well as the African indicine from the Asian indicine breeds. PC3 to PC7 explained between less than 2 and 1.4% of the variance, while the remaining components explained less than 1% of the total variance, thus were not taken into consideration. PC3 separated Holstein from the remaining taurine breeds. PC4 identified a group of British breeds formed by Angus, Red Angus and Hereford, and PC5 identified a component of variation specific to Hereford separating it from the other taurines. PC6 identified the Chillingham cattle as well as the Jersey breed, and PC7 identified a specific component of variation to Chillingham separating it from the remaining taurine breeds. The combination of PC1 and PC2 identify the major groups of individuals described here in terms of species and geographic component (**Figure 3A**) with the hybrid breeds occurring between these. As expected from the Admixture analysis, the breeds Kuri, Borgou, and

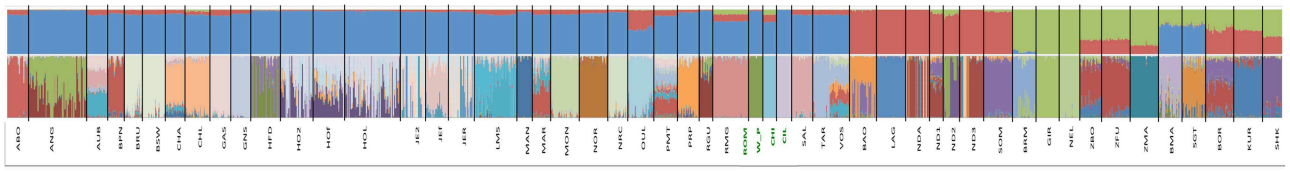


FIGURE 2 | Analysis of Population Structure. Results of the analysis of population structure conditioning the dataset to 3 clusters (top row) and to 51 clusters (bottom row). Each animal is represented by a straight bar that is colored. The amount of a color reflects the individual's proportion of genetic variation originating in the cluster of that color. Each breed is labeled in the center of its box on the bottom of each plot. In the top row the non-African taurine breeds are labeled in blue, the African taurine in red and the indicine (both African and not African in green). The new datasets produced for this project are highlighted in green.

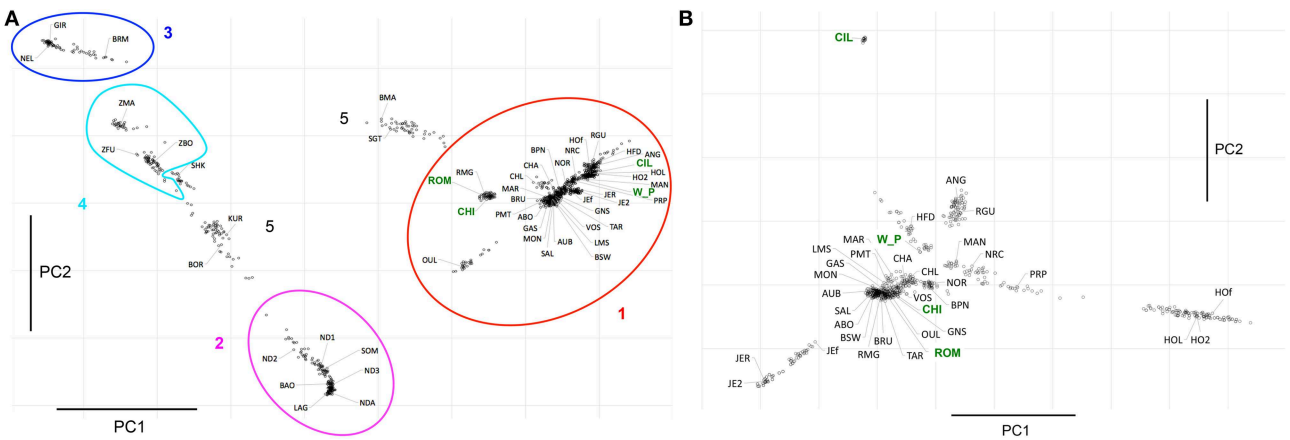


FIGURE 3 | Principal Component Analyses. (A) This plots shows PC1 and PC2, which together explain ~15% of the variance in the dataset. The ellipses represent: Red=non-African taurine, Purple=African taurine, Dark Blue=Asian indicine, Light Blue=African indicine. Three letter codes refer to the breed abbreviations (**Table 1**). **(B)** Plot showing PC1 and PC2 of the analysis of the non-African taurine breeds, which together explain ~6.5% of the variance in that dataset. The new datasets produced for this project are highlighted in green in both **(A,B)**.

Sheko are placed with these principal components between the group of African taurine and indicine. Similarly, Beef Master and Santa Gertrudis are placed between the group of non-African taurine and indicine breeds, although from this analysis it is not clear whether their indicine genetic component is more similar to African indicine or Asian indicine.

The methods described above identified differences between breeds but did not allow us to determine how these relate to each other. We therefore estimated a NeighbourNet network with these breeds using Reynold's distance. The torso of the network shows multiple pathways of connection between breeds reflecting the relatively recent divergence between many populations (**Figure 4**). However, the topology of the network resulted in the pattern expected from the previous analyses (Admixture and PCA), where a clear separation between taurine and indicine occurs, as well as the distinct separation between the African taurine breeds from the non-African taurines. Similarly to the PCA and in contrast to the Bayesian approach of Admixture, the network also resolved the African indicine breeds separately from their Asian counterparts. The dendrogram depicting the clustering of breeds using Reynold's distance recovered the grouping of breeds identified in the PCA with most branches (83%) supported with high statistical support (bootstrap values higher than 70%; SM Figure 2). The clustering of the outgroups was not well resolved, except for the grouping between bison and yak. The indicine breeds clustered together with the Asian indicine showing a greater similarity to each other, and then to the Madagascan zebu. The two African indicine clustered together and then clustered with the remaining indicines. The taurine breeds separated into two clusters, one representing the African taurines and the other the non-African taurines. The hybrids Beefmaster and Santa Gertrudis clustered together within the non-African taurine group reflecting their larger taurine genetic component. Contrastingly, the African hybrids (Sheko, Borgou, and Kuri) clustered with the indicine breeds suggesting a higher indicine genetic component in their genotype.

A comparison between the Welsh White Park cattle and the other British breeds (Chillingham, Hereford, and Jersey) and Italian breeds (Chianina, Romagnola, Piedmontese) was carried out to establish whether the UK white cattle have a genetic link with Roman white cattle, as is popularly believed. The Principal Components Analysis of this datasets identified the White Park breed clustering near the other British breeds (i.e., HFD, ANG, RGU), instead of closer to the Italian breeds (**Figure 3B**), as would be expected if White Park cattle had a strong Italian genetic

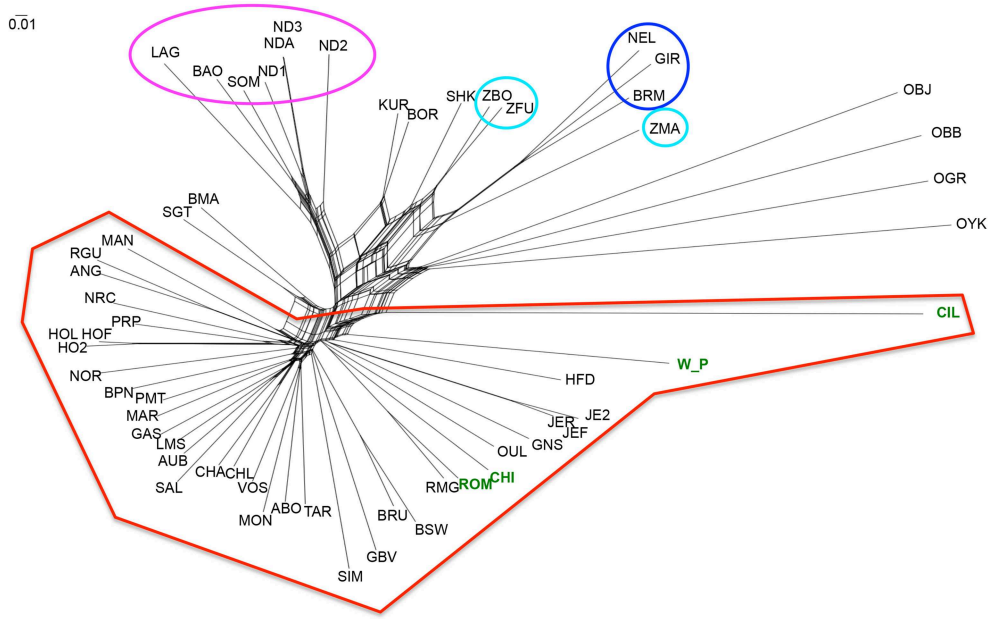


FIGURE 4 | NeighbourNet depicting the relationships between cattle breeds. Three letter codes refer to the breed abbreviations (**Table 1**). The groups marked with ellipses in **Figure 3A** were marked with the same colors here as well. The new datasets produced for this project are highlighted in green.

component due to its speculated Roman origin. Similarly, in the dendrogram, the Italian breeds clustered together and separately from the Welsh White Park and Chillingham breeds, the latter two not clustering with any of the other taurine breeds suggesting their genetic distinctiveness. The Treemix analysis revealed a similar topology with the Welsh White Park and Chillingham clustering together and then clustering to Hereford and Angus, rather than clustering with the Italian breeds (**Figure 5**, SM Figures 3, 4). Adding up to seven or eight migration edges to the tree topology improved the amount of the variance explained by the phylogenetic model, albeit marginally (f statistic ~ 0.99965; **Figure 5**; Pickrell and Pritchard, 2012). The model without migration edges already presents an f statistic of 0.9981 (**Figure 5** inset), a value above the threshold described previously, and above which the phylogram was not better explained by adding additional migration events (Pickrell and Pritchard, 2012; Decker et al., 2014). However, more migration edges did not seem to further increase this variance as the f statistic reached an asymptote between 7 and 8 migration edges (**Figure 5** inset). The largest increase in variance occurred by adding the first migration edge to the graph from Brahman to the ancestor of Chianina and Romagnola, consistently with the Admixture results that suggest these Italian breeds have a minor (~10%) indicine genetic component. With the exception of two edges linking Brahman, Angus and Somba, the remaining migration edges were between taurine breeds. All migration edges had a weight less than ~0.2, with the exception of those from Jersey into Hereford, and from the ancestor of Hereford, White Park and Chillinghan, into Charolais, both of which have a weight above 0.4, suggesting the source population made a substantial genetic contribution to the recipient breeds.

Demographic History

The trend in historic effective population sizes for each breed was estimated using the distribution of linkage disequilibrium across the genome. All breeds showed a declining effective population size over the last ~2000 generations (i.e., approximately 8000 years assuming an average generation length of 4 years throughout these species history; **Figure 6**). Interestingly the African breeds showed an average higher effective population size for each species than the non-African breeds. Consistent with their hybrid nature, the five hybrid breeds presented a larger effective population size, probably reflecting the artificial increase in heterozygosity deriving from the admixture event (**Figure 6**).

Signals of Selection

A replicated approach using XP-EHH was used to identify consistent SNPs showing signatures of selection in the dataset (**Figure 1**). Four pairwise comparisons between taurine and indicine breeds were carried out resulting in 10,150 SNPs that passed the MAF threshold within population and which were tested in at least three out of the four pairwise comparisons. Among these SNPs, 3029 (~30%) showed consistent signatures of selection in the taurine breeds and 2385 (~24%) in the indicine breeds. The comparison between African and non-African indicine breeds resulted in 9388 SNPs tested, out of which 1768 (~19%) showed signatures of selection specific to the Asian breeds and 2049 (~22%) showed signatures of selection specific to the African breeds. For the comparison between taurine breeds, a similar number of SNPs was tested, 9938 SNPs. Of these 1872 (~19%) showed signatures of selection in European cattle breeds and 1739 (~17%) in African breeds. The SNPs showing consistent signatures of selection were linked to neighboring

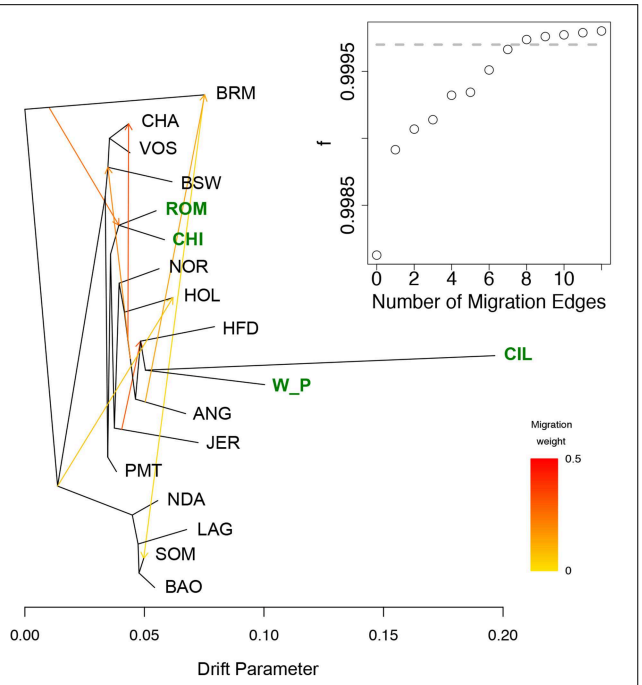


FIGURE 5 | Phylogenetic network of the inferred relationships between 18 cattle breeds. The phylogenetic network inferred by Treemix of the relationships between Welsh White Park cattle breed and 17 other breeds is shown with Brahman (BRM) as outgroup. Migration edges between breeds are shown with arrows pointing in the direction toward the recipient breed of the migrants, and colored according to the ancestry percent received from the donor breed. The new datasets produced for this project are highlighted in green. The inset shows the f index representing the fraction of the variance in the sample covariance matrix ($\widehat{W}$) accounted for by the model covariance matrix (W), as a function of the number of modeled migration events; the gray dashed line marks the number of migrant edges beyond which the f statistic asymptots.

genes that occurred up to 50,000 bp downstream or upstream of the SNP position. No significantly enriched GO categories were found for any of the sets of breeds analyzed here. However, some patterns were observed, e.g., for the candidate SNPs under selection in Asian indicine several genes related to immunity were identified, e.g., LIPH (involved in platelet aggregation), Tespa1 (involve in thymocyte development), CCL14 (activates monocytes) (SM Table 1). However, interestingly, several RNA types were identified for several of the groups of breeds tested (**Table 2**). For the comparison between taurine and indicine breeds 6 snoRNA and 10 snRNA were identified among the taurine breeds, and two miRNA and 1 snoRNA in the indicine breeds. For the comparison between Asian and African indicine, a total of three miRNA, 1 snoRNA and 2 snRNA were identified for the Asian breeds and none for the African, while for the taurine comparison, the opposite was true, i.e., 1 snoRNA and 1 snRNA were identified for the African breeds and none for the European breeds. Overall this suggests that a substantial amount of selection signature occurred near non-coding RNAs that modify transcription of genes or of other RNAs.

Discussion

This dataset represents one of the most comprehensive analyses of *B. taurus* and *B. indicus* cattle breed genetic data. Overall a higher observed genetic variation was observed in taurine breeds that originated or derived from European cattle breeds (∼0.29), than in the other cattle breeds. Both African taurine and indicine breeds presented an intermediate level of observed heterozygosity (∼0.22 in both sets) and the Asian indicine breeds presented the lowest observed heterozygosity (∼0.17). While the observed difference could have a generic demographic explanation, e.g., a larger effective population size in the taurine breeds of European descent, such pattern is not consistent with what has been described with other types of molecular markers such as microsatellites (Rothammer et al., 2013). The most likely explanation for this pattern is ascertainment bias (Gautier et al., 2009; Matukumalli et al., 2009). In particular, for those breeds not included in the panel of breeds used to identify SNPs in the BovineSNP50 chip assay, minimum allele frequencies are expected to be substantially lower and the genetic variation captured with the chosen SNPs may not be reflective of the true underlying genetic variation in those breeds. In this case, SNP discovery was based on the genome sequence reads of five taurine breeds (Holstein, Angus, Limousin, Jersey, Norwegian Red) and one indicine (Brahman) breed compared to the reference genome of a Hereford cow (Bovine Hapmap et al., 2009). Ascertainment bias deriving from the design of this SNP array was previously described for a subset of the dataset analyzed here, where up to 30% of the SNPs had a MAF > 0.3 in taurine breeds, while only 19% of the SNPs had an equivalent MAF in indicine breeds (Nelore, Brahman, and Gir) (Bovine Hapmap et al., 2009). Similarly, the analysis of Holstein, Angus and Brahman samples carried out with a subset of the SNPs in the BovineSNP50 chip assay ascertained in each of the breeds separately, showed how Brahman animals presented a lower heterozygosity using the taurine SNPs (p-value < 0.05 and p-value $= 0.07$, respectively) than using the indicine SNPs (Neto and Barendse, 2010). Contrastingly, in the same analysis, for the Holstein and Angus derived SNPs, the taurine samples presented an observed heterozygosity typical to that of any other taurine breed. Furthermore, as no African taurine breeds and most indicine breeds were not included in the SNP ascertainment, the observed levels of genetic variation are likely to be an underestimate of the true genetic variation in these breeds (Matukumalli et al., 2009), and the associated inbreeding coefficient may consequently be an overestimate of the real ones. Thus, comparisons of genetic diversity discussed here are only likely to be meaningful within major genetic groups.

Among all cattle, the British Chillingham breed possessed unusually low observed heterozygosity (0.026) and an extremely high inbreeding coefficient (0.92). The Chillingham occurs in Northumberland, UK, and for the last ∼300 years has lived in a feral state with almost no human intervention and no immigration (Visscher et al., 2001). Additionally, this breed passed through a bottleneck in 1947 where the herd reduced to five males and eight females. Previous analysis of genetic variation using 25 microsatellites identified only one marker with

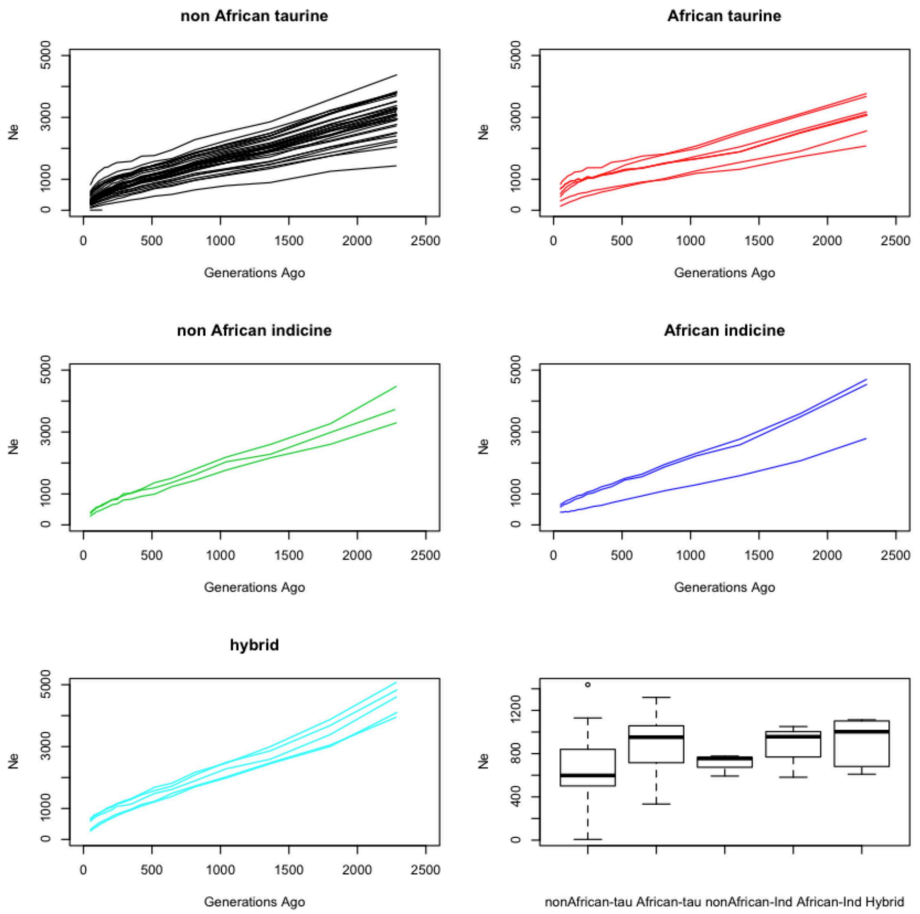


FIGURE 6 | Demographic history reconstruction. Each plot shows the trend in effective population size for a set of cattle breeds between ∼50 and ∼2300 generations ago. The bottom right graph shows boxplots for the harmonic-mean of the effective population sizes trends estimated with SNeP for the various groups of cattle breeds analyzed.

polymorphism (Visscher et al., 2001). Additionally, sequences of the full mitochondria of eight animals (sequencing depth ∼2935x using next generation sequencing) revealed a single haplotype in the herd that differentiated them from other taurine breeds by three mutations, and one mutation (likely to be a recurrent mutation) that related them to indicine breeds and yak (Hudson et al., 2012). These studies suggest that Chillingham probably survives because the population has purged most deleterious mutations. The results shown here represent the largest study of genetic variation in this breed so far and did find similarly low levels of genetic variation in the breed (∼900 polymorphic SNPs out of ∼35,000). An extended analysis including more individuals and genetic markers in this breed may identify genetic variation of evolutionary importance not captured in this study, as well as shedding light on the distribution of homozygous and heterozygous tracts in the breed that could be used to monitor its genetic health.

Population Structure

The analyses of population structure easily identified the taurine and indicine genetic components in the dataset (**Figure 2**). However, when conditioning the number of clusters in the data to $K = 2$, the African taurine cattle presented a genetic composition similar to that of the American hybrid breeds Beef Master and Santa Gertrudis, the latter two are breeds of recent origin (less than 100 years old). Increasing the number of clusters to $K = 3$ clearly separates the African taurines in their own cluster, suggesting that despite the strong non-African taurine and indicine genetic component (as seen for $K = 2$), African taurines (both longhorn and shorthorn) have a substantial third genetic component specific to Africa which may derive from cross breeding with local aurochs (Mctavish et al., 2013). Consistent with previous observations, the African taurine N'Dama (ND1 and ND2) and Somba (SOM) breeds presented a minor indicine genetic component (Gautier et al., 2010). Among the hybrids, Beefmaster (BMA) and Santa Gertrudis (STG) presented a major genetic component of non-African taurine origin and the remaining of indicine origin, in accordance with their breed formation. Beefmaster and Santa Gertrudis are both hybrid breeds of North American origin resulting from the cross between European taurine cattle of British origin (Hereford and Shorthorn) and Brahman cattle (Mctavish et al.,

TABLE 2 | Classes of molecules under selection.

Molecule Class	Taurine vs. Indicine	
	Taurine	**Indicine**
Protein Coding	15	14
RNA (mi, sno, sn)	0,6,10	2,1,0
rRNA	0	0
Pseudogene	1	1
	Indicine	
	Asian	African
Protein coding	46	2
RNA (mi, sno, sn)	3,1,2	0
rRNA	3	0
Pseudogene	3	0
	Taurine	
	European	African
Protein coding	5	1
RNA (mi, sno, sn)	0	0,1,1
rRNA	0	0
pseudogene	0	0

DNA sequences coding for molecules in close proximity to SNPs under selection were grouped as Protein Coding, RNA types (mi, sno, and sn), rRNA and pseudogenes. The number of SNPs under selection close to each of these classes is shown for each of the three analyses presented.

2013). On the other hand, the remaining three hybrid breeds, the Borgou, Kuri, and Sheko present a genetic make-up of African taurine and indicine origin, with the first two having a somewhat larger taurine proportion than the Sheko breed. Among the taurine breeds, the Moroccan Oulmès Zaer (OUL) breed of unknown origin, presented approximately ∼3/5 of its genotype of non-African taurine type and the remaining of African taurine type (**Figure 2**), consistent with previous results suggesting an influence of African longhorn taurine into the otherwise European taurine genetic background of this breed (Gautier et al., 2010; Decker et al., 2014). Interestingly, the non-African taurine cattle presented around 10% or less of their genetic make-up showing African taurine ancestry (**Figure 2**), possibly deriving form cross Mediterranean movement of cattle during the last millennia (Dürst, 1899; Do Vale, 1907; Felius et al., 2011). A further indicine genetic component was also detected in the French Charolais and Gascon breeds (<5%), and the Italian Romagnola and Chianina breeds (∼10%). These results support observations by Decker et al. (2014) that suggest indicine admixture in European taurine breeds (the bulk of non-African taurines in this study) is rare instead of widespread (Mctavish et al., 2013). The genetic variation of African taurine is clearly distinct from that of other cattle breeds, and the PCA analysis suggests that the three main cattle groups (African and non-African taurine, and indicine) are almost as different from each other. While these results do not provide evidence for or against the hypothesis of a third domestication event in Africa (Felius et al., 2011), it suggests that African taurine cattle present a unique characteristics in their gene-pool.

The analysis of population structure using Admixture resulted in asymptoting CV values reflecting problems in identifying the true number of clusters in the data, possibly due to the presence of family structure within the breeds, or relatively low divergence between the estimated clusters. Thus, we also carried out principal components analysis, and as expected from the problems with Admixture, there is relatively low variation among cattle groups, i.e., the largest principal component (PC1) only explains ∼10% of the variance in the dataset (separating taurine form indicine breeds), while PC2 explains ∼5% of the variance (separating African from non-African taurines). Consistent with the Admixture result, the combination of PC1 and PC2 places the hybrid breeds Kuri, Borgou and Sheko between the African taurine and African indicine clusters, with the Sheko much closer to the indicine cluster reflecting its larger indicine background than Kuri and Borgou. The American Beef Master and Santa Gertrudis are placed between the non-African taurine and Asian indicine, but somewhat closer to the taurine cluster reflecting the larger taurine component of their genetic variation.

Lastly, the NeighbourNet analysis and the dendrogram identify the Madagascan zebu as more similar to the Asian indicine than to the other two African indicine. While this placing may be biased due to the ascertainment bias in the SNP array, it is tempting to suggest that the Madagascan population share a more recent history with Asian indicine breeds than the continental African breeds (e.g., a more recent establishment of the Madagascan population).

Roman Origin of Welsh White Park

Within the United Kingdom, traditional knowledge suggests that the Welsh White Park were brought to Britain by the Romans (Felius et al., 2011). While there is no evidence of the Romans having introduced cattle to Britain (Visscher et al., 2001), PCA places the White Park almost in the middle of the group of non-African taurine breeds (**Figure 3A**), consistently with the results by Decker et al. (2014) based on genotypes from five individuals. In particular, a comparison between Welsh White Park and the three Italian breeds in this dataset, Chianina, Romagnola and Piedmontese, revealed that the Italian breeds are more closely related to each other than to White Park, and the Welsh breed is more closely related to the other British breeds (**Figures 3B**, **5**). Nevertheless, while this suggests that Welsh White Park are not similar to the Italian breeds it still does not answer the question regarding their origin. Both the NeighbourNet (**Figure 4**) and the phylogenetic network (**Figure 5**) show that the Welsh White Park and Chillingham cluster near each other (**Figure 4**). Such observation could be the outcome of genetic drift or inbreeding (or an interplay of these two parameters) in these breeds driving the divergence of their allelic frequencies (particularly in the case of Chillingham). Consistently with this hypothesis, the phylogenetic network analysis modeling genetic drift suggests that Welsh White Park and Chillingham have been through substantial drift in comparison to other taurines (**Figure 5**). Additionally, the Treemix analysis suggests that no migration events connect either the Welsh White Park or Chillingham with any of the Italian breeds, unless an additional edge that marginally increases the f-statistic by 0.01% (i.e., from 99.96 to 99.97% of the variance explained) is added; however, that additional migration edge is from Welsh White

Park into Romagnola (contrary to common believe of an Italian ancestry of White Park) and has a weight near 0 (SM Figure 4) indicating that a very small fraction (if at all) of the Romagnola allelic variation may be of Welsh origin. While such idea could be appealing to some, it does not necessarily reflect a real historic event as the inferred edge provides an almost negligible contribution to the variance in the phylogram and corresponds to a minimal contribution of Welsh allelic variants into Romagnola. Additionally, the question of whether the inferred patterns of migration by Treemix are somehow dependent on the data analyzed could be raised. Further studies on the genetic history of these iconic breeds should include additional breeds such as the Podolian Hungarian Gray, or breeds close to the domestication center in order to better characterize the origin of this animals.

Demographic History

The demographic history of a population can be traced using the distribution of LD across the genome (e.g., Hill and Robertson, 1968), and if the cattle domestication event was marked by a strong bottleneck, it is expected that such a pattern should be visible with this dataset. However, it is important to note that the analyses presented here might provide a biased estimate of the true extent of demographic history, as LD calculated over SNP array markers is expected to decay at a lower rate than when the same estimates are obtained from whole-genome sequence data (Qanbari et al., 2014).

All the breeds analyzed present trends marked by a decline in their effective population size during the last ∼2000 generations (**Figure 6**). Assuming an average generation length of 4 years, that suggests a constant decline across all breeds since the onset of the domestication process approximately 8000 years ago (Gautier et al., 2010). These results are consistent with observations describing a general decrease in breeds such as Angus, Holstein, Hereford, N'Dama and Brahman over the same period of time (Gautier et al., 2007; De Roos et al., 2008; Flury et al., 2010). Interestingly, overall, the harmonic-mean effective population size of the non-African taurine breeds was smaller than that of the non-African indicine, but both of those means were much smaller than that of the African taurines and indicines (**Figure 6**). These results suggest that either crossing between African taurine and indicine breeds has increased the effective population size in these breeds by moving genetic variation between breeds, and/or that more benign management practices in Africa have resulted in a milder reduction in effective population size. Although the second hypothesis cannot be ruled out, consistent with the first hypothesis, the hybrid breeds present overall a higher average effective population size.

Signatures of Selection

The unique nature of the data presented here provided us with a potentially powerful statistical framework to detect signatures of selection. In contrast to other studies, we examined multiple breeds for each species, accounting for their distribution in and out of Africa to search for signatures of selection in the equivalent of a replicated experimental design, where instead of identifying selection in a particular breed, we used several breeds of each species in each continent as biological replicas. The test used is an extension of Sabeti's extended haplotype homozygosity (Sabeti et al., 2002), which is based on the principle that a site under positive selection will rapidly increase in frequency, bringing neighboring variants to a high frequency in a process called hitchhicking (Smith and Haigh, 1974), thereby removing genetic variation around the site under selection (a selective sweep) and leaving a track of extended homozygosity. The magnitude of the effect of the selected site on the neighboring variation largely depends on how strong and recent the selective pressure is. If the process is slow, because selection is weak, recombination events separate the selected site from neighboring variants breaking down linkage and consequently resulting in a short area of extended homozygosity (Sabeti et al., 2002). On the contrary, if the process is fast because selection is strong, the area of extended homozygosity will be large, as there is not enough time for recombination to break down linkage. Consequently, identifying old signatures of positive selection using the EHH or any of its derived versions (e.g., XP-EHH) is rather difficult as the expected pattern of linkage disequilibrium is likely to have disappeared (Sabeti et al., 2007; Chen et al., 2010). However, despite the ancient divergence between *B. taurus* and *B. indicus,* i.e., approximately 250,000 years old (Bradley et al., 1996; Bovine Hapmap et al., 2009), a consistent signature of selection was found for several thousands of SNPs (5385 SNPs) probably reflecting relatively recent selective events rather than ancient ones.

Consistent with the expectation for the comparison between breeds in both species, a similar number of SNPs in each species showed signatures of selection (i.e., 3029 in taurine and 2385 in indicine). For the African vs. non-African comparison of taurine breeds 1872 of the SNPs under selection were on European breeds and 1739 were in African breeds. In contrast, for the comparison of African vs. non-African indicine breeds, 1768 of the SNPs under selection were on the Asian indicine breeds and 2049 were in the African breeds. The difference between the comparison of taurine breeds and indicine breeds in and out of Africa may reflect the shorter period of time that the later have spent in Africa since their introduction to that continent. The first domestic cattle to have entered Africa were taurines approximately 7000 years go (Payne and Hodges, 1997; Hassan, 2000; Felius et al., 2014). These domestic animals probably entered via a route crossing through Egypt and moving into northern Africa. In contrast, indicine breeds entered Africa much later via routes leading to East Africa via Egypt some time between ∼3500 and 2500 years ago (Payne and Hodges, 1997; Hanotte et al., 2002). The timing of these migration events into Africa implies that taurine breeds have been in Africa between two and three times longer than indicine breeds.

At least three non-exclusive hypotheses can be derived from the taurine colonization history of Africa. Firstly, African taurine have become clearly differentiated from other non-African taurines as genetic drift exacerbated differences between the taurine propagules that spread from the domestication center into Europe and Africa. Secondly, African taurines had a longer time to hybridize with local aurochs, as possibly was the case for the European taurines, which hybridized with European aurochs (Decker et al., 2014). Consequently, African taurines are expected

to carry unique genetic variation (deriving from local aurochs) that differentiates them from non-African taurines, as African aurochs were likely to genetically differ from European aurochs due to isolation by distance (Perez-Pardal et al., 2010; Mctavish et al., 2013). Thirdly, by having spent a longer time in Africa, African taurines are expected to have had a longer time to adapt to the local environment and although some of their signatures of local adaptation may have dissipated, they may have left a trace of genetic identity that helps differentiating African taurine form other cattle breeds. In contrast, *B. indicus* was introduced to Africa approximately ∼2000 years ago, and during this short period of time African indicine breeds may have not had enough time to differentiate from other indicine breeds, as much as the African taurines did. However, as the presence of indicine breeds in Africa is recent it is reasonable to expect that these carry more signatures of selection than African taurines (∼22 vs. ∼17%, respectively).

The analysis of selection carried out here sought to identify SNPs under selection using the equivalent of a replicated experiment. For that purpose it was necessary that several pairwise comparisons identified the same SNP under selection showing the same sign in the XP-EHH analysis (e.g., a SNP showing selection for the European taurine variant in the three pairwise comparisons). The rationale for this is that genetic drift can cause extreme differences in allele frequencies between populations, however, the chance that the same allele changes allele frequency in the same direction in multiple populations is low (e.g., A gets fixed by chance in the three replicates). While our approach is expected to be robust as it combines the power of the track of linkage disequilibrium left after positive selection and its replicability across populations occupying similar environments (e.g., three European breeds tested against three African breeds), it is possible that weaker signatures of selection were not identified. Nevertheless, it is interesting to note that for the pairwise comparisons where several genes were identified to be in close proximity to the SNPs under selection some chromosomes seems to be more frequent under selection than others. For example, among the Asian indicine, eight genes occurred on the first chromosome, and 11 on the 19th chromosome (SM Table 1), while three out of the four genes identified among the European taurine breeds occurred on the 10th chromosome. Lastly, it is expected that short term adaptive processes may be mediated through modifications of regulatory pathways (Karim et al., 2011; Visscher and Goddard, 2011; David et al., 2013; Furusawa and Kaneko, 2013). Consistently with such hypothesis we find that several SNPs under selection occur in close proximity to RNA coding sequences involved in transcription regulation (e.g., miRNAs) and modification of other RNAs (e.g., snoRNAs).

The domestication of cattle in Eurasia had profound implications for society (Ajmone-Marsan et al., 2010; Groeneveld et al., 2010; Felius et al., 2014). Cattle are one of the major livestock species in the world and represent a large assemblage of widespread breeds as well as locally adapted breeds (Ajmone-Marsan et al., 2010; Felius et al., 2011, 2014). Studying the genetic resource of cattle breeds around the world is an on going work (e.g., Bovine Hapmap et al., 2009; Gautier et al., 2010; Mctavish et al., 2013; Decker et al., 2014) and approaches like the ones used here show the power that multiple comparisons have to derive particular trends and patterns that have characterized these breeds' histories. Additionally, identifying gaps representing regions where cattle has not yet been studied, as well as breeds that have not yet been genetically characterized is of utter importance to identify the genetic variation with adaptive value (e.g., local adaptation) and that should be the target of conservation efforts. Finally, whether Welsh White Park cattle were introduced to Britain by the Romans seems unlikely with the data described here, but cannot be entirely ruled out. The current results suggest that the breed presents a distinct genetic variation possible reflecting its ancient origin, but without some sort of calibration for the molecular phylogeny, an Approximate Bayesian computation approach that estimates times of divergence between breeds, or its comparison to local breeds close to the domestication center, this question is left partially unanswered (Csillery et al., 2010a,b; Sunnaker et al., 2013).

Author Contributions

PO and MWB conceived the project; PO, MB, EN, and FB carried out analyses; WD, DW, MM, AS, PA contributed samples and reagents; PO wrote the manuscript; MB, EN, FB, and MWB revised the manuscript. All authors approved the final manuscript.

Acknowledgments

The authors acknowledge the Chillingham Wild Cattle Association for the collection of samples for DNA analysis, as well as the NextGen consortium (http://nextgen.epfl.ch/) and Genomic-Resources (European Science Foundation; http://www.esf.org/index.php?id=7009) for funding, and two reviewers for their comments.

Supplementary Material

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fgene.2015.00191/abstract

References

Abraham, G., and Inouye, M. (2014). Fast principal component analysis of large-scale genome-wide data. *PLoS ONE* 9:e93766. doi: 10.1371/journal.pone.0093766

Ajmone-Marsan, P., Garcia, J. F., and Lenstra, J. (2010). On the origin of cattle: how aurochs became cattle and colonized the world. *Evol. Anthropol.* 19, 148–157. doi: 10.1002/evan.20267

Alexander, D. H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19, 1655–1664. doi: 10.1101/gr.094052.109

Barbato, M., Orozco-Terwengel, P., Tapio, M., and Bruford, M. (2015). SNeP: a tool to estimate trends in recent effective population size trajectories using genome-wide SNP data. *Front. Genet.* 6:109. doi: 10.3389/fgene.2015.00109

Beja-Pereira, A., Caramelli, D., Lalueza-Fox, C., Vernesi, C., Ferrand, N., Casoli, A., et al. (2006). The origin of European cattle: evidence from

modern and ancient DNA. *Proc. Natl. Acad. Sci. U.S.A.* 103, 8113–8118. doi: 10.1073/pnas.0509210103

Biffani, S., Marusi, M., Biscarini, F., and Canavesi, F. (2005). Developing a genetic evaluation for fertility using angularity and milk yield as correlated traits. *Interbull. Bull.* 33, 63–66.

Bollongino, R., Elsner, J., Vigne, J. D., and Burger, J. (2008). Y-SNPs do not indicate hybridisation between European aurochs and domestic cattle. *PLoS ONE* 3:e3418. doi: 10.1371/journal.pone.0003418

Bovine Hapmap, C., Gibbs, R. A., Taylor, J. F., Van Tassell, C. P., Barendse, W., Eversole, K. A., et al. (2009). Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Science* 324, 528–532. doi: 10.1126/science.1167936

Bradley, D. G., Machugh, D. E., Cunningham, P., and Loftus, R. T. (1996). Mitochondrial diversity and the origins of African and European cattle. *Proc. Natl. Acad. Sci. U.S.A.* 93, 5131–5135. doi: 10.1073/pnas.93.10.5131

Bulmer, M. G. (1971). The effect of selection on genetic variability. *Am. Nat.* 105, 201–211. doi: 10.1086/282718

Chen, H., Patterson, N., and Reich, D. (2010). Population differentiation as a test for selective sweeps. *Genome Res.* 20, 393–402. doi: 10.1101/gr.100545.109

Cobb, M. (2006). Heredity before genetics: a history. *Nat. Rev. Genet.* 7, 953–958. doi: 10.1038/nrg1948

Corbin, L. J., Liu, A. Y., Bishop, S. C., and Woolliams, J. A. (2012). Estimation of historical effective population size using linkage disequilibria with marker data. *J. Anim. Breed. Genet.* 129, 257–270. doi: 10.1111/j.1439-0388.2012.01003.x

Csillery, K., Blum, M. G. B., Gaggiotti, O. E., and Francois, O. (2010b). Invalid arguments against ABC: reply to AR Templeton. *Trends Ecol. Evol.* 25, 490–491. doi: 10.1016/j.tree.2010.06.011

Csillery, K., Blum, M. G., Gaggiotti, O. E., and Francois, O. (2010a). Approximate Bayesian Computation (ABC) in practice. *Trends Ecol. Evol.* 25, 410–418. doi: 10.1016/j.tree.2010.04.001

David, L., Ben-Harosh, Y., Stolovicki, E., Moore, L. S., Nguyen, M., Tamse, R., et al. (2013). Multiple genomic changes associated with reorganization of gene regulation and adaptation in yeast. *Mol. Biol. Evol.* 30, 1514–1526. doi: 10.1093/molbev/mst071

Decker, J. E., Mckay, S. D., Rolf, M. M., Kim, J., Molina Alcala, A., Sonstegard, T. S., et al. (2014). Worldwide patterns of ancestry, divergence, and admixture in domesticated cattle. *PLoS Genet.* 10:e1004254. doi: 10.1371/journal.pgen.1004254

De Roos, A. P., Hayes, B. J., Spelman, R. J., and Goddard, M. E. (2008). Linkage disequilibrium and persistence of phase in Holstein-Friesian, Jersey and Angus cattle. *Genetics* 179, 1503–1512. doi: 10.1534/genetics.107.084301

Do Vale, M. (1907). *Bovídeos Protuguezes* Lisbon: La Bécarre.

Dürst, J. (1899). *Die Rinder von Babylonien, Assyrien un Ägypten und ihr Zusammenhang mit den Rindern der Alten Welt*. Berlin: Gerog Reimber. doi: 10.5962/bhl.title.86398

Eden, E., Navon, R., Steinfeld, I., Lipson, D., and Yakhini, Z. (2009). GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* 10:48. doi: 10.1186/1471-2105-10-48

Edwards, C. J., Ginja, C., Kantanen, J., Perez-Pardal, L., Tresset, A., Stock, F., et al. (2011). Dual origins of dairy cattle farming - evidence from a comprehensive survey of european Y-Chromosomal Variation. *PLoS ONE* 6:e15922. doi: 10.1371/journal.pone.0015922

Fao. (2007). "Global plan of action for animal genetic resources and the Interlaken Declaration," in *International Technical Conference on Animal Genetic Resources for Food and Agriculture* (Rome).

Felius, M., Beerling, M., Buchanan, D., Theunissen, B., Koolmees, P., and Lenstra, J. (2014). On the history of cattle genetic resources. *Diversity* 6, 705–750. doi: 10.3390/d6040705

Felius, M., Koolmees, P., Theunissen, B., Consortium, E. C. G. D., and Lenstra, J. (2011). On the breeds of cattle - historc and current classifictions. *Diversity* 3, 660–692. doi: 10.3390/d3040660

Felsenstein, J. (1993). *PHYLIP (Phylogeny Inference Package) version 3.5c*. 3.69 ed. (Seattle, WA: Department of Genetics, University of Washington).

Flori, L., Gonzatti, M. I., Thevenon, S., Chantal, I., Pinto, J., Berthier, D., et al. (2012). A quasi-exclusive European ancestry in the Senepol tropical cattle breed highlights the importance of the slick locus in tropical adaptation. *PLoS ONE* 7:e36133. doi: 10.1371/journal.pone.0036133

Flury, C., Tapio, M., Sonstegard, T. S., Drögemüller, T., Leeb, H., Simianer, H., et al. (2010). Effective population size of an indigenous Swiss cattle breed estimated fromlinkage disequilibirium. *J. Anim. Breed. Genet.* 127, 339–347. doi: 10.1111/j.1439-0388.2010.00862.x

Furusawa, C., and Kaneko, K. (2013). Epigenetic feedback regulation accelerates adaptation and evolution. *PLoS ONE* 8:e61251. doi: 10.1371/journal.pone.0061251

Gautier, M., Faraut, T., Moazami-Goudarzi, K., Navratil, V., Foglio, M., Grohs, C., et al. (2007). Genetic and haplotypic structure in 14 European and African cattle breeds. *Genetics* 177, 1059–1070. doi: 10.1534/genetics.107.075804

Gautier, M., Flori, L., Riebler, A., Jaffrezic, F., Laloe, D., Gut, I., et al. (2009). A whole genome Bayesian scan for adaptive genetic divergence in West African cattle. *BMC Genomics* 10:550. doi: 10.1186/1471-2164-10-550

Gautier, M., Laloe, D., and Moazami-Goudarzi, K. (2010). Insights into the genetic history of French cattle from dense SNP data on 47 worldwide breeds. *PLoS One* 5:e13038. doi: 10.1371/journal.pone.0013038

Gotherstrom, A., Anderung, C., Hellborg, L., Elburg, R., Smith, C., Bradley, D. G., et al. (2005). Cattle domestication in the Near East was followed by hybridization with aurochs bulls in Europe. *Proc. Biol. Sci.* 272, 2345–2350. doi: 10.1098/rspb.2005.3243

Groeneveld, L. F., Lenstra, J. A., Eding, H., Toro, M. A., Scherf, B., Pilling, D., et al. (2010). Genetic diversity in farm animals–a review. *Anim. Genet.* 41(Suppl. 1), 6–31. doi: 10.1111/j.1365-2052.2010.02038.x

Hanotte, O., Bradley, D. G., Ochieng, J. W., Verjee, Y., Hill, E. W., and Rege, J. E. (2002). African pastoralism: genetic imprints of origins and migrations. *Science* 296, 336–339. doi: 10.1126/science.1069878

Hassan, F. (2000). *The Origins and Development of African Livestock: Archaelogy, Genetics, Linguistics and Ethnography*. London: University College London Press.

Herrero, M., and Thornton, P. K. (2013). Livestock and global change: emerging issues for sustainable food systems. *Proc. Natl. Acad. Sci. U.S.A.* 110, 20878–20881. doi: 10.1073/pnas.1321844111

Hill, W. G., and Robertson, A. (1968). Linkage disequilibrium in finite populations. *Theor. Appl. Genet.* 38, 226–231. doi: 10.1007/BF01245622

Hudson, G., Wilson, I., Payne, B. I., Elson, J., Samuels, D. C., Santibanez-Korev, M., et al. (2012). Unique mitochondrial DNA in highly inbred feral cattle. *Mitochondrion* 12, 438–440. doi: 10.1016/j.mito.2012.05.003

Huson, D. H., and Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* 23, 254–267. doi: 10.1093/molbev/msj030

Kantanen, J., Olsaker, I., Holm, L. E., Lien, S., Vilkki, J., Brusgaard, K., et al. (2000). Genetic diversity and population structure of 20 North European cattle breeds. *J. Hered.* 91, 446–457. doi: 10.1093/jhered/91.6.446

Karim, L., Takeda, H., Lin, L., Druet, T., Arias, J. A., Baurain, D., et al. (2011). Variants modulating the expression of a chromosome domain encompassing PLAG1 influence bovine stature. *Nat. Genet.* 43, 405–413. doi: 10.1038/ng.814

Lewis, J., Abas, Z., Dadousis, C., Lykidis, D., Paschou, P., and Drineas, P. (2011). Tracing cattle breeds with principal components analysis ancestry informative SNPs. *PLoS ONE* 6:e18007. doi: 10.1371/journal.pone.0018007

Loftus, R. T., Machugh, D. E., Bradley, D. G., Sharp, P. M., and Cunningham, P. (1994). Evidence for two independent domestications of cattle. *Proc. Natl. Acad. Sci. U.S.A.* 91, 2757–2761. doi: 10.1073/pnas.91.7.2757

Machugh, D. E., Loftus, R. T., Cunningham, P., and Bradley, D. G. (1998). Genetic structure of seven European cattle breeds assessed using 20 microsatellite markers. *Anim. Genet.* 29, 333–340. doi: 10.1046/j.1365-2052.1998.295330.x

Matukumalli, L. K., Lawley, C. T., Schnabel, R. D., Taylor, J. F., Allan, M. F., Heaton, M. P., et al. (2009). Development and characterization of a high density SNP genotyping assay for cattle. *PLoS ONE* 4:e5350. doi: 10.1371/journal.pone.0005350

Mctavish, E. J., Decker, J. E., Schnabel, R. D., Taylor, J. F., and Hillis, D. M. (2013). New World cattle show ancestry from multiple independent domestication events. *Proc. Natl. Acad. Sci. U.S.A.* 110, E1398–1406. doi: 10.1073/pnas.1303367110

Medjugorac, I., Kustermann, W., Lazar, P., Russ, I., and Pirchner, F. (1994). Marker-derived phylogeny of european cattle supports demic expansion of agriculture. *Anim. Genet.* 25, 19–27. doi: 10.1111/j.1365-2052.1994.tb00399.x

Neto, L. R. P., and Barendse, W. (2010). Effect of SNP origin on analyses of genetic diversity in cattle. *Anim. Product. Sci.* 50, 792–800. doi: 10.1071/AN10073

Payne, W., and Hodges, J. (1997). *Tropical Cattle, Origins, Breeds and Breeding Policies.* Oxford: Blackwell Science Ltd.

Pellecchia, M., Negrini, R., Colli, L., Patrini, M., Milanesi, E., Achilli, A., et al. (2007). The mystery of Etruscan origins: novel clues from *Bos taurus* mitochondrial DNA. *Proc. Biol. Sci.* 274, 1175–1179. doi: 10.1098/rspb.2006.0258

Perez-Pardal, L., Royo, L. J., Beja-Pereira, A., Curik, I., Traore, A., Fernandez, I., et al. (2010). Y-specific microsatellites reveal an African subfamily in taurine (*Bos taurus*) cattle. *Anim. Genet.* 41, 232–241. doi: 10.1111/j.1365-2052.2009.01988.x

Perry, B. D., Grace, D., and Sones, K. (2013). Current drivers and future directions of global livestock disease dynamics. *Proc. Natl. Acad. Sci. U.S.A.* 110, 20871–20877. doi: 10.1073/pnas.1012953108

Pickrell, J. K., and Pritchard, J. K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* 8:e1002967. doi: 10.1371/journal.pgen.1002967

Pryce, J., Royal, M., Garnsworthy, P., and Maod, I. (2004). Fertility in high-producing dairy cow. *Livestock Product. Sci.* 86, 125–135. doi: 10.1016/S0301-6226(03)00145-3

Pryce, J., Veerkamp, R., Thompson, R., Hill, W. G., and Simm, G. (1997). Genetic aspects of common health disorders and measures of fertility in Holstein Friesian dairy cattle. *Anim. Sci.* 65, 353–360. doi: 10.1017/S1357729800008559

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795

Qanbari, S., Pausch, H., Jansen, S., Somel, M., Strom, T. M., Fries, R., et al. (2014). Classic selective sweeps revealed by massive sequencing in cattle. *PLoS Genet.* 10:e1004148. doi: 10.1371/journal.pgen.1004148

Rischkowsky, B., and Pilling, D. (2007). *The State of the World's Animal Genetic Resources for Food and Agriculture.* Rome: Food and Agriculture Organization of the United Nations.

Rothammer, S., Seichter, D., Forster, M., and Medugorac, I. (2013). A genome-wide scan for signatures of differential artificial selection in ten cattle breeds. *BMC Genomics* 14:908. doi: 10.1186/1471-2164-14-908

Russell, N. (2007). *Like Engend'ring Like. Heredity and Animal Breeding in Early Modern England.* Cambridge, MA: Cambridge University Press.

Sabeti, P. C., Reich, D. E., Higgins, J. M., Levine, H. Z., Richter, D. J., Schaffner, S. F., et al. (2002). Detecting recent positive selection in the human genome from haplotype structure. *Nature* 419, 832–837. doi: 10.1038/nature01140

Sabeti, P. C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., et al. (2007). Genome-wide detection and characterization of positive selection in human populations. *Nature* 449, 913–918. doi: 10.1038/nature06250

Scheet, P., and Stephens, M. (2006). A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *Am. J. Hum. Genet.* 78, 629–644. doi: 10.1086/502802

Simm, G. (1998). *Genetic Improvement of Cattle and Sheep.* Wallingford: Farming Press.

Smith, J. M., and Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genet. Res.* 23, 23–35. doi: 10.1017/S0016672300014634

Sunnaker, M., Busetto, A. G., Numminen, E., Corander, J., Foll, M., and Dessimoz, C. (2013). Approximate Bayesian computation. *PLoS Comput. Biol.* 9:e1002803. doi: 10.1371/journal.pcbi.1002803

Sved, J. A. (1971). Linkage disequilibrium and homozygosity of chromosome segments in finite populations. *Theor. Popul. Biol.* 2, 125–141. doi: 10.1016/0040-5809(71)90011-6

Sved, J. A., Mcrae, A. F., and Visscher, P. M. (2008). Divergence between human populations estimated from linkage disequilibrium. *Am. J. Hum. Genet.* 83, 737–743. doi: 10.1016/j.ajhg.2008.10.019

Szpiech, Z. A., and Hernandez, R. D. (2014). selscan: an efficient multithreaded program to perform EHH-based scans for positive selection. *Mol. Biol. Evol.* 31, 2824–2827. doi: 10.1093/molbev/msu211

Taberlet, P., Coissac, E., Pansu, J., and Pompanon, F. (2011). Conservation genetics of cattle, sheep, and goats. *C. R. Biol.* 334, 247–254. doi: 10.1016/j.crvi.2010.12.007

Thornton, P. K. (2010). Livestock production: recent trends, future prospects. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 365, 2853–2867. doi: 10.1098/rstb.2010.0134

Troy, C. S., Machugh, D. E., Bailey, J. F., Magee, D. A., Loftus, R. T., Cunningham, P., et al. (2001). Genetic evidence for Near-Eastern origins of European cattle. *Nature* 410, 1088–1091. doi: 10.1038/35074088

Visscher, P. M., and Goddard, M. E. (2011). Cattle gain stature. *Nat. Genet.* 43, 397–398. doi: 10.1038/ng.819

Visscher, P. M., Smith, D., Hall, S. J., and Williams, J. L. (2001). A viable herd of genetically uniform cattle. *Nature* 409, 303. doi: 10.1038/35053160

Wendorf, F., and Schild, R. (1994). Are the early holocene cattle in the eastern Sahara domestic or wild? *Evol. Anthropol.* 3, 118–128. doi: 10.1002/evan.1360030406

Wood, R. J. (1973). Robert Bakewell (1725-1795), pioneer animal breeder, and his influence on Charles Darwin. *Folia Mendeliana* 58, 231–242.

Xu, L., Bickhart, D. M., Cole, J. B., Schroeder, S. G., Song, J., Tassell, C. P., et al. (2014). Genomic signatures reveal new evidences for selection of important traits in domestic cattle. *Mol. Biol. Evol.* doi: 10.1093/molbev/msu333

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Orozco-terWengel, Barbato, Nicolazzi, Biscarini, Milanesi, Davies, Williams, Stella, Ajmone-Marsan and Bruford. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in GENETICS

ORIGINAL RESEARCH ARTICLE
published: 04 March 2015
doi: 10.3389/fgene.2015.00068

Microsatellite genotyping of medieval cattle from central Italy suggests an old origin of Chianina and Romagnola cattle

Maria Gargani[1]*, Lorraine Pariset[1], Johannes A. Lenstra[2], Elisabetta De Minicis[3], European Cattle Genetic Diversity Consortium[†] and Alessio Valentini[1]

[1] *Department for Innovation in Biological, Agro-food and Forest systems, University of Tuscia, Viterbo, Italy*
[2] *Faculty of Veterinary Medicine, Utrecht University, Utrecht, Netherlands*
[3] *Department of Sciences of Cultural Heritage (DISBEC), University of Tuscia, Viterbo, Italy*

Edited by:
Stéphane Joost, École Polytechnique Fédérale de Lausanne, Switzerland

Reviewed by:
Pablo Orozco ter Wengel, Cardiff University, UK
Letizia Nicoloso, Università degli Studi di MIlano, Italy
Florian Alberto, CNRS, France

***Correspondence:**
Maria Gargani, Department for Innovation in Biological, Agro-food and Forest systems, University of Tuscia, Via S. Camillo de Lellis, 01100 Viterbo, Italy
e-mail: maria.gargani@unitus.it

Analysis of DNA from archeological remains is a valuable tool to interpret the history of ancient animal populations. So far most studies of ancient DNA target mitochondrial DNA (mtDNA), which reveals maternal lineages, but only partially the relationships of current breeds and ancient populations. In this study we explore the feasibility of nuclear DNA analysis. DNA was extracted from 1000-years old cattle bone collected from Ferento, an archeological site in central Italy. Amplification of 15 microsatellite FAO-recommended markers with PCR products yielded genotypes for four markers. Expected heterozygosity was comparable with values of modern breeds, but observed heterozygosity was underestimated due to allelic loss. Genetic distances suggested a position intermediate between (1) Anatolian, Balkan, Sicilian and South-Italian cattle and (2) the Iberian, North-European and Central-European cattle, but also a clear relationship with two central-Italian breeds, Chianina and Romagnola. This suggests that these breeds are derived from medieval cattle living in the same area. Our results illustrate the potential of ancient DNA for reconstructing the history of local cattle husbandry.

Keywords: microsatellite, ancient DNA, cattle, Ferento, NeighborNet

INTRODUCTION

The study of ancient DNA (aDNA) has developed since about 28 years from the analysis of short segments of mitochondrial DNA (mtDNA) to the spectacular whole-genome analysis of close relatives of the *Homo sapiens* (Higuchi et al., 1984; Pääbo, 1985; Reich et al., 2010). It complements archeological and historical investigations by providing ancestral links between ancient samples and present organisms (Vernesi et al., 2004; Caramelli et al., 2007). The majority of aDNA studies still target mtDNA, which is facilitated by its high copy number per cell. This allowed to determine the sequences of mtDNA control regions extracted from cattle remains excavated at several sites (Lenstra et al., 2014), but also the entire mitochondrial genome of two *Bos primigenius* specimens (Edwards et al., 2010; Lari et al., 2011). However, the genealogy inferred from the mtDNA not necessarily reflect that inferred with nuclear markers. Analysis of nuclear DNA from ancient remains is more challenging, but has been reported in several studies (Greenwood et al., 1999, 2001; Noonan et al., 2005; Pariset et al., 2007; Green et al., 2010; Reich et al., 2010). Ancient genotypes of nuclear microsatellites markers are informative for genetic diversity, subdivision and geographical origin (Edwards et al., 2003; Allentoft et al., 2011; Ishida et al., 2012; Nyström et al., 2012). Moreover, there are several studies carried out with the so called FAO list of microsatellites in almost every continent on a huge amount of individuals of several breeds (Ajmone-Marsan and The GLOBALDIV Consortium, 2010). This valuable database could be related to ancient remains in order i. to discover the closeness of ancient individuals to extant populations to understand the type of husbandry and production of the time and ii. to infer the origin, migration and admixture of present breeds.

Authentic cattle breeds in central and southern Italy are of the Podolian type (Felius, 1995). However, it is not yet clear if these are of ancient origin or have been imported (Felius et al., 2014). The aim of this study was to gain insight into history of Italian cattle by comparing 1000-year old cattle DNA

[†]The following members of the European Cattle Genetic Diversity Consortium contributed to this study: K. Moazami-Goudarzi, D. Laloë, INRA, Jouy-en-Josas, France; J. L. Williams (presently at the Parco Tecnologica Padano, Lodi, Italy), P. Wiener, Roslin Institute, UK; D. G. Bradley, Trinity College, Dublin, Ireland; G. Erhardt, Justus-Liebig Universität, Giessen, Germany; B. Harlizius, School of Veterinary Medicine, Hannover, Germany; I. Medugorac, Ludwig-Maximilians-Universität, Munich C. Looft, E. Kalm, Christian-Albrechts-Universität, Kiel, Germany; J. Cañón, S. Dunner, Universidad Complutense de Madrid, Spain; C. Rodellar, I. Martín-Burriel, Veterinary Faculty, Zaragoza, Spain; A. Sanchez, Universitat Autonoma de Barcelona; A. M. Martínez, Universidad de Córdoba, Córdoba, Spain; C. Ginja, Instituto Superior de Agronomia, Lisboa, Portugal; P. Ajmone-Marsan, Universitá Cattolica del S. Cuore, Piacenza, Italy; F. Pilla, A. Bruzzone, Universitá del Molise, Campobasso, Italy; D. Marletta, Università degli Studi di Catania, Catania, Italy; L. E. Holm, Danish Institute of Agricultural Sciences, Tjele, Denmark; G. Dolf, University of Berne, Switzerland.

of Italian origin with DNA from modern cattle. DNA was extracted from bones collected in Ferento, an archeological site near Viterbo in central Italy inhabited since Bronze Age, but developed mainly during Roman and Medieval ages. It rises on a triangular plateau made of tuff stone [Pianicara, IGM F.137 II NE] and delimited by two ditches: Guzzarella (or Vezzarella) and Acqua Rossa, subaffluents of the Tiber. The archeological investigations in Ferento revealed five areas, named assay I, assay II, assay III, assay IV, and assay V. The assay I comprises the area northward the Decumano in the medieval quarter; the assay II is localized on west area at the limit of the plateau, where an important necropolis was found; the assay III was occupied with a residential building on the southern side of the Decumano and with an important system of tanks, leant against the Theater; the assay IV, placed in the northwestern area of the plateau, and the assay V, near the assay I close to the fortification.

We compared the microsatellite genotypes obtained for four microsatellites with those of modern cattle breeds in order to uncover the origins of the Central-Italian medieval cattle.

MATERIALS AND METHODS

ANCIENT AND MODERN SAMPLES

Thirty bone samples of medieval cattle (∼1000 BP) were collected at Ferento, an archeological site near Viterbo in central Italy (**Table 1**). The analyzed remains were recovered from five different excavation areas. The dating was made on the basis of pottery found in the same layer. Moreover, one samples were sent to CEDAD (CEntre for DAting and Diagnostics) belonging to the University of Salento (Department of Innovation Engineering) and were carbon-dated using high resolution mass spectrometry. The results confirmed the age estimated by pottery layers. Species identification, for the selection of the samples to be analyzed, was based on morphology and dimensions of the specimens.

In order to avoid contamination and degradation by sudden change in environmental conditions, samples were picked up using latex gloves, put immediately in vacuum-pack, stored at −20°C (Pariset et al., 2007) and transferred to a lab dedicated to ancient DNA analysis. In a dedicated laboratory room, the outer surface was removed with sand paper and bone powder was collected by drilling. DNA was extracted according to Yang et al. (1998) and quantified using a DTX Multimode Detector

Table 1 | Medieval samples.

Sample	Element	Age	Excavation areas	GPS	Amplification
Ferento 1	Astragalus	∼1000 BP	5	42°29′21.2″N, 12°07′55.8″E	No
Ferento 2	Phalanx II	∼1000 BP	3	42°29′18.4″N, 12°07′53.9″E	Yes
Ferento 3	Phalanx I	∼1000 BP	4	42°29′19.2″N, 12°07′52.4″E	No
Ferento 4	Heel	∼1000 BP	4	42°29′19.2″N, 12°07′52.4″E	Yes
Ferento 5	Molar 3	∼1000 BP	3	42°29′18.4″N, 12°07′53.9″E	Yes
Ferento 6	Phalanx I	∼1000 BP	4	42°29′19.2″N, 12°07′52.4″E	Yes
Ferento 7	Astragalus	∼1000 BP	4	42°29′19.2″N, 12°07′52.4″E	Yes
Ferento 8	Astragalus sx	∼1000 BP	4	42°29′19.2″N, 12°07′52.4″E	Yes
Ferento 9	Jawbone	∼1000 BP	3	42°29′18.4″N, 12°07′53.9″E	No
Ferento 10	Phalanx II	∼1000 BP	4	42°29′19.2″N, 12°07′52.4″E	Yes
Ferento 11	Jawbone	∼1000 BP	4	42°29′19.2″N, 12°07′52.4″E	Yes
Ferento 12	Heel	∼1000 BP	4	42°29′19.2″N, 12°07′52.4″E	No
Ferento 13	Jawbone	∼1000 BP	2	42°29′19.4″N, 12°07′47.2″E	Yes
Ferento 14	Metatarsus	∼1000 BP	3	42°29′18.4″N, 12°07′53.9″E	No
Ferento 15	Heel	∼1000 BP	5	42°29′21.2″N, 12°07′55.8″E	Yes
Ferento 16	Phalanx I	∼1000 BP	4	42°29′19.2″N, 12°07′52.4″E	No
Ferento 17	Astragalus	∼1000 BP	1	42°29′20.6″N, 12°07′56.7″E	No
Ferento 18	Phalanx I	∼1000 BP	1	42°29′20.6″N, 12°07′56.7″E	Yes
Ferento 19	Phalanx I	∼1000 BP	1	42°29′20.6″N, 12°07′56.7″E	No
Ferento 20	Metatarsus	∼1000 BP	4	42°29′19.2″N, 12°07′52.4″E	Yes
Ferento 21	Phalanx I	∼1000 BP	4	42°29′19.2″N, 12°07′52.4″E	Yes
Ferento 22	Metatarsus	∼1000 BP	4	42°29′19.2″N, 12°07′52.4″E	Yes
Ferento 23	Phalanx II	∼1000 BP	3	42°29′18.4″N, 12°07′53.9″E	Yes
Ferento 24	Heel	∼1000 BP	5	42°29′21.2″N, 12°07′55.8″E	Yes
Ferento 25	Metatarsus	∼1000 BP	2	42°29′19.4″N, 12°07′47.2″E	No
Ferento 26	Phalanx I	∼1000 BP	3	42°29′18.4″N, 12°07′53.9″E	No
Ferento 27	Humerus	∼1000 BP	2	42°29′19.4″N, 12°07′47.2″E	No
Ferento 28	Phalanx II	∼1000 BP	1	42°29′20.6″N, 12°07′56.7″E	No
Ferento 29	Phalanx I	∼1000 BP	2	42°29′19.4″N, 12°07′47.2″E	No
Ferento 30	Metatarsus	∼1000 BP	4	42°29′19.2″N, 12°07′52.4″E	No

880 (Beckman) with Picogreen method (Quant-iT PicoGreen, Invitrogen). At least two independent DNA extractions were performed. One single sample per day was extracted with a negative control. Comprehensive datasets of modern samples of genotypes of cattle breeds from Europe, Africa and Asia (Table S1) have been described previously (European Cattle Genetic Diversity Consortium, 2006; Medugorac et al., 2009; Laloë et al., 2010; Felius et al., 2011; Delgado et al., 2012).

MICROSATELLITE GENOTYPING

Fifteen microsatellites were selected from the FAO list of recommended microsatellites (Table S2) (FAO, 2011). Only for three of these markers the published primers generated PCR products shorter than 160 bp. In order to optimize the amplification with aDNA templates, we redesigned primers for the other 12 microsatellites (Table S2) using Primer3 software. In a separate laboratory room, a first-round of PCR was carried out using a 20 µl reaction volume containing PCR buffer with 2.5 mM $MgCl_2$, 200 µM of each dNTPs, 0.05 µM of both the forward (fluorescent labeled dyes) and the reverse primer (Proligo), 0.2 units of Bio-x-act short Taq Polymerase (Bioline) and 10 ng of DNA. A 5 min denaturation step at 94°C was followed by 30 cycles of 30 s denaturation at 94°C, 1 min at the respective annealing temperatures (Table S2) and 1 min of extension at 68°C, followed by a final 5 min extension step at 68°C. Products were purified with exosap (Promega). A second round of PCR was performed with as template the product (1 µl) of the first PCR reaction. During each PCR experiment both extraction and the PCR negative controls were checked for the absence of amplification products. At least four amplifications per sample were performed in order to monitor for allelic drop out. PCR products were separated by electrophoresis on a CEQ 8800 sequencer (Beckman Coulter) and sized with the standard CEQ™ DNA Size Standard Kit—400 (Beckman Coulter). The alleles were scored using the proprietary CEQ fragment analysis software.

Microsatellite dataset from modern cattle breeds were obtained from the EU Resgen project. In order to compare the allele lengths of the ancient cattle with those of the modern cattle database, we have included in all plates subject to PCR three modern samples (ITROM10, ITROM11, ITROM20) as a reference. The ancient cattle alleles were recalculated for comparison with the reference individuals. Genotypes were scored as heterozygous when the electropherogram showed a clear biallelic profile or when two different mono-allelic profiles were observed in different PCRs (Allentoft et al., 2011). The allelic profiles have been deposited in the Dryad database (http://dx.doi.org/10.5061/dryad.d4500). Allele frequencies and observed and expected heterozygosity were calculated for each population using MSToolkit. Genetic distances were calculated by using the program MSAT2 (http://genetics.stanford.edu/hpgl/projects/microsat/) and visualized in NeighborNet graphs with SplitsTree4 program (http://www.splitstree.org/). The NeighborNet graphs were simplified by combining several breeds in regional pools as approximations of ancestral gene populations. Thus, we made metapopulations of the dairy breeds from the Northwest-European Lowland, Central Brown cattle (including Swiss-Brown, Rendena, Bruna, and Cabannina), a West-Central cluster (Simmental and its derivatives); Iberian cattle, Sicilian Modicana and Cinesara, Balkan Busha, Anatolian cattle and Indopakistan zebu breeds, respectively. Principal component analysis of a matrix of F_{ST} genetic distances was performed using the program Genalex (Peakall and Smouse, 2006). Unsupervised model-based clustering was done with the program Structure (Pritchard et al., 2000).

RESULTS

Extraction of genomic DNA yielded for all 30 ancient samples quantifiable DNA and for 16 of these also amplification products. From the 15 microsatellites, four (INRA5, HEL1, MM12, CSRM60) could be amplified successfully in at least one sample. A total of 760 amplifications were performed on ancient samples. The success rate PCR per sample and per locus is shown in Table S3. Markers INRA5, HEL1, MM12, and CSRM60 yielded genotypes in 6, 8, 10, and 10 samples, respectively. All extraction and PCR controls for different DNA isolation, purification and amplification steps did not yield PCR products. In 35 PCR reactions we observed allelic dropout of 50% of heterozygous genotypes for HEL1, 60% for CSRM60, 75% for MM12, 73% for INRA5 and 64% averaged of the four loci, but never more than two alleles per marker in a given sample. The expected (He) and observed (Ho) heterozygosity values for a panel of European breeds calculated on the basis of 4 microsatellites ranged from 0.48 (Lagunaire breed) to 0.771 (Istrian breed) and from 0.43 (Lagunaire breed) to 0.76 (Istrian breed) respectively. The He and Ho values for pools of breeds ranged from 0.68 (Zebu breed pool) to 0.786 (Sicilian breed pool) and from 0.62 (Zebu breed pool) to 0.79 (Anatolian breed pool) respectively. These values were in the same range of those based on 30 microsatellites: He and Ho for panel of European breeds ranged from 0.544 (Lagunaire breed) to 0.741(Podolica breed) and from 0.51(Lagunaire breed) to 0.73 (Bohemian Red breed); He and Ho for pools of breed ranged from 0.67 (Zebu breed pool) and 0.780 (Anatolian breed pool) and from 0.59 (Zebu breed pool) and 0.72 (Busha breed pool) respectively (Table S1). The expected heterozygosity (0.698) in Ferento samples is within the range of the values found for modern breeds. We observed a clear heterozygote deficit (observed heterozygozity 0.588) related to allelic dropout for ancient samples.

Principal component analysis (Figure S1) as well as model-based clustering of Ferento cattle (not shown) combined with a panel of European breeds typed with the same markers differentiated three clusters: (1) African cattle, (2) Northwestern-European Lowland dairy breeds, and (3) Central-European and Iberian cattle as well as the ancient samples from Ferento. However, PCA did not establish relationships between Ferento cattle and individual breeds, since relative positions of breeds could not be reproduced with other panels of four microsatellites (not shown). We are aware that four microsatellites may have a low power in resolving diversity in populations that diverged since very few generations, however, we could not increase the number of markers that successfully and repeatedly amplified ancient DNA within the FAO list that permits to relate ancient to extant breeds.

In order to get a more informative comparison of ancient Ferento genotypes with those of other cattle, we constructed NeighborNet graphs. **Figure 1** shows NeighborNet graphs of the clusters with several modern Podolian breeds and ancient

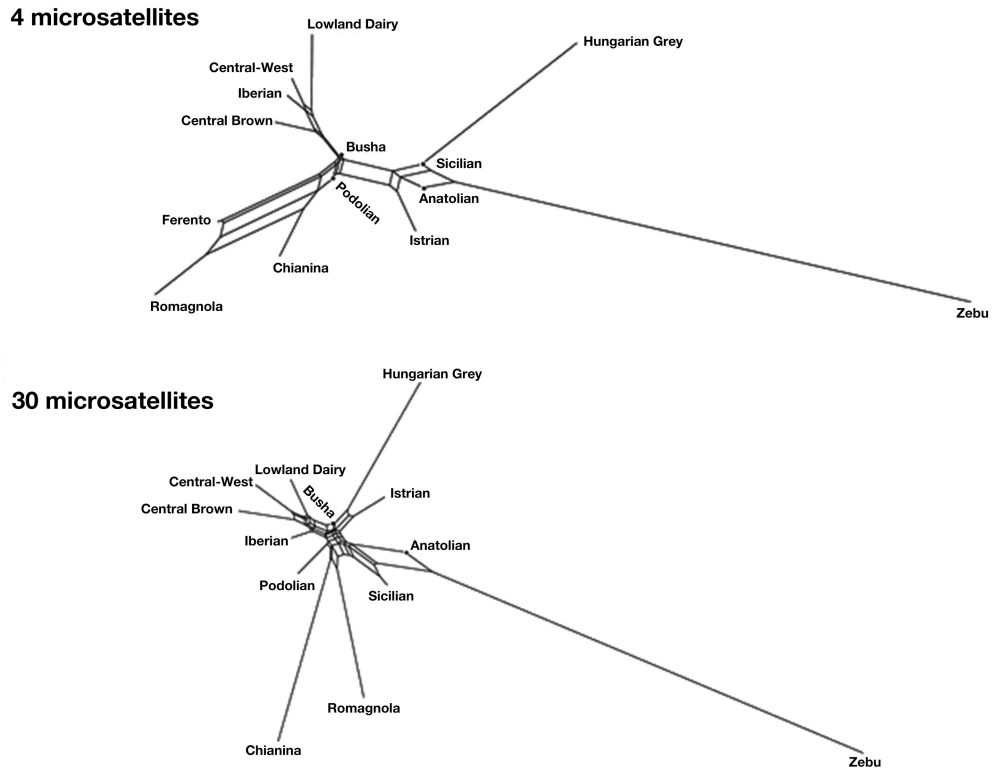


FIGURE 1 | NeighborNet graphs of Reynold's distances of Ferento cattle with modern Podolian breeds, and European, Anatolian and Indian breed clusters based on 4 or 30 microsatellites. Ferento cattle is only shown with 4 microsatellites.

Ferento cattle. Although distances on the basis of only four microsatellites are not expected to be accurate, these distances reproduce the clustering that are generated by the 30 microsatellites. Interestingly, Ferento cattle is clearly linked with the Tuscanian Chianina and the related Romagnola cattle. The ancient and modern central Italian cattle is intermediate between (1) the Anatolian, Sicilian and Balkan Podolian breeds, and (2) the Iberian, Central European and Lowland dairy breeds. We observed that none of the individuals from modern breeds were as close to Ferento as the Chianina and Romagnola breeds.

DISCUSSION

In this study we demonstrate the feasibility of genotyping microsatellites for the study of cattle of 1000 years ago. The rates of amplification we have obtained (between 11 and 19%) appear to be lower than those found in in previous studies (Edwards et al., 2003; Nyström et al., 2012), which is probably due to relatively high temperature in central Italy. Although this limits a more elaborate analysis of the ancient genotypes, our results reconstruct the history of Italian cattle. The origin of the Podolian cattle in Italy is controversial (Felius et al., 2014). The first cattle arrived in the Italian peninsula at the start of the Neolithic (ca. 8000 BP). These long-horned cattle were largely replaced by short-horned cattle around 4500–5000 BP. Neolithic cattle gradually decreased in size and, as in the rest of Europe, Bronze and Iron age cattle in Italy had wither heights of typically 110 cm. However, cattle with wither heights of 115–135 cm were imported from Epirus in the centuries before the Roman era and are probably the ancestors of the large Roman cattle with wither heights of 150 cm or more (Kron, 2002). The analysis of mitochondrial DNA from modern taurine cattle has detected five mtDNA haplogroup with a specific geographic distribution (Lenstra et al., 2014). The relatively high frequencies of the T and T2 haplogroups of Italian Podolian cattle (Bonfiglio et al., 2010; Lenstra et al., 2014) may be explained by these imports of cattle. In an alternative scenario Anatolian cattle were brought to Tuscany by the founders of the Etruscan civilization (Pellecchia et al., 2007). After the Roman era, the large cattle disappeared from the European fossil record and cattle became even smaller than in the Iron age with a typical wither heights of 95 cm for the medieval Ferento cattle. Since the Renaissance cattle became larger and the Tuscan Chianina is now the largest cattle of the world. This breed and two other central-Italian breeds, Romagnola and Marchigiana, are related to the South Italian and Balkan long-hornd Podolian breeds (Felius, 1995). These cattle are believed to have emerged after the 12th century by local selection for long horns (Bökönyi, 1974). There have been several opportunities for gene flow between the Balkans and Italy, e.g., the consecutive migrations of Wisigoths, Osthrogoths and Lombards and the large scale import of Hungarian Gray cattle as "meat on the hoof" after the renaissance via Venice. The last imports have been well documented and involved not only oxes, but also fertile bulls (Appuhn, 2010). The deviating mtDNA haplogroup distribution of the Italian Podolian cattle indicates that

any introgression of Balkan cattle was via bulls only (Lenstra et al., 2014). Alternatively, it has also been proposed that Italian Podolian cattle descends from ancient local cattle or even local aurochs (Ciani and Matassino, 2001).

The present data contributes to a historic reconstruction by indicating a link of the modern large, white and short-horned Chianina and Romagnola with local cattle living in the same area 1000 years earlier. This argues against an origin of the Chianina, Romagnola and the close relative Marchigiana from the postmedieval imports. Since the Germanic invasions are not likely to have replaced completely the local cattle, it is plausible that Ferento cattle descends from cattle kept during the Roman era. The large white central-Italian and Podolian cattle do not form a tight cluster in the network, which leaves open the possibility of an eastern origin of the long-horned Italian Podolian cattle. This may include the long-horned and feral Maremmana, for which no genotypes are available for all four microsatellites, but is in other networks close to the Hungarian Gray.

In conclusion our study demonstrates the feasibility and the potentiality of the analysis of nuclear DNA from ancient livestock remains.

AUTHOR CONTRIBUTIONS

MG designed the study, carried out microsatellite amplifications experiments, and drafted the manuscript, LP contributed to the design of the work, JL contributed substantially to the analysis and interpretation of data and revised the work, ED provided ancient samples bones, ECGDC provided modern dataset, AV participated in developing ideas and revised the manuscript.

ACKNOWLEDGMENTS

This work has been supported by the European Commission (project ResGEN 98-118). The content of this publication does not represent the views of the Commission or its services.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://www.frontiersin.org/journal/10.3389/fgene.2015.00068/abstract

REFERENCES

Ajmone-Marsan, P., and The GLOBALDIV Consortium. (2010). A global view of livestock biodiversity and conservation—GLOBALDIV. *Anim. Genet.* 41, 1–5. doi: 10.1111/j.1365-2052.2010.02036.x

Allentoft, M. E., Oskam, C., Houston, J., Hale, M. L., Gilbert, M. T. P., Rasmussen, M., et al. (2011). Profiling the dead: generating microsatellite data from fossil bones of extinct megafauna—protocols, problems, and prospects. *PLoS ONE* 6:e16670. doi: 10.1371/journal.pone.0016670

Appuhn, K. (2010). Ecologies of beef: eighteenth-century epizootics and the environmental history of early modern Europe. *Environ. Hist.* 15, 268–287. doi: 10.1093/envhis/emq049

Bökönyi, S. (1974). *History of Domestic Mammals in Central and Eastern Europe*. Budapest: Akadémiai Kiadó.

Bonfiglio, S., Achilli, A., Olivieri, A., Negrini, R., Colli, L., Liotta, L., et al. (2010). The enigmatic origin of bovine mtDNA haplogroup R: sporadic interbreeding or an independent event of Bos primigenius domestication in italy? *PLoS ONE* 5:e15760. doi: 10.1371/journal.pone.0015760

Caramelli, D., Vernesi, C., Sanna, S., Sampietro, L., Lari, M., Castrì, L., et al. (2007). Genetic variation in prehistoric Sardinia. *Hum. Genet.* 122, 327–336. doi: 10.1007/s00439-007-0403-6

Ciani, F., and Matassino, D. (2001). Il bovino grigio allevato in Italia; origine. Nota 1: il bovino Macrocero. *Taurus Speciale* 13, 89–99.

Delgado, J. V., Martínez, A. M., Acosta, A., Alvarez, L. A., Armstrong, E., Camacho, E., et al. (2012). Genetic characterization of Latin-American Creole cattle using microsatellite markers. *Anim. Genet.* 43, 2–10. doi: 10.1111/j.1365-2052.2011.02207.x

Edwards, C. J., Connellan, J., Wallace, P. F., Park, S. D. E., McCormick, F. M., Olsaker, I., et al. (2003). Feasibility and utility of microsatellite markers in archaeological cattle remains from a Viking Age settlement in Dublin. *Anim. Genet.* 34, 410–416. doi: 10.1046/j.0268-9146.2003.01043.x

Edwards, C. J., Magee, D. A., Park, S. D. E., McGettigan, P. A., Lohan, A. J., Murphy, A., et al. (2010). A complete mitochondrial genome sequence from a mesolithic wild aurochs (Bos primigenius). *PLoS ONE* 5:9255. doi: 10.1371/journal.pone.0009255

European Cattle Genetic Diversity Consortium. (2006). Marker-assisted conservation of European cattle breeds: an evaluation. *Anim. Genet.* 37, 475–481. doi: 10.1111/j.1365-2052.2006.01511.x

FAO. (2011). *Molecular Genetic Characterization of Animal Genetic Resources*. Rome: FAO Animal Production and Health Guidelines, No. 9.

Felius, M. (1995). *Cattle Breeds-An Encyclopedia*. Doetinchem: Misset.

Felius, M., Beerling, M-L., Buchanan, D. S., Theunissen, B., Koolmees, P. A., and Lenstra, J. A. (2014). On the history of cattle genetic resources. *Diversity* 6, 705–750. doi: 10.3390/d6040705

Felius, M., Koolmees, P. A., Theunissen, B., European Cattle Genetic Diversity Consortium, and Lenstra, J. A. (2011). On the breeds of cattle—historic and current classifications. *Diversity* 3, 660–692. doi: 10.3390/d3040660

Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., et al. (2010). A draft sequence of the Neandertal genome. *Science* 328, 710–722. doi: 10.1126/science.1188021

Greenwood, A. D., Capelli, C., Possnert, G., and Pääbo, S. (1999). Nuclear DNA sequences from late Pleistocene megafauna. *Mol. Biol. Evol.* 16, 1466–1473. doi: 10.1093/oxfordjournals.molbev.a026058

Greenwood, A. D., Lee, F., Capelli, C., DeSalle, R., and Tikhonov, A. (2001). Evolution of endogenous retro-virus-like elements of the woolly mammoth (Mammuthus primigenius) and its relatives. *Mol. Biol. Evol.* 18, 840–847. doi: 10.1093/oxfordjournals.molbev.a003865

Higuchi, R., Bowman, B., Freiberger, M., Ryder, O. A., and Wilson, A. C. (1984). DNA sequences from the quagga, an extinct member of the horse family. *Nature* 312, 282–284. doi: 10.1038/312282a0

Ishida, Y., Roca, A. L., Fratpietro, S., and Greenwood, A. D. (2012). Successful Genotyping of Microsatellites in the Woolly Mammoth. *J. Hered.* 103, 459–464. doi: 10.1093/jhered/esr139

Kron, G. (2002). Archaeozoological evidence for the productivity of Roman livestock farming. *MBAH* 21, 53–73.

Laloë, D., Moazami-Goudarzi, K., Lenstra, J. A., Ajmone-Marsan, P., Azor, P., Rodero, E., et al. (2010). Spatial trends of domestic ruminants in Europa. *Diversity* 2, 932–945. doi: 10.3390/d2060932

Lari, M., Rizzi, E., Mona, S., Corti, G., Catalano, G., Chen, K., et al. (2011). The Complete Mitochondrial Genome of an 11,450-year-old Aurochsen (Bos primigenius) from Central Italy. *BMC Evol. Biol.* 11:32. doi: 10.1186/1471-2148-11-32

Lenstra, J. A., Ajmone-Marsan, P., Beja-Pereira, A., Bollongino, R., Bradley, D. G., Colli, L., et al. (2014). Meta-analysis of mitochondrial DNA reveals several population bottlenecks during worldwide migrations of cattle. *Diversity* 6, 178–187. doi: 10.3390/d6010178

Medugorac, I., Medugorac, A., Russ, I, Veit-Kensch, C. E., Taberlet, P., Luntz, B., et al. (2009). Genetic diversity of European cattle breeds highlights the conservation value of traditional unselected breeds with high effective population size. *Mol. Ecol.* 18, 3394–3410. doi: 10.1111/j.1365-294X.2009.04286.x

Noonan, J. P., Hofreiter, M., Smith, D., Priest, J. R., and Rohland, N. (2005). Genomic sequencing of Pleistocene cave bears. *Science* 309, 597–599. doi: 10.1126/science.1113485

Nyström, V., Humphrey, J., Skoglund, P., Mc Keown, N. J., Vartanyan, S., Shaw, P. W., et al. (2012). Microsatellite genotyping reveals end-Pleistocene decline in mammoth autosomal genetic variation. *Mol. Ecol.* 21, 3391–3402. doi: 10.1111/j.1365-294X.2012.05525.x

Pääbo, S. (1985). Molecular cloning of ancient Egyptian mummy DNA. *Nature* 314, 644–645. doi: 10.1038/314644a0

Pariset, L., Gabbianelli, F., De Bernardis, D., Fortunato, M. T., Alhaique, F., De Minicis, E., et al. (2007). Ancient DNA: genomic amplification of roman and medieval bovine bones. *Ital. J. Anim. Sci.* 6, 179–181. doi: 10.4081/ijas.2007.1s.179

Peakall, R., and Smouse, P. E. (2006). GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol. Ecol. Notes* 6, 288–295. doi: 10.1111/j.1471-8286.2005.01155.x

Pellecchia, M., Negrini, R., Colli, L., Patrini, M., Milanesi, E., Achilli, A., et al. (2007). The mystery of Etruscan origins: novel clues from Bos taurus mitochondrial DNA. *Proc. Biol. Sci.* 274, 1175–1179. doi: 10.1098/rspb.2006.0258

Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959.

Reich, D., Green, R. E., Kircher, M., Krause, J., Patterson, N., Durand, E. Y., et al. (2010). Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* 468, 1053–1060. doi: 10.1038/nature09710

Vernesi, C., Caramelli, D., Dupanloup, I., Bertorelle, G., Lari, M., Cappellini, E., et al. (2004). The Etruscans: a population-genetic study. *Am. J. Hum. Genet.* 74, 694–704. doi: 10.1086/383284

Yang, D. Y., Eng, B., Waye, J. S., Dudar, J. C., and Saunders, S. R. (1998). Technical note: improved DNA extraction from ancient bones using silica-based spin columns. *Am. J. Phys. Anthropol.* 105, 539–543.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 25 November 2014; accepted: 10 February 2015; published online: 04 March 2015.

Citation: Gargani M, Pariset L, Lenstra JA, De Minicis E, European Cattle Genetic Diversity Consortium and Valentini A (2015) Microsatellite genotyping of medieval cattle from central Italy suggests an old origin of Chianina and Romagnola cattle. Front. Genet. ***6****:68. doi: 10.3389/fgene.2015.00068*

This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics.

Copyright © 2015 Gargani, Pariset, Lenstra, De Minicis, European Cattle Genetic Diversity Consortium and Valentini. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in GENETICS

ORIGINAL RESEARCH ARTICLE
published: 05 January 2015
doi: 10.3389/fgene.2014.00442

Hybrid origin of European commercial pigs examined by an in-depth haplotype analysis on chromosome 1

Mirte Bosse*, Ole Madsen, Hendrik-Jan Megens, Laurent A. F. Frantz, Yogesh Paudel, Richard P. M. A. Crooijmans and Martien A. M. Groenen

Animal Breeding and Genomics Centre, Wageningen University, Wageningen, Netherlands

Edited by:
Johann Sölkner, University of Natural Resources and Life Sciences Vienna (BOKU), Austria

Reviewed by:
Georgia Hadjipavlou, Agricultural Research Institute, Cyprus
Zhi-Liang Hu, Iowa State University, USA

****Correspondence:***
Mirte Bosse, Animal Breeding and Genomics Centre, Wageningen University, Radix Building number 107, Droevendaalsesteeg 1, 6708 PB Wageningen, Netherlands
e-mail: mirte.bosse@wur.nl

Although all farm animals have an original source of domestication, a large variety of modern breeds exist that are phenotypically highly distinct from the ancestral wild population. This phenomenon can be the result of artificial selection or gene flow from other sources into the domesticated population. The Eurasian wild boar (*Sus scrofa*) has been domesticated at least twice in two geographically distinct regions during the Neolithic revolution when hunting shifted to farming. Prior to the establishment of the commercial European pig breeds we know today, some 200 years ago Chinese pigs were imported into Europe to improve local European pigs. Commercial European domesticated pigs are genetically more diverse than European wild boars, although historically the latter represents the source population for domestication. In this study we examine the cause of the higher diversity within the genomes of European commercial pigs compared to their wild ancestors by testing two different hypotheses. In the first hypothesis we consider that European commercial pigs are a mix of different European wild populations as a result of movement throughout Europe, hereby acquiring haplotypes from all over the European continent. As an alternative hypothesis, we examine whether the introgression of Asian haplotypes into European breeds during the Industrial Revolution caused the observed increase in diversity. By using re-sequence data for chromosome 1 of 136 pigs and wild boars, we show that an Asian introgression of about 20% into the genome of European commercial pigs explains the majority of the increase in genetic diversity. These findings confirm that the Asian hybridization, that was used to improve production traits of local breeds, left its signature in the genome of the commercial pigs we know today.

Keywords: *Sus scrofa*, hybridization, domestication, introgression, genetic variation, haplotype homozygosity

INTRODUCTION

Domestication is a complex process that has major implications for both phenotypic and genetic variation. It is not an exception that the domesticated form appears to be very different from the wild species in terms of phenotype and genetic makeup. Examples include multiple crop species (Doebley et al., 2006), dogs (vonHoldt et al., 2010) and farm animals (Andersson, 2001; Dobney and Larson, 2006). The differences are caused mainly by two phenomena: (1) selection for particular traits in the domesticated population including domestication genes, which can either facilitate the maintenance of the species in question or have commercial interest; (2) hybridization with individuals from highly divergent populations to improve selected traits. The domesticated pig (*Sus scrofa*) is a good example of such a species, since the domesticated form, as well as its wild relatives, is widespread across the Eurasian continent although phenotypically it can be highly distinct. Domestication of the pig is known to have its origin independently in the Near East and in Asia roughly 10,000 years ago (ya), which led to at least two distinct domestication clades (Kijas and Andersson, 2001; Larson et al., 2005).

Strong artificial selection after the initial domestication led to a wide variety of breeds, each with distinct phenotypes, and selective signatures in the genome (Rubin et al., 2012; Wilkinson et al., 2013). Breed formation and artificial selection for particular traits can drastically reduce genetic diversity, which has been shown for multiple species (Kristensen and Sorensen, 2005; Taberlet et al., 2008). Surprisingly, in pigs, the commercial breeds in Europe are generally more diverse than their wild counterparts (Groenen et al., 2012; Bosse et al., 2014a). In this research we examine which process contributed most to the difference in genetic diversity between European commercial breeds and European wild boars.

In Europe, pig domestication did not occur as a single, unique event, but rather was a continuous process of domestication, isolation and hybridization that led to the domestic European pigs seen today (Larson et al., 2007). Furthermore, glaciations likely had a major impact on the genetic diversity in European wild boar (Scandura et al., 2008). It has been suggested that there were multiple refugia in Europe during the last glaciation, resulting in many private haplotypes for the separate populations (Alves et al., 2010). In the drawn-out process of domestication of the pig in Europe, the mixing of wild boar genetic variation from different regions in Europe, might explain the high diversity found in modern European pigs. Although variation has been

lost locally in most European wild populations, the combined genetic diversity from geographically isolated populations should display similar patterns of genetic diversification as is shown for European commercial haplotypes. The first hypothesis we test, therefore, is that the European breeds are a combination of separate European populations that have been amalgamated into a single population, resulting in higher levels of variation.

Introgression from Asian pigs into European breeds was first demonstrated with molecular data by Giuffra et al. (2000), and indeed multiple international breeds have subsequently been found to contain Far Eastern mitochondrial haplotypes (Clop et al., 2004; Fang and Andersson, 2006). Ramirez et al. (2009) suggested that this introgression was mostly female driven, because of the predominance of the European HY1 Y-chromosomal haplotype in European domestic pigs. An Asian origin for multiple commercially important phenotypes has been shown to be the result of this hybridization (Ojeda et al., 2008; Wilkinson et al., 2013; Bosse et al., 2014b; Hidalgo et al., 2014). Alves et al. (2003) showed that not all European local domestic breeds, such as Iberian pigs, contain mtDNA of Asian origin, and based on studies of genomic DNA, varying levels of admixture in local breeds have been suggested (Herrero-Medrano et al., 2013). We recently found that regions in the genome of Large White pigs that contain DNA that is shared with Asian pigs are generally more diverse than regions that do not share DNA with Asian haplotypes (Bosse et al., 2014a). However, it is unknown whether this is a direct result of the introgression (rather than, for example, incomplete lineage sorting). Moreover, how much the introgressed Asian haplotypes contributed to variation in the genome of European commercial pigs remains an unanswered question. Thus, the second hypothesis we test is that the Asian introgression has led to higher diversity in the European commercial pigs.

For prioritizing farm animal genetic resources (FanGR) for conservation, it is important to know the distribution and the origin of variation in the (domesticated) species (Groeneveld et al., 2010). With this work, we make a contribution by analyzing the details of genetic diversity on chromosome 1 within and between groups of pigs and wild boars in Asia and Europe.

MATERIALS AND METHODS

DATA

The data used for this paper consists of all variants on chromosomes 1, 2, and 18 that were observed in 136 pigs. These variants have previously been deposited into dbSNP (release 138). The data was obtained by aligning Illumina paired-end 100bp reads to the *Sus scrofa* reference genome (build 10.2) using Mosaik Aligner (V.1.1.0017). Reads were trimmed to a minimum base PHRED quality of 20 averaged over 3 consecutive bases and only mate pairs with both reads at least 45 bp in length were included. Each individual was sequenced to $\sim 10\times$ depth of coverage. SNPs were called separately per individual with SAMtools (V. 0.1.13) pileup with a minimum coverage of 4x and with at least 2 reads supporting the alternative allele. Sites were filtered for a minimum genotype and mapping PHRED quality of 20. Most of our analyses were based on all 2,747,210 variants called on chromosome 1. From the original matrix containing all variable sites in all 136 pigs, indels were excluded and SNP loci were retained if called in >80% of all individuals. The minimum coverage of genotypes called within each group of pigs was set to >80%, resulting in 410,237 high-quality SNPs on chromosome 1. All individuals were imputed and phased for these 410,237 SNPs with Beagle v.3.3.2. Although it is unknown whether the two haplotypes represent the actual phases, we considered them to be one full-length haplotype, as uncertainties in phase should balance out when homozygosity rates are calculated for all haplotype pairs in the dataset. We pooled the haplotypes from pigs belonging to the 8 groups listed in **Table 1**.

Table 1 | Number and haplotypes per group and background of sequenced individuals.

Group	No. haplotypes	Codes	Population details
Outgroup	4	INDO	(wild) Sumatran *Sus scrofa*
European local	32	AS,BB,BK,BS,GO,LB,LE,LS,MW,TA,NS	Heritage breeds (Old British breeds), Less global breeds (Linderodsvin, Bunte Bentheimer, Angler Sattelschwein, Leicoma, Nera Siciliana)
European Iberian	22	CA,CM,CS,CT, MA,NI	Pigs from the Iberian peninsula (Mangalica, Negro Iberico, Casertana, Chato Murciano, Calabrese, Cinta Senese)
European commercial	94	DU,HA,LR,LW,PI	Widespread commercial breeds (Duroc, Hampshire, Landrace, Large White, Pietrain)
European wild	52	WB21,22,25,26,28,31,32,33,42,44,72	Wild boar from Western, South-Eastern and Southern Europe (Netherlands, France, Spain, Italy, Switzerland, Greece, Samos, Armenia)
Asian commercial	30	JQ,MS,XI	Asian breeds known to be commercially important (Meishan, Xiang, Jianquahai)
Asian local	18	JI,LSP,TH,WS,ZA	Local breeds and wild pigs (Jinhua, Leping spotted, Wannan spotted, Zhang, Thai)
Asian wild	20	WB20,29,30	North China, South China, Japan

The group name of the pigs under "group" is how this group of individuals is referred to in the rest of the text. The codes of all pigs correspond to their labels in ***Figure 1****. The details of the populations or breeds that the pigs belong to are summarized in the column "Population details." Note that information for the European local and Asian local individuals can be limited, and therefore these are rather heterogeneous groups.*

PHYLOGENETIC ANALYSIS

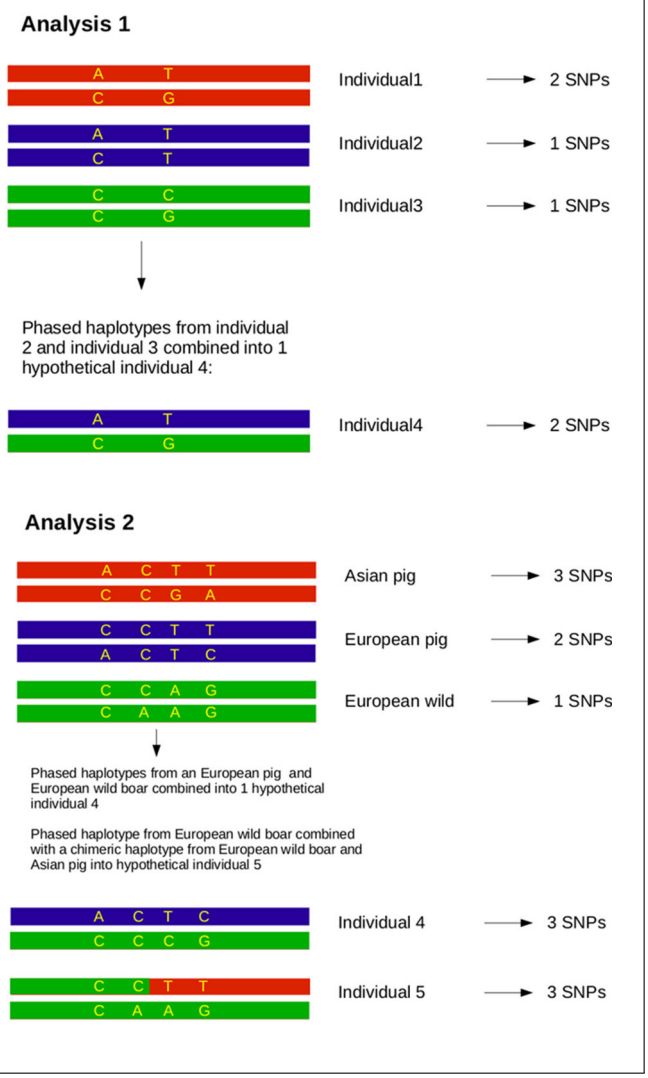
To assess the relationship of haplotypes in our dataset, we constructed a phylogenetic tree based on the phased haplotypes. Each haplotype was considered as an independent sample, so that haplotypes belonging to the same individual do not necessarily need to cluster together. Because missing sites were imputed with Beagle, no missing alleles were present in the phased haplotypes. Sites with more than two alleles were removed from the data and a distance matrix was constructed in PLINK (Purcell et al., 2007). NEIGHBOR (PHYLIP V. 3.695; Felsenstein, 2005) was used to build a neighbor-joining tree for all haplotypes using two Sumatran *Sus scrofa* as outgroup, and the tree was depicted using FIGTREE (http://tree.bio.ed.ac.uk/software/figtree/).

HAPLOTYPE HOMOZYGOSITY ANALYSIS

Analysis 1

After individuals were phased for the full length of chromosome 1, the homozygosity was analyzed between two haplotypes spanning the full chromosome for all possible combinations of two haplotypes in the dataset. Haplotype homozygosity is defined as the proportion of homozygous sites between two paired haplotypes, and ranged from 0 to 1. We calculated haplotype homozygosity as the proportion of all sites (410,237) that occurred in homozygous state, so that 0 represents only heterozygous loci and 1 represents complete homozygosity between both haplotypes. We then paired all possible combinations of two haplotypes in the dataset and determined the homozygosity of these hypothetical diploid individuals in R (see **Box 1**). Haplotype homozygosity was pooled for pairs of haplotypes belonging to the same group (**Table 1**), so that we ended up with a distribution of homozygosity within a group that represents the full range of variation between haplotypes in a group. Within-group haplotype homozygosity was then compared between the different groups. In the second part of this analysis haplotypes from two different groups were paired and the haplotype homozygosity for these mixed pairs was computed to obtain a distribution of homozogosities between haplotypes. This distribution was then compared with the distribution of homozygosity between haplotypes from two other groups.

Box 1 | Principles of the analyses.



Analysis 2

Previous estimates on the fraction of Asian DNA ranged from 20 to a maximum of 35% (Groenen et al., 2012; Bosse et al., 2014b). In the second analysis we wanted to assess the influence of Asian introgression into a European haplotype. In order to do this, we simulated introgression by transferring 15, 20, and 25% of a haplotype belonging to the Asian commercial group into a haplotype that belongs to the European wild group (see **Box 1**). We used a custom perl script to construct these chimeric haplotypes in which 15, 20, or 25% of the alleles coming from an Asian commercial haplotype replace the alleles in a European wild haplotype. All possible pairs between European wild and Asian commercial haplotypes to construct a chimeric haplotype were included. Then, these chimeric haplotypes were again paired with all possible European wild haplotypes (except for the one that the chimeric haplotype is constructed of) and the homozygosity between the two haplotypes was calculated as described for analysis 1. These haplotype homozygosities were pooled so that a distribution of haplotype homozygosity in the artificially created Asian-European hybrids was obtained.

CONSISTENCY OVER CHROMOSOMES

All analyses presented in this paper are based on haplotypes spanning the full length of chromosome 1. We selected this chromosome because it is the longest pig chromosome and therefore the introgression signals are probably most representative for the full genome and less prone to occasional aberrations due to a limited recombination/drift. However, to check whether chromosome 1 is representative for the complete genome, we compared the haplotype homozygosities for the same pairs of individuals between chromosome 1 and two other chromosomes: chromosome 2 (the second longest chromosome), and the shortest and acrocentric chromosome 18. We tested the correlation coefficient between the haplotype homozygosities of the different chromosomes with Pearson's product-moment correlation in R.

RUNS OF HOMOZYGOSITY

We extracted runs of homozygosity (ROH) from all combinations of paired haplotypes coming from the European pigs and wild boars. ROHs were called with the –homozyg option using PLINK v1.07, allowing for one heterozygous site in the ROH and a minimum ROH size of 10Kb.

RESULTS AND DISCUSSION

VARIATION WITHIN GROUPS

We analyzed the phylogenetic relationship of all haplotypes spanning chromosome 1 by constructing a neighbor-joining tree (**Figure 1**). The Asian and European haplotypes form two distinct clusters, which is consistent with the hypothesis of independent domestication (Kijas and Andersson, 2001; Larson et al., 2005; Groenen et al., 2012; Ramírez et al., 2014). European wild boars constitute a monophyletic clade within the European commercial pigs. The pig reference genome sequence (Groenen et al., 2012) clusters within a group of Duroc pigs, which is expected because the reference genome is based on a female Duroc. The Chinese commercial and local haplotypes cluster with the Northern and Southern Chinese wild haplotypes. The only exception is the Zhang pig, which is closer to European pigs (labeled "ZA" in **Figure 1**). This individual is possibly introgressed with European breeds and therefore we mention explicitly when this individual is included in the analysis. Haplotypes from the same individual generally cluster together, but within the European commercial group this is not always the case, showing the close relationship of these individuals. The Japanese wild boar (WB20) and the Mangalica pigs (MA) are the most inbred individuals, with homozygosity between the two haplotypes within each individual above 0.99.

Branches within the Asian cluster are longer than those for European haplotypes. When the homozygosity between two haplotypes from individuals with the same background is measured, the variation (within groups) between two Asian haplotypes is indeed higher than between two European haplotypes from the same group, except for the Japanese wild boar (**Figure 2**). This is congruent with previous findings that *Sus scrofa* has its origin in Asia (Groenen et al., 2012; Frantz et al., 2013) and that European pigs experienced a stronger bottleneck during the last glaciation, resulting in reduced variation (Bosse et al., 2012). Independent domestication should lead to Asian local and commercial pigs being more variable than European pigs, which has been shown previously based on microsatellite data (Megens et al., 2008) and sequence data (Bosse et al., 2012) and is also supported by our analysis (**Figure 2**).

Before and even after the establishments of modern breeds, hybridization between different European populations was common practice. Therefore European commercial pigs are all thought to contain Asian haplotypes. However, this is not necessarily the case for all local breeds in Europe. Our results show that variation between haplotypes from European local breeds is lower than between European commercial haplotypes, which could be due to less Asian introgression or because they have a less mixed European origin (Herrero-Medrano et al., 2013). Some breeds from the Iberian peninsula and old British heritage breeds cluster with the European wild boar (**Figure 1**) which suggests that the source population for domestication more closely resembles these breeds and wild boar, and that genetic differentiation between those pigs is low as recently described by Ramírez et al. (2014). In line with our expectations, we find that variation between two European wild haplotypes is generally lower than between two European commercial haplotypes, especially when variation within individuals is not considered. These findings serve as initial concept of our further analyses.

CONSISTENCY OVER CHROMOSOMES

We did an in-depth analysis of haplotypes on chromosome 1, but first verified whether chromosome 1 is actually a representative model for the rest of the (autosomal) genome. The correlation between haplotype homozygosity for pairs of haplotypes of chromosome 1 and haplotype homozygosity for chromosome 18 is 0.9848, and between chromosome 1 and chromosome 2 is 0.9874. Looking at the homozygosities for pairs of haplotypes on chromosome 1 and pairs of haplotypes on chromosome 18 (**Figure S1**), two small clouds of dots stand out: one having a higher homozygosity on chromosome 18 (red) and the other having a lower homozygosity on chromosome 18 compared to chromosome 1 (orange). These clouds actually represent the haplotypes from only two Asian pigs WS01U03 (red) and ZA01U02 (orange) in combination with all European haplotypes, suggesting a different level of European introgression into the different chromosomes for these two pigs. Since the overall correlation coefficients are so high for the rest of the paired haplotypes in the dataset, we conducted the rest of the analyses only on chromosome 1 and excluded these two individuals from further analyses.

VARIATION IN WILD BOARS

Sus scrofa probably originated in South-East Asia. To assess the full width of variation that is present within the species in the wild, we measured variation for all possible pairs of haplotypes in the dataset. The lowest homozygosity between haplotypes is observed when a haplotype is paired with an outgroup haplotype (the peak at ∼0.72 in **Figure S1**). The geographic region closest to the center of origin is often the richest in genetic diversity, as shown for other species like dogs and humans (Long and Kittles, 2003; vonHoldt et al., 2010). Indeed, our analysis corroborate that the divergence between haplotypes is larger when at least one haplotype is Asian than when no Asian haplotypes are present (**Figures 2**, **3**). Eastern and Western *Sus scrofa* diverged around 1.2 Mya and this divergence resulted in a multitude of fixed differences between both wild populations (Groenen et al., 2012). Naturally, this divergence also contributes to genetic variation within the species, and to quantify the unique contributions of both continents to variation within the species we looked at the difference in homozygosity between paired haplotypes from the same continent and paired haplotypes from Europe and Asia. For mainland *Sus scrofa*, most divergence between haplotypes is found when a European wild haplotype is pooled with an Asian haplotype, regardless its domestication status. The fact that we do not find a significant difference in homozygosity between an Asian wild or Asian local and commercial haplotype paired with a European wild haplotype suggests that the time since the most recent common ancestor is similar and that generally no or very

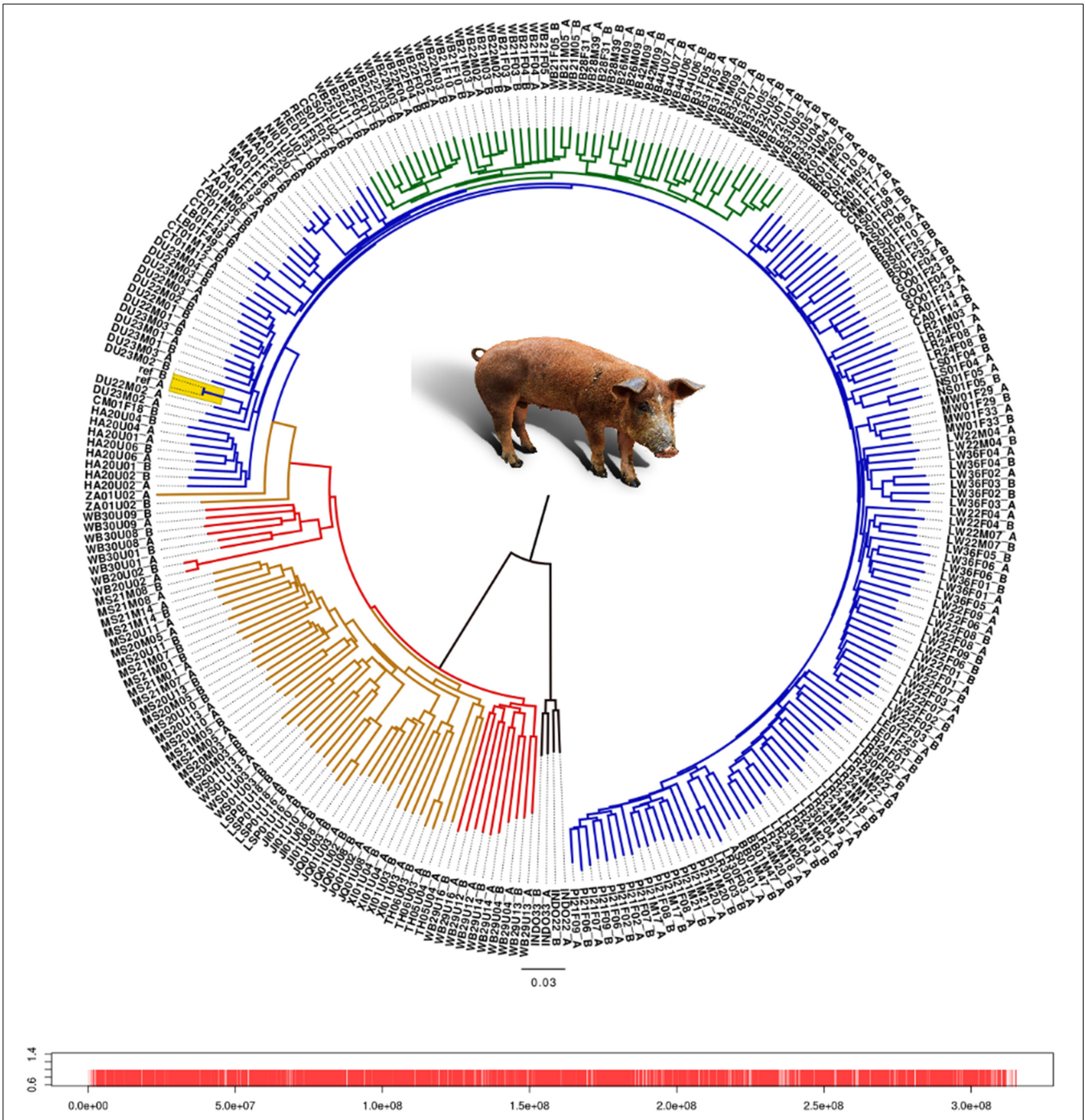


FIGURE 1 | Neighbor-joining tree of all haplotypes of chromosome 1. Each individual has 2 haplotypes, one labeled after the name of the individual with the suffix "A" and the second haplotype contains the suffix "B." Red line, Asian wild haplotype; orange line, Asian commercial or local haplotype; blue line, European commercial or local haplotype; green line, European wild haplotype. Locations of the markers on chromosome 1 are indicated by red bars. Alleles from the pig reference genome are included as two separate haplotypes without variation between them, and are highlighted in yellow.

little introgression from Europe into our sampled Asian domesticated breeds has occurred. The homozygosity of European wild haplotypes paired with Asian wild is lower than that of two Asian wild haplotypes (averages of 0.825 and 0.84, **Figure 3**), but the difference is far less pronounced than the difference in homozygosity between two European wild haplotypes and the mixture between European and Asian (0.94 vs. 0.825, **Figure 3**). This indicates that the largest source of variation comes from the Asian wild boars, and that despite the ∼1.2 My divergence between European and Asian populations, the European clade contributes marginally to the genetic diversity of the species as a whole. The finding that populations further away from the source

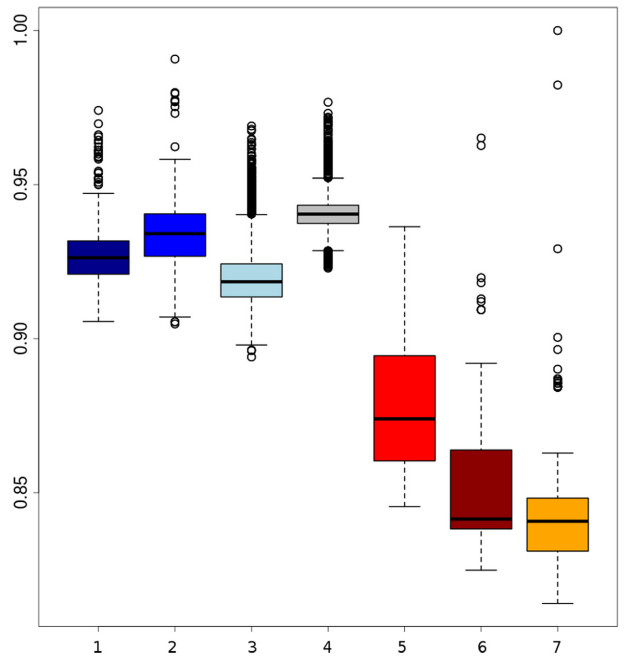


FIGURE 2 | Boxplots of homozygosity between two randomly paired haplotypes within groups. (1) Darkblue, two European local haplotypes; (2) blue, two European Iberian haplotypes; (3) lightblue, two European commercial haplotypes; (4) gray, two European wild haplotypes; (5) red, Asian commercial haplotypes; (6) brown, Asian local haplotypes; (7) orange, Asian wild haplotpyes (the highest dot indicates haplotype homozygosity within the Japanese wild boar).

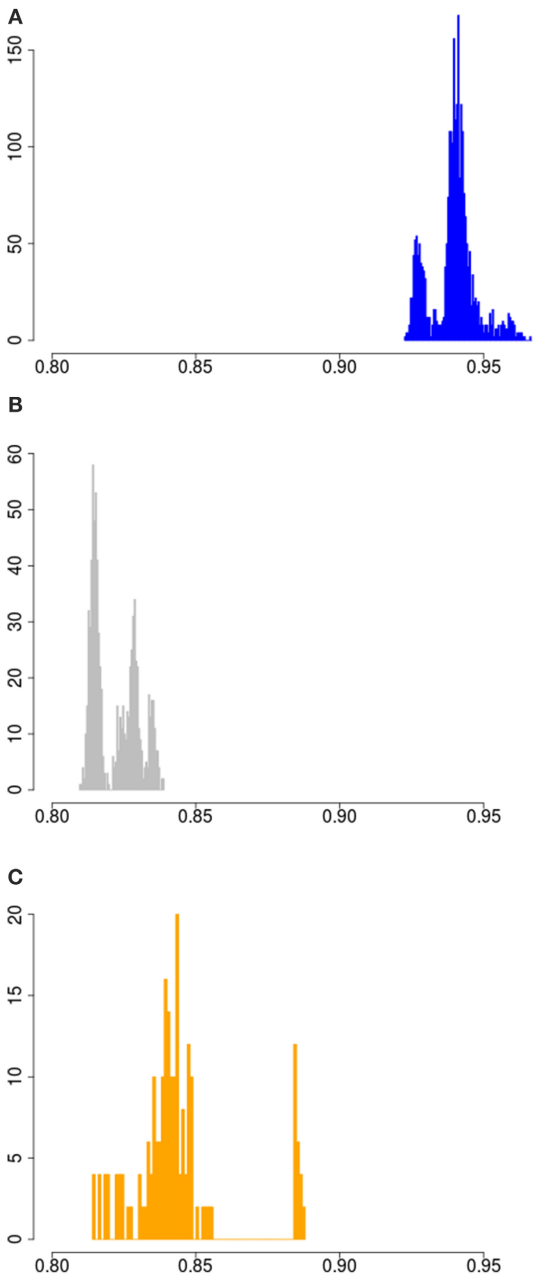


FIGURE 3 | Homozygosity between paired wild haplotypes. (A) Haplotype homozygosity between all possible pairs of European wild haplotypes. **(B)** Haplotype homozygosity between all possible pairs of 1 European wild and 1 Asian wild haplotye. **(C)** Haplotype homozygosity between all possible pairs of two Asian wild haplotypes.

population capture less genetic diversity is consistent with other species.

VARIATION BETWEEN EUROPEAN HAPLOTYPES

We had a closer look at the cause of the difference in variation within Europe. One of our hypothesis was that if the higher variation in the commercial lines is mainly caused by a mixture of different European populations, the distribution of variation between two European haplotypes should overlap with the distribution of variation between European commercial haplotypes. The European wild boars used in the current study are derived from different glaciation refugial origins and should therefore represent well extant wild boar variation throughout Europe. All possible pairs of haplotypes from European wild origin should therefore result in a distribution that exceeds the lowest haplotype homozygosity of all pairs of European commercial haplotypes, because the most divergent haplotypes from Europe are included in the European wild distribution. The far tail of the distribution of European wild haplotypes with most variation does not even overlap the mean of variation between two commercial European haplotypes (**Figure 4A**), indicating that two wild European haplotypes show more homozygosity than two random European commercial haplotypes, even if these wild haplotypes are sampled from very divergent populations. This suggests that the variation within the European commercial group cannot be completely explained by a mixture of European wild haplotypes. Therefore, it is highly unlikely that the relatively high degree of variation (compared to European wild boar) that is generally found within the European commercial breeds, is due to a mixture of European wild haplotypes, as assumed in hypothesis 1. The distributions for paired haplotypes within the European local and European Iberian group have lower means than the European wild group as well, and their extremes also exceed the European wild distribution. These findings suggest that even some local breeds may contain introgressed haplotypes.

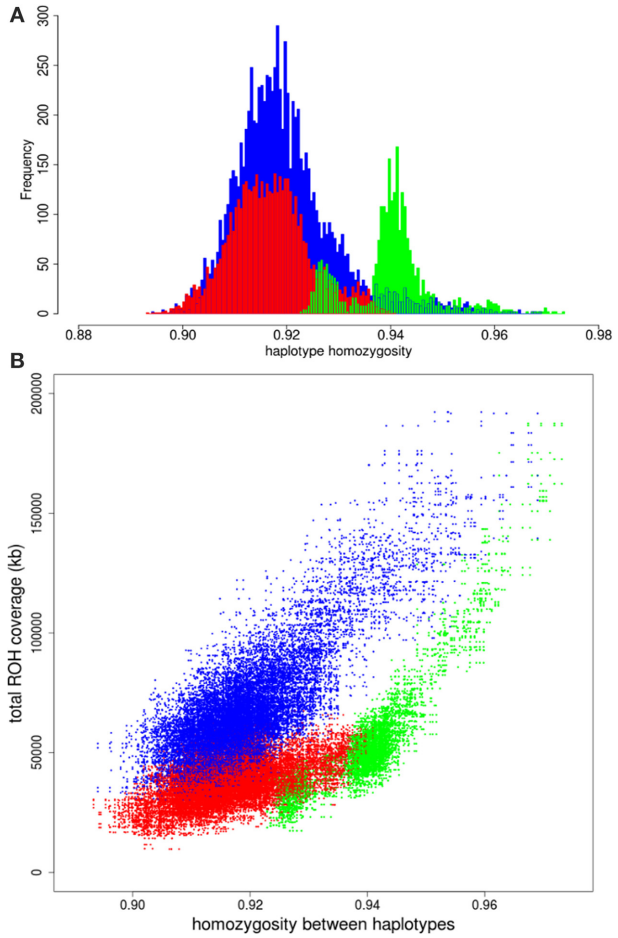


FIGURE 4 | Homozygosity between paired haplotypes in Europe. **(A)** Homozygosity between two European wild haplotypes is displayed in green. Homozygosity between two European commercial haplotypes is in red and the blue bars indicate homozygosity between one European wild and one European domesticated haplotype. **(B)** Homozygosity between haplotypes over the full chromosome on the x-axis is plotted against total ROH coverage between haplotypes on the y-axis for three combinations: two European commercial haplotypes (blue); two European wild haplotypes (green); one European wild and one European commercial haplotype (red).

Runs of homozygosity (ROH)

Another possibility of the higher variation in European commercial breeds is that European wild boar populations experienced strong recent bottlenecks and associated loss of diversity after the split with European domestic pigs (domestication). We compared the correlation between total ROH coverage on chromosome 1 (as inferred from PLINK) and homozygosity between haplotypes for the European commercial breeds and European wild boar. ROHs between two commercial European haplotypes are slightly more abundant and longer than ROHs between one European commercial and one European wild haplotype (**Figures S2A,B**). By contrast, more ROHs are found between two European wild haplotypes than between a European wild and a European commercial haplotype (**Figures S2C,D**). The average length of the ROHs between two European wild haplotypes is generally the same as between a European wild and a European commercial haplotype, unless haplotypes belong to the same European wild population (e.g., within the Netherlands). If the higher level of homozygosity between European wild haplotypes would have been caused by recent inbreeding, the coverage of ROH on chromosome 1 should be higher between two European wild haplotypes than between two European commercial haplotypes. As can be seen in **Figure 4B**, the haplotype homozygosity between two European wild haplotypes is higher than between two European commercial haplotypes with the same level of ROH coverage. These findings suggest that recent inbreeding (i.e., the occurrence of ROH) does not explain the higher homozygosity between wild haplotypes compared to commercial haplotypes.

THE EFFECT OF INTROGRESSION

Pairing with Asian haplotypes

Although the hypothesis that different source populations in Europe caused the higher diversity in commercial pigs can be rejected based on these previous analyses, our second hypothesis, that Asian introgression caused the higher diversity, is not immediately confirmed. In a previous study (Bosse et al., 2014b) we showed that within the genome of a commercial European pig, the variation is higher when at least one Asian haplotype is present. This observation however does not confirm the role of Asian introgression either, since the presence of an Asian haplotype can be due to incomplete lineage sorting or recent introgression. Another potential cause of the increased variation is hybridization with an unknown population, so called "ghost admixture." Introduced haplotypes from an unknown source are likely to increase variation in the European commercial population. Since this source should be unrelated to any of the pig groups here studied, pairing of a commercial European haplotype and an Asian haplotype should not result in less variation than an European wild haplotype paired with an Asian haplotype. If, however, the higher variation in European commercial genomes is due to Asian introgression, pairing with an Asian haplotype should result in higher homozygosity when a commercial European haplotype is used than when a wild European haplotype is used. We do find a small but significant difference between the European wild and European commercial haplotypes when they are paired with a commercial Asian haplotype (**Figure 5A**). As expected, the pairing with a European commercial haplotype results in less variation than the European wild haplotypes. Together with the lower haplotype homozygosity in the European commercial group, these findings indeed suggest that the introgression is Asian derived, or at least that the introgressed haplotypes are genetically more similar to Asian haplotypes.

Variation with chimeric haplotypes

In order to test whether the influx of Asian haplotypes caused the increase in homozygosity between the Asian wild and European commercial group, and to quantify this amount, we created composite haplotypes that contained 15, 20, and 25% of an Asian commercial breed haplotype and 85, 80 and 75% of a European wild haplotype as described in **Box 1**. These percentages were chosen because the introgression fraction from Asia into the European commercial pigs has previously been estimated to be between 15 and 35% (Fang and Andersson, 2006; Groenen et al., 2012; Bosse et al., 2014a). If the percentage of introgression is around 20%, then the distribution of haplotype

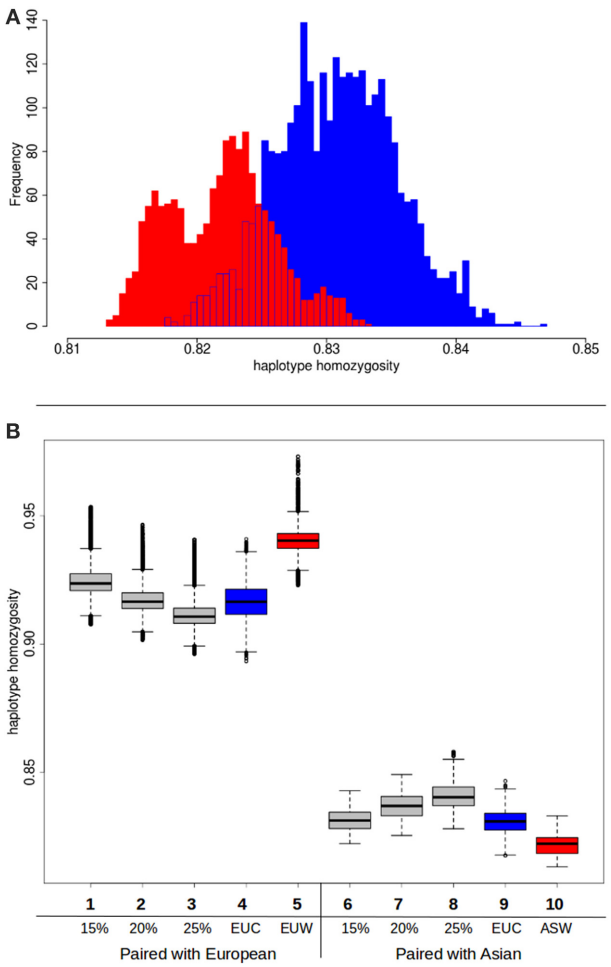


FIGURE 5 | Haplotype homozygosity with Asian introgression. **(A)** Homozygosity between haplotypes when Asian commercial haplotypes are paired with European commercial (red) or European wild (blue). **(B)** Boxplots of haplotype homozygosity. Haplotypes are paired with European wild haplotypes (left) or Asian commercial haplotypes (right). Red boxes indicate haplotypes paired with European wild haplotypes. Blue boxes represent haplotypes that are paired with European commercial haplotypes. Gray boxes represent the distribution of homozygosity when the haplotype is paired with a chimeric haplotype that is a combination of a European wild haplotype and a Asian commercial haplotype (see also **Box 1** in Supplementary material). (1) European wild paired with 15% Asian chimeric haplotype (2) European wild paired with 20% Asian chimeric haplotype (3) European wild paired with 25% Asian chimeric haplotype (4) European wild paired with European commercial (5) European wild paired with European wild (6) Asian commercial paired with 15% Asian chimeric haplotype (7) Asian commercial paired with 20% Asian chimeric haplotype (8) Asian commercial paired with 25% Asian chimeric haplotype (9) Asian commercial paired with European commercial (10) Asian commercial paired with European wild.

homozygosity when a European commercial haplotype is paired with a European wild haplotype should strongly overlap the distribution when a composite haplotype containing 20% Asian commercial alleles is paired with a European wild haplotype. On top of that, the distribution of the chimeric haplotype paired with an Asian haplotype should overlap that of a European commercial haplotype paired with a Asian haplotype. The results show (**Figure 5B**) that pairing of a chimeric haplotype of European wild and Asian commercial with a European wild haplotype indeed results in a similar distribution of homozygosity as a pair between a European wild and a European commercial haplotype. Mean haplotype homozygosity shifts from 0.941 to 0.917, suggesting 20% introgression of Asian haplotypes. Our results confirm the previous estimates of around 20% admixture and demonstrate that the Asian introgression decreased haplotype homozygosity within Europe. In addition, we show that the haplotype homozygosity when a chimeric haplotype is paired with an Asian commercial haplotype increases compared to a European wild haplotype paired with an Asian commercial haplotype. The mean of the 15% Asian chimeric haplotypes is closest to the mean of a European commercial haplotype paired with an Asian commercial haplotype (**Figure 5B**), supporting the hypothesis that the introgression indeed comes from Asia.

CONCLUSIONS

We confirmed Asia as the biggest source of genetic variation in *Sus scrofa*, in line with its geographical origin. The higher variation in the European commercial pigs compared to the European wild boar is largely explained by introgression of Asian haplotypes, rather than a mixture of European backgrounds.

ACKNOWLEDGMENTS

This work was financially supported by the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement number 249894. We thank Bert Dibbits and Kimberley Laport for labwork, and Kyle Schachtschneider for editing the manuscript. Images were produced by B. van de Water.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://www.frontiersin.org/journal/10.3389/fgene.2014.00442/abstract

Figure S1 | Haplotype homozygosity and consistency over chromosomes. **(A)** Distribution of haplotype homozygosity between all possible pairs of haplotypes on chromosome 1. The first peak around 0.725 contains all haplotypes paired with a haplotype from Sumatra. The second peak at 0.825 represents all haplotypes paired with a Chinese wild or Chinese commercial/local haplotype. The third peak round 0.92 shows all paired European haplotypes. **(B)** Consistency over chromosomes. The x-axis displays homozygosity between haplotypes from chromosome 1, and the y-axis shows the homozygosity between the same pairs of haplotypes for chromosome 18.

Figure S2 | Runs of homozygosity between paired haplotypes. ROHs on chromosome 1 are recorded between pairs of haplotypes that belong to the European wild group (green) or the European commercial group (blue). **(A)** Number of ROH and average ROH length when a haplotype is paired with a European commercial haplotype. **(B)** Number of ROH and total ROH length when a haplotype is paired with a European commercial haplotype. **(C)** Number of ROH and average ROH length when a haplotype is paired with a European wild haplotype. **(D)** Number of ROH and total ROH length when a haplotype is paired with a European wild haplotype.

REFERENCES

Alves, E., Ovilo, C., Rodriguez, M. C., and Silio, L. (2003). Mitochondrial DNA sequence variation and phylogenetic relationships among Iberian pigs and other domestic and wild pig populations. *Anim. Genet.* 34, 319–324. doi: 10.1046/j.1365-2052.2003.01010.x

Alves, P. C., Pinheiro, I., Godinho, R., Vicente, J., Gortazar, C., and Scandura, M. (2010). Genetic diversity of wild boar populations and domestic pig breeds (*Sus scrofa*) in South-western Europe. *Biol. J. Linn. Soc.* 101, 797–822. doi: 10.1111/j.1095-8312.2010.01530.x

Andersson, L. (2001). Genetic dissection of phenotypic diversity in farm animals. *Nat. Rev. Genet.* 2, 130–138. doi: 10.1038/35052563

Bosse, M., Megens, H. J., Frantz, L. A. F., Madsen, O., Larson, G., Paudel, Y., et al. (2014a). Genomic analysis reveals selection for Asian genes in European pigs following human-mediated introgression. *Nat. Commun.* 5:4392. doi: 10.1038/ncomms5392

Bosse, M., Megens, H. J., Madsen, O., Frantz, L. A. F., Paudel, Y., Crooijmans, R. P. M. A., et al. (2014b). Untangling the hybrid nature of modern pig genomes: a mosaic derived from biogeographically distinct and highly divergent Sus scrofa populations. *Mol. Ecol.* 23, 4089–4102. doi: 10.1111/mec.12807

Bosse, M., Megens, H. J., Madsen, O., Paudel, Y., Frantz, L. A., Schook, L. B., et al. (2012). Regions of homozygosity in the porcine genome: consequence of demography and the recombination landscape. *PLoS Genet.* 8:e1003100. doi: 10.1371/journal.pgen.1003100

Clop, A., Amills, M., Noguera, J. L., Fernandez, A., Capote, J., Ramon, M. M., et al. (2004). Estimating the frequency of Asian cytochrome B haplotypes in standard European and local Spanish pig breeds. *Genet. Sel. Evol.* 36, 97–104. doi: 10.1186/1297-9686-36-1-97

Dobney, K., and Larson, G. (2006). Genetics and animal domestication: new windows on an elusive process. *J. Zool.* 269, 261–271. doi: 10.1111/j.1469-7998.2006.00042.x

Doebley, J. F., Gaut, B. S., and Smith, B. D. (2006). The molecular genetics of crop domestication. *Cell* 127, 1309–1321. doi: 10.1016/j.cell.2006.12.006

Fang, M. Y., and Andersson, L. (2006). Mitochondrial diversity in European and Chinese pigs is consistent with population expansions that occurred prior to domestication. *Proc. Biol. Sci.* 273, 1803–1810. doi: 10.1098/rspb.2006.3514

Felsenstein, J. (2005). *PHYLIP (Phylogeny Inference Package) version 3.6*. Seattle, WA: Department of Genome Sciences, University of Washington.

Frantz, L. A., Schraiber, J. G., Madsen, O., Megens, H. J., Bosse, M., Paudel, Y., et al. (2013). Genomic sequencing provides fine scale inference of evolutionary history. *Genome Biol.* 14:R107. doi: 10.1186/gb-2013-14-9-r107

Giuffra, E., Kijas, J. M. H., Amarger, V., Carlborg, O., Jeon, J. T., and Andersson, L. (2000). The origin of the domestic pig: independent domestication and subsequent introgression. *Genetics* 154, 1785–1791.

Groenen, M. A., Archibald, A. L., Uenishi, H., Tuggle, C. K., Takeuchi, Y., Rothschild, M. F., et al. (2012). Analyses of pig genomes provide insight into porcine demography and evolution. *Nature* 491, 393–398. doi: 10.1038/nature11622

Groeneveld, L. F., Lenstra, J. A., Eding, H., Toro, M. A., Scherf, B., Pilling, D., et al. (2010). Genetic diversity in farm animals - a review. *Anim. Genet.* 41, 6–31. doi: 10.1111/j.1365-2052.2010.02038.x

Herrero-Medrano, J. M., Megens, H. J., Groenen, M. A. M., Ramis, G., Bosse, M., Perez-Enciso, M., et al. (2013). Conservation genomic analysis of domestic and wild pig populations from the Iberian Peninsula. *BMC Genet.* 14:106. doi: 10.1186/1471-2156-14-106

Hidalgo, A. M., Bastiaansen, J. W. M., Harlizius, B., Megens, H. J., Madsen, O., Crooijmans, R. P. M. A., et al. (2014). On the relationship between an Asian haplotype on chromosome 6 that reduces androstenone levels in boars and the differential expression of SULT2A1 in the testis. *BMC Genet.* 15:4. doi: 10.1186/1471-2156-15-4

Kijas, J. M. H., and Andersson, L. (2001). A phylogenetic study of the origin of the domestic pig estimated from the near-complete mtDNA genome. *J. Mol. Evol.* 52, 302–308. doi: 10.1007/s002390010158

Kristensen, T. N., and Sorensen, A. C. (2005). Inbreeding - lessons from animal breeding, evolutionary biology and conservation genetics. *Anim. Sci.* 80, 121–133. doi: 10.1079/ASC41960121

Larson, G., Albarella, U., Dobney, K., Rowley-Conwy, P., Schibler, J., Tresset, A., et al. (2007). Ancient DNA, pig domestication, and the spread of the Neolithic into Europe. *Proc. Natl. Acad. Sci. U.S.A.* 104, 15276–15281. doi: 10.1073/pnas.0703411104

Larson, G., Dobney, K., Albarella, U., Fang, M., Matisoo-Smith, E., Robins, J., et al. (2005). Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. *Science* 307, 1618–1621. doi: 10.1126/science.1106927

Long, J. C., and Kittles, R. A. (2003). Human genetic diversity and the nonexistence of biological races. *Hum. Biol.* 75, 449–471. doi: 10.1353/hub.2003.0058

Megens, H. J., Crooijmans, R. P. M. A., Cristobal, M. S., Hui, X., Li, N., and Groenen, M. A. M. (2008). Biodiversity of pig breeds from China and Europe estimated from pooled DNA samples: differences in microsatellite variation between two areas of domestication. *Genet. Sel. Evol.* 40, 103–128. doi: 10.1051/gse:2007039

Ojeda, A., Huang, L. S., Ren, J., Angiolillo, A., Cho, I. C., Soto, H., et al. (2008). Selection in the making: a worldwide survey of haplotypic diversity around a causative mutation in porcine IGF2. *Genetics* 178, 1639–1652. doi: 10.1534/genetics.107.084269

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795

Ramirez, O., Ojeda, A., Tomas, A., Gallardo, D., Huang, L. S., Folch, J. M., et al. (2009). Integrating Y-Chromosome, mitochondrial, and autosomal data to analyze the origin of pig breeds. *Mol. Biol. Evol.* 26, 2061–2072. doi: 10.1093/molbev/msp118

Ramírez, O., Burgos-Paz, W., Casas, E., Ballester, M., Bianco, E., Olalde, I., et al. (2014). Genome data from a sixteenth century pig illuminate modern breed relationships. *Heredity* 131, 46–52. doi: 10.1038/hdy.2014.81

Rubin, C. J., Megens, H. J., Barrio, A. M., Maqbool, K., Sayyab, S., Schwochow, D., et al. (2012). Strong signatures of selection in the domestic pig genome. *Proc. Natl. Acad. Sci. U.S.A.* 109, 19529–19536. doi: 10.1073/pnas.1217149109

Scandura, M., Iacolina, L., Crestanello, B., Pecchioli, E., Di Benedetto, M. F., Russo, V., et al. (2008). Ancient vs. recent processes as factors shaping the genetic variation of the European wild boar: are the effects of the last glaciation still detectable? *Mol. Ecol.* 17, 1745–1762. doi: 10.1111/j.1365-294X.2008.03703.x

Taberlet, P., Valentini, A., Rezaei, H. R., Naderi, S., Pompanon, F., Negrini, R., et al. (2008). Are cattle, sheep, and goats endangered species? *Mol. Ecol.* 17, 275–284. doi: 10.1111/j.1365-294X.2007.03475.x

vonHoldt, B. M., Pollinger, J. P., Lohmueller, K. E., Han, E. J., Parker, H. G., Quignon, P., et al. (2010). Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication. *Nature* 464, 898–902. doi: 10.1038/nature08837

Wilkinson, S., Lu, Z. H., Megens, H. J., Archibald, A. L., Haley, C., Jackson, I. J., et al. (2013). Signatures of diversifying selection in European pig breeds. *PLoS Genet.* 9:e1003453. doi: 10.1371/journal.pgen.1003453

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 10 October 2014; paper pending published: 07 November 2014; accepted: 03 December 2014; published online: 05 January 2015.

*Citation: Bosse M, Madsen O, Megens H-J, Frantz LAF, Paudel Y, Crooijmans RPMA and Groenen MAM (2015) Hybrid origin of European commercial pigs examined by an in-depth haplotype analysis on chromosome 1. Front. Genet. **5**:442. doi: 10.3389/fgene.2014.00442*

This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics.

Copyright © 2015 Bosse, Madsen, Megens, Frantz, Paudel, Crooijmans and Groenen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

TECHNOLOGY REPORT
published: 20 March 2015
doi: 10.3389/fgene.2015.00109

SNeP: a tool to estimate trends in recent effective population size trajectories using genome-wide SNP data

Mario Barbato[1]*, *Pablo Orozco-terWengel*[1], *Miika Tapio*[2] and *Michael W. Bruford*[1]

[1] School of Biosciences, Cardiff University, Cardiff, UK, [2] MTT Agrifood Research Finland, Biotechnology and Food Research, Jokioinen, Finland

Effective population size (N_e) is a key population genetic parameter that describes the amount of genetic drift in a population. Estimating N_e has been subject to much research over the last 80 years. Methods to estimate N_e from linkage disequilibrium (LD) were developed ~40 years ago but depend on the availability of large amounts of genetic marker data that only the most recent advances in DNA technology have made available. Here we introduce *SNeP,* a multithreaded tool to perform the estimate of N_e using LD using the standard PLINK input file format (.ped and.map files) or by using LD values calculated using other software. Through *SNeP* the user can apply several corrections to take account of sample size, mutation, phasing, and recombination rate. Each variable involved in the computation such as the binning parameters or the chromosomes to include in the analysis can be modified. When applied to published datasets, *SNeP* produced results closely comparable with those obtained in the original studies. The use of *SNeP* to estimate N_e trends can improve understanding of population demography in the recent past, provided a sufficient number of SNPs and their physical position in the genome are available. Binaries for the most common operating systems are available at https://sourceforge.net/projects/snepnetrends/.

Keywords: effective population size, linkage disequilibrium, SNPChip, demography, large scale genotyping

OPEN ACCESS

Edited by:
Paolo Ajmone Marsan, Università Cattolica del Sacro Cuore, Italy

Reviewed by:
David MacHugh, University College Dublin, Ireland
Yuri Tani Utsunomiya, Universidade Estadual Paulista(UNESP), Brazil

***Correspondence:**
Mario Barbato, School of Biosciences, Cardiff University, Sir Martin Evans Building, Museum Avenue CF10 3AX Cardiff, UK
barbatom@cardiff.ac.uk

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: *15 November 2014*
Accepted: *03 March 2015*
Published: *20 March 2015*

Citation:
Barbato M, Orozco-terWengel P, Tapio M and Bruford MW (2015) SNeP: a tool to estimate trends in recent effective population size trajectories using genome-wide SNP data. Front. Genet. 6:109. doi: 10.3389/fgene.2015.00109

Introduction

Effective population size (N_e) is an important genetic parameter that estimates the amount of genetic drift in a population, and has been described as the size of an idealized Wright–Fisher population expected to yield the same value of a given genetic parameter as in the population under study (Crow and Kimura, 1970). N_e sizes can be influenced by fluctuations in census population size (N_c), by the breeding sex ratio and the variance in reproductive success.

N_e estimation can be achieved using approaches that fall into three methodological categories: demographic, pedigree-based, or marker-based (Flury et al., 2010). Pedigree data have been traditionally used to obtain N_e estimates in livestock. However, reliable estimates of N_e depend on the pedigree being complete. This state of knowledge is feasible in some domestic populations, the demographic parameters of which have been accurately monitored for a sufficiently large number of generations. However, in practice, the applicability of this approach remains limited to a few cases involving highly managed breeds (Flury et al., 2010; Uimari and Tapio, 2011).

One solution to overcome the limitation of an incomplete pedigree is to estimate the recent trend in N_e using genomic data. Several authors have recognized that N_e could be estimated from information on linkage disequilibrium (LD) (Sved, 1971; Hill, 1981). LD describes the non-random association of alleles in different loci as a function of the recombination rate between the physical positions of the loci in the genome. However, LD signatures can also result from demographic processes such as admixture and genetic drift (Wright, 1943; Wang, 2005), or through processes such as "hitchhiking" during selective sweeps (Smith and Haigh, 1974) or background selection (Charlesworth et al., 1997). In such scenarios alleles at different loci become associated independently of their proximity in the genome. Assuming that a population is closed and panmictic, the LD value calculated between neutral unlinked loci depends exclusively on genetic drift (Sved, 1971; Hill, 1981). This occurrence can be used to predict N_e due to the known relationship between the variance in LD (calculated using allele frequencies) and effective population size (Hill, 1981).

Recent advances in genotyping technology (e.g., using SNP bead arrays with tens of thousands of DNA probes) have enabled the collection of vast amounts of genome-wide linkage data ideal for estimating N_e in livestock and humans among others (e.g., Tenesa et al., 2007; de Roos et al., 2008; Corbin et al., 2010; Uimari and Tapio, 2011; Kijas et al., 2012). However, a software tool that enables estimation of N_e from LD is lacking, and researchers currently rely on a combination of tools to manipulate data, infer LD, and tend to use bespoke scripts to perform the appropriate calculations and estimate N_e.

Here we describe *SNeP*, a software tool that allows the estimation of N_e trends across generation using SNP data that corrects for sample size, phasing and recombination rate.

Materials and Methods

The method *SNeP* uses to calculate LD depends on the availability of phased data. When the phase is known the user can select Hill and Robertson (1968) squared correlation coefficient that makes use of haplotype frequencies to define LD between each pair of loci (Equation 1). However, in the absence of a known phase, squared Pearson's product-moment correlation coefficient between pairs of loci can be selected. While these two approaches are not the same, they are highly comparable (McEvoy et al., 2011):

$$r^2 = \frac{\left(p_{AB}-p_A p_B\right)^2}{p_A\left(1-p_A\right)p_B\left(1-p_B\right)} \quad (1)$$

$$r^2_{X,Y} = \frac{\left[\sum_{i=1}^{n}\left(X_i-\overline{X}\right)\left(Y_i-\overline{Y}\right)\right]^2}{\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2\sum_{i=1}^{n}\left(Y_i-\overline{Y}\right)^2} \quad (2)$$

where p_A and p_B are respectively the frequencies of alleles A and B at two separate loci (X, Y) measured for n individuals, p_{AB} is the frequency of the haplotype with alleles A and B in the population studied, $\overline{X}$ and $\overline{Y}$ are the mean genotype frequencies for the first and second locus respectively, X_i is the genotype of individual i at the first locus and Y_i is the genotype of individual i at the second locus. Equation (2) correlates the genotypic allele counts instead of the haplotype frequencies and is not influenced by double heterozygotes (this approach results in the same estimates as the --r2 option in PLINK).

SNeP estimates the historic effective population size based on the relationship between r^2, N_e, and c (recombination rate), (Equation 3—Sved, 1971), and enabling users to include corrections for sample size and uncertainty of the gametic phase (Equation 4—Weir and Hill, 1980):

$$E\left(r^2\right) = \left(1+4N_e c\right)^{-1} \quad (3)$$

$$r^2_{adj} = r^2 - (\beta n)^{-1} \quad (4)$$

where n is the number of individual sampled, $\beta = 2$ when the gametic phase is known and $\beta = 1$ if instead the phase is not known.

Several approximations are used to infer the recombination rate using the physical distance (δ) between two loci as a reference and translating it into linkage distance (d), which is usually described as $Mb(\delta) \approx cM(d)$. For small values of d the latter approximation is valid, but for larger values of d the probability of multiple recombination events and interference increases, moreover the relationship between map distance and recombination rate is not linear, as the maximum recombination rate possible is 0.5. Thus, unless using very short δ, the approximation $d \approx c$ is not ideal (Corbin et al., 2012). We therefore implemented mapping functions to translate the estimated d into c, following Haldane (1919), Kosambi (1943), Sved (1971), and Sved and Feldman (1973). Initially *SNeP* infers d for each pair of SNPs as directly proportional to δ according to $d = k\delta$ where k is a user defined recombination rate value (default value is 10^{-8} as in $Mb = cM$). The inferred value of δ can then be subjected to one of the available mapping functions if required by the user.

Solving Equation (3) for N_e and including all the corrections described, allows the prediction of N_e from LD data using (Corbin et al., 2012):

$$N_{T(t)} = \left(4f\left(c_t\right)\right)^{-1}\left(E\left[r^2_{adj}|c_t\right]^{-1} - \alpha\right) \quad (5)$$

where N_t is the effective population size t generations ago calculated as $t = \left(2f(c_t)\right)^{-1}$ (Hayes et al., 2003), c_t is the recombination rate defined for a specific physical distance between markers and optionally adjusted with the mapping functions mentioned above, r^2_{adj} is the LD value adjusted for sample size and $\alpha := \{1, 2, 2.2\}$ is a correction for the occurrence of mutations (Ohta and Kimura, 1971). Therefore, LD over greater recombinant distances is informative on recent N_e while shorter distances provide information on more distant times in the past. A binning system is implemented in order to obtain averaged r^2 values that reflect LD for specific inter-locus distances. The binning system implemented uses the following formula to define the minimum and maximum values for each bin:

$$b_i^{min} = minD + (maxD - minD)\left(\frac{b_i - 1}{totBins}\right)^x \quad (6a)$$

$$b_i^{max} = minD + (maxD - minD)\left(\frac{b_i}{totBins}\right)^x \quad (6b)$$

Where b_i ($\mathbb{N}^1$) is the i^{th} bin of the total number of bins (*totBins*), *minD*, and *maxD* are respectively the minimum and the maximum distance between SNPs and x is a positive real number ($\mathbb{R}^0$) When x equals 1, the distribution of distances between the bins is linear and each bin has the same distance range. For larger values of x the distribution of distances changes allowing a larger range on the last bins and a smaller range on the first bins. Varying this parameter allows the user to have a sufficient number of pairwise comparisons to contribute to the final N_e estimate for each bin.

Example Application

We tested *SNeP* with two published datasets that had been previously used to describe trends in N_e over time using LD, *Bos indicus* [54,436 SNPs of 423 East African Shorthorn Zebu (SHZ)–Mbole-Kariuki et al., 2014, data available at Dryad Digital Repository: doi:10.5061/dryad.bc598.] and *Ovis aries* [49,034 SNPs genotyped in 24 Swiss White Alpine (SWA), 24 Swiss Black-Brown Mountain sheep (SBS), 24 Valais Blacknose sheep (VBS), 23 Valais Red sheep (VRS), 24 Swiss Mirror sheep (SMS) and 24 Bundner Oberländer sheep (BOS)–Burren et al., 2014]. The r^2 estimates for the cattle datasets were obtained by the authors using GenABLE (Aulchenko et al., 2007) using a minimum allele frequency (MAF) < 0.01 and adjusting the recombination rate using Haldane's mapping function (Haldane, 1919). The r^2 estimates of the sheep data were calculated by the authors using PLINK-1.07 (Purcell et al., 2007), with a MAF < 0.05 and no further corrections. For both autosomal datasets r^2 estimates where corrected for sample size using equation (4) with $\beta = 2$. For these comparative analyses the *SNeP* command line included the same parameters used for the published data apart from the r^2 estimates, calculated through genotype count and the use of *SNeP*'s novel binning strategy.

Results

SNeP is a multithreaded application developed in C++ and binaries for the most common operating systems (Windows, OSX, and Linux) can be downloaded from https://sourceforge.net/projects/snepnetrends/. The binaries are accompanied by a manual describing the step-by-step use of *SNeP* to infer trends in N_e as described here. *SNeP* produces an output file with tab delimited columns showing the following for each bin that was used to estimate N_e: the number of generations in the past that the bin corresponds to (e.g., 50 generations ago), the corresponding N_e estimate, the average distance between each pair of SNPs in the bin, the average r^2 and the standard deviation of r^2 in the bin, and the number of SNPs used to calculate r^2 in the bin. This file can be easily imported in Microsoft Excel, R or other software to plot the results. The plots shown here (**Figures 1**, **3**) correspond to the columns of generations ago and N_e from the output file. The column with the r^2 standard deviation is provided for users to inspect the variance in the N_e estimate in each bin, particularly for those bins reflecting older time estimates and which are less reliable as the number of SNPs used to estimate r^2 becomes smaller.

The format required for the input files is the standard PLINK format (ped and map files) (Purcell et al., 2007). *SNeP* allows the users to either calculate LD on the data as described above, or use a custom precalculated LD matrix to estimate N_e using Equation (5).

The software interface allows the user to control all parameters of the analysis, e.g., the distance range between SNPs in bp, and the set of chromosomes used in the analysis (e.g., 20–23). Additionally, *SNeP* includes the option to choose a MAF threshold (default 0.05), as it has been shown that accounting for MAF results in unbiased r^2 estimates irrespective of sample size

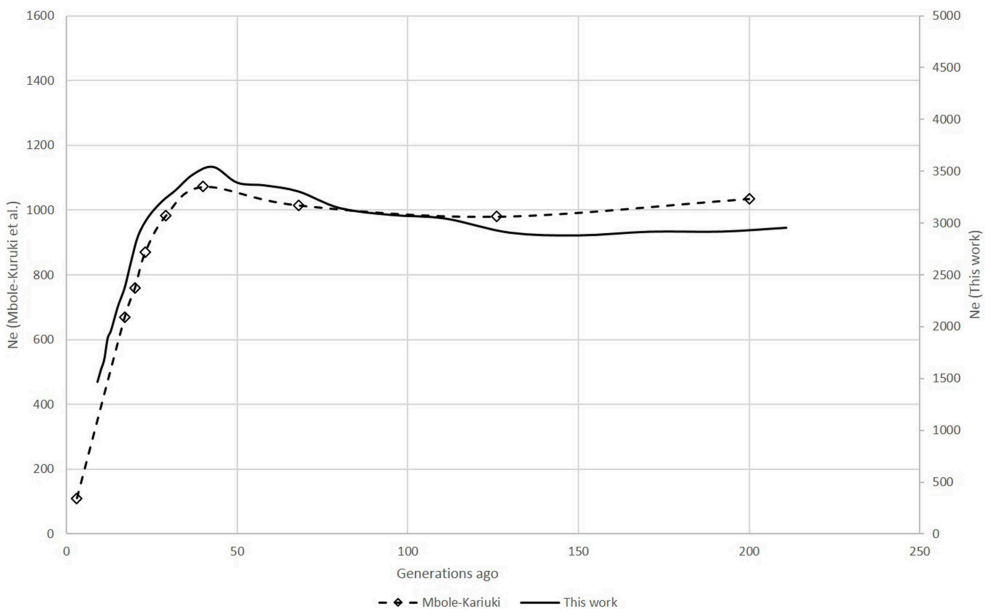


FIGURE 1 | Comparison of N_e trends of six Swiss sheep breeds according to Burren et al. (2014) (dashed lines) and this work (solid lines).

(Sved et al., 2008). *SNeP*'s multithreaded architecture allows fast computation of large datasets (we tested up to ∼100K SNPs for a single chromosome), for example the BOS data described here was analyzed with one processor in 2′43″, the use of two processors reduced the time to 1′43″, four processors reduced the analysis time to 1′05″.

Zebu Example

For the zebu analysis, the shapes of the N_e curves obtained with *SNeP* and their published data trends showed the same trajectory with a smooth decline until around 150 generations ago, followed by an expansion with a peak around 40 generations ago and ending in a steep decline on the most recent generations (**Figure 1**). However, while the trends in both curves were the same, the two approaches resulted in different N_e estimates, with *SNeP*'s values being approximately three-fold larger than those in the original paper. While we attempted to use the authors' parameters in our analyses, some differences were inevitable, i.e., the original publication of the cattle data estimated r^2 with a different approach to that implemented in *SNeP*. Analyses with *SNeP* were based on genotypes, while the original analysis was based on inferred two locus haplotypes, which results in the published data showing an expected r^2 of 0.32 at the minimum distance, while our estimates was 0.23. Similarly, Mbole-Kariuki et al. (2014) obtained a background level $r^2 = 0.013$ around 2 Mb, while our estimate at the same distance was 0.0035 (data not shown). Consequently, as our estimates of LD were consistently smaller than Mbole-Kariuki et al. (2014) it is expected that our N_e estimates should be larger. While this observation highlights the importance of a careful choice of the parameters and their thresholds, it is important to highlight that although the absolute magnitude of the N_e values is different, the trends are almost identical.

Swiss Sheep Example

The six Swiss sheep breeds analyzed with *SNeP* produced comparable results with those from the original paper (**Figure 2**), with mostly overlapping N_e trend curves (**Figure 3**). However, the general trend in N_e showed a decline toward the present. *SNeP* produced slightly larger values of N_e for the more distant past (700–800 generations). This is due to the different binning system used in *SNeP*, which allows the user to obtain a more even distribution of pairwise comparisons within each bin (i.e., the number of SNP pairwise comparisons within each bin is comparable). For the time span extending beyond 400 generations ago, Burren et al. (2014) used only three bins in their analysis (centered at 400, 667, and 2000 generations ago) while for the same time span *SNeP* used 5 bins with a number of pairwise comparisons dependent to the range defined with formulae 6a,b. Consequently, Burren and colleagues' approach ends with a higher density of data describing the most recent generations than describing the oldest generations. Therefore, the use of fewer bins tends to increase the presence of smaller values of N_e in each bin, consequently lowering the average N_e value for each bin. The N_e values for the recent past, compared at the 29th generation in the past, gave very similar results. The largest difference (50) was obtained for the SBS breed.

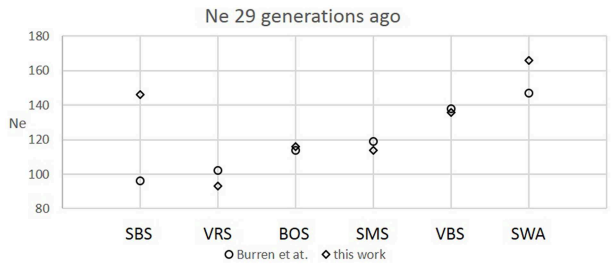


FIGURE 2 | Comparison between recent N_e values calculated at the 29th generation in this work and Burren et al. (2014) for six Swiss sheep breeds.

Discussion

Analysis of N_e using LD data was first demonstrated 40 years ago, and has been applied, developed and improved since (Sved, 1971; Hayes et al., 2003; Tenesa et al., 2007; de Roos et al., 2008; Corbin et al., 2012; Sved et al., 2013). The traditionally small number of SNPs analyzed is no longer a limitation, since SNP Chips comprise an extremely large number of SNPs, available in a short time and at a reasonable price. This has boosted the use of the method, which has been applied to humans (Tenesa et al., 2007; McEvoy et al., 2011) as well as to several domesticated species (England et al., 2006; Uimari and Tapio, 2011; Corbin et al., 2012; Kijas et al., 2012). Along with these improvements, methodological limitations have become apparent and have been addressed here, with the majority of the efforts pointing to the correct estimation of recent N_e. Yet, the quantitative value of the estimate is highly dependent on sample size, the type of LD estimation and the binning process (Waples and Do, 2008; Corbin et al., 2012), while its qualitative pattern depends more on the genetic information than on data manipulation.

So far this method has been applied using a variety of software, no standardized approach exists to bin the results and each study has applied a more or less arbitrary approach, e.g., binning for generation classes in the past (Corbin et al., 2012), binning for distance classes with a constant range for each bin (Kijas et al., 2012) or binning per distance classes in a linear fashion but with larger bins for the more recent time points (Burren et al., 2014). To our knowledge the only software available that estimate N_e through LD is NeEstimator (Do et al., 2014), an upgraded version of the former LDNE (Waples and Do, 2008) allowing the analysis of large dataset (as 50k SNPChip). Importantly, while *SNeP* focuses on estimating historical N_e trends, NeEstimator's aim is to produce contemporary unbiased N_e estimates, the latter should therefore be considered as a complementary tool while investigating demography through LD.

We used *SNeP* to analyze two datasets where the method was previously applied. The results we obtained for the sheep data were both quantitatively and qualitatively comparable with those obtained by Burren et al. (2014), while for the Zebu data we obtained a N_e trend estimate that closely matched that of

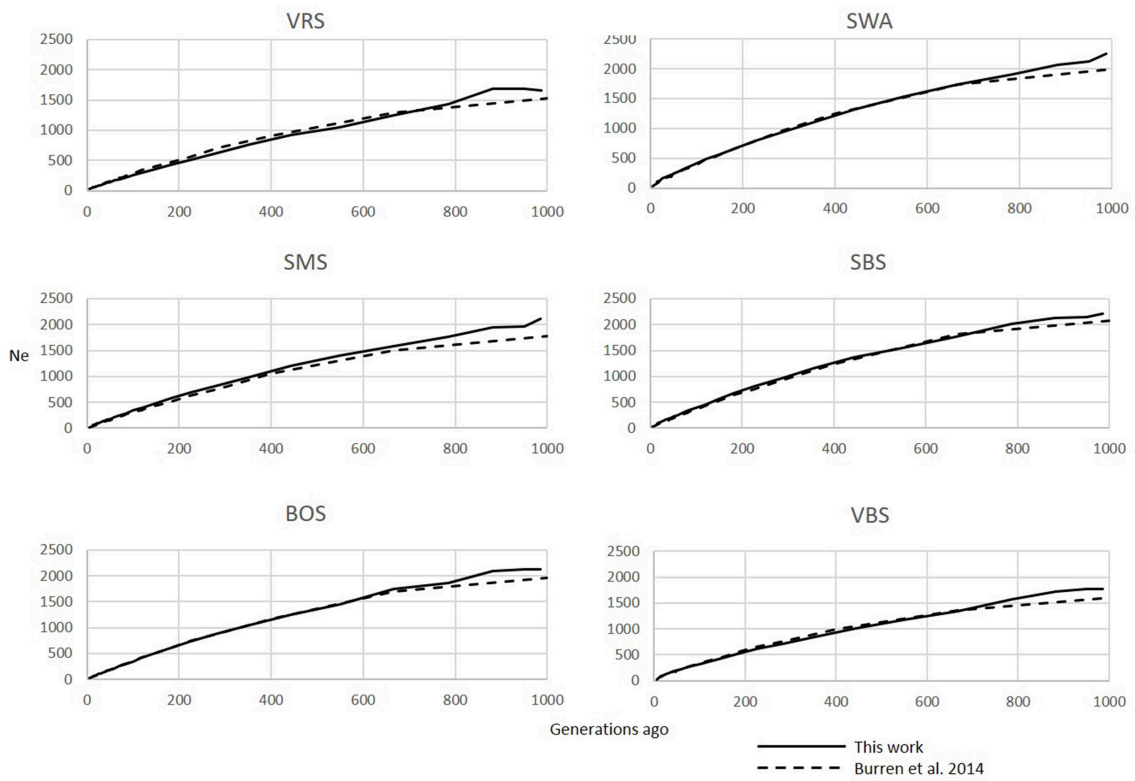


FIGURE 3 | Comparison of N_e trends for the last 250 generations in the SHZ data obtained by Mbole-Kariuki et al. (2014) (dashed line) and using *SNeP* (solid line).

Mbole-Kariuki et al. (2014) although our point estimates of N_e were larger than those described for the data (Mbole-Kariuki et al., 2014). The discrepancy between these two results reflects that Burren and colleagues produced their r^2 estimates using PLINK (the standard software for large scale SNP data manipulation) which uses the same approach used to estimate r^2 by *SNeP*, while Mbole-Kariuki et al. followed Hao et al. (2007) for r^2 estimation. The use of different estimates for LD is critical for the quantitative aspect of the N_e curve, where due to the hyperbolic correlation between N_e and r^2, a decrease in r^2 on its range closer to 0 can lead to a very large change in N_e estimates, while differences in estimates are less significant when the r^2 value is high, i.e., closer to 1. Therefore, although in one of the datasets the N_e values where substantially different, in both cases the N_e curves overlapped with those originally published.

As already suggested by other authors, the reliability of the quantitative estimates obtained with this method must be taken with caution, especially for N_e values related to the most recent and the oldest generations (Corbin et al., 2012) because for recent generations, large values of c are involved, not fitting the theoretical implications that Hayes proposed to estimate a variable N_e over time (Hayes et al., 2003). Estimates for the oldest generations might also be unreliable as coalescent theory shows that no SNP can be reliably sampled after $4N_e$ generations in the past (Corbin et al., 2012). Further, N_e estimates, and especially those related to generations further in the past, are strongly affected by data manipulation factors, such as the choice of MAF and alpha values. Additionally, the binning strategy applied can interfere with the general precision of the method, for example where an insufficient number of pairwise comparisons are used to populate each bin.

One of the applications of method is to compare breed demographies. In this case the shape of the N_e curves would be the optimal tool to differentiate different demographic histories, more than their numerical values, by using them as a potential demographic fingerprint for that breed or species, yet taking into consideration that mutation, migration, and selection can influence the N_e estimation through LD (Waples and Do, 2010). Additionally, careful consideration of the data analyzed with *SNeP* (and other software to estimate N_e) is very important, as the presence of confounding factors such as admixture, may result in biased estimates of N_e (Orozco-terWengel and Bruford, 2014).

The aim of *SNeP* is therefore to provide a fast and reliable tool to apply LD methods to estimate N_e using high throughput genotypic data in a more consistent way. It allows two different r^2 estimation approaches plus the option of using r^2 estimates from external software. The use of *SNeP* does not overcome the limits of the method and the theory behind it, yet it allows the user to apply the theory using all corrections suggested to date.

Author Contributions

MB conceived and wrote the software and the manuscript. MB, MT, and POtW tested the software and performed the analyses. MT, POtW, and MWB revised the manuscript. All authors approved the final manuscript.

Acknowledgments

We thank Christine Flury for providing the sheep data and for useful discussion. We also thank the two reviewers for useful suggestions to improve this paper. MB was supported by the program Master and Back (Regione Sardegna).

References

Aulchenko, Y. S., Ripke, S., Isaacs, A., and van Duijn, C. M. (2007). GenABEL: an R library for genome-wide association analysis. *Bioinformatics* 23, 1294–1296. doi: 10.1093/bioinformatics/btm108

Burren, A., Signer-Hasler, H., Neuditschko, M., Tetens, J., Kijas, J. W., Drögemüller, C., et al. (2014). Fine-scale population structure analysis of seven local Swiss sheep breeds using genome-wide SNP data. *Anim. Genet. Resour.* 55, 67–76. doi: 10.1017/S2078633614000253

Charlesworth, B., Nordborg, M., and Charlesworth, D. (1997). The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genet. Res.* 70, 155–174. doi: 10.1017/S0016672397002954

Corbin, L. J., Blott, S. C., Swinburne, J. E., Vaudin, M., Bishop, S. C., and Woolliams, J. A. (2010). Linkage disequilibrium and historical effective population size in the Thoroughbred horse. *Anim. Genet.* 41(Suppl. 2), 8–15. doi: 10.1111/j.1365-2052.2010.02092.x

Corbin, L. J., Liu, A. Y. H., Bishop, S. C., and Woolliams, J. A. (2012). Estimation of historical effective population size using linkage disequilibria with marker data. *J. Anim. Breed. Genet.* 129, 257–270. doi: 10.1111/j.1439-0388.2012.01003.x

Crow, J. F., and Kimura, M. (1970). *An Introduction to Population Genetics Theory*. New York, NY: Harper and Row.

de Roos, A. P. W., Hayes, B. J., Spelman, R. J., and Goddard, M. E. (2008). Linkage disequilibrium and persistence of phase in Holstein-Friesian, Jersey and Angus cattle. *Genetics* 179, 1503–1512. doi: 10.1534/genetics.107.084301

Do, C., Waples, R. S., Peel, D., Macbeth, G. M., Tillett, B. J., and Ovenden, J. R. (2014). NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. *Mol. Ecol. Resour.* 14, 209–214. doi: 10.1111/1755-0998.12157

England, P. R., Cornuet, J.-M., Berthier, P., Tallmon, D. A., and Luikart, G. (2006). Estimating effective population size from linkage disequilibrium: severe bias in small samples. *Conserv. Genet.* 7, 303–308. doi: 10.1007/s10592-005-9103-8

Flury, C., Tapio, M., Sonstegard, T., Drögemüller, C., Leeb, T., Simianer, H., et al. (2010). Effective population size of an indigenous Swiss cattle breed estimated from linkage disequilibrium. *J. Anim. Breed. Genet.* 127, 339–347. doi: 10.1111/j.1439-0388.2010.00862.x

Haldane, J. B. S. (1919). The combination of linkage values, and the calculation of distances between the loci of linked factors. *J. Genet.* 8, 299–309.

Hao, K., Di, X., and Cawley, S. (2007). LdCompare: rapid computation of single- and multiple-marker r2 and genetic coverage. *Bioinformatics* 23, 252–254. doi: 10.1093/bioinformatics/btl574

Hayes, B. J., Visscher, P. M., McPartlan, H. C., and Goddard, M. E. (2003). Novel multilocus measure of linkage disequilibrium to estimate past effective population size. *Genome Res.* 13, 635–643. doi: 10.1101/gr.387103

Hill, W. G. (1981). Estimation of effective population size from data on linkage disequilibrium. *Genet. Res.* 38, 209–216. doi: 10.1017/S0016672300020553

Hill, W. G., and Robertson, A. (1968). Linkage Disequilibrium in Finite Populations. *Theor. Appl. Genet.* 38, 226–231. doi: 10.1007/BF01245622

Kijas, J. W., Lenstra, J. A., Hayes, B. J., Boitard, S., Porto Neto, L. R., San Cristobal, M., et al. (2012). Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. *PLoS Biol.* 10:e1001258. doi: 10.1371/journal.pbio.1001258

Kosambi, D. D. (1943). The estimation of map distances from recombination values. *Ann. Eugen.* 12, 172–175. doi: 10.1111/j.1469-1809.1943.tb02321.x

Mbole-Kariuki, M. N., Sonstegard, T., Orth, A., Thumbi, S. M., Bronsvoort, B. M. D. C., Kiara, H., et al. (2014). Genome-wide analysis reveals the ancient and recent admixture history of East African Shorthorn Zebu from Western Kenya. *Heredity* 113, 297–305. doi: 10.1038/hdy.2014.31

McEvoy, B. P., Powell, J. E., Goddard, M. E., and Visscher, P. M. (2011). Human population dispersal "Out of Africa" estimated from linkage disequilibrium and allele frequencies of SNPs. *Genome Res.* 21, 821–829. doi: 10.1101/gr.119636.110

Ohta, T., and Kimura, M. (1971). Linkage disequilibrium between two segregating nucleotide sites under the steady flux of mutations in a finite population. *Genetics* 68, 571–580.

Orozco-terWengel, P., and Bruford, M. W. (2014). Mixed signals from hybrid genomes. *Mol. Ecol.* 23, 3941–3943. doi: 10.1111/mec.12863

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795

Smith, J. M., and Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genet. Res.* 23, 23–35. doi: 10.1017/S0016672300014634

Sved, J. A. (1971). Linkage disequilibrium and homozygosity of chromosome segments in finite populations. *Theor. Popul. Biol.* 141, 125–141. doi: 10.1016/0040-5809(71)90011-6

Sved, J. A., Cameron, E. C., and Gilchrist, A. S. (2013). Estimating effective population size from linkage disequilibrium between unlinked loci: theory and application to fruit fly outbreak populations. *PLoS ONE* 8:e69078. doi: 10.1371/journal.pone.0069078

Sved, J. A., and Feldman, M. W. (1973). Correlation and probability methods for one and two loci. *Theor. Popul. Biol.* 4, 129–132. doi: 10.1016/0040-5809(73)90008-7

Sved, J. A., McRae, A. F., and Visscher, P. M. (2008). Divergence between human populations estimated from linkage disequilibrium. *Am. J. Hum. Genet.* 83, 737–743. doi: 10.1016/j.ajhg.2008.10.019

Tenesa, A., Navarro, P., Hayes, B. J., Duffy, D. L., Clarke, G. M., Goddard, M. E., et al. (2007). Recent human effective population size estimated from linkage disequilibrium. *Genome Res.* 17, 520–526. doi: 10.1101/gr.6023607

Uimari, P., and Tapio, M. (2011). Extent of linkage disequilibrium and effective population size in Finnish Landrace and Finnish Yorkshire pig breeds. *J. Anim. Sci.* 89, 609–614. doi: 10.2527/jas.2010-3249

Wang, J. (2005). Estimation of effective population sizes from data on genetic markers. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360, 1395–1409. doi: 10.1098/rstb.2005.1682

Waples, R. S., and Do, C. (2008). LDNE: a program for estimating effective population size from data on linkage disequilibrium. *Mol. Ecol. Resour.* 8, 753–756. doi: 10.1111/j.1755-0998.2007.02061.x

Waples, R. S., and Do, C. (2010). Linkage disequilibrium estimates of contemporary Ne using highly variable genetic markers: a largely untapped resource for applied conservation and evolution. *Evol. Appl.* 3, 244–262. doi: 10.1111/j.1752-4571.2009.00104.x

Weir, B. S., and Hill, W. G. (1980). Effect of mating structure on variation in linkage disequilibrium. *Genetics* 95, 477–488.

Wright, S. (1943). Isolation by distance. *Genetics* 28, 114–138.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Barbato, Orozco-terWengel, Tapio and Bruford. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

LOCAL BREEDS

frontiers in GENETICS

MINI REVIEW ARTICLE
published: 25 February 2015
doi: 10.3389/fgene.2015.00033

Challenges and opportunities in genetic improvement of local livestock breeds

*Filippo Biscarini[1], Ezequiel L. Nicolazzi[1], Alessandra Stella[1,2], Paul J. Boettcher[3] and Gustavo Gandini[4]**

[1] *Parco Tecnologico Padano, Lodi, Italy*
[2] *Institute of Agricultural Biology and Biotechnology, National Research Council, Milan, Italy*
[3] *Animal Production and Health Division, Food and Agriculture Organization of the United Nations, Rome, Italy*
[4] *Department of Veterinary Sciences and Public Health, University of Milan, Milan, Italy*

Edited by:
Göran Andersson, Swedish University of Agricultural Sciences, Sweden

Reviewed by:
Tad S. Sonstegard, Agricultural Research Service, United States Department of Agriculture, USA
Daniel Fischer, MTT Agrifood Research Finland, Finland
Maja Ferencakovic, University of Zagreb, Croatia

****Correspondence:***
Gustavo Gandini, Department of Veterinary Sciences and Public Health, University of Milan, Via Celoria 10, Milan 20133, Italy
e-mail: gustavo.gandini@unimi.it

Sufficient genetic variation in livestock populations is necessary both for adaptation to future changes in climate and consumer demand, and for continual genetic improvement of economically important traits. Unfortunately, the current trend is for reduced genetic variation, both within and across breeds. The latter occurs primarily through the loss of small, local breeds. Inferior production is a key driver for loss of small breeds, as they are replaced by high-output international transboundary breeds. Selection to improve productivity of small local breeds is therefore critical for their long term survival. The objective of this paper is to review the technology options available for the genetic improvement of small local breeds and discuss their feasibility. Most technologies have been developed for the high-input breeds and consequently are more favorably applied in that context. Nevertheless, their application in local breeds is not precluded and can yield significant benefits, especially when multiple technologies are applied in close collaboration with farmers and breeders. Breeding strategies that require cooperation and centralized decision-making, such as optimal contribution selection, may in fact be more easily implemented in small breeds.

Keywords: local breeds, genetic diversity, selection, genomics, phenotyping

INTRODUCTION: THE FOCUS

Local breeds contribute across-breed genetic diversity to global animal genetic resources (AnGR). Unfortunately, many local breeds have a small population size which puts them at risk of extinction, according to the FAO (2013) system of categorization. Economics are a significant driver for loss of these breeds, as they tend to be less productive than common international transboundary breeds. Adding market value to local livestock breeds is a recognized strategy in conservation of AnGR (FAO, 2007), but the genetic improvement of breeds' traits is also a concrete option for increasing their profitability. However, formal selection has rarely been implemented in local breeds. In this regard, we argue that the limited efforts observed are not due to the presence of insurmountable constraints. This paper reviews and discusses challenges and opportunities in implementing genetic improvement for local livestock breeds. We address our analysis to breeds categorized as being at risk because of their small population sizes.

CHALLENGES

Although the constraints that breeds of small census size are faced with are not insurmountable, they are not trivial either. A first major challenge is simply a strong inertia built up by a typical small breed's history; small breeds are usually not small by accident. Some breeds are small because they were developed and adapted to a given region and isolated by geographical constraints from expansion into wider markets. Others are small because of either deliberate policies to remain exclusive or a lack of effort or success in promotion and marketing. Others have never been as productive as the best breeds and were thus not competitive enough economically to appeal to a wide number of breeders. Small breeds suffer because they cannot take advantage of economies of scale in breeding and marketing programs. Larger breeds have a greater opportunity to increase selection response, because a larger number of individuals allows for greater selection differentials, especially when artificial insemination and other reproductive biotechnologies can increase the number of offspring per individual. Breeding companies also have more interest in larger breeds because the potential market is greater and because the truly superior animals are more extreme and thus more valuable. This vicious cycle basically condemns breeds to remain small under "standard" conditions. However, this reality does not preclude that particular interventions can be undertaken to overcome these obstacles.

THE NEED FOR SELECTION PROGRAMS IN SMALL BREEDS

As explained, the inferior economic performance of a breed leads to decreased interest among farmers and eventually extinction. Therefore, some approach to selection is needed to increase economic performance: more output, less costs. In principle, all livestock breeds should be able to benefit from the advances in

animal breeding and improvement. Obviously, differences exist across world regions and among breeds. For each breed, a critical first step is to undertake a SWOT (strengths, weaknesses, opportunities and threats; Martín-Collado et al., 2013) or similar analysis to identify logical breeding objectives and strategies to achieve those objectives. Some local breeds already have sufficiently high production to achieve profitability, but low performance with regard to secondary characters, such as with the udder conformation in the Reggiana cattle from Italy (Gandini et al., 2007). Other breeds may obtain greater benefits by improving output while maintaining their characteristic secondary traits, such as adaptation to the environment. For example, the Valdostana cattle are uniquely adapted to their mountainous production environment, but have relatively low milk yield. A well-designed selection program may seek to improve milk yield, but do this gradually to avoid creating energy demands that cannot be met by mountain pastures. At the same time, selection must avoid increasing body size and maintain leg conformation because these cows must go up to mountain pastures and negotiate steep slopes, meaning their center of gravity must remain low. In less economically developed countries, the importation of exotic breeds is often still proposed as a quick solution, but without adaptation failure is likely to occur, especially during years when climatic extremes occur. On the other hand, the time and capital invested in selection within breed will eventually lead to adapted and profitable populations (FAO, 2013). FAO (2013) has collected examples of successful selection programs in local breeds.

GENETIC VARIATION WITHIN BREEDS

In conservation genetics, maintenance of both within-breed and across-breed genetic diversity are primary aims (Ollivier and Foulley, 2005), as they play different but critical roles in sustaining animal production. Selection can negatively affect both these components, and breeding programs should cautiously monitor genetic variation. Within a breed, the level and rate of inbreeding, which is negatively correlated with effective population size, is generally used as parameter of within breed variation. Index selection based on information from relatives leads to the reduction of effective population size and increases the probability of co-selecting close relatives (Wray and Thompson, 1990). This is particularly true for traits with low heritability. In designing selection programs for small populations, a main challenge is to maximize genetic gain at an acceptable inbreeding rate. Inbreeding rate per generation should be below 1.0% (Meuwissen and Woolliams, 1994), which may preclude selection when the population size is particularly small. However, as the population size increases, selection intensity can be increased, resulting in a continuum of situations with respect to selection differential. Different strategies are available to achieve genetic response with control of inbreeding, ranging from the earlier methods using sub-optimal criteria of selection (e.g., Grundy et al., 1994; Villanueva et al., 1994), to considering genetic relationships among selected animals (e.g., Brisbane and Gibson, 1995), to the most-sophisticated selection with optimal contributions (OC; Meuwissen, 1997; Grundy et al., 1998).

Various studies have compared OC selection retrospectively with the observed values after truncation selection. Potential for up to 30% more genetic gain at a given inbreeding rate was revealed in Meatlinc sheep and Aberdeen Angus populations (Avendaño et al., 2003). Koenig and Simianer (2006) observed 13% more genetic gain from OC selection compared to the actual breeding program for German Holstein bulls, under the same average relationship constraint. These studies determined that a lack of inbreeding control had resulted in unbalanced use of ancestors in these populations and underline the benefits of OC selection. This is particularly true in local breeds with limited population size, where inbreeding rates are expected to be high using conventional selection.

When acceptable rates of inbreeding are not known *a priori*, it is possible to generate a response surface of inbreeding versus genetic gain to facilitate the choice of the inbreeding level to be adopted (Brisbane and Gibson, 1995; Meuwissen, 1997).

GENETIC VARIATION AMONG BREEDS

The genetic diversity of a livestock species is generally addressed by keeping a sufficient number of breeds. However, species-wide diversity must also be considered during selection within a breed. In general, FAO (2013) suggests that selection should conserve breeds as genetically and cultural distinct genetic resources. Selection for increased output while ignoring traits correlated to traits of conservation interest such as adaptation, specific genetic variants, and quality of products, can reduce breed distinctiveness and between-breed variation. Identification of selection traits in local breeds should be accurate and based upon knowledge of the trait biology. Advances in genomics and bioinformatics have allowed the identification of genomic similarities/differences among livestock breeds (see de Simoni Gouveia et al., 2014 for a review). Some of these genomic signatures may contribute to explain the phenotypic uniqueness of breeds (Huson et al., 2014; Somavilla et al., 2014) and facilitate prioritization and the use of genomic breeding tools to preserve these important traits. A further option is the landscape genomics approach, whereby the association between alleles and geographic locations and/or climatic variables is targeted and assumed to be suggestive of signatures of adaptation, giving information on the environmental forces acting on the genome (Joost et al., 2013).

ECONOMIC SUSTAINABILITY

Genetic improvement programs require significant investments. Although well-designed genetic improvement can be expected to eventually provide positive returns on investments, in local breeds the costs will often be relatively high on a per-animal basis. The breeding strategy and system that maximize genetic response may not be optimal from an economic standpoint. Recording of performance and pedigrees may not be economically sustainable, even if restricted to a portion of the population (e.g., the nucleus). Lack of infrastructure in marginal areas where local breeds are often found may impair the introduction of genetic improvement programs, and the development of the infrastructure may be costly. When considering artificial insemination, the production of few semen doses per donor, as expected in small populations, can substantially increase per-dose semen costs relative to large populations. Therefore, a cost–benefit analyses should be conducted before implementing selection in local

breeds to determine the optimal approach. In genetic improvement programs, economic returns should be evaluated in the long term, given generation intervals and genetic cumulative effects. Additionally, organizational and infrastructural shortcomings are often associated to local breeds: these could be circumvented by taking advantage of existing organizations and infrastructures developed by larger breeding organizations: examples are the genetic evaluation for Meuse–Rhine–Yssel cattle carried out by CRV[1] in the Netherlands, and regional breeds cattle data managed by ICBF[2] in Ireland.

OPPORTUNITIES

GENOMICS

In the last decade, low- and high-density genomic tools (Lenstra et al., 2012; Utsunomiya et al., 2014) have been broadly used to study and characterize the genetic diversity and population structure of livestock (FAO, 2011; Rothschild and Plastow, 2014). However, genomic information has contributions beyond characterization to make to the management and conservation of AnGR. In addition to identifying genomic regions subject to natural selection, signatures of artificial selection have been identified applying statistical approaches to genomic data (Stella et al., 2010; Randhawa et al., 2014). Other uses of genomic information with potentially high impacts on management include: the estimation of genome-based relationships and inbreeding coefficients, from single nucleotide polymorphisms (SNPs; e.g., Manichaikul et al., 2010; VanRaden et al., 2011a), from runs of homozygosity (e.g., Purfield et al., 2012), or unifying different sources of information (e.g., Wang and Da, 2014); genetic approaches for breed-based product identification and traceability, for authentication and quality assurance (Nicoloso et al., 2013); and the identification of recessive lethals or other specific mutations of interest (VanRaden et al., 2011b; Pirola et al., 2013). All of the above can be applied to small breeds at relatively low cost (e.g., using low-density or custom panels). However, as is the case for the analysis of any breed that had not been considered in the development of the SNP chips, ascertainment bias should be carefully taken into account (Albrechtsen et al., 2010; Lachance and Tishkoff, 2013).

In terms of economics, the greatest impact of genomics in livestock has been its application to breeding (i.e., genomic selection, GS). However, a large number of phenotyped (i.e., directly or through daughter/relative performances) and genotyped individuals are necessary to obtain accurate genomic breeding values (Goddard and Hayes, 2009). Breeding schemes combining traditional and genomic information have been proven to obtain good results in medium-scale breeds (Thomasen et al., 2014). However, considering the low number of individuals in small populations, only small gains from the use of genomic information can be expected (Pryce and Daetwyler, 2012). One possible solution is multi- or across-breed prediction, although its utility depends on various factors, including genetic distances among populations and the trait(s) considered (Lund et al., 2014). Such strategies have been successfully tested on relatively large dairy cattle breeds (Lund et al., 2011; Hozé et al., 2014), but are yet untested on small breeds. Given that application of GS in small breeds would require high-density genotyping, once again a specific cost–benefit analysis should be carefully considered. To face this issue, statistical methods that integrate genotyped and ungenotyped individuals could be adopted (Misztal et al., 2009). Another option, especially in populations where pedigree information is known, could be to genotype key ancestors at high-density and the rest of the population at low-medium density (Hozé et al., 2014). Imputation methods, also with a multi-breed reference population, can then be applied to obtain high genotyping accuracies for all animals (Berry et al., 2014).

INNOVATIVE PHENOTYPES

Technological developments in agriculture have impacted not only the field of genomics, but also the collection of phenotypes. Animal phenotyping has mainly taken two directions: on one hand, the measurement of an increasingly large array of new phenotypes (Houle et al., 2010); on the other, the development of systems for automatic trait measurement and recording (e.g., Berry et al., 2012).

Genomic selection, in which trait measurement is limited to the reference population, contributed to put emphasis on the collection of novel phenotypes (e.g., Schöpke, 2014). For instance, milk quality traits now include not only total protein and fat content but also sub-components like lactoferrin and fatty acids (see RobustMilk, 2012). Mid- and near-infrared spectroscopy (MIR/NIR) allow quantitative evaluation of the composition of biological samples and have found wide application in dairy cattle breeding (e.g., De Marchi et al., 2014). Health-related traits represent yet another field of phenotypic investigation and include direct veterinary records, indirect measures of mastitis (e.g., milk electrical conductivity, milk mineral content) and female fertility (e.g., milk hormone assays, physical activity), and traits related to lameness or metabolic syndromes (Egger-Danner et al., 2014). Growing interest is being placed on behavioral traits like cow temperament (König et al., 2006) or feather pecking in laying hens (e.g., Biscarini et al., 2010).

Falling genotyping prices have left trait measurement as the major economically limiting factor in livestock selection schemes, thereby motivating active research in the (semi)automatic acquisition of phenotypes on a large scale. Automated milk-recording systems are becoming popular (e.g., Biscarini et al., 2012). The industry has been developing sensors to automatically measure many traits of direct or indirect interest. Pedometers for ambulatory activity indirectly measure fertility and lameness, and rumination tags monitor rumination time, which is related to metabolic activity and methane emissions (Soriani et al., 2012; Methagene, 2014). Image and video analysis can yield predictors of meat yield and quality (Pabiou et al., 2010) or body condition (Berry et al., 2012).

The combination of novel phenotypes and automatic trait recording is a powerful tool to improve both herd management and breeding schemes. Unfortunately, similar to GS (and conventional selection, for that matter), economies of scale are usually important in making the application of these technologies affordable and, therefore, innovative phenotyping is currently affecting mainly large commercial livestock populations. Nevertheless, this

[1]https://www.crv4all.nl/service/mrij
[2]http://www.icbf.com/

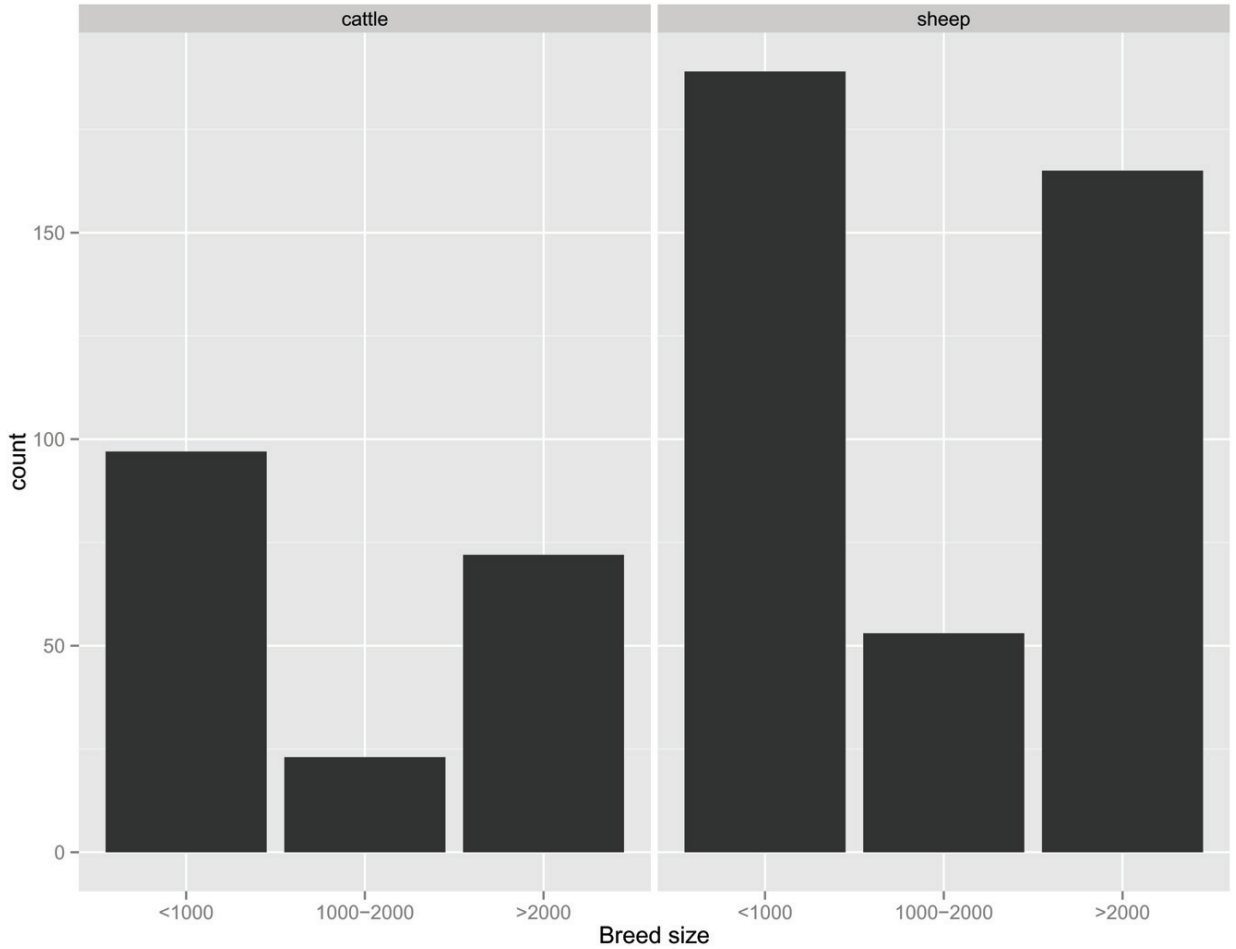


FIGURE 1 | Distribution of world cattle and sheep breeds according to the number of breeding females (data source DAD-IS: www.dad.fao.org, accessed on January 13, 2015): <1,000, 1,000–2,000, >2,000.

trend can still be a very promising development for smaller local breeds, inasmuch as the new technologies can be developed and perfected in larger populations, increasing their efficiency and decreasing costs so that can eventually help fill the gap toward optimized breeding and management practices in small breeds.

BREEDING STRATEGIES

The small census size of local breeds does bring potential advantages. For example, implementation of OC selection requires decision-making to be centralized, which is impossible in large commercial populations where autonomy is widely dispersed and breeding organizations compete. In local breeds, fewer stakeholders are involved, so such coordination may be possible. Haile-Mariam et al. (2007) proposed a practical method to maximize genetic gain and minimize inbreeding by selection of dairy bulls on the genetic index of their progeny weighted by the cost of their expected inbreeding, a method that—due to its simplicity—could be promoted in local breeds. However, to date the OC method has been mainly analyzed in simplified populations with high selection intensities and rarely in multi stage selection schemes (Hinrichs and Meuwissen, 2011) and under the conditions encountered in local breeds (Gourdine et al., 2012; Gandini et al., 2014a,b).

Gourdine et al. (2012) simulated a breeding program aimed to improve meat quality in a local pig breed farmed in low-input systems with a given herd structure. OC selection at inbreeding rates around 0.001 per generation was shown to achieve reasonable gains. In dairy cattle populations with 500–6,000 females it has been shown that substantial genetic gain—about 50–70% of that achievable in large populations—can be obtained at an inbreeding rate per generation of about 0.001 (Gandini et al., 2014a). FAO (2013) advices selection for production to be implemented in breeds categorized as "vulnerable"; in cattle and sheep this translates to breeds with a number of breeding females between 1,000 and 2,000. Breeds with a larger breeding population are not considered at risk, while breeds with less than 1,000 breeding females are regarded as endangered: in this latter group of breeds selection programs are not advised. **Figure 1** shows the number of cattle and sheep breeds registered in the FAO database DAD-IS[3] under the vulnerable category, corresponding to 23 and 53

[3]www.dad.fao.org, accessed on January 13, 2015

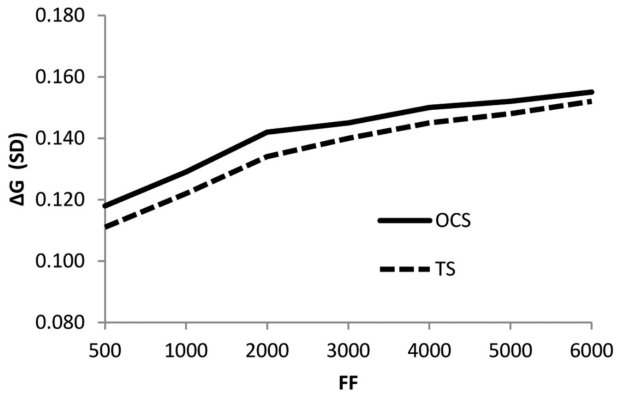


FIGURE 2 | Annual genetic gain (ΔG) (standard deviation units) with optimum contribution selection (OCS) and truncation selection (TS) in populations from 500 to 6,000 females (FF) with inbreeding control (Gandini et al., 2014b).

potential candidate breeds for selection programs in cattle and sheep, respectively.

Figure 2 shows the genetic gain as a function of population size under a young bull scheme with OC selection, compared to truncation selection at the same rate of inbreeding of 0.001 per generation (Gandini et al., 2014a). Other studies have reported much greater potential advantages of OC selection, from 16 to 44% relative to truncation selection; these were due to the much higher selection intensity and index accuracy used in these simulations (Gandini et al., 2014a). Kosgey et al. (2006), based on an analysis of selection for small ruminants in the tropics, underline that the success of breeding programs is mainly determined by their compatibility with the farming conditions and the involvement of the farmers, and that simplicity and applicability of the systems should be a major criterion in designing the breeding scheme. Close cooperation with farmers will also facilitate adoption of complementary practices, such as niche marketing and exploitation of breeds' environmental services (FAO, 2013), which may be as important as genetic improvement in achieving profitability.

When pedigree and performance recording is limited, as often it occurs in the local breed scenario, genetic improvement can be generated in a small fraction of the population, the nucleus (e.g., Roden, 1994) and then disseminated to the whole population, with or without the use of artificial insemination. Annual genetic gains ranging from a minimum of 0.073 SD/generation (100-female nucleus for a commercial population of 500 females) to a maximum of 0.138 SD/generation (400-female nucleus for a commercial population of 5,000 females) have been simulated in small ruminant populations (Gandini et al., 2014a). A limitation with some nucleus schemes can be the genetic lag occurring between the commercial population and the nucleus (Bichard, 1971).

CONCLUSION

Small local breeds face many challenges in maintaining or increasing their population sizes in order to avoid extinction. Their inferior productivity (relative to larger breeds) is a common reason why many small breeds are small. In turn, their size inhibits small breeds from exploiting economies of scale afforded to large breeds in genetic selection. Nevertheless, some form of genetic improvement is almost necessary for small breeds to have any hope for long-term survival. Although small breeds may not be able to fully utilize technological advances such as GS and innovative phenotyping, they can benefit from the ongoing development and use of these technologies in large breeds. Advanced methods to optimize genetic response and maintenance of diversity may actually be more easily applied in breeds with small population sizes and fewer stakeholders. Application of a battery of genetic tools, along with close cooperation with breeders and utilization of other tools such as innovative product marketing (FAO, 2013) may allow small breeds to not only survive, but also thrive.

REFERENCES

Albrechtsen, A., Nielsen, F. C., and Nielsen, R. (2010). Ascertainment biases in SNP chips affect measures of population divergence. *Mol. Biol. Evol.* 27, 2534–2547. doi: 10.1093/molbev/msq148

Avendaño, S., Villanueva, B., and Woolliams, J. A. (2003). Expected increases in genetic merit from using optimized contributions in two livestock populations of beef cattle and sheep. *J. Anim. Sci.* 81, 2964–2975.

Berry, D. P., Cromie, A. R., McHugh, N., Burke, M., Pabiou, T., MacCarthy, J., et al. (2012). "New traits for dairy cattle breeding," in *Proceedings 38th International Committee for Animal Recording (ICAR) Meeting*, Cork, Ireland.

Berry, D. P., McClure, M. C., and Mullen, M. P. (2014). Within- and across-breed imputation of high-density genotypes in dairy and beef cattle from medium- and low-density genotypes. *J. Anim. Breed. Genet.* 13, 165–172. doi: 10.1111/jbg.12067

Bichard, M. (1971). Dissemination of genetic improvement through a livestock industry. *Anim. Prod.* 13, 401–411. doi: 10.1017/S0003356100010606

Biscarini, F., Bovenhuis, H., Van Der Poel, J., Rodenburg, T. B., Jungerius, A. P., and Van Arendonk, J. A. M. (2010). Across-line SNP association study for direct and associative effects on feather damage in laying hens. *Behav. Genet.* 40, 715–727. doi: 10.1007/s10519-010-9370-0

Biscarini, F., Nicolazzi, E. L., Stella, A., and the Prozoo Team. (2012). "Automated milk-recording systems: an experience in Italian dairy cattle farms," in *Proceedings of the Book of Abstracts of the 63rd Annual Meeting of the EAAP*, 27–31 August 2012, Bratislava, Slovakia, 95.

Brisbane, J. R., and Gibson, J. P. (1995). Balancing selection response and rate of inbreeding by including genetic relationship in selection decisions. *Theor. Appl. Genet.* 91, 421–431.

De Marchi, M., Toffanin, V., Cassandro, M., and Penasa, M. (2014). Mid-infrared spectroscopy as phenotyping tool for milk traits. *J. Dairy Sci.* 97, 1171–1186. doi: 10.3168/jds.2013-6799

de Simoni Gouveia, J. J., Barbosa da Silva, M. V. G., Rezende Paiva, S., and Pinheiro de Oliveira, S. M. (2014). Identification of selection signatures in livestock species. *Genet. Mol. Biol.* 37, 330–342. doi: 10.1590/S1415-47572014000300004

Egger-Danner, C., Cole, J. B., Pryce, J. E., Gengler, N., Heringstad, B., Bradley, A., et al. (2014). Overview of new traits and phenotyping strategies in dairy cattle with a focus on functional traits. *Animal* 9, 191–207. doi: 10.1017/S1751731114002614

FAO. (2007). *The Global Strategy for the Management of Farm Animal Genetic Resources and the Interlaken Declaration*. Available at: ftp://ftp.fao.org/docrep/fao/meeting/014/j9556e.pdf (accessed October 10, 2014).

FAO. (2011). *Guidelines: Molecular Genetic Characterization of Animal Genetic Resources*. Available at: http://www.fao.org/docrep/014/i2413e/i2413e00.htm (accessed October 10, 2014).

FAO. (2013). *In Vivo Conservation of Animal Genetic Resources. FAO Animal Production and Health Guidelines.* No. 14. Rome: FAO.

Gandini, G., Del Corvo, M., Biscarini, F., and Stella, A. (2014a). Genetic improvement of small ruminant local breeds with nucleus and inbreeding control: a simulation study. *Small Rumin. Res.* 120, 196–203. doi: 10.1016/j.smallrumres.2014.06.004

Gandini, G., Stella, A., Del Corvo, M., and Jansen, G. B. (2014b). Selection with inbreeding control in simulated young bull schemes for local dairy cattle breeds. *J. Dairy Sci.* 97, 1790–1798. doi: 10.3168/jds.2013-7184

Gandini, G., Maltecca, C., Pizzi, F., Bagnato, A., and Rizzi, R. (2007). Comparing local and commercial breeds on functional traits and profitability: the case of Reggiana dairy cattle. *J. Dairy Sci.* 90, 2004–2011. doi: 10.3168/jds.2006-204

Goddard, M. E., and Hayes, B. J. (2009). Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nat. Rev. Genet.* 10, 381–391. doi: 10.1038/nrg2575

Gourdine, J. L., Sorenson, A. C., and Rydhmer, L. (2012). There is room for selection in a small local pig breed when using optimum contribution selection: a simulation study. *J. Anim. Sci.* 90, 76–84. doi: 10.2527/jas.2011-3898

Grundy B., Caballero, A., Santiago, E., and Hill, W. G. (1994). A note on using biased parameter values and non-random mating to reduce rates of inbreeding in selection programmes. *Anim. Prod.* 59, 465–468. doi: 10.1017/S0003356100008011

Grundy, B., Villanueva, B., and Woolliams, J. A. (1998). Dynamic selection procedures for constrained inbreeding and their consequences for pedigree development. *Genet. Res.* 72, 159–168. doi: 10.1017/S0016672398003474

Haile-Mariam, M., Bowman, P. J., and Goddard, M. (2007). A practical approach for minimising inbreeding and maximising genetic gain in dairy cattle. *Genet. Sel. Evol.* 30, 369–389. doi: 10.1186/1297-9686-39-4-369

Hinrichs, D., and Meuwissen, T. H. E. (2011). Analyzing the effect of different approaches of penalized relationship in multi stage selection schemes. *J. Anim. Sci.* 89, 3426–3432. doi: 10.2527/jas.2010-3621

Houle, D., Govindaraju, D. R., and Omholt, S. (2010). Phenomics: the next challenge. *Nat. Rev. Genet.* 11, 855–866. doi: 10.1038/nrg2897

Hozé, C., Fritz, S., Phocas, F., Boichard, D., Ducrocq, V., and Croiseau, P. (2014). Efficiency of multi-breed genomic selection for dairy cattle breeds with different sizes of reference population. *J. Dairy Sci.* 97, 3918–3929. doi: 10.3168/jds.2013-7761

Huson, H. J., Kim, E. S., Godfrey, R. W., Olson, T. A., McClure, M. C., Chase, C. C., et al. (2014). Genome-wide association study and ancestral origins of the slick-hair coat in tropically adapted cattle. *Front. Genet.* 5:101. doi: 10.3389/fgene.2014.00101

Joost, S., Vuilleumier, S., Jensen, J. D., Schoville, S., Leempoel, K., Stucki, S., et al. (2013). Uncovering the genetic basis of adaptive change: on the intersection of landscape genomics and theoretical population genetics. *Mol. Ecol.* 22, 3659–3665. doi: 10.1111/mec.12352

Koenig, S., and Simianer, H. (2006). Approaches to the management of inbreeding and relationship in the German Holstein dairy cattle population. *Livest. Sci.* 103, 40–53. doi: 10.1016/j.livsci.2005.12.009

König, S., Köhn, F., Kuwan, K., Simianer, H., and Gauly, M. (2006). Use of repeated measures analysis for evaluation of genetic background of dairy cattle behavior in automatic milking systems. *J. Dairy Sci.* 89, 3636–3644. doi: 10.3168/jds.S0022-0302(06)72403-1

Kosgey, I. S., Baker, R. L., Udo, H. M. J., and Van Arendonk, J. A. M. (2006). Successes and failures of small ruminant breeding programmes in the tropics: a review. *Small Rumin. Res.* 61, 13–28. doi: 10.1016/j.smallrumres.2005.01.003

Lachance, J., and Tishkoff, S. A. (2013). SNP ascertainment bias in population genetic analyses: why it is important, and how to correct it. *Bioessays* 35, 780–786. doi: 10.1002/bies.201300014

Lenstra, J. L., Groeneveld, L. F., Eding, H., Kantanen, J., Williams, J. L., Taberlet, P., et al. (2012). Molecular tools and analytical approaches for the characterization of farm animal genetic diversity. *Anim. Genet.* 43, 483–502. doi: 10.1111/j.1365-2052.2011.02309.x

Lund, M. S., de Roos, A. P. V., de Vries, A. G., Druet, T., Ducrocq, V., Fritz, S., et al. (2011). A common reference population from four European Holstein populations increases reliability of genomic predictions. *Genet. Sel. Evol.* 43, 43. doi: 10.1186/1297-9686-43-43

Lund, M. S., Su, G., Janss, L., Guldbrandtsen, B., and Brondum, R. F. (2014). Genomic evaluation of cattle in a multi-breed context. *Livest. Sci.* 166, 101–110. doi: 10.1016/j.livsci.2014.05.008

Manichaikul, A., Mychaleckyj, J. C., Rich, S. S., Daly, K., Sale, M., and Chen, W. M. (2010). Robust relationship inference in genome-wide association studies. *Bioinformatics* 26, 2867–2873. doi: 10.1093/bioinformatics/btq559

Martín-Collado, D., Díaz, C., Mäki-Tanila, A., Colinet, F., Duclos, D., Hiemstra, S. J., et al. (2013). The use of SWOT analysis to explore and prioritize conservation and development strategies for local cattle breeds. *Animal* 7, 885–894. doi: 10.1017/S175173111200242X

Methagene. (2014). *Large-scale Methane Measurements on Individual Ruminants for Genetic Evaluations*. Available at: http://www.methagene.eu/ (accessed October 15, 2014).

Meuwissen, T. H. E. (1997). Maximizing the response of selection with predefined rate of inbreeding. *J. Anim. Sci.* 75, 934–940.

Meuwissen, T. H. E., and Woolliams, J. A. (1994). Effective size of livestock populations to prevent a decline in fitness. *Theor. Appl. Genet.* 89, 10119–11026.

Misztal, I., Legarra, A., and Aguilar, I. (2009). Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. *J. Dairy Sci.* 92, 4648–4655. doi: 10.3168/jds.2009-2064

Nicoloso, L., Crepaldi, P., Mazza, R., Ajmone-Marsan, P., and Negrini, R. (2013). Recent advance in DNA-based traceability and authentication of livestock meat PDO and PGI products. *Recent Pat. Food Nutr. Agric.* 5, 9–18. doi: 10.2174/2212798411305010004

Ollivier, L., and Foulley, J. (2005). Aggregate diversity: new approach combining within and between breed genetic diversity. *Livest. Prod. Sci.* 95, 247–254. doi: 10.1016/j.livprodsci.2005.01.005

Pabiou, T., Fikse, W. F., Cromie, A. R., Kean, M. G., Näsholm, A., and Berry, D. P. (2010). Use of digital images to predict carcass cut yields in cattle. *Livest. Sci.* 137, 130–140. doi: 10.1016/j.livsci.2010.10.012

Pirola, Y., Della Vedova, G., Bonizzoni, P., Stella, A., and Biscarini, F. (2013). "Haplotype-based prediction of gene alleles using pedigrees and SNP genotypes," in *Proceedings of the International Conference on Bioinformatics, Computational Biology and Biomedical Informatics (BCB'13)*. New York, NY: ACM, 9 pages.

Pryce, J. E., and Daetwyler, H. D. (2012). Designing dairy cattle breeding schemes under genomic selection: a review of international research. *Anim. Prod. Sci.* 52, 107–114. doi: 10.1071/AN11098

Purfield, D. C., Berry, D. P., McParland, S., and Bradley, D. G. (2012). Runs of homozygosity and population history in cattle. *BMC Genet.* 13:70. doi: 10.1186/1471-2156-13-70

Randhawa, I. A. S., Khatkar, M. S., Campbell Thomson, P., and Raadsma, H. W. (2014). Composite selection signals can localize the trait specific genomic regions in multi-breed populations of cattle and sheep. *BMC Genet.* 15:34. doi: 10.1186/1471-2156-15-34

RobustMilk. (2012). *Innovative and Practical Breeding Tools for Improved Dairy Products from More Robust Dairy Cattle*. Available at: http://www.robustmilk.eu/ (accessed October 15, 2014).

Roden, J. A. (1994). Review of the theory of open nucleus breeding systems. *Anim. Breed. Abstr.* 62, 151–157.

Rothschild, M. F., and Plastow, G. S. (2014). Applications of genomics to improve livestock in the developing world. *Livest. Sci.* 166, 76–83. doi: 10.1016/j.livsci.2014.03.020

Schöpke, K. (2014). "*Assembling a Reference Population—From Genetic Architecture to New Phenotypes*," in *Proceedings of the 10th WCGALP*, Vancouver.

Somavilla, A. L., Sonstegard, T. S., Higa, R. H., Rosa, A. N., Siqueira, F., Silva, L. O. C., et al. (2014). A genome-wide scan for selection signatures in Nellore cattle. *Anim. Genet.* 45, 771–781. doi: 10.1111/age.12210

Soriani, N., Trevisi, E., and Calamari, L. (2012). Relationships between rumination time, metabolic conditions, and health status in dairy cows during the transition period. *J. Anim. Sci.* 90, 4544–4554. doi: 10.2527/jas.2011-5064

Stella, A., Ajmone-Marsan, P., Lazzari, B., and Boettcher, P. (2010). Identification of selection signatures in cattle breeds selected for dairy production. *Genetics* 185, 1451–1461. doi: 10.1534/genetics.110.116111

Thomasen, J. R., Egger-Danner, C., Willam, A., Guldbrandtsen, B., Lund, M. S., and Sorensen, A. C. (2014). Genomic selection strategies in a small dairy cattle population evaluated for genetic gain and profit. *J. Dairy Sci.* 97, 458–470. doi: 10.3168/jds.2013-6599

Utsunomiya, Y. T., Bomba, L. B., Lucente, G., Colli, L., Negrini, R., Lenstra, J. A., et al. (2014). Revisiting AFLP fingerprinting for an unbiased assessment of genetic structure and differentiation of taurine and zebu cattle. *BMC Genet.* 15:47. doi: 10.1186/1471-2156-15-47

VanRaden, P. M., Olson, K. M., Wiggans, G. R., Cole, J. B., and Tooker, M. E. (2011a). Genomic inbreeding and relationships among Holsteins, Jerseys, and Brown Swiss. *J. Dairy Sci.* 94, 5673–5682. doi: 10.3168/jds.2011-4500

VanRaden, P. M., Olson, K. M., Null, D. J., and Hutchison, J. L. (2011b). Harmful recessive effects on fertility detected by absence of homozygous haplotypes. *J. Dairy Sci.* 94, 6153–6161. doi: 10.3168/jds.2011-4624

Villanueva, B., Woolliams, J. A., and Simm, G. (1994). Strategies for controlling rates of inbreeding in MOET nucleus schemes for beef cattle. *Genet. Sel. Evol.* 25, 517–535. doi: 10.1186/1297-9686-26-6-517

Wang, C., and Da, Y. (2014). Quantitative genetics model as the unifying model for defining genomic relationship and inbreeding coefficient. *PLoS ONE* 9:e114484. doi: 10.1371/journal.pone.0114484

Wray, N. R., and Thompson, R. (1990). Prediction of rates of inbreeding in selected populations. *Genet. Res.* 55, 41–54. doi: 10.1017/S0016672300025180

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 02 December 2014; paper pending published: 31 December 2014; accepted: 25 January 2015; published online: 25 February 2015.

*Citation: Biscarini F, Nicolazzi EL, Stella A, Boettcher PJ and Gandini G (2015) Challenges and opportunities in genetic improvement of local livestock breeds. Front. Genet. **6**:33. doi: 10.3389/fgene.2015.00033*

This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics.

Copyright © 2015 Biscarini, Nicolazzi, Stella, Boettcher and Gandini. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in
GENETICS

MINI REVIEW ARTICLE
published: 12 February 2015
doi: 10.3389/fgene.2015.00024

Recent advances in understanding the genetic resources of sheep breeds locally-adapted to the UK uplands: opportunities they offer for sustainable productivity

Dianna Bowles [1,2]*

[1] Department of Biology, University of York, York, UK
[2] The Sheep Trust, University of York, York, UK

Edited by:
Stéphane Joost, Swiss Federal Institute of Technology in Lausanne, Switzerland

Reviewed by:
Christine Flury, Swiss College of Agriculture, Switzerland
Licia Colli, Università Cattolica del Sacro Cuore, Italy

***Correspondence:**
Dianna Bowles, Department of Biology, University of York, York YO10 5DD, UK
e-mail: dianna.bowles@york.ac.uk

Locally adapted breeds of livestock are of considerable interest since they represent potential reservoirs of adaptive fitness traits that may contribute to the future of sustainable productivity in a changing climate. Recent research, involving three hill sheep breeds geographically concentrated in the northern uplands of the UK has revealed the extent of their genetic diversity from one another and from other breeds. Results from the use of SNPs, microsatellites, and retrovirus insertions are reviewed in the context of related studies on sheep breeds world-wide to highlight opportunities offered by the genetic resources of locally adapted hill breeds. One opportunity concerns reduced susceptibility to Maedi Visna, a lentivirus with massive impacts on sheep health and productivity globally. In contrast to many mainstream breeds used in farming, each of the hill breeds analyzed are likely to be far less susceptible to the disease threat. A different opportunity, relating specifically to the Herdwick breed, is the extent to which the genome of the breed has retained primitive features, no longer present in other mainland breeds of sheep in the UK and offering a new route for discovering unique genetic traits of use to agriculture.

Keywords: biodiversity, adaptive fitness, sheep, genomics, farm animal genetic resources

INTRODUCTION AND CONTEXT

The importance of locally adapted livestock breeds is becoming recognized through their contribution to food security in marginal land areas of the world. Reviews by the FAO and other international agencies highlight the crucial significance of these breeds, both as sustainable resources of food but also as living reservoirs of biodiversity, providing genetic adaptive fitness traits to improve mainstream agriculture (FAO, 2009; Hoffmann, 2013).

In the UK, the locally adapted hill breeds of sheep are typically farmed on Disadvantaged or Severely Disadvantaged Areas for agricultural production (Defra, 2012). These breeds are characterized by their ability to thrive in the harsh environments of mountains, fells, and moorlands, rearing lambs on low inputs of feed and management. The breeds are not rare and exist in many tens of thousands. But, due to their adaptation to local environments, each breed is concentrated within a single region of the UK (Carson et al., 2009). In these regions, the sheep continue to be commercially farmed and contribute to the heritage and economies of their guardian communities.

The extent of the breed's geographical concentration, leads to their susceptibility and disproportionate losses if an infectious disease enters their region. This issue was first recognized in the UK during the Foot and Mouth Disease (FMD) epidemic of 2001 which started in the northwest of England and disproportionately affected livestock farmed in that region [Archive.defra.gov.uk/foodfarm/farmanimal/diseases/atoz/fmd/2001; www.royalsoc.ac.uk (2004). Policy document 25/04 ISBN 0854036105 Royal Society Infectious Disease in Livestock Inquiry: Follow Up review]. Losses to the Herdwick sheep breed were of particular concern, which led to the setting up of the first national gene bank for sheep and a government commitment to survey the precise level of geographical concentration that existed amongst commercially farmed regional breeds.

The Sheep Trust, a national UK charity, working with Sheep Breeder Associations (SBAs) and more than 1000 breeders, georeferenced individual flocks of 12 regional breeds (Carson et al., 2009). The data showed certain of these Heritage Breeds were indeed extremely concentrated with for example, up to 95% of the Herdwicks numbering some 47,000 animals, tightly clustered within 23 km of the breed's mean center in the Lake District National Park of northern England.

Endangerment of the genetic resources of livestock may arise through disease but significantly, for the hill breeds of sheep throughout Europe, changes in the policies of national governments and EU regulations are also leading to new risks (www.publications.parliament.uk/pa/cm201011/cmselect/cmenvfru/556/556.pdf; http://publications.naturalengland.org.uk/publication/410700). Historically, sheep numbers increased as a reflection of headage payments in the EC's Common Agricultural Policy (CAP) and this undoubtedly led to some overgrazing of the uplands. Changes to CAP then switched to

subsidy payments based on land area rather than numbers of animals farmed. This is leading to major reductions in hill sheep numbers, since farmers are no longer being paid on flock sizes. In some areas, agri-environment schemes to promote rejuvenation of the vegetation are also decreasing sheep numbers further with zero stocking of the hills for certain times of the year.

Recognition of risk to a sheep breed's genetic resources through geographical concentration has recently been accepted in policies for breed protection and conservation (www.gov.uk/government/uploads/system/uploads/attachment_data/file/254127/uk-breeds-at-risk.pdf). However, it is also crucial to determine the genetic distinctiveness of each breed. This is an indicator of the biodiversity deserving protection and is also a potential route to identifying adaptive fitness traits that can underpin the future sustainability and security of food production.

This mini review describes new insights gained from a recent genetic study of Herdwicks, Rough Fell and Dalesbred, three hill breeds of the UK (Bowles et al., 2014). These locally adapted breeds, each geographically concentrated in the northern uplands of England have not been previously analyzed. The new findings add to our current understanding of sheep genetic resources world-wide.

ANALYTICAL TOOLS

Molecular genetic studies of sheep breeds have used a number of approaches to analyze breed origins, relatedness, distinctiveness and selection (reviewed in Groeneveld et al., 2010; Lenstra et al., 2012).

These include single nucleotide polymorphisms (SNPs) using a candidate gene approach, as well as the increasingly widespread application of the *OvineSNP50BeadChip* that is leading to considerable advances (for example, Pariset et al., 2006a,b; Kijas et al., 2009, 2012; Johnston et al., 2011, 2013; Heaton et al., 2012, 2013, 2014; Riggio et al., 2013; Zhang et al., 2013; Fariello et al., 2014; Gutiérrez-Gil et al., 2014; Hazard et al., 2014; Lv et al., 2014; McRae et al., 2014; Periasamy et al., 2014), including those addressing global sheep diversity undertaken by the International Sheep Genomics Consortium (Kijas et al., 2009, 2012). These studies complement and extend the earlier use of microsatellites to investigate variation and population structure (for example, Alvarez et al., 2005; Tapio et al., 2005a,b, 2007; Lawson Handley et al., 2007; Peter et al., 2007), particularly in the study of European breeds. Also, an unrelated approach based on retrotyping, has enabled the origins and routes of domestication to be investigated through analyzing insertions of a panel of endogenous Jaagsiekte sheep retroviruses (enJSRVs) into defined sites of the genome (Arnaud et al., 2007; Chessa et al., 2009).

A selection of these approaches was applied to study the genetic resources of the three geographically concentrated hill breeds, each farmed in adjacent regions of the UK northern uplands. (Bowles et al., 2014; http://www.herdwick-sheep.com; www.dalesbredsheep.co.uk; www.roughfellsheep.co.uk). Equal numbers of registered rams from each of the breeds comprised the populations selected because provenance, familial origins and lack of relatedness could be confirmed. Also, it was reasoned that rams contribute the major genetic resource to a breed, whilst forming only a small fraction of the total breed numbers.

INSIGHTS GAINED

REDUCED SUSCEPTIBILITY TO MAEDI VISNA

Small ruminant lentiviruses infect millions of animals globally, with Maedi-Visna (MV) a major disease, reducing sheep health and productivity as well as lifespan. Studies by Heaton et al. (2012, 2013) have clearly demonstrated that variations in the gene encoding TMEM154, an ovine transmembrane protein, are associated with infection of sheep to the lentivirus. Polypeptide variants with the presence of glutamate (E) at position 35 in the ancestral, full length version of the protein are associated with increased susceptibility to the lentivirus, whereas lysine (K) at position 35, or deletion mutants are associated with reduced susceptibility (Heaton et al., 2012).

Heaton reported the average frequency of the highly susceptible *TMEM154* alleles was 0.51 across 74 sheep breeds world-wide (Heaton et al., 2013). Significantly, >25% of those analyzed, including some mainstream breeds such as the Scottish Texel, showed a frequency of above 0.8 indicating a major latent health and welfare risk for the global sheep industry. This contrasts with the recent data on the three hill breeds (Bowles et al., 2014), with frequencies for the susceptible allele ranging from 0.26 to 0.42, suggesting a lower than average risk of MV infection and a much greater likely resilience to the disease threat.

In the UK, flocks of continental breeds such as Texels (www.texel.co.uk), stringently accredited for the absence of MV, represent a major component of the mainstream sheep industry. Genetic resistance of the hill breeds to lentivirus infections offers an alternative route to achieve an infection-free status for the national flock and an important reason to protect and continue to make use of the genetic resources of these locally adapted breeds.

LACK OF IN-BREEDING AMONGST LOCALLY ADAPTED BREEDS

Panels of microsatellites have been used with great effect to study many aspects of sheep breeds world-wide (Alvarez et al., 2005; Tapio et al., 2005a,b, 2007; Lawson Handley et al., 2007; Peter et al., 2007). These have included large-scale comparisons such as 57 breeds across Europe and the Middle East by Peter and coworkers of the ECONOGENE Consortium (Peter et al., 2007) and 32 breeds of Northern Europe by Tapio et al. (2005a), as well as studies on breeds farmed in close geographical proximity, such as in regions of the Baltic and in northern Spain (Alvarez et al., 2005; Tapio et al., 2005b, 2007).

In the study of the three UK hill breeds, a subset of microsatellites were applied to investigate the possibility of in-breeding within the populations analyzed (Bowles et al., 2014). This is relevant since within the extensive hill farming systems of the uplands, detailed written pedigrees are not maintained and for some breeds, such as the Herdwicks, only rams are formally registered. This informal management of genetic resources is a risk and differs from monitoring in the commercial breeding of mainstream breeds and in the conservation breeding of numerically scarce rare breeds.

However, when the occurrence of heterozygote deficiency within populations per marker per breed was evaluated, no significant inbreeding was revealed. The average local inbreeding coefficient (F_{IS}) was weakly positive in all cases, with the lowest

value of 0.059 in Herdwick and highest in Rough Fell (0.162). These values are within the range previously found for two other northern UK breeds with substantially greater population sizes (Peter et al., 2007), the Swaledale (0.070) and the Scottish Blackface (0.031), indicating the informal practices supported by Breed Societies are currently sufficient to maintain diversity within the genetic pool of each breed.

ORIGINS AND OPPORTUNITIES OF A PRIMITIVE GENOME

The use of enJSRVs as genetic markers (Arnaud et al., 2007; Chessa et al., 2009) enabled Chessa and co-workers to explore the historical origins and routes of domestication of sheep breeds across the world. They reasoned that given each endogenous retrovirus in a host genome arises from a single irreversible integration event, populations sharing the retrovirus DNA in the same genomic location must be related phylogenetically. Four common retrovirus insertions enabled them to design a classification scheme of "retrotypes" that was applied to 133 breeds globally. Genomes with none of the common insertions were rare and defined as "primitive" since they preceded any of the integration events. Only 10 breeds analyzed world-wide were found to have an abundance of individuals with this primitive characteristic (Chessa et al., 2009).

In the recent study of the hill breeds, very surprisingly, the Herdwick population contained a high proportion of animals with a "primitive" (R0) genome. In contrast, R0 was absent from Rough Fell and Dalesbred, both of which contained abundant retrotypes more typical of other UK breeds.

The presence of R0 individuals retained in the present day Herdwick population, provides an indicator of origins of domestication as well as farming practices over the centuries. The data suggest Herdwicks originate from a common ancestral founder flock to other "primitive" breeds retaining the R0 retrotype. Within the UK only the North Ronaldsay of the Orkney Islands are known to have R0's and to an extreme proportion. Elsewhere, substantial levels of R0's were found in populations in Sweden, Finland and Iceland (Chessa et al., 2009). The fact that R0 individuals continue in these present day breeds most probably reflects the degree of isolation with which they have been farmed since their first introduction. In turn, this may also reflect the original extent of their "fitness" to those harsh northern, upland environments and this fitness has continued to be maintained over the centuries by farming practices.

A second surprising feature shared both by the Herdwicks and the Rough Fell, was the presence of two extremely rare retrovirus insertions in their populations. These insertions were so rare that they were not included in the global retrotype classification and in fact in that earlier study, the pair was only found in Texel breed populations (Chessa et al., 2009). Present day Texels originate from a number of breeds (www.texel.co.uk), none of which were shown to have the rare retrovirus insertions (Chessa et al., 2009). This suggests therefore that it may be a very distant historical association between the two UK hill breeds and current Texels, such as involving a common origin in the form of the Pin-tail ancestral population of Texel Island in the Wadden Sea. This region, now designated as a World Heritage site, was once above sea level and from archeological remains is known to have been a major trading post for Vikings as they moved outwards to colonize the south and west (Besteman, 2004; Knole, 2010). Intriguingly, local folklore in Cumbria has always linked the arrival of Herdwicks to the early Viking settlers (Brown, 2009). The new genetic data suggest that link extends to the Rough Fell, although the lack of any R0 retrotypes in today's Rough Fell population implies a subsequent greater divergence from the common founder flock, compared to the Herdwicks.

CONCLUSIONS AND WAY FORWARD

Genetic diversity in farm animals is becoming recognized as a highly significant resource. With the rapid progress in DNA sequencing technologies and the availability of SNP chips for many major species, genetic breeding is increasingly being used for improvement. As yet however, many of these practical applications continue to focus on production traits and the mainstream breeds of livestock.

Predicted impacts of climate change, increases in food requirement for the world population and economic costs of farming are all contributing to renewed interest in the genetic resources of those breeds that are locally adapted to their farmed habitats. This interest is of relevance to improving a sustainable form of agricultural productivity in both the developed and developing worlds. Low inputs with good feed conversion, hardiness to harsh environments, enhanced resistance to parasites and pathogens, and low management costs are all characteristics of local breeds when farmed extensively in the habitats to which they are adapted (FAO, 2009; Hoffmann, 2013).

In this context, locally adapted sheep breeds are production systems of our food on land that often cannot support any other forms of agriculture. Their ability benefits current food security and provides opportunities for future farming systems that will be required to maintain productivity in the predicted harsh conditions of climate change. Despite the benefits offered by these breeds, the very real risk of their genetic erosion requires urgent action before these living reservoirs of adaptive fitness traits become lost forever.

The first step is to analyze their genetic distinctiveness to demonstrate to policy makers the biodiversity that the sheep breeds contain. The three UK breeds recently studied have not been analyzed previously, but extend earlier molecular studies of other locally adapted and mainstream breeds. A surprisingly broad genetic range was observed for the hill breeds, even though they were farmed in close proximity within adjacent regions of the northern uplands of England.

Next generation sequencing, the use of the *OvineSNP50BeadChip* and global collaborative networks such as the International Sheep Consortium, will provide many more opportunities for large-scale genetic characterization of sheep breeds. Already there are examples of those technologies used to explore traits relevant to adaptation, such as growth characteristics (Zhang et al., 2013), milk production (Gutiérrez-Gil et al., 2014), and social behavior (Hazard et al., 2014). Attention is also focusing on genetic factors important for resisting gastrointestinal (GI) parasites (Riggio et al., 2013; McRae et al., 2014; Periasamy et al., 2014). Infections from GI nematodes have major impacts on sheep health, welfare and productivity through to

economic costs for management input and use of control agents (Nieuwhof and Bishop, 2005) and the problem is worsening through emerging resistance to the major anthelmintic classes (Beech et al., 2011). Selective breeding, whether for resistance or tolerance (Bishop, 2012) is likely to be a major output from future genomic studies.

However, many more data are required linking the different phenotypic traits for potential use in genetic improvement to geographies and the specific ecologies and parasite challenges to which the livestock are adapted. Recent advances in landscape genomics and the bringing together of large data sets combining phenotypic recording, DNA analyses and environmental parameters are providing a new platform to achieve this holistic understanding (Joost et al., 2007, 2013; Colli et al., 2014). Significantly, use of the *OvineSNP50BeadChip* has recently been applied in a large project studying 32 sheep breeds adapted to a wide spectrum of different regional climates (Lv et al., 2014). Two hundred and thirty SNPs were identified with evidence for selection likely due to climate-mediated pressure and 17 strong candidate genes were highlighted to be under environmental adaptive selection. This study confirms the utility of a genomic approach to understand adaptation of sheep to their environments and if now applied to geographically concentrated breeds, such as the Herdwick with its potential abundance of primitive features, there is real potential to uncover the genetic basis of the breed's unique adaptations.

ACKNOWLEDGMENT

Emma Rankine is greatly thanked for her computing skills and patience.

REFERENCES

Alvarez, I., Gutierrez, J. P., Royo, L. J., Fernandez, I., Gomez, E., Arranz, J. J., et al. (2005). Testing the usefulness of the molecular coancestry information to assess genetic relationships in livestock using a set of Spanish sheep breeds. *J. Anim. Sci.* 83, 737–744.

Arnaud, F., Caporale, M., Varela, M., Biek, R., Chessa, B., Alberti, A., et al. (2007). A paradigm for virus–host coevolution: sequential counter-adaptations between endogenous and exogenous retroviruses. *PLoS Pathog.* 3:e170. doi: 10.1371/journal.ppat.0030170

Beech, R. N., Skuce, P., Bartley, D. J., Martin, R. J., Prichard, R. K., Gilleard, J. S. (2011). Anthelmintic resistance: markers for resistance or susceptibility? *Parasitology* 138, 160–174. doi: 10.1017/S0031182010001198

Besteman, J. (2004). "Two Viking hordes from the former island Wieringen (The Netherlands) Viking relations with Fresia in an archeological perspective," in *Land, Sea and Home, Settlement in the Viking period*, eds J. Hines, A. Lane, and M. Redknap (Leeds: Maney Publishing), 93–108.

Bishop, S. C. (2012). A consideration of resistance and tolerance for ruminant nematode infections. *Front. Genet.* 3:168. doi: 10.3389/fgene.2012.00168

Bowles, D., Carson, A., and Isaac, P. (2014). Genetic distinctiveness of the herdwick sheep breed and two other locally adapted hill breeds of the UK. *PLoS ONE* 9:e87823. doi: 10.1371/journal.pone.0087823

Brown, G. (2009). *Herdwicks – Herdwick Sheep and the English Lake District.* Kirkby Stephen: Hayloft Publishing.

Carson, A., Elliott, M., Groom, J., Winter, A., and Bowles, D. (2009). Geographical isolation of native sheep breeds in the UK – evidence of endemism as a risk factor to genetic resources. *Livest Sci.* 123, 288–299. doi: 10.1016/j.livsci.2008.11.026

Chessa, B., Pereira, F., Arnaud, F., Amorim, A., Goyache, F., Mainland, I., et al. (2009). Revealing the history of sheep domestication using retrovirus integrations. *Science* 324, 532–536. doi: 10.1126/science.1170587

Colli, L., Joost, S., Negrini, R., Nicoloso, L., Crepaldi, P., Ajmone-Marsan, P., et al. (2014). Assesing the spatial dependence of adaptive loci in 43 European and Western Asian goat breeds using AFLP markers. *PLoS ONE* 9:e86668. doi: 10.1371/journal.pone.00866

Defra. (2012). *UK Country Report on Farm Animal Genetic Resources.* UK Government.

FAO. (2009). Livestock keepers – guardians of biodiversity. *Animal Production and Health Paper No. 167*. Rome.

Fariello, M., Servin, B., Tosser-Klopp, G., Rupp, G., Moreno, C., San Cristobal, M., et al. (2014). Selection signatures in worldwide sheep populations. *PLoS ONE* 9:e103813. doi: 10.1371/journal.pone.0103813

Groeneveld, L. F., Lenstra, J. A., Eding, H., Toro, M. A., Scherf, B., Pilling, D., et al. (2010). Genetic diversity in farm animals – a review. *Anim. Genet.* 41, 6–31. doi: 10.1111/j.1365-2052.2010.02038.x

Gutiérrez-Gil, B., Arranz, J., Pong-Wong, R., García-Gámez, E., Kijas, J., Wiener, P. (2014). Application of selection mapping to identify genomic regions associated with dairy production in sheep. *PLoS ONE* 9:e94623. doi: 10.1371/journal.pone.0094623

Hazard, D., Moreno, C., Foulquié, D., Delval, E., François, D., Bouix, J., et al. (2014). Identification of QTLs for behavioral reactivity to social separation and humans in sheep using the OvineSNP50 BeadChip. *BMC Genomics* 15:778. doi: 10.1186/1471-2164-15-778

Heaton, M. P., Clawson, M. L., Chito-McKown, C. G., Leymaster, K. A., Smith, T. P. L., Harhay, G. P., et al. (2012). Reduced lentivirus susceptibility in sheep with TMEM154 mutations. *PLoS Genet.* 8:e1002467. doi: 10.1371/journal.pgen.1002467

Heaton, M. P., Kalbfleisch, T. S., Petrik, D. T., Simpson, B., Kijas, J. W., Clawson, M. L., et al. (2013). Genetic testing for THEMM154 mutations associated with lentivirus susceptibility in sheep. *PLoS ONE* 8:e55490. doi: 10.1371/journal.pone.0087823

Heaton, M. P., Leymaster, K. A., Kalbfleisch, T. S., Kijas, J. W., Clarke, S. M., McEwan, J., et al. (2014). SNPs for parentage testing and traceability in globally diverse breeds of sheep. *PLoS ONE* 9:e94851. doi: 10.1371/journal.pone.0094851

Hoffmann, I. (2013). Adaptation to climate change – exploring the potential of locally adapted breeds. *Animal* 7, 346–362. doi: 10.1017/S1751731113000815

Johnston, S. E., Gratten, J., Berenos, C., Pilkington, J. G., Clutton-Brock, T. H., Pemberton, J. M., et al. (2013). Life history trade-offs at a single locus maintain sexually selected genetic variation. *Nature* 502, 93–95. doi: 10.1038/nature12489

Johnston, S. E., McEwan, J. C., Pickering, N. K., Kijas, J. W., Beraldi, D., Pilkington, J. G., et al. (2011). Genome-wide association mapping identifies the genetic basis of discrete and quantitative variation in sexual weaponry in a wild sheep population. *Mol. Ecol.* 20, 2555–2566. doi: 10.1111/j.1365-294X.2011.05076.x

Joost, S., Bonin, A., Bruford, M. W., Depres, L., Conord, C., Erhardt, G., et al. (2007). A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. *Mol. Ecol.* 16, 3955–3969. doi: 10.1111/j.1365-294X.2007.03442.x

Joost, S., Vuilleumier, S., Jensen, J., Schoville, S., Leempoel, K., Stucki, S., et al. (2013). Uncovering the genetic basis of adaptive change: on the intersection of landscape genomics and theoretical population genetics. *Mol. Ecol.* 22, 3659–3665. doi: 10.1111/mec.12352

Kijas, J. W., Lenstra, J. A., Hayes, B., Boitard, S., Porto Neto, L. R., San Magali, C., et al. (2012). Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. *PLoS Biol.* 10:e1001258. doi: 10.1371/journal.pbio.1001258

Kijas, J. W., Townley, D., Dalrymple, B. P., Heaton, M. P., Maddox, J. F., McGrath, A., et al. (2009). A genome wide survey of SNP variation reveals the genetic structure of sheep breeds. *PLoS ONE* 4:e4668. doi: 10.1371/journal.pone.0004668

Knole, E. (2010). "Frisia in Carolingian times," in *Viking Trade and Settlement*, ed I. S. Klaesoe (Copenhagen: Museum Tusculanum Press), 43–60.

Lawson Handley, L. J., Byrne, K., Santucci, F., Townsend, S., Taylor, M., Bruford, M., et al. (2007). Genetic structure of European sheep breeds. *Heredity* 99, 620–631. doi: 10.1038/sj.hdy.6801039

Lenstra, J. A., Groeneveld, L. F., Eding, H., Kantanen, J., Williams, J. L., Taberlet, P., et al. (2012). Molecular tools and analytical approaches for the characterisation of farm animal genetic resources. *Anim. Genet.* 43, 483–502. doi: 10.1111/j.1365-2052.2011.02309.x

Lv, F.-H., Agfa, S., Kantanen, J., Colli, L., Stucki, S., Kijas, J. W., et al. (2014). Adaptations to climate-mediated selective pressures in sheep. *Mol. Biol. Evol.* 31, 3324–3343 doi: 10.1093/molbev/msu264

McRae, K. M., McEwan, J. C., Dodds, K. G., Gemmell, N. (2014). Signatures of selection in sheep bred for resistance or susceptibility to gastrointestinal nematodes. *BMC Genomics* 15:637. doi: 10.1186/1471-2164-15-637 PMCID: PMC4124167

Nieuwhof, G. J., and Bishop, S. C. (2005). Costs of the major endemic diseases of sheep in Great Britain and the potential benefits of reduction in disease impact. *Anim. Sci.* 81, 23–29 doi: 10.1079/ASC41010023

Pariset, L., Cappuccio, I., Ajmone-Marsan, P., Bruford, M., Dunner, S., Cortes, O., et al. (2006b). Characterisation of 37 breed-specific single nucleotide polymorphisms in sheep. *J. Hered.* 97, 531–534. doi: 10.1093/jhered/esl020

Pariset, L., Cappuccio, I., Joost, S., D'Andrea, M. D., Marletta, D., Ajmone Marsan, P., et al. (2006a). Characterisation of SNPs in sheep and their variation as evidence of selection. *Anim. Genet.* 37, 290–292. doi: 10.1111/j.1365-2052.2006.01441.x

Periasamy, K., Pichler, R., Poli, M., Cristel, S., Cetrá, B., Medus, D., et al. (2014). Candidate gene approach for parasite resistance in sheep – variation in immune pathway genes and association with fecal egg count. *PLoS ONE* 9:e88337. doi: 10.1371/journal.pone.0088337

Peter, C., Bruford, M., Perez, T., Dalamitra, S., Hewitt, G., Erhardt, G. (2007). Genetic diversity and subdivision of 57 European and Middle-Eastern sheep breeds. *Anim. Genet.* 38, 37–44. doi: 10.1111/j.1365-2052.2007.01561.x

Riggio, V., Matika, O., Pong-Wong, R., Stear, M. J., Bishop, S. C. (2013). Genome-wide association and regional heritability mapping to identify loci underlying variation in nematode resistance and body weight in Scottish Blackface lambs. *Heredity* 110, 420–429. doi: 10.1038/hdy.2012.90

Tapio, I., Tapio, M., Grislis, Z., Holm, L. E., Jeppsson, S., Kantanen, J., et al. (2005b). Unfolding of population structure in Baltic sheep breeds using microsatellite analysis. *Heredity* 94, 448–456. doi: 10.1038/sj.hdy.6800640

Tapio, M., Ozerov, M., Viinalass, H., Kiseliova, T., Kantanen, J. (2007). Molecular genetic variation in sheep of the central Volga area inhabited by Finno-Ugric peoples. *Agric. Food Sci.* 16, 157–169. doi: 10.2137/145960607782219346

Tapio, M., Tapio, I., Grislis, Z., Holm, L. E., Jeppsson, S., Kantanen, J., et al. (2005a). Native breeds demonstrate high contributions to the molecular variation in northern European sheep. *Mol. Ecol.* 14, 3951–3963. doi: 10.1111/j.1365-294X.2005.02727.x

Zhang, L., Liu, J., Zhao, F., Ren, H., Xu, L., Lu, J., et al. (2013). Genome-wide association studies for growth and meat production traits in sheep. *PLoS ONE* 8:e66569. doi: 10.1371/journal.pone.0066569

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 14 November 2014; accepted: 16 January 2015; published online: 12 February 2015.

Citation: Bowles D (2015) Recent advances in understanding the genetic resources of sheep breeds locally-adapted to the UK uplands: opportunities they offer for sustainable productivity. Front. Genet. ***6****:24. doi: 10.3389/fgene.2015.00024*

This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics.

Copyright © 2015 Bowles. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in GENETICS

PERSPECTIVE ARTICLE
published: 05 March 2015
doi: 10.3389/fgene.2015.00073

A case study on strains of Buša cattle structured into a metapopulation to show the potential for use of single-nucleotide polymorphism genotyping in the management of small, cross-border populations of livestock breeds and varieties

Elli T. Broxham[1]*, Waltraud Kugler[1] and Ivica Medugorac[2]

[1] *SAVE Foundation, St. Gallen, Switzerland*
[2] *Chair of Animal Genetics and Husbandry, Faculty of Veterinary Medicine, Ludwig-Maximilians-University Munich, Munich, Germany*
**Correspondence: elli.broxham@gmail.com*

Edited by:
Stéphane Joost, École polytechnique fédérale de Lausanne, Switzerland

Reviewed by:
Ino Curik, University of Zagreb, Croatia
Levente Czegledi, University of Debrecen, Hungary

Keywords: cross-border, small populations, metapopulations, cattle, Balkans

For over 20 years, SAVE Foundation has concerned itself with the conservation of European livestock breeds. Often, this work has occurred in small populations of cross-border breeds. Livestock has developed over centuries of selection by farmers. The selection has been based on the ecological, climatic, and economic conditions of the local environments. One of these factors, the economic, is affected by the political situation of the country. Country borders as well as human populations have changed many times over the centuries. In contrast, the ecological and climate conditions have generally remained more stable. This means that highly similar locally adapted breeds can often be found both sides of current political borders and are generally grouped within the geo-ecological regions they developed in Druml et al. (2007).

Progressive displacement crossing by highly selected cosmopolitan breeds caused additional fragmentation of locally well-adapted but, in current economic environments, less successful breeds. Remnants from these formerly large populations may not amount to enough stock to make up a nucleus herd for conservation breeding within each single country. However, together with highly similar strains in bordering countries, there may be enough animals to create a combined gene pool and to conserve the locally adapted breed. Up to now, strategies for implementing this conservation method have relied heavily on historical and phenotypical information. Many projects conducted by SAVE over the years have been based on research into historical literature and data, field trips and discussions with experts with local knowledge. The high cost of genotyping combined with the relatively low usefulness of the results has meant that many projects carried out by SAVE and its partners have not used genotyping as a basis for conservation. The advent of SNP genotyping has changed this situation. The relatively low cost for genome-wide genotyping led to a substantial new situation where a higher level of accuracy can be brought into projects without creating a massive financial burden.

An example of this in action is the *BushaLive* project, funded under the "*Funding Strategy for the Implementation of the Global Plan of Action for Animal Genetic Resources.*" The *BushaLive* project targets the autochthonous Buša (also written as Busha) cattle breed of the Balkans. This breed survives in small, highly endangered, sub-populations that are additionally separated from each other due to the break-up of Yugoslavia into smaller states. The Buša is the collective term for small and robust shorthorn (brachyceros) cattle of the Balkans. Residuals from this formerly large Population are currently present in Albania, Bosnia and Herzegovina, Bulgaria, Croatia, Greece, Kosovo[1], Montenegro, Serbia and The Former Yugoslav Republic of Macedonia. Therefore, it is a typical example of the situation of cross-border breeds (SAVE Foundation, 2014).

The Buša is hardy and well-adapted to extensive farming in challenging environments with low managerial input. The Buša, with their withers' height at around 100 cm, show high fertility and modest production in harsh environments (Kompan et al., 2008). It is an important part of the local identity, but will be lost if conservation measures are not put in place to protect it. Stakeholders across the various nationalities and religions present in the Balkans share a common willingness to collaborate in conserving the breed.

Within the first phase of the *BushaLive* project blood samples have been taken from 254 animals in Albania, Bosnia and Herzegovina, Bulgaria, Croatia, Kosovo (see Footnote 1), Montenegro, Serbia, and The Former Yugoslav Republic of Macedonia. A minimum of 20 samples of possibly unrelated animals per country and strain was requested for the study. To obtain unbiased estimates of the diversity parameters, the population

[1]This designation is without prejudice to positions on status, and is in line with UNSCR 1244 and ICJ advisory opinion on the Kosovo declaration of independence.

history, genetic differentiation, and the degree of admixture in distinct Buša strains we performed the genome-wide SNP genotyping (the BovineSNP50 BeadChip). To assess admixture and diversity parameters, genotypes of eight reference populations have been included. These populations represent possible sources of admixture as well as having been subject to different levels of artificial selection. Four Buša strains sampled in former studies have also been included. These old samples complement the newly collected material (SAVE Foundation, 2014).

The achieved results provide a framework of future breeding actions and decisions which will be discussed between stakeholders involved in the conservation programme of this fragmented metapopulation (Medugorac et al., 2009). Even if different clusters of strains of Buša cattle are determined as well as different degrees of admixture in some of strains and animals within strains, the final conclusions as well as an estimation of the effective population size will only be possible after completion of all the analyses. However, the results obtained so far show that locally well-adapted strains that have never been intensively managed and differentiated into standardized breeds show large haplotype diversity. This suggests the need for a conservation and recovery strategy that does not rely exclusively on searching for the original native genetic background, but rather on the identification and removal of common introgressed haplotypes (SAVE Foundation, 2014).

Additional to the genotyping further information on each of the sampled animals has been collected via a comprehensive survey targeting their phenotypic characteristics and husbandry systems, as well as the products and services that they provide. This information, together with the genetic data, will be used to provide a basis for the development of a regional strategy for the management of the breed, spanning all stakeholder levels from farmers to governments. The project will also explore the potential for more effective marketing of the breeds' products. The next steps will be the establishment of basic recording systems and support for the development of breeding organizations and common breeding goals. The project will close with a stakeholder workshop for people working at all levels in the conservation of the breed. The event will provide an opportunity to pass on the information gathered and the strategies developed during the project to those who will use them in the future. All the results and data will be published, as they become available, online on: http://agrobiodiversity.net/balkan/ topic_network/Bushalive.asp.

REFERENCES

Druml, T., Curik, I., Baumung, R., Aberle, K., Distl, O., and Sölkner, J. (2007). Individual-based assessment of population structure and admixture in Austrian, Croatian and German draught horses. *Heredity (Edinb)*. 98, 114–122. doi: 10.1038/sj.hdy.6800910

Kompan, D., Cividini, A., and Simčič, M. (2008). *Current Status of the Brachycerous Cattle Populations in the South Eastern European Countries and Strategies for Their Sustainable Conservation*. Slovenia: Biotechnical Faculty, Zootechnical Department.

Medugorac, I., Medugorac, A., Russ, I., Veit-Kensch, C. E., Taberlet, P., Luntz, B., et al. (2009). Genetic diversity of European cattle breed highlights the conservation value of traditional unselected breeds with high effective population size. *Mol. Ecol.* 18, 3394–3410. doi: 10.1111/j.1365-294X.2009.04286.x

SAVE Foundation. (2014). *BushaLive Interim Activity Report, Switzerland*. Available online at: http://save-foundation.net/.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 15 September 2014; accepted: 12 February 2015; published online: 05 March 2015.

Citation: Broxham ET, Kugler W and Medugorac I (2015) A case study on strains of Buša cattle structured into a metapopulation to show the potential for use of single-nucleotide polymorphism genotyping in the management of small, cross-border populations of livestock breeds and varieties. Front. Genet. ***6****:73. doi: 10.3389/fgene.2015.00073*

This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics.

Copyright © 2015 Broxham, Kugler and Medugorac. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 13 May 2015
doi: 10.3389/fgene.2015.00173

Genomic analysis for managing small and endangered populations: a case study in Tyrol Grey cattle

Gábor Mészáros[1], Solomon A. Boison[1], Ana M. Pérez O'Brien[1], Maja Ferenčaković[2], Ino Curik[2], Marcos V. Barbosa Da Silva[3], Yuri T. Utsunomiya[4], Jose F. Garcia[4] and Johann Sölkner[1]*

[1] Division of Livestock Sciences, University of Natural Resources and Life Sciences, Vienna, Austria, [2] Department of Animal Science, University of Zagreb, Zagreb, Croatia, [3] Empresa Brasileira de Pesquisa Agropecuária, Juiz de Fora, Brazil, [4] UNESP-Universidade Estadual Paulista, Jaboticabal, Brazil

OPEN ACCESS

Edited by:
Paolo Ajmone Marsan, Università Cattolica del Sacro Cuore, Italy

Reviewed by:
Ikhide G. Imumorin, Cornell University, USA
Mahdi Saatchi, Iowa State University, USA
Paul Boettcher, Food and Agriculture Organization of the United Nations, Italy

***Correspondence:**
Gábor Mészáros, Division of Livestock Sciences, University of Natural Resources and Life Sciences, Augasse 2-6, A-1090 Vienna, Austria
gabor.meszaros@boku.ac.at

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: *15 November 2014*
Accepted: *20 April 2015*
Published: *13 May 2015*

Citation:
Mészáros G, Boison SA, Pérez O'Brien AM, Ferenčaković M, Curik I, Da Silva MVB, Utsunomiya YT, Garcia JF and Sölkner J (2015) Genomic analysis for managing small and endangered populations: a case study in Tyrol Grey cattle. Front. Genet. 6:173. doi: 10.3389/fgene.2015.00173

Analysis of genomic data is increasingly becoming part of the livestock industry. Therefore, the routine collection of genomic information would be an invaluable resource for effective management of breeding programs in small, endangered populations. The objective of the paper was to demonstrate how genomic data could be used to analyse (1) linkage disequlibrium (LD), LD decay and the effective population size (Ne_{LD}); (2) Inbreeding level and effective population size (Ne_{ROH}) based on runs of homozygosity (ROH); (3) Prediction of genomic breeding values (GEBV) using small within-breed and genomic information from other breeds. The Tyrol Grey population was used as an example, with the goal to highlight the potential of genomic analyses for small breeds. In addition to our own results we discuss additional use of genomics to assess relatedness, admixture proportions, and inheritance of harmful variants. The example data set consisted of 218 Tyrol Grey bull genotypes, which were all available AI bulls in the population. After standard quality control restrictions 34,581 SNPs remained for the analysis. A separate quality control was applied to determine ROH levels based on Illumina GenCall and Illumina GenTrain scores, resulting into 211 bulls and 33,604 SNPs. LD was computed as the squared correlation coefficient between SNPs within a 10 mega base pair (Mb) region. ROHs were derived based on regions covering at least 4, 8, and 16 Mb, suggesting that animals had common ancestors approximately 12, 6, and 3 generations ago, respectively. The corresponding mean inbreeding coefficients (F_{ROH}) were 4.0% for 4 Mb, 2.9% for 8 Mb and 1.6% for 16 Mb runs. With an average generation interval of 5.66 years, estimated Ne_{ROH} was 125 ($Ne_{ROH>16\,Mb}$), 186 ($Ne_{ROH>8\,Mb}$) and 370 ($Ne_{ROH>4\,Mb}$) indicating strict avoidance of close inbreeding in the population. The LD was used as an alternative method to infer the population history and the Ne. The results show a continuous decrease in Ne_{LD}, to 780, 120, and 80 for 100, 10, and 5 generations ago, respectively. Genomic selection was developed for and is working well in large breeds. The same methodology was applied in Tyrol Grey cattle, using different reference populations. Contrary to the expectations, the accuracy of GEBVs with very small within breed reference populations were very high, between 0.13–0.91 and 0.12–0.63, when estimated breeding values and deregressed breeding values were used as pseudo-phenotypes, respectively. Subsequent analyses confirmed the high

accuracies being a consequence of low reliabilities of pseudo-phenotypes in the validation set, thus being heavily influenced by parent averages. Multi-breed and across breed reference sets gave inconsistent and lower accuracies. Genomic information may have a crucial role in management of small breeds, even if its primary usage differs from that of large breeds. It allows to assess relatedness between individuals, trends in inbreeding and to take decisions accordingly. These decisions would be based on the real genome architecture, rather than conventional pedigree information, which can be missing or incomplete. We strongly suggest the routine genotyping of all individuals that belong to a small breed in order to facilitate the effective management of endangered livestock populations.

Keywords: breed management, endangered breeds, SNP chip, linkage disequilibrium, runs of homozygosity, genomic selection

Introduction

In the last decade, technological advancement has allowed for the genotyping of large numbers of single nucleotide polymorphisms (SNP) in the genome. The increase in SNP density was accompanied with decrease in price for the commercial SNP-chips, standard sets of SNPs selected, and sold by genotyping companies in large numbers, dominating animal, and plant breeding research in many countries.

Traditionally microsatellite markers were used for genotyping animals in population genetics studies. A popular set of microsatellites endorsed by the Food and Agriculture Organization (FAO) is widely used to evaluate genetic diversity in farm animals, especially endangered breeds (Baumung et al., 2004; Groeneveld et al., 2010). As a technological follow up, a set of SNPs could be used for a similar purpose. An advantage of the SNP markers is their occurrence on standard genotyping panels. Pooling of genotypes and comparison of different populations is feasible, contrary to the microsatellites, where (partially) different panels could be genotyped each time. The disadvantage of the current SNP panels from breed diversity perspective is their development in direction of commercial application in the most common species and breeds, with little research undertaken to prepare assays to replace the ISAG-FAO microsatellite panels.

The application of the SNP markers in animal breeding however, goes beyond population genetics. The early adopters were the large breeding organizations managing breeds with many animals and large financial capital. After the general success of the genomic selection approach (Meuwissen et al., 2001) the utilization of the genomic information has increased considerably. Today genomic breeding values (GEBV) are routinely used for making selection decisions, which has resulted in reducing the generation interval and increasing genetic gain compared to classical progeny testing systems in dairy cattle populations (Hutchison et al., 2014).

Genomic selection was an incentive to genotype nearly every young bull in many large cattle populations. This incentive is missing in smaller breeds because a large population size is generally perceived as a requirement to estimate reliable GEBV. Although there were numerous studies using SNP data in many small breeds, these are rather isolated efforts to demonstrate an interesting phenomenon or describe other interesting general aspect of a particular breed. Even though there are a relatively low number of animals to be genotyped in small populations, there is a general lack of routine genotyping in small breeds.

The objective of the paper is to demonstrate how genomic data could be used to ascertain population structure in small and endangered breeds, evaluate GEBV, and assess the range of potential applications from the breed management perspective. To tackle this goal, we have used the Tyrol Grey breed as an example to demonstrate some of the potential uses of genomic data in small and endangered populations. Some of the potential uses of genomic data are not applicable to the Tyrol Grey breed, but they are still extremely useful in the breed management context. We discuss such uses in the last part of the paper in order to give a comprehensive overview about the potential of genomics in small and endangered breeds.

Material and Methods

Data and Quality Control

The Tyrol Grey cattle is an endangered cattle breed with population size of consisting of 3785 breeding animals as of 2013 (ÖNGENE, www.oengene.at, 2014). We were able to genotype all available sires due to its small population size. From the available 218 Tyrol Grey AI bulls, we have genotyped 99 animals with the Illumina® BovineSNP50 BeadChip (50 K) and 119 animals with the Illumina® BovineHD BeadChip (HD) with about 770 K SNPs.

Only the 49,394 SNPs appearing on both 50 K and HD chips were retained, as the 50 K chip is a standard genotyping platform used for routine genotyping in taurine cattle. SNP markers with unknown positions and those on sex chromosome were excluded.

Two separate quality checks of the data were undertaken. The first quality check was done and the data used for estimating linkage disequilibrium (LD), effective population size (Ne) and also in genomic prediction approaches. The second quality check followed the approach of Ferenčaković et al. (2013). This was a more stringent quality check to help reduce error that might occur when estimating genomic inbreeding from runs of homozygosity (ROH).

The first quality check of the available SNP data was undertaken with the following criteria. SNPs with call rate less than 90% and Hardy-Weinberg equilibrium Fishers's exact *P*-value below 10^{-6} were removed using PLINK v1.07 (Purcell et al., 2007). SNP markers with minor allele frequency (MAF) < 0.05 and those mapped to the same physical positions were also deleted. Samples with call rate lower than 90% were also discarded. After quality control, 34,581 SNPs remained. The second quality check has been applied in the calculation of ROH as we did not exclude SNPs with low MAF, with high LD or those deviating from HWE. Genotyping errors were reduced by discarding SNPs with Illumina GenCall score ≤ 0.7, SNPs with Illumina GenTrain score ≤ 0.4 and animals with more than 5% missing genotypes. The same quality control settings has been used in Ferenčaković et al. (2013). The analyses were based on 211 bulls each genotyped for the same 33,604 SNPs with average distance of 73.655 kb between adjacent SNPs (from 23 to 1,955,291 bp), all placed on 29 autosomes. ROH segments were identified as a part of the genome in which 15 or more consecutive homozygous SNPs at a density of one SNP on every 100 kb are not more than 1 Mb apart. ROH calculations were done using the algorithm implemented in SNP and Variation Suite (v7.6.8 Win 64; Golden Helix, Bozeman, MT, USA www.goldenhelix.com).

Linkage Disequilibrium

The squared correlation (r^2) was used to measure the LD. The r^2 values were calculated using PLINK as pairwise comparisons of markers on the same chromosome, separated by less than 10 Mb. The decay of LD was analyzed using bins of 100 kb for the maximum distance between SNP pairs. Marker bins below 100 kb for a 50 K SNP panel generally generate very small numbers of pairwise LD values. PLINK calculates as r^2 between two SNPs:

$$r^2_{LD} = \left(\frac{\sum_{i=1}^{n} (g_{ij} - \bar{g}_j)(g_{im} - \bar{g}_m)}{\sqrt{\sum_{i=1}^{n} (g_{ij} - \bar{g}_j)^2} * \sqrt{\sum_{i=1}^{n} (g_{im} - \bar{g}_m)^2}} \right)^2$$

Where n = number of individuals with non-missing genotype; g is the genotype allele count of 0, 1, and 2 for AA, AB, and BB, respectively for individual i of SNP j and SNP m.

The calculations of effective population size (Ne_{LD}) were based on McEvoy et al. (2011). The Ne was based on the LD values:

$$E\left(r^2_{LD}\right) \approx \frac{1}{\alpha + 4Ne_{LD}c}$$

Where $E(r_{LD}{}^2)$ is the expected squared correlation of allele frequencies at a pair of loci, α is 2 when the impact of mutation is considered and 1 otherwise. Variable c is the genetic distance between loci in Morgans. The Ne was calculated as:

$$Ne_{LD} \approx \frac{1}{4c} * \left(\frac{1}{r^2_{LD}} - \alpha \right)$$

Assuming that the population has been constant in size, the approximation of Ne_{LD} is true for t generations ago, where $t = 1.2c$ (Hayes et al., 2003). It has been noted that LD patterns from shorter inter marker distances were informative about the Ne in the more distant past, while markers separated by longer distances are informative about the recent Ne. The relationships describing the historical development of Ne_{LD} should be considered only as an approximation, as LD patterns might be affected by variety of factors (de Roos et al., 2008).

Runs of Homozygosity

The inbreeding coefficients (F_{ROH}) were calculated from the formula proposed by McQuillan et al. (2008); $F_{ROH} = L_{ROH}/L_{AUTOSOME}$, where L_{ROH} is the total length of all ROH in the genome of an individual while $L_{AUTOSOME}$ refers to the specified length of the autosomal genome covered by SNPs on the chip (here 2,499,624,571 bp). We calculated three coefficients; $F_{ROH>4\,Mb}$, $F_{ROH>8\,Mb}$, and $F_{ROH>16\,Mb}$ defined by the minimum ROH lengths being higher than 4, 8, or 16 Mb, respectively. Under the assumption that 1 cM = 1 Mb, calculated inbreeding coefficients are expected to correspond to the reference ancestral population being remote approximately 12 ($F_{ROH>4\,Mb}$), 6 ($F_{ROH>8\,Mb}$), and 3 ($F_{ROH>16\,Mb}$) generations. For more detailed explanation see Howrigan et al. (2011) and Curik et al. (2014).

The calculation of the effective population size (Ne_{ROH}) was based on the equation $Ne_{ROH} = 1/2\Delta F$, where ΔF was calculated as regression coefficient b, representing change of $F_{ROH>4\,Mb}$ per 1 year (regression of $F_{ROH>4\,Mb}$ on the birth year), multiplied by the generation interval 5.66, previously calculated from the Tyrol Grey pedigree by using the software ENDOG, version 4.6 (Gutiérrez and Goyache, 2005), together with pedigree inbreeding coefficient (F_{PED}).

Autozygosity islands were defined as regions where SNPs had extreme ROH frequency (outliers according to boxplot distribution, see **Figure 6**).

Computation of descriptive statistics (PROC Means), bootstrap confidence intervals (SAS Macro), regression analysis (PROC REG), and graphical illustrations (PROC Boxplot, SAS Macro) were performed by the SAS software v 9.3.

Genomic Selection in Small Breeds

Single and multi-breed scenarios were considered to derive within and across breed GEBV. For the multi-breed scenarios the German-Austrian genotype pool of 6730 Fleckvieh and 1415 Brown Swiss bulls was used to extend the training set, as Fleckvieh is the major breed in Austria and Brown Swiss has common history with the Tyrol Grey. Breeding values (EBV), deregressed breeding values (dEBV) and their corresponding reliabilities for 10 major production and functional traits in Austria were provided by *Zuchtdata EDV- Dienstleistungen GmbH,* Austria. The deregression procedure of the EBVs removed the contribution of relatives other than daughters to the breeding value, based on the methodology of Garrick et al. (2009).

GEBV was estimated by fitting a polygenic effect assuming that every marker has a constant variance (GBLUP) (Meuwissen et al., 2001) i.e., assuming that each marker explains an equal proportion of the total genetic variance (σ^2_g). The GBLUP model was:

$$y = 1_n\mu + Zg + e$$

y = EBV or dEBV;
1_n = vector of 1 s;
μ = overall mean;
Z = design matrix allocating records to breeding values;
G = vector of random additive genetic effect using the genomic relationship matrix (G)
coming from $N(0, G\sigma_g^2)$;
e = vector of random residual errors $N(0, R\sigma_e^2)$, where R was diagonal matrix with weight calculated as $r^2/(1-r^2)$

To study the predictability of the above model, three strategies were used to group the animals into reference and validation sets, with main focus on the genomic evaluation of the Tyrol Grey breed.

1. Single breed scenario: Only Tyrol Grey bulls were used. The validation sets consisted from young bulls born after 2003, with the older Tyrol Grey bulls in the reference set.
2. Multi-breed scenario: The validation sets consisted from young Tyrol Grey bulls born after 2003, the reference sets included the rest of the Tyrol Grey bulls, as well as the Brown Swiss or the Fleckvieh or both Brown Swiss and Fleckvieh.
3. Across breed scenario: All Tyrol Grey bulls were put into the validation set. The reference set consisted of the Fleckvieh and/or Brown Swiss.

To be able to discuss the results obtained from the multi-breed and across breeds' scenarios, Eigen vectors, and values are computed on an estimated genomic relationship matrix with the three breeds. Principal component analysis plots are provided in **Figure 1**.

Prediction accuracy was measured as the correlations between the resulting GEBVs and pseudo-phenotype EBV. Bootstrapping procedure (sampling with replacement) was used to calculate the standard error of the correlation between the GEBV and EBV. The estimated GEBV were bootstrapped 10,000 times (this value appeared to give stable results) and the bootstrap GEBVs are correlated to the EBVs. The standard errors were calculated from the 10,000 estimated accuracies. This procedure gives us a fair estimate of the degree of dispersion of the estimated correlations. Although other cross validation procedure like random splitting procedures could have been employed, we chose to use forward prediction which is more relevant in breeding. In addition, limited number of individuals in the validation set also supports the idea of bootstrapping to calculate standard errors of the correlation estimates.

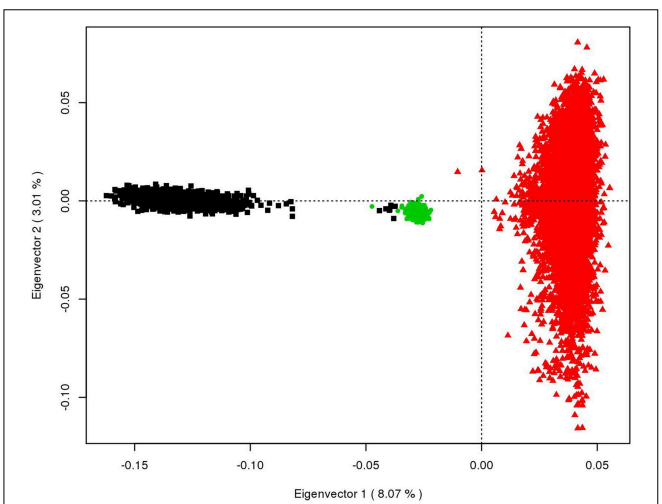


FIGURE 1 | Principal component analysis of the Tyrol Grey (in green), Brown Swiss (in black), and Fleckvieh (in brown) breeds; The amount of explained variance by the first two eigenvectors is shown in brackets.

Results

Linkage Disequilibrium and Effective Population Size

LD was computed as squared correlations ($r_{LD}{}^2$) for all SNP pairs within chromosomes. The LD was high for markers close to each other, but decayed quickly with growing inter marker distance (**Figure 2**). The $r_{LD}{}^2$ was around 0.7 for very short inter-marker distances below 10 kb, but was 0.1 for marker distances at ~150 kb. After this it followed by a moderate decay until 10 Mb ($r^2_{LD} = 0.03$). The average inter-marker distance in our study was ~75 kb, with average $r_{LD}{}^2$ of 0.192 ± 0.254 for adjacent markers.

The LD was used to estimate historical Ne (**Figure 3**). The method of calculation allows varying genetic distance and mutation occurrence, leading to slightly different results. In our case we calculated historical Ne based on the most likely scenario, i.e., considering mutations to occur and genetic distance per unit of physical distance (cM/Mb) of 1.25, according to Arias et al. (2009). The Ne_{LD} was around 200 about 20 generations in the past and declined to about 80 in the following 15 generations. The standard deviations of estimates show the uncertainty caused by slightly differing results from multiple LD windows pointing to the same generation.

Runs of Homozygosity

There was a considerable difference among animals in number of ROH segments and the length of the genome covered by these ROH segments (**Figure 4**). For example when there was only a single ROH segment, this could be 8 to 60 Mb long. The cumulative length of the ROH segments of 60 Mb could be due to a single ROH segment, or the sum of 10 smaller ROH segments (**Figure 4**). Similar distributions were observed for other animals, with higher differences between total lengths of homozygous regions as the number of ROH increased. The age of inbreeding is defined as the time to the common ancestor and is quantified with the length of the ROH segments. Thus, the minimal ROH length of 4 Mb implies a common ancestor dating 12 generations in the past. Similarly the minimal ROH length of 8 and 16 Mb implies a common ancestor dating 6 and 3 generations ago, respectively.

The summary statistics for the three ROH ($F_{ROH>4\,Mb}$, $F_{ROH>8\,Mb}$,and $F_{ROH>16\,Mb}$) and one pedigree (F_{PED}) inbreeding coefficients are presented in **Table 1** and in **Figure 5**. Considering pedigree depth of the Tyrol Grey population, the values obtained are in agreement with the assumption that $F_{ROH>4\,Mb}$, $F_{ROH>8\,Mb}$, and $F_{ROH>16\,Mb}$ correspond to the reference ancestral population where the common ancestors

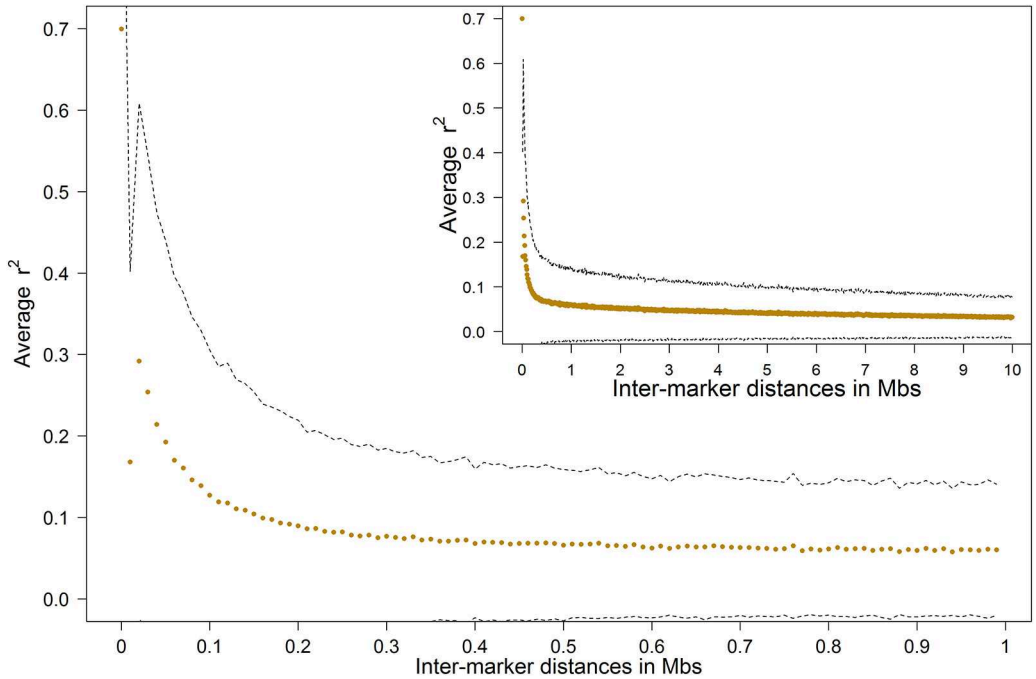


FIGURE 2 | Average LD decay in the Tyrol Grey population, dashed lines show the standard deviation boundary.

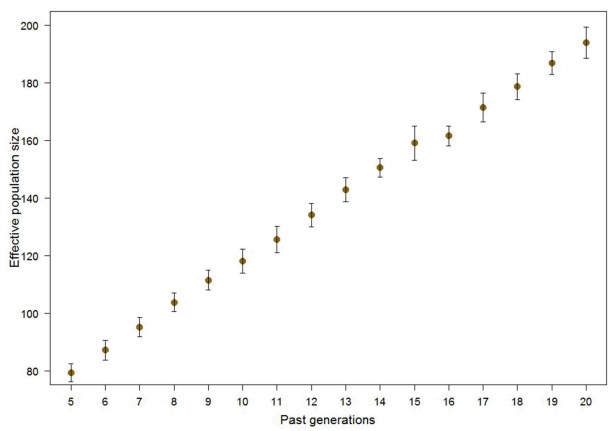


FIGURE 3 | Means and standard deviations for the historical Ne_{LD} in Tyrol Grey cattle, accounting for mutation and ratio of genetic per physical distance of 1.25.

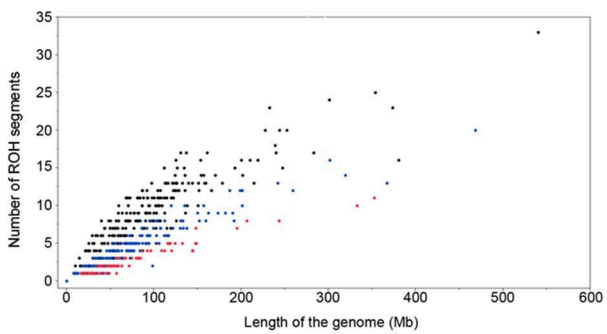


FIGURE 4 | Number of ROH segments and the length of the genome covered by ROH segments (minimum ROH length set to 4 Mb in black, 8 Mb in blue and 16 Mb in red).

are approximately considered to be 12, 6, and 3 generations remote as well as with values obtained in other populations (Ferenčaković et al., 2013). Animals with extreme pedigree inbreeding, for example after the threshold where $F_{PED} > 0.05$, were precisely identified by $F_{ROH>8\,Mb}$. The two peaks in **Figure 5** are caused by the fact that $F_{ROH>16\,Mb}$ values cannot be smaller than 0.006, i.e., 16 Mb divided by genome covered by SNPs on the chip.

The relationship between the number of ROH segments and the length of the genome covered by ROH is shown in **Figure 4**. A considerable difference among animals has been found in number of ROH segments and the length of the genome covered. Animals with the same total ROH inbreeding ($F_{ROH>4\,Mb}$) might have a different number of ROH segments but with different lengths, which is a consequence of the different distances from the common ancestors.

Ne_{ROH} derived from change of inbreeding levels per generation (ΔF) is lowest when estimated from pedigree information and increases with restriction to longer ROH segments (see **Table 1**). The very high Ne_{ROH} ($Ne_{ROH>16\,Mb} = 370$) indicates strict avoidance of close inbreeding (like half sib, parent-offspring or first cousin mating) by the Tyrol Grey cattle breeders.

We have identified three regions with outlying ROH frequencies (4 Mb threshold) on chromosomes 5, 6, and 8 (**Figure 6**). Regions with increased ROH frequencies, the highest genomic autozygosity, are most likely consequences of selection as shown by Kim et al. (2013) and in computer simulations by Curik et al. (2002). The first region with the highest signal on BTA 8 starts at position 32,450,361 (BTB-00258020) and ends at position 46,041,080 (SNP BTA-28204-no-rs). Second region positioned on BTA 6 starts at position 36,277,967 (BTA-97637-no-rs) and ends at position 41,123,393 (SNP BTB-00406718). Finally, the region on BTA 5 starts at position 34,101,843 (BTB-01495784) and ends at position 42,918,584 (BTA-73464-no-rs). There are 147, 23, and 38 genes with known or unknown function within the signal regions on BTA 8, 6, and 5, respectively.

TABLE 1 | Levels of inbreeding (F) with lower and upper 95% confidence intervals (L95CI, U95CI), change of inbreeding per generation (ΔF) and inbreeding effective population size [Ne, with Ne = 1/(2ΔF)].

Statistic	ROH > 4 Mb	ROH > 8 Mb	ROH > 16 Mb	Pedigree
F	0.040	0.029	0.016	0.024
L95CI	0.036	0.025	0.014	0.021
U95CI	0.044	0.032	0.019	0.027
ΔF	0.004	0.003	0.001	0.005
Ne	125	186	370	102

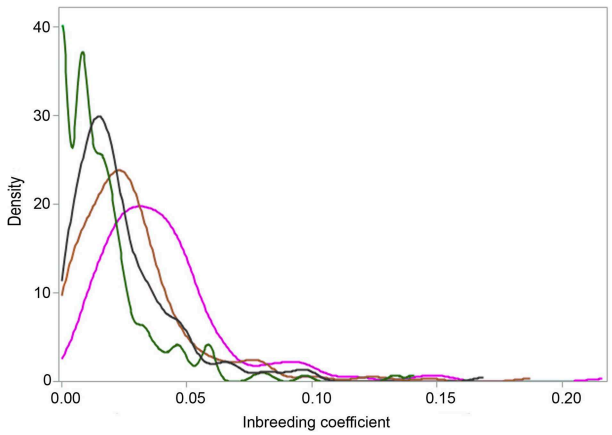


FIGURE 5 | Distributions of three ROH ($F_{ROH>4\,Mb}$ in magenta; $F_{ROH>8\,Mb}$ in brown and $F_{ROH>16\,Mb}$ in green) and one pedigree (F_{PED} in black) inbreeding coefficients for the Tyrol Grey cattle.

Genomic Selection

Genomic breeding values for Tyrol Grey bulls using major production and functional traits were computed. For the production traits the breeding values (EBV) and deregressed breeding values (dEBV) for milk yield, fat yield and content, protein yield and content were considered. For functional traits EBVs and dEBVs for longevity, persistency, maternal fertility, somatic cell count, and milking speed were included.

Single breed evaluations were used with only old bulls and young bulls born before 2003 as reference population. The validation animals consisted of bulls born after 2003. The number of animals in the validation set differed based on the trait, but in general they were between 36 and 49 when EBVs were used, and between 29 and 42 when dEBVs were used as response variable. The results are shown in **Figures 7A,B**. In general the average accuracies ranged from 0.13 to 0.91. Accuracy for the production traits in kg (fat kg, protein kg) was lower than using their direct counterpart measured in percent (fat% and protein%). This has largely been attributed to higher heritabilities for fat and protein content, compared to fat and protein production. For almost all functional traits however, the correlations were much

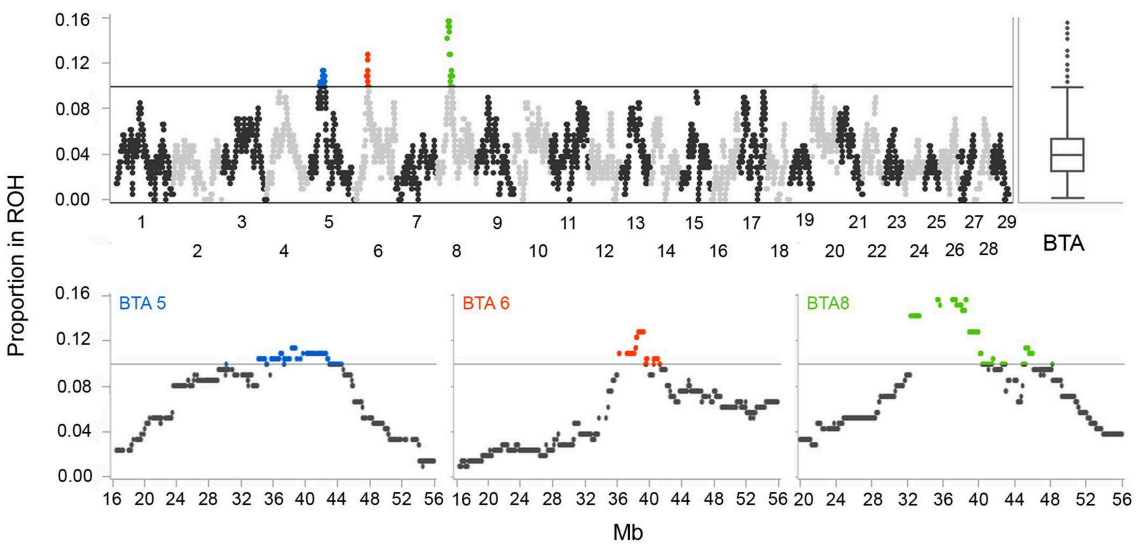


FIGURE 6 | Autozygosity islands, regions with extreme ROH frequency in Tyrol Grey cattle (minimum ROH length set to 4 Mb).

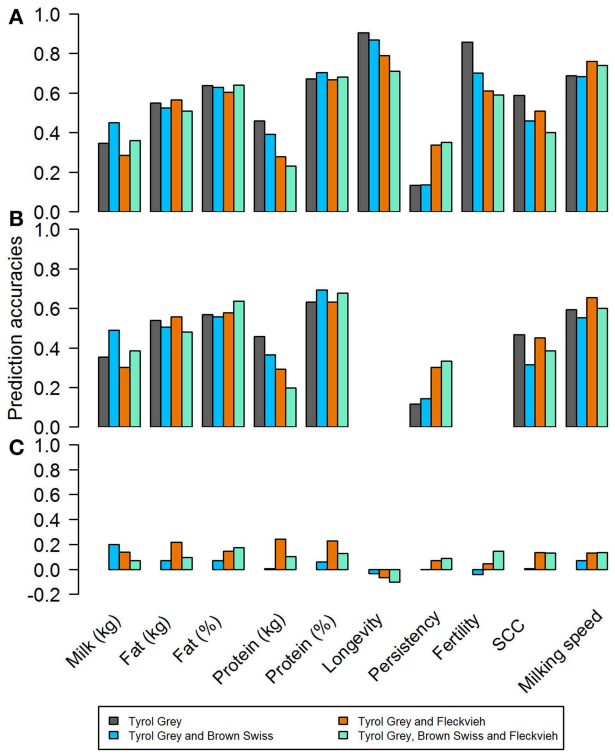


FIGURE 7 | Correlations between estimated breeding values and genomic breeding values based on (A) EBV for single and multi-breed reference sets; (B) dEBV for single and multi-breed reference sets; (C) EBV for across breed reference sets.

higher compared to any of the previously reported results in the literature. In fact on average they were even higher than that of the production traits. The follow up bootstrapping generated large standard errors for all traits.

In order to improve accuracies for production traits, a multi-breed approach was undertaken by adding genotypes to reference population from other breeds. Just like in the single breed scenario, validation individuals consisted of bulls born after 2003. In theory the increase in the size of reference population should increase the prediction accuracy. However, the gains and losses in accuracies varied considerably, depending on the trait, when either Brown Swiss, Fleckvieh or both populations) were added to the Tyrol Grey reference. In general, for all functional traits, adding other breeds resulted in lower accuracies, except for persistency.

When EBVs were used as response variable (**Figure 7A**) the impact of adding Fleckvieh into the reference set was favorable for persistency. For all other traits the results did not improve or were even worse with mixed reference sets. The benefits, if any, were not consistent across traits, showing a different pattern for each trait. The longevity and maternal fertility traits could not be evaluated due to lack of bulls with deregressed breeding values with reliability over 0.3 (**Figure 7B**).

An additional scenario to predict GEBV of Tyrol Grey bulls from another breed (Fleckvieh and Brown Swiss bulls) was studied. Unlike the multi-breed approach, an across breed scenario meant that, the reference population to estimate marker effect were either Fleckvieh or Brown Swiss bulls, while the validation set was the entire population of Tyrol Grey bulls. The correlations between the EBVs and estimated $GEBV_{EBV}$ were very low (**Figure 7C**). In general, the correlations were somewhat higher with Fleckvieh bulls in the reference set, but still remained below 0.25 in all cases. For longevity, predicting GEBVs from both estimates of marker effects from Brown Swiss and Fleckvieh resulted in negative accuracies. Moreover, the accuracies obtained in with this scenario were lower than that of the single breed or multi-breed approach.

Bootstrapping was used to assess the degree of confidence in the GEBV accuracies. It showed very wide confidence intervals for estimated GEBV for almost all traits (**Table 2**). Contrary to expectations, the confidence intervals for both longevity and fertility remained very high.

In addition to estimating the correlation between GEBVs and EBVs, we also computed the correlation between GEBVs and parent averages (**Table 3**). High correlations signify that the estimated GEBVs only predict the part of EBVs estimated as parent averages [0.5 (*EBVsire* + *EBVdam*)]. The correlations between the GEBVs and the parent averages for EBVs and dEBVs were very moderate to high. These correlations are the highest for longevity and fertility, indicating that the high GEBV accuracies were driven by parent averages. In other words, there is no advantage of GEBVs over parent averages for longevity and fertility, and relatively little advantage for other traits in the Tyrol Grey population.

Discussion

Genomic Analysis in Tyrol Grey Cattle

Traditionally the research interests in small and endangered populations are in genetic diversity parameters and breed conservation efforts. The justification is to describe breeding resources which might be important for coping with future needs and for facilitating the sustainable use of marginal areas (Toro et al., 2009). Microsatellites were a popular tool to describe genetic diversity (Baumung et al., 2004). In addition to microsatellites, SNP markers have been used to describe genetic diversity via parameters like allelic richness, heterozygosity/homozygosity levels (Makina et al., 2014), or LD and the associated Ne_{LD} (Hill, 1981; Hayes et al., 2003; Tenesa et al., 2007; Medugorac et al., 2009; Flury et al., 2010).

LD, measured as the correlation between alleles, is a fundamental concept in molecular genetics, while a large number of genomic methodologies are highly dependent on it (McKay et al., 2007; Pérez O'Brien et al., 2014). A typical LD pattern was observed in our study, with high LD for markers close to each other, quickly decaying with increasing inter-marker distance. Similar patterns were observed also in other studies (de Roos et al., 2008; Flury et al., 2010; Qanbari et al., 2010). In addition to the genome wide scale it is possible to utilize the LD information on the gene level. An example of this

TABLE 2 | Mean accuracies of GEBVs computed from EBVs and dEBVs from the single breed scenario and their 95% confidence intervals computed based on 10,000 bootstrap samples.

Traits	r(EBV,GEBV)			r(dEBV,GEBV)		
	Mean	Lower CI	Upper CI	Mean	Lower CI	Upper CI
Milk yield	0.345	0.027	0.675	0.354	0.050	0.668
Fat yield	0.549	0.349	0.755	0.539	0.335	0.755
Fat content	0.637	0.403	0.902	0.569	0.311	0.864
Protein yield	0.460	0.205	0.727	0.458	0.193	0.741
Protein content	0.673	0.493	0.869	0.632	0.445	0.830
Longevity	0.906	0.859	0.950	–	–	–
Persistency	0.133	−0.101	0.364	0.117	-0.139	0.373
Fertility	0.858	0.777	0.944	–	–	–
SCC	0.589	0.383	0.804	0.468	0.188	0.747
Milking speed	0.688	0.521	0.856	0.594	0.343	0.849

TABLE 3 | Average reliabilities of validation animals and correlations between parent averages based on EBV and GEBV/PA based on dEBV and GEBV.

Traits	Average EBV reliability	r(PA-EBV, GEBV-EBV)	r(PA-dEBV, GEBV-dEBV)
Milk yield	0.74	0.67	0.64
Fat yield	0.74	0.77	0.73
Fat content	0.74	0.79	0.71
Protein yield	0.74	0.73	0.69
Protein content	0.74	0.82	0.76
Longevity	0.26	0.81	NA
Persistency	0.70	0.39	0.35
Fertility	0.30	0.92	NA
SCC	0.58	0.75	0.62
Milking speed	0.57	0.88	0.86

approach was the description of the entire genetic variability of a meat tenderness gene with only 16 polymorphic SNPs and 18 haplotypes in three French cattle breeds (Marty et al., 2010).

LD can be used to calculate Ne_{LD} (Hill, 1981), even when the pedigree information is missing or it is incomplete. As the Ne size is sometimes used as a criterion to determine the endangerment status of a breed and thus always of interest. The Ne_{LD} relies on assumed impact of mutation and recombination distance (McEvoy et al., 2011), thus neglecting mutation rate or approximating the recombination distance to 1 Mb $\approx$ 1 cM leads to different outcomes (Corbin et al., 2012).

We note here that, calculation of Ne based on genomic data is deemed controversial. The Ne_{LD} was nearly the same in two Finnish pig populations when compared to pedigree data (Uimari and Tapio, 2011), much lower in a Swiss cattle breed (Flury et al., 2010) and strongly biased upwards in a Spanish pony population (Goyache et al., 2011). Simulation studies showed a downward bias for Ne_{LD} (Sved et al., 2013). As there are several theoretical conflicts in the estimation procedure, extreme caution is advised when calculating Ne_{LD} (Goyache et al., 2011; Corbin et al., 2012).

Several other methods were developed to overcome some of the limitations of Ne_{LD}. The most popular approaches are chromosome segment homozygosity (Hayes et al., 2003) and calculation of Ne based on inbreeding rate per generation calculated from ROH (MacLeod et al., 2013, 2009; Curik et al., 2014). Here we have presented the estimation of several Ne_{ROH} depending on the three ROH length thresholds. The method is direct extension of the estimation of Ne based on pedigree inbreeding coefficient and has not been evaluated empirically. While the values obtained for $Ne_{ROH>4\,Mb}$ are close to, although somewhat higher, Ne_{LD} and Ne_{PED} broader empirical evaluation of the method is required for its comprehensive understanding. We would like to point out that Ne_{ROH} and Ne_{LD} are two conceptually different estimates that can supplement and/or substitute Ne_{PED} estimates and provide useful information for the conservation management of a population in question. The Ne based on genomic information could be directly applied, for example in determining the risk status of breeds, as the current method used by FAO relies solely on number of male and female animals.

In the genomic selection era, SNP information is predominantly used for breeding value estimation. The popularity of genomic selection (Meuwissen et al., 2001; Hayes et al., 2009) resulted to the routine genotyping of young bulls in several large breeds (e.g., Holstein, Fleckvieh). These genotypes accumulate to an ever growing reference population which is subsequently re-used to estimate SNP effects to improve the accuracy of GEBV (Van Raden et al., 2011a). These large reference populations allow the genomic selection to be so successful (Misztal, 2011). Given the small population size of the Tyrol Grey cattle, and many other small breeds, the size of the reference population will not be high enough to meet standards of large breeds, especially in reference population size. In order to increase the reference population size other breeds are sometimes added to the breed of interest (multi-breed) or

used entirely alone (across breed) to calculate marker effects in genomic evaluations.

Estimated breeding values and deregressed breeding values for a range of traits were used to assess the feasibility of genomic selection is the Tyrol Grey breed. Surprisingly high accuracy of GEBV was obtained in the single breed evaluations (**Figure 7A**) and especially for longevity and fertility. Based on our criteria of discarding records with dEBV reliability below 0.30, GEBV were not estimated for longevity and fertility when using dEBVs as pseudo-phenotypes. The reason for the high accuracy of prediction for $GEBV_{EBV}$ was that, reliabilities of EBVs for young bulls were low. With low reliabilities, EBVs were similar to parent averages. This was affirmed by the high correlation between EBVs and parent averages (**Table 3**). Similarly large reliabilities were reported by Morota et al. (2014) when GEBV were correlated to low reliable EBVs.

In small breeding populations, the opportunities to obtain a sufficiently large number of daughters to generate highly reliable EBVs in a progeny testing scheme are limited. The problem is compounded especially for lowly heritable traits. For example, with trait heritability of 0.34 (milk yield), about 100 daughter records are need achieve reliabilities of 0.90. With heritability of 0.12 for longevity and 0.02 for fertility in our data set, about 300 and 1840 daughter records would be need. As shown in this study, predictive ability of a forward prediction scheme using young bulls as validation set was unusually high from PA driven EBVs. Lower reliabilities has been reported for the same traits with similar heritability in other large population breeds such as Fleckvieh (Ertl et al., 2014) or Holstein (Olson et al., 2012). The results from the study affirms the idea that, validation animals should have reliable EBVs if predictive ability is computed based on the correlation between GEBV and EBV. An alternative to this problem would be to use a single-step GBLUP approach (Legarra et al., 2014). Reliabilities would be computed based on the inverse of the diagonal element of the MME (Henderson, 1975). Reliabilities computed using the single step GBLUP approach could be compared to reliabilities of parent averages. Potential benefit of use of genomic information could be directly measured.

Multi-breed reference populations for genomic prediction are highly dependent on the LD and structure and genetic distance between breeds. The accuracy of genomic prediction could be substantially improved when the breeds are genetically very close or when animals of the same breed from multiple countries are pooled (Lund et al., 2014). Also using Bayesian variable selection instead of the BLUP approach could be beneficial in case of more distantly related breeds (Erbe et al., 2012; Bolormaa et al., 2013; Zhou et al., 2014). Lower accuracies in multi-breed genomic evaluation can be attributed to extent and differences in LD between markers and QTL (Goddard and Hayes, 2009), phase and allele substitution effects of QTLs (Spelman et al., 2002; Thaller et al., 2003).

In order to demonstrate the across breed genomic evaluation in Tyrol Grey cattle the German-Austrian Fleckvieh and Brown Swiss genotype pools were used in the evaluation. Using the large reference population composed of these two breeds to estimate GEBV was not successful. As shown in **Figure 7C**, the accuracies were very low for all traits. The accuracies have improved when part of the Tyrol Grey bulls were included into the reference set. This multi-breed reference however, did not have a clear advantage over the accuracies from the small breed reference set, similarly to Karoui et al. (2012).

A large population size is generally perceived as a requirement to estimate reliable GEBV, as we highlighted in the introduction of this paper. When the population is below of a critical mass the EBVs will be driven by parent averages, therefore genomic selection techniques will bring little new information into genetic evaluation of small breeds, as demonstrated in our paper. On the other hand, if reliable pedigrees are not available in a certain breed, i.e., no conventional breeding value estimation can be done, the breeding values estimated based on genomic data are a secure way to improve the breed.

Additional Uses of Genomic Data for Management of Small and Endangered Breeds

Even if genomic selection methods produce uncertain results in small breeds, there are a number of other reasons why a routine genotyping of the population would be beneficial. The identification of relatedness and inbreeding levels in the population has one of the biggest advantages from the practical perspective. The genomic relationship matrix can uncover family structures and infer relatedness within the population (Supplement Figure 1), even if the pedigree information is missing or it is incomplete (Calus et al., 2011). The correctness of the existing pedigrees can be verified comparing genomic information, e.g., by checking for Mendelian inconsistencies to identify incorrect parent-offspring relationships.

Similarly to the genomic relatedness it is possible to calculate the inbreeding coefficient. Compared to non-genomic approaches, here the knowledge of the pedigree is not needed, and so equally good results can be produced for animals, whose pedigree is dubious, incomplete, or entirely missing. Runs of homozygosity extend the analysis of relatedness between two individuals by identifying long homozygous segments, supposedly coming from the same ancestor somewhere in the past, inferring the individual's inbreeding coefficient. Based on the length of the segments it is possible to identify the number of generations to the common ancestor. In this study, the mean level of inbreeding estimated from three different ROH lengths ($Ne_{ROH>4\,Mb}$, $Ne_{ROH>8\,Mb}$, and $Ne_{ROH>16\,Mb}$) ranged from 0.016 to 0.029 and were comparable to other studies. In addition, there were only four individuals with outlying inbreeding coefficients higher than 0.125, indicating that potential risks could have been even more reduced with genomic information available. The utilization of genomic information to control inbreeding as well as to reduce early embryonic loss or appearance of congenital genetic defects due to recessive haplotypes in homozygous state (see more detailed discussion below) seems promising.

Crossbreeding is a very common strategy to increase the productivity of a breed or to introduce a desirable quality from another breed. The levels of crossbreeding are traditionally computed based on pedigree information. The pedigree approach assumes that the genetic composition of individuals with the

same type of ancestry information is equal. This assumption does not hold however, as recombinations alter the composition of ancestral chromosomes, resulting into different admixture levels (Bryc et al., 2010). The Girolando cattle for example were bred to achieve a 5/8 of Holstein and 3/8 of Gir cross. Based on the pedigree information the expected Holstein admixture level is 62.5%. The real admixture levels based on SNP data can vary as much as 49–85% (Orazietti et al., 2014, unpublished). The adaptability of breeds can be also increased by producing optimal composites for a specific region. For example, introducing the alleles that are responsible for the trypanotolerance in Baoule cattle into the genomes of the trypano susceptible zebu populations in Burkina Faso would be a great advantage (Soudré et al., 2013). In other small populations the crossbreeding with large commercial populations could be a concern due to the loss of purebred stock. In all cases SNP chip data provides reliable estimates of the admixture levels which facilitates the selection of the desirable genotypes for breeding purposes (Frkonja et al., 2012). Furthermore, the genomic information could be used to purge the foreign genome from a small population (Amador et al., 2014).

A less frequent, but a much more critical utilization of genomic data is detection of lethal or sub-lethal genotypes. The obvious case is when a disorder is found in a population and an attempt is made to discover its source and genetic background by *ad hoc* genotyping of affected individuals. A very good example for this *ad hoc* approach was the disorder similar to bovine progressive degenerative myeloncephalopathy (weaver syndrome) in the Tyrol Grey population. As the purebred population is small, the disorder would have had a devastating effect. The region with the causative mutation was identified combining homozygosity mapping (Charlier et al., 2008) and other genome wide association techniques in 14 affected and 27 control animals. More detailed analysis allowed pinpointing the causal mutation in the mitofusin (MFS2) gene. Routine genotyping of breeding animals identifies any carriers and will purge the population from this mutation within a short period (Drögemüller et al., 2011).

The previous case demonstrated an efficient an identification of causal variants for a known disorder. If the disorder itself or its symptoms were less obvious however, the detection of affected animals may be much more difficult. To detect these cases it is possible to screen the whole population genotype data. Alleles with relatively high heterozygote frequency in the population, but without the occurrence of both homozygotes indicate lethal variants. Eleven candidate haplotypes were detected using this technique in the North American Holstein, Jersey, and Brown Swiss population, some of them with confirmed phenotypic effects (Van Raden et al., 2011b). Similar technique was used to identify homozygote deficient haplotypes with potentially negative effects on fertility traits in Nordic Holstein (Sahana et al., 2013) and Jersey (Sonstegard et al., 2013). In most cases the frequency of carrier animals with harmful genomic regions in heterozygous state is relatively low, but it can also be surprisingly widespread. In Finnish Red cattle a region associated with embryonic death had a frequency of 32% in the population, due to its positive effect on milk yield (Kadri et al., 2014). In general, the genotype screening allows the detection of new disorders or to confirm the causative sites of known defects. These disorders and defects can be then avoided in subsequent generations by planned mating of carriers and non-carriers, or even eradication of certain disorders from the breed by restricted usage of carrier genotypes.

Conclusions

In a very short time, high-throughput molecular information has become a standard tool in animal breeding. Routine genotyping of the entire male population in small breeds is often not in place, although it would be feasible due to the small population size and extreme reduction in genotyping price. Our results suggest that genomic selection is not readily applicable in small breeds even with very large reference populations in a multi breed setting. There are numerous other utilizations of the genomic information however, that make routine genotyping not only beneficial but outright desirable for the management of small breeds. Apart from various genetic diversity measures, the identification of regions identical by descent instead of approximations according to the pedigree will help to better understand relatedness and inbreeding in the population. Furthermore, the pool of genotypes for the entire breed enables to continuously scan the population and allow a swift reaction in identifying carriers of lethal or potentially harmful haplotypes. The new information can be used to eliminate undesirable alleles through the mating process. Similarly, the breed proportions due to admixture could be estimated with the goal to fix a desirable ratio or to preserve the purity of the breed.

While our paper describes an example from the Tyrol Grey population, we would like to stress that the recommendations are valid for all small and endangered breeds. The genotyping of SNP markers is a mature and well understood technology, with uses that can complement, improve or even replace approaches for breed management. Therefore, we suggest the continuous collection of genotypes and their use in breed monitoring and improvement.

Acknowledgments

In memory of Otto Hausegger (1964–2014), managing director of the Tiroler Grauviehzuchtverband. The authors would like to thank the Tiroler Grauviehzuchtverband for the cooperation in the genotyping of the Tyrol Grey population. We are also grateful to Förderverein Biotechnologieforschung, Rinderbesammungsgenossenschaft Memmingen, Gesellschaft zur Förderung der Fleckviezucht in Niederbayern, Nutzvieh GmbH Miesbach, Rinderunion Baden-Württemberg eG, Zentrale Arbeitsgemeinschaft Österreichischer Rinderzüchter, Arbeitsgemeinschaft Süddeutscher Rinderzucht- und Besamungsorganisationen for providing the Fleckvieh and Brown Swiss genotype data and ZuchtData EDV-Dienstleistungen GmbH for providing the phenotype information. The Vienna Scientific Cluster is acknowledged

for providing computing resources for part of the analyses. The authors are grateful for the financial support of the Austrian Ministry for Transport, Innovation and Technology (BMVIT) and the Austrian Science Fund (FWF) via the project TRP46-B19.

Supplementary Material

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fgene.2015.00173/abstract

References

Amador, C., Hayes, B. J., and Daetwyler, H. D. (2014). Genomic selection for recovery of original genetic background from hybrids of endangered and common breeds. *Evol. Appl.* 7, 227–237. doi: 10.1111/eva.12113

Arias, J. A., Keehan, M., Fisher, P., Coppieters, W., and Spelman, R. (2009). A high density linkage map of the bovine genome. *BMC Genet*. 10:18. doi: 10.1186/1471-2156-10-18

Baumung, R., Simianer, H., and Hoffmann, I. (2004). Genetic diversity studies in farm animals – a survey. *J. Anim. Breed. Genet*. 121, 361–373. doi: 10.1111/j.1439-0388.2004.00479.x

Bolormaa, S., Pryce, J. E., Kemper, K., Savin, K., Hayes, B. J., Barendse, W., et al. (2013). Accuracy of prediction of genomic breeding values for residual feed intake and carcass and meat quality traits in Bos taurus, Bos indicus, and composite beef cattle. *J. Anim. Sci*. 91, 3088–3104. doi: 10.2527/jas.2012-5827

Bryc, K., Velez, C., Karafet, T., Moreno-Estrada, A., Reynolds, A., Auton, A., et al. (2010). Genome-wide patterns of population structure and admixture among Hispanic/Latino populations. *Proc. Natl. Acad. Sci. U.S.A.* 107, 8954–8961. doi: 10.1073/pnas.0914618107

Calus, M. P. L., Mulder, H. A., and Bastiaansen, J. W. M. (2011). Identification of Mendelian inconsistencies between SNP and pedigree information of sibs. *Genet. Sel. Evol.* 43:34. doi: 10.1186/1297-9686-43-34

Charlier, C., Coppieters, W., Rollin, F., Desmecht, D., Agerholm, J. S., Cambisano, N., et al. (2008). Highly effective SNP-based association mapping and management of recessive defects in livestock. *Nat. Genet*. 40, 449–454. doi: 10.1038/ng.96

Corbin, L. J., Liu, A. Y. H., Bishop, S. C., and Woolliams, J. A. (2012). Estimation of historical effective population size using linkage disequilibria with marker data. *J. Anim. Breed. Genet*. 129, 257–270. doi: 10.1111/j.1439-0388.2012.01003.x

Curik, I., Ferenčaković, M., and Sölkner, J. (2014). Inbreeding and runs of homozygosity: a possible solution to an old problem. *Livest. Sci. Genomics Appl. Livest. Prod.* 166, 26–34. doi: 10.1016/j.livsci.2014.05.034

Curik, I., Sölkner, J., and Stipic, N. (2002). Effects of models with finite loci, selection, dominance, epistasis and linkage on inbreeding coefficients based on pedigree and genotypic information. *J. Anim. Breed. Genet*. 119, 101–115. doi: 10.1046/j.1439-0388.2002.00329.x

Drögemüller, C., Reichart, U., Seuberlich, T., Oevermann, A., Baumgartner, M., Kühni Boghenbor, K., et al. (2011). An unusual splice defect in the Mitofusin 2 gene (MFN2) is associated with degenerative axonopathy in tyrolean grey cattle. *PLoS ONE* 6:e18931. doi: 10.1371/journal.pone.0018931

de Roos, A. P. W., Hayes, B. J., Spelman, R. J., and Goddard, M. E. (2008). Linkage disequilibrium and persistence of phase in Holstein–Friesian, Jersey and Angus Cattle. *Genetics* 179, 1503–1512. doi: 10.1534/genetics.107.084301

Erbe, M., Hayes, B. J., Matukumalli, L. K., Goswami, S., Bowman, P. J., Reich, C. M., et al. (2012). Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. *J. Dairy Sci*. 95, 4114–4129. doi: 10.3168/jds.2011-5019

Ertl, J., Edel, C., Emmerling, R., Pausch, H., Fries, R., and Götz, K.-U. (2014). On the limited increase in validation reliability using high-density genotypes in genomic best linear unbiased prediction: observations from Fleckvieh cattle. *J. Dairy Sci*. 97, 487–496. doi: 10.3168/jds.2013-6855

Ferenčaković, M., Sölkner, J., and Curik, I. (2013). Estimating autozygosity from high-throughput information: effects of SNP density and genotyping errors. *Genet. Sel. Evol*. 45:42. doi: 10.1186/1297-9686-45-42

Flury, C., Tapio, M., Sonstegard, T., Drögemüller, C., Leeb, T., Simianer, H., et al. (2010). Effective population size of an indigenous Swiss cattle breed estimated from linkage disequilibrium. *J. Anim. Breed. Genet*. 127, 339–347. doi: 10.1111/j.1439-0388.2010.00862.x

Frkonja, A., Gredler, B., Schnyder, U., Curik, I., and Sölkner, J. (2012). Prediction of breed composition in an admixed cattle population. *Anim. Genet*. 43, 696–703. doi: 10.1111/j.1365-2052.2012.02345.x

Garrick, D. J., Taylor, J. F., and Fernando, R. L. (2009). Deregressing estimated breeding values and weighting information for genomic regression analyses. *Genet. Sel. Evol*. 41:55. doi: 10.1186/1297-9686-41-55

Goddard, M. E., and Hayes, B. J. (2009). Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nat. Rev. Genet*. 10, 381–391. doi: 10.1038/nrg2575

Goyache, F., Álvarez, I., Fernández, I., Pérez-Pardal, L., Royo, L. J., and Lorenzo, L. (2011). Usefulness of molecular-based methods for estimating effective population size in livestock assessed using data from the endangered black-coated Asturcón pony. *J. Anim. Sci*. 89, 1251–1259. doi: 10.2527/jas.2010-3620

Groeneveld, L. F., Lenstra, J. A., Eding, H., Toro, M. A., Scherf, B., Pilling, D., et al. (2010). Genetic diversity in farm animals – a review. *Anim. Genet*. 41, 6–31. doi: 10.1111/j.1365-2052.2010.02038.x

Gutiérrez, J. P., and Goyache, F. (2005). A note on ENDOG: a computer program for analysing pedigree information. *J. Anim. Breed. Genet.* 122, 172–176. doi: 10.1111/j.1439-0388.2005.00512.x

Hayes, B. J., Bowman, P. J., Chamberlain, A. J., and Goddard, M. E. (2009). Invited review: genomic selection in dairy cattle: progress and challenges. *J. Dairy Sci*. 92, 433–443. doi: 10.3168/jds.2008-1646

Hayes, B. J., Visscher, P. M., McPartlan, H. C., and Goddard, M. E. (2003). Novel multilocus measure of linkage disequilibrium to estimate past effective population size. *Genome Res*. 13, 635–643. doi: 10.1101/gr.387103

Henderson, C. R. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics* 31, 423–447. doi: 10.2307/2529430

Hill, W. G. (1981). Estimation of effective population size from data on linkage disequilibrium. *Genet. Res*. 38, 209–216. doi: 10.1017/S0016672300020553

Howrigan, D. P., Simonson, M. A., and Keller, M. C. (2011). Detecting autozygosity through runs of homozygosity: a comparison of three autozygosity detection algorithms. *BMC Genomics* 12:460. doi: 10.1186/1471-2164-12-460

Hutchison, J. L., Cole, J. B., and Bickhart, D. M. (2014). Short communication: use of young bulls in the United States. *J. Dairy Sci*. 97, 3213–3220. doi: 10.3168/jds.2013-7525

Kadri, N. K., Sahana, G., Charlier, C., Iso-Touru, T., Guldbrandtsen, B., Karim, L., et al. (2014). A 660-Kb deletion with antagonistic effects on fertility and milk production segregates at high frequency in Nordic Red cattle: additional evidence for the common occurrence of balancing selection in livestock. *PLoS Genet*. 10:e1004049. doi: 10.1371/journal.pgen.1004049

Karoui, S., Carabaño, M. J., Díaz, C., and Legarra, A. (2012). Joint genomic evaluation of French dairy cattle breeds using multiple-trait models. *Genet. Sel. Evol*. 44:39. doi: 10.1186/1297-9686-44-39

Kim, E.-S., Cole, J. B., Huson, H., Wiggans, G. R., Van Tassell, C. P., Crooker, B. A., et al. (2013). Effect of artificial selection on runs of homozygosity in U.S. Holstein Cattle. *PLoS ONE* 8:e80813. doi: 10.1371/journal.pone.0080813

Legarra, A., Christensen, O. F., Aguilar, I., and Misztal, I. (2014). Single Step, a general approach for genomic selection. *Livest. Sci. Genomics Appl. Livest. Prod.* 166, 54–65. doi: 10.1016/j.livsci.2014.04.029

Lund, M. S., Su, G., Janss, L., Guldbrandtsen, B., and Brøndum, R. F. (2014). Genomic evaluation of cattle in a multi-breed context. *Livest. Sci. Genomics Appl. Livest. Prod.* 166, 101–110. doi: 10.1016/j.livsci.2014.05.008

MacLeod, I. M., Larkin, D. M., Lewin, H. A., Hayes, B. J., and Goddard, M. E. (2013). Inferring demography from runs of homozygosity in whole-genome sequence, with correction for sequence errors. *Mol. Biol. Evol*. 30, 2209–2223. doi: 10.1093/molbev/mst125

MacLeod, I. M., Meuwissen, T. H. E., Hayes, B. J., and Goddard, M. E. (2009). A novel predictor of multilocus haplotype homozygosity: comparison

with existing predictors. *Genet. Res.* 91, 413–426. doi: 10.1017/S0016672309990358

Makina, S. O., Muchadeyi, F. C., van Marle-Koster, E., MacNeil, M. D., and Maiwashe, A. (2014). Genetic diversity and population structure among six cattle breeds in South Africa using a whole genome SNP panel. *Front Genet.* 5:333. doi: 10.3389/fgene.2014.00333

Marty, A., Amigues, Y., Servin, B., Renand, G., Levéziel, H., and Rocha, D. (2010). Genetic variability and linkage disequilibrium patterns in the bovine DNAJA1 gene. *Mol. Biotechnol.* 44, 190–197. doi: 10.1007/s12033-009-9228-y

McEvoy, B. P., Powell, J. E., Goddard, M. E., and Visscher, P. M. (2011). Human population dispersal "Out of Africa" estimated from linkage disequilibrium and allele frequencies of SNPs. *Genome Res.* 21, 821–829. doi: 10.1101/gr.119636.110

McKay, S. D., Schnabel, R. D., Murdoch, B. M., Matukumalli, L. K., Aerts, J., Coppieters, W., et al. (2007). Whole genome linkage disequilibrium maps in cattle. *BMC Genet.* 8:74. doi: 10.1186/1471-2156-8-74

McQuillan, R., Leutenegger, A.-L., Abdel-Rahman, R., Franklin, C. S., Pericic, M., Barac-Lauc, L., et al. (2008). Runs of homozygosity in European populations. *Am. J. Hum. Genet.* 83, 359–372. doi: 10.1016/j.ajhg.2008.08.007

Medugorac, I., Medugorac, A., Russ, I., Veit-Kensch, C. E., Taberlet, P., Luntz, B., et al. (2009). Genetic diversity of European cattle breeds highlights the conservation value of traditional unselected breeds with high effective population size. *Mol. Ecol.* 18, 3394–3410. doi: 10.1111/j.1365-294X.2009.04286.x

Meuwissen, T. H. E., Hayes, B. J., and Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819–1829.

Misztal, I. (2011). FAQ for genomic selection. *J. Anim. Breed. Genet.* 128, 245–246. doi: 10.1111/j.1439-0388.2011.00944.x

Morota, G., Boddhireddy, P., Vukasinovic, N., Gianola, D., and De Nise, S. (2014). Kernel-based variance component estimation and whole-genome prediction of pre-corrected phenotypes and progeny tests for dairy cow health traits. *Front Genet.* 5:56. doi: 10.3389/fgene.2014.00056

Olson, K. M., Van Raden, P. M., and Tooker, M. E. (2012). Multibreed genomic evaluations using purebred Holsteins, Jerseys, and Brown Swiss. *J. Dairy Sci.* 95, 5378–5383. doi: 10.3168/jds.2011-5006

Pérez O'Brien, A. M., Mészáros, G., Utsunomiya, Y. T., Sonstegard, T. S., Garcia, J. F., Van Tassell, C. P., et al. (2014). Linkage disequilibrium levels in Bos indicus and Bos taurus cattle using medium and high density SNP chip data and different minor allele frequency distributions. *Livest. Sci. Genomics Appl. Livest. Prod.* 166, 121–132. doi: 10.1016/j.livsci.2014.05.007

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795

Qanbari, S., Pimentel, E. C. G., Tetens, J., Thaller, G., Lichtner, P., Sharifi, A. R., et al. (2010). The pattern of linkage disequilibrium in German Holstein cattle. *Anim. Genet.* 41, 346–356. doi: 10.1111/j.1365-2052.2009.02011.x

Sahana, G., Nielsen, U. S., Aamand, G. P., Lund, M. S., and Guldbrandtsen, B. (2013). Novel harmful recessive haplotypes identified for fertility traits in nordic holstein cattle. *PLoS ONE* 8:e82909. doi: 10.1371/journal.pone.0082909

Sonstegard, T. S., Cole, J. B., Van Raden, P. M., Van Tassell, C. P., Null, D. J., Schroeder, S. G., et al. (2013). Identification of a nonsense mutation in CWC15 associated with decreased reproductive efficiency in Jersey cattle. *PLoS ONE* 8:e54872. doi: 10.1371/journal.pone.0054872

Soudré, A., Ouédraogo-Koné, S., Wurzinger, M., Müller, S., Hanotte, O., Ouédraogo, A. G., et al. (2013). Trypanosomosis: a priority disease in tsetse-challenged areas of Burkina Faso. *Trop. Anim. Health Prod.* 45, 497–503. doi: 10.1007/s11250-012-0248-4

Spelman, R. J., Ford, C. A., McElhinney, P., Gregory, G. C., and Snell, R. G. (2002). Characterization of the DGAT1 gene in the New Zealand dairy population. *J. Dairy Sci.* 85, 3514–3517. doi: 10.3168/jds.S0022-0302(02)74440-8

Sved, J. A., Cameron, E. C., and Gilchrist, A. S. (2013). Estimating effective population size from linkage disequilibrium between unlinked loci: theory and application to fruit fly outbreak populations. *PLoS ONE* 8:e69078. doi: 10.1371/journal.pone.0069078

Tenesa, A., Navarro, P., Hayes, B. J., Duffy, D. L., Clarke, G. M., Goddard, M. E., et al. (2007). Recent human effective population size estimated from linkage disequilibrium. *Genome Res.* 17, 520–526. doi: 10.1101/gr.6023607

Thaller, G., Krämer, W., Winter, A., Kaupe, B., Erhardt, G., and Fries, R. (2003). Effects of DGAT1 variants on milk production traits in German cattle breeds. *J. Anim. Sci.* 81, 1911–1918.

Toro, M. A., Fernández, J., and Caballero, A. (2009). Molecular characterization of breeds and its use in conservation. *Livest. Sci. Special Issue: Anim. Genet. Resour.* 120, 174–195. doi: 10.1016/j.livsci.2008.07.003

Uimari, P., and Tapio, M. (2011). Extent of linkage disequilibrium and effective population size in Finnish Landrace and Finnish Yorkshire pig breeds. *J. Anim. Sci.* 89, 609–614. doi: 10.2527/jas.2010-3249

Van Raden, P. M., O'Connell, J. R., Wiggans, G. R., and Weigel, K. A. (2011a). Genomic evaluations with many more genotypes. *Genet. Sel. Evol.* 43:10. doi: 10.1186/1297-9686-43-10

Van Raden, P. M., Olson, K. M., Null, D. J., and Hutchison, J. L. (2011b). Harmful recessive effects on fertility detected by absence of homozygous haplotypes. *J. Dairy Sci.* 94, 6153–6161. doi: 10.3168/jds.2011-4624

Zhou, L., Heringstad, B., Su, G., Guldbrandtsen, B., Meuwissen, T. H. E., Svendsen, M., et al. (2014). Genomic predictions based on a joint reference population for the Nordic Red cattle breeds. *J. Dairy Sci.* 97, 4485–4496. doi: 10.3168/jds.2013-7580

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Mészáros, Boison, Pérez O'Brien, Ferenčaković, Curik, Da Silva, Utsunomiya, Garcia and Sölkner. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in GENETICS

ORIGINAL RESEARCH ARTICLE
published: 17 February 2015
doi: 10.3389/fgene.2015.00044

Characterization of genetic diversity and gene mapping in two Swedish local chicken breeds

Anna M. Johansson[1]* and Ronald M. Nelson[2]

[1] *Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala, Sweden*
[2] *Department of Clinical Sciences, Swedish University of Agricultural Sciences, Uppsala, Sweden*

Edited by:
Juha Kantanen, Natural Resources Institute Finland, Finland

Reviewed by:
Menghua Li, Chinese Academy of Sciences, China
Mervi Sisko Honkatukia, MTT Agrifood Research Finland, Finland

****Correspondence:***
Anna M. Johansson, Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Ulls väg 26, PO Box 7053, SE-75007 Uppsala, Sweden
e-mail: anna.johansson@slu.se

The aim of this paper is to study genetic diversity in the two Swedish local chicken breeds Bohuslän-Dals svarthöna and Hedemorahöna. The now living birds of both of these breeds (about 500 for Bohuslän-Dals svarthöna and 2600 for Hedemorahöna) originate from small relicts of earlier larger populations. An additional aim was to make an attempt to map loci associated with a trait that are segregating in both these breeds. The 60k SNP chip was used to genotype 12 Bohuslän-Dals svarthöna and 22 Hedemorahöna. The mean inbreeding coefficient was considerably larger in the samples from Hedemorahöna than in the samples from Bohuslän-Dals svarthöna. Also the proportion of homozygous SNPs in individuals was larger in Hedemorahöna. In contrast, on the breed level, the number of segregating SNPs were much larger in Hedemorahöna than in Bohuslän-Dals svarthöna. A multidimensional scaling plot shows that the two breeds form clusters well-separated from each other. Both these breeds segregate for the dermal hyperpigmentation phenotype. In Bohuslän-Dals svarthöna most animals have dark skin, but some individuals with lighter skin exists (most easily detected by their red comb). An earlier study of the Fm locus showed that this breed has the same complex rearrangement involving the *EDN3* gene as Silkie chicken and two other studied Asian breeds. In the breed Hedemorahöna, most individuals have normal skin pigmentation (and red comb), but there are some birds with darker skin and dark comb. In this study the involvement of the *EDN3* gene is confirmed also in Hedemorahöna. In addition we identify a region on chromosome 21 that is significantly associated with the trait.

Keywords: chicken, SNP, comb color, genetic diversity, dermal hyperpimentation, EDN3

INTRODUCTION

There are 11 Swedish local chicken breeds. The conservation of the Swedish local chicken breeds is coordinated by the Swedish association for local poultry (Svenska Lanthönsklubben). The Swedish local chicken breeds originate from different parts of Sweden and only a small number of birds remained when they were rescued from extinction. An interesting aspect is that, although they have gone through severe bottle-necks, several of the breeds still retain much phenotypic diversity. This makes these breeds promising to use for mapping of the segregating traits. The genetic diversity is probably limited which will make the loci involved in the segregating traits easier to identify. Sequencing of the mtDNA D-loop in nine of these breeds showed limited diversity but revealed multiple maternal origins in several of the breeds (Englund et al., 2015). Due to the limited mtDNA diversity, it is necessary to analyze autosomal markers in these breeds, in order to study the relatedness between breeds. Analysis of a set of dense autosomal markers would also have the added value of giving the possibility to do genetic mapping of interesting traits in these breeds.

In this paper we chose to study the breeds Bohuslän-Dals svarthöna and Hedemorahöna. We chose these breeds since both breeds segregate for the dermal hyperpigmentation phenotype. We have chosen the comb color as phenotype to study, since this is an easy phenotype to accurately collect during the short visit to the chicken owners. Bohuslän-Dals svarthöna originates from the northern part of Bohuslän in western Sweden (close to the Norwegian border). Around 1899, a woman got the ancestors of the current population of this breed as a wedding gift. Her two sons inherited the flock and in 1958 a man got the birds from the two brothers (Olsson, 2004). These black birds is said to have be common in this area in old times and there is a legend stating that they are the descendants of birds that sailors brought from a foreign country. The current population size is approximately 500. Most individuals have black feathers but also black beaks, legs, skin, and dark colored meat. Some individuals with lighter skin exists (most easily detected by their red comb). The black pigmentation was recently shown to be due to the same complex rearrangement involving the *Endothelin 3* (*EDN3*) locus as Silkie chicken and two other studied Asian breeds (Dorshorst et al., 2011). The Hedemorahöna originates from the area close to Hedemora in the Dalarna county. They were rediscovered in the 1980s when a woman saw a flock that reminded her of the chickens her parent had during her childhood (Olsson, 2004). The population size is approximately 2600. Hedemorahöna have a wide range of plumage colors. Most individuals have normal skin

pigmentation (and red comb), but some birds have been observed to have darker skin and dark comb.

What we have seen from our collected data, is that the comb color can be categorized into three categories: red, semi-dark, and dark. This has also been noted in several generations by the owners of these birds. It is not a continuous scale and it is easy to type each bird into one of these categories. Males have all three comb colors. However, females only have red or dark combs. This points to the involvement of the Z chromosome in the trait. Additional observations from a bird owner, who is interested in genetics, show that the inheritance of comb color can mostly be explained as a Z-linked inheritance, but that there are exceptions that do not fit with a single Z-linked locus (Thomas Englund, Pers. Commun.). The *Fm* and *id* loci are known to be involved in dark pigmentation phenotype in the chicken breed Silkie (Bateson and Punnett, 1911; Dorshorst et al., 2010). The *Fm* locus is located on chromosome 20 and the causative mutation have been shown to be a complex rearrangement involving the *EDN3* gene (Dorshorst et al., 2011). The *id* locus is located at the Z chromosome. A region have been identified (Dorshorst et al., 2010; Siwek et al., 2013), but the gene have not yet been identified. A recent study on dermal shank pigmentation identified three significant SNPs in a 0.7 Mb region on (Li et al., 2014). Looking in detail at regions reported to contain these loci was therefore an obvious choice, although the Z chromosome was difficult to analyze correctly. A complicating factor is that both males and females of Hedemorahöna with light plumage color (such as wheaten, Columbian, salmon) rarely have very dark comb, i.e., even if they from the comb color in their offspring seem to have only the "dark allele" at the Z chromosome, their comb is not as dark as the birds with darker plumage colors but instead look more semi-dark. If they have completely white plumage color, they always have red comb, even if they have parents with dark comb (Thomas Englund, Pers. Commun.). This points to an interaction between the id locus and loci involved in plumage color.

Table 1 | Description of the samples and result on inbreeding and homozygosity in the two breeds.

	Hedemorahöna	Bohuslän-Dals svarthöna
Number of samples	22	12
Number of sampled flocks	3	2
Mean inbreeding coefficient (*F*)	0.17	-0.09
Mean number of homozygous SNPs in an individual	24,743	14,525
Comb color among samples	nine dark; two semi-dark; one red	two dark; 19 red; one unknown

Table 2 | Number of fixed and segregating SNPs on chromosomes 1–28 in the two breeds, both in the whole dataset and in 100 replicates of taking a subset of 12 Hedemorahöna at random to compare with the 12 Bohuslän-Dals svarthöna.

	Whole dataset (22 Hedemorahöna + 12 Bohuslän-Dals svarthöna)	Mean of 100 replicates of 12 random Hedemorahöna and all 12 Bohuslän-Dals svarthöna
Genotypes present in both lines	50,617	50,615.27
Fixed for different alleles	767	970.14
Fixed for the same allele	10,326	10,965.43
Hedemorahöna fixed, Bohuslän-Dals svarthöna segregating	5583	6333.84
Bohuslän-Dals svarthöna fixed, Hedemorahöna segregating	16,166	15,322.71

The aim of this paper is to study genetic diversity in the two Swedish local chicken breeds Bohuslän-Dals svarthöna and Hedemorahöna. The now living birds in both of these breeds originate from small relicts of earlier larger populations. Here we map loci associated with the comb color trait that are segregating in within these breeds. To do this we genotyped 12 Bohuslän-Dals svarthöna and 22 Hedemorahöna with the 60k SNP chip produced by Illumina for the GWMAS Consortium (Groenen et al., 2011). An association study was performed on the genotypes given the recorded comb color phenotype.

MATERIAL AND METHODS

SAMPLING AND GENOTYPING

The blood samples were collected by visits to the flocks by a veterinary student with special training for taking blood samples from chicken, or taken by a local veterinarian and sent to us. The blood sample was taken from the wing vein with a small needle and mixed with EDTA in an Eppendorf tube. Ethical permission (number C247/6) for the collection of blood samples was obtained from the Uppsala ethical board for animal research (name in Swedish: Uppsala djurförsöksetiska nämnd) prior to the collection of samples.

In this study 12 samples (three males and nine females) from Bohuslän-Dals svarthöna and 22 samples (six males and 16 females) from Hedemorahöna were included (Supplementary Table 1). The samples from Bohuslän-Dals svarthöna came from two different flocks and the samples from Hedemorahöna came from three different flocks.

For genotyping the 60k SNP chip, produced by Illumina for the GWMAS Consortium, was used (Groenen et al., 2011). The genotyping was done by the company DNA LandMarks in Canada. All positions given in this paper are from the SNP chip data. It is based on the genome assembly Gallus_gallus-2.1.

DIVERSITY ANALYSIS

The calculations of homozygosity of individuals and inbreeding were done with the function *het* in the program PLINK (Purcell et al., 2007, URL: http://pngu.mgh.harvard.edu/purcell/

plink/) and were based on the autosomal markers (chromosomes 1–28). The analyses of fixation in the breeds were done by custom R scripts (available from the authors upon request).

SNP QUALITY CONTROL

Genome-wide analysis was done with GenAbel version 1.8-0 (Aulchenko et al., 2007). The data was filtered in GenAbel by iteratively excluding: markers with a call rate below 90%; markers with a minor allele frequency below 5%; markers that are not in Hardy-Weinberg equilibrium ($p < 1 \times 10^{-8}$) and individuals with a call rate below 85%. Additionally individuals with high autosomal heterozygosity (FDR <1%) and individuals with a pairwise Identity By State (IBS) value above 0.9 were removed. In total 34,955 markers (out of 53,312) and 31 individuals (out of 34) passed all the criteria.

GWAS

A genome wide association was performed using the filtered markers and the comb color phenotype using the *qtscore* function in GenAbel. The comb color phenotype, measured for each individual, was encoded as 1, 2, and 3 (red, semi-dark, and dark, respectively). For each position on the genome an association analysis of the SNP genotype with the comb color phenotype was

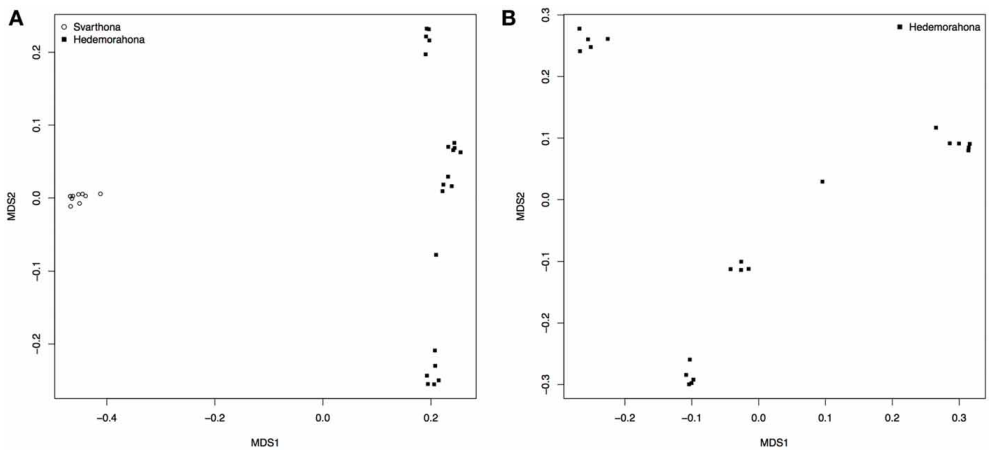


FIGURE 1 | Population stratification. (A) MDS analysis on population indicating stratification by breed. **(B)** MDS on Hedemorahöna to estimate substructuring within this breed.

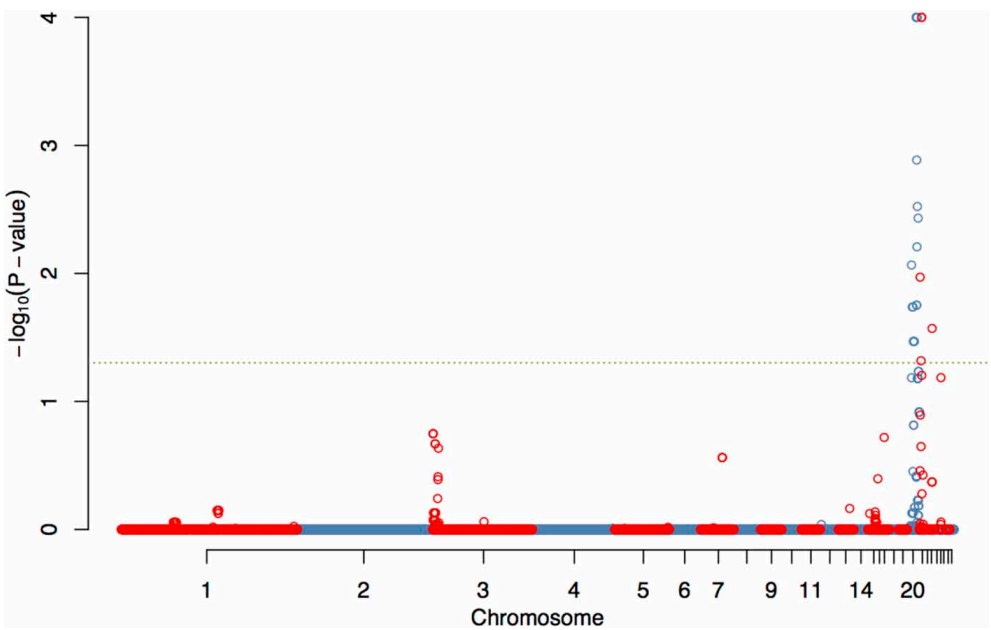


FIGURE 2 | Genome wide association of comb color phenotype. Manhattan plot showing FDR-adjusted *p*-values for genome-wide associations of comb color, corrected for kinship.

performed on all the individuals, as implemented in the *qtscore* function. Simultaneously population structure and relatedness in the data were corrected for. We calculated the genomic kinship matrix prior to the association analysis. The covariates, resulting from the relatedness, was subsequently calculated using the *polygenic* function and included by implementing the GRAMMAS method, in the *qtscore* function. The association of the comb color phenotype at each SNP included (and was thus corrected for) kinship and population effects. The analysis was permuted 10,000 times to obtain a FDR corrected, empirical genome wide significance.

To estimate the level of population stratification we performed a multidimensional scaling (MDS) analysis using the *cmdscale* function in R. This function was applied to the distances computed from the kinship matrix. We performed this analysis on the full dataset (i.e., both populations), as well as on the Hedemorahöna separately to check for substructure in this population.

RESULTS

DIVERSITY AND INBREEDING IN THE TWO BREEDS

The mean inbreeding coefficient (F) was considerably larger in the samples from Hedemorahöna than in the samples from Bohuslän-Dals svarthöna (**Table 1**). Most Bohuslän-Dals svarthöna individuals were not inbred, i.e., F were less than or close to 0 (Supplementary Table 1). The negative inbreeding coefficients show that these birds are more heterozygous than expected under Hardy–Weinberg equilibrium. However, half the Hedemorahöna individuals had $F > 0.1$ and most of these had $F > 0.2$ (Supplementary Table 1). The most inbred individual of Hedemorahöna had $F = 0.45$. The inbreeding varied between the three sampled flocks of Hedemorahöna with one flock having considerable higher inbreeding (mean $F = 0.31$) than the other two (mean $F = 0.09$ and 0.13, respectively). In the Bohuslän-Dals svarthöna, the difference between the flocks was smaller.

There was a very large difference in homozygosity between the breeds. The number of homozygous SNPs in an individual was larger in Hedemorahöna (mean 24,743) than in Bohuslän-Dals svarthöna (mean 14,528). The most homozygous sample from Bohuslän-Dals svarthöna was much less homozygous than the least homozygous sample from Hedemorahöna.

In contrast to the large number of homozygous SNPs in each individual, our Hedemora samples have considerable more segregating SNPs within the breed than Bohuslän-Dals svarthöna (**Table 2**). In Hedemora about two thirds of the SNPs are segregating, whereas in Bohuslän-Dals svarthöna less than half of the SNPs are segregating. Twenty percent of the SNPs in the two breeds are fixed for the same allele.

A multidimensional scaling plot shows that the two breeds form separate clusters well-separated from each other (**Figure 1A**). All the samples from Bohuslän-Dals svarthöna were close together, whereas the samples from Hedemorahöna

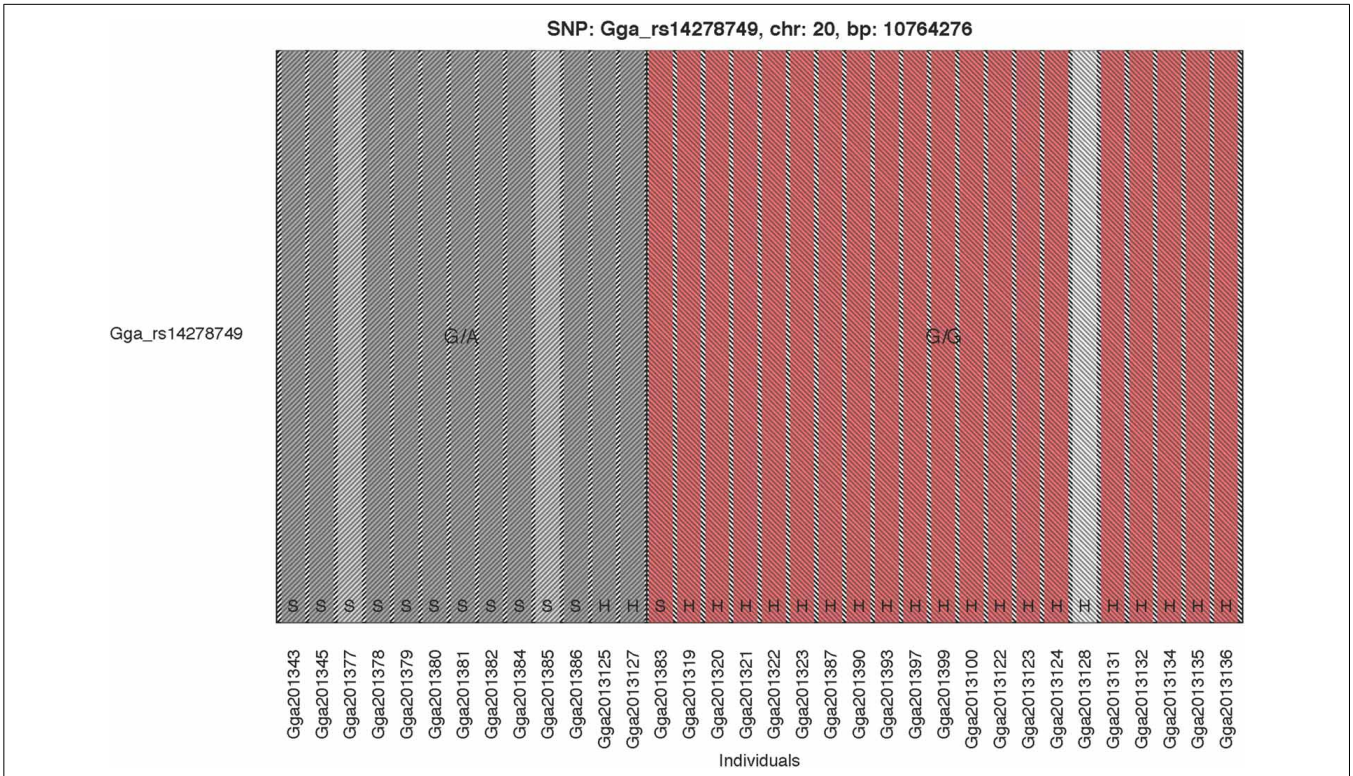


FIGURE 3 | Genotype Phenotype Map. The genotype phenotype map for the SNP Gga_rs14278749 (position 10,764,276 on chromosome 20). Bar colors indicate comb color of individuals, with the white indicating the individual with unknown phenotype. The two shaded areas indicate the two observed genotypes, respectively. The breeds are indicated at the bottom of the bars with H and S for Hedemorahöna and Bohuslän-Dals Svarthöna, respectively.

was more spread out and showed some population structure (**Figures 1A,B**). Since the MDS-plots showed that there was clear differentiation between the two breeds (**Figure 1A**), we corrected for population structure in the association analyses.

GWAS ON AUTOSOMAL MARKERS IN OUR DATA

The GWAS on the two populations clearly identify regions on chromosomes 20 and 21 as associated with comb color (**Figure 2**, Supplementary Table 2). The associated markers on chromosome

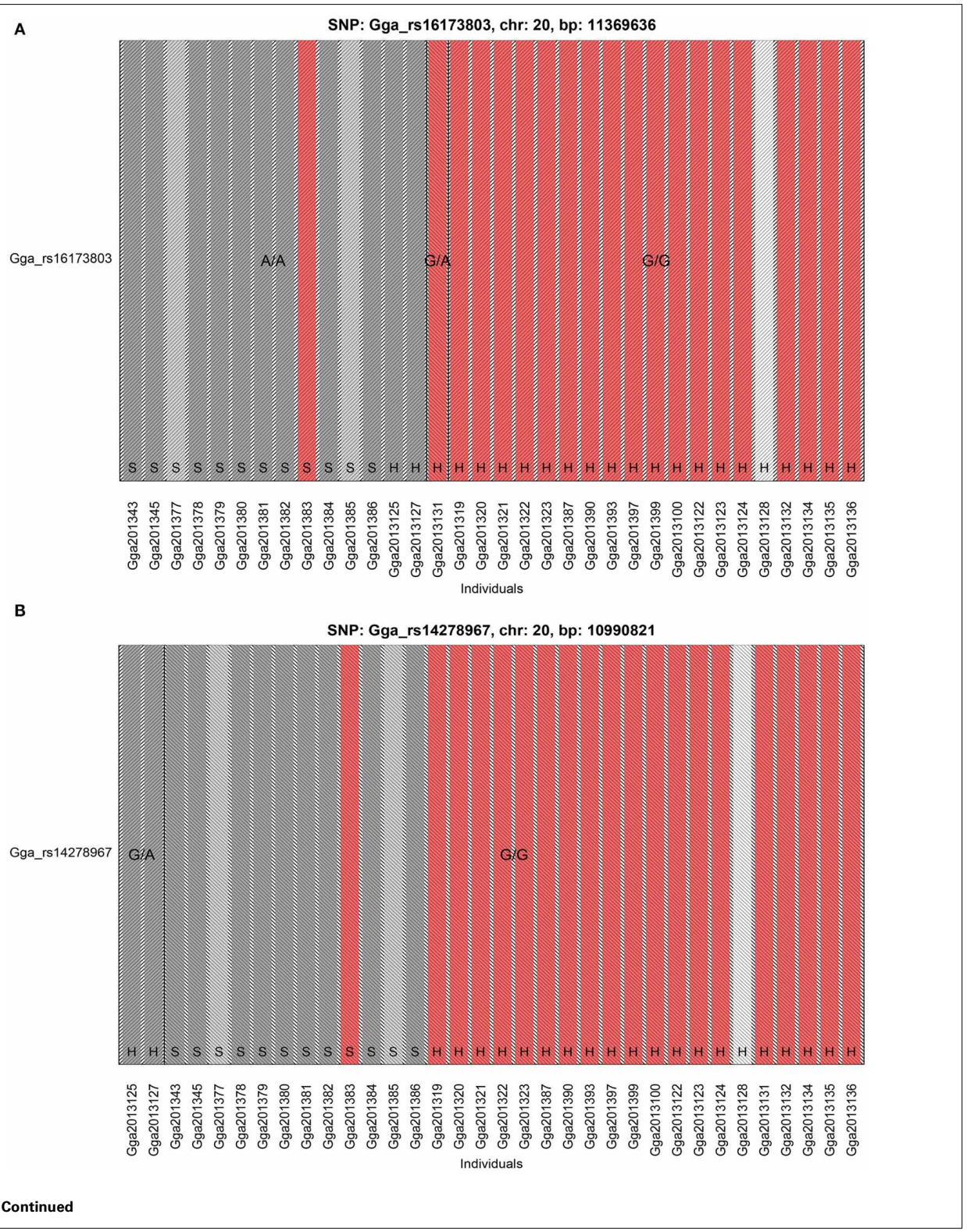
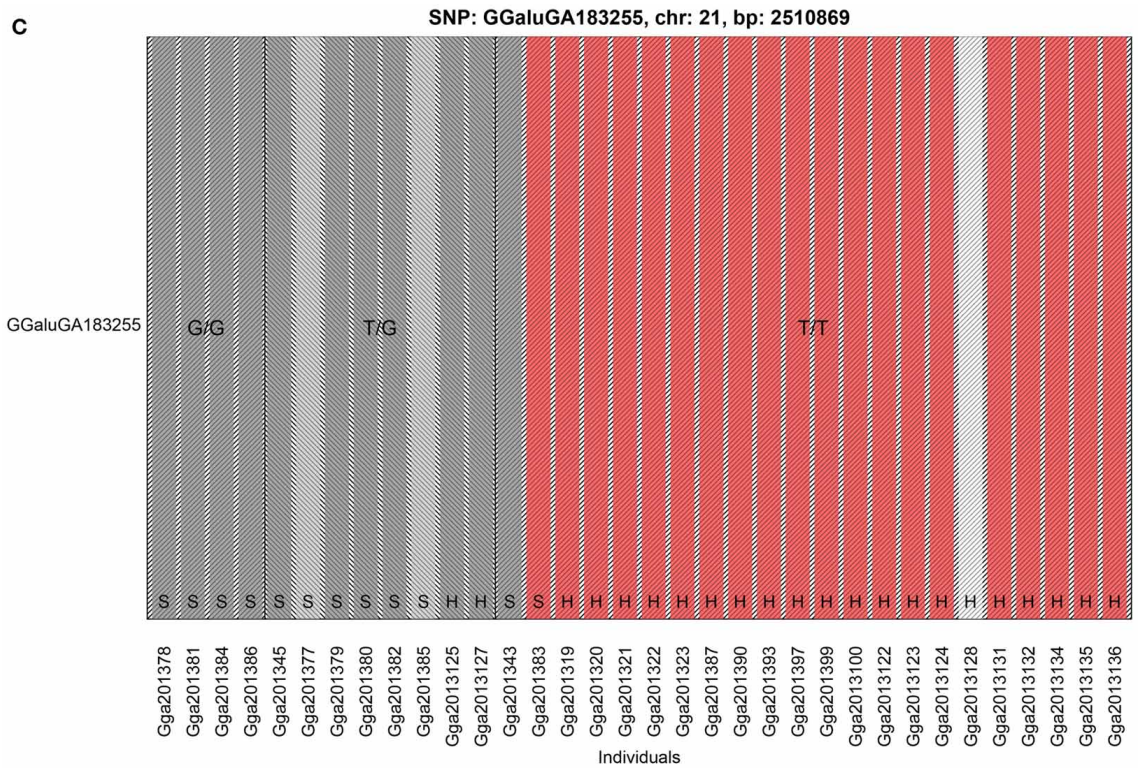


FIGURE 4 | Continued



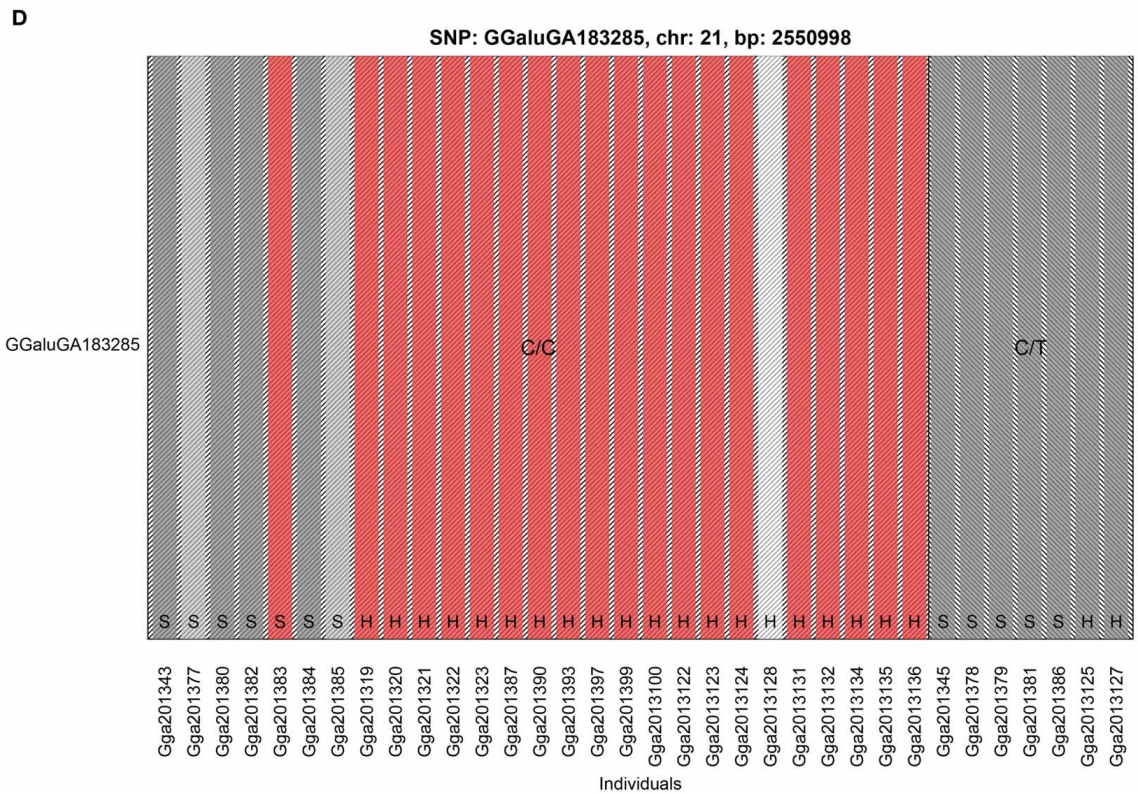


FIGURE 4 | Genotype Phenotype Maps. The genotype phenotype maps for the SNPs **(A)** Gga_rs16173803 and, **(B)** Gga_rs14278967 on chromosome 20 as well as, **(C)** GGaluGA183255 and, **(D)** GGaluGA183285 on chromosome 21. Bar colors indicate comb color of individuals with, the white indicating the individual with unknown phenotype. Note that the two SNPs on chromosome 20 shown in this figure was not included in the GWAS, because they failed the quality control check. The reasons are that Gga_rs14278967 has a minor allele frequency of less than 5% while Gga_rs16173803 is not in Hardy-Weinberg-equilibrium (at *p*-level of 1e-08).

20 are located between positions 10,194,286 and 12,375,897. This region encompasses the *EDN3* locus previously reported to be associated with dermal hyperpigmentation by Dorshorst et al. (2011). The SNP Gga_rs14278749 at position 10,764,276 on chromosome 20 was most significant (FDR $p = 9.9 \times 10^{-5}$). The genotype-phenotype map clearly shows that this SNP completely explain the red vs. dark comb color (**Figure 3**). Note that this SNP is located in the first duplicated region identified by Dorshorst et al. (2011). The interpretation of our genotype data should therefore be that the "heterozygotes" do have the duplication and that they have one allele in one copy of the segment and the other allele in the other copy of the segment. Since the duplication is segregating, the amount of heterozygotes was not high enough to make the SNP fail the quality control. On chromosome 21 the SNP GGaluGA183285 (FDR $p = 9.9 \times 10^{-5}$) and the SNP GGaluGA183255 (FDR $p = 1.0 \times 10^{-04}$) was most significant, these are located close to each other (position 2,550,998 and 2,510,869, respectively, Supplementary Table 2). Genotype–phenotype maps for these two SNPs on chromosome 21 can be seen in **Figures 4C,D**.

RESULTS ON SNPs ON CHROMOSOME 20 SHOWING ASSOCIATION WITH PIGMENTATION IN OTHER STUDIES

In the study by Dorshorst et al. (2011), the Silkie and Ayam Cemani samples were all heterozygous for five SNPs in the first duplicated region (positions 10,717,294–10,846,232) on chromosome 20. Among our 12 samples from Bohuslän-Dals svarthöna, all except one (which was the one with red comb) where heterozygous for two of these SNPs and two additional SNPs in this region. The two individuals from Hedemorahöna with dark comb were heterozygous for three of the five SNPs identified in Silkie and Ayam Cemani. The 11 samples from Bohuslän-Dals svarthöna with dark or semi-dark comb were also all heterozygous for three SNPs in the second duplicated region.

The SNP Gga_rs16173803 in the second duplicated region shows the same homozygous genotype in all the 12 samples from Bohuslän-Dals svarthöna and the two samples from Hedemorahöna with dark comb, one other Hedemora is heterozygous and all the rest are homozygous for the other allele (**Figure 4A**).

The SNP Gga_rs14278967 at position 10,990,821 at chromosome 20 is associated with Fm phenotype in Siwek et al. (2013). In our study, the two Hedemora with dark comb are heterozygous for this SNP, all other samples are homozygous for the same allele (**Figure 4B**). Note that also all genotyped individuals from Bohuslän-Dals svarthöna were homozygous for the major allele.

Z CHROMOSOME

On chromosome Z, six markers in the region from position 107,457 to position 465,329 segregate perfectly together with the comb color phenotype in the breed Bohuslän-Dals svarthöna. The hen with the red comb have one allele while all the individuals with dark comb have the other allele. The two males with semi-dark comb are heterozygotes. This is not the same region on chromosome Z that has been shown to be associated with the id locus in other breeds previously (Dorshorst et al., 2010; Siwek et al., 2013). The SNP Gga_rs14685750 at position 72,985,598 at chromosome Z is associated with the id+/id+ genotype at Silkie and Green-legged Partridgelike fowl in Siwek et al. (2013). At this SNP, there is no variation in our samples from the Hedemorahöna and Bohuslän-Dals svarthöna. One adjacent SNP is fixed for one allele in Bohuslän-Dals svarthöna and another allele in Hedemorahöna, but many other such SNPs occur elsewhere on the Z chromosome. In Bohuslän-Dals svarthöna, several SNPs in this region have one allele in the red combed female and two dark combed females and the other allele in the other dark combed birds. The two semi dark males are heterozygous. In Hedemorahöna there is no variation in the SNPs closest to position 72,985,598. There are 21 SNPs showing a somewhat interesting pattern in this breed earlier on the chromosome (position 3,622,633–7,440,378). Here the two individuals with dark comb, the one with unknown comb color and one with red color have one allele and the rest of the individuals with red comb have the other allele. Interestingly, for these SNPs, the Bohuslän-Dals svarthöna is fixed for the dark comb allele.

DISCUSSION

Our results showed that the diversity, with respect to the breed, is larger in Hedemorahöna than in Bohuslän-Dals svarthöna. However, the individuals within Hedemorahöna have less diversity due to inbreeding.

A possibility is that the larger number of segregating SNPs in Hedemorahöna might explained by the fact that we had more samples from Hedemorahöna than from Bohuslän-Dals svarthöna. We therefore also analyzed the number of segregating SNPs in random subsets of 12 samples of Hedemorahöna. We replicated the random sampling 100 times and calculated the average number of fixed and segregating SNPs in the breeds over the replicates. The result was very similar to the results from the whole dataset (**Table 2**), strengthening the conclusion that there are much more segregating SNPs in Hedemorahöna.

As a comparison, Bohuslän-Dals svarthöna has about the same amount of segregating SNPs as the Low Line after 40 generations in the body weight selection experiment in Virginia (Johansson et al., 2010). The sample size of the Low line from the body weight selection experiment (20) is similar to our Hedemorahöna sample size (22).

Our GWAS and genotype–phenotype map confirmed that the region with the EDN3 gene (Fm) is associated with dark pigmentation in both breeds. The variation at a single SNP in this region completely explain if an individual has dark or red comb. However, the dark vs. semi-dark comb cannot be explained by this region. We also identified a region on chromosome 21 that is significantly associated with comb color. A possibility is that the region we identified on chromosome 21 together with id on the Z chromosome can explain if an individual have dark or semi-dark comb. However, our limited sample size with very few semi-dark individuals did not allow us to further explore this. With the model we used to correct for kinship in the GWAS it was not possible to analyze the Z chromosome. We also analyzed the data including the Z chromosome with a more simple model (data not shown), but did not find any significant results on the Z chromosome. We did find markers with interesting segregation patterns on the Z chromosome, although there were different markers in

the two breeds. At the associated SNP in Siwek et al. (2013), there is no variation in any of our two breeds. The small sample size or lack of SNP variation might be the reason that we did not get any significant results on the Z chromosome, or our interesting patterns where some markers segregated with comb color in a breed might be just due to chance.

CONCLUSION

In conclusion, we showed that the genetic diversity on the level of an individual is lower in our samples from Hedemorahöna than in Bohuslän-Dals svarthöna. In contrast, there is more diversity within the breed in Hedemorahöna than in Bohuslän-Dals svarthöna. The high inbreeding detected in many of our samples from Hedemorahöna is a cause for concern. When owners of this breed are getting a new breeding animal to their flock, they should aim to get an individual with low relatedness to the existing birds in the flock. We have shown here, using comb color as an example, that it is possible to map loci for traits segregating using a limited number of individuals, in these breeds. Regarding the comb color, we have shown that one locus on chromosome 20 (*EDN3*) determines if an individual will have red or dark comb, and that at least one other locus is probably involved in determining if a dark individual get a dark or semi dark comb.

ACKNOWLEDGMENTS

We thank the Swedish association for local poultry (Svenska Lanthönsklubben) for providing contact information to chicken owners and information about the breeds. We thank Thomas Englund for information about inheritance of comb color in Hedemorahöna. AMJ received funding for of the collection of blood samples and genotyping from KSLA and Erik Philip-Sörensen foundation.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://www.frontiersin.org/journal/10.3389/fgene.2015.00044/abstract

REFERENCES

Aulchenko, Y. S., Ripke, S., Isaacs, A., and van Duijn, C. M. (2007). GenABEL: an R library for genome-wide association analysis. *Bioinformatics* 23, 1294–1296. doi: 10.1093/bioinformatics/btm108

Bateson, W., and Punnett, R. (1911). The inheritance of the peculiar pigmentation of the silky fowl. *J. Genet.* 1, 185–203. doi: 10.1007/BF02981551

Dorshorst, B., Molin, A.-M., Rubin, C.-J., Johansson, A. M., Strömstedt, L., Pham, M.-H., et al. (2011). A complex genomic rearrangement involving the *Endothelin 3* locus causes dermal hyperpigmentation in the chicken. *PLoS Genet.* 7:e1002412. doi: 10.1371/journal.pgen.1002412

Dorshorst, B., Okimoto, R., and Ashwell, C. (2010). Genomic regions associated with dermal hyperpigmentation, polydactyly and other morphological traits in the silkie chicken. *J. Hered.* 101, 339–350. doi: 10.1093/jhered/esp120

Englund, T., Strömstedt, L., and Johansson, A. M. (2015). Relatedness and diversity of nine Swedish local chicken breeds as indicated by the mtDNA D-loop. *Hereditas* 151, 229–233. doi: 10.1093/jhered/esp120

Groenen, M. A., Megens, H. J., Zare, Y., Warren, W. C., Hillier, L. W., Crooijmans, R. P., et al. (2011). The development and characterization of a 60K SNP chip for chicken. *BMC Genomics* 12:274. doi: 10.1186/1471-2164-12-274

Johansson, A. M., Pettersson, M. E., Siegel, P. B., and Carlborg, Ö. (2010). Genome-wide effects of long-term divergent selection. *PLoS Genet.* 6:e1001188. doi: 10.1371/journal.pgen.1001188

Li, G., Li, D., Yang, N., Qu, L., Hou, L., Zheng, J., et al. (2014). A genome-wide association study identifies novel single nucleotide polymorphisms associated with dermal shank pigmentation in chickens. *Poult. Sci.* 93, 2983–2987. doi: 10.3382/ps.2014-04164

Olsson, R. (2004). *Genbanker och Rasramar för Våra Svenska Lantraser av Fjäderfän*. Strängnäs: Svenska lanthönsklubben.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., et al. (2007). PLINK: a toolset for whole-genome association and population-based linkage analysis. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795

Siwek, M., Wragg, D., Sławińska, A., Malek, M., Hanotte, O., and Mwacharo, M. (2013). Insights into the genetic history of green-legged Partridgelike fowl: mtDNA and genome-wide SNP analysis. *Anim. Genet.* 44, 522–532. doi: 10.1111/age.12046

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 31 October 2014; accepted: 30 January 2015; published online: 17 February 2015.

*Citation: Johansson AM and Nelson RM (2015) Characterization of genetic diversity and gene mapping in two Swedish local chicken breeds. Front. Genet. **6**:44. doi: 10.3389/fgene.2015.00044*

This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics.

Copyright © 2015 Johansson and Nelson. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in
GENETICS

ORIGINAL RESEARCH ARTICLE
published: 27 February 2015
doi: 10.3389/fgene.2015.00062

Morphological and genetic characterization of an emerging Azorean horse breed: the Terceira Pony

Maria S. Lopes, Duarte Mendonça, Horst Rojer, Verónica Cabral, Sílvia X. Bettencourt and Artur da Câmara Machado*

Biotechnology Centre of Azores, Department of Agriculture, University of Azores, Angra do Heroísmo, Portugal

Edited by:
Johann Sölkner, University of Natural Resources and Life Sciences Vienna, Austria

Reviewed by:
Jeff Silverstein, United States Department of Agriculture – Agricultural Research Service, USA
Michelle Martinez-Montemayor, Universidad Central del Caribe, USA

***Correspondence:**
Artur da Câmara Machado, Biotechnology Centre of Azores, Department of Agriculture, University of Azores, Rua Capitão João D'Ávila, 9700-042 Angra do Heroísmo, Portugal
e-mail: amachado@uac.pt

The Terceira Pony is a horse indigenous to Terceira Island in the Azores. These horses were very important during the colonization of the island. Due to their very balanced proportions and correct gaits, and with an average withers height of 1.28 m, the Terceira Pony is often confused with a miniature pure-bred Lusitano. This population was officially recognized as the fourth Portuguese equine breed by the national authorities in January, 2014. The aim of this study was to analyze the morphology and the genetic diversity by means of microsatellite markers of this emerging horse breed. The biometric data consisted of 28 body measurements and nine angles from 30 animals (11 sires, 19 dams). The Terceira Pony is now a recognized horse breed and is gaining in popularity amongst breeders and the younger riding classes. The information obtained from this study will be very useful for conservation and management purposes, including maximizing the breed's genetic diversity, and solidifying the desirable phenotypic traits.

Keywords: Azores, equine, morphotype, SSR, genetic diversity, management, conservation

INTRODUCTION

Horses have played an important role in the development of Portugal throughout the centuries. Nowadays there are four officially recognized Portuguese autochthonous horse breeds – Lusitano, Sorraia, Garrano, and Terceira Pony. The Terceira Pony is autochthonous to the Azorean archipelago and was recognized by the national authorities on January 27, 2014. These animals are believed to have descended from horses first brought from mainland Portugal to the islands during the 15th century and were selected for size and adaptation to the local conditions. These extremely hardy and well adapted horses were very important during the colonization for transportation of goods and people, for agriculture and, if necessary, for meat and milk (Luciano da Silva, 1971). Although they contributed substantially to the development of the islands, due to new agricultural practices and the introduction of horses from other origins, their importance gradually declined and therefore the sustainability of the Terceira Pony now depends on a shift toward new market needs.

The Terceira Pony has very correct and balanced proportions with physical and personality characteristics that conform to the image of a modern riding horse, making it popular for riding by children. The existing population of the Terceira Pony comprises about 100 animals living on Terceira island, many of them descending from 14 founders (six sires and eight dams) while a few horses are dispersed on the other islands, mainly Faial, Graciosa, S. Jorge, and S. Miguel. In order to preserve the existing animals and increase their number, the Association of Breeders and Friends of the Terceira Pony was established in 2010.

In general for a horse to be accepted to the studbook it has to conform to specific phenotypic characteristics. Body conformation is used in distinguishing, evaluating, and comparing breeds (Lawrence, 2001). The relationships among body dimensions also affect the horses' beauty and performance in sports (Evans, 2000; Lawrence, 2001; Parker, 2002; Sadek et al., 2006). As the Terceira Pony has been recently recognized as a breed and the stud-book is in the process of being established, horses from other breeds can therefore still be used as founders provided that they fit the body conformation, gaits, and breeding goals set forth by the Association of Breeders and Friends of the Terceira Pony.

Nowadays, due to the reduced number of animals the Terceira Pony is evolving from deliberate crosses of specific sires and dams, which can lead to a loss of genetic diversity. Such a breeding program has to be carefully monitored to ensure that the decrease in variability does not adversely affect the beneficial characteristics of this breed. Microsatellites are accepted as the most suitable molecular markers to investigate breed genetic diversity (Takezaki and Nei, 1996) and differentiation, to estimate the genetic structure of populations (Pritchard et al., 2000) and comprise an attractive potential source of information about population histories and evolutionary processes (Aberle and Distl, 2004). The evaluation of genetic variability is one of the first steps in the process of species genetic conservation aimed at preserving both genetic variability and population viability (Bömcke et al., 2011).

The objective of the present study was to gather data to assist in management improvements and conservation planning by accessing the body conformation of the Terceira Pony by measuring

30 fully grown animals so that external breeders can match the required phenotype, estimate morphometric indices from these measures, assess biometric data and evaluate functional classification standards, and by analyzing 52 animals with 15 polymorphic microsatellite loci, that will enable quantifying the genetic variability within the breed.

MATERIALS AND METHODS

BIOMETRIC STUDY

For the biometric study of the Terceira Pony only fully grown animals, 11 sires and 19 dams, were measured. A total of 37 measurements (**Table 1**) were taken from the left side of the horse while standing on a horizontal surface in standard position, except for measurements of the head, with zoometric stick, and tape. Angles were measured with a protractor from the Animal Measuring System developed by ISOMED. Additionally the approximated live weight (W) was estimated from body measurement data as defined by Santos (1981).

Based on these 37 measurements, 19 indexes (**Table 2**) were calculated to evaluate the proportions of the animals and to define its type.

GENETIC STUDY

A total of 52 animals, corresponding to the 14 founders (six sires and eight dams) and their descendants were genotyped in this study. Fresh blood was collected for DNA extraction, by a veterinarian and following good veterinary practices, by jugular venipuncture and placed in sterile tubes with EDTA. Tubes were kept in a refrigerator box for transport to the laboratory. DNA was extracted following a modification of the Miller et al. (1988) salting out protocol described elsewhere (Lopes, 2011).

Genotyping was performed using the 15 autosomal microsatellite markers included in the Food and Agriculture Organization/International Society of Animal Genetics Measurement of Domestic Animal Diversity panel (Hoffmann et al., 2004) [AHT4, AHT5 (Binns et al., 1995), ASB2, ASB17, ASB23 (Breen et al., 1997), HMS2, HMS3, HMS6, HMS7 (Guérin et al., 1994), HTG4, HTG6 (Ellegren et al., 1992), HTG7, HTG10 (Marklund et al., 1994), LEX33 (Coogle et al., 1996), and VHL20 (van Haeringen et al., 1994)]. Amplification was carried out in volumes of 20 μl containing 20 ng total DNA, 160 μM of each dNTP, 2–10 pmol of each primer and 1 U Taq DNA polymerase (Fermentas) in reaction buffer containing $MgCl_2$. All forward primers were fluorescence labeled at the 5′ end with FAM, TET, HEX, NED, VIC, or PET, to allow multiplexing and simultaneous separation of the amplified products. Reactions were performed in a UNO II Biometra thermocycler with a first cycle of 5 min denaturation at 96°C, 40 s annealing and 40 s elongation at 72 C, followed by 35 cycles of 40 s denaturation, 40 s annealing, 40 s elongation and a final cycle of 40 s denaturation, 40 s annealing and 30 min elongation to maximize Taq DNA polymerases' ability to catalyze non-templated nucleotide addition (Smith et al., 1995), thus minimizing the potential for genotyping error attributable to the modified "plus-A" product. The PCR products were size fractionated by capillary electrophoresis using an automated sequencer (ABI PRISM 310 Genetic Analyzer,

Table 1 | Mean and SD of body measurements and approximate live weight of dams and sires of the Terceira Pony.

Measurements		Dams	Sires
Heights (cm)	Withers[a]	127.11 ± 2.90	129.70 ± 2.36
	Sub-sternal flank[b]	72.40 ± 2.30	71.40 ± 0.89
	Back[a]	122.44 ± 3.51	123.00 ± 2.29
	Rump[a]	127.44 ± 3.52	129.40 ± 2.99
Depth (cm)	Chest[a]	54.67 ± 2.54	58.30 ± 2.66
Lengths (cm)	Head[a]	50.11 ± 3.22	49.10 ± 1.02
	Neck[a]	68.00 ± 4.39	66.50 ± 3.20
	Body[a]	127.33 ± 1.87	128.00 ± 0.71
	Barrel[a]	60.44 ± 4.45	55.50 ± 3.91
	Rearquarters[a]	38.33 ± 6.00	42.30 ± 3.56
	Shoulder[a]	50.61 ± 3.59	53.40 ± 3.13
	Humerous[a]	24.17 ± 1.41	23.70 ± 1.48
	Radius[a]	32.50 ± 1.56	33.90 ± 1.34
	Metacarpus[a]	19.61 ± 0.89	20.60 ± 1.29
	Fore phalanx[a]	10.67 ± 0.50	10.20 ± 0.27
	Small trunk[c]	89.00 ± 6.44	85.70 ± 3.60
	Pelvis[a]	27.67 ± 1.30	25.70 ± 3.38
	Femur[a]	30.94 ± 1.88	31.50 ± 1.41
	Tibia[a]	38.06 ± 1.81	40.50 ± 1.41
	Metatarsus[a]	26.33 ± 1.50	29.00 ± 1.87
	Hind phalanx[a]	10.67 ± 0.61	10.60 ± 0.42
Widths (cm)	Skull[d]	17.17 ± 1.06	16.40 ± 0.55
	Chest[a]	28.89 ± 2.22	31.00 ± 1.58
	Hips[a]	37.89 ± 1.90	38.90 ± 2.30
	Thurls[a]	35.78 ± 1.99	37.90 ± 2.36
Circumferences (cm)	Chest[a]	145.11 ± 5.23	149.20 ± 4.55
	Forelimb cannon bone[a]	15.72 ± 0083	16.10 ± 0.55
	Hindlimb cannon bone[a]	17.44 ± 0.81	17.60 ± 0.42
Angles (°)	Shoulder[a]	55.78 ± 4.41	46.80 ± 2.77
	Shoulder joint[a]	42.22 ± 4.94	41.60 ± 4.34
	Fore fetlock joint[a]	49.56 ± 3.13	53.20 ± 2.28
	Fore hoof wall[a]	48.89 ± 4.91	51.20 ± 2.28
	Croup[a]	25.44 ± 4.03	20.60 ± 1.95
	Femur[a]	52.89 ± 6.79	52.60 ± 5.73
	Hock joint[a]	56.00 ± 5.66	51.00 ± 9.17
	Hind fetlock joint[a]	52.11 ± 3.44	53.80 ± 1.48
	Hind hoof wall[a]	52.11 ± 3.44	53.40 ± 2.41
Approximate live weight (W) (kg)		244.45 ± 26.42	265.70 ± 24.22

[a] *Following the description by Zechner et al. (2001)*
[b] *Following the description by McManus et al. (2005)*
[c] *Following the description by Komosa and Purzyc (2009)*
[d] *Following the description by Solé et al. (2013).*

Table 2 | Conformation indices determined based on the measurements taken for dams and sires of the Terceira Pony.

Conformation indices	Dams	Sires
Body Index (BI)[a]	0.877	0.857
Dactilo-Thoracic Index (DTI)[a]	0.108	0.108
Chest Index (CI)[a]	<0	<0
Body Racio (BR)[a]	0.99	1
Quadratic Index (QI)[b]	99.83	101.32
Compact Index1 (CPI1)[a]	1.92	2.05
Compact Index 2 (CPI2)[a]	9.02	8.95
Index of Arm Trunk[c]	22.65	23.83
Index of Scapula[c]	38.82	41.17
Index of Arm (humerous)[c]	19.02	18.27
Index of Forearm (radius)[c]	25.57	26.14
Index of Metacarpus Length[c]	15.43	15.88
Index of Trunk (greater)[c]	100	98.69
Index of Trunk (smaller)[c]	70.23	66.08
Index of Femur[c]	24.28	24.34
Index of Croup[c]	30.1	33.05
Index of Chest[c]	28.82	31.3
Index of Metatarsus Lenght[c]	20.66	22.41
Index of Metacarpus Circumference[c]	12.37	12.41

[a] *Following the description by Martin-Rosset (1983)*
[b] *Following the description by Druml et al. (2008)*
[c] *Following the description by Komosa and Purzyc (2009).*

PE Applied Biosystems) and fragment lengths were determined with the help of internal size standards (GeneScan 350 TAMRA Size Standard and GeneScan 500 LIZ Size Standard, PE Applied Biosystems).

Genetic variability was measured by estimating total number of alleles (TNA), effective number of alleles (Ne), observed (Ho), expected (He), unbiased expected heterozygosities (uHe), inbreeding coefficient (F_{IS}), and average exclusion probability (PE) calculated with GenAlex 6.5 (Peakall and Smouse, 2012) and polymorphism information content (PIC) determined with Cervus 3.0.3 (Marshal et al., 1998). The software Identity (Wagner and Sefc, 1999) was used to calculate the probability of paternity exclusion (PE). Excess and deficiency of heterozygotes and deviations from Hardy–Weinberg equilibrium (Weir and Cockerham, 1984) were estimated using GENEPOP (Raymond and Rousset, 1995) using the Markov chain algorithm with 1000 dememorization steps for every 400 batches and 1000 iterations per batch. Samples in which a single allele per locus was detected were considered homozygous genotypes, instead of heterozygous with a null allele, for the purpose of computing genetic diversity parameters.

Genetic structure of the population was inferred with the Bayesian approach of STRUCTURE (Pritchard et al., 2000). A 20,000 initial burn-in was used to minimize the effect of the starting configurations, followed by 100,000 MC iterations, as recommended by Falush et al. (2007) with 10 independent replicates each. Several sets of inferred clusters where tested to determine the most appropriate number of clusters for modeling the data. The most likely number of clusters (K) was estimated by using the maximal value of L(K) and by calculating ΔK (Evanno et al., 2005). All runs used an admixture model with correlated frequencies and the parameter of individual admixture alpha set to be the same for all clusters and with a uniform prior. To access and visualize the distribution of different animals based on genetic distances, a three-dimensional graphic using data from the first, second and third principal coordinates (PCoA) was constructed in NTSYS-PC software package (Rohlf, 1992).

Table 3 | Number of alleles (TNA), number of effective alleles (Ne), information index (I), observed heterozygosity (Ho), expected heterozygosity (He), unbiased expected heterozygosity (uHe), polymorphism information content (PIC) and heterozygote deficiency (Fis).

Locus	TNA	Ne	I	Ho	He	uHe	PIC	Fis
AHT4	8	3.337	1.423	0.725	0.700	0.707	0.654	−0.036
AHT5	7	3.463	1.449	0.824	0.711	0.718	0.667	−0.158
ASB2	8	2.869	1.384	0.596	0.651	0.658	0.615	0.085
ASB17	11	4.222	1.828	0.846	0.763	0.771	0.743	−0.109
ABS23	8	3.659	1.528	0.827	0.727	0.734	0.686	−0.138
HMS2	7	3.411	1.489	0.667	0.707	0.714	0.672	0.057
HMS3	7	3.174	1.433	0.625	0.685	0.692	0.646	0.087
HMS6	6	2.078	1.057	0.596	0.519	0.524	0.482	−0.149
HMS7	6	3.598	1.473	0.725	0.722	0.729	0.678	−0.005
HTG4	5	3.382	1.338	0.706	0.704	0.711	0.652	−0.002
HTG6	5	2.316	1.042	0.654	0.568	0.574	0.512	−0.151
HTG7	4	1.380	0.590	0.269	0.275	0.278	0.263	0.022
HTG10	7	3.367	1.483	0.809	0.703	0.711	0.666	−0.150
LEX33	8	4.125	1.654	0.769	0.758	0.765	0.727	−0.015
VHL20	8	5.520	1.830	0.857	0.819	0.827	0.794	−0.047
Mean	7	3.327	1.400	0.700	0.668	0.674	0.630	−0.047

RESULTS

BIOMETRIC STUDY

Mean values and SD of all 37 measurements, calculated with Excel, taken from the Terceira Pony dams, and sires, as well as the approximated live weight, are presented in **Table 1**. **Table 2** presents the conformation indexes determined based on the measurements taken. These indices allow comparison between breeds and studies in terms of proportions of body segments, regardless of length differences (Komosa and Purzyc, 2009).

GENETIC STUDY

A total of 105 alleles were identified across the 15 loci, ranging from 4 for HTG7 to 11 for ASB17 and with a mean of 7 alleles per locus. The least informative locus for this population is HTG7 with the lowest values obtained for TNA, Ne, Ho, He, and PIC, and the most informative SSR marker is VHL20 with the highest TNA, Ne, He, and PIC values (**Table 3**). The total observed heterozygosity was higher than the expected heterozygosity (0.700 against 0.674, respectively), as 11 out of the 15 markers revealed observed heterozygosity values higher than the expected ones.

Tests for Hardy–Weinberg equilibrium revealed significant deficit of heterozygotes for HMS2 ($p = 0.0015$) with the corresponding F_{IS} value of 0.057. In general all loci were highly informative with the PIC values higher than 0.5 except for HTG7. The PE was 99.99 % using this set of microsatellites, indicating its usefulness in parentage testing for this breed.

The Bayesian analysis carried out in STRUCTURE demonstrated that there were no genetic clusters, as average mean values of LnP(K) did not show any substantial increases when K varied from 1 to 7 (data not shown). This result was supported by the PCoA based on the matrix of individual genotypes where no sub-grouping is observed, though a few individuals, all descending from the same founder mare (PT8), appear at the right margins of the figure. In the PCoA the first three PCoAs explain 28.20% (axis 1 = 11.25%; axis 2 = 8.78%; axis 3 = 8.17%) of the variation in the data (**Figure 1**)

DISCUSSION

The objective of this work was to determine the morphotype of the Terceira Pony and to analyze its genetic structure.

Horse breeding aims at influencing functional conformation for the improvement of traits such as sport performance. Identifying patterns of equine anatomy for different horse breeds could be helpful in predicting how successful the animals will be in performing different tasks. Part of the beauty of the Terceira Pony depends on its body conformation, body measurements and the relationships among the dimensions; therefore, metric features of the exterior can become a selection tool for the Terceira Pony (Koenen et al., 2004).

Among the standard measurements used in horse breeding, the most significant exterior variable is height at withers (Komosa and Purzyc, 2009). An average of height at withers of 128.00 cm for the population under study classifies these horses as ponies (Hendricks, 2007).

The present study indicated that from a morphological and zoometric point of view the Terceira Pony may be regarded as a mediline (BI = 0.877 for dams and 0.857 for sires; Oom and da Costa Ferreira, 1987), eumetric (DTI = 0.108; Cabral et al., 2004), elipometric ($W < 350$ kg; McManus et al., 2008), dolichocephalic and "far from ground" (CI < 0; Ribeiro, 1988; McManus et al., 2008) animal. Also the BR and QI indicate that these animals are well proportioned, with withers height, croup height, and body length approximately equal. Due to these proportions the Terceira Pony can therefore be considered as a well-proportioned saddle horse suitable for sports (Zamborlini, 2001).

The CPI defines how compact horses are and gives an indication on their aptitude for traction (McManus et al., 2005, 2008). Values obtained for the Terceira Pony are contrasting: while CPI1 showed that animals were more adapted for riding, CPI2 indicated animals adapted for light traction. The difference observed between the two indices reveals the versatility and potential of these animals, which is also supported by the variation of the angle of the shoulder. Shoulders should be efficient to transform speed into driving force transmitted by the hind limbs and therefore variations of the shoulder angle may be an indication of the potential use of the animals (Camargo and Chieffi, 1971). The longer and more obliquely positioned scapula results in a longer and more swinging gait. For the Terceira Pony the shoulder angle of sires ranges from 45 to 55° and of dams is 55° suggesting that sires can be used for draft and saddle and dams for racing (Camargo and Chieffi, 1971).

The long bones that form the limbs and thus affect height are of special importance, not only to the appearance of an animal but also to the quality of gaits and practical predispositions. From the anatomical point of view, hind limbs are used to start a stride and fore limbs are used mainly to support the body mass during movement (Mawdsley et al., 1996). The Terceira Pony has a long hind limb with indices of the metacarpus length and the metatarsus length higher than those reported for other Pony breeds (Komosa and Purzyc, 2009); this predisposition was previously associated with jumping in Polish Half-bred horses (Komosa et al., 2013).

Longer croups are desirable in racing, jumping and also in marchers and are associated with elongated and stronger muscles capable of powerful contractions necessary for speed and which facilitate propulsion (Jones, 1987). A short croup is tolerated only in draft horses, but this reduction in length must be compensated by greater muscle development (Nascimento, 1999). Also the slope of the croup influences the fitness of the horse. A croup with a horizontal direction (12–25°) is conducive to speed, inclined (25–35°) is suitable for light traction, jumping and riding, oblique (35–45°) should only be tolerated for heavy traction, and too steep (45 and 55°) is always undesirable (Nascimento, 1999). For the Terceira Pony, while males have croups with a horizontal direction (20.60°) favorable for speed, females stand out from the males for presenting an inclined croup (25.44°) suitable for jumping and riding.

According to Komosa et al. (2013) among the indices used for describing the differences in the exterior conformation of horses, the scapula index deserves particular attention. The scapula belongs to the flat bones and plays a considerable role in the movement of a horse (Komosa and Purzyc, 2009). By comparison with other Pony horses, the Hucul and the Konik (Komosa and Purzyc, 2009), the Terceira Pony showed generally lower values for the following indices: arm trunk, arm, forearm, greater trunk, smaller trunk, femur, and metacarpus circumference. The only exceptions were for the indices of scapula, chest, and croup where the Hucul horses showed slightly lower values; and for the index of metacarpus and metatarsus lengths where the Terceira Pony showed higher values than the other two breeds. The scapula, greater trunk, smaller trunk, and metacarpus circumference indices of the Terceira Pony were closer to those of the Polish Half-bred horse and the Thoroughbred (Komosa et al., 2013) than the Hucul and Konik breeds. Higher values of greater trunk index were reported for the Pantaneiro horse (McManus et al., 2008) and lower for the Alter Real (Oom and da Costa Ferreira, 1987) and the Mangalarga Marchador (Cabral et al., 2004) when compared with the Terceira Pony.

Based on this study, it can be stated that due to its height at withers, the Terceira Pony is different from the Thoroughbred and Polish Half-bred horses, but its exterior conformation is more similar to these riding breeds than to other primitive ponies like the Hucul and Konik horses.

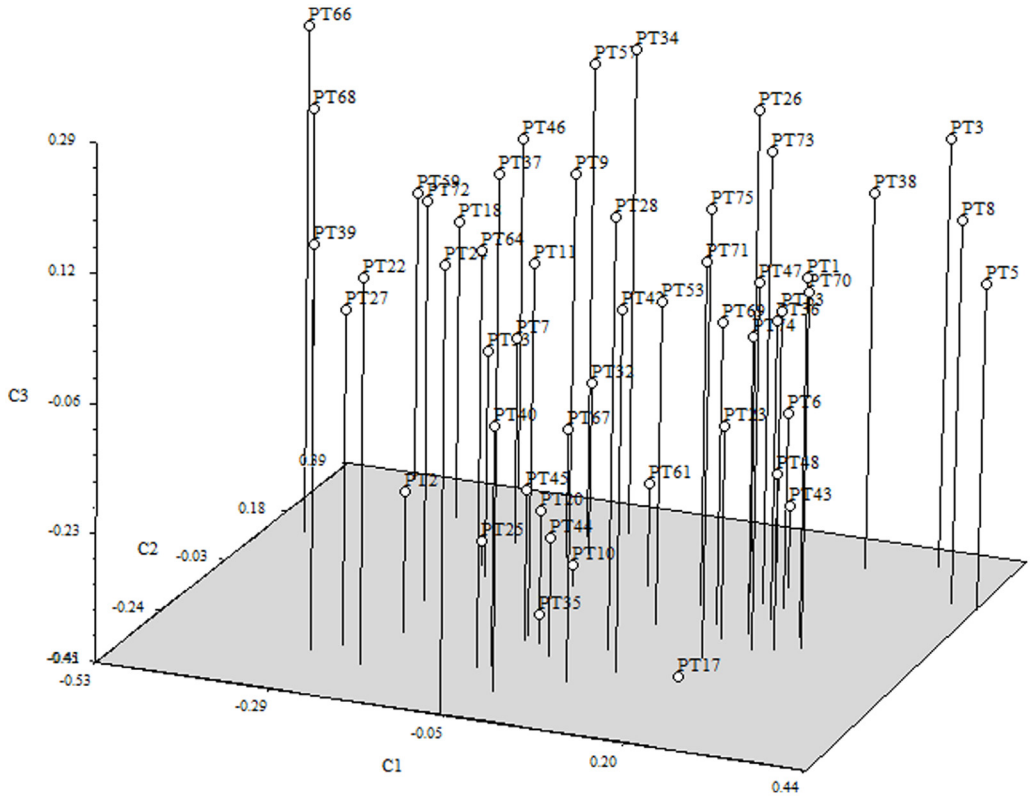


FIGURE 1 | Tri-dimensional representation of the first three axes of the principal coordinate analysis (PCoA) from the matrix of genetic distances of the Terceira Pony population.

FIGURE 2 | Trovão – Founder stallion of the Terceira Pony (height at withers 1.29 m).

Although these results may seem unexpected, the genetic background of the Terceira Pony is closer to the Polish Half-bred horse and to the Thoroughbred than to the Konik and Hucul horses. The Terceira Pony is believed to be a representative of the horses living in the Iberian Peninsula during the Portuguese and Spanish discoveries, that contributed to the development of many other European modern horse breeds that were later introduced and dispersed throughout the Americas, founding numerous breeds in the new world (Rodero et al., 1992; Luís et al., 2007; Lopes, 2011). A preliminary study conducted with 64 worldwide horse breeds including the Terceira Pony showed closest genetic similarity with breeds form Iberian origin (Lopes et al., 2011).

The genetic analysis showed that the Terceira Pony presents levels of genetic diversity similar to other, older breeds from the Iberian Peninsula, breeds from South America of Iberian origin, as well as breeds from Asia, Europe and North and South America (Luís et al., 2007; Felicetti et al., 2010; Cothran et al., 2011; Lopes, 2011). However, allelic frequencies are heterogeneous and equal to or higher than 40% for all loci except VHL20. This may be explained by the different founder lineages that the actual population may have. A census made in 2001 in Terceira island identified 144 horses with a height at withers equal or lower than 139 cm out of 726 horses (Braga, personal communication) which are believed to be the founders of the actual population. Ponies with the same phenotypic characteristics were also identified in other islands and recently introduced in the herd, who at the time this work was conducted, had few or no descendants. Nevertheless, although some founders were more represented than others, no sub division of the population was observed by either analysis methods, a result typical of a homogeneous population. According to Pedersen (1999) and Proschowsky et al. (2003) lower values for genetic diversity were expected as until now phenotypic selection has been based on a few isolated sires and dams.

The Terceira Pony has been associated with husbandry and social activities of the people from Terceira island over 100s of years. Presumably the animals that were more robust and tractable and of the "desirable type" have been bred by farmers. Therefore the Terceira Pony is a breed autochthonous of the Azorean archipelago and very well adapted to the local conditions. However, with a reduced number of animals, it still sustains relatively high genetic diversity that needs to be taken into account in future breeding strategies, for conservation purposes and in the management of the studbook to avoid genetic erosion.

The data presented in this study define this horse breed as well-proportioned with a small and narrow head, a long neck well placed between the long shoulders and leaving the withers without any convexity. The overall average of biometric variables and trends classify the Terceira Pony as having small shape, low weight, shortline, eumetric, or well balanced with strong and resistant legs considered far from ground (**Figure 2**). It is a fast animal, smart, extremely docile and easy to manage and therefore ideal for teaching equestrian sports to young children and for physical therapy or work with disabled people. Due to its phenotypic homogeneity in body conformation, balanced gaits, personality, and cultural importance and geographical location (FAO, 1999) the Terceira Pony was recently recognized as a breed. However, as the Terceira Pony still has an open stud book allowing the entrance of horses from other origins with a specific morphological type, and as breeds tend to change based on the function they are breed for, these results will serve as baseline data for future follow-up studies of the breed. The standard established for the Terceira Pony and the genetic data presented in this work are therefore of utmost importance, as the harmonization of selection and preservation is difficult and they may be opposed to one other (Bodó, 1990).

AUTHOR CONTRIBUTIONS

ACM conceived the experiment. HR and VC collected measurements and blood. MSL performed the genetic analysis. MSL, DM, and SXB analyzed the data. MSL, SXB, and ACM wrote the manuscript. All authors read the manuscript and agreed to submit it.

ACKNOWLEDGMENTS

LA-IBB-CBA-UAc, is supported by the Portuguese Foundation for Science and Technology (FCT, PEst-OE/EQB/LA0023/2013) and DRCT. The following authors were supported by FRC: MSL (M3.1.7/F/023/2011), DM (M3.1.7/F/010A/2009), HR (M3.1.7/F/007/2010), VC (M3.1.2/F/046/2011), and SXB (M3.1.7/F/026/2011).

REFERENCES

Aberle, K. S., and Distl, O. (2004). Domestication of the horse: results based on microsatellite and mitochondrial DNA markers. *Arch. Tierz.* 47, 517–535.

Binns, M. M., Holmes, N. G., Holliman, A., and Scott, A M., (1995). The identification of polymorphic microsatellite loci in the horse and their use in thoroughbred parentage testing. *Br. Vet. J.* 151, 9–15. doi: 10.1016/S0007-1935(05)80057-0

Bodó, I. (1990). "Methods and experiences with in situ preservation of farm animals," in *Animal Genetic Resources*, ed. G. Wiener (Rome: FAO Animal Production and Health Paper 80), 85–102.

Bömcke, E., Gengler, N., and Cothran, E. G. (2011). Genetic variability in the Skyros pony and its relationship with other Greek and foreign horse breeds. *Genet. Mol. Biol.* 34, 68–76. doi: 10.1590/S1415-47572010005000113

Breen, M., Lindgren, G., Binns, M. M., Norman, J., Irvin, Z., Bell, K., et al. (1997). Genetical and physical assignments of equine microsatellites-first integration of anchored markers in horse genome mapping. *Mamm. Genome* 8, 267–273. doi: 10.1007/s003359900407

Cabral, G. C., de Almeida, F. Q., Quirino, C. R., de Azevedo, P. C. N., Batista Pinto, L. F., and Santos, E. M. (2004). Morphometric evaluation of Mangalarga Marchador horse: conformation index and body proportions. *R. Bras. Zootec.* 33, 1798–1805. doi: 10.1590/S1516-35982004000700018

Camargo, M. X., and Chieffi, A. (1971). *Ezoognózia.* São Paulo: Instituto de Zootecnia, 320.

Coogle, L., Reid, R., and Bailey, E. (1996). Equine dinucleotide repeat loci from LEX025 to LEX033. *Anim. Genet.* 27, 289–290. doi: 10.1111/j.1365-2052.1996.tb00500.x

Cothran, E. G., Canelon, J. L., Luís, C., Conant, E., and Juras, R. (2011). Genetic analysis of the Venezuelan Criollo horse. *Genet. Mol. Res.* 10, 2394–2403. doi: 10.4238/2011.October.7.1

Druml, T., Baumung, R., and Sölkner, J. (2008). Morphological analysis and effect of selection for conformation in the Noreiker draught horse population. *Livest. Sci.* 115, 118–128. doi: 10.1016/j.livsci.2007.06.015

Ellegren, H., Johansson, M., Sandberg, K., and Andersson, L. (1992). Cloning of highly polymorphic microsatellites in the horse. *Anim. Genet.* 23, 133–142. doi: 10.1111/j.1365-2052.1992.tb00032.x

Evanno, G., Regnaut, S., and, Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* 14, 2611–2620. doi: 10.1111/j.1365-294X.2005.02553.x

Evans, J. W. (2000). *Horses: A Guide to Selection, Care and Enjoyment.* New York: W. H. Freeman and Company, 896.

Falush, D., Stephens, M., and Pritchard, J. K. (2007). Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Mol. Ecol. Notes* 7, 574–578. doi: 10.1111/j.1471-8286.2007.01758.x

FAO. (1999). *The Global Strategy for the Management of Farm Animal Genetic Resources*. Rome: Executive Brief, 43.

Felicetti, M., Lopes, M. S., Verini-Supplizi, A., da Câmara Machado, A., Silvestrelli, M., Mendonça, D., et al. (2010). Genetic diversity in the Maremmano horse and relationships with other European breeds. *Anim. Genet.* 41, 53–55. doi: 10.1111/j.1365-2052.2010.02102.x

Guérin, G., Bertaud, M., and Amiques, Y. (1994). Characterization of seven new horse microsatellites: HMS1, HMS2, HMS3, HMS5, HMS6, HMS7 and HMS8. *Anim. Genet.* 25, 62. doi: 10.1111/j.1365-2052.1994.tb00072.x

Hendricks, B. L. (2007). *International Encyclopedia of Horse Breeds.* Norman: University of Oklahoma Press, 496.

Hoffmann, I., Ajmone Marsan, P., Barker, J. S. F., Cothran, E. G., Hanotte, O., Lenstra, J. A., et al. (2004). "New MoDAD marker sets to be used in diversity studies for the major farm animal species: recommendations of a joint ISAG/ FAO working group," in *Proceedings of the 29th International Conference on Animal Genetics*, 123, Tokyo.

Jones, W. E. (1987). *Genética e Criação de Cavalos*. São Paulo: Roca, 666.

Koenen, E. P. C., Aldridge, L. I., and Philipsson, J. (2004). An overview of breeding objectives for warmblood sport horses. *Livest. Prod. Sci.* 88, 77–84. doi: 10.1016/j.livprodsci.2003.10.011

Komosa, M., Frackowiak, H., Purzyc, H., Wojnowska, M., Gramacki, A., and Gramacki, J. (2013). Differences in exterior conformation between primitive, Half-bred, and Thoroughbred horses: anatomic-breeding approach. *J. Anim. Sci.* 91, 1660–1668. doi: 10.2527/jas.2012-5367

Komosa, M., and Purzyc, H. (2009). Konik and Hucul horses: a comparative study of exterior measurements. *J. Anim. Sci.* 87, 2245–2254. doi: 10.2527/jas.2008-1501

Lawrence, L. A. (2001). *Horse Conformation Analysis.* Pullman, WA: Washington State University, 10.

Lopes, M. S. (2011). *Molecular Tools for the Characterisation of the Lusitano Horse.* Ph.D. thesis, University of Azores, Azores, 216.

Lopes, M. S., Cothran, G., Juras, R., Mendonça, D., and da Câmara Machado, A. (2011). "Origin of the endangered Terceira pony assessed by microsatellites and mitochondrial DNA sequence variation," in *Proceedings of the Ninth Dorothy Russell Havemeyer International Equine Genome Mapping Workshop*, Chaska, MN.

Luciano da Silva, M. (1971). *Portuguese Pilgrims and Dighton Rock*, Chap. 11. Available at: http://www.dightonrock.com/pilgrim_chapter_11.htm

Luís, C., Juras, R., Oom, M. M., and Cothran, E. G. (2007). Genetic diversity and relationships of Portuguese and other horse breeds based on protein and microsatellite loci variation. *Anim. Genet.* 38, 2–27. doi: 10.1111/j.1365-2052.2006.01545.x

Marklund, S., Ellegren, H., Eriksson, S., Sandberg, K., and Andersson, L. (1994). Parentage testing and linkage analysis in the horse using a set of highly polymorphic microsatellites. *Anim. Genet.* 25, 19–23. doi: 10.1111/j.1365-2052.1994.tb00050.x

Marshal, T. C., Slate, J., Kruuk, L., and Pemberton, J. M. (1998). Statistical confidence for likelihood-based paternity inference in natural populations. *Mol. Ecol.* 7, 639–655. doi: 10.1046/j.1365-294x.1998.00374.x

Martin-Rosset, W. (1983). Particularités de la croissance et du développment du cheval. *Ann. Zootec.* 32, 373–380. doi: 10.1051/animres:19830108

Mawdsley, A., Kelly, E. P. Smith, F. H., and Brophy, P. O. (1996). Linear assessment of the thoroughbred horse: an approach to conformation evaluation. *Equine Vet. J.* 28, 461–467. doi: 10.1111/j.2042-3306.1996.tb01618.x

McManus, C. M., Falcão, R. A., Spritze, A., Costa, D., Louvandini, H., Dias, L. T., et al. (2005). Caracterização morfológica de equinos de raça Campeiro. *R. Bras. Zootec.* 34, 1553–1562. doi: 10.1590/S1516-35982005000500015

McManus, C. M., Santos, S. A., da Silva, J. A., Louvandini, H., Avreu, U. G. P., Sereno, J. R. B., et al. (2008). Body indices for the Pantaneiro horse. *Braz. J. Vet. Res. Anim. Sci.* 45, 362–370.

Miller, S. A, Dykes, D. D., and Polesky, H. F. (1988). A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* 16, 1215. doi: 10.1093/nar/16.3.1215

Nascimento, J. F. (1999). *Mangalarga Marchador: Tratado Morfofuncional.* Belo Horizonte: Associação Brasileira dos Criadores de Cavalos Mangalarga Marchador, 577.

Oom, M. M., and da Costa Ferreira, J. (1987). Estudo biométrico do cavalo Alter. *Rev. Port. de Ciênc. Vet.* 82, 101–148.

Parker, R. (2002). *Equine Science*. New York: Thomson Delmar Learning, 688.

Peakall, R., and Smouse, P. E. (2012). GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics* 28, 2537–2539. doi: 10.1093/bioinformatics/bts460

Pedersen, N. C. (1999). A review of immunologic diseases of the dog. *Vet. Immunol. Immunopathol.* 69, 251–342. doi: 10.1016/S0165-2427(99)00059-8

Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959.

Proschowsky, H. F., Rugbjerg, H., and Ersboll, A. K. (2003). Mortality of purebred and mixed-breed dogs in Denmark. *Prev. Vet. Med.* 58, 63–74. doi: 10.1016/S0167-5877(03)00010-2

Raymond, M., and Rousset, F. (1995). GENEPOP (Version 1.2): population genetics software for exact tests and ecumenicism. *J. Hered.* 86, 248–249.

Ribeiro, D. B. (1988). *O Cavalo: Raças, Qualidade e Defeitos*. São Paulo: Editora Globo, 318.

Rodero, A., Delgado, J. V., and Rodero, E. (1992). Primitive Andalusian livestock and their implications in the discovery of America. *Arch. Zootec.* 41, 383–400.

Rohlf, F. J. (1992). *Numerical Taxonomy and Multivariable Analysis System (Version 2.11f)*. New York: Applied Biostatistics Inc., 44.

Sadek, M. H., Al-Aboud, A. Z., and Ashmawy, A. A. (2006). Factor analysis of body measurements in Arabian horses. *J. Anim. Breed. Genet.* 123, 369–377. doi: 10.1111/j.1439-0388.2006.00618.x

Santos, R. F. (1981). *O Cavalo de Sela Brasileiro e Outros Equídeos*. Botucatu: Editora Varela, 288.

Smith, J. R., Carpten, J. D., Brownstein, M. J., Ghosh, S., Magnuson, V. L., Gilbert, D. A., et al. (1995). Approach to genotyping errors caused by non-templated addition by Taq DNA Polymerase. *Genome Res.* 5, 312–317. doi: 10.1101/gr.5.3.312

Solé, M., Gómez, M. D., Molina, A., Peña F., and Valera, M. (2013). Analyses of conformational performance differentiation among functional breeding goals in the Menorca horse breed. *Arch. Tierz.* 37, 367–379. doi: 10.7482/0003-9438-56-037

Takezaki, N., and Nei, M. (1996). Genetic distance and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics* 144, 389–399.

van Haeringen, H., Bowling, A. T., Scott, M. L., Lenstra, J. A., and Zwaagstra, K. A. (1994). A highly polymorphic horse microsatellite locus: VHL20. *Anim. Genet.* 25, 207. doi: 10.1111/j.1365-2052.1994.tb00129.x

Wagner, H. W., and Sefc, K. M. (1999). *IDENTITY 4.0. Centre for Applied Genetics.* Vienna: University of Agricultural Sciences.

Weir, B. S., and Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution (N. Y.)* 38, 1358–1370.

Zamborlini, L. C. (2001). *Estudo Genético-Quantitativo da Raça Mangalarga Marchador*. Ph.D. thesis, Escola de Veterinária, Universidade Federal de Minas Gerais, Belo Horizonte, 39.

Zechner, P., Zohman, F., Sölkner, J., Bodo, I., Habe, F., Marti, E., et al. (2001). Morphological description of the Lipizzan horse population. *Livest. Prod. Sci.* 69, 163–177. doi: 10.1016/S0301-6226(00)00254-2

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 23 September 2014; accepted: 08 February 2015; published online: 27 February 2015.

Citation: Lopes MS, Mendonça D, Rojer H, Cabral V, Bettencourt SX and da Câmara Machado A (2015) Morphological and genetic characterization of an emerging Azorean horse breed: the Terceira Pony. Front. Genet. ***6****:62. doi: 10.3389/fgene.2015.00062*

This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics.

Copyright © 2015 Lopes, Mendonça, Rojer, Cabral, Bettencourt and da Câmara Machado. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 24 March 2015
doi: 10.3389/fgene.2015.00105

Fecal egg counts for gastrointestinal nematodes are associated with a polymorphism in the MHC-DRB1 gene in the Iranian Ghezel sheep breed

Rahman Hajializadeh Valilou[1], Seyed A. Rafat[1], David R. Notter[2], Djalil Shojda[1], Gholamali Moghaddam[1] and Ahmad Nematollahi[3]*

[1] Department of Animal Science, Faculty of Agriculture, University of Tabriz, Tabriz, Iran, [2] Department of Animal and Poultry Sciences, Virginia Tech, Blacksburg, VA, USA, [3] Department of Pathobiology, College of Veterinary Medicine, University of Tabriz, Tabriz, Iran

OPEN ACCESS

Edited by:
Stéphane Joost, École Polytechnique Fédérale de Lausanne, Switzerland

Reviewed by:
Francisco Ruiz-Fons, Spanish National Wildlife Research Institute – Spanish Research Council, Spain
Yuri Tani Utsunomiya, Universidade Estadual Paulista, Brazil

***Correspondence:**
Seyed A. Rafat, Department of Animal Science, Faculty of Agriculture, University of Tabriz, Boulevard 29 Bahman, Tabriz 5166616471, Iran
rafata@tabrizu.ac.ir

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: *16 November 2014*
Accepted: *28 February 2015*
Published: *24 March 2015*

Citation:
Valilou RH, Rafat SA, Notter DR, Shojda D, Moghaddam G and Nematollahi A (2015) Fecal egg counts for gastrointestinal nematodes are associated with a polymorphism in the MHC-DRB1 gene in the Iranian Ghezel sheep breed. Front. Genet. 6:105. doi: 10.3389/fgene.2015.00105

Genetic variation among sheep breeds in resistance to gastrointestinal nematodes (GIN) has been demonstrated in several production environments. Relationships between the ovine major histocompatibility complex and resistance to GIN have been studied, but few studies have systematically examined this issue in less-developed and semi-arid regions. The aim of the current study was to explore associations between fecal worm egg counts (FEC) for several GIN and polymorphisms in the DRB1 gene. One hundred male lambs were selected at 4–6 months of age from weaned animals in five flocks ($n = 20$ per flock). Body weights were determined, FAMACHA scores based on color of the ocular mucous membranes were assigned as an indicator of anemia, and blood and fecal samples were collected twice to evaluate FEC and blood packed cell volume (PCV) and for DNA isolation. A repeated-measures analysis of variance was used to test effects of genotype on FEC. The model included fixed effects of flock, genotype, time of measurement (1 or 2), and flock × time and genoype × time interactions, and a random (repeated) effect of lamb. Two genotypes (A1A1 and A1A2) were observed following digestion of Region 1 of Ovar-DRB1 with PstI. Genotypic frequencies were 0.73 for A1A1 and 0.27 for A1A2. FEC differed between Ovar_DRB1 genotypes A1A1 and A1A2 for *Marshallagia marshalli*, Strongyle, and total nematode FEC. Observed FEC were 30–41% lower for genotype A1A1. Differences among genotypes were consistent across measurement times, with no effect of genotype × measurement time interaction for any parasite class ($P \geq 0.34$). A significant association was observed between FAMACHA scores and lamb PCV, and the residual correlation between these two variables was −0.51 ($P < 0.001$). FAMACHA scores can thus be used to detect differences among lambs in PCV, and polymorphic markers of Ovar-DRB1 have potential value as an indicator of parasite resistance in applied animal breeding programs on sheep farms in this region.

Keywords: nematodes, genetic resistance, FAMACHA, MHC-DRB1, PCR-RFLP, sheep

Introduction

The Ghezel sheep is one of the 27 mainly fat-tailed native breeds of Northwestern Iran (Tavakolian, 2000). Animals of this breed graze for much of the year and are therefore continuously exposed to natural nematode infections. Gastrointestinal nematodes (GIN) of sheep and goats are widespread, diverse, and highly pathogenic, and can also infect other ruminant species such as cattle and reindeer (Jacquiet et al., 1998; Achi et al., 2003; Hrabok et al., 2006). Effects of GIN are most extreme in young animals, and therefore represent a real threat to the sheep industry (Waller, 2006).

The development of multi-drug resistance by GIN has driven research into alternative control measures, including selection of sheep that are genetically resistant to GIN infection. Genetic variation among breeds in resistance to GIN has been demonstrated in a variety of production environments. Genetically resistant sheep (either representing resistant local breeds or developed by selection within commercial breeds) are increasingly being used to improve animal production and well-being (Amarante et al., 2009). Genetically resistance sheep types also provide an opportunity to study novel mechanisms of resistance that may not be present in susceptible commercial breeds (Piedrafita et al., 2010). However, to date the mechanisms underlying genetic resistance of sheep to GIN infections are not precisely known.

Evidence for host genetic variation in aspects of disease resistance has now been documented for many diseases and in all major domestic livestock species (Bishop, 2005). In particular, small ruminants are notable for the large number of diseases where host genetic variation has been documented. Because parasite resistance in sheep has a moderate heritability (0.2–0.6; Baker, 1998; Stear et al., 2007), selective breeding has been used successfully with several breeds of sheep in different climates (Gray, 1987). Most sheep breeding programs for GIN resistance are based on recording of faecal egg counts (FEC), but this type of phenotype measurements is costly and difficult to collect on a large scale. In these situations, use of molecular genetic information is an interesting option. Use of molecular markers of resistance to GIN in sheep breeding programs has shown some promise, but difficulties remain, mainly because effects or previously identified quantitative trait loci (QTL) have not been consistent across breeds (Matika et al., 2011).

In sheep, class II genes of the major histocompatibility complex (MHC) are located on chromosome 20 and encode polymorphic glycoproteins composed of nine covalently linked subunits. Gruszczynska et al. (2000) found significant effect of OLA–DRB1 (MHC class II) on body weight at birth of Polish Heath sheep. There is also a body of scientific literature linking genes in the sheep MHC with the ability of sheep to resist infection by GIN as measured by FEC (Schwaiger et al., 1995; Buitkamp et al., 1996; Stear et al., 1996; Paterson et al., 1998). These findings are perhaps not surprising given the role of these genes in controlling specific immune responses. This "MHC effect" is never-the-less thought to be relatively small, accounting for an estimated 11% of total phenotypic variation in traits associated with GIN resistance (Buitkamp et al., 1996), although it accounts for a somewhat larger proportion of the additive genetic variation (Stear et al., 1997). These results have led to speculation that the MHC contains genes that could be used as markers for breeding to reduce FEC but that do not fully explain genetic resistance to GIN.

Few studies have addressed the relationships between the ovine MHC and resistance to GIN in less-developed and semi-arid regions. The current study is part of a multi-national collaborative project sponsored by the International Atomic Energy Agency and the Food and Agricultural Organization of the United Nations and designed to study genetic control of resistance to GIN in local sheep breeds. A particular focus of the study is the abomasal GIN *Haemonchus contortus* and was justified by this parasite's ability to produce large numbers of eggs, resulting in extensive pasture contamination; the blood-sucking nature of this nematode, which can causes life-threatening levels of anemia; and the associated potential for very significant reductions in lamb performance and survival (Gatongi et al., 1998; Waller et al., 2004). In Iran, Ashrafi et al. (2014) demonstrated polymorphism in exon 2 of MHC gene OLA-DRB1 in the Makui sheep breed of Northwestern Iran. In this study, our aim was thus to explore possible associations between nematode resistance and polymorphism of DRB1. Because *H. contortus* is only one of several GIN known to be present in the temperate regions of Iran, we likewise focused our study on a variety of different GIN known to be important in the region (Garedaghi and Bahavarnia, 2013; Moradpour et al., 2013; Yagoob et al., 2013).

Material and Methods

Study Area

The study area is shown in **Figure 1**. The districts of Eastern and Western Azerbaijan provinces are agro-ecological zones and these zones are the site of origin and the habitat of the Ghezel sheep breed in Iran (Tavakolian, 2000). This breed was therefore selected for the present study. The study region is located at latitude 35–38.8°North and longitude 46–48°East and receives annual rainfall of 150–350 mm. The temperature is highest in June, before the onset of the monsoon season. During late spring and early to mid-summer, the daily maximum temperature rarely declines below 22°C. Relative humidity is lowest during April and May and rises during the monsoon season. The year is commonly divided into four seasons: winter (December–February), spring (March–April), summer (May–September), and autumn (October–November). The summer also includes the monsoon season (July–August; www.irimo.ir). This study was conducted in May and June, a time when GIN numbers were anticipated to be elevated.

Animals and Scheduling of Phenotype Sampling

One hundred male lambs at an age of 4–6 months were selected for this study. Each lamb was randomly selected from weaned animals within five flocks ($n = 20$ per flock). After deworming to eliminate existing nematode infection and when a parasite-free condition was confirmed (28 days later), the 20 lambs in each flock were allowed to graze together with untreated

FIGURE 1 | The study region was located in Northwestern Iran between 35 and 39° North latitude and between 46 and 48° East longitude.

contemporaries from the same flock for at least 28 days without deworming. From day 31 post-infection, body weights were determined and blood and fecal samples were collected twice, 1 week apart, to evaluate fecal parasite egg counts and blood packed cell volume (PCV) and for DNA isolation. A single observer assigned FAMACHA scores for all flocks and both sampling times using a scale of 1–5. Scores were based on the color of the ocular mucous membranes surrounding the eye using procedures and color charts described by Vatta et al. (2001). All experimental procedures were approved by the University of Tabriz Animal Care and Ethics Committee.

Sample Processing

Individual fecal samples were collected from the rectum, processed to determine FEC using the modified McMaster technique, and reported as eggs per gram of feces. Observed parasite ova in the feces were categorized by parent species as: (1) Strongyles, (2) *Nematodirus* sp., (3) *Trichuris* sp., and (4) *Marshallagia marshalli*. The Strongyle group potentially included a number of common abomasal and intestinal sheep nematodes typical of mixed nematode infections in small ruminants such as *H. contortus*, *Teladorsagia circumcincta*, *Ostertagia occidentalis*, and *Trichostrongylus axei*, *colubriformis*, *vitrinus*, and *rugatus*. Fecal egg counts (FEC) were also summed across the four parasite classes for each lamb to derive a total nematode egg count. FEC were determined by the Clayton Lane technique.

Blood was obtained from the jugular vein with sterile vacuum tubes with anticoagulant (EDTA). For each sample, the PCV (%) was determined on the day of collection using the micro-hematocrit method. Blood was then mixed with 0.5 M of EDTA (pH = 8), and frozen at −20°C. DNA was isolated from blood using the protocol of Samadi Shams et al. (2011). Sequences of forward and reverse primers for amplification of the Ovar MHC-DRB1 (Region 2) gene are shown in **Table 1** (Amills et al., 1995).

The PCR was performed in a 25 μl reaction using the master mix kit (Ampliqon Company) in a T-Personal thermo-cycler (BiometeraPersonal Cycler Version 3.26 co., Germany). The PCR mixture contained: 50–100 ng of DNA, 2.5 μl of 10X PCR buffer (200 mM (NH4)2SO4, 0.1 mM Tween 20%, 750 mM Tris-HCl (pH 8.8), 2.5 mM MgCl2, 200 μM dNTPs, and 3 μl mix of oligonucleotids (10 pmol from each primer), 1U Taq DNA

TABLE 1 | Characteristics of PCR primers, nucleotide substitutions, restriction enzymes, amino acid changes, PCR producte size, and digested sequence sizes (allele) sizes for RFLP polymorphisms in Exon 2 of the Ovar-DRB1 gene.

Characteristic	Palindrome 2	Palindrome 2
Primer sequence	5-TAT CCC GTC TCT GCA GCA CAT TTC-3 5-TCG CCG CTG CAC ACT GAA ACT CTC-3	
Nucleotide substitution	(CTGCAG to GTGCAG or TTGCAG)	(TCGA to TACG)
Restriction enzyme	PstI	TaqI
Amino acid change	Phe and Tyr to Val and Cys	Phe to Tyr
PCR product(bp)	285	285
Allele size(bp)	(+/+): 241 and 44	(+/+): 163 and 122

polymerase (Dream Taq polymerase, Ampliqon company) and 11 μl ddH2O. A total of 35 cycles was adapted for denaturation at 94°C/1 min, annealing at 61°C /1 min and polymerization at 72°C/2 min (**Table 2**). The PCR products were electrophoresed at 85 V for 45 min in 2.5% agarose gels, and visualized under UV light. The power supply for electrophoresis was a PAC1000 (Bio-Rad company; USA). The size of the alleles was determined based on a 100 bp DNA size standard (Ampliqon Company) using the computer software BIO 1D++. The PCR product for each sample was digested with 10 units of PstI and TaqI enzymes at 37 and 65°C, respectively. The characterization of each has been given in **Table 3**. The digested products were separated in a 2 and 3% agaros gel for 1 h at 85 V. The gels were stained with ethidium bromide.

Statistical Analysis

Fecal egg counts were analyzed separately for each class of parasite and for the total nematode egg count. For each parasite class, FEC values were expressed as residual deviations from flock × time subclass means, and distributions of residuals was tested for skewness (ω) and kurtosis (κ). If FEC were not normally distributed, FEC values were transformed before further analysis using Box–Cox transformations of the form $[(FEC^{\lambda} - 1)/\lambda]$; Box and Cox, 1964]. Optimum values of λ for a range of values between −2 and 2 were determined using a maximum-likelihood criterion (Draper and Smith, 1981) in the TransReg Procedure of SAS.

A repeated-measures analysis of variance was conducted using the MIXED Procedure of SAS to test effects of genotype on transformed FEC, body weights, and PCV. The model included fixed effects of flock, genotype, time of measurement (1 or 2), and flock × time and genoype × time interactions and a random (repeated) effect of lamb. This model was also fitted to untransformed FEC using the GLIMMIX Procedure of SAS and assuming a negative binomial distribution of FEC (O'Hara and Kotze, 2010). Associations between FAMACHA scores and measured variables were evaluated by adding effects of FAMACHA score and associated two-way interactions with other fixed effects to this mixed model.

TABLE 2 | PCR temperatures program.

PCR step	Temperature (C^0)	Time	Cycles
Elementary denaturation	95	4 min	1
Secondary denaturation	95	1 min	35
Annealing	61	50 s	
Extension	72	1 min	
Termination	72	5 min	1

Results

Two genotypes (A1A1 and A1A2) were observed following digestion of Region 1 of Ovar-DRB1 with PstI (**Figures 2** and **3**). Two genotypes (B1B1 and B2B2) were likewise obtained following digestion with TaqI, but only one lamb had genotype B2B2 and effects of the TaqI polymorphism were therefore not considered further. For the PstI polymorphism, both genotypes were present in each of the five flocks, with overall genotypic frequencies

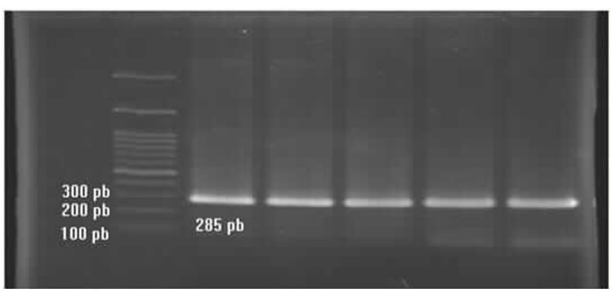


FIGURE 2 | PCR products of the Ovar-DRB1 gene.

TABLE 3 | Characteristics and features of TaqI and PstI enzymes and restriction sites.

Enzyme	Origination	Unit	Concentration	Restriction site	Optimal performance temperature (C^0)	Incubation time
PstI	*Providencia stuarti*	3000	10 U/μL	5′…C ▼TGCAG…3′3′…G ACG ▲TT…5′	37	1–16 h
TaqI	*Thermusa quaticus* YT-1	3000	10 U/μL	5′…T ▼CG A…3′ 3′…A GC ▲T…5′	65	1–16 h

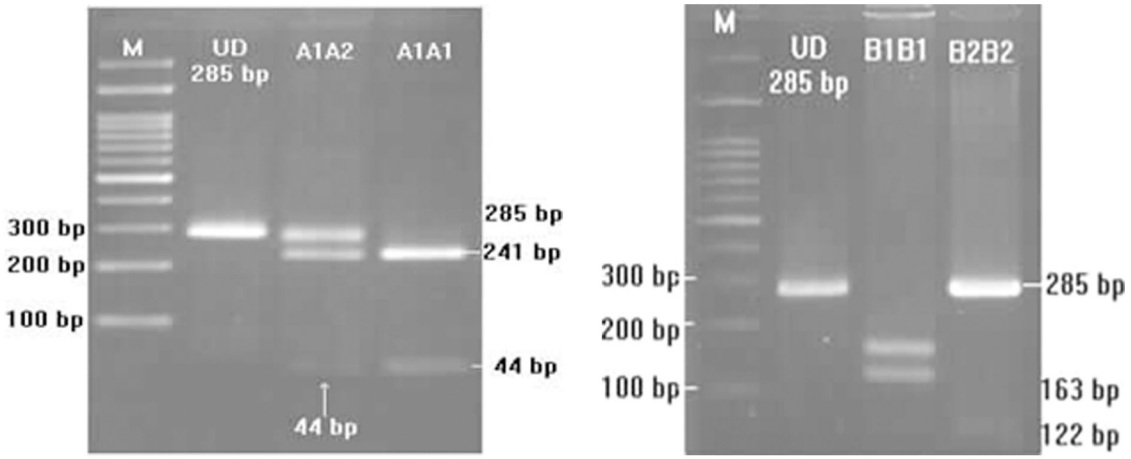


FIGURE 3 | Genotypes and PCR-RFLP pattern of the digested Ovar-DRB1 gene digested (Left: digested with PstI; Right: digested with TaqI).

of 0.73 for A1A1 and 0.27 for A1A2 (**Table 4**). The observed number of alleles at these two loci was similar to that pervious reported by Amills et al. (1995). The PCR product was sequenced ($n = 4$; Bioneer, Munpyeongseo-ro, Daedeok-gu, Daejeon 306–220, Republic of Korea). **Figure 4** illustrates the sequence.

Descriptive statistics for measured variables are shown in **Table 5**. The incidence of *Trichurus* infection in these data was very low. Eggs of *Trichurus* sp. were observed in fewer than 7% of the samples and FEC for Trichurus sp. were therefore excluded from the statistical analysis. Means and medians for FEC for remaining parasite classes are shown by flock and measurement time in **Figure 5**. Means for Strongyle FEC were very low in flocks 1 and 2, and means for *M. marshalli* FEC were very low in flocks 2, 3, and 4. The FEC records for these flocks for these parasite classes were therefore excluded from the final analyses.

After removing FEC records for flocks 1 and 2 and expressing FEC as residual deviations from flock × time subclass means, the distribution of *M. marshalli* FEC was somewhat skewed to the right ($\omega = 0.49$; $P < 0.05$) but did not exhibit significant kurtosis ($\kappa = -0.04$). This result was in agreement with the similarity between means and medians for *M. marshalli* FEC in **Figure 5**. A Box–Cox transformation with $\lambda = 0.5$ reduced the observed level of skewness ($\omega = -0.28$), and a square-root transformation was used in the final analysis of *M. marshalli* FEC. As expected from differences between means and medians in **Figure 5**, *Nematodirus*, Strongyle, and total nematode FEC were not normally distributed. For these parasite classes, distributions of FEC were strongly skewed to the right ($\omega \geq 1.28$) and were leptokurtic ($\kappa \geq 3.51$). Distributions of *Nematodirus* and Strongyle FEC also exhibited clumping at zero. Estimates of λ for Box–Cox transformations were close to zero for these parasite classes, with $\lambda = 0.1$ for total nematode and Strongyle FEC and $\lambda = -0.1$ for *Nematodirus* FEC. The Box–Cox transformation is undefined at $\lambda = 0$ but asymptotically approaches a logarithmic transformation as λ approaches zero. We therefore used a simple logarithmic transformation [ln(FEC + 1)] for these parasite classes.

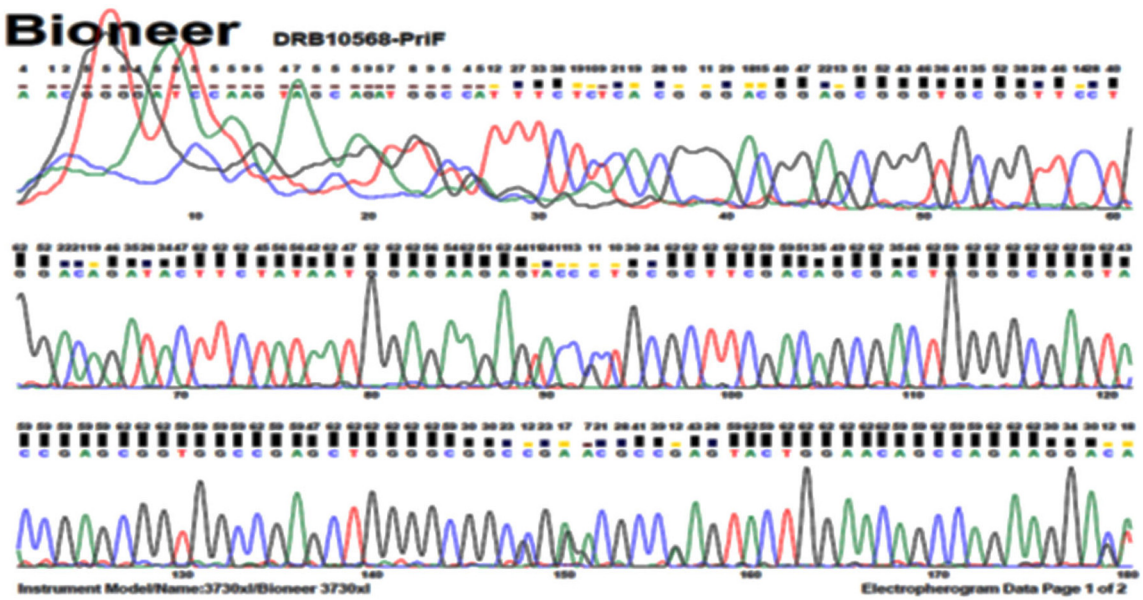
Results of the repeated-measures analysis (**Table 6**) indicated that flock effects on FEC were large ($P < 0.001$) for all remaining parasite classes. Significant differnces were observed between Ovar_DRB1 genotypes A1A1 and A1A2 for *M. marshalli*, Strongyle, and total nematode FEC (**Figure 6**). Means for *M. marshalli* FEC in **Figure 6** were based on untransformed FEC and were 40% lower for lambs of genotype A1A1 compared to lambs of genotype A1A2. Means for Strongyle, *Nematodirus*, and total nematode FEC were backtransformed from means of transformed variables (m) as ($e^{m} - 1$), and SEs for backtransformed means were approximated by assuming that SEMs for log-transformed FEC were approximately equal to coefficients of variation of backtransformed means. Backtransformed means for these parasite classes indicated that Strongyle and total nematode

TABLE 4 | Genotype and gene frequencies.

Genotypes[1]	Numbers of lambs					Overall genotypic frequencies
	Flock 1	Flock 2	Flock 3	Flock 4	Flock 5	
A1A1	15	10	16	17	15	0.73
A1A2	5	10	4	3	5	0.27

[1]*Alleles A1 and A2 were identified following digestion of Ovar-DRB1 with PstI.*
*Alleles B1 and B2 were identified following digestion with TaqI (**Figure 3**).*
Corresponding overall allelic frequencies were 0.865 and 0.135 for A1 and A2, respectively.

TATCCCGTCT CTGCAGCACA TTTCAACGGG GATCCAAGTA GCAGATGGCC ATTTCTCTCA CGGGACGGAG
CGGGTGCGGT TCCTGGACAG ATACTTCTAT AATGGAGAAG AGTACCCTGC GCTTCGACAG CGACTGGGGC
GAGTACCGAG CGGTGGCCGA GCTGGGGCGG CCGAACGCCG AGTACTGGAA CAGCCAGAAG GACATCCTGG
AGCAGACGCG GGCCGACGTG GACACGTACT GCAGACACAA CTACGGGGTC GTTGAGAGTT TCAGTGTGCA
GCGGCGAA**GA GAGTTTCAGT GTGCAGCGGC GA**



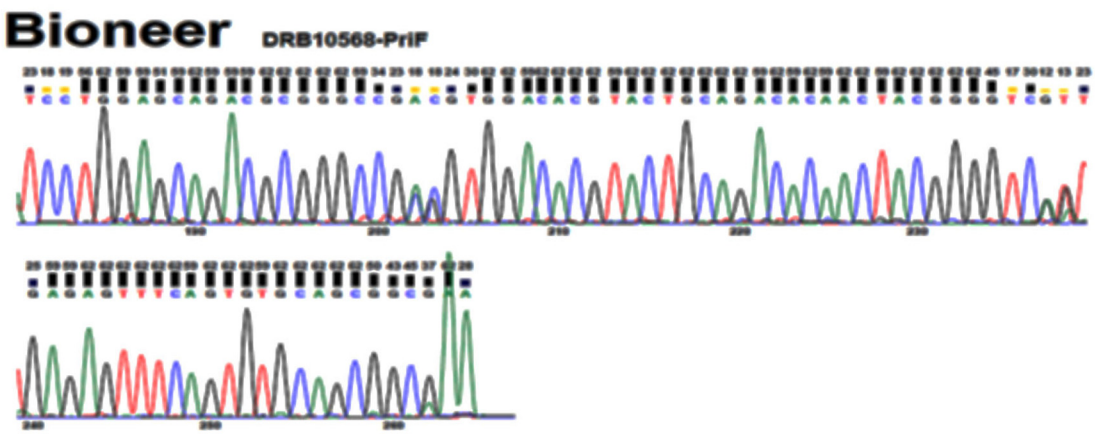


FIGURE 4 | Nucleotide sequence of the DRB1 allele in one of the samples (animal no 10568). Primer complementary regions are indicated in bold type while the PstI sites are underlined.

TABLE 5 | Descriptive statistics for fecal egg counts (FEC; eggs per gram of feces) for various classes of gastrointestinal nematodes and performance for 100 Ghezel lambs from five flocks.

Measurement	Mean	Median	SD	Minimum value	Maximum value
Strongyle FEC	4.44	0	9.38	0	61.25
Nematodirus FEC	8.41	3	13.55	0	73.50
Trichurus FEC	0.22	0	1.00	0	8.75
Marshallagia FEC	4.06	0	7.30	0	31.50
Total nematode FEC	171.28	105	202.24	0	997.50
Body weight, kg	31.84	31	5.89	19.7	47.50
PCV, %	33.04	33	3.93	21.0	46.00
FAMACHA score	1.78	2	0.74	1.0	4.00

Two fecal samples were obtained from each lamb. Samples were taken weekly beginning at approximately 31 days after deworming.

FEC were 41 and 30% lower, respectively, for lambs of genotype A1A1 compared to lambs of genotype A1A2. No effect of genotype was observed for *Nematodirus* FEC Differences among genotypes were consistent among flocks and measurement times, with no effect of genotype × flock ($P \geq 0.23$) or genotype × measurement time interaction ($P \geq 0.34$ for any parasite class.

The impact of logarithmic transformation of Strongyle and total nematode FEC can be seen by comparing differences among genotypes for untransformed FEC. For untransformed FEC, means for lambs of genotype A1A1 were 29 ($P = 0.20$) and 23% ($P = 0.12$) lower compared to lambs of genotype A1A2. These differences in effect of genotype reflect the greater impact of occasional large FEC from the skewed right tail of the FEC distribution on mean differences between genotypes and on the residual variance. This issue could be addressed in untransformed data by removing records with very high FEC as outliers. We considered recoding of extreme values to be preferable to removing them,

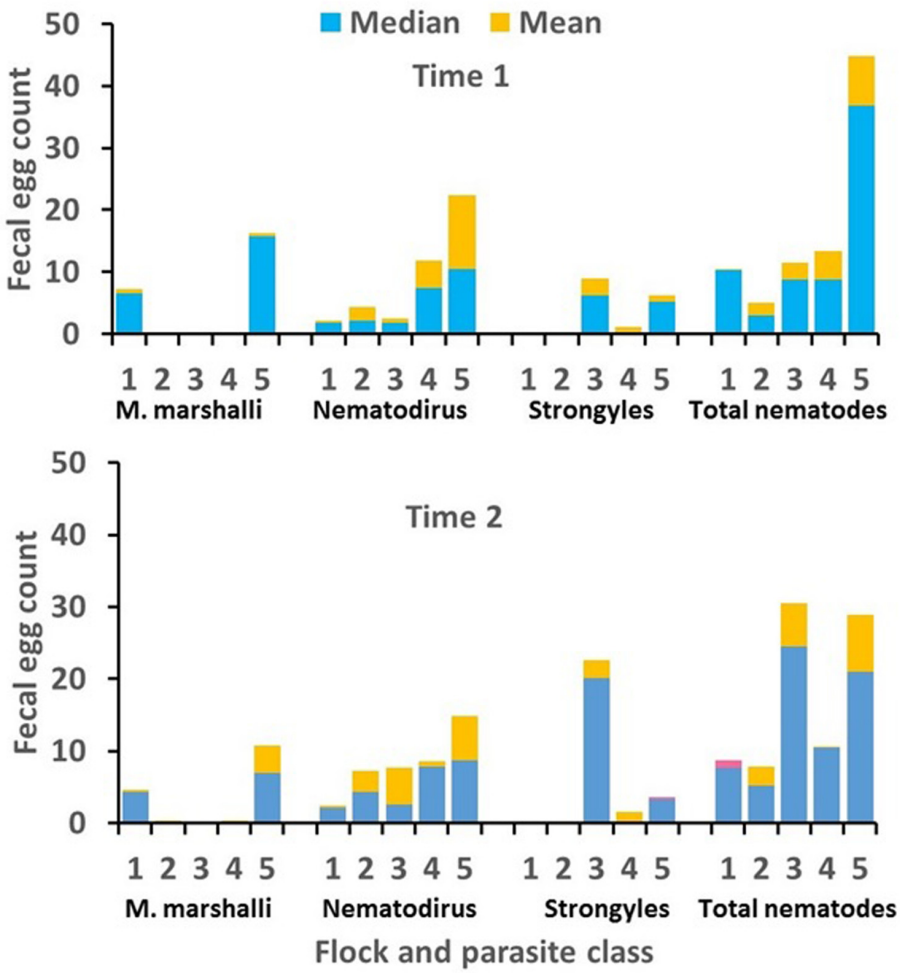


FIGURE 5 | Means and medians for Strongyle, *Nematodirus* sp., *Marshallagia marshalli*, and total nematode fecal egg counts (FEC; eggs/gram of feces) by flock and time of sampling. Medians are generally shown by the height of the blue bars and means are shown by the total height of the blue plus orange bars. In rare cases where the median exceeded the mean, the median is shown by the combined height of blue plus pink bars and the mean is shown by the height of the blue bars.

TABLE 6 | Results of analysis of variance for fixed effects of flock, genotype, time of measurement, and their two-way interactions on FEC for each parasite class.

FEC for:	Effect	*F*-value	Pr > *F*
Total nematodes	Flock	20.13	< 0.001
	Genotype	4.42	0.04
	Time	0.56	0.46
	Genotype × flock	0.39	0.81
	Time × flock	4.31	< 0.01
	Genotype × time	0.02	0.90
Strongyles	Flock	17.98	< 0.001
	Genotype	6.24	0.02
	Time	1.68	0.20
	Genotype × flock	1.11	0.34
	Time × flock	12.14	< 0.001
	Genotype × time	0.93	0.34
Marshallagia marshalli	Flock	20.66	< 0.001
	Genotype	6.57	0.01
	Time	2.27	0.14
	Genotype × flock	1.48	0.23
	Time × flock	0.28	0.60
	Genotype × time	0.28	0.60
Nematodirus	Flock	6.16	< 0.001
	Genotype	0.73	0.39
	Time	1.55	0.22
	Genotype × flock	0.25	0.91
	Time × flock	1.29	0.28
	Genotype × time	0.01	0.93

Total nematode, Strongyle, and Nematodirus FEC were transformed as ln(FEC + 1) and M. marshalli FEC were transformed as the square root of FEC before analysis.

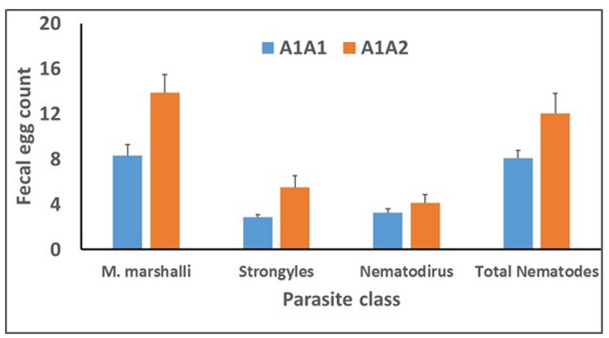


FIGURE 6 | Genotype means for FEC (egg/gram of feces) for the PstI restriction-fragment polymorphism in the Ovar-DRB1 gene for several parasite classes. Values for Strongyle, *Nematodirus* sp., and total nematode FEC were transformed as ln(FEC + 1) for statistical analysis; least-squares means (m) were back-transformed to the original scale as ($e^{m} - 1$) for presentation and are geometric, rather than arithmetic, means. Values for *M. marshalli* FEC were transformed as the square root of FEC for analysis; least-squares means shown here are based on untransfomed FEC.

and use of a normalizing transformation provides an objective strategy to account for the presence of extreme values.

Results from analysis of untransformed FEC assuming a negative binomial distribution did not differ greatly from those from the analysis of transformed FEC. Means for untransformed *M. marshalli*, Strongyle, and total nematode FEC assuming a negative binomial distribution were 40 ($P = 0.02$), 18 ($P = 0.01$), and 43% ($P = 0.15$) lower, respectively, for lambs of genotype A1A1 compared to lambs of genotype A1A2. A significant effect of genotype was thus confirmed for *M. marshalli* and Strongyle FEC, but not for total nematode FEC, perhaps in association with different contributions of the various parasite classes to total nemadode FEC among flocks and measurement times.

Consistent effects of measurement time on FEC were not observed for any parasite class, but flock × measurement time interaction was significant for Strongyle and total nematode FEC (**Table 6**). The interaction was explained by a threefold increase in Strongyle FEC between the first and second measurement time in flock 3 (**Figure 5**). Significant difference in FEC between measurements times were not observed in other flocks.

Effects of genotype were not observed for lamb body weight or PCV. The distribution of FAMACHA scores revealed little evidence of anemia in these lambs, with frequencies across flocks and measurement times for FAMACHA scores of 1 through 5 of 40, 44,15, 1, and 0%, respectively. No association was observed beteween FAMACHA scores and lamb body weights or FEC for any parasite class. However, a significant association was observed between FAMACHA scores and lamb PCV (**Figure 7**). The PCV declined linearly as FAMACHA scores declined from 1 to 3 and were much lower for the two lambs that received a FAMACHA score of four. After adjusting for effects of herd and measurement time, a residual correlation of −0.51 ($P < 0.001$) was observed between PCV and the FAMACHA score.

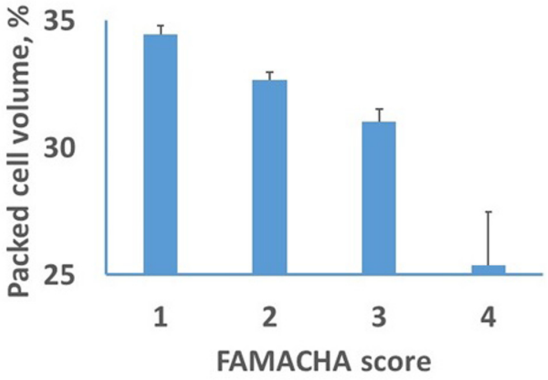


FIGURE 7 | Least-squares means and SEs for effects of FAMACHA score on packed cell volume (PCV).

Discussion

Breeding for resistance to nematode infection can complement the use of anthelmintics in sheep husbandry. Resistant animals can be selected on the basis of low FEC (Eady et al., 2003; Kahn et al., 2003), and measurement of FEC is generally considered to be the standard method for assessment of the level of resistance to GIN. The number of eggs is easy to measure,

TABLE 7 | Comparision of of results of the current study with other investigations of Ovar-DRB1 or nearby genes.

Breed	Country	Method	Gene	Association	Reference
Suffolk	Ireland	PCR products sequenced	Ovar-DRB1	Low FEC	Sayers et al. (2005)
Romanov	Poland	Microsatellite	Ovar-DRB1	Low FEC	Charon et al. (2002)
Scotish Blackface	Scotland	Simple Tandom Repeat	Ovar-DRB1	Low FEC	Schwaiger et al. (1995)
Pelibuey	Mexico	Microsatellite	Ovar-DRB1	LowFEC	Figueroa Castillo et al. (2011)
Corriedale	Brazil	Microsatellite	IL- 4	Low FEC	Benavides et al. (2009)
Texel	Ireland	Microsatellite	IFN-γ	Low FEC	Sayers et al. (2005)
Soay	Scotland	Microsatellite	IFN-γ	Low FEC	Coltman et al. (2001)
Ghezel	**Iran**	**PCR-RFLP**	**Ovar-DRB1**	**Low FEC**	**Current study**

indicates parasitism *per se*, and correlates well to the number of adult nematodes present in lambs (Douch et al., 1996; Baker, 1999). However, FEC is affected by several factors, such as parasite fecundity and egg-laying patterns, variations in egg distribution in feces, diet composition, intestinal transit time, and the level of immunity (Douch et al., 1996). A disadvantage of FEC as a marker of resistance is the requirement that animals be infected, either naturally or artificially, to determine the FEC value. The effort and cost in obtaining FEC measurements can also be a disadvantage, especially under extensive production conditions. Under some production conditions, it is therefore difficult to assess the resistance status of animals for breeding programs and there is, consequently, considerable interest in the evaluation of phenotypic and genetic markers associated with parasite resistance. Other phenotypic measures such as degree of anemia, circulating eosinophil counts, antibody levels to larval, or adult stages, and plasma pepsinogen concentrations can be used to predict worm burdens and resistance levels in infected sheep, but phenotypic markers that allow accurate prediction of an individual's resistance status in the absence of infection are generally not available (Albers et al., 1987; Beh and Maddox, 1996).

A number of studies around the world have attempted to identify relationships between genetic resistance to GIN and various genes and genetic markers. Some of these studies have been summarized in **Table 7**. Polymorphisms within the ovine MHC complex were associated with resistance to *T. colubriformis* (Douch and Outteridge, 1989), *T. circumcincta* (Schwaiger et al., 1995), and *H. contortus* (Luffau et al., 1990; Outteridge et al., 1996). However, Cooper et al. (1989), Blattman et al. (1993) and Hulme et al. (1993) did not find evidence of an association between polymorphisms in the ovine MHC locus and resistance or susceptibility to *H. contortus*. A list of QTL for GIN resistance in sheep was provided by Dominik (2005). This review suggests that there is considerable evidence for an important role for the MHC in parasite susceptibility and resistance to *H. contortus*.

Sallé et al. (2012) reported four QTL regions on sheep chromosomes (OAR) 5, 12, 13, and 21 in Romane X Martinik Black Belly backcross lamb that had an important role in genetic resistance to *H. contortus*. Riggio et al. (2014) suggested other regions of OAR1, 3, 4, 5, 7, 19, 20, and 24 that were involved in GIN resistance. Davies et al. (2006) likewise found evidences for QTL on chromosomes 2, 3, 14, and 20 that were associated with parasitic infections in Scottish blackface sheep.

Results of the current study provided additional evidence of an association between polymorphism in the DRB1 gene and GIN FEC, and the first indication of an effect of this locus on *M. marshalli* FEC. Screening of GIN FEC under natural infection is informative because it corresponds to typical conditions in the production environment. However, possible interactions among effects of different parasite classes and variation among flocks and measurement times in parasite loads preclude a deeper understanding of mechanisms driving the observed associations. More intersive studies, involving controlled infections with individual species of GIN are thus required to confirm hypothesized effects of DRB1 genotype on GIN parasite resistance and to confirm the specificity or generality of observed associations among parasite classes.

A significant negative association between PCV and FAMACHA score confirmed that the FAMACHA score can be used to diagnose differences in PCV in lambs and was consistent with previous results (Kaplan et al., 2004). Recommendations for veterinary intervention based on FAMACHA scores (Vatta et al., 2001) suggest that lambs with scores of one or two do not require attention, but that lambs with scores of four or five require immediate attention. Recommendations for animals with a score of three depend upon the age and nutritional status of the lamb and anticipated cause(s) of anemia. Intervention is generally recommended for lambs, but not adults, with a FAMACHA score of three. In the current study, the lack of association between FAMACHA scores and FEC, low overall FEC levels, and limited evidence for parasitism by *H. contortus* (the only GIN, of those evaluated, that causes blood loss and anemia) suggest that parasitism probably is not the main cause of subclinical anemia in these flocks. Nonetheless, results in **Figure 7** indicate that FAMACHA scores can be used to detect mean differences in PCV among lambs.

Conclusion

Our results reinforce previous evidence that some alleles of the ovine MHC are involved in determining levels of susceptibility or resistance to infection with GIN. This result provides the opportunity to use these alleles as genetic markers of resistance to GIN, leading to the development of that are better adapted to parasite infestations in the environment.

Implication of this research are that FAMACHA tests and polymorphic markers of Ovar-DRB1 can be used in applied animal breeding programs on sheep farms of the region, especially in animals infected with GIN and located in the temperate regions of Asia. Assessment of the precision of genetic evaluations based on molecular information has potential to provide a new perspective on the design of sheep breeding schemes and selection programs (Assenza et al., 2014).

Acknowledgments

We acknowledge the assistance of the FAO/IAEA program, Contract No 16101. We acknowledge A. Javanmard, A. Barzgari, S. Radmand, and H. Cheraghi for help in this research. We appreciate constructive comments by two anonymous reviewers regarding options for statistical analysis of FEC.

References

Achi, Y. L., Zinsstag, J., Yao, K., Yeo, N., Dorchies, P., and Jacquiet, P. (2003). Host specificity of *Haemonchus* spp. for domestic ruminants in the savanna in northern Ivory Coast. *Vet. Parasitol.* 116, 151–158. doi: 10.1016/S0304-4017(03)00258-9

Albers, A. J., Gray, G. D., Piper, L. R., Le Jambre, L. F., and Barger, I. A. (1987). The genetics of resistance and resilience to *Haemonchus contortus* infection in young merino sheep. *Int. J. Parasitol.* 17, 1355–1363. doi: 10.1016/0020-7519(87)90103-2

Amarante, A. F. T., Susin, I., Rocha, R. A., Silva, M. B., Mendes, C. Q., and Pires, A. V. (2009). Resistance of santa ines and crossbred ewes to naturally acquired gastrointestinal nematode infections. *Vet. Parasitol.* 165, 273–280. doi: 10.1016/j.vetpar.2009.07.009

Amills, M., Francino, O., and Sanchez, A. (1995). Nested PCR allows the characterization of TaqI and PstI RFLPs in the second exon of the caprine MHC class II DRB gene. *Vet. Immunol. Immunopathol.* 48, 313–321. doi: 10.1016/0165-2427(95)05442-9

Ashrafi, F., Hashemi, A., Mardani, K., and Darvishzadeh, R. (2014). Study on genetic variability in MHC-DRB1 second exon in makuie sheep breed population. *Genetika* 46, 269–275. doi: 10.2298/GENSR1401269A

Assenza, F., Elsen, J. M., Legarra, A., Carré, C., Sallé, G., Robert-Granié, C., et al. (2014). Genetic parameters for growth and faecal worm egg count following *Haemonchus contortus* experimental infestations using pedigree and molecular information. *Genet. Sel. Evol.* 46, 13. doi: 10.1186/1297-9686-46-13

Baker, R. L. (1998). A review of genetic resistance to gastrointestinal nematode parasites in sheep and goats in the tropics and evidence for resistance in some sheep and goat breeds in sub-humid coastal Kenya. *Anim. Gen. Res. Inform. Bull.* 24, 13–30. doi: 10.1017/S1014233900001103

Baker, R. L. (1999). Genetics of resistance to endoparasites and ectoparasites. *Int. J. Parasitol.* 29, 73–75. doi: 10.1016/S0020-7519(98)00181-7

Beh, K. J., and Maddox, J. F. (1996). Prospects for development of genetic markersf for resistance to gastrointestinal parasite infection in sheep. *Int. J. Parasitol.* 26, 879–897. doi: 10.1016/S0020-7519(96)80060-9

Benavides, M. V., Weimer, T. A., Borba, M. F. S., and Sacco, A. M. S. (2009). Genetic analyses of polymorphisms on ovine chromosomes 5 and 20 and their effect on resistance to internal parasites. *Small Rumin. Res.* 83, 67–73. doi: 10.1016/j.smallrumres.2009.03.006

Bishop, S. C. (2005). "Disease resistance: genetics," in *Encyclopedia of Animal Science*, eds W. G. Pond and A. W. Bell (New York: Marcel Dekker), 288–290.

Blattman, A. N., Hulme, D. J., Kinghorn, B. P., Woolaston, R. R., Gray, G. D., and Beh, K. J. (1993). A search for associations between major histocompatibility complex restriction fragment length polymorphism bands and resistance to *Haemonchus confortus* infection in sheep. *Anim. Genet.* 24, 277–282. doi: 10.1111/j.1365-2052.1993.tb00311.x

Box, G. E. P., and Cox, D. R. (1964). An analysis of transoformations. *J. R. Stat. Soc. Ser. B* 26, 211–252.

Buitkamp, J., Filmether, P., Stear, M. J., and Epplen, J. T. (1996). Class I and class II major histocompatibility complex alleles are associated with faecal egg counts following natural, predominantly *Ostertagia circumcincta* infection. *Parasitol. Res.* 82, 693–696. doi: 10.1007/s004360050187

Charon, K. M., Moskwa, B., Rutkowski, R., Gruszczyñska, J., and Świderek W. (2002). Microsatellite polymorphism in DRB1 gene (MHC class II) and its relation to nematode faecal egg count in Polish Heath Sheep. *J. Anim. Feed Sci.* 11, 47–58.

Coltman, D. W., Wilson, K., Pilkington, J. G., Stear, M. J., and Pemberton, J. M. (2001). A microsatellite polymorphism in the gamma interferon gene is associated with resistance to gastrointestinal nematodes in a naturally-parasitized population of Soay sheep. *Parasitology* 122, 571–582. doi: 10.1017/S0031182001007570

Cooper, D. W., Van Oorschot, R. A. H., Piper, L. R., and Le Jambre, L. F. (1989). No association between the ovine leucocyte antigen (OLA) system in the australian merino and susceptibility to *Haemonchuscontortus* infection. *Int. J. Parasitol.* 19, 695–697. doi: 10.1016/0020-7519(89)90051-9

Davies, G., Stear, M. J., Benothman, M., Abuagob, O., Kerr, A., Mitchell, S., et al. (2006). Quantitative trait loci associated with parasitic infection in Scottish blackface sheep. *Heredity* (*Edinb*) 96, 252–258. doi: 10.1038/sj.hdy.6800788

Dominik, S. (2005). Quantitative trait loci for internal nematode resistance in sheep: a review. *Genet. Sel. Evol.* 37(Suppl. 1), S83–S96. doi: 10.1051/gse:2004027

Douch, P. G. C., Green, R. S., Morris, C. A., McEwan, J. C., and Windon, R. G. (1996). Phenotypic markers for selection of nematode-resistant sheep. *Int. J. Parasitol.* 26, 899–911. doi: 10.1016/S0020-7519(96)80062-2

Douch, P. G. C., and Outteridge, P. M. (1989). The relationship between ovinelymphocyte antigens and parasitological and production parameters in Romney sheep. *Int. J. Parasitol.* 19, 35–41. doi: 10.1016/0020-7519(89)90019-2

Draper, N. R., and Smith, H. (1981). *Applied Regression Analysis*, 2nd Edn. New York: John Wiley & Sons.

Eady, S. J., Woolaston, R. R., and Barger, I. A. (2003). Comparison of genetic and nongenetic strategies for control of gastrointestinal nematodes of sheep. *Livest. Prod. Sci.* 81, 11–23. doi: 10.1016/S0301-6226(02)00197-5

Figueroa Castillo, J. A., Mendez Mendina, R. D., BerruecosVillalbos, J. M., Gayosso-Vazquez, A., Ulloa-Arvizu, R., Acosta Rodriguez, R., et al. (2011). Association between major histocompatibility complex microsatellites, fecal egg count, blood packed cell volume and blood eosinophilia in Peibuey sheep infected with *Haemonchus contortus*. *Vet. Parasitol.* 177, 339–344. doi: 10.1016/j.vetpar.2010.11.056

Garedaghi, Y., and Bahavarnia, S. R. (2013). Prevalence and species composition of abomasal nematodes in sheep and goats slaughtered at Tabriz town, Iran. *J. Anim. Sci. Adv.* 3, 37–41. doi: 10.5455/jasa.20130219031609

Gatongi, P. M., Prichard, R. K., Ranjan, S., Gathuma, J. M., Munyua, W. K., Cheruyot, H., et al. (1998). Hypobiosis of *Haemonchus contortus* in natural infections of sheep and goats in a semi-arid area of Kenya. *Vet. Parasitol.* 77, 49–61. doi: 10.1016/S0304-4017(97)00207-0

Gray, G. D. (1987). Genetic resistance to haemonchosis in sheep. *Parasitol. Today* 3, 253–255. doi: 10.1016/0169-4758(87)90155-4

Gruszczynska, J., Charon, K. M., Kitlinska, J., and Szydlowski, M. (2000). The influence of OLA-DRB1 (MHC class II) gene polymorphism on lamb body weight and weight gain in Polish Heath Sheep. *J. Appl. Genet.* 41, 101–112.

Hrabok, J. T., Oksanen, A., Nieminen, M., and Waller, P. J. (2006). Prevalence of gastrointestinal nematodes in winter slaughtered reindeer of northern Finland. *Rangifer* 27, 133–139. doi: 10.7557/2.27.2.194

Hulme, D. J., Nicholas, F. W., Windon, R. G., Brown, S. C., and Beh, K. J. (1993). The MHC class II region and resistance to an intestinal parasite in sheep. *J. Anim. Breed. Gen.* 110, 459–472. doi: 10.1111/j.1439-0388.1993.tb00758.x

Jacquiet, P., Cabaret, J., Thiam, E., and Cheikh, D. (1998). Host range and the maintenance of *Haemoncus* spp. in an adverse arid climate. *Int. J. Parasitol.* 28, 253–261. doi: 10.1016/S0020-7519(97)00185-9

Kahn, L. P., Knox, M. R., Walkden- Brown, S. W., and Lea, J. M. (2003). Regulation of the resistance to nematode parasites of single- and twin-bearing Merino

ewes through nutrition and genetic selection. *Vet. Parasitol.* 114, 15–31. doi: 10.1016/S0304-4017(03)00099-2

Kaplan, R. M., Burke, J. M., Terrill, T. H., Miller, J. E., Getz, W. R., Mobini, S., et al. (2004). Validation of the FAMACHA© eye color chart for detecting clinical anemia on sheep and goat farms in the southern United States. *Vet. Parasitol.* 123, 105–120. doi: 10.1016/j.vetpar.2004.06.005

Luffau, G., Vu TienKhang, J., Bouix, J., Nguyen, T. G., Cullen, P., and Ricordeau, G. (1990). Resistance to experimental infections with *Haemonchus contortus* in Romanov sheep. *Genet. Sel. Evol.* 22, 205–229. doi: 10.1186/1297-9686-22-2-205

Matika, O., Pong-Wong, R., Wooliams, J. A., and Bishop, S. C. (2011). Confirmation of two quantitative trait loci regions for nematode resistance in commercial British terminal sire breeds. *Animal* 5, 1149–1156. doi: 10.1017/S175173111100022X

Moradpour, N., Borji, H., Razmi, G., Maleki, M., and Kazimi, H. (2013). Pathophysiology of *Marshallagia marshalli* in experimentally infected lambs. *Parasitology* 140, 1762–1767. doi: 10.1017/S0031182013001042

O'Hara, R. B., and Kotze, D. J. (2010). Do not log-transform count data. *Meth. Ecol. Evol.* 1, 118–122. doi: 10.1111/j.2041-210X.2010.00021.x

Outteridge, P. M., Anderson, L., Douch, P. G. C., Green, R. S., Gwakisa, P. S., Hohenhaus, M. A., et al. (1996). The PCR typing of MHC-DRB genes in the sheep using primers for an intronic microsatellite: application to nematode parasite resistance. *Immunol. Cell Biol.* 74, 330–336. doi: 10.1038/icb.1996.59

Paterson, S., Wilson, K., and Pemberton, J. M. (1998). Major histocompatibility complex variation associted with juvenile survival and prasite resistance in a large ungulate population (*Ovis aries* L.). *Proc. Natl. Acad. Sci. U.S.A.* 95, 3714–3719. doi: 10.1073/pnas.95.7.3714

Piedrafita, D., Raadsma, H. W., Gonzalez, J., and Meeusen, E. (2010). Increased production through parasite control: can ancient breeds of sheep teach us new lessons? *Trends Parasitol.* 26, 568–573. doi: 10.1016/j.pt.2010.08.002

Riggio, V., Pong-Wong, R., Salle, G., Usai, M. G., Casu, S., Moreno, C. R., et al. (2014). A joint analysis to identify loci underlying variation in nematode resistance in three European sheep populations. *J. Anim. Breed. Genet.* 131, 426–436. doi: 10.1111/jbg.12071

Sallé, G., Jacquiet, P., Gruner, L., Cortet, J., Sauvé, C., Prévot, F., et al. (2012). A genome scan for QTL affecting resistance to *Haemonchus contortus* in sheep. *J. Anim. Sci.* 90, 4690–4705. doi: 10.2527/jas.2012-5121

Samadi Shams, S., Zununi Vahed, S., Soltanzad, F., Kafil, V., Barzegari, A., Atashpaz, S., et al. (2011). Highly effective DNA extraction method from fresh, frozen, dried and clotted blood samples. *BioImpacts* 1, 183–187. doi: 10.5681/bi.2011.025

Sayers, G., Good, B., Hanrahan, J. P., Ryan, M., Angles, J. M., and Sweeney, T. (2005). Major histocompatibility complex DRB1 gene: its role in nematode resistance in Suffolk and Texel sheep breeds. *Parasitology* 131, 403–409. doi: 10.1017/S0031182005007778

Schwaiger, F. W., Gostomski, D., Stear, M. J., Duncan, J. L., Mckellar, Q. A., Epplen, J. T., et al. (1995). An-ovine major histocompatibility complex DRB1 allele is associated with low faecal egg counts following natural, predominantly Ostertagia circumcincta infection. *Int. J. Parasitol.* 25, 815–822. doi: 10.1016/0020-7519(94)00216-B

Stear, M. J., Bairden, K., Bishop, S. C., Buitkamp, J., Duncan, J. L., Gettinby, G., et al. (1997). The genetic basis of resistance to *Ostertagia circumcincta* in lambs. *Vet. J.* 154, 111–119. doi: 10.1016/S1090-0233(97)80049-4

Stear, M. J., Doligalska, M., and Donskow Schmelter, K. (2007). Alternatives to anthelmintics for the control of nematodes in livestock. *Parasitology* 134, 139–151. doi: 10.1017/S0031182006001557

Stear, M. J., Park, M., and Bishop, S. C. (1996). The key components of resistance to *Ostertagia circumcincta* in lambs. *Parasitol. Today* 12, 438–441. doi: 10.1016/0169-4758(96)10069-7

Tavakolian, J. (2000). *An Introduction to Genetic Resources of Native Farm Animals in Iran*. Tehran: Animal Science Genetic Research Institute Press.

Vatta, A. F., Letty, B. A., van der Linde, M. J., van Wijk, E. F., Hansen, J. W., and Krecek, R. C. (2001). Testing for clinical anaemia caused by *Haemonchus* spp. in goats farmed under resource-poor conditions in South Africa using an eye colour chart developed for sheep. *Vet. Parasitol.* 99, 1–14. doi: 10.1016/S0304-4017(01)00446-0

Waller, P. J. (2006). Sustainable nematode parasite control strategies for ruminant livestock by grazing management and biological control. *Anim. Feed Sci. Tech.* 126, 277–289. doi: 10.1016/j.anifeedsci.2005.08.007

Waller, P. J., Schwan, O., Ljungström, B.-L., Rydzik, A., and Yeates, G. W. (2004). Evaluation of biological control of sheep parasites using *Duddingtonia flagrans* under commercial arming conditions on the island of Gotland, Sweden. *Vet. Parasitol.*126, 299–315. doi: 10.1016/j.vetpar.2004.08.008

Yagoob, G., Hossein, H., and Asso, F. (2013). Prevalence of abomasal nematodes in sheep slaughtered at Baneh town. *Amer. J. Anim. Vet. Sci.* 8, 142–145. doi: 10.3844/ajavsp.2013.142.145

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Valilou, Rafat, Notter, Shojda, Moghaddam and Nematollahi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

APPROACHES AND TOOLS FOR BREEDING PROGRAMS

frontiers in Genetics

REVIEW
published: 13 May 2015
doi: 10.3389/fgene.2015.00167

An interpretive review of selective sweep studies in *Bos taurus* cattle populations: identification of unique and shared selection signals across breeds

Beatriz Gutiérrez-Gil[1]*, *Juan J. Arranz*[1] and *Pamela Wiener*[2]

[1] Departamento de Producción Animal, Universidad de León, León, Spain, [2] Division of Genetics and Genomics, Roslin Institute and R(D)SVS, University of Edinburgh, Midlothian, UK

OPEN ACCESS

Edited by:
Michael William Bruford, Cardiff University, UK

Reviewed by:
Yuri Tani Utsunomiya, Universidade Estadual Paulista, Brazil
Dominique Rocha, Institut National de la Recherche Agronomique, France

***Correspondence:**
Beatriz Gutiérrez-Gil, Departamento de Producción Animal, Facultad de Veterinaria, Universidad de León, Campus de Vegazana, 24071 León, Spain
beatriz.gutierrez@unileon.es

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: 31 December 2014
Accepted: 13 April 2015
Published: 13 May 2015

Citation:
Gutiérrez-Gil B, Arranz JJ and Wiener P (2015) An interpretive review of selective sweep studies in Bos taurus cattle populations: identification of unique and shared selection signals across breeds. Front. Genet. 6:167. doi: 10.3389/fgene.2015.00167

This review compiles the results of 21 genomic studies of European *Bos taurus* breeds and thus provides a general picture of the selection signatures in taurine cattle identified by genome-wide selection-mapping scans. By performing a comprehensive summary of the results reported in the literature, we compiled a list of 1049 selection sweeps described across 37 cattle breeds (17 beef breeds, 14 dairy breeds, and 6 dual-purpose breeds), and four different beef-vs.-dairy comparisons, which we subsequently grouped into core selective sweep (CSS) regions, defined as consecutive signals within 1 Mb of each other. We defined a total of 409 CSSs across the 29 bovine autosomes, 232 (57%) of which were associated with a single-breed (Single-breed CSSs), 134 CSSs (33%) were associated with a limited number of breeds (Two-to-Four-breed CSSs) and 39 CSSs (9%) were associated with five or more breeds (Multi-breed CSSs). For each CSS, we performed a candidate gene survey that identified 291 genes within the CSS intervals (from the total list of 5183 BioMart-extracted genes) linked to dairy and meat production, stature, and coat color traits. A complementary functional enrichment analysis of the CSS positional candidates highlighted other genes related to pathways underlying behavior, immune response, and reproductive traits. The Single-breed CSSs revealed an over-representation of genes related to dairy and beef production, this was further supported by over-representation of production-related pathway terms in these regions based on a functional enrichment analysis. Overall, this review provides a comparative map of the selection sweeps reported in European cattle breeds and presents for the first time a characterization of the selection sweeps that are found in individual breeds. Based on their uniqueness, these breed-specific signals could be considered as "divergence signals," which may be useful in characterizing and protecting livestock genetic diversity.

Keywords: cattle, breeds, selection signals, candidate genes, diversity, selective sweep, domestication

Introduction

The genetic diversity of livestock species is an economical and cultural inheritance from our ancestors, and an indispensable resource to meet the unpredictable needs of our future (Larson et al., 1992). The history of this diversity involves the spread of livestock populations from their centers of domestication as small samples of the original domesticated populations. Under new environments and the effects of genetic drift and natural selection, the different groups developed into distinct local populations (FAO, 2013). Associated with later advances in animal husbandry and breeding, more specialized breeds and breeding lines were developed. During the past 250 years, there has been a development of individually uniform but collectively highly diverse and distinguishable populations, which are known as "standardized breeds" (FAO, 2013).

In livestock populations, approximately half of the genetic diversity is shared across breeds while the other half is observed within single breeds (Sponenberg and Bixby, 2007). Hence, the substantial loss of biodiversity associated with the loss of a breed means that effective management of breeds is essential to managing the overall biodiversity of domesticated species.

During the establishment of modern livestock breeds, the genomes of domestic animal species have been subjected to multiple human-imposed selection events influencing traits of concern to agriculturists. In comparison to natural selection, artificial selection has the ability to rapidly change the genome. Selection not only affects the favored mutation but it produces a "hitchhiking" effect on the frequency of neutral alleles at linked loci (Maynard Smith and Haigh, 1974; Kaplan et al., 1989). Selection-mapping or hitchhiking mapping approaches exploit this phenomenon by searching for genomic regions of reduced variability as signatures of strong positive selection, with the aim of identifying causal mutations controlling selected phenotypes (e.g., Kohn et al., 2000; Harr et al., 2002; Storz et al., 2004; Pollinger et al., 2005; Voight et al., 2006). The different methods developed for detection of selection signatures through the analysis of genetic markers are based either on the distribution of allelic frequencies or the properties of haplotypes segregating within populations, or on the distribution of genetic differentiation between populations (reviewed by Hohenlohe et al., 2010).

In recent years, the availability of high-density, genome-wide single nucleotide polymorphism (SNP) arrays and parallel progress in statistical techniques have allowed the identification of genomic regions that have been subjected to positive artificial selection in livestock species ("selection scans"). While identifying a selection signature in the same region in different breeds gives support to the hypothesis that a particular genomic region has undergone selection for a given trait, many selection signatures appear to be breed-specific. By comparing the results of the studies that have searched for selection signatures in different cattle breeds, this review provides a map of selection footprints that could be considered a source of genetic diversity in these domestic populations and therefore represent a valuable resource that may be worth protecting independently of the productive ability of the breed(s) involved.

Genetic Diversity and Selection Signature Studies in *Bos taurus* Cattle

Present day cattle breeds are the result of years of human selection, adaptation to different environments and cross-breeding, as well as demographic effects such as bottlenecks and migration, all of which contribute to the current patterns of genetic diversity (Bruford et al., 2003; Laloe et al., 2010). Human-mediated selective processes include those related to domestication, breed formation, and ongoing selection to enhance performance and productivity. In 2009, the Bovine HapMap Consortium presented the first detailed genome-wide characterization of the genetic variability of 19 geographically and phylogenetically diverse bovine breeds, based on the analysis of 37,470 SNPs. This study showed that taurine breeds (*Bos taurus*) showed a lower genetic diversity than indicine breeds (*Bos indicus*), probably due to a lower diversity within the pre-domestication ancestral population and/or post-domestication effects of stronger bottlenecks at breed formation and stronger selection for docility and productivity (Bovine HapMap Consortium, 2009). The authors concluded that despite the decline in effective population size (N_e) of some breeds, overall genetic diversity in cattle was "not low" and the between-breed differences in diversity were due to events at and before breed formation rather than differences in the intensity of natural or artificial selection post-domestication. This study was the first to perform a high-resolution, genome-wide examination of the structure of the cattle genome in different breeds and reported selection signatures in regions involving genes known to harbor causal mutations related to production traits (e.g., *GDF-8* and *ABCG2,* in relation to muscle conformation and milk composition, respectively) and genes associated with food conversion efficiency (e.g., *R3HDM1*). Since this initial analysis, many studies have followed, with the common aim of identifying specific genomic regions influenced by artificial selection in cattle breeds.

This review compiles the results of 21 genomic studies of European-related *Bos taurus* populations and thus provides a general picture of the selection signatures in taurine cattle identified by genome-wide selection-mapping scans. By performing a systematic comparison of the results reported in the literature, we have identified those regions that are found in several breeds showing the same production characteristics, and that therefore are very likely to harbor mutations with significant effects on production traits. In general, these are the regions that have already been highlighted by the different authors, as they show the highest statistical support for the presence of a positive selection signature, and because in many cases they contain genes related to the shared production characteristics that can be viewed as selection candidates. We also show that in many cases selection signatures are also shared by breeds showing different production characteristics. These may be regions of interest in relation to metabolic homeostasis or other general traits such as disease resistance and behavior. But one of the main objectives of the interpretive survey presented herein is to highlight those regions that have been reported in a single breed. In general, results of this type are not discussed in detail by the authors, and in some cases are not presented to the reader, such that a large portion

of the biological information generated through these genomic studies is never interpreted. However, we hypothesize herein that these single-breed sweeps may indicate genomic sources of unique phenotypic characteristics of the target breed for which the selection signal has been detected. Although determining the phenotype associated with these single-breed sweeps may be particularly difficult, the identification and characterization of these regions as "divergence signals" may be of value as an initial step to protect, from a genomic point of view, the wealth of livestock diversity.

General Overview of the Reviewed Studies

As an initial attempt to perform a systematic review of the available literature on cattle selection signals, this review targets the genome-wide selective sweep scans described in *Bos taurus* breeds of European origin and mainly focuses on the interpretation of selection sweeps associated with dairy and beef production specialization. Hence, some studies have not been considered, including studies limited to specific chromosomes (e.g., Hayes et al., 2008; Prasad et al., 2008), or studies mainly addressing *Bos indicus* (Somavilla et al., 2014), African taurine cattle breeds (Gautier et al., 2009) or cross-bred cattle (Flori et al., 2012) or studies focusing on *Bos taurus-Bos indicus* comparisons (Porto-Neto et al., 2013; Utsunomiya et al., 2013a). Exceptions were four studies that included in their larger-scale analysis some *Bos indicus* and hybrid breeds (Bovine HapMap Consortium, 2009; Qanbari et al., 2011; Ramey et al., 2013; Porto-Neto et al., 2014), although we have only considered the results reported for the European *Bos taurus* breeds. Details of the 21 studies compiled in this review are provided in **Table 1**, including information about the breeds analyzed and their production characteristics, the statistical method(s) used for the identification of selection signatures, the SNP-chip or dataset analyzed, and other technical details such as the version of the reference genome on which the study was based.

Depending on the number of breeds analyzed, we classify the studies as those that focus on: (i) a single breed (Qanbari et al., 2010, 2014; Glick et al., 2012; Boitard and Rocha, 2013; Lee et al., 2013b; Lim et al., 2013; Pan et al., 2013); (ii) a pair-wise comparison of closely-related populations with divergent production characteristics (mostly beef vs. dairy breeds, e.g., Hayes et al., 2009; Wiener et al., 2011; Hosokawa et al., 2012; Pintus et al., 2014) and (iii) several breeds, from three (Flori et al., 2009) to 19 breeds (Bovine HapMap Consortium, 2009), of the same or different production characteristics, and for which both across- and within-population analyses are performed. Overall, the selection sweeps considered in this review involved 37 breeds (including 17 beef breeds, 14 dairy breeds, and six dual-purpose breeds), and four different beef-vs.-dairy comparisons (Australian Holstein vs. Australian Angus, Charolais vs. Holstein, Japanese Black vs. Japanese Holstein, Piedmontese vs. Italian Brown) (Supplementary Table S1 in Supplementary Material). In addition, we have considered those selection sweeps reported for Holstein populations from specific geographic regions, such as Italian, Israeli or Chinese Holstein cattle, for Angus and Australian Angus cattle, and for Simmental and its German strain Fleckvieh, as related to "distinct" breeds in order to investigate whether there is evidence for geographical region-specific sweeps for the same breed.

The genotyping platforms used in the considered studies demonstrate the rapid development of livestock genomic tools over the last few years (**Table 1**). The earliest study included, that of Hayes et al. (2009), involved the analysis of 9323 SNPs genotyped by Parallele TM or Affymetrix TM and the Bovine HapMap study (2009) generated data using a custom Affymetrix 10K genotyping chip and Illumina 1536 BeadArray assays (Taylor, personal communication). Additional analyses of the original Bovine HapMap dataset were reported later by Stella et al. (2010) and Wiener et al. (2011). But most of the studies compiled in this review (10 out of 17) are based on the medium density SNP-array platform (∼50K SNPs) provided by the Illumina Bovine SNP50 Genotyping BeadChip (Matukumalli et al., 2009). This SNP-array provides an initial dataset of 54,001 SNPs of which quality control filtering left between 29,848 (Mancini et al., 2014) and 47,651 (Rothammer et al., 2013) markers available for analysis in the 10 studies considered (**Table 1**). The studies of Druet et al. (2013) and Porto-Neto et al. (2014) involved genotyping with the Illumina BovineHD genotyping assay (>770K SNPs), which, after quality control filtering, resulted in 680,000 and 725,293 markers, respectively. In Kemper et al. (2014) the genotypes obtained with the Illumina BovineHD chip were used to perform imputation of a second dataset generated with the Illumina Bovine SNP50v2.0 (Erbe et al., 2012), yielding a total of 616,350 and 692,527 SNPs for analysis within the groups of dairy and beef breeds, respectively. Ramey et al. (2013) used the Illumina's Bovine SNP50 Genotyping BeadChip and a prescreening assay comprising almost 2.8 million SNPs that were used as an initial marker panel in the design of the Affymetrix Axiom Genome-wide BOS 1 assay (AFFXB1P). Finally, some of the most recent studies have used data generated by large-scale sequencing. Lee et al. (2013b) analyzed more than 15 million SNPs identified through the sequencing of 12 genomes of Hanwoo cattle, whereas Qanbari et al. (2014) performed a sequence-based imputation, from a 50K SNP panel bridged by a high-density panel to the full genome sequence of Fleckvieh individuals.

The reports reviewed here have applied different but complementary statistics to detect selection signatures (**Table 1**). We classify the studies in the following categories: (i) studies that have estimated differences in allele frequencies by contrasting pair of breeds through F_{ST} (or related statistics) or by differences in allelic frequencies (Flori et al., 2009; Hayes et al., 2009; Wiener et al., 2011; Hosokawa et al., 2012; Mancini et al., 2014; Pintus et al., 2014; Porto-Neto et al., 2014; the across-breed results of Stella et al., 2010); (ii) studies based on extended regions of low diversity or the calculation of extended haplotype homozygosity (EHH) or variants of this statistic such as Relative Extended Haplotype Homozygosity (REHH), the long-range haplotype (LRH) test, and integrated Haplotype Homozygosity Score (iHS) (Qanbari et al., 2010; Glick et al., 2012; Lim et al., 2013; Pan et al., 2013; Ramey et al., 2013; Rothammer et al., 2013); and (iii) studies based on the allele frequency spectrum, in which regions with outlying allele frequency patterns within

TABLE 1 | Summary of the genome-wide selection mapping studies included in this review.

Study	Original reference genome	Genotyping platform[1]	Number of analyzed SNPs	Number of regions considered in this review[2]	Number of breeds	Breeds[3]	Production category	Analysis method[4]	Origin of data from original study
Boitard and Rocha, 2013	UMD3.1	50K-chip	35,564	3	1	BLO	Beef	HMM-allele frequency	Table 1
Bovine HapMap Consortium, 2009	Btau4.0	AfC-10K 1536-chip	33,326	67	19	BMAS*, ANG, REDA, HER, NORW, HOL, LIM, CHA, BRSW, JER, GUE, PIE, ROM NDAM*, SHE*, NEL*, BRA*, GIR*, SGER*	Dairy, Beef, Dual-purpose	F_{ST}, iHS	Table 1; Table S8
Druet et al., 2013	UMD3.1	HD-chip	725,293	147[a]	12	BB, DBB, HOL, JER, LIM, HER, ANG, CHA, GUE, PIE, ROM, BRSW	Dairy, Beef, Dual-purpose	HMM-average heterozygosity	Tables S1–S12
Flori et al., 2009	Btau4.0	50K-chip	42,486	6	3	HOL, NORM, MONT	Dairy	F_{ST}	Table 2
Glick et al., 2012	Btau4.0	50K-chip	41,814	25	1	ISR-HOL	Dairy	REHH	Table 3
Hayes et al., 2009	UMD3.1	TM-9323	9,323	15	2	AUSHOL-vs.-AUSANG	Beef vs. Dairy	Differences in allele frequencies	Table 2
Hosokawa et al., 2012	Btau4.0	50K-chip	40,635	11	2	JAPBL-vs.-JAHOL	Beef vs. Dairy	Differences in allele frequencies	Table 2
Kemper et al., 2014	UMD3.1	HD-chip (real and imputed)	610,123	30	8	HOL, JER, ANG, CHA, HER, LIM, MUG, SHOR	Dairy, Beef	HAPH, iHS, F_{ST}	Table 2, Table 6
Lee et al., 2013b	UMD3.1	WGS	15,125,420	15	1	HAN	Beef (high intramuscular fat content)	LD-ω, CLR	Table 1
Lim et al., 2013	UMD3.1	AfM-10K	4,522	2		HAN	Beef (high intramuscular fat content)	EHH-iES	Table 1
Mancini et al., 2014	Btau4.0	50K-chip	29,848	5	5	IT-BR, IT-HOL, PIE, MAR, IT-PEZZ	Dairy, Beef, Dual-purpose	F_{ST}	Table 1
Pan et al., 2013	Btau4.0	50K-chip	40,130	16	1	CHI-HOL	Dairy	EHH	Table 3
Pintus et al., 2014	Btau4.0	50K-chip	42,514	53	2	PIE-vs.-ITBR	Beef vs. Dairy	F_{ST} (LOWESS regression correction)	Table 1
Porto-Neto et al., 2014	UMD3.1	HD-chip	680,000	55	9	HAN (against six European breeds, NEL, NDAM)	Beef (high intramuscular fat content)	F_{ST}	Table 2
Qanbari et al., 2010	Btau4.0	50K-chip	41,398	12	1	HOL	Dairy	REHH	Table 4
Qanbari et al., 2011	Btau4.0	50K-chip	40,595	14	10	HOL, BRSW, SIM, AUS-ANG, BELR*, HER, MUG, SGER, SHOR, BRA*	Dairy, Beef, Dual-purpose	iHS, F_{ST}	Table 2
Qanbari et al., 2014	Btau4.0	WGS-based imputation	Sequenced data (15,182,131 SNPs), medium density panel (39,304), high density panel (645,189).	140	1	FLE	Dual-purpose	iHS, CLR	Tables S3, S4

(Continued)

TABLE 1 | Continued

Study	Original reference genome	Genotyping platform[1]	Number of analyzed SNPs	Number of regions considered in this review[2]	Number of breeds	Breeds[3]	Production category	Analysis method[4]	Origin of data from original study
Ramey et al., 2013	UMD3.1	50K-chip AFFXB1P-chip-2787037	52,942 (50kchip dataset); 2,575,339 SNPs (AFFXB1P)	132	15	ANG, BRA, CHA, HAN, HER, LIM, SAL, SHOR, SIM, BRSW, FINA, HOL, JER, WAG, BRA*	Dairy, Beef, Dual-purpose	Extended low diversity haplotypes	Tables 2, 3, 4
Rothammer et al., 2013	UMD3.1	50K-chip	47,651	61[b]	8	GAL, BB, RED-HOL, OBRA, MUR, FRGE, FLE, BRA*	Beef	XP-EEH	Tables S4–S13
Stella et al., 2010	Btau4.0	HMap-data	32,689	215[c]	5	BRSW, GUE, HOL, NORW, JER	Dairy	CLL	Within-breed results[5]
Wiener et al., 2011	Btau4.0	HMap-data	31,312	30[d]	2	CHA-vs.-HOL	Beef vs. Dairy	Difference in allele frequencies	Table S2

[1] High-throughput genotyping arrays used for genotyping in the different studies 50k chip, Illumina's Bovine SNP50 Genotyping BeadChip; 1536-chip, Illumina 1536 BeadArray assays; HD-chip, Illumina Bovine HD chip genotyping assay; AfC-10K, Affymetrix Custom 10K; TM-9323, 9323 SNPs (Parallele TM/Affymetrix TM); HMap-data, data from Bovine HapMap Consortium; WGS, Whole genome sequence; AfM-10K, Affymetrix MegAllele GeneChip Bovine Mapping 10K SNP array; AFFXB1P-chip-2787037, Affymetrix Axiom Genome-wide BOS 1 assay = 2,787,037 SNPs.

[2] Where indicated the initial significant positions reported by the original authors were filtered in this review as described below (number of raw reported sweeps):

[a] From an initial list of 398 positions, only those showing an average heterozygosity < 0.003 were considered herein.

[b] From an initial list of 1251 breed-specific significant positions, only those with consecutive XP-EHH values (gap < 1 Mb) higher than 4 were considered herein.

[c] From the initial results from the within-breed analysis results kindly provided by the authors (a total of 13,077 significant positions at P < 0.01, genome-wide), we selected, for each breed, the top 10% of the positions showing a significant CLL, and then we grouped the positions in signatures regions. Following the authors' criteria, adjacent signature regions were considered "distinct" if they were separated by at least 3 consecutive windows with non-significant CLL (P < 0.01 genome-wide).

[d] From the initial list of 121 significant positions of top 1% MA_d values, the positions were grouped in regions of subsequent significant regions with no more than 1 Mb gaps.

[3] Breed abbreviations: ANG, Angus; AUS-ANG, Australian Angus; BB, Belgian Blue; BLO, Blonde d'Aquitaine; CHA, Charolais; GAL, Galloway; HER, Hereford; HAN, Korean Hanwoo; LIM, Limousin; MAR, Marchigiana; MUG, Murray Gray; PIE, Piedmontese; REDA, Red Angus; ROM, Romagnola; SAL, Salers; SHOR, Shorthorn; WAG, Wagyu; AUSHOL-vs.-AUSANG, Australian Holstein vs. Australian Angus; CHA-vs.-HOL, Charolais vs. Holstein; JAPBL-vs.-JAHOL, Japanese Black vs. Japanese Holstein; PIE-vs.-ITBR, Piedmontese vs. Italian Brown; BRA, Braunvieh; BRSW, Brown Swiss; CHI-HOL, Chinese Holstein; FINA, Finnish Ayrshire; GUE, Guernsey; HOL, Holstein; ISR-HOL, Israeli Holstein; IT-BR, Italian Brown; IT-HOL, Italian Holstein; IT-PEZZ, Italian Pezzeta rosa; JER, Jersey; MONT, Montbéliarde; NORM, Normande; NORW, Norwegian Red; RED-HOL, Red Holstein; DBB, Dual-purpose Belgian blue; FLE, Fleckvieh; FRGE, Franken Gelbvieh; MUR, Murnau-Werdenfelser; OBRA, Original Braunvieh; SIM, Simmental.

[4] Statistical methods used in the original studies reviewed in this work: CLL, Parametric composite log likelihood of the differences in allelic frequencies; CLR, Composite likelihood ratio of a model considering a section sweep as defined by Nielsen et al. (2005); EHH, Extended haplotype homozygosity; F_{ST}, Population differentiation statistic; HAPH, Haplotype homozygosity measurement; HMM allele frequency, Hidden Markov model-based test, local variations in the allele frequency; iHS, Integrated haplotype homozygosity score; LD ω, LD-based omega (ω) statistics; REHH, Relative extended haplotype homozygosity; XP-EEH, Cross population extended haplotype homozygosity.

[5] Boettcher and Stella, personal communication.

[*] Additional breeds analyzed in the original studies but not considered in this review: BMAS, Beefmaster; BELR, Belmont Red; NDAM, N'Dama; SHE, Sheko; NEL, Nelore; BRA, Brahman; GIR, Gir; SGER, Santa Gertrudis.

a single population are identified through various tests (e.g., the CLR, Composite likelihood-ratio test; CLL, parametric composite log likelihood; and HMM, Hidden Markov Model-based test) (Boitard and Rocha, 2013; Druet et al., 2013; the within-breed results of Stella et al., 2010). The studies based on F_{ST} and related statistics (category i) detect diversifying selection between breeds. Of within-breed studies, those based on differences in allele frequency patterns (category iii) have greatest power to detect completed selection (fixation of alleles) whereas the haplotype-based procedures (category ii) have greatest power to detect ongoing selection, as they explore the structure of haplotypes and essentially identify unusually long haplotypes carrying the ancestral and derived alleles (Qanbari et al., 2014). Some of the studies implement two or three different selective sweep mapping methods that fall into multiple categories (Bovine HapMap Consortium, 2009; Qanbari et al., 2011, 2014; Lee et al., 2013b; Kemper et al., 2014) (**Table 1**).

Filtering Criteria and Comparative Approach

In order to look for independent identification of the same regions and to identify those single-breed sweeps that could be uniquely associated with individual breeds, we compiled all the selection signals as reported in the different studies. For some studies reporting both regions identified in specific breeds and also across-breed analyses (Flori et al., 2009; Stella et al., 2010; Qanbari et al., 2011), we only considered the regions reported for specific breeds. The only exception to this criterion was the inclusion in our reviewed dataset of the 12 autosomal regions with extreme F_{ST} value across all populations reported by the Bovine HapMap Consortium (2009). In all cases except one, the details of the selection signatures (Start-End of the region; candidate genes included) were obtained from the original publications (tables in the main text or Supplementary Material); the only exception was the results reported for Stella et al. (2010). In this case, we compiled the genomic positions of the 13,000 significant positions ($P < 0.001$) identified for the five individual breeds (kindly provided by the authors, Boettcher and Stella, personal communication).

For four of the studies for which the original list of significant regions/positions included the results of all the positions/windows exceeding the significance threshold (Stella et al., 2010; Wiener et al., 2011; Druet et al., 2013; Rothammer et al., 2013), we applied additional filtering criteria by selecting the most significant regions or those on the top/bottom of the distribution and/or by grouping close significant positions (within 1 Mb of distance or a distance criteria previously applied by the authors) under the same sweep signals (see **Table 1** for details about the additional filtering applied to these four studies).

An important issue when comparing the results of genomic studies in cattle is related to the use of different versions of the bovine genome assembly. Nine studies were based on the UMD_3.1 reference sequence, the version currently available at Ensembl (http://www.ensembl.org/Bos_taurus/Info/Index) and the UCSC browser (genome.ucsc.edu/). Eleven out of the remaining studies provided results with reference to the previous Btau_4.0 version of the assembly (currently available at http://aug2010.archive.ensembl.org/Bos_taurus/Info/Index) whereas Qanbari et al. (2014) referred to the Btau_4.6.1 version. To make the genomic positions reported by the different studies comparable across studies, we used *LiftOver* (https://genome.ucsc.edu/cgi-bin/hgLiftOver) to translate all genomic positions to the UMD_3.1 assembly. Using default parameters, we automatically obtained the correspondence between Btau4.0/Btau4.6.1 (hereafter referred to as Btau_4) and UMD_3.1 coordinates for 403 out of the 612 regions. For the 209 other Btau_4-based regions for which the *LiftOver* analysis did not yield appropriate UMD_3.1 coordinates, we performed a manual search to provide approximate UMD_3.1 genomic positions (using the closest genes to the positions flanking the selection signal in the Btau_4 region as markers to localize the region in the UMD_3.1 reference genome).

Finally, the list of all reported selection sweeps across the 21 studies, which included a total of 1049 selection sweep regions, was sorted by UMD_3.1 genomic position. With the aim of generating an interpretable set of results, the initial 1049 selection signals were subsequently grouped into core selective sweep (CSS) regions, which were defined as signals within 1 Mb of each other. This criterion was established following a detailed analysis of the regions harboring genes such as *GDF*-8, *MC1R*, and *DGAT1*, with large phenotypic effect and previously identified as being subjected to positive selection. The flanking intervals of the defined CSSs were based on the most proximal and most distal positions of the individual selection signals included in each CSS; the breeds for which individual selection signals were included in each CSS were also noted.

Interpretative Analysis of Selection Sweeps Reported in Cattle

The number of detected selective sweeps varied across the 21 studies reviewed here (**Table 1**). Of the 1049 selection sweeps identified, the greatest number of regions, 215 (∼20%), were obtained from the filtered data of the within-breed analysis provided by Stella et al. (2010) for five specialized dairy cattle breeds. The study from which the next highest number of regions was obtained was Druet et al. (2013) (147 regions; ∼14%), who studied 12 breeds with different production characteristics (dairy, beef, dual-purpose). In contrast, four studies contributed fewer than 10 selective sweeps each to the total list (16 regions in total, ∼1.5% of signals all together). These were based on breeds without wide distributions, such as French dairy breeds (Flori et al., 2009), Blonde d'Aquitaine (Boitard and Rocha, 2013), Hanwoo cattle (Lim et al., 2013), and Italian breeds (Mancini et al., 2014).

By grouping the consecutive selection sweeps reported by the different authors, (allowing gaps no greater than 1 Mb), we defined a total of 409 CSSs across the 29 bovine autosomes (Supplementary Table S2 in Supplementary Material; **Figure 1**), 232 (57%) of which were associated with a single-breed (Single-breed

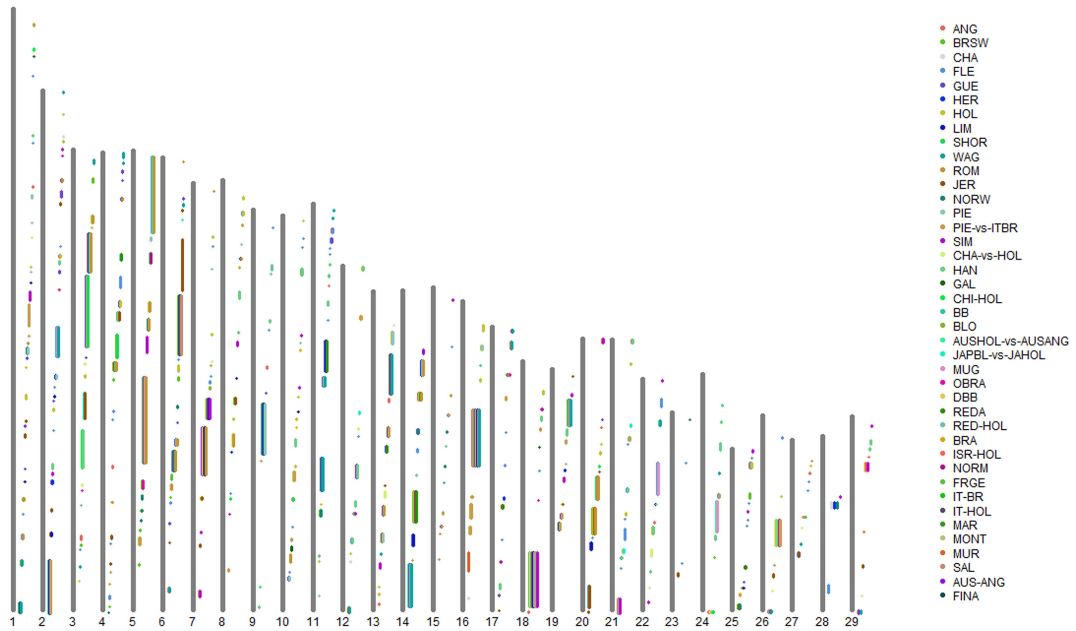


FIGURE 1 | Graphical representation of the Core Selection Sweep (CSS) regions defined in the present work across the 29 bovine autosomes, based on the selection sweeps reported by the 21 genome-wide selection-mapping scans reviewed in the present study. Note that the proximal end of each chromosome (centromere) is represented at the bottom of the plot. ANG, Angus; AUS-ANG, Australian Angus; AUSHOL-vs.-AUSANG, Australian Holstein vs. Australian Angus; BB, Belgian Blue; BLO, Blonde d'Aquitaine; BRA, Braunvieh; BRSW, Brown Swiss; CHA, Charolais; CHA-vs.-HOL, Charolais vs. Holstein; CHI-HOL, Chinese Holstein; DBB, Dual-purpose Belgian blue; FINA, Finnish Ayrshire; FLE, Fleckvieh; FRGE, Franken Gelbvieh; GAL, Galloway; GUE, Guernsey; HER, Hereford; HOL, Holstein; ISR-HOL, Israeli Holstein; IT-BR, Italian Brown; IT-HOL, Italian Holstein; JAPBL-vs.-JAHOL, Japanese Black vs. Japanese Holstein; JER, Jersey; HAN, Korean Hanwoo; LIM, Limousin; MAR, Marchigiana; MONT, Montbéliarde; MUR, Murnau-Werdenfelser; MUG, Murray Gray; NORM, Normande; NORW, Norwegian Red; OBRA, Original Braunvieh; PIE, Piedmontese; PIE-vs.-ITBR, Piedmontese vs. Italian Brown; REDA, Red Angus; RED-HOL, Red Holstein; ROM, Romagnola; SAL, Salers; SHOR, Shorthorn; SIM, Simmental; WAG, Wagyu.

CSSs). For the remaining CSSs, we distinguished between 134 CSSs (33%) associated with a limited number (from 2 to 4) of breeds (76 two-breed CSSs, 42 three-breed CSSs, and 16 four-breed CSSs) and 39 CSSs (9.5%) that were associated with five or more breeds (from 5 to 19 breeds) (Supplementary Table S2 in Supplementary Material). We will refer to these two categories as Two-to-Four-breed CSSs and Multi-breed CSSs, respectively. In addition, four identified CSS regions were only detected in the across-breed F_{ST} analyses reported by the Bovine HapMap Consortium (2009), and will henceforth be referred to as HapMap-Unique CSSs. These four groups of CSSs are indicated by different cell color backgrounds in Supplementary Table S2, which also includes the genes that were highlighted by the original studies as possible candidate targets of the identified selection sweep.

We have also performed a thorough search of plausible candidate genes for the defined CSSs. This involved a systematic extraction of the annotated genes included in the corresponding interval of the UMD_3.1 bovine assembly using BioMart (http://www.ensembl.org/biomart/martview/). Subsequently, a systematic search for functional candidate genes was conducted by searching within CSSs for genes from four lists of genes related to phenotypes for which cattle breeds have been subjected to strong positive selection (a total of 1255 genes). These lists comprised: (i) the database of 449 genes (considering only unique genes) related to milk production and mastitis provided by Ogorevc et al. (2009); (ii) a list of 519 candidate genes for meat production and meat quality derived from the EU funded GemQual project (QLK5 – CT2000-0147; Williams et al., 2009; Sevane et al., 2013, 2014); (iii) a list of 176 genes related to coat color in cattle and other mammals (http://homepage.usask.ca/~schmutz/colors.html; Olson, 1999; Montoliu et al., 2014), and (iv) a list of 111 genes associated with stature and body size in humans and cattle (Pryce et al., 2011; Guo et al., 2012; Kemper et al., 2012) (See Supplementary Table S3 in Supplementary Material for a complete list of the candidate genes associated with the four phenotype groups). Note that some of the genes appear in more than one of the candidate gene lists (e.g., *PPARGC1A*, *MC1R*).

In addition, for the genes extracted for the three main CSS categories (Single-breed, Two-to-Four-breed, and Multi-breed CSSs), we carried out a functional enrichment analysis with WebGestalt (http://bioinfo.vanderbilt.edu/webgestalt/; Wang et al., 2013), using pathways defined by WikyPathways (http://www.wikipathways.org/index.php/Download_Pathways), and selecting "hsapiens_entrezgene_protein-coding" as the reference set and hypergeometric *p*-value and the "top10" options.

Plausible Candidate Genes Underlying CSS Regions

Overall Results

The BioMart analysis extracted a total of 5182 genes from the 409 CSSs (Supplementary Table S4 in Supplementary Material). The number of genes extracted for the CSSs was proportional to the length of the genomic intervals involved in the CSSs (**Table 2**). Hence, a larger number of genes was extracted for the Multi-breed CSSs (2440 genes), which spanned a total of 264.05 Mb across 20 out of the 29 bovine autosomes. From the Two-to-Four-breed CSSs, which involved 202.95 Mb across all the bovine autosomes except BTA23, we extracted 1886 genes (Supplementary Table S4 in Supplementary Material).

Although more than half of the defined CSSs were associated with a single breed, and these were located across all the autosomes, the breed-specific selection sweeps spanned a shorter genomic length (73.01 Mb) and thus included a smaller number of genes (839). Seventeen of the extracted genes were located within the four regions uniquely detected in the HapMap project (**Table 2**). Of the 5183 genes found within CSSs, 291 of them were included in the four lists of phenotype-related candidate genes. The number of candidate genes mapping within the CSS categories defined were two (HapMap-unique CSSs), 67 (Single-breed CSSs), 83 (Two-to-Four-breed CSSs), and 139 (Multi-breed CSSs) (**Table 2**). The number of candidate genes for dairy, beef or body-size related traits was similar among the Single-, Two-to-Four-, and Multi-breed CSS categories, whereas coat color genes were mainly detected in the CSSs involving more than one breed (18 in the Multi-breed-CSS group and 19 in the Two-to-Four-breed category) (**Table 2**). Considering the three main CSS categories (Single-breed, Two-to-Four-breed, and Multi-breed CSSs), the candidate genes were not over-represented in the genes located within CSSs although the subset of dairy-related genes was slightly over-represented (but not significantly, based on Fisher's Exact Test). When the same analysis was done separately at the Single-breed CSSs and other CSSs (merging the Two-to-Four-breed and the Multi-breed CSSs together), the Single-breed category was significantly enriched (Fisher's Exact Test, $p = 0.006$) for production genes (beef and dairy) (0.07 of genes are production genes versus 0.05 for the genome overall), whereas the CSSs involving more than one breed were not significantly enriched for production genes (only 0.04 of genes are production genes).

The candidate genes highlighted by this survey are detailed in Supplementary Table S2 within the corresponding CSS where they are included. The gene symbol is indicated with different font color depending on the database of candidates from which it was identified (blue = "dairy-related," red = beef-related," green = "coat-color-related," and pink = "stature/body-size-related").

Single-Breed CSSs

The 232 single-breed CSSs identified corresponded to selection signals reported in beef (54), dairy (87) and dual-purpose (64) cattle breeds, and 28 of them were reported in a beef vs. dairy pair-wise comparison (Supplementary Table S2 in Supplementary Material). Fleckvieh showed the largest number (49) of these breed-specific selection sweeps, followed by Holstein (33 CSSs), Korean Hanwoo cattle (22), Jersey (14), Guernsey (13), and Simmental (10). Most of the Fleckvieh-specific CSSs breed were reported by Qanbari et al. (2014), and the Korean Hanwoo-related ones were reported by Porto-Neto et al. (2014). The uniqueness of these regions may be biased due to the higher marker density of these studies, which were based on whole genome sequence and the HD-chip dataset, respectively, compared with studies performed in other breeds, which were based on the lower-density SNP panels. The 33 Holstein-specific regions, however, were also extracted from studies based on lower density panels (Qanbari et al., 2010; Stella et al., 2010) and thus their abundance does not appear to be an artifact of the methodology. The large number of such CSSs may be directly related to the very strong selection and resulting high level of dairy specialization in this breed. It is not possible to present a detailed discussion of the each of the single-breed regions and associated candidate genes, but we discuss below some of the regions for which plausible candidate genes could be identified.

A number of dairy-related candidate genes were identified in Holstein-specific regions. For example, several genes related to the immune response were located in Holstein-specific regions,

TABLE 2 | Characterization of the four categories of core selection sweeps (CSSs) defined in this review based on the number of associated breeds.

	Single-breed CSS	Two-to-Four-breed-CSS	Multibreed-CSS	HapMap-Unique-CSS	Total
Number of CSS	232	134	39	4	409
Average CSS-interval length (Mb)	73.01	202.95	264.05	1.21	541.22
Total number of extracted genes[a]	839	1886	2440	17	5182
Number of candidate genes[b]	67	83	139	2	291
Color	3	19	18	–	40
Dairy-Mastitis	36	38	54	1	129
Beef	21	19	55	1	96
Stature/Body-size	7	7	12	–	26

[a] *Genes extracted with the web-based BioMart tool available at ensembl.org and based on the UMD_3.1 bovine genome assembly.*

[b] *Candidate genes identified through the candidate gene survey performed in this study for four phenotype classes under putative selection in modern cattle breeds.*

including *IL12B* (Subunit beta of interleukin 12), (CSS-149), which is a cytokine expressed by activated macrophages that has been found to be expressed in milk somatic cells during intramammary infections (Lee et al., 2006), and *TLR4* (CSS-171), for which polymorphisms have been associated with mastitis (Wang et al., 2007; de Mesquita et al., 2012) and somatic cell score in cattle (Li et al., 2014b; Wang et al., 2014a). Finally, CSS-298 includes two genes expressed in the mammary gland, *G0S2* and *LAMB3* (Ron et al., 2007). These two genes are also associated with fat metabolism in cattle (Lee et al., 2013a; Ahn et al., 2014) and therefore may be linked to the fat mobilization related to high dairy production.

For the Chinese and Israeli Holsteins, four and seven population-specific CSSs were observed, respectively. None of the Single-breed CSSs were linked to Italian Holsteins, whose selection sweeps were shared with the general Holstein population. Despite the world-wide spread of the Holstein breed, the different conditions in which the animals are reared in some of the countries e.g., resistance to heat stress in Israeli Holstein (Flamenbaum and Galon, 2010), may underlie some of these population-specific CSSs. Apart from the candidate genes suggested by Pan et al. (2013) in the Chinese study, our candidate gene survey did not detect any additional genes associated with known cattle phenotypes. Regarding this point, it should be noted that the different Holstein subpopulations shared CSSs involving major dairy candidate genes, such as the *DGAT1* (CSS-251), *ABCG2* (CSS-123), and *PLAG1* (CSS-254) genes, all of them classified as Multibreed-CSSs. For the other CSSs involving more than one breed, the Chinese Holstein was found independently of other Holstein populations in two cases included in the Two-to-Four-breed category: CSS-110 (shared with Brown Swiss, Fleckvieh, Simmental) and CSS-34 (shared with Jersey), and the Multibreed CSS-67 region (shared with Guernsey, Jersey, Korean Hanwoo, and Angus).

For two other dairy breeds, Jersey and Guernsey, several breed-specific selection sweeps were identified (14 and 13, respectively). One Guersey-related sweep (CSS-118) includes the *NFKB1* gene, whose liver expression is altered in response to prepartum energy intake and post-partum intramammary inflammatory challenge in dairy cows (Graugnard et al., 2013). A Jersey-related sweep (CSS-383) includes *PTEN* (phosphatase and tensin homolog) which encodes a tumor suppressor gene regulating many cellular processes, including growth, adhesion, and apoptosis. *PTEN* has also recently been shown to function as an inhibitor during mammary gland development and lactation in dairy cows (Wang et al., 2014b). At the pathway level, the PTEN-AKT pathway is required for the initiation of lactation through the induction of autocrine prolactin (Chen et al., 2012). In addition, *PTEN* has been shown also to play a vital role in regulating fatty acid metabolism (Fu et al., 2012).

A number of Single-breed CSSs were identified in beef cattle breeds. There were several beef-related candidate genes located in the Angus-associated CSS-63, including *CTSK, CTSS* (cathepsin K and S), and *TMOD4* (tropomodulin 4). CTSS (cathepsin S) is known to be involved in antigen presentation and also cleaves some extracellular matrix proteins. Through its physiological role, which is to degrade type I collagen, CTSK appears to regulate adipocyte differentiation in adipose tissues of obese patients and animal models (Xiao et al., 2006; Han et al., 2009).

Beef-related candidate genes are located in several of the Korean Hanwoo-specific selection sweeps, including *ITGB3* (β3 integrin; CSS-313), which is involved in cytoskeletal organization and plays a role in the adhesion between the cell cytoskeleton and cell extracellular matrix. During postmortem aging, degradation of integrin has been found to be associated with increased drip loss in pork (Lawson, 2004), suggesting it may also be related to meat quality traits in cattle. Furthermore, *MC2R* (adrenocorticotropin receptor) and *MC5R* (melanocortin 5 receptor) are located in the Hanwoo-specific CSS-365. *MC2R* encodes a receptor for the adrenocorticotropic hormone which plays a crucial role in the regulation of glucocorticoid secretion, while *MC5R* is involved in lipid metabolism, exocrine function, and proinflammatory activity (reviewed by Switonski et al., 2013). In addition, *MC5R* expression down-regulates leptin secretion in cultured adipocytes and in humans *MC5R* polymorphisms were reported to be associated with obesity (Switonski et al., 2013). In pigs, *MC2R* is located within a QTL region for intramuscular fat content and back fat thickness (Jacobs et al., 2002) and *MC5R* is close to a QTL influencing fatness and meat quality. Several reports have confirmed an association between porcine back fat thickness or feed intake and variants of the *MC5R* gene (Emnett et al., 2001; Kováčik et al., 2012).

Several Fleckvieh-specific CSSs also include functional candidate genes. The *MFGE8* (milk fat globule-EGF factor 8 protein), located in CSS-334, has been reported to be associated with an index assessing productivity and functional and conformation traits (Fontanesi et al., 2014), which may be relevant to the dual-purpose production characteristics of this breed. The same CSS also includes *ISG20* (interferon stimulated exonuclease gene 20 kDa), which is involved in cumulus oocyte growth and may be related to fertility (Puglisi et al., 2013).

Another Fleckvieh-specific region (CSS-352) includes the *ATP2B2* gene (plasma membrane Ca(2+)-ATPase). The protein encoded by this gene is involved in the transport of calcium across the mammary cell apical membrane. This protein is related to calcium-mediated cell death and has been suggested to play a part in early signaling of mammary gland involution (Reinhardt and Lippolis, 2009).

Two-to-Four-Breed CSSs

Mapping within the Two-to-Four-breed CSS intervals, we found a high proportion of coat-color related genes (22% of the 93 candidate genes associated with these regions) including *KITLG* (*KIT*-ligand, also known as mast cell growth factor) (CSS-103; identified in Hereford, Holstein, Normande, and a Piedmontese vs. Italian Brown comparison) and *MITF* (microphthalmia-associated transcription factor) (CSS-350; identified in Fleckvieh and Murray Gray), both of which are known to be associated with coat color in cattle (Seitz et al., 1999; Hayes et al., 2010).

Several other genes that have been associated with coat color phenotypes in species other than cattle fall in Two-to-Four breed CSSs, including *HS2ST1* (CSS-68; identified in Fleckvieh, Guernsey, Japanese Black vs. Japanese Holstein, Jersey), *AP3B1*

(CSS-183; Guernsey, Piedmontese), *MAP2K1* (CSS-185; Holstein, Piedmontese, Romagnola), *MYC* (CSS-252; identified in Holstein, Piedmontese vs. Italian Brown), *PTS* (CSS-263; identified in Guernsey, Piedmontese vs. Italian Brown), *PDPK1* (CSS-368; identified in Brown Swiss, Jersey, Norwegian Red), and *ERCC2* (CSS-303; Angus, Simmental).

An interesting region (CSS-131) identified within the Two-to-Four-breed CSS category is that harboring the bovine casein gene cluster on BTA6 (84.66–97.99 Mb). The selection sweeps included in this CSS were identified in three dairy breeds: Braunvieh, Israeli Holstein, and Jersey. Caseins (CSN1S1, CSN1S2, CSN2, CSN3, etc.) represent the primary protein constituents of cow's milk (approximately 80%). The amount and allelic variants of caseins are associated with clotting properties and cheese yield (Wedholm et al., 2006). Due to the importance of caseins in milk production, it is intriguing that only three out of 14 dairy breeds included in this study show a selective sweep near the casein cluster. Nevertheless, this observation agrees with the discordant results reported in the 1980s and 1990s regarding the association of specific casein alleles with production traits, which appear to be breed-specific (reviewed by Caroli et al., 2009).

The *LEP* (leptin) gene appears as a strong candidate gene underlying the selection sweeps reported for one dairy (Guernsey) and two beef breeds (Piedmontese, Red Angus) (CSS-96). Leptin regulates feed intake and energy balance in mammals (Houseknecht et al., 1998) and is involved in the regulation of nutritional status and reproductive functions. Polymorphisms in the bovine *LEP* gene are associated with feed intake (Lagonigro et al., 2003) as well as production traits in both beef (Woronuk et al., 2012) and dairy cattle (Liefers et al., 2002).

Multi-breed CSSs

Only 39 of the 409 CSSs defined herein involved at least five breeds. As was observed for the Two-to-Four-breed CSSs, the number of CSSs generally decreased as the number of breeds associated with the CSS increased. Hence, we found 12 five-breed CSSs, six six-breed CSSs, three seven-breed CSSs, three eight-breed CSS, and five nine-breed CSSs, and 10 CSSs involving 10–19 breeds.

The two CSSs involving the largest number of breeds were located on BTA6 (CSS-123) and BTA16 (CSS-278), and included 18 and 19 breeds (or pair of breeds), respectively, out of the 41 breed/breed pairs considered in this study (Supplementary Table S2 in Supplementary Material). CSS-123 involved selection sweeps reported in a large number of dairy breeds (Brown Swiss, Chinese Holstein, Guernsey, Holstein, Italian Brown, Italian Holstein, Jersey, Norwegian Red, Montbéliarde) but also beef production (Angus, Hereford, Romagnola, Piedmontese, Marchigiana) and dual-purpose (Fleckieh, MurnauWerdenfelser, OriginalBraunvieh) breeds (a selective signal was also identified for Piedmontese vs. Italian Brown, a beef-dairy comparison). This CSS includes the *ABCG2* (ATP-Binding Cassette, Sub-Family G Member 2) gene, which harbors a QTN for milk composition previously reported in cattle (Olsen et al., 2007). The precise role this gene plays in milk compositions was not initially understood but a later study suggested that *ABCG2* plays a role in mammary epithelial cell proliferation and that functional polymorphisms in this gene may influence the cellular compartment of the mammary gland and potentially milk production (Wei et al., 2012). This interval also includes the *SPP1* (osteopontin) gene, which has been shown to have significant role in the modulation of milk protein gene expression (Sheehy et al., 2009) and whose allelic variants have also been shown to be associated with variation in milk compositions (Leonard et al., 2005; Khatib et al., 2007). Possibly due to its role as a cytokine, osteopontin has been shown to be beneficial for reducing the incidence of infection during the transition period in lactating cows (Dudemaine et al., 2014). As mentioned above, CSS-123 was also identified in major beef production breeds. In this regard, the *NCAPG* gene, also located in this genomic region, harbors a causal mutation (I442M) related to fetal growth, carcass performance, and body frame size in cattle (Eberlein et al., 2009; Setoguchi et al., 2009, 2011). Interestingly, a later study has also shown a possible association of this polymorphism on milk production traits (Weikard et al., 2012). *NCAPG* overlaps with the *LCORL* (ligand dependent nuclear receptor corepressor-like) gene and in many cases these two genes are jointly referred to as *LCORL/NCAPG*. The *LCORL/NCAPG* locus influences feed intake, gain, meat and carcass traits in beef cattle (Lindholm-Perry et al., 2011) and has been associated with human height (Soranzo et al., 2009; Lango-Allen et al., 2010) and withers height in horses (Tetens et al., 2013). Another notable gene located within CSS-123 is *LAP3* (leucine aminopeptidase 3), which has been associated with milk production traits (Zheng et al., 2011). The region involving *NCAPG*, *LCORL*, and *LAP3* genes has been associated with calving ease in Norwegian Red dairy cows (Olsen et al., 2008) and in Piedmontese beef cattle (Bongiorni et al., 2012). The results in the latter breed suggest that selection on *LAP3* for better calving ease is driving the selection signature in this region. Therefore, the large number of breeds included in CSS-123 probably results from the presence of multiple genes influencing various traits of economic interest in cattle.

The region associated with 19 different breeds was CSS-278 on BTA16 (38.500–53.307 Mb). Although it involves selection sweeps reported in a large number of beef-related breeds (Angus, Australian Angus, Charolais, Hereford, Korean Hanwoo, Limousin, Piedmontese, Red Angus, Salers, Shorthorn), it was also related to dairy (Brown Swiss, Guernsey, Holstein, Jersey) and dual-purpose (Simmental, Fleckvieh, FrankenGelbvieh, Braunvieh) breeds. The BioMart extraction for this CSS interval included 253 annotated genes among which we did not identify any gene with known major effects. The genes suggested by the corresponding authors for the selection sweeps included in this CSS involve several genes related to different biological functions: immune response (*PIK3CD*, *SPSB1, ISG15, TNFRSF9*), development *(RERE*), lipid transportation (*GLTPD1)* muscle physiology (*AGRN)*, and apoptosis (programmed cell death) (*FASLG, TNFRSF1B, DFFB, TNFRSF25, DFFA, CASP9*). Among these, *CASP9* (caspase 9) is the strongest candidate as it belongs to a subgroup of proteases involved in the phase of apoptosis initiation that occurs in the post-mortem conditioning period and that, together with the calpain system, influences the ultimate meat tenderness (Ouali et al., 2006).

Our candidate gene survey in relation to CSS-278 also identified one dairy-related gene (*PEX14*, peroxisomal biogenesis factor 14), genes related to muscle physiology within the beef-candidate list (*SLC2A5,* solute carrier family 2 member 5; *TNNT2*, troponin T type 2, cardiac; *TNNI1*, troponin I type 1, skeletal, slow; *SKI*, v-ski avian sarcoma viral oncogene homolog: and *CTRC*, caldecrin), and one gene related to coat color (*ZBTB17*, Zinc Finger And BTB Domain Containing 17). *ZBTB17* is required for hair follicle structure and hair morphogenesis, and mutations in the murine gene are associated with darkened coat, dark skin, dark dermis around hairs, and abnormal follicles.

Two Multi-breed CSSs regions on BTA14 were identified in 13 and 14 breeds. One of these regions was located on the proximal end of the chromosome (CSS-251, 1.657–12.713 Mb) and involved both dairy and beef breeds (Angus, Australian Holstein vs. Australian Angus, Charolais, Charolais vs. Holstein, Chinese Holstein, Guernsey, HapMap project, Hereford, Holstein, Jersey, Korean Hanwoo, Limousin, Norwegian Red, Piedmontese, Wagyu). It is highly likely that CSS-251 incorporates the selection sweep reported in relation to the *DGAT1* (diacylglycerol O-acyltransferase 1) gene for many dairy cattle breeds, based on the causal role of the mutation K232A on milk composition (Grisart et al., 2002). In addition, the *DGAT1* gene has also been associated with carcass and meat quality traits in beef cattle (Thaller et al., 2003; Wu et al., 2012; Avilés et al., 2013). However, the *DGAT1* gene is located at the very proximal end of the chromosome (1.795–1.805 Mb), indicating that the large CSS-251 interval incorporates selection sweeps related to other genes, such as that reported near the *TG* (thyroglobulin; located at 9.262–9.509 Mb) gene. *TG* is known to influence carcass and meat quality traits in beef cattle (Gan et al., 2008; Bennett et al., 2013). Another gene highlighted by our candidate gene query in this region is *CYP11B1* (Cytochrome P450, Family 11, Subfamily B, Polypeptide 1), which influences energy metabolism. A study in German Holstein cattle has shown that SNPs in this gene are associated with milk production traits and somatic cell score independently of the *DGAT1* genotype (Kaupe et al., 2007).

The other CSS on BTA14 associated with a large number of breeds was CSS-254 (23.885–31.847 Mb). The related breeds included both dairy (Brown Swiss, Chinese Holstein, Holstein, Jersey, Montbéliarde, Normande, Norwegian Red) and beef (Angus, Braunvieh, Charolais, Limousin, Piedmontese, Red Angus) breeds as well as the dual-purpose Fleckvieh. Within this region we found the *PLAG1* (pleiomorphic adenoma gene 1) gene, which has been shown to be associated with stature in Jersey × Holstein crosses (Karim et al., 2011) but also shows pleiotropic effects on fertility such that the *PLAG1* allele associated with increased height and weight was also associated with reduced fat, greater feed intake, less residual feed intake, later puberty in both sexes, and longer post-partum interval before reconceiving in cows (Fortes et al., 2013). This region also encompasses a cluster of genes, including *CHCHD7* (coiled-coil-helix-coiled-coil-helix domain containing 7), *SDR16C5* (short chain dehydrogenase/reductase family 16C, member 5), *MOS* (v-mos Moloney murine sarcoma viral oncogene homolog), *LYN* (v-yes-1 Yamaguchi sarcoma viral related oncogene homolog), *PENK* (proenkephalin), and *RPS20* (ribosomal protein S20), that have been associated with stature in cattle and humans (Utsunomiya et al., 2013b). In particular, a polymorphism ablating a polyadenylation signal of *RPS20* has been proposed as the candidate causal mutation of a QTL influencing calving ease and stillbirth incidence in the Fleckvieh breed (Pausch et al., 2011). Another possible candidate for that CSS is *NCOA2* (nuclear receptor coactivator 2), which encodes a transcriptional coactivator for steroid receptors and nuclear receptor and has been found to influence puberty in tropical breeds of beef cattle (Fortes et al., 2011).

Also on BTA6, CSS-130 (67.850–83.375 Mb) involved selection signatures identified in 13 different breeds, involving dairy, beef, and dual-purpose cattle breeds. This region includes a cluster of tyrosine kinase receptor genes (*PDGFRA, KIT*, and *KDR*). The *KIT* (the Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog) gene, which is centered in the CSS interval (71.796–71.917 Mb), explains a considerable proportion of the variation in patterned pigmentation (Hayes et al., 2010), such as the characteristic spotting phenotype of Holstein and other dairy breeds. Close to the *KIT* gene, at 71.374–71.421 Mb, *PDGFRA* (platelet-derived growth factor alpha receptor), has recently been identified as the strongest positional candidate for the non-*MC1R*-related reddening phenotype in an F2 Nellore-Angus population (Hanna et al., 2014). Other coat color genes were also located in Multi-breed CSSs. CSS-297, which contains the *MC1R* gene, was identified in 15 different breeds. Polymorphisms in this gene are related to the production of eumelanin and phaeomelanin pigments and determine the red-black axis in cattle coat color (Robbins et al., 1993). In addition, *MC1R*, through a competitive relationship for alpha-melanocyte stimulating hormone (α-MSH) with the Melanocortin 4 Receptor (appetite suppressing receptor), has been associated with growth and carcass traits (McLean and Schmutz, 2009). Some of the selection signals of the large CSS-297 interval are likely to be related to other genes. A large list of genes has been suggested by the corresponding authors, and our candidate gene survey also highlighted candidates for dairy traits (*SLC7A5*, solute carrier family 7 member 5), meat production (*CTRB1*; chymotrypsinogen B1; *FOXC2*, forkhead box C2; *CDH15*, cadherin 15, type 1, M-cadherin), and stature (*GALNS*; galactosamine (N-acetyl)-6-sulfate sulfatase).

Another coat color gene, *SILV* (*silver*), also known as *PMEL* (premelanosome protein), is located within CSS-109, related to 11 breeds (breed pairs). This gene has been associated with the white coat color characteristic of the Charolais breed (Gutiérrez-Gil et al., 2007; Kuehn and Weikard, 2007a) and pale Highland cattle (Schmutz and Dreger, 2013). It also may be involved in the gray coat phenotype of the Murray Gray breed. In addition, it has been suggested that this gene, based on its multiple splice variants expressed in a variety of tissues independent of pigmentation, could have functions other than melanosome development (Kuehn and Weikard, 2007b). The proximal section of CSS-109 also includes the *IFNG* (interferon, gamma) gene, of interest due to its relationship with the immune response.

The CSS-248 region, which was identified in seven breeds including beef breeds (Korean Hanwoo, Marchigiana, Piedmontese, Shorthorn, Simmental, Wagyu), one dairy breed (Holstein)

and a beef vs. dairy comparison (Japanese Black vs. Japanese Holstein), includes the *ASIP* (agouti signaling protein) gene. Although this is a color-related gene in many species, mutations in the *ASIP* coding region have not been found to play an important role in coat color variation in cattle (Royo et al., 2005). However, a transcript variant of *ASIP* has been assumed to be the causal variant for the brindle coat color of Normande cattle (Girardot et al., 2006) and due to the expression of this gene in adipocytes and its implication in the obese yellow mouse, this transcript has also been suggested to be related to the milk composition traits in this dairy breed (Girardot et al., 2006) and intramuscular fat content in other breeds (Albrecht et al., 2012).

In addition to coat color and patterning, the presence or absence of horns is a breed hallmark in European *B. taurus*. The locus controlling the polled phenotype, *POLLED*, is located within CSS-1, on BTA1 (0.198–2.60 Mb) (Brenneman et al., 1996), which was identified in 10 breeds. The molecular basis of this phenotype has proven to be complex and the existence of allelic heterogeneity has been suggested for this locus, with the candidate causal mutations located outside known genes or regulatory regions (Drögemüller et al., 2005; Medugorac et al., 2012; Allais-Bonnet et al., 2013). Recently, a long intergenic noncoding RNA has been suggested as the most probable cause of horn bud agenesis for one of the defined allelic variants (Allais-Bonnet et al., 2013).

Another Multi-breed CSS including a gene with a major effect on a bovine qualitative phenotype is CSS-32, located at the proximal end of BTA2, and including selection sweeps described in Belgian Blue, Blonde d'Aquitaine, Limousin, Piedmontese, and a Piedmontese vs. Italian Brown breed-comparison. These are all breeds known to show disruptive or missense mutations in the myostatin (*GDF8* or *MSTN*) gene, associated with muscle conformation and in extreme cases, "double muscling" (Grobet et al., 1997; McPherron and Lee, 1997; Smith et al., 2000; Boitard and Rocha, 2013). CSS-32 encompasses the myostatin gene (2: 6.213–6.220 Mb) but also extends over a large region of BTA2 (0–13.850 Mb), due to the long selective sweep reported by Kemper et al. (2014) for Limousin, known to show very high frequency (∼94.2%) of the *GDF8*-F94L mutation (Vankan et al., 2010), whereas for the other breeds, the selective sweep was closer to the *GDF*-8 location. As a result of the large size of this CSS, selective sweeps that originally did not include the *GDF*-8 gene have been incorporated under the CSS-32 label; these include two sweeps described in a Piedmontese vs. Italian Brown comparison (Pintus et al., 2014) 6.717–9.760 Mb, including the *SLC40A1* (solute carrier family 40, member 1), *COL5A2* (collagen, type V, alpha 2), *COL3A1* (collagen, type III, alpha 1), *CALCRL* (calcitonin receptor-like) and *ITGAV* (integrin, alpha V) genes, and a Fleckvieh selective sweep located in a gene-desert region. Polymorphisms in the *SLC40A1* gene have been related to beef iron content (Duan et al., 2012).

Other Multi-breed regions include genes associated with production traits. These include the short CSS-124 region, which was identified in eight breeds and includes the *PPARGC1A* (peroxysome proliferator-activated receptor-γ coactivator-1α) gene, which mediates expression of genes involved in oxidative metabolism, adipogenesis, and gluconeogenesis (Puigserver and Spiegelman, 2003). Expression of this gene has been suggested to be required for the initiation and development of lactation in dairy cattle (Weikard et al., 2005). *PPARGC1A* has also been shown to be associated with milk composition (Weikard et al., 2005; Khatib et al., 2007; Schennink et al., 2009), reproduction (Komisarek and Walendowska, 2012), growth (Li et al., 2014a), carcass traits (Shin and Chung, 2013; Ramayo-Caldas et al., 2014), and meat quality (Sevane et al., 2013).

CSS-72 (identified in eight breeds) includes *LEPR (leptin receptor),* which due to its interaction with leptin, may be a target of selection in relation to a wide range of economically relevant traits, including growth (Guo et al., 2008), milk production (Suchocki et al., 2010), and calving interval (Trakovická et al., 2013). Finally, CSS-314 (identified in six breeds) includes the *FASN* gene, which has been associated with milk and beef fatty acid composition (Roy et al., 2006; Morris et al., 2007; Zhang et al., 2008). CSS-322 (identified in five breeds) includes *GHR* (growth hormone receptor), which has been shown to harbor a causal mutation of a QTL influencing milk yield and composition (Blott et al., 2003) and *FST* (follistatin), which encodes a protein related to ovary function and has also been suggested to play a key role in regulating bovine mammary branching morphogenesis and epithelial differentiation (Bloise et al., 2010).

We acknowledge that our candidate gene survey did not take into account genes related to the immune response and behavior, which are found in various CSSs, as reported by many of the original studies reviewed here. For example, the list of Biomart-extracted genes from all the CSS defined in this study (Supplementary Table S4 in Supplementary Material) includes genes directly associated with the immune response. Hence, the list includes 34 genes encoding proteins related to interferon and interleukin responses (16, 12, and five of them belonging to the Multi-breed, Two-to-Four-breed, and Single-breed CSSs, respectively). The CSSs also include 128 genes encoding olfactory receptors (106 of them within Multi-breed CSSs, and 11 in each of the two other categories) and 28 encoding olfactory receptor-like proteins (22 in the Multi-breed CSSs, five in the Single-breed category and one in the Two-to-Four CSSs), which are proposed to be associated with behavioral traits modified through domestication in cattle (Ramey et al., 2013) and other livestock species (Bovine HapMap Consortium, 2009).

Although our survey and the original papers have identified clear candidates for some of the Multibreed CSS regions (mainly genes influencing morphological traits but also some genes with large effects on production traits), it is worth noting that due to our method of merging multiple selection signals at similar positions under the same label, some of these CSSs involve a much larger region than that directly related to the gene with the major gene effect (as discussed above regarding the myostatin region).

Enrichment Analysis

In an attempt to highlight genes influencing traits other than those considered in our candidate gene survey and to identify the functional biological pathways that are over-represented in the genes included in the CSSs, we performed a complementary functional enrichment analysis for the genes extracted from the

Single-breed, Two-to-Four-breed and Multi-breed CSS regions (Supplementary Table S5 in Supplementary Material).

Among the top 10 significant pathway terms in the Single-breed CSSs, five terms were related to the immune response [regulation of Toll-like receptor signaling, IL-1 and IL-4 signaling, and leucocyte-related validated miRNA (defined by tarBase database) pathways], and the others were linked to global metabolism (leptin signaling pathway), bone and muscle physiology (RNAKL-RANK signaling, osteopontin signaling pathways) and to one of the most important intracellular signal transduction pathways (MAPK signaling pathway).

To further explore these results and assess whether single-breed CSSs are linked to genes underlying the physiology of the production specialization for which they have been selected, we performed the functional enrichment analysis of the Single-breed CSSs, separately for the beef and dairy breeds and also for the dual-purpose breeds (**Table 3**). Whereas pathway terms related to the general immune response were found at similar proportions within the 10 top terms of the three subcategories (although Toll-like receptor signaling pathways terms were only identified in the analysis of the dairy CSSs), other pathway terms appeared to be subcategory-specific. For example, bone and muscle physiology-related terms constituted the majority (5/10) of the top 10 significant terms for beef breed CSSs (i.e., osteopontin signaling, RANKL-RANK signaling pathway, endochondral ossification, osteoclast signaling, striated muscle contraction) whereas those related to major metabolic pathways (leptin signaling pathway, vitamin D synthesis, insulin signaling) were found within the top 10 significant terms only in the dairy-related CSSs. For the Single-breed CSSs associated with dual-purpose breeds, the top 10 significant terms were mainly related to cell signaling pathways involving two important mitogen-activated protein kinases (*MAP2K2, MAPK3*), which are linked to pathways involving receptors of serotonin and histamine (see **Table 3**). Whereas serotonin is a local regulator in the mammary gland that regulates lactation and initiates the transition into the earliest phases of the involution process related to the return of the mammary gland to morphologically near pre-pregnant state (Horseman and Collier, 2014), the histamine receptors may, in addition to their involvement in local immune responses, also show central effects on modulation of behavior related to the biological function of histamine as a neurotransmitter in the central nervous system (Schneider et al., 2014). The analysis of the dual-purpose-related Single-breed CSSs also revealed over-representation of genes involved in myometrial relaxation and contraction pathways, which could be related to the selection of females that are good dairy cows and can also give birth to calves with meat-production characteristics (e.g., large size).

In the Two-to-Four-Breed CSSs the functional gene enrichment analysis (Supplementary Table S5 in Supplementary Material) highlighted three pathway terms related to global metabolism (insulin signaling, glucuronidation, and metapathway biotransformation; the latter term involves several enzymes from the cytochrome P450 superfamily of enzymes, sulfotransferases, and glucuronosyltransferases) and others were related to the immune response [regulation of the Toll-like receptor signaling pathway and lymphocyte-validated miRNAs (TarBase)], cell adhesion mechanisms (integrin-mediated cell adhesion, focal adhesion), and specific cell physiology pathways [MAPK signaling, epithelium-related validated miRNA (TarBase), and microRNAs in cardiomyocyte hypertrophy]. The enrichment analysis performed for the Multi-breed CSSs highlighted among the top 10 significant terms, two related to the immune response (complement activation, classical pathway, complement and coagulation cascades), two related to overall lipid-metabolism (adipogenesis, SREBF, and miR33 in cholesterol and lipid homeostasis), and others related to skeleton and reproductive physiology (regulation of actin cytoskeleton, and myometrial relaxation and contraction pathways).

Because of the large number of genes highlighted by this functional analysis, we do not present here a detailed discussion about the known effects of these genes in cattle. This could be the objective of future studies focusing on some of the CSS regions presented here.

Overall Conclusions

Compilation of the results from many selection sweep mapping studies in cattle provides an ideal opportunity to investigate how artificial selection has influenced the variability and architecture of the bovine genome. Selection is likely to have eroded the levels of genetic variation that existed in the original domesticated population. At the same time, selection on a livestock breed has tended to fix specific variants that have become distinctive genetic signals of that breed compared with others. Strong selection for improvement of productivity, such as milk or beef production traits, has led to specialization of cattle breeds. It might be expected that breeds that share the same production characteristics would show a similar picture of selection sweeps related to such specialization, and conversely, that divergently specialized breeds would share few selection sweeps. However, our review shows that in many cases selection signatures are also shared by breeds showing different production characteristics. These may include regions containing genes associated with metabolic homeostasis or other general traits such as disease resistance and behavior, but may also reflect the pleiotropic effects of genes on traits relevant to both beef and dairy production. Because of the large number of selective sweeps compared here, we have not performed a detailed analysis of all genes included within the CSSs, although in a number of cases, it was possible to speculate as to which gene or genes could be the targets of selection.

This review presents an initial comparative map of the selection sweeps reported in European *Bos taurus* cattle breeds and provides an integrated dataset that can also incorporate results from future studies and thus allow the researchers to perform systematic comparisons of selection sweeps reported in cattle. This type of comparative tool is essential to properly interpret the results of individual studies for such a complex topic as selection sweeps across different breeds of the same species.

Considering the three main CSS categories defined here, the Single-breed and the Two-to-Four-breed CSS groups together accounted for about 90% of the CSSs, whereas only 9.5% of the CSSs were identified in five or more breeds (**Figure 1**). These

TABLE 3 | Results from the gene enrichment analysis performed using WikiPathway analysis (WebGestalt software; Wang et al., 2013) individually for the three production-based subcategories (beef, dairy, dual-purpose) of the Single-breed core selective sweep (CSSs).

Core selective sweep (CSS) subcategory (number of genes extracted with BioMart)	Pathway name	Number of genes	Gene names	Statistics[a]
Single-breed-CSSs involving BEEF breeds (304 genes)	Interleukin-11 signaling pathway	3	*TGFB1*, *RELA*, *IKBKB*	C = 49; O = 3; E = 0.30; *R* = 9.93; rawP = 0.0035; adjP = 0.0540
	Osteopontin signaling	2	*RELA*, *IKBKB*	C = 18; O = 2; E = 0.11; R = 18.01; rawP = 0.0054; adjP = 0.0540
	RANKL-RANK signaling pathway	3	*CTSK*, *RELA*, *IKBKB*	C = 66; O = 3; E = 0.41; R = 7.37; rawP = 0.0080; adjP = 0.0540
	Corticotropin-releasing hormone	4	*TGFB1*, *FOSL1*, *RELA*, *NCOA2*	C = 123; O = 4; E = 0.76; R = 5.27; rawP = 0.0072; adjP = 0.0540
	Endochondral ossification	3	*TGFB1*, *PTCH1*, *PLAT*	C = 69; O = 3; E = 0.42; R = 7.08; rawP = 0.0089; adjP = 0.0534
	Osteoclast signaling	2	*CTSK*, *ITGB3*	C = 19; O = 2; E = 0.12; R = 17.07; rawP = 0.0060; adjP = 0.0540
	B Cell receptor signaling pathway	4	*PIP5K1A*, *GSK3A*, *RELA*, *IKBKB*	C = 114; O = 4; E = 0.70; R = 5.69; rawP = 0.0056; adjP = 0.0540
	NOD pathway	3	*RELA, IKBKB, NLRC4*	C = 39; O = 3; E = 0.24; R = 12.47; rawP = 0.0018; adjP = 0.0540
	Androgen receptor signaling pathway	4	*KAT5, RELA, KAT7, NCOA2*	C = 91; O = 4; E = 0.56; R = 7.13; rawP = 0.0025; adjP = 0.0540
	Striated muscle contraction	2	*MYL4, ACTN3*	C = 38; O = 2; E = 0.23; R = 8.53; rawP = 0.0230; adjP = 0.0690
Single-breed-CSSs involving DAIRY breeds (239 genes)	Regulation of toll-like receptor signaling pathway	6	*MAPK13, TLR4, MAPK14, PELI2, IL12B, NFKB1*	C = 154; O = 6; E = 0.69; R = 8.75; rawP = 7.00e-05;adjP = 0.0033
	Estrogen signaling pathway	3	*ESR1, MAPK14, NFKB1*	C = 30; O = 3; E = 0.13; R = 22.46; rawP = 0.0003; adjP = 0.0047
	Toll-like receptor signaling pathway	5	*MAPK13, TLR4, MAPK14, IL12B, NFKB1*	C = 116; O = 5; E = 0.52; R = 9.68; rawP = 0.0002; adjP = 0.0047
	Leptin signaling pathway	4	*ESR1, PDE3B, MAPK14, NFKB1*	C = 81; O = 4; E = 0.36; R = 11.09; rawP = 0.0005; adjP = 0.0059
	Vitamin D synthesis	2	*CYP27A1, CYP2R1*	C = 10; O = 2; E = 0.04; R = 44.92; rawP = 0.0009; adjP = 0.0070
	TFs Regulate miRNAs related to cardiac hypertrophy	2	*RASGRF1, NFKB1*	C = 10; O = 2; E = 0.04; R = 44.92; rawP = 0.0009; adjP = 0.0070
	Focal Adhesion	5	*COL6A2, SHC3, PTEN, LAMA5, LAMB3*	C = 185; O = 5; E = 0.82; R = 6.07; rawP = 0.0015; adjP = 0.0088
	Alpha 6 Beta 4 signaling pathway	3	*MAPK14, LAMA5, LAMB3*	C = 50; O = 3; E = 0.22; R = 13.48; rawP = 0.0015; adjP = 0.0088
	Complement activation, classical pathway	2	*C1QA, C1QC*	C = 17; O = 2; E = 0.08; R = 26.43; rawP = 0.0026; adjP = 0.0136
	Insulin signaling	4	*MAPK13, SHC3, PTEN, MAPK14*	C = 163; O = 4; E = 0.73; R = 5.51; rawP = 0.0062; adjP = 0.0291
Single-breed-CSSs involving DUAL-PURPOSE breeds (195 genes)	Serotonin receptor 4-6-7 and NR3C Signaling	3	*HTR4, MAP2K2, MAPK3*	C = 18; O = 3; E = 0.06; R = 46.67; rawP = 3.50e-05; adjP = 0.0014
	Calcium regulation in the cardiac cell	5	*RGS18, RGS2, RGS1, ATP2B2, PRKAR1B*	C = 151; O = 5; E = 0.54; R = 9.27; rawP = 0.0002; adjP = 0.0041
	Serotonin Receptor 2 and ELK-SRF-GATA4 signaling	2	*MAP2K2, MAPK3*	C = 16; O = 2; E = 0.06; R = 35.01; rawP = 0.0015; adjP = 0.0205
	IL-7 signaling pathway	2	*MAP2K2, MAPK3*	C = 25; O = 2; E = 0.09; R = 22.40; rawP = 0.0036; adjP = 0.0246
	IL-9 signaling pathway	2	*MAP2K2, MAPK3*	C = 25; O = 2; E = 0.09; R = 22.40; rawP = 0.0036; adjP = 0.0246

(Continued)

TABLE 3 | Continued

Core selective sweep (CSS) subcategory (number of genes extracted with BioMart)	Pathway name	Number of genes	Gene names	Statistics[1]
	Myometrial relaxation and contraction pathways	4	*RGS18, RGS2, RGS1, PRKAR1B*	C = 162; O = 4; E = 0.58; R = 6.91; rawP = 0.0028; adjP = 0.0246
	Monoamine GPCRs	2	*HTR4, HRH1*	C = 33; O = 2; E = 0.12; R = 16.97; rawP = 0.0062; adjP = 0.0261
	Serotonin HTR1 group and FOS pathway	2	*MAP2K2, MAPK3*	C = 33; O = 2; E = 0.12; R = 16.97; rawP = 0.0062; adjP = 0.0261
	Hypothetical network for drug addiction	2	*MAP2K2, MAPK3*	C = 35; O = 2; E = 0.12; R = 16.00; rawP = 0.0070; adjP = 0.0261
	EPO receptor signaling	2	*MAP2K2, MAPK3*	C = 35; O = 2; E = 0.12; R = 16.00; rawP = 0.0070; adjP = 0.0261

[a]*Wikipathway analysis Statistics: C, the number of reference genes in the category; O, the number of genes in the gene set and also in the category; E, the expected number in the category; R, ratio of enrichment; rawP, p-value from hypergeometric test; adjP, p-value adjusted by the multiple test adjustment.*

Multi-breed CSSs appear to encompass the sweeps involving the limited number of genes that have large phenotypic effects across different breeds and also, in part due to the long Multi-breed CSS intervals resulting from our CSS-labeling approach, other putatively selected genes with small effect sizes, some of which are breed-specific. Regarding the large phenotypic effects linked to the Multi-breed CSSs, many of them appear to relate to physical rather than production traits, consistent with a simpler genetic architecture (i.e., fewer genes involved in determination of the phenotype) for the former. The putatively strongly selected phenotypes include physical hallmarks that define a breed, such as coat color and patterning (*MC1R*, *KIT*) or obvious morphological traits such as lack of horns (*POLL* locus) and stature (*PLAG1*). The strong signals of selection in relation to morphological traits (e.g., body size and color-patterning traits) are consistent with the theory of the "domestication syndrome" in mammals, which suggests that selective pressure for tameness during the initial stages of domestication involved a developmental reduction in neural crest cell populations and led to multiple phenotypic changes shared by various domesticated animals species (e.g., depigmentation, floppy and reduced ears, shorter muzzles, docility, smaller brain, or cranial capacity) (Wilkins et al., 2014).

In addition to the Multi-breed CSS regions including genes that influence physical traits, there are also several genomic regions that show evidence of selection across many breeds and appear to be driven by selection on production-related genes such as *ABCG2, DGAT1, NCAPG*, and *GHR*. For the CSSs including genes with large effects, there was a correspondence between the production profiles of the breeds associated with these CSSs and the known effects of the putative target gene. It is interesting that some of these genes for which the initial major effect was related to a specific specialization (e.g., *DGAT1* for milk and *NCAPG* for growth traits), latter studies have shown that they also have effects on traits of interest in the alternative production group (e.g., *DGAT1* for beef composition and *NCAPG* for milk traits). These observations for genes with known major effects provide insights into the complexity of the relationship between genes and phenotypes; this complexity may be even more pronounced for genes of small effect.

In addition to these genes with major effect, the Multi-breed CSS intervals also included other potential selection candidates related to production (dairy and beef) traits (Supplementary Table S2 in Supplementary Material), which could represent some of the small size effect genes underlying the complex genetic architecture of quantitative traits. The functional enrichment analysis for these genomic regions suggested that genes related to the immune response and reproduction traits may also have been selection targets shared by many breeds. We also found a significant over-representation of genes related to olfactory receptors (protein coding and pseudogenes) in the Multi-breed CSSs. The abundance of these genes within selection sweep intervals, which has previously been highlighted (Bovine HapMap Consortium, 2009; Ramey et al., 2013; Qanbari et al., 2014), suggests that these behavior-related loci may have played a role in cattle domestication, whereas newly evolving functions have been suggested for these genes based on their reported duplication in the cattle genome (Elsik et al., 2009). Regarding the large number of olfactory receptor genes included in the Multi-breed CSS regions, it should be taken into account that this gene family shows one of the highest frequency of somatic mutations in their coding regions due to low expression levels, late replication time during the cell cycle and high regional non-coding mutation rate (Lawrence et al., 2013). This observation may suggest these genes as false positive results in GWAS analyses, as pointed out by Lawrence et al. (2013), and may also be relevant in interpreting results from selection signature analyses.

As mentioned above, about 90% of the CSSs defined involved a single breed (57%) or a limited number of breeds (33%, Two-to-Four CSSs). The Single-breed CSSs included an over-representation of genes related to dairy and beef production; this observation was supported by the functional enrichment analysis, which highlighted production-related pathway terms associated with these regions (**Table 3**). Hence, the Single-breed CSS

regions may include genes with small effects that influence quantitative traits of economic interest. This also suggests that similar selective pressures on different breeds, for example, for milk and meat production traits, can result in allele frequency changes in different genomic regions. This interpretation agrees with the hypothesis that many genes influence the complex traits under selection in cattle and that few of them show large phenotypic effects (Hayes et al., 2010). Alternatively, although within the same production category (dairy, beef, dual-purpose), the breeds may have been selected for subtly different production characteristics or have been subjected to differential natural (environmental) selection. In any case, each breed retains its own unique signature of its selection history. The functional enrichment analysis performed for the dual-purpose breeds, for which extremely strong selection has not been performed on either dairy or beef traits, primarily revealed genes related to reproduction traits and behavior-physiological pathways. Overall, the Single-breed CSSs pinpoint specific regions that appear to have been uniquely selected in the corresponding breeds. We propose these regions as potential markers of unique diversity and further studies focusing on the molecular basis of these selection sweeps are recommended. Furthermore, we acknowledge that a more comprehensive review also covering *Bos indicus* and African *Bos taurus* cattle would provide an enhanced overview of the impact of artificial and natural selection on the cattle genome. For example, for a selection sweep that appears to be related to short, slick hair coat (which in turn is associated with heat-stress tolerance) in tropical Senepol cattle (Flori et al., 2012), a mutation in the *PRLR* (prolactin receptor) has been identified as the putative causal mutation (Littlejohn et al., 2014). The identification of this effect, associated with a gene of major importance in lactation, provides a clear example of pleiotropy and the complex genetic architecture of physiological traits and suggests that examining selection sweeps in a broader range of cattle breeds could help to dissect the genetic architecture of traits of economic relevance.

This large-scale review of selection sweeps in European cattle reveals the historical impacts of long-term selection pressures on a species of great importance in human history. This review also presents for the first time a characterization of the selection sweeps that are breed-specific, and suggests that based on their uniqueness, these could be considered as "divergence signals," which may be important for the management and prioritization of livestock diversity.

Acknowledgments

The authors would like to acknowledge Milagros Sánchez for providing graphical assistance for the CSS representation. BGG is funded through the Spanish "Ramón y Cajal" Programme (RYC-2012-10230) from the Spanish Ministry of Economy and Competitiveness (State Secretariat for Research, Development and Innovation). The authors gratefully acknowledge the initial support for collaborative work between these two groups from the European Science Foundation through the GENOMIC-RESOURCES Exchange Grant awarded to Beatriz Gutierrez (EX/3723). This work is also included in the framework of the project AGL2012-34437 funded by the Spanish Ministry of Economy and Competitiveness (MINECO). The Roslin Institute receives core funding from the British Biotechnology and Biological Sciences Research Council (BBSRC).

Supplementary Material

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fgene.2015.00167/abstract

References

Ahn, J., Li, X., Choi, Y. M., Shin, S., Oh, S. A., Suh, Y., et al. (2014). Differential expressions of G0/G1 switch gene 2 and comparative gene identification-58 are associated with fat content in bovine muscle. *Lipids* 49, 1–14. doi: 10.1007/s11745-013-3866-3

Albrecht, E., Komolka, K., Kuzinski, J., and Maak, S. (2012). Agouti revisited: transcript quantification of the ASIP gene in bovine tissues related to protein expression and localization. *PLoS ONE* 7:e35282. doi: 10.1371/journal.pone.0035282

Allais-Bonnet, A., Grohs, C., Medugorac, I., Krebs, S., Djari, A., Graf, A., et al. (2013). Novel insights into the bovine polled phenotype and horn ontogenesis in Bovidae. *PLoS ONE*8:e63512. doi: 10.1371/journal.pone.0063512

Avilés, C., Polvillo, O., Peña, F., Juárez, M., Martínez, A. L., and Molina, A. (2013). Associations between DGAT1, FABP4, LEP, RORC, and SCD1 gene polymorphisms and fat deposition in Spanish commercial beef. *J. Anim. Sci.* 91, 4571–4577. doi: 10.2527/jas.2013-6402

Bennett, G. L., Shackelford, S. D., Wheeler, T. L., King, D. A., Casas, E., and Smith, T. P. (2013). Selection for genetic markers in beef cattle reveals complex associations of thyroglobulin and casein1-s1 with carcass and meat traits. *J. Anim. Sci.* 91, 565–571. doi: 10.2527/jas.2012-5454

Bloise, E., Cassali, G. D., Ferreira, M. C., Ciarmela, P., Petraglia, F., and Reis, F. M. (2010). Activin-related proteins in bovine mammary gland: localization and differential expression during gestational development and differentiation. *J. Dairy Sci.* 93, 4592–4601. doi: 10.3168/jds.jds.2009-2981

Blott, S., Kim, J. J., Moisio, S., Schmidt-Küntzel, A., Cornet, A., Berzi, P., et al. (2003). Molecular dissection of a quantitative trait locus: a phenylalanine-to-tyrosine substitution in the transmembrane domain of the bovine growth hormone receptor is associated with a major effect on milk yield and composition. *Genetics* 163, 253–266. Available online at: http://www.genetics.org/content/163/1/253.long

Boitard, S., and Rocha, D. (2013). Detection of signatures of selective sweeps in the Blonde d'Aquitaine cattle breed. *Anim. Genet.* 44, 579–583. doi: 10.1111/age.12042

Bongiorni, S., Mancini, G., Chillemi, G., Pariset, L., and Valentini, A. (2012). Identification of a short region on chromosome 6 affecting direct calving ease in Piedmontese cattle breed. *PLoS ONE* 7:e50137. doi: 10.1371/journal.pone.0050137

Brenneman R. A., Davis, S. K., Sanders, J. O., Burns, B. M., Wheeler, T. C., Turner, J. W., et al. (1996). The polled locus maps to BTA1 in a *Bos indicus* × *Bos taurus* cross. *J. Hered.* 87, 156–161. doi: 10.1093/oxfordjournals.jhered.a022975

Bruford, M. W., Bradley, D. G., and Luikart, G. (2003). DNA markers reveal the complexity of livestock domestication. *Nat. Rev. Genet.* 4, 900–910. doi: 10.1038/nrg1203

Caroli, A. M., Chessa, S., and Erhardt, G. J. (2009). Invited review: milk protein polymorphisms in cattle: effect on animal breeding and human nutrition. *J. Dairy Sci.* 92, 5335–5352. doi: 10.3168/jds.2009-2461

Chen, C. C., Stairs, D. B., Boxer, R. B., Belka, G. K., Horseman, N. D., Alvarez, J. V., et al. (2012). Autocrine prolactin induced by the Pten-Akt pathway is required for lactation initiation and provides a direct link between the Akt and Stat5 pathways. *Genes Dev.* 26, 2154–2168. doi: 10.1101/gad.197343.112

Bovine HapMap Consortium. (2009). Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Science* 324, 528–532. doi: 10.1126/science.1167936

de Mesquita, A. Q. E., Rezende, C. S., de Mesquita, A. J., Jardim, E. A., and Kipnis, A. P. (2012). Association of TLR4 polymorphisms with subclinical mastitis in Brazilian Holsteins. *Braz. J. Microbiol.* 43, 692–697. doi: 10.1590/S1517-83822012000200034

Drögemüller, C., Wöhlke, A., Mömke, S., and Distl, O. (2005). Fine mapping of the polled locus to a 1-Mb region on bovine chromosome 1q12. *Mamm. Genome* 16, 613–620. doi: 10.1007/s00335-005-0016-0

Druet, T., Pérez-Pardal, L., Charlier, C., and Gautier, M. (2013). Identification of large selective sweeps associated with major genes in cattle. *Anim. Genet.* 44, 758–762. doi: 10.1111/age.12073

Duan, Q., Tait, R. G. Jr., Mayes, M. S., Garrick, D. J., Liu, Q., Van Eenennaam, A. L., et al. (2012). Genetic polymorphisms in bovine transferrin receptor 2 (TFR2) and solute carrier family 40 (iron-regulated transporter), member 1 (SLC40A1) genes and their association with beef iron content. *Anim. Genet.* 43, 115–122. doi: 10.1111/j.1365-2052.2011.02224.x

Dudemaine, P. L., Thibault, C., Alain, K., and Bissonnette, N. (2014). Genetic variations in the SPP1 promoter affect gene expression and the level of osteopontin secretion into bovine milk. *Anim. Genet.* 45, 629–640. doi: 10.1111/age.12176

Eberlein, A., Takasuga, A., Setoguchi, K., Pfuhl, R., Flisikowski, K., Fries, R., et al. (2009). Dissection of genetic factors modulating fetal growth in cattle indicates a substantial role of the non-SMC condensin I complex, subunit G (NCAPG) gene. *Genetics* 183, 951–964. doi: 10.1534/genetics.109.106476

Elsik, C. G., Tellam, R. L., and Worley, K. C. (2009). The genome sequence of taurine cattle: a window to ruminant biology and evolution. *Science* 324, 522–528. doi: 10.1126/science.1169588

Emnett, R., Moeller, S., Irwin, K., Rothschild, M. F., Plastow, G., and Goodwin, R. (2001). "Association studies with leptin receptor, melanocortin-4 receptor, melanocortin-5 receptor, and peroxisome proliferator activated receptor-γ," in *Research and Reviews: Swine 2001 OARDC Special Circular,* eds M. L. Eastridge, W. L. Bacon, C. L. Knipe, D. L. Meeker, T. B. Turner, and D. L. Zartman (Columbus: The Ohio State University), 185, 57–63. Available online at: http://ohioline.osu.edu/sc185/index.html

Erbe, M., Hayes, B. J., Matukumalli, L. K., Goswami, S., Bowman, P. J., Reich, C. M., et al. (2012). Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. *J. Dairy Sci.* 95, 4114–4129. doi: 10.3168/jds.2011-5019

FAO. (2013). *In Vivo Conservation of Animal Genetic Resources.* Rome: FAO Animal Production and Health Guidelines. No. 14.

Flamenbaum, I., and Galon, N. (2010). Management of heat stress to improve fertility in dairy cows in Israel. *Reprod. Dev.* 56 Suppl. S36–S41. doi: 10.1262/jrd.1056S36

Flori, L., Fritz, S., Jaffrézic, F., Boussaha, M., Gut, I., Heath, S., et al. (2009). The genome response to artificial selection: a case study in dairy cattle. *PLoS ONE* 4:e6595. doi: 10.1371/journal.pone.0006595

Flori, L., Gonzatti, M. I., Thevenon, S., Chantal, I., Pinto, J., Berthier, D., et al. (2012). A quasi-exclusive European ancestry in the Senepol tropical cattle breed highlights the importance of the slick locus in tropical adaptation. *PLoS ONE* 7:e36133. doi: 10.1371/journal.pone.0036133

Fontanesi, L., Calò, D. G., Galimberti, G., Negrini, R., Marino, R., Nardone, A., et al. (2014). A candidate gene association study for nine economically important traits in Italian Holstein cattle. *Anim. Genet.* 45, 576–580. doi: 10.1111/age.12164

Fortes, M. R., Reverter, A., Nagaraj, S. H., Zhang, Y., Jonsson, N. N., Barris, W., et al. (2011). A single nucleotide polymorphism-derived regulatory gene network underlying puberty in 2 tropical breeds of beef cattle. *J. Anim. Sci.* 89, 1669–1683. doi: 10.2527/jas.2010-3681

Fortes, M. R. S., Kemper, K. E., Sasazaki, S., Reverter, A., Pryce, J. E., Barendse, W., et al. (2013). Evidence for pleiotropism and recent selection in the PLAG1 region in Australian beef cattle. *Anim. Genet.* 44, 636–647. doi: 10.1111/age.12075

Fu, S., Deng, Q., Yang, W., Ding, H., Wang, X., Li, P., et al. (2012). Increase of fatty acid oxidation and VLDL assembly and secretion overexpression of PTEN in cultured hepatocytes of newborn calf. *Cell. Physiol. Biochem.* 30, 1005–1013. doi: 10.1159/000341477

Gan, Q. F., Zhang, L. P., Li, J. Y., Hou, G. Y., Li, H. D., Gao, X., et al. (2008). Association analysis of thyroglobulin gene variants with carcass and meat quality traits in beef cattle. *J. Appl. Genet.* 49, 251–255. doi: 10.1007/BF03195621

Gautier, M., Flori, L., Riebler, A., Jaffrézic, F., Laloé, D., Gut, I., et al. (2009). A whole genome Bayesian scan for adaptive genetic divergence in West African cattle. *BMC Genomics* 10:550. doi: 10.1186/1471-2164-10-550

Girardot, M., Guibert, S., Laforet, M. P., Gallard, Y., Larroque, H., and Oulmouden, A. (2006). The insertion of a full-length *Bos taurus* LINE element is responsible for a transcriptional deregulation of the Normande Agouti gene. *Pigment Cell Res.* 19, 346–355. doi: 10.1111/j.1600-0749.2006.00312.x

Glick, G., Shirak, A., Uliel, S., Zeron, Y., Ezra, E., Seroussi, E., et al. (2012). Signatures of contemporary selection in the Israeli Holstein dairy cattle. *Anim. Genet.* 43 Suppl. 1, 45–55. doi: 10.1111/j.1365-2052.2012.02348.x

Graugnard, D. E., Moyes, K. M., Trevisi, E., Khan, M. J., Keisler, D., Drackley, J. K., et al. (2013). Liver lipid content and inflammometabolic indices in peripartal dairy cows are altered in response to prepartal energy intake and postpartal intramammary inflammatory challenge. *J. Dairy Sci.* 96, 918–935. doi: 10.3168/jds.2012-5676

Grisart, B., Coppieters, W., Farnir, F., Karim, L., Ford, C., Berzi, P., et al. (2002). Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the bovine *DGAT1* gene with major effect on milk yield and composition. *Genome Res.* 12, 222–231. doi: 10.1101/gr.224202

Grobet, L., Martin, L. J., Poncelet, D., Pirottin, D., Brouwers, B., Riquet, J., et al. (1997). A deletion in the bovine myostatin gene causes the double-muscled phenotype in cattle. *Nat. Genet.* 17, 71–74. doi: 10.1038/ng0997-71

Guo, J., Jorjani, H., and Carlborg, Ö. (2012). A genome-wide association study using international breeding-evaluation data identifies major loci affecting production traits and stature in the Brown Swiss cattle breed. *BMC Genet.* 13:82. doi: 10.1186/1471-2156-13-82

Guo, Y., Chen, H., Lan, X., Zhang, B., Pan, C., Zhang, L., et al. (2008). Novel SNPs of the bovine LEPR gene and their association with growth traits. *Biochem. Genet.* 46, 828–834. doi: 10.1007/s10528-008-9197-z

Gutiérrez-Gil, B., Wiener, P., and Williams, J. L. (2007). Genetic effects on coat colour in cattle: dilution of eumelanin and phaeomelanin pigments in an F2-Backcross Charolais x Holstein population. *BMC Genet.* 8:56. doi: 10.1186/1471-2156-8-56

Han, J., Luo, T., Gu, Y., Li, G., Jia, W., and Luo, M. (2009). Cathepsin K regulates adipocyte differentiation: possible involvement of type I collagen degradation. *Endocr. J.* 56, 55–63. doi: 10.1507/endocrj.K08E-143

Hanna, L. L., Sanders, J. O., Riley, D. G., Abbey, C. A., and Gill, C. A. (2014). Identification of a major locus interacting with MC1R and modifying black coat color in an F2 Nellore-Angus population. *Genet. Sel. Evol.* 46:4. doi: 10.1186/1297-9686-46-4

Harr, B., Kauer, M., and Schlotterer, C. (2002). Hitchhiking mapping: a population-based fine-mapping strategy for adaptive mutations in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. U.S.A.* 99, 12949–12954. doi: 10.1073/pnas.202336899

Hayes, B. J., Chamberlain, A. J., Maceachern, S., Savin, K., McPartlan, H., MacLeod, I., et al. (2009). A genome map of divergent artificial selection between *Bos taurus* dairy cattle and *Bos taurus* beef cattle. *Anim. Genet.* 40, 176–184. doi: 10.1111/j.1365-2052.2008.01815.x

Hayes, B. J., Lien, S., Nilsen, H., Olsen, H. G., Berg, P., Maceachern, S., et al. (2008). The origin of selection signatures on bovine chromosome 6. *Anim. Genet.* 39, 105–111. doi: 10.1111/j.1365-2052.2007.01683.x

Hayes, B. J., Pryce, J., Chamberlain, A. J., Bowman, P. J., and Goddard, M. E. (2010). Genetic architecture of complex traits and accuracy of genomic prediction: coat colour, milk-fat percentage, and type in Holstein cattle as contrasting model traits. *PLoS Genet.* 6:e1001139. doi: 10.1371/journal.pgen.1001139

Hohenlohe, P. A., Phillips, P. C., and Cresko, W. A. (2010). Using population genomics to detect selection in natural populations: key concepts and methodological considerations. *Int. J. Plant Sci.* 171, 1059–1071. doi: 10.1086/656306

Horseman, N. D., and Collier, R. J. (2014). Serotonin: a local regulator in the mammary gland epithelium. *Annu. Rev. Anim. Biosci.* 2, 353–374. doi: 10.1146/annurev-animal-022513-114227

Hosokawa, D., Ishii, A., Yamaji, K., Sasazaki, S., Oyama, K., and Mannen, H. (2012). Identification of divergently selected regions between Japanese Black and Holstein cattle using bovine 50k SNP array. *Anim. Sci. J.* 83, 7–13. doi: 10.1111/j.1740-0929.2011.00927.x

Houseknecht, K. L., Baile, C. Z., Matteri, R. L., and Spurlock, M. C. (1998). The biology of leptin: a review. *J. Anim. Sci.* 76, 1405–1420.

Jacobs, K., Van Poucke, M., Mattheeuws, M., Chardon, P., Yerle, M., Peelman, L. J. et al. (2002). Characterization of the porcine melanocortin 2 receptor gene (MC2R). *Anim. Genet.* 33, 415–421. doi: 10.1046/j.1365-2052.2002.00899.x

Kaplan, N. L., Hudson, R. R., and Langley, C. H. (1989). The hitchhiking effect revisited. *Genetics* 123, 887–899.

Karim, L., Takeda, H., Lin, L., Druet, T., Arias, J. A. C., Baurain, D., et al. (2011). Variants modulating the expression of a chromosome domain encompassing PLAG1 influence bovine stature. *Nat. Genet.* 43, 405–413. doi: 10.1038/ng.814

Kaupe, B., Brandt, H., Prinzenberg, E. M., and Erhardt, G. (2007). Joint analysis of the influence of CYP11B1 and DGAT1 genetic variation on milk production, somatic cell score, conformation, reproduction, and productive lifespan in German Holstein cattle. *J. Anim. Sci.* 85, 11–21. doi: 10.2527/jas.2005-753

Kemper, K. E., Saxton, S. J., Bolormaa, S., Hayes, B. J., and Goddard, M. E. (2014). Selection for complex traits leaves little or no classic signatures of selection. *BMC Genomics* 15:246. doi: 10.1186/1471-2164-15-246

Kemper, K. E., Visscher, P. M., and Goddard, M. E. (2012). Genetic architecture of body size in mammals. *Genome Biol.* 13:244. doi: 10.1186/gb-2012-13-4-244

Khatib, H., Zaitoun, I., Wiebelhaus-Finger, J., Chang, Y. M., and Rosa, G. J. (2007). The association of bovine PPARGC1A and OPN genes with milk composition in two independent Holstein cattle populations. *J. Dairy Sci.* 90, 2966–2970. doi: 10.3168/jds.2006-812

Kohn, M. H., Pelz, H. J., and Wayne, R. K. (2000). Natural selection mapping of the warfarin-resistance gene. *Proc. Natl. Acad. Sci. U.S.A.* 97, 7911–7915. doi: 10.1073/pnas.97.14.7911

Komisarek, J., and Walendowska, A. (2012). Analysis of the PPARGC1A gene as a potential marker for productive and reproductive traits in cattle. *Folia Biol.* 60, 171–174. doi: 10.3409/fb60_3-4.171-174

Kováčik, A., Bulla, J., Trakovická, A., Žitný, J., and Rafayová, A. (2012). The effect of the porcine melanocortin-5 receptor (MC5R) gene associated with feed intake, carcass and physic chemical characteristics. *J. Microbiol. Biotechnol. Food Sci.* 1, 498–506. Available online at: http://www.jmbfs.org/wp-content/uploads/2012/01/Kovacik.pdf

Kuehn, C., and Weikard, R. (2007a). An investigation into the genetic background of coat colour dilution in a Charolais x German Holstein F2 resource population. *Anim. Genet.* 38, 109–113. doi: 10.1111/j.1365-2052.2007.01569.x

Kuehn, C., and Weikard, R. (2007b). Multiple splice variants within the bovine silver homologue (SILV) gene affecting coat color in cattle indicate a function additional to fibril formation in melanophores. *BMC Genomics* 8:335. doi: 10.1186/1471-2164-8-335

Lagonigro, R., Wiener, P., Pilla, F., Woolliams, J. A., and Williams, J. L. (2003). A new mutation in the coding region of the bovine leptin gene associated with feed intake. *Anim. Genet.* 34, 371–374. doi: 10.1046/j.1365-2052.2003.01028.x

Laloe, D., Moazami-Goudarzi, K., Lenstra, J. A., Ajmone Marsan, P., Azor, P., Baumung, R., et al. (2010). Spatial trends of genetic variation of domestic ruminants in Europe. *Diversity* 2, 932–945. doi: 10.3390/d2060932

Lango-Allen, H., Estrada, K., Lettre G., Berndt, S. I., Weedon, M. N., Rivadeneira, F., et al. (2010). Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* 467, 832–838.

Larson, J., Swanson, J., Berry, D'. A., and Smith, C. (1992). *Selected Readings on the History and Use of Old Livestock Breeds*. Beltsville, MD: American Minor Breeds Conservancy.

Lawrence, M. S., Stojanov, P., Polak, P., Kryukov, G. V., Cibulskis, K., Sivachenko, A., et al. (2013). Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* 499, 214–218. doi: 10.1038/nature12213

Lawson, M. A. (2004). The role of integrin degradation in post-mortem drip loss in pork. *Meat Sci.* 68, 559–566. doi: 10.1016/j.meatsci.2004.05.019

Lee, H. J., Jang, M., Kim, H., Kwak, W., Park, W., Hwang, J. Y., et al. (2013a). Comparative transcriptome analysis of adipose tissues reveals that ecm-receptor interaction is involved in the depot-specific adipogenesis in cattle. *PLoS ONE* 8:e66267. doi: 10.1371/journal.pone.0066267

Lee, J. W., Bannerman, D. D., Paape, M. J., Huang, M. K., and Zhao, X. (2006). Characterization of cytokine expression in milk somatic cells during intramammary infections with *Escherichia coli* or *Staphylococcus aureus* by real-time PCR. *Vet. Res.* 37, 219–229. doi: 10.1051/vetres:2005051

Lee, T., Cho, S., Seo, K. S., Chang, J., Kim, H., and Yoon, D. (2013b). Genetic variants and signatures of selective sweep of Hanwoo population (Korean native cattle). *BMB Rep.* 46, 346–351. doi: 10.5483/BMBRep.2013.46.7.211

Leonard, S., Khatib, H., Schutzkus, V., Chang, Y. M., and Maltecca, C. (2005). Effects of the osteopontin gene variants on milk production traits in dairy cattle. *J. Dairy Sci.* 88, 4083–4086. doi: 10.3168/jds.S0022-0302(05)73092-7

Li, Q., Wang, Z., Zhang, B., Lu, Y., Yang, Y., Ban, D., et al. (2014a). Single nucleotide polymorphism scanning and expression of the pig PPARGC1A gene in different breeds. *Lipids* 49, 1047–1055. doi: 10.1007/s11745-014-3928-1

Li, Z., Zhang, H., Wang, H., Chen, L., Wang, L., Liu, X., et al. (2014b). Polymorphism in the promoter of TLR4 gene by PCR-RFLP and its association with somatic cell score in Chinese Holstein. *Archiv. Tierzucht* 57, 1–6. doi: 10.7482/0003-9438-57-006

Liefers, S. C., te Pas, M. F., Veerkamp, R. F., and van der Lende, T. (2002). Associations between leptin gene polymorphisms and production, live weight, energy balance, feed intake, and fertility in Holstein heifers. *J. Dairy Sci.* 85, 1633–1638. doi: 10.3168/jds.S0022-0302(02)74235-5

Lim, D., Gondro, C., Park, H. S., Cho, Y. M., Chai, H. H., Seong, H. H., et al. (2013). Identification of recently selected mutations driven by artificial selection in Hanwoo (Korean cattle). *Asian-Australas J. Anim. Sci.* 26, 603–608. doi: 10.5713/ajas.2012.12456

Lindholm-Perry, A. K., Sexten, A. K., Kuehn, L. A., Smith, T. P., King, D. A., Shackelford, S. D., et al. (2011). Association, effects and validation of polymorphisms within the NCAPG - LCORL locus located on BTA6 with feed intake, gain, meat and carcass traits in beef cattle. *BMC Genet.* 12:103. doi: 10.1186/1471-2156-12-103

Littlejohn, M. D., Henty, K. M., Tiplady, K., Johnson, T., Harland, C., Lopdell, T., et al. (2014). Functionally reciprocal mutations of the prolactin signalling pathway define hairy and slick cattle. *Nat. Commun.* 5:5861. doi: 10.1038/ncomms6861

Mancini, G., Gargani, M., Chillemi, G., Nicolazzi, E. L., Marsan, P. A., Valentini, A., et al. (2014). Signatures of selection in five Italian cattle breeds detected by a 54K SNP panel. *Mol. Biol. Rep.* 41, 957–965. doi: 10.1007/s11033-013-2940-5

Matukumalli, L. K., Lawley, C. T., Schnabel, R. D., Taylor, J. F., Allan, M. F., Van Tassell, C. P. et al. (2009). Development and characterization of a high density SNP genotyping assay for cattle. *PLoS ONE* 4:e5350. doi: 10.1371/journal.pone.0005350

Maynard Smith, J., and Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genet. Res.* 23: 23–35. doi: 10.1017/S0016672300014634

McLean, K. L., and Schmutz, S. M. (2009). Associations of melanocortin 1 receptor genotype with growth and carcass traits in beef cattle. *Can. J. Anim. Sci.* 89, 295–300. doi: 10.4141/CJAS08094

McPherron, A. C., and Lee, S. J. (1997). Double muscling in cattle due to mutations in the myostatin gene. *Proc. Natl. Acad. Sci. U.S.A.* 94, 12457–12461. doi: 10.1073/pnas.94.23.12457

Medugorac, I., Seichter, D., Graf, A., Russ, I., Blum, H., Göpel, K. H., et al. (2012). Bovine polledness–an autosomal dominant trait with allelic heterogeneity. *PLoS ONE* 7:e39477. doi: 10.1371/journal.pone.0039477

Montoliu, L., Oetting, W. S., and Bennett, D. C. (2014). *Color Genes. European Society for Pigment Cell Research. World Wide Web.* Available online at: http://www.espcr.org/micemut.

Morris, C. A., Cullen, N. G., Glass, B. C., Hyndman, D. L., Manley, T. R., Hickey, S. M., et al. (2007). Fatty acid synthase effects on bovine adipose fat and milk fat. *Mamm. Genome* 18, 64–74. doi: 10.1007/s00335-006-0102-y

Nielsen, R., Williamson, S., Kim, Y., Hubisz, M. J., Clark, A. G., and Bustamante, C. (2005). Genomic scans for selective sweeps using SNP data. *Genome Res.* 15, 1566–1575. doi: 10.1101/gr.4252305

Ogorevc, J., Kunej, T., Razpet, A., and Dovc, P. (2009). Database of cattle candidate genes and genetic markers for milk production and mastitis. *Anim. Genet.* 40, 832–851. doi: 10.1111/j.1365-2052.2009.01921.x

Olsen, H. G., Meuwissen, T. H., Nilsen, H., Svendsen, M., and Lien, S. (2008). Fine mapping of quantitative trait Loci on bovine chromosome 6 affecting calving difficulty. *J. Dairy Sci.* 91, 4312–4322. doi: 10.3168/jds.2008-1000

Olsen, H. G., Nilsen, H., Hayes, B., Berg, P. R., Svendsen, M., Lien, S., et al. (2007). Genetic support for a quantitative trait nucleotide in the ABCG2 gene affecting milk composition of dairy cattle. *BMC Genet.* 8:32. doi: 10.1186/1471-2156-8-32

Olson, T. A. (1999). "Genetics of color variation," in *The Genetics of Cattle,* eds R. Fries and A. Ruvinsky (Wallingford: CABI), 33–53.

Ouali, A., Herrera-Mendez, C. H., Coulis, G., Becila, S., Boudjellal, A., Aubry, L., et al. (2006). Revisiting the conversion of muscle into meat and the underlying mechanisms. *Meat Sci.* 74, 44–58. doi: 10.1016/j.meatsci.2006.05.010

Pan, D., Zhang, S., Jiang, J., Jiang, L., Zhang, Q., and Liu, J. (2013). Genome-wide detection of selective signature in Chinese Holstein. *PLoS ONE* 8:e60440. doi: 10.1371/journal.pone.0060440

Pausch, H., Flisikowski, K., Jung, S., Emmerling, R., Edel, C., Götz, K. U., et al. (2011). Genome-wide association study identifies two major loci affecting calving ease and growth-related traits in cattle. *Genetics* 187, 289–297. doi: 10.1534/genetics.110.124057

Pintus, E., Sorbolini, S., Albera, A., Gaspa, G., Dimauro, C., Steri, R., et al. (2014). Use of locally weighted scatterplot smoothing (LOWESS) regression to study selection signatures in Piedmontese and Italian Brown cattle breeds. *Anim. Genet.* 45, 1–11. doi: 10.1111/age.12076

Pollinger, J. P., Bustamante, C. D., Fledel-Alon, A., Schmutz, S., Gray, M. M., and Wayne, R. K. (2005). Selective sweep mapping of genes with large phenotypic effects. *Genome Res.* 15, 1809–1819. doi: 10.1101/gr.4374505

Porto-Neto, L. R., Lee, S. H., Sonstegard, T. S., Van Tassell, C. P., Lee, H. K., Gondro, C. et al. (2014). Genome-wide detection of signatures of selection in Korean Hanwoo cattle. *Anim. Genet.* 45, 180–190. doi: 10.1111/age.12119

Porto-Neto, L. R., Sonstegard, T. S., Liu, G. E., Bickhart, D. M., Da Silva, M. V. B., Machado, M., et al. (2013). Genomic divergence of zebu and taurine cattle identified through high-density SNP genotyping. *BMC Genomics* 14:876. doi: 10.1186/1471-2164-14-876

Prasad, A., Schnabel, R. D., McKay, S. D., Murdoch, B., Stothard, P., Kolbehdari, D., et al. (2008). Linkage disequilibrium and signatures of selection on chromosomes 19 and 29 in beef and dairy cattle. *Anim. Genet.* 39, 597–605. doi: 10.1111/j.1365-2052.2008.01772.x

Pryce, J. E., Hayes, B. J., Bolormaa, S., and Goddard, M. E. (2011). Polymorphic regions affecting human height also control stature in cattle. *Genetics* 187, 981–984. doi: 10.1534/genetics.110.123943

Puglisi, R., Cambuli, C., Capoferri, R., Giannino, L., Lukaj, A., Duchi, R., et al. (2013). Differential gene expression in cumulus oocyte complexes collected by ovum pick up from repeat breeder and normally fertile Holstein Friesian heifers. *Anim. Reprod. Sci.* 141, 26–33. doi: 10.1016/j.anireprosci.2013.07.003

Puigserver, P., and Spiegelman, B. M. (2003). Peroxisome proliferator-activated receptor-gamma coactivator 1 alpha (PGC-1 alpha): transcriptional coactivator and metabolic regulator. *Endocr. Rev.* 24, 78–90. doi: 10.1210/er.2002-0012

Qanbari, S., Gianola, D., Hayes, B., Schenkel, F., Miller, S., Moore, S., et al. (2011). Application of site and haplotype-frequency based approaches for detecting selection signatures in cattle. *BMC Genomics* 12:318. doi: 10.1186/1471-2164-12-318

Qanbari, S., Pausch, H., Jansen, S., Somel, M., Strom, T. M., Fries, R., et al. (2014). Classic selective sweeps revealed by massive sequencing in cattle. *PLoS Genet.* 10:e1004148. doi: 10.1371/journal.pgen.1004148

Qanbari, S., Pimentel, E. C., Tetens, J., Thaller, G., Lichtner, P., Sharifi, A. R., et al. (2010). A genome-wide scan for signatures of recent selection in Holstein cattle. *Anim. Genet.* 41, 377–389. doi: 10.1111/j.1365-2052.2009.02016.x

Ramayo-Caldas, Y., Fortes, M. R., Hudson, N. J., Porto-Neto, L. R., Bolormaa, S., Barendse, W., et al. (2014). A marker-derived gene network reveals the regulatory role of PPARGC1A, HNF4G, and FOXP3 in intramuscular fat deposition of beef cattle. *J. Anim. Sci.* 92, 2832–2845. doi: 10.2527/jas.2013-7484

Ramey, H. R., Decker, J. E., McKay, S. D., Rolf, M. M., Schnabel, R. D., and Taylor, J. F. (2013). Detection of selective sweeps in cattle using genome-wide SNP data. *BMC Genomics* 14:382. doi: 10.1186/1471-2164-14-382

Reinhardt, T. A., and Lippolis, J. D. (2009). Mammary gland involution is associated with rapid down regulation of major mammary Ca2+-ATPases. *Biochem. Biophys. Res. Commun.* 378, 99–102. doi: 10.1016/j.bbrc.2008.11.004

Robbins, L. S., Nadeau, J. H., Johnson, K. R., Kelly, M. A., Roselli-rehfuss, L., Baack, E., et al. (1993). Pigmentation phenotypes of variant extension locus alleles result from point mutations that alter MSH receptor function. *Cell* 72, 827–834. doi: 10.1016/0092-8674(93)90572-8

Ron, M., Israeli, G., Seroussi, E., Weller, J. I., Gregg, J. P., Shani, M., et al. (2007). Combining mouse mammary gland gene expression and comparative mapping for the identification of candidate genes for QTL of milk production traits in cattle. *BMC Genomics* 8:183. doi: 10.1186/1471-2164-8-183

Rothammer, S., Seichter, D., Förster, M., and Medugorac, I. (2013). A genome-wide scan for signatures of differential artificial selection in ten cattle breeds. *BMC Genomics* 14:908. doi: 10.1186/1471-2164-14-908

Roy, R., Ordovas, L., Zaragoza, P., Romero, A., Moreno, C., Altarriba, J., et al. (2006). Association of polymorphisms in the bovine FASN gene with milk-fat content. *Anim. Genet.* 37, 215–218. doi: 10.1111/j.1365-2052.2006.01434.x

Royo, L. J., Alvarez, I., Fernández, I., Arranz, J. J., Gómez, E., and Goyache, F. (2005). The coding sequence of the ASIP gene is identical in nine wild-type coloured cattle breeds. *J. Anim. Breed. Genet.* 122, 357–360. doi: 10.1111/j.1439-0388.2005.00541.x

Schennink, A., Bovenhuis, H., Léon-Kloosterziel, K. M., van, Arendonk, J. A., and Visker, M. H. (2009). Effect of polymorphisms in the FASN, OLR1, PPARGC1A, PRL and STAT5A genes on bovine milk-fat composition. *Anim. Genet.* 40, 909–916. doi: 10.1111/j.1365-2052.2009.01940.x

Schmutz, S. M., and Dreger, D. L. (2013). Interaction of MC1R and PMEL alleles on solid coat colors in Highland cattle. *Anim. Genet.* 44, 9–13. doi: 10.1111/j.1365-2052.2012.02361.x

Schneider, E. H., Neumann, D., and Seifert, R. (2014). Modulation of behaviour by the histaminergic system: lessons from HDC-, H3R- and H4R-deficient mice. *Neurosci. Biobehav. Rev.* 47C, 101–121. doi: 10.1016/j.neubiorev.2014.07.020

Seitz, J. J., S., Schmutz, M., Thue, T. D., and Buchanan, F.C. (1999). A missense mutation in the bovine MGF gene is associated with the roan phenotype in Belgian blue and Shorthorn cattle. *Mamm. Genome* 10, 710–712. doi: 10.1007/s003359901076

Setoguchi, K., Furuta, M., Hirano, T., Nagao, T., Watanabe, T., Sugimoto, Y., et al. (2009). Cross-breed comparisons identified a critical 591-kb region for bovine carcass weight QTL (CW-2) on chromosome 6 and the Ile-442-Met substitution in NCAPG as a positional candidate. *BMC Genet.* 10:43. doi: 10.1186/1471-2156-10-43

Setoguchi, K., Watanabe, T., Weikard, R., Albrecht, E., Kühn, C., Kinoshita, A., et al. (2011). The SNP c.1326T>G in the non-SMC condensin I complex, subunit G (NCAPG) gene encoding a p.Ile442Met variant is associated with an increase in body frame size at puberty in cattle. *Anim. Genet.* 42, 650–655. doi: 10.1111/j.1365-2052.2011.02196.x

Sevane, N., Armstrong, E., Cortés, O., Wiener, P., Wong, R. P., Dunner, S., et al. (2013). Association of bovine meat quality traits with genes included in the PPARG and PPARGC1A networks. *Meat Sci.* 94, 328–335. doi: 10.1016/j.meatsci.2013.02.014

Sevane, N., Armstrong, E., Wiener, P., Pong, W. R., Dunner, S., and GemQual Consortium (2014). Polymorphisms in twelve candidate genes are associated with growth, muscle lipid profile and meat quality traits in eleven European cattle breeds. *Mol. Biol. Rep.* 41, 4721–4731. doi: 10.1007/s11033-014-3343-y

Sheehy, P. A., Riley, L. G., Raadsma, H. W., Williamson, P., and Wynn, P. C. (2009). A functional genomics approach to evaluate candidate genes located in a QTL interval for milk production traits on BTA6. *Anim. Genet.* 40, 492–498. doi: 10.1111/j.1365-2052.2009.01862.x

Shin, S., and Chung, E. (2013). Novel SNPs in the bovine ADIPOQ and PPARGC1A genes are associated with carcass traits in Hanwoo (Korean cattle). *Mol. Biol. Rep.* 40, 4651–4660. doi: 10.1007/s11033-013-2560-0

Smith, J. A., Lewis, A. M., Wiener, P., and Williams, J. L. (2000). Genetic variation in the bovine myostatin gene in UK beef cattle: allele frequencies and haplotype analysis in the South Devon. *Anim. Genet.* 31, 306–309. doi: 10.1046/j.1365-2052.2000.00521.x

Somavilla, A. L., Sonstegard, T. S., Higa, R. H., Rosa, A. N., Siqueira, F., Silva, L. O., et al. (2014). A genome-wide scan for selection signatures in Nellore cattle. *Anim. Genet.* 45, 771–781. doi: 10.1111/age.12210

Soranzo, N., Rivadeneira, F., Chinappen-Horsley, U., Malkina, I., Richards, J. B., Hammond, N., et al. (2009). Meta-analysis of genome-wide scans for human adult stature identifies novel loci and associations with measures of skeletal frame size. *PLoS Genet.* 5:e1000445. doi: 10.1371/journal.pgen.1000445

Sponenberg, D. P., and Bixby, D. E. (2007). *Managing Breeds for a Secure Future: Strategies for Breeders and Breed Associations*. Pittsboro, NC: American Livestock Breeds Conservancy.

Stella, A., Ajmone-Marsan, P., Lazzari, B., and Boettcher, P. (2010). Identification of selection signatures in cattle breeds selected for dairy production. *Genetics* 185, 1451–1461. doi: 10.1534/genetics.110.116111

Storz, J. F., Payseur, B. A., and Nachman, M. W. (2004). Genome scans of DNA variability in humans reveal evidence for selective sweeps outside of Africa. *Mol. Biol. Evol.* 21, 1800–1811. doi: 10.1093/molbev/msh192

Suchocki, T., Komisarek, J., and Szyda, J. (2010). Testing candidate gene effects on milk production traits in dairy cattle under various parameterizations and modes of inheritance. *J. Dairy Sci.* 93, 2703–2717. doi: 10.3168/jds.2009-2550

Switonski, M., Mankowska, M., and Salamon, S. (2013). Family of melanocortin receptor (MCR) genes in mammals-mutations, polymorphisms and phenotypic effects. *J. Appl. Genet.* 54, 461–472. doi: 10.1007/s13353-013-0163-z

Tetens, J., Widmann, P., Kühn, C., and Thaller, G. (2013). A genome-wide association study indicates LCORL/NCAPG as a candidate locus for withers height in German Warmblood horses. *Anim. Genet.* 44, 467–471. doi: 10.1111/age.12031

Thaller, G., Kühn, C., Winter, A., Ewald, G., Bellmann, O., Wegner, J., et al. (2003). DGAT1, a new positional and functional candidate gene for intramuscular fat deposition in cattle. *Anim. Genet.* 34, 354–357. doi: 10.1046/j.1365-2052.2003.01011.x

Trakovická, A., Moravčíková, N., and Kasarda, R. (2013). Genetic polymorphisms of leptin and leptin receptor genes in relation with production and reproduction traits in cattle. *Acta Biochim. Pol.* 60, 783–787. http://www.actabp.pl/pdf/4_2013/783.pdf

Utsunomiya, Y. T., do Carmo, A. S., Carvalheiro, R., Neves, H. H., Matos, M. C., Zavarez, L. B., et al. (2013b). Genome-wide association study for birth weight in Nellore cattle points to previously described orthologous genes affecting human and bovine height. *BMC Genet.* 14:52. doi: 10.1186/1471-2156-14-52

Utsunomiya, Y. T., Pérez O'Brien, A. M., Sonstegard, T. S., Van Tassell, C. P., do Carmo, A. S., Mészáros, G., et al. (2013a). Detecting loci under recent positive selection in dairy and beef cattle by combining different genome-wide scan methods. *PLoS ONE* 8:e64280. doi: 10.1371/journal.pone.0064280

Vankan, D. M., Waine, D. R., and Fortes, M. R. (2010). Real-time PCR genotyping and frequency of the myostatin F94L mutation in beef cattle breeds. *Animal* 4, 530–534. doi: 10.1017/S175173110999139X

Voight, B. F., Kudaravalli, S., Wen, X., and Pritchard, J. K. (2006). A map of recent positive selection in the human genome. *PLoS Biol.* 4: 446–458. doi: 10.1371/journal.pbio.0040154

Wang, J., Duncan, D., Shi, Z., and Zhang, B. (2013). WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013. *Nucleic Acids Res.* 41, W77–W83. doi: 10.1093/nar/gkt439

Wang, X. P., Luoreng, Z. M., Gao, S. X., Guo, D. S., Li, J. Y., Gao, X., et al. (2014a). Haplotype analysis of TLR4 gene and its effects on milk somatic cell score in Chinese commercial cattle. *Mol. Biol. Rep.* 41, 2345–2351. doi: 10.1007/s11033-014-3088-7

Wang, X., Xu, S., Gao, X., Ren, H., and Chen, J. B. (2007). Genetic polymorphism of TLR4 gene and correlation with mastitis in cattle. *J. Genet. Genomics* 34, 406–412. doi: 10.1016/S1673-8527(07)60044-7

Wang, Z., Hou, X., Qu, B., Wang, J., Gao, X., and Li, Q. (2014b). *Pten* regulates development and lactation in the mammary glands of dairy cows. *PLoS ONE* 9:e102118. doi: 10.1371/journal.pone.0102118

Wedholm, A., Larsen, L. B., Lindmark-Månsson, H., Karlsson, A. H., and Andrén, A. (2006). Effect of protein composition on the cheese-making properties of milk from individual dairy cows. *J. Dairy Sci.* 89, 3296–3305. doi: 10.3168/jds.S0022-0302(06)72366-9

Wei, J., Geale, P. F., Sheehy, P. A., and Williamson, P. (2012). The impact of ABCG2 on bovine mammary epithelial cell proliferation. *Anim. Biotechnol.* 23, 221–224. doi: 10.1080/10495398.2012.696567

Weikard, R., Kuehn, C., Goldammer, T., Freyer, G., and Schwerin, M. (2005). The bovine PPARGC1A gene: molecular characterization and association of an SNP with variation of milk fat synthesis. *Physiol. Genomics* 21, 1–13. doi: 10.1152/physiolgenomics.00103.2004

Weikard, R., Widmann, P., Buitkamp, J., Emmerling, R., and Kuehn, C. (2012). Revisiting the quantitative trait loci for milk production traits on BTA6. *Anim. Genet.* 43, 318–323. doi: 10.1111/j.1365-2052.2011.02258.x

Wiener, P., Edriss, M. A., Williams, J. L., Waddington, D., Law, A., Woolliams, J. A., et al. (2011). Information content in genome-wide scans: concordance between patterns of genetic differentiation and linkage mapping associations. *BMC Genomics* 12:65. doi: 10.1186/1471-2164-12-65

Wilkins, A. S., Wrangham, R. W., and Fitch, W. T. (2014). The "domestication syndrome" in mammals: a unified explanation based on neural crest cell behaviour and genetics. *Genetics* 197, 795–808. doi: 10.1534/genetics.114.165423

Williams, J. L., Dunner, S., Valentini, A., Mazza, R., Amarger, V., Checa, M. L., et al. (2009). Discovery, characterization and validation of single nucleotide polymorphisms within 206 bovine genes that may be considered as candidate genes for beef production and quality. *Anim. Genet.* 40, 486–491. doi: 10.1111/j.1365-2052.2009.01874.x

Woronuk, G. N., Marquess, F. L., James, S. T., Palmer, J., Berryere, T., Deobald, H., et al. (2012). Association of leptin genotypes with beef cattle characteristics. *Anim. Genet.* 43, 608–610. doi: 10.1111/j.1365-2052.2012.02320.x

Wu, X. X., Yang, Z. P., Shi, X. K., Li, J. Y., Ji, D. J., Mao, Y. J., et al. (2012). Association of SCD1 and DGAT1 SNPs with the intramuscular fat traits in Chinese Simmental cattle and their distribution in eight Chinese cattle breeds. *Mol. Biol. Rep.* 39, 1065–1071. doi: 10.1007/s11033-011-0832-0

Xiao, Y., Junfeng, H., Tianhong, L., Lu, W., Shulin, C., Yu, Z., et al. (2006). Cathepsin K in adipocyte differentiation and its potential role in the pathogenesis of obesity. *J. Clin. Endocrinol. Metab.* 91, 4520–4527. doi: 10.1210/jc.2005-2486

Zhang, S., Knight, T. J., Reecy, J. M., and Beitz, D. C. (2008). DNA polymorphisms in bovine fatty acid synthase are associated with beef fatty acid composition. *Anim. Genet.* 39, 62–70. doi: 10.1111/j.1365-2052.2007.01681.x

Zheng, X., Ju, Z., Wang, J., Li, Q., Huang, J., Zhang, A., et al. (2011). Single nucleotide polymorphisms, haplotypes and combined genotypes of LAP3 gene in bovine and their association with milk production traits. *Mol. Biol. Rep.* 38, 4053–4061. doi: 10.1007/s11033-010-0524-1

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Gutiérrez-Gil, Arranz and Wiener. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 07 April 2015
doi: 10.3389/fgene.2015.00127

Artificial selection with traditional or genomic relationships: consequences in coancestry and genetic diversity

Silvia Teresa Rodríguez-Ramilo[1], *Luis Alberto García-Cortés*[1] *and María Ángeles Rodríguez de Cara*[2]*

[1] Departamento de Mejora Genetica Animal, Instituto Nacional de Investigacion y Tecnologia Agraria y Alimentaria, Madrid, Spain, [2] Laboratoire d'Eco-anthropologie et Ethnobiologie, Museum National d'Histoire Naturelle, Paris, France

OPEN ACCESS

Edited by:
Michael William Bruford, Cardiff University, UK

Reviewed by:
Mario Calus, Wageningen UR Livestock Research, Netherlands
Hendrik-Jan Megens, Wageningen University, Netherlands

***Correspondence:**
Angeles de Cara, Laboratoire d'Eco-anthropologie et Ethnobiologie, UMR 7206 Centre National de la Recherche Scientifique/Museum National d'Histoire Naturelle/Universite Paris 17 place du Trocadéro, F-75116 Paris, France
adecara@mnhn.fr

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: *31 October 2014*
Accepted: *17 March 2015*
Published: *07 April 2015*

Citation:
Rodríguez-Ramilo ST, García-Cortés LA and de Cara MÁR (2015) Artificial selection with traditional or genomic relationships: consequences in coancestry and genetic diversity. Front. Genet. 6:127. doi: 10.3389/fgene.2015.00127

Estimated breeding values (EBVs) are traditionally obtained from pedigree information. However, EBVs from high-density genotypes can have higher accuracy than EBVs from pedigree information. At the same time, it has been shown that EBVs from genomic data lead to lower increases in inbreeding compared with traditional selection based on genealogies. Here we evaluate the performance with BLUP selection based on genealogical coancestry with three different genome-based coancestry estimates: (1) an estimate based on shared segments of homozygosity, (2) an approach based on SNP-by-SNP count corrected by allelic frequencies, and (3) the identity by state methodology. We evaluate the effect of different population sizes, different number of genomic markers, and several heritability values for a quantitative trait. The performance of the different measures of coancestry in BLUP is evaluated in the true breeding values after truncation selection and also in terms of coancestry and diversity maintained. Accordingly, cross-performances were also carried out, that is, how prediction based on genealogical records impacts the three other measures of coancestry and inbreeding, and viceversa. Our results show that the genetic gains are very similar for all four coancestries, but the genomic-based methods are superior to using genealogical coancestries in terms of maintaining diversity measured as observed heterozygosity. Furthermore, the measure of coancestry based on shared segments of the genome seems to provide slightly better results on some scenarios, and the increase in inbreeding and loss in diversity is only slightly larger than the other genomic selection methods in those scenarios. Our results shed light on genomic selection vs. traditional genealogical-based BLUP and make the case to manage the population variability using genomic information to preserve the future success of selection programmes.

Keywords: genomic selection, coancestry, inbreeding, breeding value, genetic diversity

1. Introduction

Best linear unbiased prediction (BLUP) is possibly the most common selection method in animal and plant breeding, where it is used to calculate estimated breeding values (EBVs). BLUP

evaluations maximize the genetic gain given the data by increasing the accuracy of the predictions (Henderson, 1984). This method relies on both the additive relationship matrix between the individuals in the population, which are traditionally obtained from pedigree records, and on phenotypic records of the candidates to selection. Such is the power of BLUP that it is actually not only used in breeding programmes, but also in evolutionary ecology to estimate the strength of selection and evolutionary change (see Hadfield et al., 2010 for a review) and more recently in human genetics for the prediction of complex traits (Makowsky et al., 2011).

With the advent of high-throughput genotyping techniques and the development of chips containing thousands of single nucleotide polymorphisms (SNPs) at a reasonable cost, the implementation of genome-wide evaluations (Meuwissen et al., 2001; Goddard and Hayes, 2007) is routinely used in many breeding programs, and conventional BLUP selection based on pedigrees is now migrating to genomic selection.

Genome-based EBV (estimated breeding values based on high-density marker data across the genome) have generally yielded a higher accuracy than pedigree-based EBV (Meuwissen et al., 2001; Goddard, 2009; Hayes et al., 2009; Sonesson et al., 2012; Rodriguez-Ramilo et al., 2014). This is because genetic markers provide a more accurate relationship matrices than pedigree data (Goddard, 2009), which accounts for the expected genetic relationships. For example, while the genealogical relationship between two full-sibs is 0.5, using molecular markers like high-density SNP chips, a more accurate value can be obtained, thus showing that the true relationship deviates from 0.5 (Visscher et al., 2006) and varies among pairs of sibs, depending on the segregation of the parental chromosomes (Garcia-Cortes et al., 2013).

Genomic selection can therefore lead to high levels of accuracy at an early age and generation intervals can be shortened leading to faster genetic gains within a specific breeding program. Furthermore, genomic selection not only has increased the accuracy in the breeding values, but also the increase in inbreeding per generation is lower than that obtained with conventional pedigree-based BLUP selection (Daetwyler et al., 2007; Sonesson et al., 2012). However, both traditional and genomic selection increase the levels of both inbreeding and coancestry, thus decreasing the pool of genetic diversity. This has wide-ranging consequences, as it is clear that such variation is needed for selection but also to avoid leading the population into extinction (Frankham et al., 2002). A crucial issue thus is a thorough understanding of the measures of coancestry between individuals and how they are affected by the relationship matrix used in the selection process, i.e., pedigree or genomic-based coancestries.

Traditionally, genealogical measures from pedigree records were used to calculate coancestry. As molecular markers became commonly used, estimates of genealogical coancestry from these markers were developed (Weir et al., 2006). It is only with the high-density panels that replacing genealogical coancestry with marker-based coancestry has become accepted as leading to more accurate predictions (Meuwissen et al., 2001; Meuwissen, 2007; Solberg et al., 2008) and to maintain more diversity in conservation programmes (de Cara et al., 2011). However, while the increase in accuracy in the EBVs using different marker types and densities is well-understood (Solberg et al., 2008; Jannink, 2010), the effect of different measures of coancestries in genomic and traditional selection has not received as much attention (Sonesson et al., 2012; Bjelland et al., 2013; Luan et al., 2014). For instance, genomic selection to estimate marker effects and predict the breeding values from them exploits the linkage disequilibrium between the markers in the panel and the causal mutations or QTL (Habier et al., 2007; de los Campos et al., 2010). When selection is performed via BLUP based on genomic relationships, the genetic gain is superior based on these relationships as compared to BLUP based on pedigree based relationships (Villanueva et al., 2005; Meuwissen, 2007) when the number of candidates for selection is large (Bastiaansen et al., 2012; Sonesson et al., 2012). Furthermore, selection based on genomic relationships also leads to lower increases in inbreeding and maintains more diversity (Sonesson et al., 2012; Liu et al., 2014).

In this study we analyse the effect of BLUP selection with four measures of coancestry on the genetic gain and on the increase in coancestry and inbreeding. For this purpose, we carry out simulations with three different genome-based relationship matrices and the matrix of genealogical relationships when inferring breeding values using BLUP. The three genomic measures of coancestry were: (1) based on shared segments of homozygosity (Fisher, 1954; Stam, 1980; Gusev et al., 2009), (2) using identity by state, that is, marker-by-marker similarity (Eding and Meuwissen, 2001; Caballero and Toro, 2002) and (3) based on a marker-by-marker count corrected by allelic frequencies (VanRaden, 2008). We measured the performance of selection with BLUP based on these four coancestries by analysing the genetic gain as measured with the true breeding values (TBVs).

2. Materials and Methods

2.1. Base Population

A base population was simulated with an effective size of 1000 individuals (half males, half females) during 10,000 generations until an equilibrium in the average genome-wide heterozygosity was reached. Every individual had a genome of 10 chromosomes of 1M with 10,100 biallelic positions each. Initially, every position in the genome carried alleles 0 or 1 at random, so that the average initial heterozygosity was 0.5. The mutation rate per position and generation was 2.5×10^{-3}. Every generation during the creation of the base population we firstly performed mutations in every individual, then chose a male and a female at random with replacement and produced an offspring with recombination. The number of recombinations per chromosome were sampled from a Poisson distribution and the recombination positions were drawn from a uniform distribution. The base populations were generated with a fortran 90 code available upon request.

2.2. Selection

We performed 100 replicates of each scenario here studied by selecting 1000 polymorphic positions from this base population to be later used as selective loci (also known as QTLs in the literature). We sampled these selective loci from positions with $0.05 < p_j < 0.95$, where p_j is the allelic frequency of allele 1

at locus *j*. Note thus that the 100 replicates are all created from one single base population by selecting different selected loci and different individuals in each replicate.

Founder individuals for each replicate were chosen at random from the base population without replacement, by drawing an equal number *N* of founder sires and dams from the base population to create generation 0. We then performed 6 generations of random mating to record the genealogy.

From generation 7 onwards we performed truncation selection for 15 generations (up to generation 21), by selecting the best 50% of the sires and 50% of the dams according to each individual's expected breeding value. These sires and dams were mated at random to produce *N* sires and *N* dams for the next generation.

The default parameters used in our simulations are $N = 50$, a marker density of 10,100 markers per chromosome and a trait with heritability of $h^2 = 0.25$. To have a thorough understanding of the dependence of the results on population size, heritability and marker density, we also studied the following scenarios: we evaluated population sizes $N = 10$ and $N = 30$, two other heritabilities of the quantitative trait ($h^2 = 0.10$ and 0.50) and two other lower marker densities (2525 and 5050 markers per chromosome). **Table 1** shows a summary of the simulated scenarios.

2.3. Calculation of Phenotypic Values and True and Estimated Breeding Values

We calculated the TBV of individual *i* as

$$TBV_i = \sum_{j=1}^{n_S} a_j \left(x_{ij} - 1\right), \tag{1}$$

where x_{ij} is the number of copies of the allele 1 that individual *i* has at the *j*-th selective locus, a_j is the effect of the allele 1 at position *j* and n_S is the number of selective loci. The values of the effects *a* were drawn from a Gaussian distribution with mean zero and variance one. The phenotypic values (y_i) of individuals were simulated as

$$y_i = \mu + TBV_i + e_i, \tag{2}$$

where e_i is an error term for individual *i*, which was normally distributed with mean zero and variance σ_e^2. The phenotypic average μ was set arbitrarily to be equal to 100, although this value does not affect the EBV. The variance σ_a^2 was calculated as the empirical variance of the TBVs in the base population and σ_e^2 was adjusted so that the heritability was the desired h^2. We had the phenotypic values for all individuals in the population.

EBV were calculated by solving Henderson's mixed model equations (Henderson, 1984) as follows:

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \frac{\sigma_e^2}{\sigma_a^2}\mathbf{A}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{EBV} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix}, \tag{3}$$

where **X** and **Z** are the incidence matrices for the fixed and random effects, respectively and **A** is the relationship matrix. We assumed the variance components to be known. Equation (3) provides the pedigree-based breeding values, while genomic based breeding values can be obtained by replacing **A** and σ_a^2 in Equation (3) by the following genomic relationships and variances.

2.3.1. Coancestry Estimates

The four following genetic relationship matrices, here defined as twice the coancestry coefficient, were used:

1. *Additive relationship matrix* (**A**): This was calculated using the coancestry coefficient between individuals *i* and *k*, $f_A(i, k)$ following (Malecot, 1948) as the probability that two alleles taken at random, one for each individual, are identical by descent (IBD).
2. *Marker-by-marker relationship matrix* (**G**): In this case, the coancestry coefficient between individuals *i* and *k*, $f_G(i, k)$, is the probability that two alleles at a given locus taken at random from each individual are equal (identical by state, IBS). In this study, $f_G(i, k)$ was calculated as $f_G(i, k) = \frac{1}{4M}\sum_{n=1}^{M}\sum_{l_i=1}^{2}\sum_{m_k=1}^{2} I_n(l_i, m_k)$ where *M* is the number of markers and $I_n(l_i, m_k)$ is the identity of gamete *l* from individual *i* with gamete *m* from individual *k* at marker *n* and takes a value of 1 if both alleles are identical and 0 otherwise.
3. *ROH-based relationship matrix* (**R**): Following the study by de Cara et al. (2013), the coancestry coefficient based on shared segments of the genome between individuals *i* and *k* was calculated as $f_R(i, k) = \frac{1}{4L}\sum_j \sum_{a_i=1}^{2}\sum_{b_k=1}^{2} L_j(a_i, b_k)$, where $L_j(a_i, b_k)$ is the length of the *j*-th shared segment measured over the gametes a_i and b_k of individuals *i* and *k* and *L* is the length of the genome. For a region to be considered a shared segment, we used a minimum length of 100 shared contiguous markers. The idea behind this segment-based relationship is that a segment shared between parents is a potential run of homozygosity (ROH) in the offspring.
4. *Marker-by-marker corrected by allele frequencies relationship matrix* (**V**): Following VanRaden (2008) a measure of coancestry $f_V(i, k)$ between individuals *i* and *k* can be

TABLE 1 | Parameters simulated for the different scenarios.

	N			h^2			SNPs		
	10	30	50	0.10	0.25	0.50	2525	5050	10,100
N	10	30	50	50	50	50	50	50	50
h^2	0.25	0.25	0.25	0.10	0.25	0.50	0.25	0.25	0.25
SNPs	10,100	10,100	10,100	10,100	10,100	10,100	2525	5050	10,100

calculated as

$$f_V(i,k) = \frac{1}{M}\sum_{n=1}^{M}\frac{(g_{in}-p_n)(g_{kn}-p_n)}{p_n(1-p_n)}, \quad (4)$$

where g_{in} refers to the gene frequency value genotypes 00, 01, and 11, coded as 1, 0.5, and 0, respectively, of individual i at locus n. Gene frequency is half the number of copies of the reference allele 1 and p_n is set at 0.5 (Forni et al., 2011).

Every generation we estimated the additive variance of the base population using restricted maximum likelihood (REML). We performed REML by using a Monte Carlo expextation-maximization (EM) algorithm (Guo and Thompson, 1991) to avoid the repeated matrix inversion required by exact algorithms (Meyer, 1991). Additive variances were estimated after six thousand iterations and discarding the first 1000. As for the base population, the fortran 90 code for the selection process is available upon request.

3. Results

As summarized in **Table 1**, we studied a combination of three population sizes, three heritabilities of the trait and three marker densities. The default case unless otherwise stated is the case of 10,100 markers per chromosome, heritability $h^2 = 0.25$ and a population size with 50 males and 50 females per generation.

3.1. Distribution of Coancestries

Most likely, the differences in our results are going to be due to the distribution of coancestries, as the different selection strategies here performed are based on the matrix of relationships between individuals. We show in **Figure 1** the distributions for the four measures of relationships prior to selection and give the variance within each figure. There we can see how the shape of the distribution of the genealogical coancestry is multimodal, given the sparse nature of the genealogical coancestry matrix and its distribution has the largest variance of all coancestry matrices, as well as the lowest mean. The distribution of coancestries f_V and f_G are fairly similar, the first one having a lower mean and a slightly larger variance although both distributions have a very small variance. Lastly, the distribution of coancestries f_R has a mean considerably lower than the other genomic coancestries f_V and f_G and a substantially larger variance.

3.2. Genetic Gain

Changes in TBVs obtained with the four relationship matrices for three population sizes $N = 10$, $N = 30$, and $N = 50$, three heritabilities $h^2 = 0.1$, $h^2 = 0.25$, and $h^2 = 0.50$, as well as three marker densities of 2525, 5050, and all 10,100 per chromosome are shown in **Figure 2** vs. generations. We only show results after generation 7, when selection starts. For a better comparison between the different coancestries here used, we show the value at each generation minus the initial value right before selection (i.e., at generation 7). Overall, all four methods performed similarly in terms of genetic gain for the sizes here studied. As expected, the final TBV increased with the number of individuals and with the heritability of the trait. The density of markers had no effect when selecting with the genealogical coancestry f_A, as expected, and, within the range of densities here studied no differences were detected in the genetic gains achieved by the genomic based estimates f_V and f_G. The most surprising result is that for a low density of markers, the genetic gain is larger performing selection based on f_R. It must be noticed that the size for a region of homozygosity to be considered as such was kept constant and thus, a ROH of 100 contiguous markers covers a much longer stretch than for 10,100 marker per chromosome. This is also surprising as it has been pointed out that the longer the ROH, the more correlated ROH-based inbreeding is with genealogical inbreeding.

3.3. Changes in Relatedness

We show in **Figure 3** results for the changes in each of the four measures used of coancestry with each selection scenario. We have used a logarithmic scale as overall, the differences between genealogical based selection and genomic based selection were very large. That is, line "A" shows the results for genealogical coancestry resulting from selecting based on this coancestry f_A and so on for scenarios G, R, and V. The results for inbreeding are not shown as they display a very similar pattern. In order to better appreciate the differences between the four measures of coancestries, we show $log\left[(1-f)/(1-f_7)\right]$ in **Figure 3**. In this way, we compare the speed of increase in each average coancestry scaled with their values at generation 7 (f_7), right before selection started. The increase in genealogical coancestry (the decay in this log scale) is the largest, followed by ROH-based coancestry. Changes in f_V and f_G are hardly distinguishable and very similar to f_R for small heritability. The smaller the population, the larger the increase in any measure of coancestry. The differences in f_G and f_V are hardly different as heritability increases from $h^2 = 0.25$ to $h^2 = 0.5$.

In **Figure 4** we show a similar plot for the change in pedigree based coancestry obtained under each selection scenario. All cases studied showed that the three genomic based selection led to lower increases in pedigree-based coancestry and the differences between the selection based on genomic relationships are hardly noticeable. The results are very similar for f_V and f_G based BLUPs on genealogical coancestry and it seems that f_R-based BLUP leads to slightly larger genealogical coancestries.

3.4. Diversity Maintained

As a measure of the diversity maintained we used f_G, as this is directly related to observed heterozygosity. In **Figure 5**, we show the changes on this marker-by-marker relatedness over generations when selection was carried out using the four strategies analyzed. As previously done for all coancestries and for genealogical coancestry, we show its rate of decrease by plotting $log(1-f_G)$ in **Figure 5**, minus this value right before starting selection $log(1-f_G(7))$ to compare all selection processes. Therefore, in this scale, the largest decrease means the largest increase in f_G.

It is important to highlight that the highest loss in genetic diversity (the largest increase in f_G) was observed for the selection based on the additive relationship matrix without exception. The fastest decay is for the smallest population size of $N = 10$ and

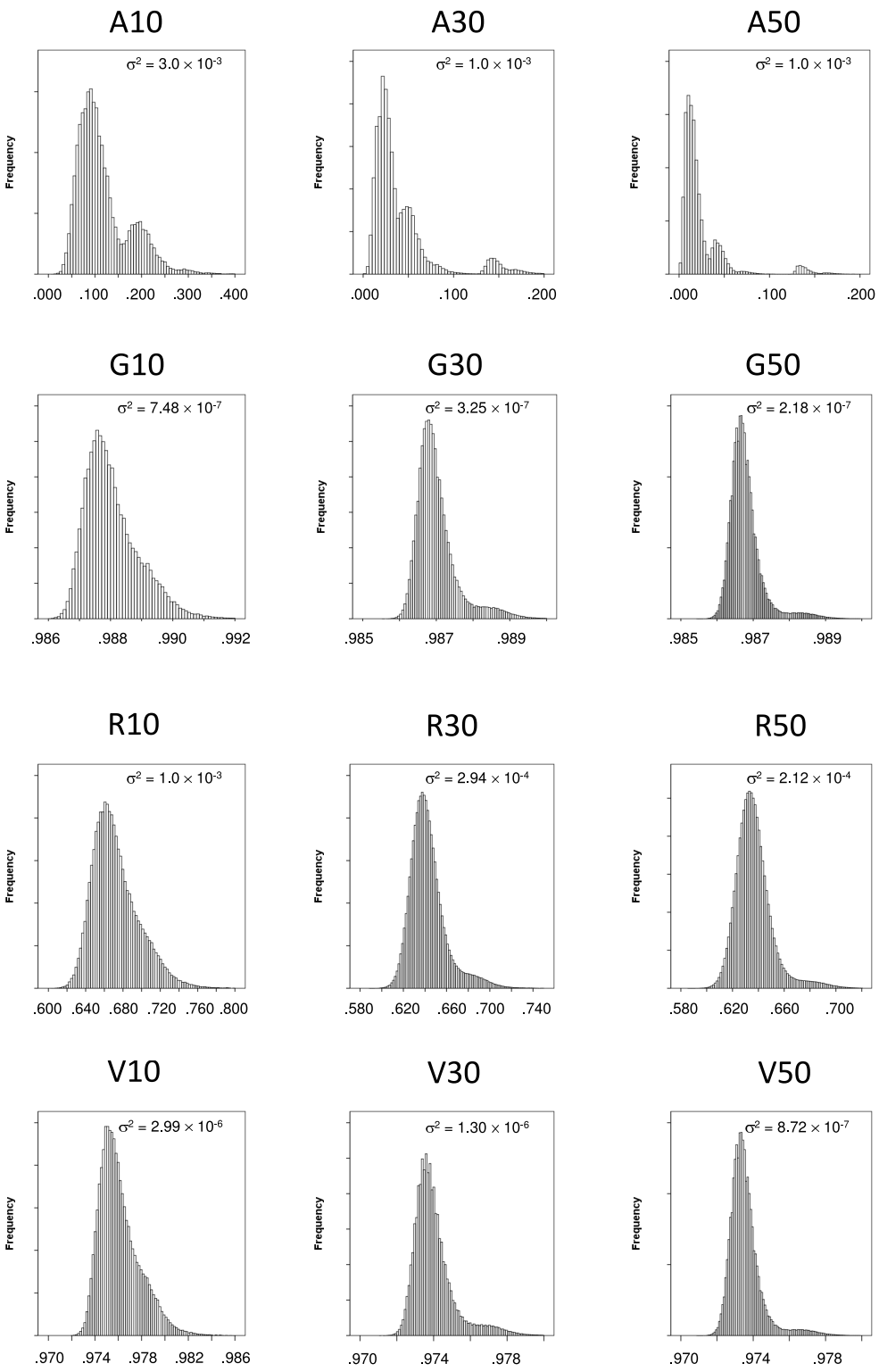


FIGURE 1 | Histograms of the coancestries at generation 6 right before selection. Top row shows the histogram for genealogical coancestry f_A for 10, 30, and 50 individuals from left to right. Similarly, the second row shows the histogram for molecular marker-by-marker coancestry f_G. The third row shows the histograms for segment-based coancestry f_R, for $N = 10$, $N = 30$, and $N = 50$ from left to right. The bottom row shows the histogram for molecular marker-by-marker coancestry corrected by allelic frequencies f_V, for $N = 10$, $N = 30$, and $N = 50$ from left to right. The variance of each histogram is given within each plot.

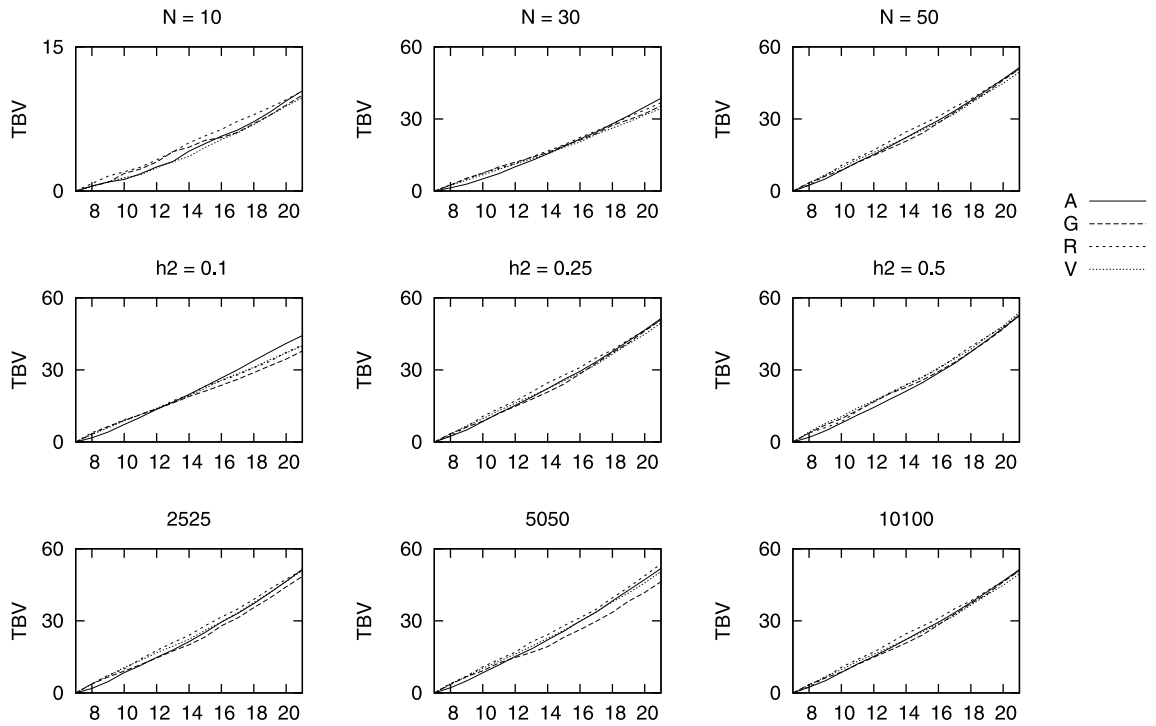


FIGURE 2 | Mean true breeding values (TBV) for different marker densities (bottom row), heritability (middle row), and population size (top row) vs. generations of selection. TBV values are shown minus the value right before truncation selection started. The default values are 50 sires and 50 dams, a heritability of 0.25 and 10, 100 markers, unless the value at the top of the figure indicates otherwise.

then for $N = 30$ and this decay is largest with decreasing population size than heritabilities or marker densities. Within each scenario, it seems that initially most diversity is maintained selecting with the genomic coancestries and the difference between f_G and f_V is small. The difference between f_G or f_V and f_R is small, though f_R-based BLUP can lead to slightly larger decreases in molecular coancestry than the other two genomic measures of relatedness, especially for small marker density. That is, f_R-based BLUP maintains slightly less genetic diversity than the other genomic based BLUPs.

4. Discussion

We have shown here results for truncation selection performed with four different measures of coancestry: f_A, f_G, f_R and f_V. All results shown are selecting the top 50% of sires and dams and we have compared results with three different population sizes, three different heritabilities of the selected trait and three different number of markers per chromosome.

We have performed 6 initial generations of random mating to have a deeper pedigree and have a fairer comparison between molecular markers which record the whole population history and genealogies, which are usually only stored when the selection programme starts.

There seems to be currently a consensus that genomic BLUP selection, whereby we mean selection based on genomic measures of relatedness, is superior to traditional pedigree-based BLUP selection (Daetwyler et al., 2007, 2010; Sonesson et al., 2012) in terms of higher genetic gain and lower increase in inbreeding. However, few studies have paid attention to the loss of genetic variability caused by each selection strategies of selection (Jannink, 2010; Bastiaansen et al., 2012; Heidaritabar et al., 2014; Liu et al., 2014). We discuss our main conclusions and the differences with these previous studies below.

4.1. On Genetic Gain

One of the main properties of BLUP is that by definition, the largest gain is obtained when the additive genetic variance of the base population is known. This is a difficult task, as for a large number of loci under selection which may be linked, the standard formula of $\sigma_a^2 = \sum_{j=1}^{n_S} 2p_j\left(1-p_j\right)a_j^2$ (Falconer and Mackay, 1996) does not apply. Furthermore, this variance is not appropriate when the performed BLUP relies on the genomic relationships f_G, f_R or f_V. Thus, we estimated the additive variance components using REML. While it is well-known that the estimates obtained with REML are more accurate for larger population sizes than the ones here studied, the differences between the four selection strategies here studied are small. We think that these differences are independent of whether the variance could have been better estimated. We believe that a more accurate estimate of the variance of the base population would lead to larger gains for all four BLUPs here performed and the differences in the trends would stay the same.

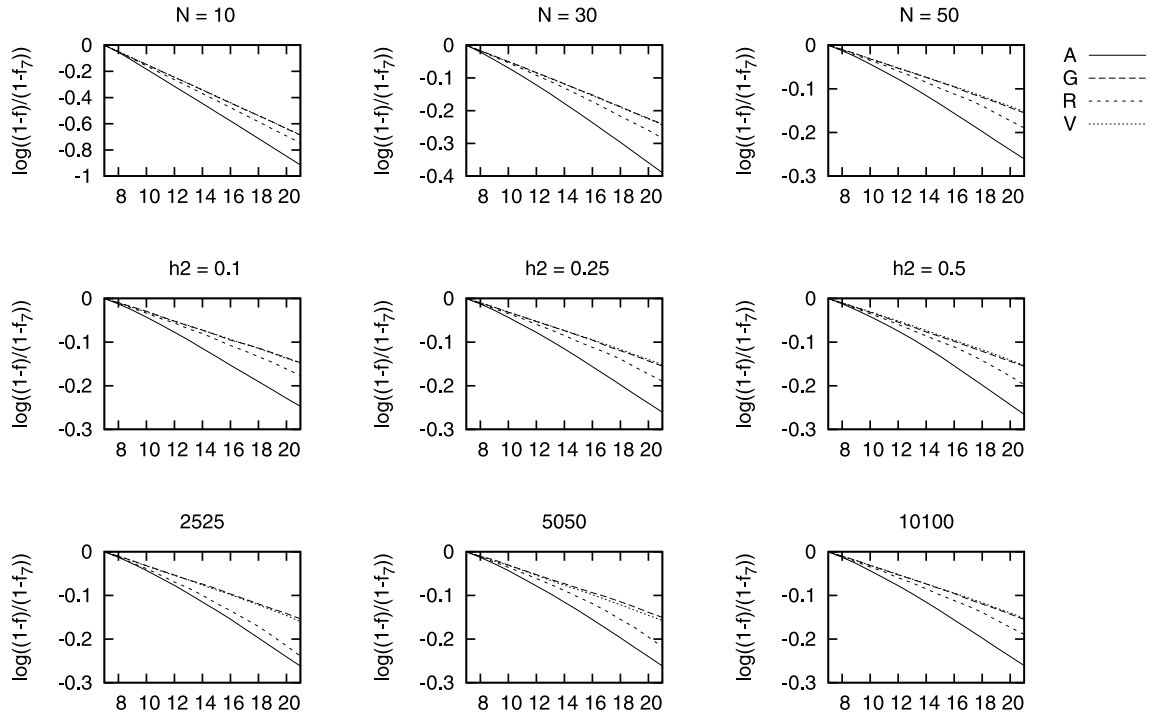


FIGURE 3 | Change in each coancestry for different marker densities (bottom row), heritability (middle row), and population size (top row) vs. generations of selection. The change in each coancestry is shown as $log\left(\frac{1-f}{1-f_7}\right)$.

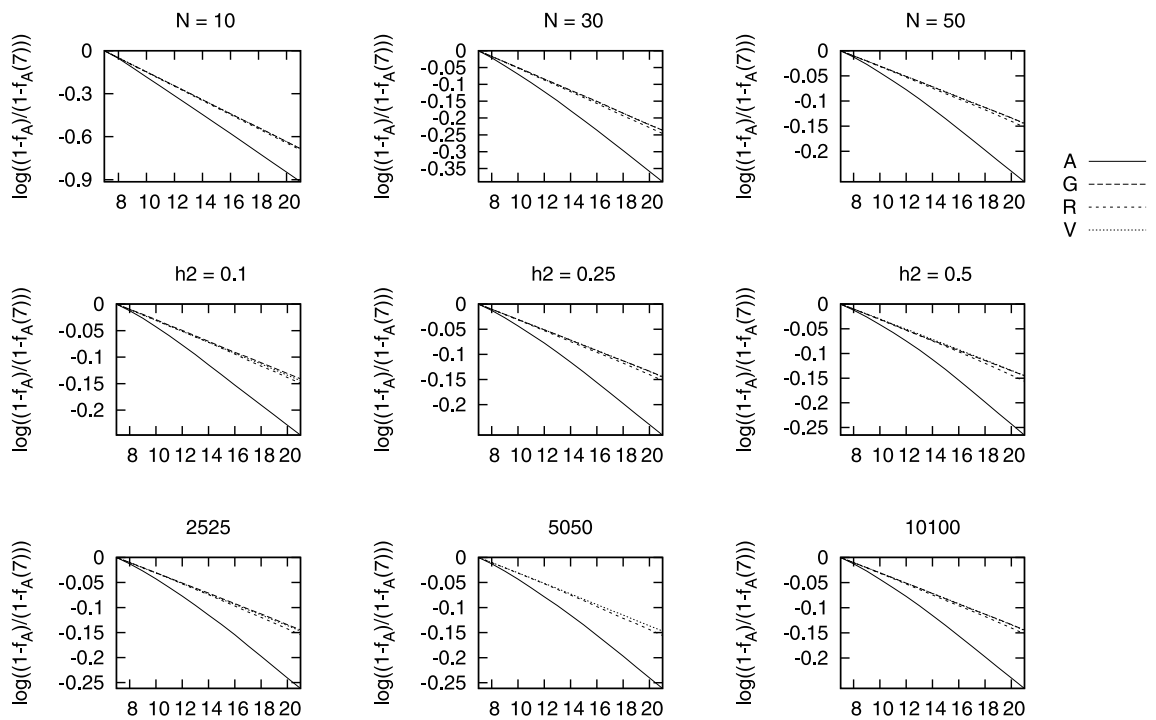


FIGURE 4 | Changes in genealogical coancestry for different marker densities (bottom row), heritability (middle row), and population size (top row) vs. generations of selection. The change in each coancestry is shown as $log\left(\frac{1-f_A}{1-f_A(7)}\right)$, where f_A is the genealogical coancestry at each generation and $f_A(7)$ is the value before selection starts.

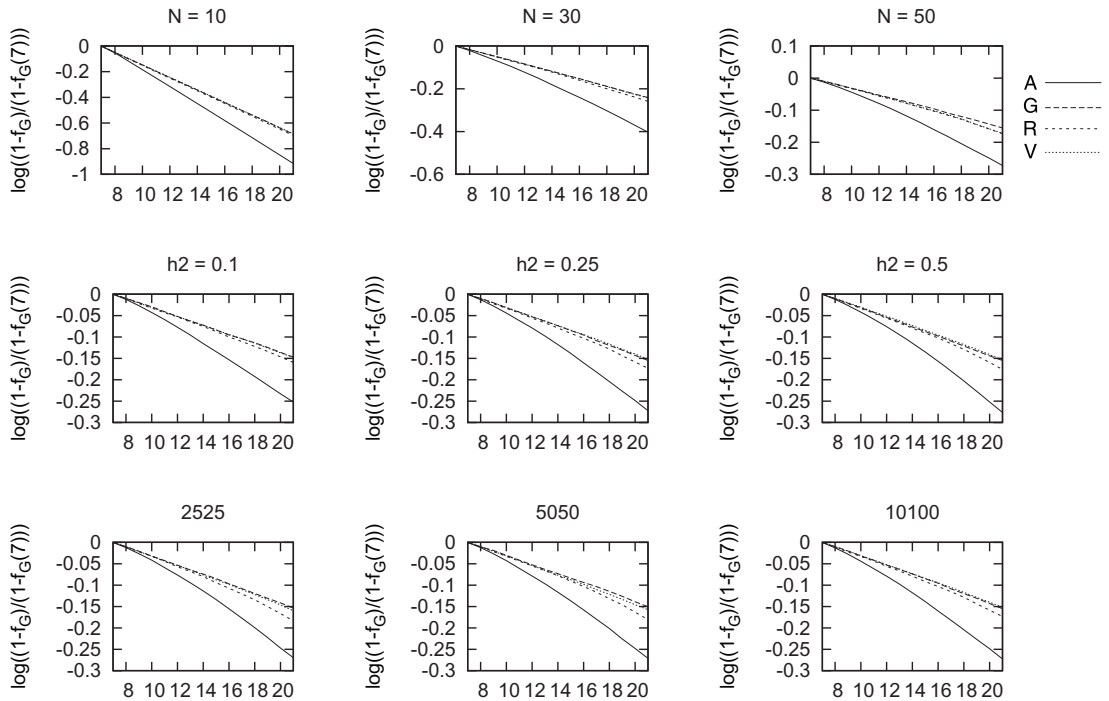


FIGURE 5 | Changes in molecular coancestry as a measure of change in diversity for different marker densities (bottom row), heritability (middle row), and population size (top row) vs. generations of selection. This change is shown as $log\left(\frac{1-f_G}{1-f_G(7)}\right)$, where f_G is the value of molecular coancestry at each generation and $f_G(7)$ is the value before selection starts.

Overall, the genetic gain was very similar with the four relationship matrices, although BLUP based on f_R performed slightly better than the other BLUPs in terms of gain for lower marker densities. It also performed somewhat better for small population size and the intermediate heritability here studied, at least up to generation 18 (i.e., after 10 generations of selection). It is worth emphasizing that for the lower marker densities here studied, we kept the same threshold size of 100 consecutive markers for a ROH to be considered as such. That means that for 2525 markers per chromosome, such ROH would cover a section of about 4 cM, while for 10,100 a ROH of 100 consecutive markers covers 1 cM. Thus, for higher marker densities, it is likely that the gains could be increased by using a larger threshold for what is considered a ROH.

As expected, the final TBVs were larger for larger population size and for higher trait heritability. This is due to the larger genetic variance for larger population sizes in which selection can act upon, while the negative effects of inbreeding are reduced with higher population sizes. It is however somewhat surprising that the differences are small in genetic gain with marker densities for the genomic relationships matrices, particularly for f_V and f_G. This could indicate that a density of 2525 markers per chromosome would give the same correlation between the true genomic relationship if we had the whole sequence and that estimated with such marker density (Rolf et al., 2010).

It is likely that the lack of differences in genetic gain between the genomic and pedigree based relationships stems from the fact that we use the marker data to infer the relationships, but not to estimate the marker effects. It is in this later scenario where genomic selection seems considerably superior to traditional pedigree BLUP, although it depends on having enough training generations where both phenotypes and genotypes are recorded, as reviewed recently by Van Eenennaam et al. (2014).

In genomic selection, markers that densely cover the genome are expected to be in complete or partial linkage disequilibrium with the trait under selection. Genomic prediction based on IBS information uses the family structure of the population (Habier et al., 2007), since the markers capture the linkage disequilibrium that arises from the family structure. Recently, Luan et al. (2014) have proposed an approach to predict genomic estimated breeding values from runs of homozygosity. This study indicates that runs of homozygosity yield a multi-locus measure of linkage disequilibrium and thus can account for larger chromosomal distances to capture linkage disequilibrium than genomic prediction based on IBS information. It is worth noting that in their study, Luan et al. (2014) used a somewhat different definition of segment that we have used here. They obtained slightly better predictions for the ROH-based scenarios than for other genomic-based scenarios. Our results seem to be in line with those obtained by Luan et al. (2014), although a more thorough analysis of both methods is required for a better comparison. The measure of ROH used by Luan et al. (2014) does not seem to require a threshold size for a run of homozygosity, but it requires knowledge of the mutation rates and the effective population size.

No significant differences were detected between the genetic gain obtained with f_G and f_V. The reason is that with the f_G approach alleles that are IBD and IBS can not be distinguished and are both included in the coancestry (and inbreeding) measures. To express both pedigree- and genomic-based estimates in the same scale several methodologies have been proposed Toro et al. (2011). However, these methods are generally inaccurate and their performances are very similar to those for f_G Toro et al. (2002).

Sonesson et al. (2012) compared breeding schemes by simulating truncation or optimum contribution selection. They estimated breeding breeding values based on genome- or pedigree-based BLUP and recorded trait information on full-sibs of the candidates. This study concluded that to control inbreeding it is necessary to account for it on the same basis as what is used to estimate breeding values. Our results are in general agreement to those of Sonesson et al. (2012) regarding the genetic gain both with genomic- and pedigree-based selection procedures and with those of Bastiaansen et al. (2012), where higher accuracies were obtained for the genomic methods than for traditional pedigree-based BLUP.

4.2. On Coancestries and Inbreeding

As we have shown in **Figure 4**, the largest increases in coancestries, and similarly for inbreeding, is for the genealogical coancestry compared to other genomic measures of coancestry. At the same time, this increase in genealogical coancestry is larger with traditional pedigree-based BLUP than for any other BLUP here performed. This is in line with what Sonesson et al. (2012) obtained using BLUP combined with optimal contributions to control the increase in inbreeding, that the rate of increase in pedigree coancestry is higher for the pedigree-based selection scenario than for the genome-based selection approaches. This can be observed regardless the population size, the true heritability, or the density of markers. Bastiaansen et al. (2012) showed similar differences between traditional pedigree-based and genomic-based BLUP. They also showed how this difference built up with generations and was hardly noticeable after one round of selection. This study showed that the increase in inbreeding hardly depended on the genomic architecture of the selected trait, which is in line with what we observe in **Figure 4**, where the increase in coancestry seems independent of the marker density or the heritability of the trait. In agreement with Bastiaansen et al. (2012), we have also shown that genomic-based BLUPs can track Mendelian sampling within families, which is not possible with genealogical-based BLUP. Our results are apparently in contrast with the recent study of Liu et al. (2014), who obtained a lower increase in inbreeding for the larger heritability 0.25 in their study compared to that obtained for $h^2 = 0.05$. This is most likely due to the fact that they looked at the results after 8 generations of selecting the top 25% candidates each generation, while we have performed selection on the top 50% candidates and looked at the increase of coancestry after 14 generations of selection. This shows the importance of understanding the dynamics at different generation intervals.

Liu et al. (2014) debated whether using genealogical records would be a good measure of inbreeding, as it reflects expected relationships and not the actual ones. They proposed measuring inbreeding then based on runs of homozygosity, and obtained that genomic-based BLUPs lead to lower increases on genealogical inbreeding as compared to phenotype BLUP, but this was not the case for inbreeding measured with ROHs. Our results for f_R are very similar to those here presented for f_A, and thus in the scenarios here studied, all genomic measures lead to lower increases in inbreeding whether we measure it with genealogies or with ROHs.

Our results show that the increase in genealogical coancestry seems slightly larger for ROH-based BLUP as compared to the other genomic-based BLUPs, although the differences are small.

4.3. On Diversity Maintained

It is well-known that selection reduces variation around the selected loci due to hitchhiking (Maynard-Smith and Haigh, 1974; Heidaritabar et al., 2014; Liu et al., 2014). Thus, if we aim at maintaining diversity while selecting favorables variants, it is important to understand which selection strategy works better overall. We evaluated f_G as a measure of diversity maintained in the selection procedures simulated in the present study indicated that all genomic estimates maintained more variability than the pedigree-based ones. This result is in agreement with those also observed using simulated data but in the context of conservation programmes (de Cara et al., 2011), and with previous results in genomic selection (Liu et al., 2014).

An interesting study by Jannink (2010) showed that more variation could be maintained by placing more weight on favorable variants that are at low frequencies. This can potentially maintain more diversity both on the selected loci and on neutral loci. According to that study, this strategy leads to larger gains in the long-term, and thus this strategy could be optimal depending on how long is the long-term. Based on this study, it would be worthwhile studying whether placing weight on rare haplotypes could lead to a compromise between genetic gains and diversity maintained.

The study by Heidaritabar et al. (2014) has shown that changes in allelic frequencies are more localized around the selected loci with genomic based BLUP, while pedigree based BLUP leads to similar changes throughout the genome. Thus, it seems that genomic selection can lead to quick losses in genetic variation in specific regions of the genome, and thus great care is required if these regions provide potential adaptation of the breed.

In agreement with Liu et al. (2014), we have obtained that a larger heritability leads to larger decreases in diversity maintained when selecting with traditional BLUP. Similarly to what happened with genealogical inbreeding, the loss of diversity does not seem to depend on heritability when selecting with genomic-based BLUPs.

Interestingly, ROH-based BLUP seems to lead to slightly larger losses in diversity than the other genomic BLUPs, but massively smaller than pedigree BLUP. Consequently, a deep study of the factors involved in the definition of a ROH could help to improve the genetic gain obtained with this estimator while also keeping the a very high genetic variability.

In conclusion, in this study conventional pedigree based selection, which has been used for decades, results in similar genetic

gains and does not maintain as much genetic variability as the genomic based selection methods. These results highlight the utility of genomic selection and also the need to manage the population variability using genomic information to preserve the future success of selection programs.

Author Contributions

All authors designed the study, performed the simulations and wrote the manuscript.

Acknowledgments

STRR and LAGC were funded by Ministerio de Economía y Competitividad grant number CGL2012-39861-C02-02. MARdC was funded by LabEx grant ANR-10-LABX-0003-BCDiv from Agence Nationale de la Recherche Investissements programme ANR-11-IDEX-0004-02. The authors are grateful to Carmen Rodriguez Valdovinos and Luis Silio for access to their server. MARdC benefited from access to the computer cluster of the Genotoul bioinformatics platform Toulouse Midi-Pyrénées.

References

Bastiaansen, J. W. M., Coster, A., Calus, M. P. L., van Arendonk, J. A. M., and Bovenhuis, H. (2012). Long-term response to genomic selection: effects of estimation method and reference population structure for different genetic architectures. *Genet. Sel. Evol.* 44:3. doi: 10.1186/1297-9686-44-3

Bjelland, D. W., Weigel, K. A., Vukasinovic, N., and Nkrumah, J. D. (2013). Evaluation of inbreeding depression in holstein cattle using whole-genome snp markers and alternative measures of genomic inbreeding. *J. Dairy Sci.* 96, 1–10. doi: 10.3168/jds.2012-6435

Caballero, A., and Toro, M. A. (2002). Analysis of genetic diversity for the management of conserved subdivided populations. *Cons. Genet.* 3, 289–299. doi: 10.1023/A:1019956205473

Daetwyler, H. D., Pong-Wong, R., Villanueva, B., and Woolliams, J. A. (2010). The impact of genetic architecture on genome-wide evaluation methods. *Genetics* 185, 1021–1031. doi: 10.1534/genetics.110.116855

Daetwyler, H. D., Villanueva, B., Bijma, P., and Woolliams, J. A. (2007). Inbreeding in genome wide selection. *J. Anim. Breed. Genet.* 124, 369–376. doi: 10.1111/j.1439-0388.2007.00693.x

de Cara, M. A. R., Fernandez, J., Toro, M. A., and Villanueva, B. (2011). Using genome-wide information to minimize the loss of diversity in conservation programmes. *J. Anim. Breed. Genet.* 128, 456–464. doi: 10.1111/j.1439-0388.2011.00971.x

de Cara, M. A. R., Fernandez, J., Toro, M. A., and Villanueva, B. (2013). Using genomic tools to maintain diversity and fitness in conservation programmes. *Mol. Ecol.* 22, 6091–6099. doi: 10.1111/mec.12560

de los Campos, G., Gianola, D., and Allison, D. B. (2010). Predicting genetic predisposition in humans: the promise of whole-genome markers. *Nat. Rev. Genet.* 11, 880–886. doi: 10.1038/nrg2898

Eding, H., and Meuwissen, T. H. E. (2001). Marker-based estimates of between and within population kinships for the conservation of genetic diversity. *J. Anim. Breed. Genet.* 118, 141–159. doi: 10.1046/j.1439-0388.2001.00290.x

Falconer, D. S., and Mackay, T. F. C. (1996). *Introduction to Quantitative Genetics*. Essex: Longman.

Fisher, R. A. (1954). A fuller theory of "junctions" in inbreeding. *Heredity* 8, 187–197.

Forni, S., Aguilar, I., and Misztal, I. (2011). Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information. *Genet. Sel. Evol.* 43:1. doi: 10.1186/1297-9686-43-1

Frankham, R., Ballou, J. D., and Briscoe, D. A. (2002). *Introduction to Conservation Genetics*. Cambridge: Cambridge University Press.

Garcia-Cortes, L. A., Legarra, A., Chevalet, C., and Toro, M. A. (2013). Variance and covariance of actual relationships between relatives at one locus. *PLoS ONE* 8:e57003. doi: 10.1371/journal.pone.0057003

Goddard, M. E. (2009). Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica* 136, 245–257. doi: 10.1007/s10709-008-9308-0

Goddard, M. E., and Hayes, B. (2007). Genomic selection. *J. Anim. Breed. Genet.* 124, 323–330. doi: 10.1111/j.1439-0388.2007.00702.x

Guo, S. W., and Thompson, E. A. (1991). Monte carlo estimation of variance component models for large complex pedigrees. *IMA J. Math. Appl. Med. Biol.* 8, 171–189.

Gusev, A., Lowe, J. K., Stoffel, M., Daly, M. J., Altshuler, D., Breslow, J. L., et al. (2009). Whole population, genome-wide mapping of hidden relatednesss. *Genome Res.* 19, 318–326. doi: 10.1101/gr.081398.108

Habier, D., Fernando, R. L., and Dekkers, J. C. M. (2007). The impact of genetic relationship information on genome-assisted breeding values. *Genetics* 177, 2389–2397. doi: 10.1534/genetics.107.081190

Hadfield, J. D., Wilson, A. J., Garant, D., Sheldon, B. C., and Kruuk, L. E. B. (2010). The misuse of blup in ecology and evolution. *Am. Nat.* 175, 116–125. doi: 10.1086/648604

Hayes, B. J., Visscher, P. M., and Goddard, M. E. (2009). Increased accuracy of artificial selection by using the realized relationship matrix. *Genet. Res.* 91, 47–60. doi: 10.1017/S0016672308009981

Heidaritabar, M., Vereijken, A., Muir, W. M., Cheng, T. M. H., Megens, H. J., Groenen, M. A. M., et al. (2014). Systematic differences in the response of genetic variation to pedigree and genome-based selection methods. *Heredity* 113, 503–513. doi: 10.1038/hdy.2014.55

Henderson, C. (1984). *Applications of Linear Models in Animal Breeding*. Guelph, ON: Guelph University Press.

Jannink, J.-L. (2010). Dynamics of long-term genomic selection. *Genet. Sel. Evol.* 42:35. doi: 10.1186/1297-9686-42-35

Liu, H., Sorensen, A. C., Meuwissen, T. H. E., and Berg, P. (2014). Allele frequency changes due to hitch-hiking in genomic selection programs. *Genet. Sel. Evol.* 46:8. doi: 10.1186/1297-9686-46-8

Luan, T., Yu, X., Dolezal, M., Bagnato, A., and Meuwissen, T. H. E. (2014). Genomic prediction based on runs of homozygosity. *Genet. Sel. Evol.* 46, 64. doi: 10.1186/s12711-014-0064-6

Makowsky, R., Pajewski, N. M., Klimentidis, Y. C., Vazquez, A. I., Duarte, C. W., Allison, D. B., et al. (2011). Beyond missing heritability: prediction of complex traits. *PLoS Genet.* 7:e1002051. doi: 10.1371/journal.pgen.1002051

Malecot, G. (1948). *Les Mathematiques de l'Heredite*. Paris: Masson et Cie.

Maynard-Smith, J., and Haigh, J. (1974). Hitch-hiking effect of a favorable gene. *Genet. Res.* 23, 23–35.

Meuwissen, T. H. E. (2007). Genomic selection: marker assisted selection on a genome wide scale. *J. Anim. Breed. Genet.* 124, 321–322. doi: 10.1111/j.1439-0388.2007.00708.x

Meuwissen, T. H. E., Hayes, B. J., and Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819–1829.

Meyer, K. (1991). Estimating variances and covariances for multivariate animal models by restricted maximum likelihood. *Genet. Sel. Evol.* 23, 67–83.

Rodriguez-Ramilo, S. T., Garcia-Cortes, L. A., and Gonzalez-Recio, O. (2014). Combining genomic and genealogical information in a reproducing kernel hilbert spaces regression model for genome-enabled predictions in dairy cattle. *PLoS ONE* 9:e93424. doi: 10.1371/journal.pone.0093424

Rolf, M. M., Taylor, J. F., Schnabel, R. D., McKay, S. D., McClure, M. C., Northcutt, S. L., et al. (2010). Impact of reduced marker set estimation of genomic relationship matrices on genomic selection for feed efficiency in angus cattle. *BMC Genet.* 11:24. doi: 10.1186/1471-2156-11-24

Solberg, T. R., Sonesson, A. K., Woolliams, J. A., and Meuwissen, T. H. E. (2008). Genomic selection using different marker types and densities. *J. Anim. Sci.* 86, 2447–2454. doi: 10.2527/jas.2007-0010

Sonesson, A. K., Woolliams, J. A., and Meuwissen, T. H. E. (2012). Genomic selection requires genomic control of inbreeding. *Genet. Sel. Evol.* 44:27. doi: 10.1186/1297-9686-44-27

Stam, P. (1980). The distribution of the fraction of the genome identical by descent in finite random mating populations. *Genet. Res.* 35, 131–155.

Toro, M. A., Barragan, C., Ovilo, C., Rodrigañez, J., Rodriguez, C., and Silio, L. (2002). Estimation of coancestry in iberian pigs using molecular markers. *Cons. Genet.* 3, 309–320. doi: 10.1023/A:1019921131171

Toro, M. A., Garcia-Cortes, L. A., and Legarra, A. (2011). A note on the rationale for estimating genealogical coancestry from molecular markers. *Genet. Sel. Evol.* 43:27. doi: 10.1186/1297-9686-43-27

Van Eenennaam, A. L., Weigel, K. A., Young, A. E., Cleveland, M. A., and Dekkers, J. C. M. (2014). Applied animal genomics: results from the field. *Ann. Rev. Anim. Biosci.* 2, 105–139. doi: 10.1146/annurev-animal-022513-114119

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91, 4414–4423. doi: 10.3168/jds.2007-0980

Villanueva, B., Pong-Wong, R., Fernandez, J., and Toro, M. A. (2005). Benefits from marker-assisted selection under an additive polygenic genetic model. *J. Anim. Sci.* 83, 1747–1752.

Visscher, P. M., Medland, S. E., Ferreira, M. A. R., Morley, K. I., Zhu, G., Cornes, B. K., et al. (2006). Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. *PLoS Genet.* 2:e41. doi: 10.1371/journal.pgen.0020041

Weir, B. S., Anderson, A. D., and Hepler, A. B. (2006). Genetic relatedness analysis: modern data and new challenges. *Nat. Rev. Genet.* 7, 771–780. doi: 10.1038/nrg1960

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Rodríguez-Ramilo, García-Cortés and de Cara. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in GENETICS

ORIGINAL RESEARCH ARTICLE
published: 09 September 2014
doi: 10.3389/fgene.2014.00307

Genome-wide association and pathway analysis of feed efficiency in pigs reveal candidate genes and pathways for residual feed intake

Duy N. Do[1], Anders B. Strathe[1,2], Tage Ostersen[2], Sameer D. Pant[1] and Haja N. Kadarmideen[1]*

[1] Section of Animal Genetics, Bioinformatics and Breeding, Department of Veterinary Clinical and Animal Sciences, Faculty of Health and Medical Sciences, University of Copenhagen, Frederiksberg, Denmark
[2] Pig Research Centre, Danish Agriculture and Food Council, Copenhagen, Denmark

Edited by:
Johannes Arjen Lenstra, Utrecht University, Netherlands

Reviewed by:
Robert John Tempelman, Michigan State University, USA
Mahdi Saatchi, Iowa State University, USA

***Correspondence:**
Haja N. Kadarmideen, Section of Animal Genetics, Bioinformatics and Breeding, Department of Veterinary Clinical and Animal Sciences, Faculty of Health and Medical Sciences, University of Copenhagen, 1870 Frederiksberg C, Denmark
e-mail: hajak@sund.ku.dk

Residual feed intake (RFI) is a complex trait that is economically important for livestock production; however, the genetic and biological mechanisms regulating RFI are largely unknown in pigs. Therefore, the study aimed to identify single nucleotide polymorphisms (SNPs), candidate genes and biological pathways involved in regulating RFI using Genome-wide association (GWA) and pathway analyses. A total of 596 Yorkshire boars with phenotypes for two different measures of RFI (RFI1 and 2) and 60k genotypic data was used. GWA analysis was performed using a univariate mixed model and 12 and 7 SNPs were found to be significantly associated with RFI1 and RFI2, respectively. Several genes such as *xin actin-binding repeat-containing protein 2* (*XIRP2*), *tetratricopeptide repeat domain 29* (*TTC29*), *suppressor of glucose, autophagy associated 1* (*SOGA1*), *MAS1*, *G-protein-coupled receptor (GPCR) kinase 5* (*GRK5*), *prospero-homeobox protein 1* (*PROX1*), *GPCR 155* (*GPR155*), and *FYVE domain containing the 26* (*ZFYVE26*) were identified as putative candidates for RFI based on their genomic location in the vicinity of these SNPs. Genes located within 50 kbp of SNPs significantly associated with RFI and RFI2 (q-value $\leq$ 0.2) were subsequently used for pathway analyses. These analyses were performed by assigning genes to biological pathways and then testing the association of individual pathways with RFI using a Fisher's exact test. Metabolic pathway was significantly associated with both RFIs. Other biological pathways regulating phagosome, tight junctions, olfactory transduction, and insulin secretion were significantly associated with both RFI traits when relaxed threshold for cut-off p-value was used ($p \leq 0.05$). These results implied porcine RFI is regulated by multiple biological mechanisms, although the metabolic processes might be the most important. Olfactory transduction pathway controlling the perception of feed via smell, insulin pathway controlling food intake might be important pathways for RFI. Furthermore, our study revealed key genes and genetic variants that control feed efficiency that could potentially be useful for genetic selection of more feed efficient pigs.

Keywords: GWAS, pigs, pathway analysis, residual feed intake

INTRODUCTION

Residual feed intake is defined as the difference between the observed feed intake and the feed intake predicted based on production traits such as average daily gain and backfat thickness. RFI is a sensitive and accurate indicator of feed efficiency in livestock that is being increasingly accepted as an alternative measure for feed efficiency in livestock species. Genetic selection for animals with reduced RFI can be advantageous from both economic and environmental perspectives (Dekkers and Gilbert, 2010; Cruzen et al., 2012; Saintilan et al., 2013). However, genetic variants and biological mechanisms regulating RFI need to be identified, which would help to improve genetic selection for this trait. GWA, a hypothesis-free approach that uses a large numbers of SNPs spread throughout the genome to identify quantitative trait loci (QTL) potentially harboring candidate loci, has been widely used to explore the genetics underlying complex traits. Past studies have led to the identification of many QTLs influencing feed conversion ratio (FCR) in pigs[1]. FCR is currently the only available measure of feed efficiency that is included in the selection index for the Danish pig breeds. However, ratio traits such as FCR are not ideal for statistical and biological reasons (Gunsett, 1984) and the accurate definition for feed efficiency in animals is still being debated. Recently, several studies have been conducted to identify QTLs and candidate genes putatively influencing RFI in pigs. Using a

Abbreviations: Bp, base pairs; dEBV, deregressed estimated breeding values; EBV, estimated breeding values; FDR, false discovery rate; GWA, genome-wide association; MAF, minor allele frequency; Mb, mega base pairs; QTL, quantitative trait locus; RFI, residual feed intake; SNP, single nucleotide polymorphism.

[1] http://www.animalgenome.org/cgi-bin/QTLdb/SS/index

Piétrain–Large White backcross population, Gilbert et al. (2010) identified QTLs on pig chromosomes (SSC) 5 and 9 for RFI in growing pigs. In Yorkshire pigs, a GWA study revealed several significant SNPs on SSC 2, 3, 5, 7, 8, 9, 14, and 15 influencing RFI (Onteru et al., 2013). A candicate gene study performed by Fan et al. (2010) validated these SNPs in *FTO* and *TCF7L2* genes as genetic markers for RFI in pigs. Recently, Shirali et al. (2013) detected novel QTLs for residual energy intake on SSC 2, 4, 7, 8, and 14 in a crossed populations (Pietrain grand-sires crossed with grand-dams bred from a three-way cross of Leicoma boars with Landrace × Large White dams). Sanchez et al. (2014) detected a SNP on SSC 6 for RFI in Large White pigs.

Recently, we have identified significant SNPs on SSC 1, 9, and 13 for RFI in Duroc pigs (Do et al., 2014). Danish Durocs, used as terminal sires in combination with crossbred LY sows (Landrace × Yorkshire), are bred with a higher emphasis on growth and feed efficiency traits compared to Yorkshire pigs, where the emphasis is considerably more on improving litter size. Given these differing emphases on selection, it is reasonable that the genetic architeture of these two breeds differs with respect to traits like feed efficiency that are targetted more intesively for selection within Durocs. In accordance with this, we have found that the genetic variation (heritability) of RFI is higher in Yorkshire compared to Duroc pigs (Do et al., 2013a). Therefore, while the biological mechanisms are likely conserved even across species (Mayr, 1963; Raff, 1996), the genetic regulation of these mechanisms is not necessarily conserved, and investigating the genetics underlying the same phenotype in a different breed could provide novel insights into the biological mechanisms underlying feed efficiency. Comparing findings of genomic investigations on different breeds that have differing linkage disequilibrium (LD) structure could also potentially assist in narrowing the boundaries of putative QTL regions.

While GWA studies have been reasonably successful, they often focus on a top few significant SNPs while ignoring other SNPs with lower significance levels that could still be biologically relevant. Gene set enrichment and pathway analyses using publicly available biological databases could potentially complement efforts to identify causal loci for complex traits, as has been shown in previous studies (Kadarmideen, 2008; Torkamani et al., 2008; Wang et al., 2010). These approaches, instead of relying solely on statistically associated genetic variants, focus on biological pathways that are mediated by genes located in the vicinity of these variants. Such approaches have been shown to provide valuable insights into the biology underlying complex phenotypes (Kadarmideen et al., 2006; Farber, 2013; Kadarmideen, 2014). Therefore, the objective of our study was to use both GWA and pathway analyses to identify SNPs, genes, and biological pathways that could potentially influence RFI in Yorkshire pigs.

MATERIALS AND METHODS

ESTIMATION OF RESIDUAL FEED INTAKE AND DEREGRESSED ESTIMATED BREEDING VALUES

Data were recorded during a 5-year period (2008–2012) and supplied by the Pig Research Centre of the Danish Agriculture and Food Council. A total of 596 Yorkshire pigs had both phenotypic (RFI) and genotypic records (based on PorcineSNP60 Illumina iSelect BeadChip). The method of calculation of RFI has been previously discussed in detail (Do et al., 2013a). In summary, RFI was computed as the difference between the observed average daily feed intake and the predicted daily feed intake using two statistical models. In the first model (RFI1), predicted daily feed intake was estimated using linear regression of daily feed intake on initial test weight (BWd) and average daily gain from 30 to 100 kg, whereas in the second model (RFI2), backfat was used as an additional regressor. The EBVs for RFI were calculated using a univariate animal model where barn–year–season were used as fixed effects and the effect of pen and the additive genetic effect were treated as random effects. The pedigree was traced back to January, 1971 and included 14,681 pigs with 1951 sires, 6766 dams. These EBVs were further deregressed as previously described (Ostersen et al., 2011; Do et al., 2013b), following the deregression procedure of Garrick et al. (2009). This procedure adjusts for ancestral information, so that the deregressed EBV (dEBVs) only includes information from individual animals and their descendants. Since our resource population consists of 5337 pigs of which only 1564 pigs had genotypic records, the use of deregressed proofs was intended to maximize use of phenotypic information from non-genotyped pigs. Because the dEBVs have unequal variances, they should be used in a weighted analysis. The weight for the *i*th animal was estimated as:

$$w_i = \frac{(1 : h^2)}{[(c + ((1 - r_i^2)/r_i^2))h^2]}$$

in which c was the part of the genetic variance that was assumed to be not explained by markers ($c = 0.1$), h^2 was the heritability of the trait, and r_i^2 was the reliability of the dEBV of the *i*th animal.

GENOTYPING AND DATA QUALITY CONTROL

The details of the resource population used and DNA collection were described in Henryon et al. (2001). For genotyping, genomic DNA was isolated from tissue by treatment with proteinase K followed by sodium chloride precipitation and SNPs were genotyped on the PorcineSNP60 Illumina iSelect BeadChip. Data quality control prior to GWA analyses was implemented by discarding animals and SNPs with a call rate <0.95, SNPs deviating from Hardy Weinberg equilibrium ($p < 0.0001$) and SNPs with a MAF < 0.05.

LINEAR MIXED MODEL USED FOR GENOME WIDE ASSOCIATION ANALYSES

A univariate linear mixed model was implemented to test the association between each SNP and RFI. The model was similar with previous GWA analysis in Duroc pigs (Do et al., 2014). In summary, the model for each SNP (analyzed individually) was as follows:

$$y = 1\mu + Za + mg + e$$

where y is the vector of dEBVs for RFI, 1is a vector of 1s with length equal to number of observations, μ is the general mean, Z is an incidence matrix relating phenotypes to the corresponding random polygenic effect, a is a vector of the random polygenic

effect $\sim N(0, A\sigma_u^2)$ where A is the additive relationship matrix and σ_u^2 is the polygenic variance, m is a vector with genotypic indicators (−1, 0, or 1) associating records to the marker effect, g is a scalar of the associated additive effect of the SNP, and e is a vector of random environmental deviates: $N(0, W^{-1}\sigma_e^2)$ where σ_e^2 is the general error variance and W is the diagonal matrix containing weights of the dEBVs. The model was fitted by restricted maximum likelihood (REML) using the DMU software (Madsen et al., 2006) and testing was done using a Wald test against a null hypothesis of $g = 0$. The Wald test was based on a t-distribution and regression coefficients and SEs were obtained by solving linear mixed model equations using DMU (Madsen et al., 2006). This test was done by the t-distribution function "pt()" in R with $p = 2$*pt[abs(β/SE), (n − 3), log = FALSE, lower.tail = FALSE; where β is regression coefficient, SE is standard error estimated based on the inverse of the mixed model coefficient matrix and n is number of SNPs in the genotype data]. Bonferroni corrected significance threshold, used to account for multiple comparisons, was estimated at a $p = 1.31e{-}06$. However, Bonferroni correction is known to be overly conservative especially when genetic data exhibits high LD, which could produce false negative results (Duggal et al., 2008). Therefore, in our analyses we considered a less conservative significance threshold of a $p = 1e{-}04$ in order to account for multiple tests. The p was chosen here based on a Bonferroni adjustment only for the number of independent tests that was in turn inferred by the number of principal components accounting for a 99% of the variance of the SNP matrix (Gao et al., 2008). Moreover, to further characterize candidate regions affecting RFI, we performed LD block analyses for the chromosomal regions with multiple (or the most) significant SNPs clustered. The blocks were defined based on criteria suggested by Gabriel et al. (2002) which implemented in Haploview (Barrett et al., 2005).

PATHWAY ANALYSES

Assignment of genes to SNPs

Assigning genes close to a few SNPs with high statistical significance, and ignoring many SNPs with lower significance levels could result in missing out on key putative candidates and associated pathways. Hence, we used the procedure of controlling false discovery rate (FDR; Benjamini and Hochberg, 1995) to select SNPs for pathways analyses. All SNPs with a FDR (or q-value) ≤0.2 were used to identify putative candidate genes. Based on previous studies, we also included pathway annotations associated with genes within 50 kb of SNPs associated with RFI at a nominal significance thresold of 0.05, in pathway analyses (Choquette et al., 2012; Peñagaricano et al., 2013). Positional candidate genes, located within 50 kb of these SNPs, were identified using function GetNeighGenes() in the NCBI2R package[2] for R program (R Development Core Team, 2008). The distance of 50 kb was used in order to capture proximal regulatory and other functional regions close to the gene. Moreover, several studies showed that the average LD was high in pigs. The average distance between adjacent SNP pairs (with average $r^2 = 0.5$) was around 72 kb in the current population (Wang et al., 2013). Therefore, the distance of 50 kb was suitable to capture the causal genes/SNPs. Individually, each gene was considered to be significantly associated with RFI if a SNP with a q-value ≤0.2 (as well as for relxed thresold at a $p \leq 0.05$) was located either inside the genic region or within 50 kb of the genes.

Assignment of genes to Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways and pathway analyses

For functional annotation, Kyoto Encyclopedia of Genes and Genomes[3] (KEGG) was used for getting pathways. To assign genes to pathways, we used the function GetPathways() in NCBI2R package; and to get number and names of genes in each pathway, the mapPathwaytoname() function[4] was used. Because RFI was production trait in pigs, the assigned pathways belonging to human diseases and drug development categories were removed. To determine whether a pathway term was significantly associated with RFI, we tested if genes significantly associated with RFI were overrepresented amongst all the genes of any given pathway. This association analysis was performed using a Fisher's exact test via the fisher.test() function in R.

RESULTS

GENOME-WIDE ASSOCIATION ANALYSES

After data quality control, a total of 37,192 SNPs and 596 pigs remained in the final dataset for GWA analyses. Eleven SNPs were significantly associated with RFI1 (**Figure 1**), and seven were associated with RFI2 (**Figure 2**); two SNPs (MARC0027992 and ASGA0039145) being associated with both traits [$p \leq 1e{-}04$ (0.0001)]. The most significant associations were found between ASGA0039145 and RFI1 ($p = 7.76e{-}06$) and between MARC0027992 and RFI2 ($p = 2.47e{-}05$). Significant SNPs associated with RFI1 were located on SSC 3, 7, 8, 9, 15, and 17 meanwhile those for RFI2 were located in SSC 1, 7, 8, 10, 14, and 15 (**Table 1**). A search for genes located in the immediate 50 kbp vicinity of these SNPs revealed *XIRP2*, *TTC29*, *SOGA1*, *MAS1*, *GRK5*, *PROX1*, *GPR155*, and *ZFYVE26* as putative candidates. Two LD blocks were detected on a candidate region (86–88 Mb on SSC8) which associated with both RFI (**Figure 3**).

PATHWAY-BASED ASSOCIATION ANALYSES

Based on results from GWA analyses, a total of 402 SNPs associated with RFI1 and another 304 SNPs associated with RFI2 (based on a FDR threshold of q-value ≤0.2) were used to locate 339 and 304 genes, respectively, within 50 kb of these SNPs for pathway analyses. A total of 21,296 genes in 50 kb flanking regions of SNPs which passed QC was used as background for enrichment test. Pathway analysis tests for KEGG pathways revealed a metabolic pathway to be associated with both RFI1 ($p = 0.008$) and RFI2 ($p = 0.002$); and an additional olfactory transduction pathway only associated with RFI2 ($p = 0.03$). Repeating pathway analyses after relaxing the significance threshold to $p \leq 0.05$, revealed 15 and 12 pathways associated with RFI1 and RFI2, respectively (Table S1 in Supplementary Materials), of

[2]http://cran.r-project.org/web/packages/NCBI2R/index.html

[3]http://www.genome.jp/kegg/

[4]http://biobeat.wordpress.com/category/r/

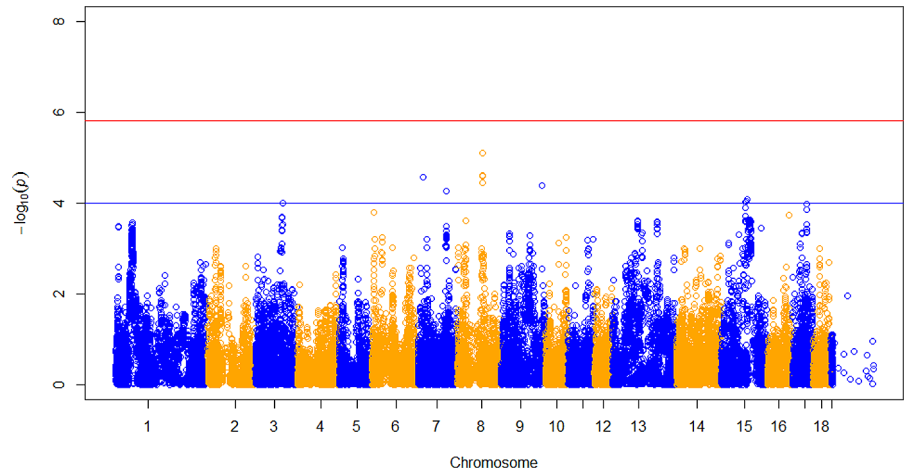


FIGURE 1 | Manhattan plot of genome-wide *p* of association for residual feed intake 1 (RFI). The horizontal red and blue line represents the Bonferroni ($p = 1.31e{-}06$) and genome-wide significance threshold ($p \leq 1e{-}04$), respectively.

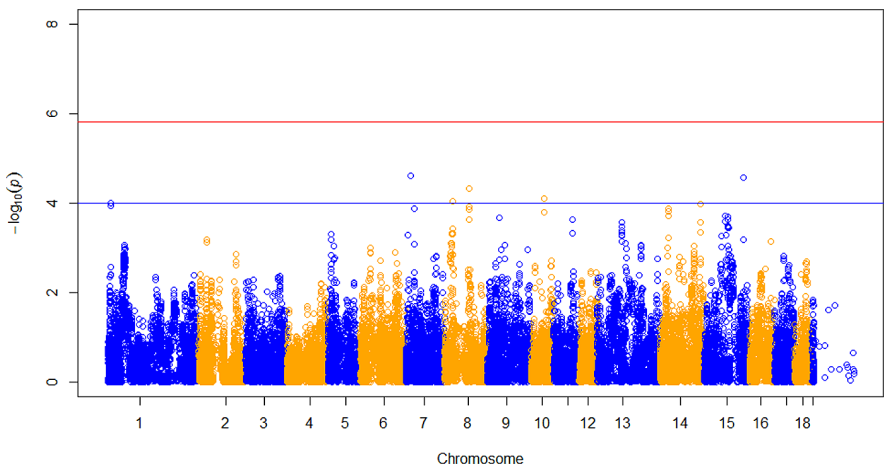


FIGURE 2 | Manhattan plot of genome-wide *p* of association for RFI 2. The horizontal red and blue line represents the Bonferroni ($p = 1.31e{-}06$) and genome-wide significance threshold ($p \leq 1e{-}04$), respectively.

which nine pathways were commonly associated with both RFI phenotypes. The pathways associated with both RFI phenotypes included metabolic pathway, olfactory transduction, tight junction and phagosome pathway that were associated with both RFI traits at $p \leq 0.01$.

DISCUSSION

The statistical and bioinformatics methods used in the current study are similar to those applied in a separate study aiming to identify QTLs influencing RFI in Duroc pigs (Do et al., 2014). However, in order to account for multiple comparisons, we used a less stringent genome-wide significance threshold ($p \leq 0.0001$) that adjusts the nominal significance threshold only for the number of independent tests that are possible given a particular dataset. The Bonferroni correction that was used in the previous study (Do et al., 2014), is known to overcompensate for multiple testing, especially when applied to correlated data. As such, the use of a relaxed threshold was expected to decrease the number of false negatives thereby increasing power. Several studies have applied pathway analysis on GWA datasets and reported pathways and GOs associated with backfat thickness (Fontanesi et al., 2012), feeding behavior (Do et al., 2013b), and RFI (Do et al., 2014) in pigs. However, these studies focused on genes in close proximity to significantly associated SNPs based on stringent genome-wide thresholds like Bonferroni, while ignoring many SNPs with lower significance levels. Therefore, we used a lower significance threshold including all genes located near SNPs associated with RFI at a q-value ≤ 0.2 in the current study. (Peñagaricano et al., 2013) also used more relaxed threshold with $p \leq 0.05$ and detected many significant pathways and network from GWA data for bull fertility traits. Moreover, since the porcine genome contains

Table 1 | Genome-wide significant SNPs for residual feed intake in Yorkshire.

Trait[1]	SNP	SSC[2]	Positions	*P*	Nearest gene	Gene positions[3]	Gene names
RFI1	ALGA0085846	15	82,835,779	9.7e−05	*XIRP2*	15:82,491,704-82,597,617	Xin actin-binding repeat-containing protein 2
RFI1	ALGA0085859	15	83,426,312	9.1e−05	*ENSSSCG00000023421*	15:83,412,046-83,526,943	Novel gene
RFI1	ALGA0114110	8	87,012,660	3.6e−05	*TTC29*	8:86,904,297-86,936,988	Tetratricopeptide Repeat Domain 29
RFI1	ASGA0039145	8	87,177,392	7.8e−06	*ENSSSCG00000009034*	8:87,200,641-87,201,588	Novel gene
RFI1	ASGA0039146	8	87,161,359	2.5e−05	*ENSSSCG00000009034*	8:87,200,641-87,201,588	Novel gene
RFI1	ASGA0089759	8	87,000,474	2.6e−05	*TTC29*	8:86,904,297-86,936,988	Tetratricopeptide Repeat Domain 29
RFI1	ASGA0103232	9	142,526,652	4.2e−05	*PROX1*	9:142,428,903-142,469,651	Prospero homeobox protein 1
RFI1	MARC0027992	7	18,810,414	2.7e−05	*ENSSSCG00000018237*	7:18,702,099-18,702,261	U1 spliceosomal RNA
RFI1	MARC0082350	15	89,623,572	8.4e−05	*GPR155*	15:89,587,688-89,695,502	G protein-coupled receptor 155
RFI1	ALGA0043644	7	97,889,360	5.4e−05	*ZFYVE26*	7:97,825,802-97,898,510	Zinc finger, FYVE domain containing 26
RFI1	H3GA0010038	3	94,756,601	1.0e−04	*ENSSSCG00000008418*	3: 95,389,252-95,546,928	Novel gene
RFI1	MARC0067053	17	45,539,648	1.1e−04	*SOGA1*	17: 45,467,693-45,539,349	Suppressor of glucose, autophagy associated 1
RFI2	ALGA0117721	10	44,751,617	8.1e−05	*ENSSSCG00000011017*	10:44,914,173-44,928,565	Novel gene
RFI2	ASGA0039145	8	87,177,392	4.7e−05	*ENSSSCG00000009034*	8:87,200,641-87,201,588	Novel gene
RFI2	DRGA0008464	8	27,798,101	9.0e−05	*RPS18*	8:28,015,649-28,016,107	Ribosomal protein S18
RFI2	H3GA0045110	15	136,035,441	2.7e−05	*ENSSSCG00000026566*	15:135,772,108-135,772,198	Novel microRNA
RFI2	MARC0027992	7	18,810,414	2.5e−05	*ENSSSCG00000018237*	7:18,702,099-18,702,261	U1 spliceosomal RNA
RFI2	MARC0072970	1	9,324,586	1.0e−04	*MAS1*	1: 9,326,800-9,327,777	MAS1 oncogene
RFI2	H3GA0042665	14	140,742,709	1.0e−04	*GRK5*	14: 140,637,968-140,875,398	G protein-coupled receptor kinase 5

[1] *RFI1: residual feed intake 1; RFI2: residual feed intake 2.*
[2] *Pig chromosome.*
[3] *Based on Sscrofa10.2 (GCA_000003025.4) at http://www.ensembl.org/Sus_scrofa/Info/Index*

many uncharacterized genes, widely used annotation servers like DAVID[5] and GOEAST[6] are of limited use, as they perform poorly in converting porcine gene IDs. Therefore, we performed functional annotation of genes by directly querying KEGG databases that are better able to handle porcine gene IDs.

Comparing results from our previous study targeting a Duroc resource population almost twice in size, we were able to identify overlapping QTLs at approximately 87 Mb on SSC 8, and at approximately 136 Mb on SSC 15 (**Table 1**; Do et al., 2014). The proportionally small number of overlapping QTLs is in agreement with other studies that have investigated QTLs for a specific trait within different pig breeds (Gregersen et al., 2012). Furthermore, despite of substantial differences in the location of QTL regions between the two breeds, pathway analyses identified many pathways (e.g., insulin regulation related pathways and cellular communication pathways) that were also identified in our previous study with Duroc. Taken together, these observations reaffirm the notion that while the biological mechanisms underlying a particular phenotype do not differ, the genetic regulation of these mechanisms can differ between breeds. This is particularly important in the context of developing marker-assisted and genomic selection strategies, as it demonstrates that improving the same trait may require different sets of markers for different lines/breeds of livestock.

CANDIDATE GENES FOR RESIDUAL FEED INTAKE

Significant associations between both RFI and SNPs on SSC 7 (MARC0027992) and 8 (ALGA0039145) were found in the present study. Here, *ENSSSCG00000018237* encoding U1 spliceosomal RNA was found near a SNP significantly associated with both RFIs on SSC 7. U1 spliceosomal RNA constitutes U1 small nuclear ribonucleoprotein that plays a role in splicing of pre-mRNAs (Zwieb, 1996). Another independent study (Onteru et al., 2013) reported a QTL region between 16 and 17 Mb on SSC 7 in Yorkshire pigs for RFI. In this study, a SNP significantly associated with RFI1 was found within an intron 4–5 of zinc finger, FYVE domain containing the 26 (*ZFYVE26*) gene that is also located on SSC 7. The gene encodes a protein containing a FYVE zinc finger binding domain which helps target the protein to membrane lipids (Laity et al., 2001). While *ZFYVE26* has been associated with autosomal recessive spastic paraplegia in humans (Herd et al.,

[5] http://david.abcc.ncifcrf.gov/
[6] http://omicslab.genetics.ac.cn/GOEAST/

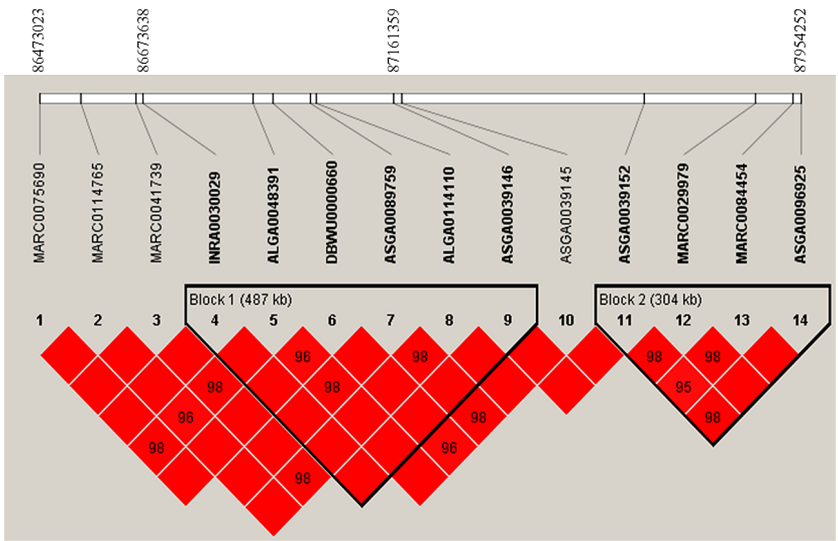


FIGURE 3 | Linkage disequilibrium (LD) pattern on a region from 86 to 88 Mb on pig chromosome 8. LD blocks are marked with triangles. Values in boxes are LD (r^2) between SNP pairs and the boxes are red indicated LOD > 2 and $D' = 1$ (LOD is the log of the likelihood odds ratio, a measure of confidence in the value of D').

2004), the precise biological function of this gene has not yet been described.

Another important association was between SNP ASGA0039145 on SSC 8 and both RFI traits. This SNP is located in a genomic region where QTLs influencing FCR have been reported in an independent study based on a Duroc resource population (Sahana et al., 2013). The *ENSSSCG00000009034* is a gene closest to this SNP; however, the gene has not been functionally characterized yet. Moreover, this SNP is tightly linked with five other SNPs to form the LD block 1 (**Figure 3**). This LD block spans 487 kb region and consists three significant SNPs for RFI1. This LD block also covers a *TTC29* gene which encodes a testis development protein that could also be an interesting candidate for further investigation. Also known as *NYD-SP14*, *TTC29* is a component of axonemal dyneins (Yamamoto et al., 2008) that have recently been demonstrated to play an important role in fat metabolism (Sohle et al., 2012).

We also found many SNPs to be associated with either RFI1 or RFI2 in our analyses. On SSC3, the SNP H3GA0010038 associated with only RFI1 located closest to a novel gene. On SSC 9, another transcriptional factor *PROX1* was found proximal to an SNP exclusively associated with RFI1. *PROX1* likely plays a fundamental role in the early development of the central nervous system (Kaltezioti et al., 2010). It also is a key regulator of lymphatic endothelial cell fate specification (Ma, 2007), ERRα mediated control of the molecular clock (Choquette et al., 2012), and modulation of insulin sensitivity and glucose handling (Fontanesi et al., 2012) that could influence energy metabolism and RFI. On SSC17, MARC0067053 is significantly associated with RFI1 ($p = 1.08e{-}04$), and is located in 5UTR′ region of *SOGA1* gene. This gene encodes a SOGA1 protein that plays a role in reducing glucose production (Cowerd et al., 2010). It also contributes to adiponectin-mediated insulin-dependent inhibition of autophagy (Forbes, 2010). Since autophagy provides biochemical intermediates for glucose production (Forbes, 2010) which influences feed consumption, SOGA1 could be an interesting candidate gene for RFI1.

Moreover, on SSC 1 and SSC 14, close to significant SNPs, we reported two members of G-protein couple receptor (*MAS1* and *GRK5*) as possible candidate genes for RFI2. For instance, the GRK5 regulates the GPCR signaling pathway and GRK5 deficiency led to insulin resistance and hepatic steatosis, or decreases diet-induced obesity and adipogenesis in mice (Wang et al., 2012b). On chromosome 10, a functionally uncharacterized gene *ENSSSCG00000011017* encoding a lysozyme-like ortholog, was located near ALGA0117721 that was significantly associated with RFI2. Other candidate genes in proximity to SNP associated exclusively with RFI1 or RFI2 were *RPS18*, *GPR155*, and *XIRP2*. Dysregulation of GPCR 155 (*GPR155*) is associated with higher feed efficiency in chicken while *RPS18*, encoding the 40S ribosomal protein S18, is involved in regulation of development (Laity et al., 2001). However, very little is known about the biological function of these genes and their relevance in the context of RFI is not apparent.

Notably, melanocortin 4 receptor (*MC4R*; SSC 1: 178,553,488-178,555,219) is perhaps most well-known candidate gene for feed efficiency or/and feed intake in pigs (Kim et al., 2000; Houston et al., 2004; Burgos et al., 2006; Davoli et al., 2012; Onteru et al., 2013). The MC4R gene codes for a G protein transmembrane receptor playing an important role in energy homeostasis control. In pigs, a SNP (missense substitution 298 Asp > Asn) in *MC4R* gene has been identified and associated to average daily gain, feed intake and fatness traits in many difference studies (Kim et al., 2000; Houston et al., 2004; Fan et al., 2009; Davoli et al., 2012). Moreover, Leptin (SSC 18: 21, 201, 786-21, 204, 671) plays a key

role in regulating energy intake and expenditure and is a candidate gene for feed efficiency in pigs (Barb et al., 2001). However, variants in these genes are not included the PorcineSNP60 Illumina iSelect BeadChip and it is unclear whether such variants could influence RFI in the current population.

STATISTICAL METHODS USED IN GWAS

One of the challenges for doing GWAS in livestock population is the large proportion of animals have phenotypic but no genotypic records, especially in dairy cattle. Recently, Wang et al. (2012a) proposed a single step GBLUP (or ssGBLUP) method that incorporates all genotypes, observed phenotypes and pedigree information jointly in one step and provides GEBVs for all animals with or without genomic data or phenotypic data or both (based on the methods of Aguilar et al., 2010; Christensen and Lund, 2010). The ssGWAS is a method based on GBLUP (Wang et al., 2014) which derives the SNP effects (or SNP variance) from GEBVs calculated from ssGBLUP. However, the use of the ssGWAS method is limited by finding appropriate number of iterations required to get marker solutions and most importantly, its inability to determine the genome-wide significance level for each SNP in the entire genomic data. However, our approach in this study of combining GWAS with pathways analysis requires a genome-wide adjusted *p*-value for each SNP for selecting the top SNPs for further downstream gene enrichment and pathway analyses using bioinformatics tools. Hence the genome-wide *p*-values for each SNP are not possible in ssGWAS method, we have adapted a mixed model GWAS and implemented using DMU package (Madsen et al., 2006). We used deregressed EBVs as a pseudo-phenotype but ssGBLUP would have also been a better choice in that it handles variability far better than the use of deregressed EBV as Wang et al. (2012a) reported differences in accuracy of genomic breeding values compared to other methods including classical GWAS.

There is still some controversy as to how to properly determine SEs of estimated SNP effects by GBLUP based-methods. Recently, Gualdron Duarte et al. (2014) provided a way to determine significance values for each SNP marker effect by linear transformations of genomic evaluations. Briefly, the likelihood ratio is calculated to test the significance of the largest effect segment of each chromosome by comparing against a reduced model with fixed effects and GEBVs. The critical value (size of the test) is adjusted by the Bonferroni correction. Moreover, we also would like to note that our GWA model could be extended to the case where the additive genetic relationship is substituted by the genomic relationship matrix like in EMMAX[7].

PATHWAYS INVOLVED IN RESIDUAL FEED INTAKE

Results from gene-set enrichment analyses are largely dependent on how gene-sets are identified or defined. In the current study, our gene-set was determined by the significance threshold that was used to declare SNPs significantly associated with RFI. Consequently, our enrichment analyses was very dependent on our choice of significance threshold. The choice of significance threshold also influences the degree of confidence that can be ascribed to results from gene-set enrichment analyses. Choosing a stringent threshold like Bonferroni will likely yield very few results with higher confidence as opposed to a lenient threshold that will likely yield more results with lower confidence. In our analyses, we decided to use a FDR based *q*-value threshold of 0.2 to balance the number of results and the degree of confidence associated with them. Applying more stringent FDR thresholds (for e.g., of 0.05 or 0.10) significantly reduced the numbers of SNP, and consequently the number of genes in the gene-set, for pathway annotation. Therefore, by setting the at a *q*-value $\leq$0.2 ($p \sim 0.001$; meaning 20% of SNPs using for pathway analyses are likely to be false), we had a reasonable number of SNPs for gene-set enrichment analyses.

Regardless to different thresholds, the metabolic pathway[8] was significantly associated with both RFI traits. Many previous studies have shown that variation in mediators of metabolic processes contribute to the variation of RFI (e.g., reviewed in Herd et al., 2004; Herd and Arthur, 2009; Hoque and Suzuki, 2009; Dekkers and Gilbert, 2010). However, the metabolic pathway is a broad overarching term that contains many specific modules (e.g., energy, carbohydrate and lipid, nucleotide and amino acid and secondary metabolism). So future investigations evaluating the contribution of specific sub-modules within this pathway to the genetic variation in RFI might be warranted. An interesting pathway associated with RFI2 (and with RFI1 when analysis is performed based on a nominal threshold of a $p \leq 0.05$) was related to olfactory transduction. Olfactory transduction pathways are responsible for the perception of odor via olfactory receptors and downstream biochemical signaling events that ultimately get transformed into electrical impulses sent to the brain (Ma, 2007). Pigs have the largest repertoire of functional olfactory receptors (Groenen et al., 2012) encoded by at least 1113 genes (Nguyen et al., 2012), 14 and 11 of which are located near SNPs significantly associated with RFI1 and RFI2, respectively, (Table S1 in Supplementary Materials). Olfaction is one of the major sensory modalities that contribute to hedonic evaluation of a food, resulting in food choice and its possible consumption. It is modulated in response to changing levels of various molecules, such as ghrelin, orexins, neuropeptide Y, insulin, leptin, and cholecystokinin (Palouzier-Paulignan et al., 2012). These molecules are known to play an important role in controlling of RFI. For example, genetic selection for low and high RFI in pigs has been shown to change leptin concentration in plasma (Lefaucheur et al., 2011). Lower RFI has also been shown to be associated with lower serum leptin concentrations in Duroc pigs (Hoque et al., 2009).

Another interesting pathway significantly associated with both RFI traits (when using a nominal threshold of a $p < 0.05$) was the insulin secretion pathway (Table S1 in Supplementary Materials). Do et al. (2014) also reported Insulin signaling pathway associated with RFI in Duroc pigs. Insulin-dependent regulation of feed intake has been described in many species including cattle (Chen et al., 2012; Rolf et al., 2012) and pigs (Colditz, 2004; Cruzen et al., 2012). The results imply that insulin secretion is possibly an intermediary pathway by which olfactory transduction influences RFI.

[7] http://genetics.cs.ucla.edu/emmax/

[8] http://www.genome.jp/kegg-bin/show_pathway?org_name=ssc&mapno=01100

Richardson and Herd (2004) indicated that differences in some plasma metabolites and hormones have been positively related to genetic and phenotypic measures of RFI in ruminants.

The genetic variants identified by GWA studies may facilitate the incorporation of marker-assisted selection in commercial breeding schemes for improvement of complex traits. Moreover, Snelling et al. (2013) and Kadarmideen (2014) have suggested that genomic selection could perform better if it is guided by network and pathway analysis. Biological pathways identified by post-GWA analyses could further our current understanding of the genetic underlying different complex traits. Therefore, our results would be of interest not only to breeders interested in using marker-assisted selection to improve feed efficiency in pigs, but also to biologists interested in better understanding the biological mechanisms influencing feed efficiency. However, it is also important to consider potential limitations of our study, such as the limited size of Yorkshire resource population, statistical model used in the estimation of RFI, and statistical models used in GWA; gene set enrichment and pathway analyses. Finally, it is also important to note that all results reported in this study are only relevant to the specific definition used in this study.

In summary, the present study describes SNPs, candidate genes and biological pathways putatively influencing RFI in Yorkshire pigs. Important candidate genes such as *XIRP2*, *TTC29*, *SOGA1*, *MAS1*, *GRK5*, *PROX1*, *GPR155*, and *ZFYVE26* were identified here that could be further investigated for harboring causal variants. Pathway analyses identified metabolic and olfactory transduction pathways to be associated with RFI. Many other pathways (such as insulin secretion, tight junction) that were found to be associated with RFI based on a lenient nominal significance threshold might be of some interest. However, more studies are required to determine their role in regulating RFI.

AUTHOR CONTRIBUTIONS

Duy N. Do did the analysis with the help of Haja N. Kadarmideen, Anders B. Strathe, Tage Ostersen, and Sameer D. Pant. Duy N. Do wrote the first draft of the manuscript. Haja N. Kadarmideen conceived and designed this project and provided supervision for Duy N. Do in all aspects of this research, including GWAS and biological interpretations. All authors contributed to writing of this manuscript, read and approved the final manuscript.

ACKNOWLEDGMENTS

We would like to thank the Department of Breeding and Genetics of the Danish Pig Research Centre for providing all data for the research reported in this study. Duy N. Do is a PhD student funded by the Department of Veterinary Clinical and Animal Sciences, Faculty of Health and Medical Sciences, University of Copenhagen. Haja N. Kadarmideen thanks EU-FP7 Marie Curie Actions – Career Integration Grant (CIG-293511) for partially funding his time spent on this research.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://www.frontiersin.org/journal/10.3389/fgene.2014.00307/abstract

REFERENCES

Aguilar, I., Misztal, I., Johnson, D., Legarra, A., Tsuruta, S., Lawlor, T. J., et al. (2010). Hot topic: a unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *J. Dairy Sci.* 93, 743–752. doi: 10.3168/jds.2009-2730

Barb, C., Hausman, G., and Houseknecht, K. (2001). Biology of leptin in the pig. *Domest. Anim. Endocrinol.* 21, 297–317. doi: 10.1016/S0739-7240(01)00123-0

Barrett, J., Fry, B., Maller, J., and Daly, M. (2005). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21, 263–265. doi: 10.1093/bioinformatics/bth457

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B (Methodological)* 57, 289–300. doi: 10.2307/2346101

Burgos, C., Carrodeguas, J. A., Moreno, C., Altarriba, J., Tarrafeta, L., Barcelona, J. A., et al. (2006). Allelic incidence in several pig breeds of a missense variant of pig melanocortin-4 receptor (MC4R) gene associated with carcass and productive traits: its relation to IGF2 genotype. *Meat Sci.* 73, 144–150. doi: 10.1016/j.meatsci.2005.11.007

Chen, Y., Arthur, P., Barchia, I., Quinn, K., Parnell, P., and Herd, R. (2012). Using gene expression information obtained by quantitative real-time PCR to evaluate Angus bulls divergently selected for feed efficiency. *Anim. Prod. Sci.* doi: 10.1215/9780822395447

Choquette, A. C., Bouchard, L., Drapeau, V., Lemieux, S., Tremblay, A., Bouchard, C., et al. (2012). Association between olfactory receptor genes, eating behavior traits and adiposity: results from the Quebec Family Study. *Physiol. Behav.* 105, 772–776. doi: 10.1016/j.physbeh.2011.10.015

Christensen, O. F., and Lund, M. S. (2010). Genomic prediction when some animals are not genotyped. *Genet. Sel. Evol.* 42, 1–8. doi: 10.1186/1297-9686-42-2

Colditz, I. (2004). Some mechanisms regulating nutrient utilisation in livestock during immune activation: an overview. *Anim. Prod. Sci.* 44, 453–457. doi: 10.1071/EA02066

Cowerd, R. B., Asmar, M. M., Alderman, J. M., Alderman, E. A., Garland, A. L., Busby, W. H., et al. (2010). Adiponectin lowers glucose production by increasing SOGA. *Am. J. Pathol.* 177, 1936–1945. doi: 10.2353/ajpath.2010.100363

Cruzen, S. M., Harris, A. J., Hollinger, K., Selsby, J. T., Gabler, N. K., Lonergan, S. M., et al. (2012). "Gilts selected for low residual feed intake have potential for decreased protein degradation," in *58th International Congress of Meat Science and Technology*, Montréal.

Davoli, R., Braglia, S., Valastro, V., Annarratone, C., Comella, M., Zambonelli, P., et al. (2012). Analysis of MC4R polymorphism in Italian Large White and Italian Duroc pigs: association with carcass traits. *Meat Sci.* 90, 887–892. doi: 10.1016/j.meatsci.2011.11.025

Dekkers, J., and Gilbert, H. (2010). "Genetic and biological aspect of residual feed intake in pigs," in *Proceedings of nineth World Congress on Genetics Applied to Livestock Production*, Leipzig.

Do, D. N., Ostersen, T., Strathe, A. B., Mark, T., Jensen, J., and Kadarmideen, H. N. (2014). Genome-wide association and systems genetic analyses of residual feed intake, daily feed consumption, backfat and weight gain in pigs. *BMC Genet.* 15:27. doi: 10.1186/1471-2156-15-27

Do, D. N., Strathe, A. B., Jensen, J., Mark, T., and Kadarmideen, H. N. (2013a). Genetic parameters for different measures of feed efficiency and related traits in boars of three pig breeds. *J. Anim. Sci.* 91, 4069–4079. doi: 10.2527/jas.2012-6197

Do, D. N., Strathe, A. B., Ostersen, T., Jensen, J., Mark, T., and Kadarmideen, H. N. (2013b). Genome wide association study reveal genetic architecture of eating behaviors in pigs and its implications for humans obesity by comparative genome mapping. *PLoS ONE* 8:e71509. doi: 10.1371/journal.pone.0071509

Duggal, P., Gillanders, E., Holmes, T., and Bailey-Wilson, J. (2008). Establishing an adjusted p-value threshold to control the family-wide type 1 error in genome wide association studies. *BMC Genomics* 9:516. doi: 10.1186/1471-2164-9-516

Fan, B., Lkhagvadorj, S., Cai, W., Young, J., Smith, R. M., Dekkers, J. C. M., et al. (2010). Identification of genetic markers associated with residual feed intake and meat quality traits in the pig. *Meat Sci.* 84, 645–650. doi: 10.1016/j.meatsci.2009.10.025

Fan, B., Onteru, S. K., Plastow, G. S., and Rothschild, M. F. (2009). Detailed characterization of the porcine MC4R gene in relation to fatness and growth. *Anim. Genet.* 40, 401–409. doi: 10.1111/j.1365-2052.2009.01853.x

Farber, C. R. (2013). Systems-level analysis of genome-wide association data. *G3 (Bethesda)* 3, 119–129. doi: 10.1534/g3.112.004788

Fontanesi, L., Schiavo, G., Galimberti, G., Calo, D., Scotti, E., Martelli, P., et al. (2012). A genome wide association study for backfat thickness in Italian Large White pigs highlights new regions affecting fat deposition including neuronal genes. *BMC Genomics* 13:583. doi: 10.1186/1471-2164-13-583

Forbes, J. M. (2010). The physiological deadlock between AMPK and gluconeogenesis: SOGA, a novel protein, may provide the key. *Am. J. Pathol.* 177, 1600–1602. doi: 10.2353/ajpath.2010.100679

Gabriel, S., Schaffner, S., Nguyen, H., Moore, J., Roy, J., Blumenstiel, B., et al. (2002). The structure of haplotype blocks in the human genome. *Science* 296, 2225–2229. doi: 10.1126/science.1069424

Gao, X., Starmer, J., and Martin, E. R. (2008). A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms. *Genet. Epidemiol.* 32, 361–369. doi: 10.1002/gepi.20310

Garrick, D. J., Taylor, J. F., and Fernando, R. L. (2009). Deregressing estimated breeding values and weighting information for genomic regression analyses. *Genet. Sel. Evol.* 41:55. doi: 10.(1186)/1297-9686-41-55

Gilbert, H., Riquet, J., Gruand, J., Billon, Y., Fève, K., Sellier, P., et al. (2010). Detecting QTL for feed intake traits and other performance traits in growing pigs in a Piétrain–Large White backcross. *Animal* 4, 1308–1318. doi: 10.1017/S1751731110000339

Gregersen, V., Conley, L., Sorensen, K., Guldbrandtsen, B., Velander, I., and Bendixen, C. (2012). Genome-wide association scan and phased haplotype construction for quantitative trait loci affecting boar taint in three pig breeds. *BMC Genomics* 13:22. doi: 10.1186/1471-2164-13-22

Groenen, M. A. M., Archibald, A. L., Uenishi, H., Tuggle, C. K., Takeuchi, Y., Rothschild, M. F., et al. (2012). Analyses of pig genomes provide insight into porcine demography and evolution. *Nature* 491, 393–398. http://www.nature.com/nature/journal/v491/n7424/abs/nature11622.html#supplementary-information

Gualdron Duarte, J., Cantet, R., Bates, R., Ernst, C., Raney, N., and Steibel, J. (2014). Rapid screening for phenotype-genotype associations by linear transformations of genomic evaluations. *BMC Bioinformatics* 15:246. doi: 10.1186/1471-2105-15-246

Gunsett, F. C. (1984). Linear index selection to improve traits defined as ratios. *J. Anim. Sci.* 59, 1185–1193.

Henryon, M., Berg, P., Jensen, J., and Andersen, S. (2001). Genetic variation for resistance to clinical and subclinical diseases exists in growing pigs. *Anim. Sci. (Glasgow)* 73, 375–388.

Herd, R., and Arthur, P. (2009). Physiological basis for residual feed intake. *J. Anim. Sci.* 87, E64–E71. doi: 10.2527/jas.2012-5687

Herd, R., Oddy, V., and Richardson, E. (2004). Biological basis for variation in residual feed intake in beef cattle. 1. Review of potential mechanisms. *Anim. Prod. Sci.* 44, 423–430. doi: 10.1071/EA02220

Hoque, M., and Suzuki, K. (2009). Genetics of residual feed intake in cattle and pigs: a review. *Asian-Aust. J. Anim. Sci.* 22, 747–755. doi: 10.5713/ajas.2009.80467

Hoque, M. A., Katoh, K., and Suzuki, K. (2009). Genetic associations of residual feed intake with serum insulin-like growth factor-I and leptin concentrations, meat quality, and carcass cross sectional fat area ratios in Duroc pigs. *J. Anim. Sci.* 87, 3069–3075. doi: 10.2527/jas.2008-1268

Houston, R. D., Cameron, N. D., and Rance, K. A. (2004). A melanocortin-4 receptor (MC4R) polymorphism is associated with performance traits in divergently selected large white pig populations. *Anim. Genet.* 35, 386-390. doi: 10.1111/j.1365-2052.2004.01182.x

Kadarmideen, H. (2008). Genetical systems biology in livestock: application to gonadotrophin releasing hormone and reproduction. *IET Syst. Biol.* 2, 423–441. doi: 10.1049/iet-syb:20070072

Kadarmideen, H. N. (2014). Genomics to Systems Biology in Animal and Veterinary Sciences: progress, lessons and opportunities. *Livest. Sci.* 166, 232–248. doi: 10.1016/j.livsci.2014.04.028

Kadarmideen, H. N., Von Rohr, P., and Janss, L. L. (2006). From genetical genomics to systems genetics: potential applications in quantitative genomics and animal breeding. *Mamm. Genome* 17, 548–564. doi: 10.1007/s00335-005-0169-x

Kaltezioti, V., Kouroupi, G., Oikonomaki, M., Mantouvalou, E., Stergiopoulos, A., Charonis, A., et al. (2010). Prox1 regulates the notch1-mediated inhibition of neurogenesis. *PLoS Biol.* 8:e1000565. doi: 10.1371/journal.pbio.1000565

Kim, K. S., Larsen, N., Short, T., Plastow, G., and Rothschild, M. F. (2000). A missense variant of the porcine melanocortin-4 receptor (MC4R) gene is associated with fatness, growth, and feed intake traits. *Mamm. Genome* 11, 131–135. doi: 10.1007/s003350010025

Laity, J. H., Lee, B. M., and Wright, P. E. (2001). Zinc finger proteins: new insights into structural and functional diversity. *Curr. Opin. Struct. Biol.* 11, 39–46. doi: 10.1016/S0959-440X(00)00167-6

Lefaucheur, L., Lebret, B., Ecolan, P., Louveau, I., Damon, M., Prunier, A., et al. (2011). Muscle characteristics and meat quality traits are affected by divergent selection on residual feed intake in pigs. *J. Anim. Sci.* 89, 996–1010. doi: 10.2527/jas.2010-3493

Ma, M. (2007). Encoding olfactory signals via multiple chemosensory systems. *Crit. Rev. Biochem. Mol. Biol.* 42, 463–480. doi: 10.1080/10409230701693359

Madsen, P., Sørensen, P., Su, G., Damgaard, L. H., Thomsen, H., and Labouriau, R. (2006). "DMU – a package for analyzing multivariate mixed models," in *Proceedings of eighth World Congress on Genetics Applied to Livestock Production*, Belo Horizonte.

Mayr, E. (1963). *Animal Species and Evolution. Animal Species and Their Evolution.* Cambridge, MA: Harvard University Press. doi: 10.4159/harvard.9780674865327

Nguyen, D., Lee, K., Choi, H., Choi, M.-K., Le, M., Song, N., et al. (2012). The complete swine olfactory subgenome: expansion of the olfactory gene repertoire in the pig genome. *BMC Genomics* 13:584. doi: 10.1186/1471-2164-13-584

Onteru, S. K., Gorbach, D. M., Young, J. M., Garrick, D. J., Dekkers, J. C. M., and Rothschild, M. F. (2013). Whole genome association studies of residual feed intake and related traits in the pig. *PLoS ONE* 8:e61756. doi: 10.1371/journal.pone.0061756

Ostersen, T., Christensen, O. F., Henryon, M., Nielsen, B., Su, G. S., and Madsen, P. (2011). Deregressed EBV as the response variable yield more reliable genomic predictions than traditional EBV in pure-bred pigs. *Genet. Sel. Evol.* 43:38. doi: 10.1186/1297-9686-43-38

Palouzier-Paulignan, B., Lacroix, M.-C., Aimé, P., Baly, C., Caillol, M., Congar, P., et al. (2012). Olfaction under metabolic influences. *Chem. Senses* 37, 769–797. doi: 10.1093/chemse/bjs059

Peñagaricano, F., Weigel, K. A., Rosa, G. J. M., and Khatib, H. (2013). Inferring quantitative trait pathways associated with bull fertility from a genome-wide association study. *Front. Genet.* 3:307. doi: 10.3389/fgene.2012.00307

R Development Core Team. (2008). *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing.

Raff, R. A. (1996). *The Shape of Life: Genes, Development, and the Evolution of Animal Form.* Chicago, IL: University of Chicago Press.

Richardson, E. C., and Herd, R. M. (2004). Biological basis for variation in residual feed intake in beef cattle. 2. Synthesis of results following divergent selection. *Aust. J. Exp. Agric.* 44, 431–440. doi: 10.1071/EA02221

Rolf, M., Taylor, J., Schnabel, R., Mckay, S., Mcclure, M., Northcutt, S., et al. (2012). Genome-wide association analysis for feed efficiency in Angus cattle. *Anim. Genet.* 43, 367–374. doi: 10.1111/j.1365-2052.2011.02273.x

Sahana, G., Kadlecová, V., Hornshøj, H., Nielsen, B., and Christensen, O. F. (2013). A genome-wide association scan in pig identifies novel regions associated with feed efficiency trait. *J. Anim. Sci.* 91, 1041–1050. doi: 10.2527/jas.2012-5643

Saintilan, R., Mérour, I., Brossard, L., Tribout, T., Dourmad, J. Y., Sellier, P., et al. (2013). Genetics of residual feed intake in growing pigs: relationships with production traits, and nitrogen and phosphorus excretion traits. *J. Anim. Sci.* 91, 2542–2554. doi: 10.2527/jas.2012-5687

Sanchez, M.-P., Tribout, T., Iannuccelli, N., Bouffaud, M., Servin, B., Tenghe, A., et al. (2014). A genome-wide association study of production traits in a commercial population of Large White pigs: evidence of haplotypes affecting meat quality. *Genet. Sel. Evol.* 46:12. doi: 10.1186/1297-9686-46-12

Shirali, M., Duthie, C.-A., Doeschl-Wilson, A., Knap, P., Kanis, E., Van Arendonk, J., et al. (2013). Novel insight into the genomic architecture of feed and nitrogen efficiency measured by residual energy intake and nitrogen excretion in growing pigs. *BMC Genet.* 14:121. doi: 10.1186/1471-2156-14-121

Snelling, W., Cushman, R., Keele, J., Maltecca, C., Thomas, M., Fortes, M., et al. (2013). Breeding and Genetics Symposium: networks and pathways to guide genomic selection. *J. Anim. Sci.* 91, 537–552. doi: 10.2527/jas.2012-5784

Sohle, J., Machuy, N., Smailbegovic, E., Holtzmann, U., Gronniger, E., Wenck, H., et al. (2012). Identification of new genes involved in human adipogenesis and fat storage. *PLoS ONE* 7:e31193. doi: 10.1371/journal.pone.0031193

Torkamani, A., Topol, E. J., and Schork, N. J. (2008). Pathway analysis of seven common diseases assessed by genome-wide association. *Genomics* 92, 265–272. doi: 10.1016/j.ygeno.2008.07.011

Wang, H., Misztal, I., Aguilar, I., Legarra, A., Fernando, R. L., Vitezica, Z., et al. (2014). Genome-wide association mapping including phenotypes from relatives without genotypes in a single-step (ssGWAS) for 6-week body weight in broiler chickens. *Front. Genet.* 5:134. doi: 10.3389/fgene.2014.00134

Wang, H., Misztal, I., Aguilar, I., Legarra, A., and Muir, W. M. (2012a). Genome-wide association mapping including phenotypes from relatives without genotypes. *Genet. Res.* 94, 73–83. doi: 10.1017/S0016672312000274

Wang, L., Shen, M., Wang, F., and Ma, L. (2012b). GRK5 ablation contributes to insulin resistance. *Biochem. Biophys. Res. Commun.* 429, 99–104. doi: 10.1016/j.bbrc.2012.10.077

Wang, K., Li, M., and Hakonarson, H. (2010). Analysing biological pathways in genome-wide association studies. *Nat. Rev. Genet.* 11, 843–854. doi: 10.1038/nrg2884

Wang, L., Sorensen, P., Janss, L., Ostersen, T., and Edwards, D. (2013). Genome-wide and local pattern of linkage disequilibrium and persistence of phase for 3 Danish pig breeds. *BMC Genet.* 14:115. doi: 10.1186/1471-2156-14-115

Yamamoto, R., Yanagisawa, H. A., Yagi, T., and Kamiya, R. (2008). Novel 44-kilodalton subunit of axonemal Dynein conserved from chlamydomonas to mammals. *Eukaryot. Cell* 7, 154–161. doi: 10.1128/EC.00341-07

Zwieb, C. (1996). The uRNA database. *Nucleic Acids Res.* 24, 76–79. doi: 10.1093/nar/24.1.76

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 06 May 2014; accepted: 18 August 2014; published online: 09 September 2014.

Citation: Do DN, Strathe AB, Ostersen T, Pant SD and Kadarmideen HN (2014) Genome-wide association and pathway analysis of feed efficiency in pigs reveal candidate genes and pathways for residual feed intake. Front. Genet. ***5****:307. doi: 10.3389/fgene.2014.00307*

This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics.

Copyright © 2014 Do, Strathe, Ostersen, Pant and Kadarmideen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in GENETICS

ORIGINAL RESEARCH ARTICLE
published: 01 October 2014
doi: 10.3389/fgene.2014.00332

Changes in variance explained by top SNP windows over generations for three traits in broiler chicken

Breno de Oliveira Fragomeni[1]*, Ignacy Misztal[1], Daniela Lino Lourenco[1], Ignacio Aguilar[2], Ronald Okimoto[3] and William M. Muir[4]

[1] Department of Animal and Dairy Science, University of Georgia, Athens, GA, USA
[2] Instituto Nacional de Investigación Agropecuaria, Las Brujas, Uruguay
[3] Cobb Vantress Inc., Siloam Springs, AR, USA
[4] Department of Animal Sciences, Purdue University, West Lafayette, IN, USA

Edited by:
Johann Sölkner, BOKU-University of Natural Resources and Life Sciences Vienna, Austria

Reviewed by:
Dirk-Jan De Koning, Swedish University of Agricultural Sciences, Sweden
Jiuzhou Song, University of Maryland, USA
Bruno Valente, University of Wisconsin – Madison, USA

***Correspondence:**
Breno de Oliveira Fragomeni, Department of Animal and Dairy Science, University of Georgia, 425 River Road, Athens, GA 30602, USA
e-mail: fragomen@uga.edu

The purpose of this study was to determine if the set of genomic regions inferred as accounting for the majority of genetic variation in quantitative traits remain stable over multiple generations of selection. The data set contained phenotypes for five generations of broiler chicken for body weight, breast meat, and leg score. The population consisted of 294,632 animals over five generations and also included genotypes of 41,036 single nucleotide polymorphism (SNP) for 4,866 animals, after quality control. The SNP effects were calculated by a GWAS type analysis using single step genomic BLUP approach for generations 1–3, 2–4, 3–5, and 1–5. Variances were calculated for windows of 20 SNP. The top ten windows for each trait that explained the largest fraction of the genetic variance across generations were examined. Across generations, the top 10 windows explained more than 0.5% but less than 1% of the total variance. Also, the pattern of the windows was not consistent across generations. The windows that explained the greatest variance changed greatly among the combinations of generations, with a few exceptions. In many cases, a window identified as top for one combination, explained less than 0.1% for the other combinations. We conclude that identification of top SNP windows for a population may have little predictive power for genetic selection in the following generations for the traits here evaluated.

Keywords: genomic selection, genome-wide association study, QTL, ssGBLUP, gene identification

INTRODUCTION

Past studies of genomics in livestock usually focused either on best estimation of breeding values (Calus, 2010) or on identification of major single nucleotide polymorphism (SNP) (Goddard and Hayes, 2009). For the latter, the purpose is exploring associations between SNP and phenotypes to better understand the genetic architecture of a trait or to use identified major SNP for genetic selection. With important SNP identified, the selection can be performed with simple tests for a few SNP.

Genetic selection using major SNP is successful if they explain a sizeable portion of the genetic variation and if their effects change little over time. Earlier simulation studies showed that linkage disequilibrium (LD) identified in one generation decays very slowly over generations (Meuwissen et al., 2001; Solberg et al., 2009). However, under strong selection the decay is much faster (Muir, 2007). Therefore, newer studies advocate continuous genotyping and recalculation of SNP effects (Habier et al., 2007; Sonesson and Meuwissen, 2009; Wolc et al., 2011). While the selection pressure would act on the largest quantitative trait loci (QTLs), it is not clear how this would impact the identification and estimation of values for the top SNP that may indicate presence of QTLs.

Identification of an individual SNP linked to a QTL is difficult because of the high collinearity of SNPs. SNPs may be in LD with a QTL so windows of consecutive SNPs can capture the effect of a QTL better than a single SNP (Habier et al., 2011). Also, SNP segments are useful to discriminate important effects from statistical noise (Sun et al., 2011). Bolormaa et al. (2010) looked at SNPs within 1 Mbp intervals. Peters et al. (2012) used windows of five adjacent SNP. In a simulation study, effects of individual QTL were best explained by the combined effect of eight adjacent SNP (Wang et al., 2012). The optimal window size may also be a function of effective population size (Goddard, 2009).

There is a shortage in studies searching for stability of marker effects across generations in production traits for broiler chicken. Despite this, in a layer population, Wolc et al. (2012) found that 1 Mbp SNP windows with large effects had consistent effects across generations, but windows that explained little variance of the trait were not validated. If a window effect is constant across generations or subsets of population, it can be indicative of a causative gene on that trait; however, if the effect is not robust, it can correspond to an unstable, sample-specific association that is not expected to provide good out-of-sample predictions.

One common issue on genome association studies is the large number of false positive gene discovery. Information from the chicken QTL database (Hu et al., 2013) shows a large number of QTL described—2,467 for growth traits, 68 for meat quality traits, and 28 for conformation—but few of these have been validated or reproduced by other studies. This can be observed not only in chicken, but in studies on all livestock species. In this way, GWAS results should be carefully interpreted before considering

an association as a causative effect. A possible causative effect should be easily accessed in further assays considering similar population structure.

The purpose of this study was to identify SNP windows that explain major portions of genetic variance and see if those values are preserved during a course of selection for growth in chicken.

MATERIALS AND METHODS

The data was provided by Cobb-Vantress Inc. (Siloam Springs, AR, USA). A total of 294,632 phenotypes from a pure line of broiler chicken collected across five consecutive generations (G1, G2, G3, G4, and G5) were used in this study. This was the sire line, selected mainly for growth rate, meat yield, feed conversion and livability, and secondarily for reproduction traits. The numerator relationship matrix included 297,017 animals. For the first two generations, animals were selected for genotyping based on body weight and conformation scores; leg defects were very unlikely. The remaining animals (from G3 to G5) were randomly selected for genotyping. The number of animals in each generation are shown in **Table 1**. The number of observations, means, and SD for all the traits are shown in **Table 2**.

Initially, genotype information from 4,922 animals in a chip with 57,635 SNPs was available (Groenen et al., 2011). The genomic data was subject to a quality control (QC) before the analysis. This QC removed SNPs with minor allele frequency <0.05, with call rates <0.9, and monomorphic SNPs. It also removed genotypes with call rates <0.9. After QC, the genotype file had 4,866 animals genotyped for 41,036 SNPs.

SNP solutions were estimated by ssGWAS (genome-wide association study using a single-step BLUP approach; Wang et al., 2012; Dikmen et al., 2013). In this methodology, the data was initially analyzed by a multi-trait single-step genomic BLUP (ssGBLUP; Misztal et al., 2009; Aguilar et al., 2010) with the same model as used for BLUP analyses (Chen et al., 2011). Effects in the model included sex, contemporary group, animal additive, and maternal permanent environmental effects. Concerning the genomic information, the genomic relationship matrix (**G**) was scaled for the average of the numerator relationship matrix for the genotyped animals ($\mathbf{A}_{22}$), which took into account the effect of non-random genotyping caused by selection (Vitezica et al., 2011). Subsequently, EBV for genotyped animals (GEBV) were converted to SNP effects and weights of SNP effect were refined iteratively. The procedure followed the S1 scenario described in Wang et al. (2012), with GEBV computed once and SNP weights refined through three iterations. The equation for predicting SNP effects using weighted genomic relationship matrix was (Wang et al., 2012):

$$\hat{\mathbf{u}} = \mathbf{DZ'}[\mathbf{ZDZ'}]^{-1}\hat{\mathbf{a}}_g$$

In which: $\hat{\mathbf{u}}$ is the vector with estimated SNP marker effects, **D** is a diagonal matrix of weights for variances of SNP effects, **Z** is a matrix relating genotypes of each locus to each individual, and $\hat{\mathbf{a}}_g$ is the additive genetic effect for genotyped animals.

The individual variance of SNP effect (the same as in **D**) was estimated as (Zhang et al., 2010):

$$\hat{\sigma}^2_{u,i} = \hat{u}^2_i 2p_i(1 - p_i)$$

In which: $\hat{u}^2_i$ is the square of the *i*th SNP marker effect, p_i is the observed allele frequency for the second allele of the *i*th marker in the current population.

When windows of *n* adjacent SNPs were used; the variances attributed to them were calculated by summing the variance of the next *n* SNPs, for each SNP. Next, the combination that contained the highest values for exclusive windows was chosen to avoid double counting. It could happen that some windows had less than *n* SNPs if they were between two windows explaining more variance or in a window at the end or beginning of a chromosome. However, those smaller windows do not explain significant part of the variance.

The analyses were performed in four scenarios: complete data set; only genotypes and phenotypes from generations G1, G2, and G3; generations G2, G3, and G4; and from generations G3, G4, and G5. Numerator relationship matrix was complete in all scenarios. All ssGWAS computations were performed using the BLUPF90 family programs (Misztal et al., 2002) modified to account for genomic information (Aguilar et al., 2010).

The choice for ssGWAS was due to its ability to support phenotypes from ungenotyped animals directly, to handle multiple trait models, and to avoid spurious solutions on SNP effects due to sampling. Sampling in Bayesian alphabet family models is strongly dependent on priors and may produce spurious SNP estimates (Gianola et al., 2009; van Hulzen et al., 2012). Comparing GWAS models in a simulated population, Wang et al. (2012) showed that ssGWAS was the most accurate method to capture the effect of potential QTLs; windows of SNP effects were used in their study.

Table 1 | Number of animals with phenotypes and genotypes in each generation.

Generation	Phenotypes	Genotypes
G1	95,770	1,142
G2	72,795	1,165
G3	66,241	754
G4	52,808	801
G5	7,018	1,004
Total	294,632	4,866

Table 2 | Number of observations, mean, and SD for the three traits.

Trait	Observations	Mean	SD
Body weight	294,632	92.66	17.2
Breast meat	75,377	45.68	7.22
Leg score	294,632	1.17	0.38

RESULTS

Preliminary results showed small individual SNP variances for all three traits, with just a few SNPs explaining more than 0.5% of the variance of the trait (**Figure 1**). Experiments with different SNP

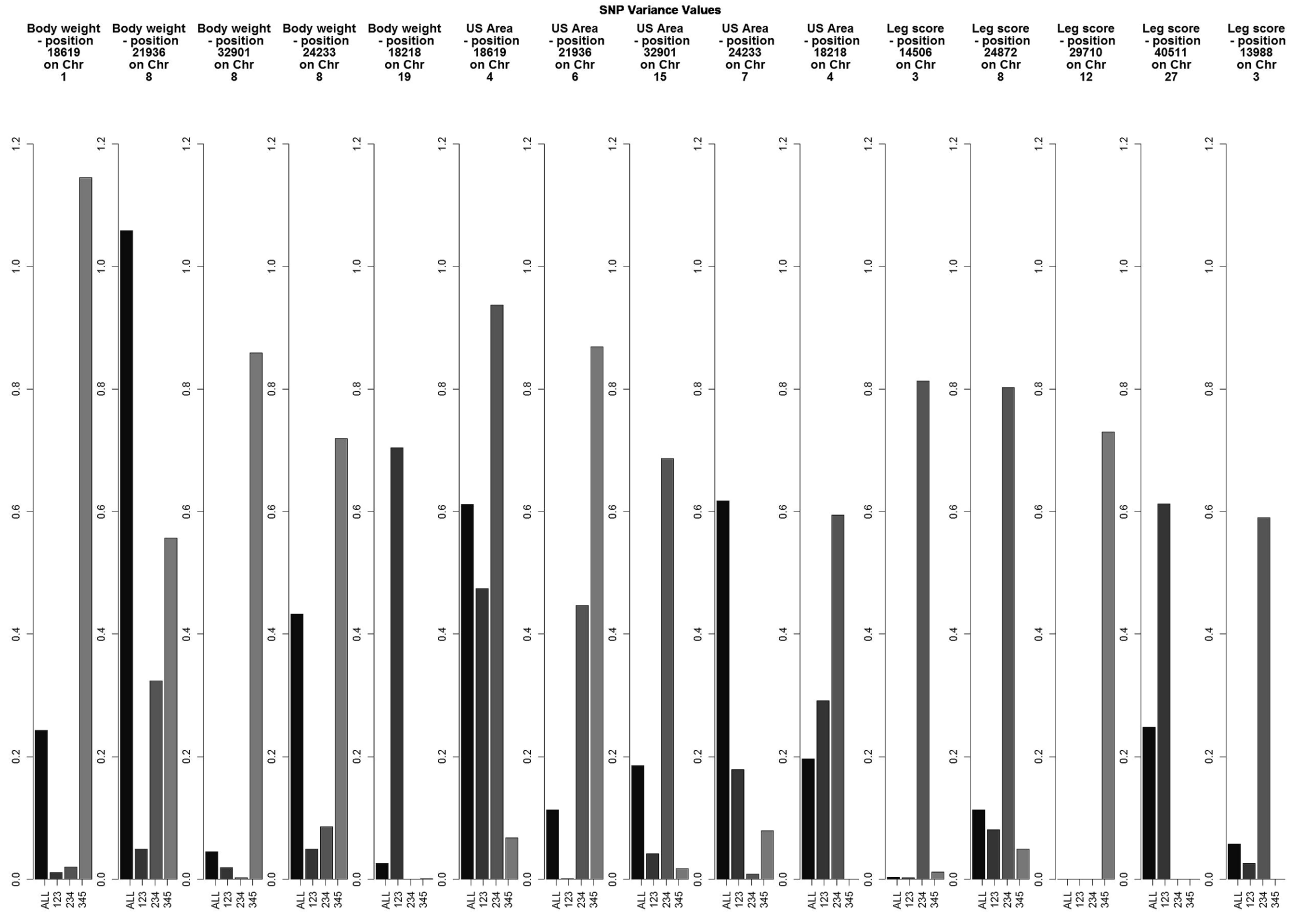


FIGURE 1 | Variance explained by the top five individual SNPs based on the combined results for all data sets for each trait.

window sizes exhibited large noise with small sizes and absence of peaks with large sizes. Subsequently, windows of 20 SNP were chosen as a reasonable size.

The variance explained by each SNP window is shown in **Figures 2–4** (corresponding to body weight, breast meat, and leg score, respectively); also, the 10 largest points were marked with a red vertical line. It is possible to see that all those traits are mainly affected by many regions with small effects, with few regions that explain more variance. These regions tended to change across the generations, but some of them retain a consistent value among the top 10 regions in all the scenarios, even though, the variance explained by those windows did not contribute significantly to the genetic variability of the trait.

For body weight, there were three regions that persisted among the top 10 in all the scenarios (**Figure 2**). Although these top three regions have been described before, the percentage of variance explained was small; only one region was above 2.5% and all the others were below 1.6%. The total variance explained by the top 10 windows summed up to 7.63%.

For breast meat, two regions were consistent among the scenarios (**Figure 3**). The window with larger effect for this trait explained 1.14% of the total variance, in the subset containing generations 3–5. The other windows explained at most 1%. The total variance explained by the top 10 windows was 6.26%.

For leg score, the value of just one region was constant across the analysis in chromosome 7 (**Figure 4**), the variance explained by this windows was 1.12% in the subset containing generations 3–5. All the other windows explained less than 1% of the genetic variance for this trait. The total variance explained by the sum of the top 10 windows was 6.01%.

DISCUSSION

In our study, the three persistent regions observed for body weight could be related with QTLs previously described in the literature. The region in chromosome 1 was consistent with the one described by Carlborg et al. (2003) that associated this with a QTL responsible for body weight. The region in chromosome 4 can be related with those found by Carlborg et al. (2004), Ikeobi et al. (2004), and Ankra-Badu et al. (2010), all of whom detected a QTL for body weight in this region. The region in chromosome 14 was close to that described by Jennen et al. (2004) and Carlborg et al. (2003) for body weight. For breast meat, the region in chromosome 3 was close to those reported by Ikeobi et al. (2004) and Uemoto et al. (2009) for pectoralis muscle mass, and to those found by Gao

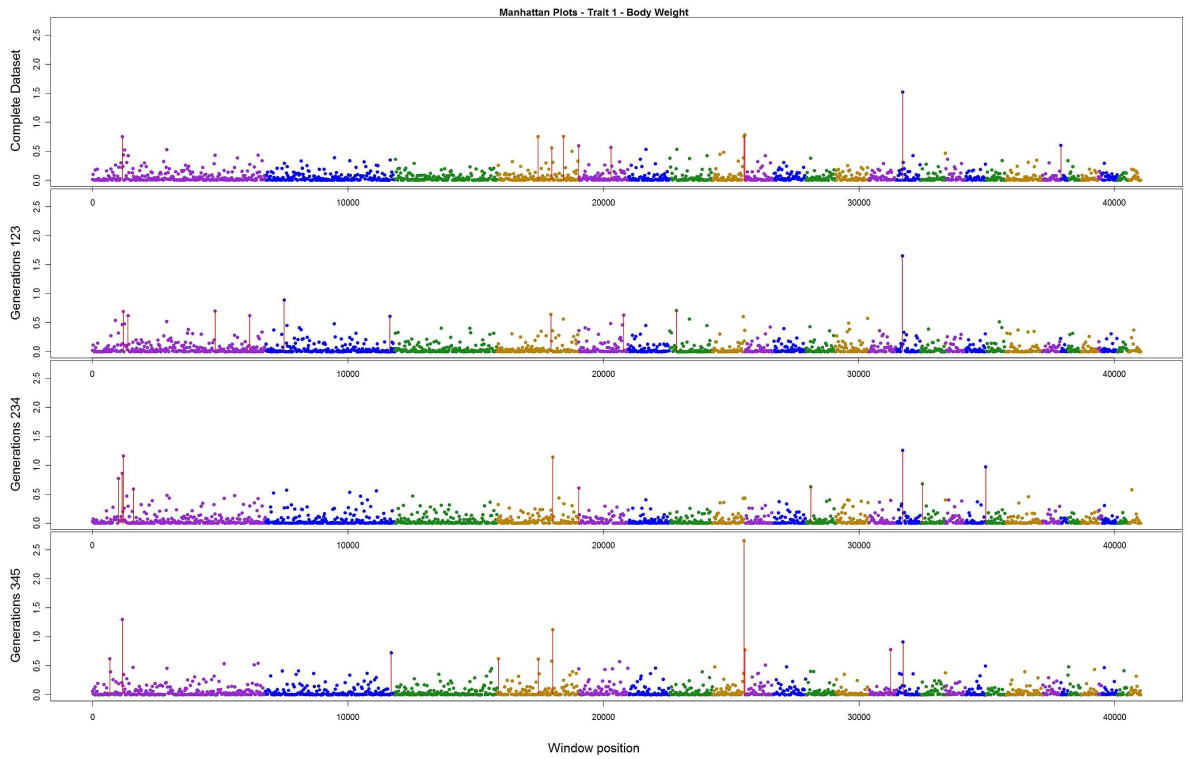


FIGURE 2 | Manhattan plots for percentage of variance explained for Body weight, performed for all the data set, and the subsets of generations.

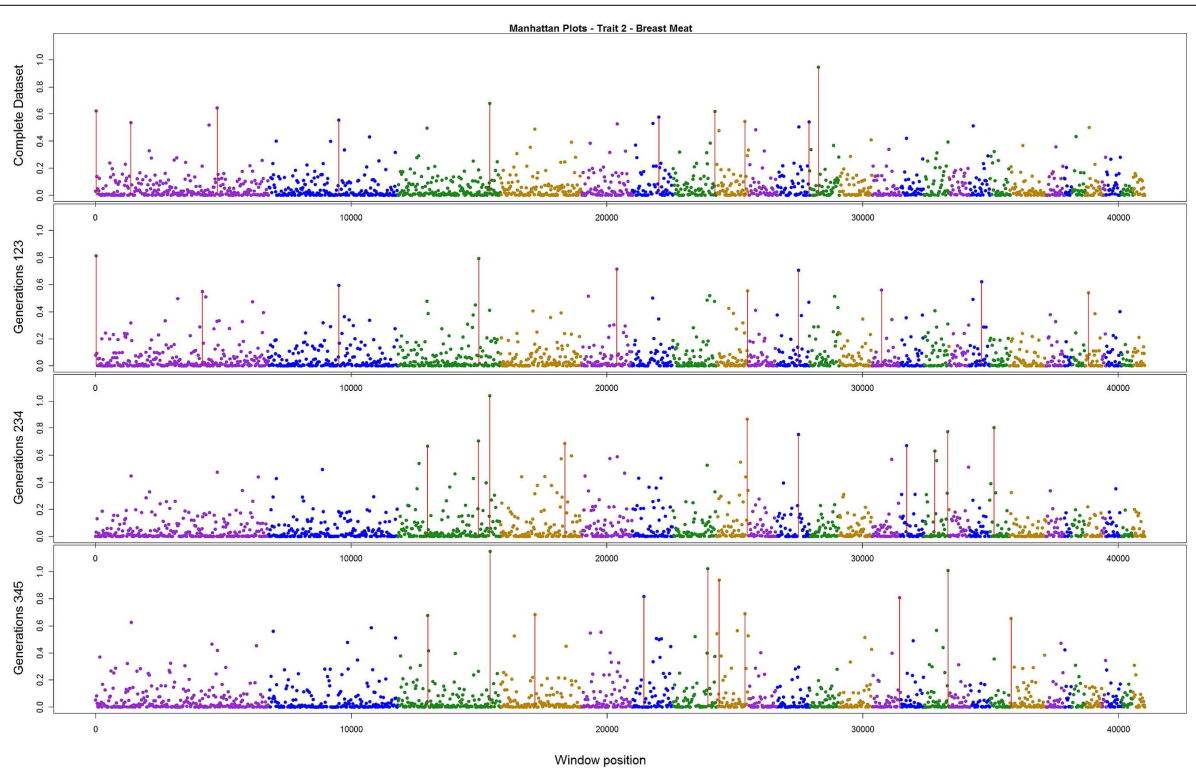


FIGURE 3 | Manhattan plots for percentage of variance explained for Breast meat, performed for all the data set, and the subsets of generations.

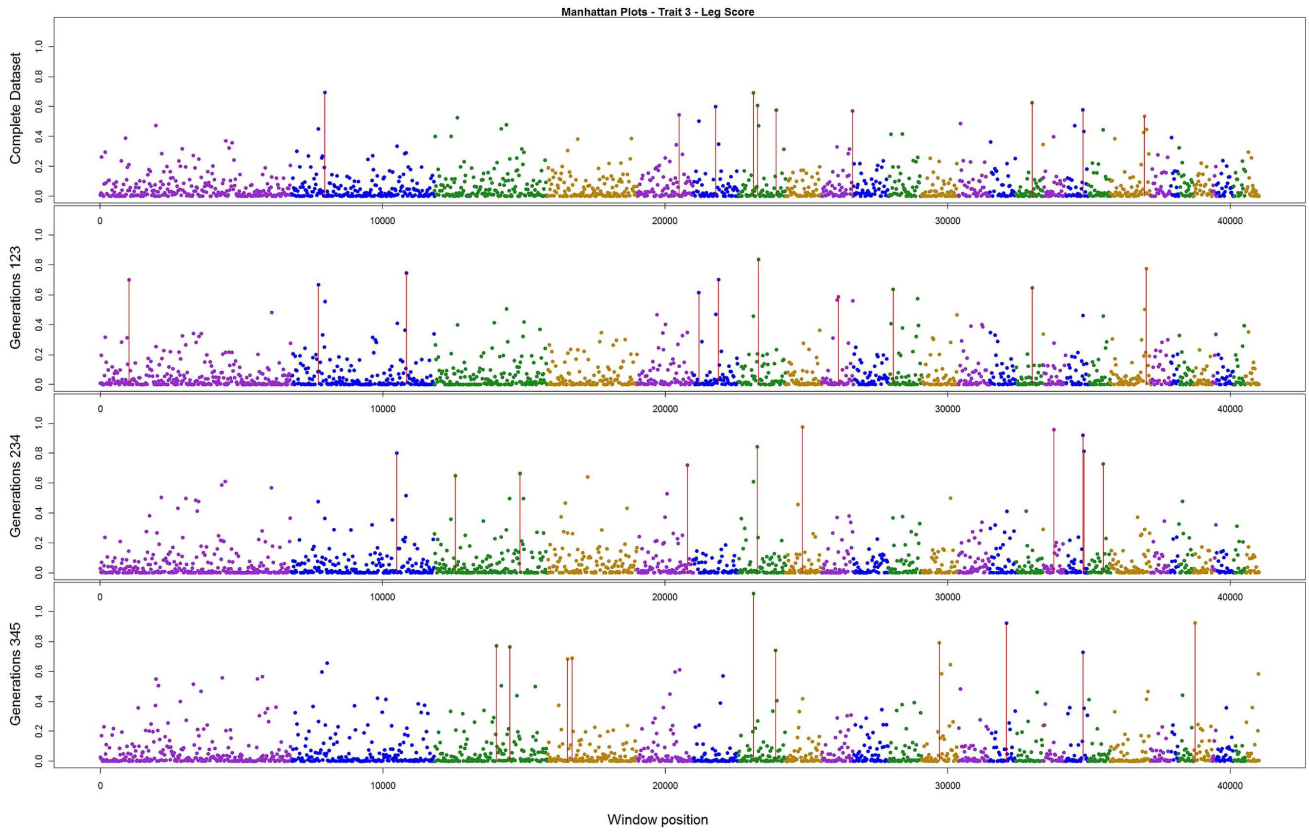


FIGURE 4 | Manhattan plots for percentage of variance explained for Leg score, performed for all the data set, and the subsets of generations.

et al. (2011) for chest width. The other region, in chromosome 8, was related by Ikeobi et al. (2004) to the pectoralis muscle mass trait. For leg score, the region in chromosome 7 had no relationship with any QTLs described previously in the literature for this trait in chicken. Nevertheless, there is a sequence of homeobox genes in the region around 16 Mbp in the same chromosome in the chicken genome. These homeobox genes (HOXD4, HOXD8, HOXD9, HOXD11, HOXD12, and HOXD13) are related with regulation of anatomical development, and might have a relationship with the leg disease score (Hillier et al., 2004). Thus, the findings in the current research are in concordance with Hayes and Goddard (2010), that a small number of markers with validated associations would explain a small portion of the genetic variance in the trait.

Wolc et al. (2012) found that for egg traits in layer chicken most of the SNPs with large effect were consistent across six generations, in both training and validating datasets. These findings could not be supported by the present results. Even though variances from three windows for body weight, two for breast meat, and one for leg disease score in the present study were stable across generations, for the other regions the results were different; it is possible that the lack of regions with larger effect on these traits, as illustrated in **Figures 2–4**, is the reason for the difference in findings. Another possible reason is the method used by the aforementioned authors; they used the Bayes B method, which assumes large effect for a few markers and is highly influenced by the prior information (Gianola et al., 2009; van Hulzen et al., 2012). In addition, the generation interval in layer chicken is a few times longer than in broiler chicken so their generations may have been overlapping. Yet, the genetic architecture could be different among the traits in the present study and in the aforementioned work.

Large changes in the variance explained by SNP windows could be indirectly due to small effective population size and subsequent low number of independent chromosome segments. According to Goddard (2009) and supported by Daetwyler et al. (2010), the number of such segments (q) is equal to $2N_eL/\log(4N_eL)$, where N_e is the effective population size and L is the length of chromosome in Morgans. Assuming $N_e = 50$ (lower range showed in Andreescu et al., 2007) and $L = 39$, $q = 435$. Subsequently there are >100 SNP per one chromosome segment, if we apply the formula to this dataset. This causes collinearity and possibly a high variance inflation factor for the estimators, amplified by changes to the effective population size during the selection. While 435 segments suggest that 435 SNP could explain nearly all variation, this is not so as the boundary between segments is fluid.

Meuwissen et al. (2001) have found a small decay in accuracy as the relationship between prediction and training generations decreases in a simulation study. According to the authors this decrease was small enough to maintain the success of breeding schemes after six generations without re-estimation of SNP effects, however, their simulation assumed random mating. Also

in a simulation study, Sonesson and Meuwissen (2009) found that re-estimating the genomic effects in every generation can maintain the accuracy of the predictions of breeding values constant. Solberg et al. (2009) also found a decrease in accuracy in further generations. They observed that with a denser panel the decay was smaller, which is probably a consequence of a higher LD between the markers and the simulated QTL. All above mentioned studies did not simulate selection in the data.

Muir (2007) showed that directional selection caused a great decline in accuracy of GEBV, demonstrating that high accuracies in the training generations were not maintained in future generations under selection. This can be a sign that the LD between marker and QTL can be lost across generations under selection, and can result in the changes observed in the present study. Alternatively, the QTL with largest effects are rapidly fixed by selection leaving SNPs with small effects remaining. In a real dataset from layer chicken, Wolc et al. (2011) demonstrated that the decay in accuracy was large enough to require a retrain of the model in every generation. Accurate estimations of genomic breeding value depend on the consistency of LD between markers and QTLs across generations (Calus, 2010), as well as proper SNP effect estimation. The LD is created and maintained by the selection process, among other factors (Lynch and Walsh, 1998). On the other hand, if a change in the allele frequency of two different loci is observed, which can be caused by selection, the LD between them can decrease (Calus, 2010). The results shown in those studies clearly display a loss of genomic prediction accuracy due to the decay of LD. This could also be extended to GWAS, and the negative impact LD decay might have on the accuracy of associations. The variation in the estimates of SNP variance in the present study can be related with those findings, because using values estimated in a different generation would lead to low predictive power if they are not constant.

The small values for SNP effect and percentage of variance explained that were obtained in this study can be related to the findings on Muir et al. (2008). The authors found significant absence of rare alleles in commercial chicken lines. Such findings were related to high inbreeding and consequently to a considerable number of alleles missing, which will reduce the allelic and genetic variability. This narrowed genetic variability can result in weaker associations for the markers, since important alleles could be lost in the process.

The short-term decay in accuracy depends more on the decrease of genomic relationships captured by markers rather than on LD (Habier et al., 2007). Therefore, the accuracy of genomic evaluation is mainly controlled by genomic relationships (Daetwyler et al., 2012; Wientjes et al., 2013). In particular, Daetwyler et al. (2012) found that 86% of the accuracy in genomic selection was retrieved by using SNP from a single chromosome. Subsequently, windows with large effects in Manhattan plots may be an artifact of relationships and not due to LD. The reason why the accuracy does not collapse completely in further generations is that some LD still persists over time, even though selection process and divergence can erode LD. Thus, the observed changes in the SNP effects across the generations in the present study can be a consequence of the changes in the relationship structure across different generations more than decay in LD.

CONCLUSION

Except for a few regions, the variation explained by the top SNP windows changes over generations. Therefore, even if SNP windows with large variance are detected in a particular data set, their usefulness for genomic selection over many generations is limited. The variance explained by an individual window is not enough to lead selection decisions based on the top regions for the studied traits.

ACKNOWLEDGMENTS

This project was supported by Agriculture and Food Research Initiative Competitive Grant no. 2009-65205-05665 from the USDA National Institute of Food and Agriculture. We would like to acknowledge Miguel Perez-Enciso for the useful comments made on this paper and Cobb-Vantress Inc. for providing the data.

REFERENCES

Aguilar, I., Misztal, I., Johnson, D. L., Legarra, A., Tsuruta, S., and Lawlor, T. J. (2010). Hot topic: a unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *J. Dairy Sci.* 93, 743–752. doi: 10.3168/jds.2009-2730

Andreescu, C., Avendano, S., Brown, S. R., Hassen, A., Lamont, S. J., and Dekkers, J. C. (2007). Linkage disequilibrium in related breeding lines of chickens. *Genetics* 177, 2161–2169. doi: 10.1534/genetics.107.082206

Ankra-Badu, G., Shriner, D., Le Bihan-Duval, E., Mignon-Grasteau, S., Pitel, F., Beaumont, C., et al. (2010). Mapping main, epistatic and sex-specific QTL for body composition in a chicken population divergently selected for low or high growth rate. *BMC Genomics* 11:107. doi: 10.1186/1471-2164-11-107

Bolormaa, S., Pryce, J. E., Hayes, B. J., and Goddard, M. E. (2010). Multivariate analysis of a genome-wide association study in dairy cattle. *J. Dairy Sci.* 93, 3818–3833. doi: 10.3168/jds.2009-2980

Calus, M. P. L. (2010). Genomic breeding value prediction: methods and procedures. *Animal* 4, 157–164. doi: 10.1017/S1751731109991352

Carlborg, O., Hocking, P. M., Burt, D. W., and Haley, C. S. (2004). Simultaneous mapping of epistatic QTL in chickens reveals clusters of QTL pairs with similar genetic effects on growth. *Genet. Res.* 83, 197–209. doi: 10.1017/S0016672304006779

Carlborg, Ö., Kerje, S., Schütz, K., Jacobsson, L., Jensen, P., and Andersson, L. (2003). A global search reveals epistatic interaction between QTL for early growth in the chicken. *Genome Res.* 13, 413–421. doi: 10.1101/gr.528003

Chen, C. Y., Misztal, I., Aguilar, I., Tsuruta, S., Aggrey, S. E., Wing, T., et al. (2011). Genome-wide marker-assisted selection combining all pedigree phenotypic information with genotypic data in one step: An example using broiler chickens. *J. Anim. Sci.* 89, 23–28. doi: 10.2527/jas.2010-3071

Daetwyler, H. D., Kemper, K. E., Van der Werf, J. H. J., and Hayes, B. J. (2012). Components of the accuracy of genomic prediction in a multi-breed sheep population. *J. Anim. Sci.* 90, 3375–3384. doi: 10.2527/jas.2011-4557

Daetwyler, H. D., Pong-Wong, R., Villanueva, B., and Woolliams, J. A. (2010). The impact of genetic architecture on genome-wide evaluation methods. *Genetics* 185, 1021–1031. doi: 10.1534/genetics.110.116855

Dikmen, S., Cole, J. B., Null, D. J., and Hansen, P. J. (2013). Genome-wide association mapping for identification of quantitative trait loci for rectal temperature during heat stress in holstein cattle. *PLoS ONE* 8:7. doi: 10.1371/journal.pone.0069202

Gao, Y., Feng, C. G., Song, C., Du, Z. Q., Deng, X. M., Li, N., et al. (2011). Mapping quantitative trait loci affecting chicken body size traits via genome scanning. *Anim. Genet.* 42, 670–674. doi: 10.1111/j.1365-2052.2011.02193.x

Gianola, D., de los Campos, G., Hill, W. G., Manfredi, E., and Fernando, R. (2009). Additive genetic variability and the Bayesian alphabet. *Genetics* 183, 347–363. doi: 10.1534/genetics.109.103952

Goddard, M. (2009). Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica* 136, 245–257. doi: 10.1007/s10709-008-9308-0

Goddard, M. E., and Hayes, B. J. (2009). Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nat. Rev. Genet.* 10, 381–391. doi: 10.1038/nrg2575

Groenen, M. A., Megens, H. J., Zare, Y., Warren, W. C., Hillier, L. W., Crooijmans, R. P., et al. (2011). The development and characterization of a 60K SNP chip for chicken. *BMC Genomics* 12:274–283. doi: 10.1186/1471-2164-12-274

Habier, D., Fernando, R. L., and Dekkers, J. C. M. (2007). The impact of genetic relationship information on genome-assisted breeding values. *Genetics* 177, 2389–2397. doi: 10.1534/genetics.107.081190

Habier, D., Fernando, R. L., Kizilkaya, K., and Garrick, D. J. (2011). Extension of the Bayesian alphabet for genomic selection. *BMC Bioinformatics* 12:186. doi: 10.1186/1471-2105-12-186

Hayes, B., and Goddard, M. (2010). Genome-wide association and genomic selection in animal breeding. *Genome* 53, 876–883. doi: 10.1139/G10-076

Hillier, L. W., Miller, W., Birney, E., Warren, W., Hardison, R. C., Ponting, C. P., et al. (2004). Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432, 695–716. doi: 10.1038/nature03154

Hu, Z. L., Park, C. A., Wu, X. L., and Reecy, J. M. (2013). Animal QTLdb: an improved database tool for livestock animal QTL/association data dissemination in the post-genome era. *Nucleic Acids Res.* 41, D871–D879. doi: 10.1093/nar/gks1150

Ikeobi, C. O. N., Woolliams, J. A., Morrice, D. R., Law, A., Windsor, D., Burt, D. W., et al. (2004). Quantitative trait loci for meat yield and muscle distribution in a broiler layer cross. *Livest. Prod. Sci.* 87, 143–151. doi: 10.1016/j.livprodsci.2003.09.020

Jennen, D. G., Vereijken, A. L., Bovenhuis, H., Crooijmans, R. P., Veenendaal, A., Van der Poel, J. J., et al. (2004). Detection and localization of quantitative trait loci affecting fatness in broilers. *Poult. Sci.* 83, 295–301. doi: 10.1093/ps/83.3.295

Lynch, M., and Walsh, B. (1998). *Genetics and Analysis of Quantitative Traits*. Sunderland: Sinauer Associates.

Meuwissen, T., Hayes, B. J., and Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819–1829.

Misztal, I., Legarra, A., and Aguilar, I. (2009). Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. *J. Dairy Sci.* 92, 4648–4655. doi: 10.3168/jds.2009-2064

Misztal, I., Tsuruta, S., Strabel, T., Auvray, B., Druet, T., and Lee, D. H. (2002). "BLUPF90 and related programs (BGF90)," in *Proceedings of the 7th World Congress on Genetics Applied to Livestock Production*, Montpellier, Communication No. 28–27.

Muir, W. M. (2007). Comparison of genomic and traditional BLUP-estimated breeding value accuracy and selection response under alternative trait and genomic parameters. *J. Animal Breed. Genet.* 124, 342–355. doi: 10.1111/j.1439-0388.2007.00700.x

Muir, W. M., Wong, G. K.-S., Zhang, Y., Wang, J., Groenen, M. A., Crooijmans, R. P., et al. (2008). Genome-wide assessment of worldwide chicken SNP genetic diversity indicates significant absence of rare alleles in commercial breeds. *Proc. Natl. Acad. Sci. U.S.A.* 105, 17312–17317. doi: 10.1073/pnas.0806569105

Peters, S. O., Kizilkaya, K., Garrick, D. J., Fernando, R. L., Reecy, J. M., Weaber, R. L., et al. (2012). Bayesian genome-wide association analysis of growth and yearling ultrasound measures of carcass traits in Brangus heifers. *J. Anim. Sci.* 90, 3398–3409. doi: 10.2527/jas.2011-4507

Solberg, T. R., Sonesson, A. K., Woolliams, J. A., Odegard, J., and Meuwissen, T. H. (2009). Persistence of accuracy of genome-wide breeding values over generations when including a polygenic effect. *Genet. Sel. Evol.* 41, 53. doi: 10.1186/1297-9686-41-53

Sonesson, A. K., and Meuwissen, T. H. E. (2009). Testing strategies for genomic selection in aquaculture breeding programs. *Genet. Sel. Evol.* 41, 37. doi: 10.1186/1297-9686-41-37

Sun, X., Habier, D., Fernando, R. L., Garrick, D. J., and Dekkers, J. C. M. (2011). Genomic breeding value prediction and QTL mapping of QTLMAS2010 data using Bayesian methods. *BMC Genet.* 5(Suppl. 3), S13. doi: 10.1186/1753-6561-5-S3-S13

Uemoto, Y., Sato, S., Odawara, S., Nokata, H., Oyamada, Y., Taguchi, Y., et al. (2009). Genetic mapping of quantitative trait loci affecting growth and carcass traits in F2 intercross chickens. *Poult. Sci.* 88, 477–482. doi: 10.3382/ps.2008-00296

van Hulzen, K. J., Schopen, G. C., van Arendonk, J. A., Nielen, M., Koets, A. P., Schrooten, C., et al. (2012). Genome-wide association study to identify chromosomal regions associated with antibody response to Mycobacterium avium subspecies paratuberculosis in milk of Dutch Holstein-Friesians. *J. Dairy Sci.* 95, 2740–2748. doi: 10.3168/jds.2011-5005

Vitezica, Z. G., Aguilar, I., Misztal, I., and Legarra, A. (2011). Bias in genomic predictions for populations under selection. *Genet. Res. (Camb.)* 93, 357–366. doi: 10.1017/S001667231100022X

Wang, H., Misztal, I., Aguilar, I., Legarra, A., and Muir, W. M. (2012). Genome-wide association mapping including phenotypes from relatives without genotypes. *Genet. Res. (Camb.)* 94, 73–83. doi: 10.1017/S0016672312000274

Wientjes, Y. C., Veerkamp, R. F., and Calus, M. P. (2013). The effect of linkage disequilibrium and family relationships on the reliability of genomic prediction. *Genetics* 193, 621–631. doi: 10.1534/genetics.112.146290

Wolc, A., Arango, J., Settar, P., Fulton, J. E., O'Sullivan, N. P., Preisinger, R., et al. (2011). Persistence of accuracy of genomic estimated breeding values over generations in layer chickens. *Genet. Sel. Evol.* 43:23. doi: 10.1186/1297-9686-43-23

Wolc, A., Arango, J., Settar, P., Fulton, J. E., O'Sullivan, N. P., Preisinger, R., et al. (2012). Genome-wide association analysis and genetic architecture of egg weight and egg uniformity in layer chickens. *Anim. Genet.* 43(Suppl 1), 87–96. doi: 10.1111/j.1365-2052.2012.02381.x.

Zhang, Z., Liu, J., Ding, X., Bijma, P., de Koning, D. J., and Zhang, Q. (2010). Best linear unbiased prediction of genomic breeding values using a trait-specific marker-derived relationship matrix. *PLoS ONE* 5:9. doi: 10.1371/journal.pone.0012648

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 11 June 2014; accepted: 04 September 2014; published online: 01 October 2014.

*Citation: Fragomeni BDO, Misztal I, Lourenco DL, Aguilar I, Okimoto R and Muir WM (2014) Changes in variance explained by top SNP windows over generations for three traits in broiler chicken. Front. Genet. **5**:332. doi: 10.3389/fgene.2014.00332*

This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics.

Copyright © 2014 Fragomeni, Misztal, Lourenco, Aguilar, Okimoto and Muir. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in GENETICS

ORIGINAL RESEARCH ARTICLE
published: 09 February 2015
doi: 10.3389/fgene.2015.00007

Genetic differentiation of Mexican Holstein cattle and its relationship with Canadian and U.S. Holsteins

Adriana García-Ruiz[1], Felipe de J. Ruiz-López[1,2] *, Curtis P. Van Tassell[3], Hugo H. Montaldo[1] and Heather J. Huson[3,4]

[1] *Facultad de Estudios Superiores Cuautitlán, Universidad Nacional Autónoma de México, Ajuchitlán, Mexico*
[2] *Centro Nacional de Investigación Disciplinaria en Fisiología y Mejoramiento Animal, Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias – Secretaría de Agricultura, Ganadería, Desarrollo Rural, Pesca y Alimentación, Ajuchitlán, Mexico*
[3] *Animal Genomics and Improvement Laboratory, Agricultural Research Service, United State Department of Agriculture, Beltsville, MD, USA*
[4] *Department of Animal Science, Cornell University, Ithaca, NY, USA*

Edited by:
Johann Sölkner, BOKU – University of Natural Resources and Life Sciences Vienna, Austria

Reviewed by:
Ikhide G. Imumorin, Cornell University, USA
Bertrand Servin, Institut National de la Recherche Agronomique, France

****Correspondence:***
Felipe de J. Ruiz-López, Centro Nacional de Investigación Disciplinaria en Fisiología y Mejoramiento Animal, Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias – Secretaría de Agricultura, Ganadería, Desarrollo Rural, Pesca y Alimentación, Km. 1 Carretera a Colón, Ajuchitlán, Querétaro 76280, Mexico
e-mail: ruiz.felipe@inifap.gob.mx

The Mexican Holstein (HO) industry has imported Canadian and US (CAN + USA) HO germplasm for use in two different production systems, the conventional (Conv) and the low income (Lowi) system. The objective of this work was to study the genetic composition and differentiation of the Mexican HO cattle, considering the production system in which they perform and their relationship with the Canadian and US HO populations. The analysis included information from 149, 303, and 173 unrelated or with unknown pedigree HO animals from the Conv, Lowi, and CAN + USA populations, respectively. Canadian and US Jersey (JE) and Brown Swiss (BS) genotypes (162 and 86, respectively) were used to determine if Mexican HOs were hybridized with either of these breeds. After quality control filtering, a total of 6,617 out of 6,836 single nucleotide polymorphism markers were used. To describe the genetic diversity across the populations, principal component (PC), admixture composition, and linkage disequilibrium (LD; r^2) analyses were performed. Through the PC analysis, HO × JE and HO × BS crossbreeding was detected in the Lowi system. The Conv system appeared to be in between Lowi and CAN + USA populations. Admixture analysis differentiated between the genetic composition of the Conv and Lowi systems, and five ancestry groups associated to sire's country of origin were identified. The minimum distance between markers to estimate a useful LD was found to be 54.5 kb for the Mexican HO populations. At this average distance, the persistence of phase across autosomes of Conv and Lowi systems was 0.94, for Conv and CAN + USA was 0.92 and for the Lowi and CAN + USA was 0.91. Results supported the flow of germplasm among populations being Conv a source for Lowi, and dependent on migration from CAN + USA. Mexican HO cattle in Conv and Lowi populations share common ancestry with CAN + USA but have different genetic signatures.

Keywords: genetic differentiation, Holstein, admixture, linkage disequilibrium

INTRODUCTION

Dairy farms in Mexico are extremely heterogeneous (e.g., different herd sizes, feeding systems, reproductive management, etc.). Conventional (Conv) dairy farms have an average herd size of 230 head and are highly mechanized and milk yield is relatively high. Cows on these farms are typically grouped in pens and rations usually include high proportions of concentrates. Low income (Lowi) systems vary with region with sizes ranging from 3 to 30 cows, with animals usually spending part of the day grazing, but may also be housed in pens (Amendola, 2002). Additionally, Lowi farms rely heavily on the use of unpaid family labor and the typical herd is smaller than Conv herds. The Mexican Holstein (HO) population has depended on US and Canada genetics for some years (Powell and Wiggans, 1991). Recently, germplasm from European populations have been introduced, however, these populations also tend to be highly influenced by US and Canadian bulls (Norman and Powell, 1999). This information was expected in the Conv Mexican systems where pedigree information was available, but little was known about the genetic influence on the Lowi system because of incomplete or unavailable pedigree information.

Before single nucleotide polymorphism (SNP) data became available, genetic diversity and relatedness was studied through the analysis of pedigree information which, unfortunately, is not always complete. This incomplete data renders results inaccurate or limits the performance of studies of individuals or populations (Boichard et al., 1997). Currently, with the availability of SNP data, it is possible to estimate breed or population compositions without the previous knowledge of ancestry information (Sölkner et al., 2010; Frkonja et al., 2012). Different analyses based on genomic information have been used to study genetic diversity of populations. In this study, principal component analysis (PCA) was performed to describe the breed or geographic allele variation, whereas admixture analysis was used to describe the structure of populations. Persistence phase of linkage disequilibrium (LD)

analysis allowed the characterization of the degree of agreement of LD across distances between populations.

Principal component analysis was first used in human populations to generate maps summarizing the allele frequency of different geographic areas (Cavalli-Sforza et al., 1994; Silva-Zolezzi et al., 2009) and more recently for understanding relationship among cattle breeds (Bovine HapMap Consortium, 2009; Lewis et al., 2011). Currently, PCA is used to control spurious genome wide association in populations structured with individuals of different geographic areas or even when geography does not explain the genetic background (Novembre and Stephens, 2008). In this study, PCA was used to identify differences among HO cattle originating from CAN + USA and two Mexican production systems.

Admixture is defined as the mixing of genomes of divergent parental origins (Buerkle and Lexer, 2008), which implies the presence of multiple genetically distinct groups or breeds in a population (Wang et al., 2005). Admixture can be studied at an individual (Tang et al., 2005) or population level (Buerkle and Lexer, 2008; Frkonja et al., 2012). These haplotype blocks vary in size because of the random nature of recombination, but become progressively shorter by further recombination with increasing generations (Winkler et al., 2010).

Linkage disequilibrium is defined as a non-random association of alleles at different loci (Du et al., 2007; Waples and England, 2011) because the recombination rate differs from that expected if the loci segregated independently. LD is common between alleles at neighboring loci that tend to be inherited together and associated in a segregating population (Du et al., 2007), but can also be associated with selection (Bulmer, 1971). Characterization of LD is used to assess whether two or more populations can be jointly analyzed in genomic studies, because markers in LD in one population may not be in LD in another population (de Roos et al., 2008), and to make meaningful inferences in populations other than the reference population will depend on the persistence of LD phase between the two populations (Dekkers and Hospital, 2002). The LD level in a population is also used for determining the required marker map resolution to be used in a genomic selection program and testing associations based on QTL scans (McKay et al., 2007). The LD of sufficiently large degree that allows the QTL scan is known as the useful LD (Lu et al., 2012).

To establish genomic evaluations in Mexico, it is important to determine whether multiple HO subpopulations exist. This information determines whether it is necessary to stratify the population by production system and if foreign genomic information, especially from Canada and US, would improve accuracy of predicted breeding values. Thus, the objective of this work was to study the genetic composition and differentiation of the Mexican HO population and the relationship between Mexican cattle with those from Canada and the US. The primary source of stratification of the Mexican population considered was the production system in which those cows performed.

MATERIALS AND METHODS

ANIMALS, BREEDS AND GENOTYPES

A total of 625 HO and HO like unrelated cows and sires, born on or after 2005 and genotyped with the Illumina BovineSNP50 or BovineLD Bead Chips were used in this analysis. A total of 149 and 303 animals assigned to the Conv and Lowi systems, respectively, and 173 were from the Canadian and US HO populations. In addition, 162 Jersey (JE) and 86 Brown Swiss (BS) sires from Canada and US were included to help determine if crossbred cows were present in the populations. The animals from the Conv system originated from 16 herds in 6 states of Mexico (Aguascalientes, Guanajuato, Estado de México, Querétaro, San Luis Potosí, and Zacatecas) while the animals of the Lowi system were from 21 herds in 4 states (Estado de México, Jalisco, Puebla, and Tlaxcala). From the 6,836 common SNP markers in both the Illumina BovineSNP50 and the BovineLD Bead Chips a total of 6,617 SNP were included in the analysis after quality control. Markers with a minor allele frequency less than 2% and call rate less than 90% were excluded. Individuals with a call rate less than 90% were also excluded. **Table 1** shows the number and frequency of markers per chromosome included in the analysis.

Table 1 | Number and frequency of single nucleotide polymorphism (SNP) per chromosome included in the analysis.

Chromosome	Number of SNP by chromosome	SNP percentage by chromosome
1	398	6.01
2	349	5.27
3	310	4.68
4	307	4.64
5	304	4.59
6	310	4.68
7	284	4.29
8	299	4.52
9	272	4.11
10	269	4.07
11	285	4.31
12	225	3.40
13	218	3.29
14	226	3.42
15	223	3.37
16	206	3.11
17	195	2.95
18	176	2.66
19	182	2.75
20	204	3.08
21	186	2.81
22	169	2.55
23	152	2.30
24	177	2.67
25	142	2.15
26	145	2.19
27	139	2.10
28	128	1.93
29	137	2.07

POPULATION COMPOSITION

The PCA proposed by Price et al. (2006) was used in this study because it models ancestry differences along continuous axes of variation. Genotypes of BS, HO, and JE animals were used. Sire country of origin was considered to explain variation among HO populations and for Mexican cattle; the production system was used as classification variable resulting in three distinct HO populations (Lowi, Conv, and CAN + USA).

GENETIC STRUCTURE OF MEXICAN HO

The ADMIXTURE package (Alexander et al., 2009) was used for this purpose because it implements a fast model-based estimation that assumes that individuals come from an admixed population with contributions from K ancestral populations. Each K population contributes a fraction q_{ik} for each individual i. Only Mexican HO animals were used for this analysis. Once the K ancestral populations were determined, the most common sire's country of origin of each ancestral population was identified based on pedigree information.

LINKAGE DISEQUILIBRIUM DIFFERENCES AMONG HO POPULATIONS

The LD, measured as r^2 for alleles at two loci was calculated as:

$$r^2 = \frac{D_{ij}^2}{p_1 p_2 q_1 q_2}$$

Where D is the difference between the observed and the expected frequency of two loci, based on population allele frequencies and assuming random assortment and can be estimated directly from the allele frequencies (Waples and England, 2011), $p_1 p_2 q_1$ and q_2 are the observed frequencies of alleles 1, 2 respectively (Hill and Robertson, 1968). The value, r^2, is considered the most robust measure of LD. Persistence phase of LD was calculated as the Pearson correlation coefficient between the root of r^2 between populations for the same pair of SNP (Badke et al., 2012).

Quality control, PCA and LD analysis were performed with SVS Golden Helix software (SNP and Variation Suite Manual v7, 2013; *Golden Helix, Inc.*), persistence phase of LD was calculated using SAS 9.2 (SAS Institute, 2009).

RESULTS

POPULATION COMPOSITION

In this study, the first three of 625 components of PCA explained 13% of the observed variation. Population differentiation is observed between the Mexican Lowi and Conv subgroups and the Canadian and US HO cattle (**Figure 1**), despite many common ancestors across these groups. The Conv system seems to be intermediate between the Lowi and CAN + USA groups. Neither Mexican system demonstrated a clustering of animals by country of origin of the sire. PCs for all North American HO cattle along with the JE and BS cattle are shown in **Figure 2**, where the individuals were color coded by breed and population of origin. Crossbred individuals derived from HO and JE or BS were represented by points located between those pure breeds. Crossbreeding was rare in the Conv herds but was much more common in the Lowi systems with a higher proportion of crossing to JE influenced animals than BS.

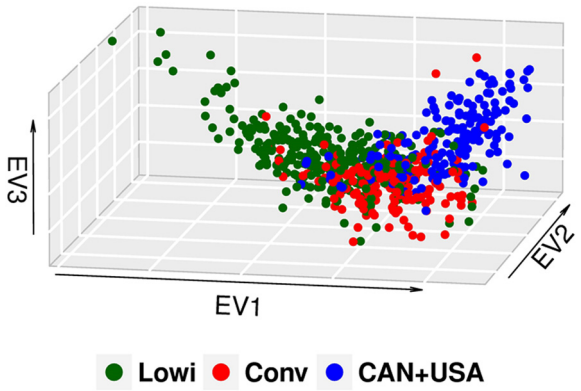


FIGURE 1 | Principal component analysis (PCA) plot of the North American (CAN + USA) and Mexican Holstein population, classifying the animals according to the production system in which they perform conventional (Conv) and low income (Lowi).

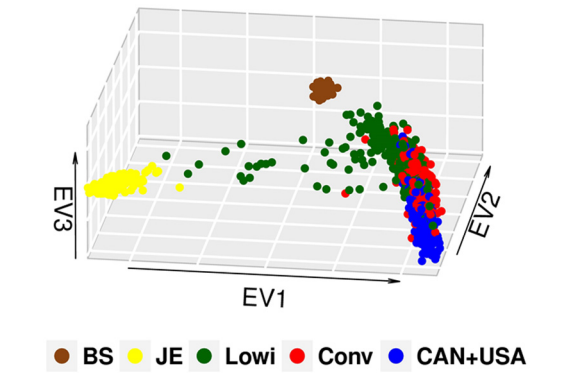


FIGURE 2 | Principal component analysis plot of the Mexican Holstein cattle of the conventional (Conv) and low income system (Lowi), the North American Holstein (CAN + USA), Jersey (JE) and Brown Swiss (BS).

GENETIC STRUCTURE OF MEXICAN HO

To identify the genetic structure of the Mexican HO systems and its dependency on the Canadian and US populations, an admixture analysis was performed first for the entire Mexican population and then for each sub-population. For both analyses, the value that best explained the stratification of the population was K = 6. Pedigree records of individual Mexican HO identified their sire's country of origin as Canadian, US, or Mexican, hence providing an indirect measure of the genetic contribution of these countries to Mexican HO. Indeed, the ADMIXTURE strata identified at $K = 6$ (**Figure 3**) correlated to the sire's country of origin. Two of the ADMIXTURE strata were directly related to sires originating from Canada and were subsequently combined into group A with the purpose of investigating country of origin influence. Group B consisted of sires most commonly from the United States (USA). The third strata, group C, consisted of sires from both Canada and US. No unique lineages or discernable characteristics were associated with group C. Another stratum was associated to sires registered in the Mexican herd book and assigned as group

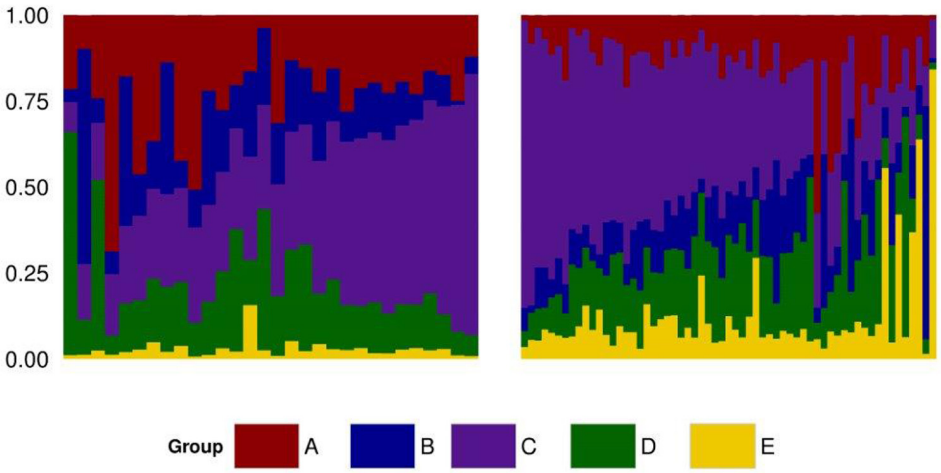


FIGURE 3 | Admixture plot of the Mexican Holstein population including the conventional (left) and low income system (right), identifying five ancestral groups (A, B, C, D, and E).

D. The last stratum included crossbred animals as evidenced by their genetic similarity to BS and JE breeds in the PC analysis and were assigned to group E. Average ancestry contributions, both overall and by production system, can be observed in **Figure 4**. Population structure showed differences in the genetic composition of the Mexican HO production systems where group C had the largest contribution to the Conv and Lowi populations at 37 and 40%, respectively. In the Conv population, group A (24%), D (19%), B (17%) and E (2.8%) followed while for Lowi animals, group C was followed by D (19%), B (14%), A (14%) and the E (13%). Overall, the contributions of the US, Canada/US, and the Mexican sires were relatively similar among the production systems. The primary differences were within the Canadian and crossbred lineages. The Conv population had approximately 1.7 fold greater contribution from Canadian lineage (group A; Conv-24.16% vs. Lowi-13.91%) and the Lowi population had ∼4.3 fold increase in crossbred lineages (group E; Conv-2.81% vs. Lowi-12.51%).

LINKAGE DISEQUILIBRIUM DIFFERENCES AMONG HO POPULATIONS

Mean LD, calculated as r^2 for different distances (at intervals of 100 Kbp) between SNP were calculated for the Conv, Lowi, and CAN + USA populations (**Figure 5**). At all distances, average r^2 was highest for CAN + USA animals, intermediate for individuals from the Mexican Conv farms, and smallest for cattle representing the Lowi systems. The differences between the CAN + USA and

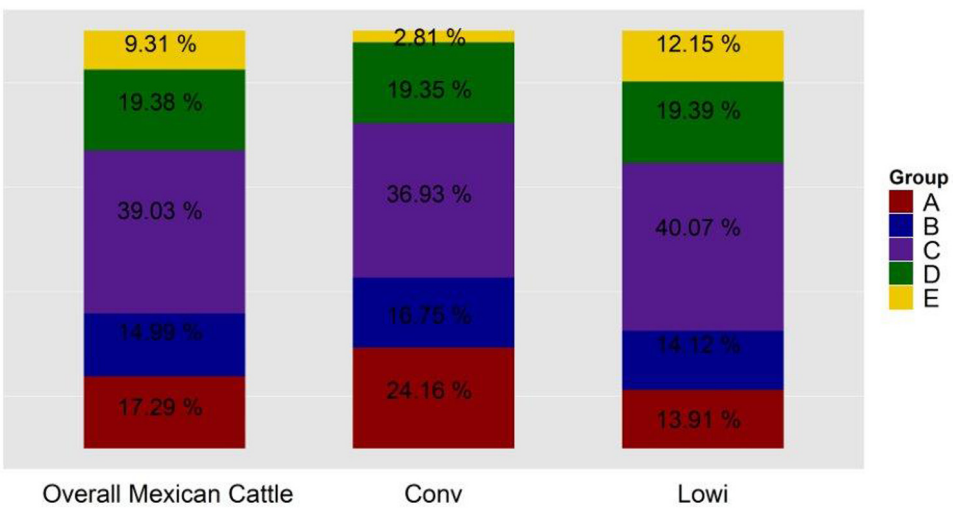


FIGURE 4 | Genetic composition of the Mexican Holstein population, overall and by the production system in which they perform: conventional (Conv) and low income (Lowi).

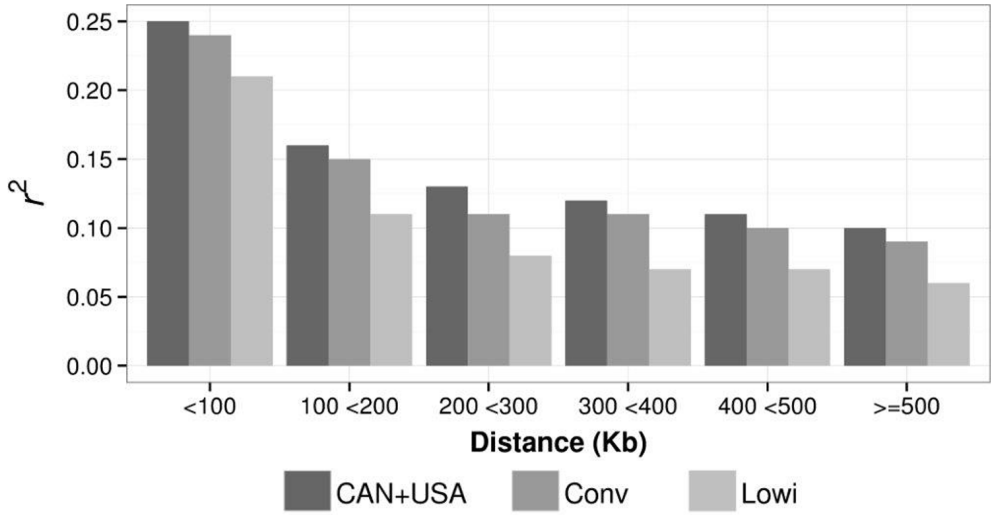


FIGURE 5 | Linkage Disequilibrium average for different genetic distances between SNP pairs in the Mexican Holstein systems [conventional (Conv) and low income (Lowi)] and the North American Holstein population (CAN + USA).

Conv populations r^2 were quite small (∼0.01) while differences between Conv and Lowi were larger and consistent, ranging from 0.03 to 0.04. The persistence of LD phase between Conv, Lowi, and CAN + USA populations was calculated at the same interval distances as was LD (**Figure 6**). As expected, the persistence of LD phase decreased when the distance between markers was increased. At all intervals, the highest correlations were between Conv and Lowi populations, with the lowest correlation being between Lowi and CAN + USA. At a distance of <100 Kb (with an average of 54.5 kb), the correlations ranged from 0.91 to 0.94 and for a distance >500 kb correlations varied from 0.75 to 0.81 between the Conv-Lowi and Lowi-CAN + USA, respectively.

DISCUSSION

The results obtained in this study, provide important details to be considered for future genetic research on the Mexican HO cattle and provided the opportunity to measure the genetic relationship of the Mexican animals with the HO populations of Canada and the US.

The Conv system seems to be intermediate in most aspects between the Lowi and CAN + USA, and this population may act as a conduit for germplasm flow from Canada and the US to Mexico.

When the PCA in the HO breed was explored, a difference in the genetic structure was observed between the Mexican production

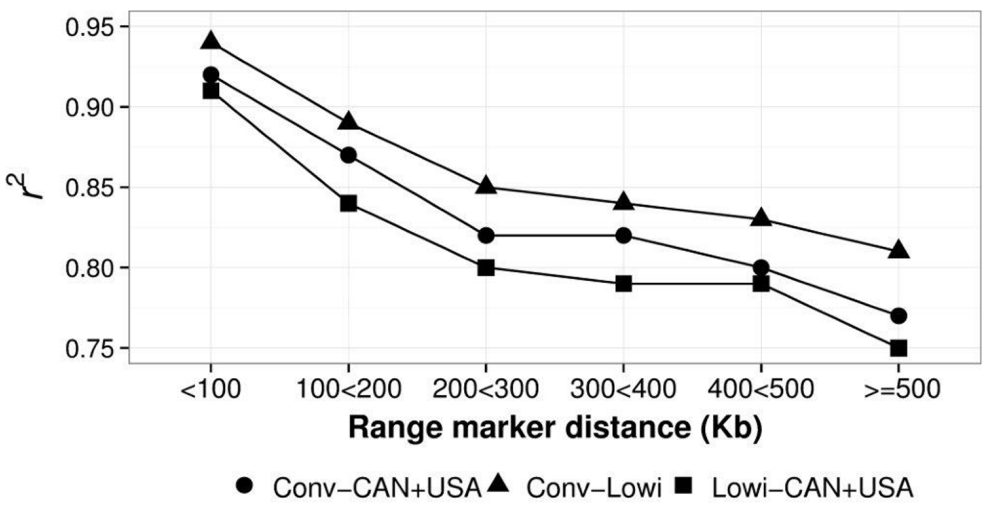


FIGURE 6 | Persistence of Linkage Disequilibrium phase between North America (CAN + USA) and Mexican Holstein Populations identified as conventional (Conv) and low income (Lowi) at different genetic distances.

systems and the CAN + USA population, although population overlap suggests that all three groups share common genetic material. In this case the Conv system seems to be a link between the Lowi system and CAN + USA population because breeders of the Conv system provide genetic material (heifers and semen) to the Lowi one and the Conv system has depended genetically from US and Canada for many years (Powell and Wiggans, 1991). The Lowi also obtains genetic material directly from CAN + USA, but at a lower proportion than the Conv one. No specific tendency to group animals for geographic area was found in the Mexican populations, but as in other studies (Novembre and Stephens, 2008), these analyses give us an idea of the genetic background of the population.

The inclusion of JE and BS resulted in individual clusters for both breeds as expected, but more importantly, they helped define the differences between the Lowi and the Conv populations. Results suggest that breeders of the Lowi system occasionally used genetic material of other breeds.

Admixture analysis showed that five different populations, linked to origin country of the sires, comprise the ancestral background of the Mexican HO populations. The two systems showed variation in the average proportion of genetic similarity to the different ancestral populations. The results show a substantial influence of the North American HOs on the Mexican population, and agree with previous results in the Conv system reported by Powell and Wiggans (1991). JE and BS breed influence is visible in both Mexican production systems, but stands out in the Lowi where the use of these dairy breeds seems to be more common. The MEX population was linked through pedigree information to registered Mexican HO bulls related to North American animals. This block also included individuals with unknown sires presumably with similar origins to those of the Mexican sires.

In line with previous studies, decay of LD was observed in this study when the distance between markers was increased (**Figure 5**; Du et al., 2007; de Roos et al., 2008; Sargolzaei et al., 2008; Badke et al., 2012). In general, average r^2 estimates for the populations in this study were in the range of those reported in other HO populations (McKay et al., 2007; de Roos et al., 2008; Sargolzaei et al., 2008; Qanbari et al., 2010; Zhou et al., 2013), although r^2 averages for the CAN + USA are slightly higher than those reported for the same population in a group of animals born after 1990 at similar distances between markers (Sargolzaei et al., 2008), except when the distance is less than 100 Kb. Note than the LD of the CAN + USA were higher than the Mexican systems at all compared genetic distances and those of the CONV are slightly higher than those of Lowi. The lowest r^2 values among the populations were for the Lowi. These lower values may be the result of breeders in the Lowi system introducing other breeds through crossbreeding and incorporating gene migration and genetic drift. This practice may explain the reduction in LD (Lynch and Walsh, 1998). The r^2 averages for the Lowi decreased rapidly when the distance increased from $<$100 to 100 and $<$200 Kb then decreased slowly at distances $>$200 Kb.

A practical application of LD is determining the number of markers necessary to perform genome wide associations studies (Gautier et al., 2007; McKay et al., 2007; de Roos et al., 2008) at a useful LD ($r^2 \geq 0.20$; de Roos et al., 2008). The useful LD of 54.5 kb found in this study, suggests that at least 53,000 markers should be used to perform genomic analysis within this population. Similar numbers of markers were suggested for other cattle populations (McKay et al., 2007; de Roos et al., 2008).

Because the persistence phase or correlation of r among populations show the genetic relationship between them (Badke et al., 2012), it was used as a measure among the HO populations included in this analysis. Results shown in **Figure 6** confirm the results of PCA, because the higher persistence phase was reported between the Conv and Lowi, followed by the one between Conv and CAN + USA ending with confirmation of the lowest relationship between Lowi and CAN + USA. As it was also reported in other studies (de Roos et al., 2008; Badke et al., 2012) the correlation of r among all populations decreased rapidly when the distance between markers increased. The difference between persistence phases among the Conv, Lowi, and CAN + USA ranged from 0.01 to 0.04, lower than that present in other breeds like Angus, Charolais, and JE (de Roos et al., 2008; Lu et al., 2012) or species like pigs (Badke et al., 2012). At distances $<$100 Kb, the persistence phase between the Conv and Lowi, Conv and CAN + USA, and CAN + USA and Lowi were lower than that reported between Chinese and Nordic HO cattle (0.97; Zhou et al., 2013) and at all measured intervals, similar values were found between Dutch black and white and Dutch red and white HO Friesian bulls and lower values were reported for Australian bulls and New Zealand Friesian cows (de Roos et al., 2008).

Results showed that the US and Canadian and the Mexican HO cattle of the Conv and Lowi have different genetic structures although these populations share much common ancestry. The main difference between the Mexican HO systems is the result of crossbreeding with other breeds, especially in the Lowi system. If joint genomic studies are to be performed between these populations, stratification of populations is recommended. Joint genetic improvement programs of HO animals across North America, i.e., including Mexico, may be established as these populations share genetic material. The useful LD founded in this populations, will determine the minimum number of SNP markers need if joint genomic studies are to be performed.

The considerable similarity between the Conv subgroup with US and Canadian populations means that integration of these groups would be straightforward and should be considered.

AUTHOR CONTRIBUTIONS

The authors have made the following declarations about their contributions: Conceived and designed the experiments Adriana García-Ruiz, Felipe de J. Ruiz-López, Curtis P. Van Tassell, and Hugo H. Montaldo. Performed experiments and analyze data: Adriana García-Ruiz Data acquisition and interpretation: Adriana García-Ruiz, Felipe de J. Ruiz-López, Curtis P. Van Tassell, Hugo H. Montaldo, and Heather J. Huson. Wrote the paper: Adriana García-Ruiz. All authors approve the manuscript final version.

ACKNOWLEDGMENTS

We would like to thank the Mexican HO Association for providing samples of the animals included in the analysis, the Council on

Dairy Cattle Breeding (*CDCB*) for providing material included in the analysis and Dr. George Wiggans, for his contribution and the edition of databases.

This study was supported by CONACYT, CONARGEN and the research projects: Study of Genetic Diversity of Mexican HO Cattle based on Genomic Information (SIGI: 1523542158) and Incorporation of Genomic information in the Genetic Evaluation Process of Mexican Dairy Cattle (SIGI: 1056821832).

REFERENCES

Alexander, D. H., Novembre, J., and Lange, K. (2009). Fast Model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19, 1655–1664. doi: 10.1101/gr.094052.109

Amendola, R. D. (2002). *A Dairy System Based on Forages, and Grazing in Temperate Mexico*. Ph.D. thesis, Wageningen University, Wageningen.

Badke, Y. M., Bates, R. O., Ernst, C. W., Schwab, C., and Steibel, J. P. (2012). Estimation of linkage disequilibrium in four US pig breeds. *BMC Genomics* 13:24. doi: 10.1186/1471-2164-13-24

Boichard, D., Maigner, L., and Verrier, É. (1997). The value of using probabilities of gene origin to measure genetic variability in a population. *Genet. Sel. Evol.* 29, 5–23. doi: 10.1051/gse:19970101

Bovine HapMap Consortium. (2009). Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Science* 324, 528–532. doi: 10.1126/science.1167936

Buerkle, C. A., and Lexer, C. (2008). Admixture as the basis for genetic mapping. *Trends Ecol. Evol.* 23, 686–694. doi: 10.1016/j.tree.2008.07.008

Bulmer, M. G. (1971). The effect of selection on genetic variability. *Am. Nat.* 105, 201–211. doi: 10.1086/282718

Cavalli-Sforza, L. L., Menozzi, P., and Piazza, A. (1994). *The History, and Geography of Human Genes*. Princeton, NJ: Princeton University Press. doi: 10.1086/282718

Dekkers, J. C., and Hospital, F. (2002). The use of molecular genetics in the improvement of agricultural populations. *Nat. Rev. Genet.* 3, 22–32. doi: 10.1038/nrg701

de Roos, A. P. W., Hayes, B. J., Spelman, R. J., and Goddard, M. E. (2008). Linkage disequilibrium and persistence of phase in Holstein-Friesian, Jersey and Angus cattle. *Genetics* 179, 1503–1512. doi: 10.1534/genetics.107.084301

Du, F. X., Clutter, A. C., and Lohuis, M. M. (2007). Characterizing linkage disequilibrium in pig populations. *Int. J. Biol. Sci.* 3, 166–178. doi: 10.7150/ijbs.3.166

Frkonja, A., Gredler, B., Schnyder, U., Curik, I., and Sölkner, J. (2012). Prediction of breed composition in an admixed cattle population. *Anim. Genet.* 43, 696–703. doi: 10.1111/j.1365-2052.2012.02345.x

Gautier, M., Faraut, T., Moazami-Goudarzi, K., Navratil, V., Foglio, M., Grohs, C., et al. (2007). Genetic and haplotypic structure in 14 European and African cattle breeds. *Genetics* 177, 1059–1070. doi: 10.1534/genetics.107.075804

Hill, W. G., and Robertson, A. (1968). Linkage disequilibrium in finite populations. *Theor. Appl. Genet.* 38, 226–231. doi: 10.1111/j.1365-2052.2012.02345.x

Lewis, J., Abas, Z., Dadousis, C., Lykidis, D., Paschou, P., and Drineas, P. (2011). Tracing cattle breeds with principal components analysis ancestry informative SNPs. *PLoS ONE* 6:e18007. doi: 10.1371/journal.pone.0018007

Lu, D., Sargolzaei, M., Kelly, M., Li, C., Vander Voort, G., Wang, Z., et al. (2012). Linkage disequilibrium in Angus, Charolais, and Crossbred beef cattle. *Front. Genet.* 3:152. doi: 10.3389/fgene.2012.00152

Lynch, M., and Walsh, B. (1998). *Genetics and Analysis of Quantitative Traits*. Sunderland, MA: Sinauer Associates.

McKay, S. D., Schnabel, R. D., Murdoch, B. M., Matukumalli, L. K., Aerts, J., Coppieters, W., et al. (2007). Whole genome linkage disequilibrium maps in cattle. *BMC Genet.* 8:74. doi: 10.1186/1471-2156-8-74

Norman, H. D., and Powell, R. L. (1999). Dairy cows of high genetic merit for yields of milk, fat and protein. *Asian Australas. J. Anim. Sci.* 12, 1316–1323. doi: 10.1186/1471-2156-8-74

Novembre, J., and Stephens, M. (2008). Interpreting principal component analyses of spatial population genetic variation. *Nat. Genet.* 40, 646–649. doi: 10.1038/ng.139

Powell, R. L., and Wiggans, G. R. (1991). Animal model evaluations for Mexican Holsteins. *J. Dairy Sci.* 74, 1420–1427. doi: 10.3168/jds.S0022-0302(91)78298-2

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* 38, 904–909. doi: 10.1038/ng1847

Qanbari, S., Pimentel, E. C. G., Tetens, J., Thaller, G., Lichtner, P., Sharifi, A. R., et al. (2010). A genome-wide scan for signatures of recent selection in Holstein cattle. *Anim. Genet.* 41, 377–389. doi: 10.1111/j.1365-2052.2009.02016.x

Sargolzaei, M., Schenkel, F. S., Jansen, G. B., and Schaeffer, L. R. (2008). Extent of linkage disequilibrium in Holstein cattle in North America. *J. Dairy Sci.* 91, 2106–2117. doi: 10.3168/jds.2007-0553

SAS Institute. (2009). *SAS Online Doc 9.2. SAS Institute Inc., Cary, NC (Online)*. Available at: http://support.sas.com/documentation/cdl_main/index.html

Silva-Zolezzi, I., Hidalgo-Miranda, A., Estrada-Gil, J., Fernandez-Lopez, J. C., Uribe-Figueroa, L., Contreras, A., et al. (2009). Analysis of genomic diversity in Mexican Mestizo populations to develop genomic medicine in Mexico. *Proc. Natl. Acad. Sci. U.S.A.* 106, 8611–8616. doi: 10.1073/pnas.0903045106

SNP and Variation Suite Manual v7. (2013). *Golden Helix, Inc., Bozeman, MT, (Online)*. Available at: www.goldenhelix.com

Sölkner, J., Frkonja, A., Raadsma, H. W., Jonas, E., Thaller, G., Gootwine, E., et al. (2010). Estimation of individual levels of admixture in crossbred populations from SNP chip data: examples with sheep and cattle populations. *Interbull Bull.* 42, 62–66.

Tang, H., Peng, J., Wang, P., and Risch, N. J. (2005). Estimation of individual admixture: analytical and study design considerations. *Genet. Epidemiol.* 28, 289–301. doi: 10.1002/gepi.20064

Wang, W. Y., Barrat, B. J., Clayton D. G., and Todd. J. A. (2005). Genome-wide association studies: theoretical and practical concerns. *Nat. Rev. Genet.* 6, 109–118. doi: 10.1038/nrg1522

Waples, R. S., and England, P. R. (2011). Estimating contemporary effective population size on the basis of linkage disequilibrium in the face of migration. *Genetics* 189, 633–44. doi: 10.1534/genetics.111.132233

Winkler, A. C., Nelson, W. G., and Smith, W. M. (2010). Admixture mapping comes of age. *Annu. Rev. Genomics Hum. Genet.* 11, 65–89. doi: 10.1146/annurev-genom-082509-141523

Zhou, L., Ding, X., Zhang, Q., Wang, Y., Lund, M. S., and Su, G. (2013). Consistency of linkage disequilibrium between Chinese and Nordic Holsteins and genomic prediction for Chinese Holsteins using a joint reference population. *Genet. Sel. Evol.* 45:7. doi: 10.1186/1297-9686-45-7

Conflict of Interest Statement: The Review Editor Ikhide G. Imumorin declares that, despite being affiliated with the same institute as the author Heather J. Huson, the review process was handled objectively. The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 17 August 2014; accepted: 09 January 2015; published online: 09 February 2015.

*Citation: García-Ruiz A, Ruiz-López FJ, Van Tassell CP, Montaldo HH and Huson HJ (2015) Genetic differentiation of Mexican Holstein cattle and its relationship with Canadian and U.S. Holsteins. Front. Genet. **6**:7. doi: 10.3389/fgene.2015.00007*

This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics.

Copyright © 2015 García-Ruiz, Ruiz-López, Van Tassell, Montaldo and Huson. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in GENETICS

ORIGINAL RESEARCH ARTICLE
published: 29 January 2015
doi: 10.3389/fgene.2015.00005

Assessment of autozygosity in Nellore cows (*Bos indicus*) through high-density SNP genotypes

Ludmilla B. Zavarez[1], Yuri T. Utsunomiya[1], Adriana S. Carmo[1], Haroldo H. R. Neves[2,3], Roberto Carvalheiro[3], Maja Ferenčaković[4], Ana M. Pérez O'Brien[5], Ino Curik[4], John B. Cole[6], Curtis P. Van Tassell[6], Marcos V. G. B. da Silva[7], Tad S. Sonstegard[6], Johann Sölkner[5] and José F. Garcia[1,8]*

[1] *Departamento de Medicina Veterinária Preventiva e Reprodução Animal, Faculdade de Ciências Agrárias e Veterinárias, UNESP – Univ Estadual Paulista, Jaboticabal, São Paulo, Brazil*
[2] *GenSys Consultores Associados, Porto Alegre, Rio Grande do Sul, Brazil*
[3] *Departamento de Zootecnia, Faculdade de Ciências Agrárias e Veterinárias, UNESP – Univ Estadual Paulista, Jaboticabal, São Paulo, Brazil*
[4] *Department of Animal Science, Faculty of Agriculture, University of Zagreb, Zagreb, Croatia*
[5] *Division of Livestock Sciences, Department of Sustainable Agricultural Systems, BOKU - University of Natural Resources and Life Sciences, Vienna, Austria*
[6] *Animal Genomics and Improvement Laboratory, United States Department of Agriculture, Agricultural Research Service, Beltsville, MD, USA*
[7] *Bioinformatics and Animal Genomics Laboratory, Embrapa Dairy Cattle, Juiz de Fora, Minas Gerais, Brazil*
[8] *Laboratório de Bioquímica e Biologia Molecular Animal, Departamento de Apoio, Produção e Saúde Animal, Faculdade de Medicina Veterinária de Araçatuba, UNESP – Univ Estadual Paulista, Araçatuba, São Paulo, Brazil*

Edited by:
Paolo Ajmone Marsan, Università Cattolica del S. Cuore, Italy

Reviewed by:
Yniv Palti, United States Department of Agriculture, USA
Nicolo Pietro Paolo Macciotta, University of Sassari, Italy

***Correspondence:**
José F. Garcia, Laboratório de Bioquímica e Biologia Molecular Animal, Departamento de Apoio, Produção e Saúde Animal, Faculdade de Ciências Agrárias e Veterinárias, UNESP – Univ Estadual Paulista, Rua Clóvis Pestana 793, Araçatuba, São Paulo, 16050-680, Brazil
e-mail: jfgarcia@fmva.unesp.br

The use of relatively low numbers of sires in cattle breeding programs, particularly on those for carcass and weight traits in Nellore beef cattle (*Bos indicus*) in Brazil, has always raised concerns about inbreeding, which affects conservation of genetic resources and sustainability of this breed. Here, we investigated the distribution of autozygosity levels based on runs of homozygosity (ROH) in a sample of 1,278 Nellore cows, genotyped for over 777,000 SNPs. We found ROH segments larger than 10 Mb in over 70% of the samples, representing signatures most likely related to the recent massive use of few sires. However, the average genome coverage by ROH (>1 Mb) was lower than previously reported for other cattle breeds (4.58%). In spite of 99.98% of the SNPs being included within a ROH in at least one individual, only 19.37% of the markers were encompassed by common ROH, suggesting that the ongoing selection for weight, carcass and reproductive traits in this population is too recent to have produced selection signatures in the form of ROH. Three short-range highly prevalent ROH autosomal hotspots (occurring in over 50% of the samples) were observed, indicating candidate regions most likely under selection since before the foundation of Brazilian Nellore cattle. The putative signatures of selection on chromosomes 4, 7, and 12 may be involved in resistance to infectious diseases and fertility, and should be subject of future investigation.

Keywords: *Bos indicus*, runs of homozygosity, selection, cattle, fertility, disease resistance

INTRODUCTION

Autozygosity is the homozygote state of identical-by-descent alleles, which can result from several different phenomena such as genetic drift, population bottleneck, mating of close relatives, and natural and artificial selection (Falconer and Mackay, 1996; Keller et al., 2011; Curik et al., 2014). In the past 20 years, the heavy use of relatively low number of sires in Brazilian Nellore breeding programs (*Bos indicus*) is deemed to have mimicked all these triggers of autozygosity, especially considering the increasing use of artificial insemination over the decades. As inbreeding has been incriminated in reduced fitness and reproductive performance in other cattle populations under artificial selection (Bjelland et al., 2013; Leroy, 2014), avoidance of mating of close relatives is a typical practice of many Nellore breeders. Therefore, there is a growing interest in characterizing and monitoring autozygosity in this breed to preserve genetic diversity and allow the long-term sustainability of breeding programs in Brazil.

Evidence from whole-genome sequencing studies in humans indicate that highly deleterious variants are common across healthy individuals (MacArthur et al., 2012; Xue et al., 2012), and although no such systematical survey has been conducted in cattle to the present date, it is highly expected that unfavorable alleles also segregate in cattle populations. Therefore, the use of ever-smaller numbers of animals as founders is expected to inadvertently increase autozygosity of such unfavorable alleles (Szpiech et al., 2013), potentially causing economic losses.

Recently, the use of high-density single nucleotide polymorphism (SNP) genotypes to scan individual genomes for contiguous homozygous chromosomal fragments has been proposed as a proxy for the identification of identical-by-descent haplotypes (Gibson et al., 2006; Lencz et al., 2007). As the length

of autozygous chromosomal segments is proportional to the number of generations since the common ancestor (Howrigan et al., 2011), the identification of runs of homozygosity (ROH) can reveal recent and remote events of inbreeding, providing invaluable information about the genetic relationships and demographic history of domesticated cattle (Purfield et al., 2012; Ferenčaković et al., 2013a; Kim et al., 2013). Also, given the stochastic nature of recombination, the occurrence of ROH is highly heterogeneous across the genome, and hotspots of ROH across a large number of samples (hereafter referred as common ROH) may be indicative of selective pressure. Moreover, the fraction of an individual's genome covered by ROH can be used as an estimate of its genomic autozygosity or inbreeding coefficient (McQuillan et al., 2008; Curik et al., 2014).

Here, we investigated the occurrence of ROH in high-density SNP genotypes in order to characterize autozygosity in the genomes of a sample of 1,278 Nellore cows under artificial selection for weight, carcass and reproductive traits. We aimed at characterizing the distribution of ROH length and genome-wide levels of autozygosity, as well as detecting common ROH that may be implicated in past events of selection.

MATERIALS AND METHODS

ETHICAL STATEMENT

The present study was exempt of the local ethical committee evaluation as genomic DNA was extracted from stored hair samples of animals from commercial herds.

GENOTYPING AND DATA FILTERING

A total of 1,278 cows were genotyped with the Illumina® BovineHD Genotyping BeadChip assay (HD), according to the manufacturer's protocol (http://support.illumina.com/array/array_kits/bovinehd_dna_analysis_kit.html). These animals comprised part of the genomic selection reference population from a commercial breeding program. These dams were born between 1993 and 2008, being under routine genetic evaluation for weight, carcass and reproductive traits by the DeltaGen program, an alliance of Nellore cattle breeders from Brazil. Data filtering was performed using *PLINK v1.07* (Purcell et al., 2007), and markers were removed from the dataset if GenTrain score lower than 70% or a call rate lower than 98% was observed. All genotyped samples exhibited call rates greater than 90%, thus no animals were filtered from further analyses. Minor allele frequency (MAF) was not used as an exclusion criterion in this analysis, so that the detection of homozygous segments was not compromised. Both autosomal and X-linked markers were included.

ESTIMATES OF GENOMIC INDIVIDUAL AUTOZYGOSITY

Genomic autozygosity was measured based on the percentage of an individual's genome that is covered by ROH. Stretches of consecutive homozygous genotypes were identified for each animal using *SNP & Variation Suite v7.6.8* (Golden Helix, Bozeman, MT, USA http://www.goldenhelix.com), and chromosomal segments were declared ROH under the following criteria: 30 or more consecutive homozygous SNPs, a density of at least 1 SNP every 100 kb, gaps of no more than 500 kb between SNPs, and no more than 5 missing genotypes across all individuals. In order to account for genotyping error and avoid underestimation of long ROH (Ferenčaković et al., 2013b), heterozygous genotype calls were allowed under conditions where there were 2 heterozygous genotypes for ROH $\geq$ 4 Mb, or no heterozygous genotypes for ROH $<$ 4 Mb. Autozygosity was estimated according to McQuillan et al. (2008):

$$F_{ROH} = \frac{\sum_{j=1}^{n} L_{ROH_j}}{L_{total}}$$

Where L_{ROH_j} is the length of ROH j, and L_{total} is the total size of the genome covered by markers, calculated from the sum of inter-marker distances in the UMD v3.1 assembly. In order to facilitate comparisons with other studies, F_{ROH} was calculated using both the genome size based on autosomal and autosomal + X chromosomes. For each animal, F_{ROH} was calculated based on ROH of different minimum lengths: 0.5, 1, 2, 4, 8 or 16 Mb, representing autozygosity events that occurred approximately 100, 50, 25, 13, 6, and 3 generations in the past, respectively (Howrigan et al., 2011; Ferenčaković et al., 2013b). Additionally, chromosome-wise F_{ROH} was also computed.

An alternative measure of autozygosity was obtained by computing the diagonal elements of a modified realized genomic relationship matrix (VanRaden, 2008; VanRaden et al., 2011), calculated as:

$$G = \frac{ZZ'}{2\sum_{l=1}^{n} p_l(1-p_l)}$$

Where Z is a centered genotype matrix and p_l is the reference allele frequency at locus l. Matrix Z is obtained by subtracting from the genotype matrix M (with genotype scores coded as 0, 1 or 2 for alternative allele homozygote, heterozygote, and reference allele homozygote, respectively) the matrix P, whose elements of column l are equal to $2p_l$. The diagonal elements of G ($G_{i,i}$) represent the relationship of an animal with itself, and thus encapsulate autozygosity information. Following VanRaden et al. (2011), $G_{i,i}$ can provide a more suitable proxy for the pedigree-based inbreeding coefficient when assuming $p_l = 0.5$, rather than using base population allele frequencies estimates (which could be difficult to estimate especially in absence of complete pedigree data). Thus, matrix G was computed using allele frequencies fixed at 0.5.

DETECTION OF COMMON RUNS OF HOMOZYGOSITY

Chromosomal segments presenting ROH hotspots were defined as ROH islands or common ROH. In order to identify such genomic regions, we used two different strategies. First, we used the clustering algorithm implemented in *SNP & Variation Suite v7.6.8*, which identifies clusters of contiguous set of SNPs with size $> s_{min}$, where every SNP has at least n_{min} samples presenting a run. Clusters were identified based on a fixed minimum cluster size of $s_{min} = 0.5$ Mb for varying minimum number of samples: 127 (10%), 255 (20%), 319 (25%), and 639 (50%). In order to assess the sensitivity of the algorithm to parameter settings in

ROH detection, we repeated the analysis using minimum numbers of 30 or 150 SNPs in a run, maximum gap sizes of 100 kb or 500 kb, and 0 or 2 heterozygous genotypes as variable parameters.

Alternatively, we calculated locus autozygosity (F_L) following Kim et al. (2013). Briefly, for each SNP, animals were scored as autozygous (1) or non-autozygous (0) based on the presence of a ROH encompassing the SNP. Then, the locus autozygosity was simply computed as:

$$F_L = \frac{\sum_{i=1}^{n} S_i}{n}$$

where S_i is the autozygosity score of individual i, and n is the number of individuals. In essence, F_L represents the proportion of animals with scores equal to 1 (i.e., that present a ROH enclosing the marker), thus it summarizes the level of local autozygosity in the sample.

RESULTS AND DISCUSSION

DISTRIBUTION OF ROH LENGTH

After filtering, 668,589 SNP marker genotypes across 1,278 animals were retained for analyses. The average, median, minimum and maximum ROH length detected across all chromosomes were 1.26, 0.70, 0.50, and 70.91 Mb, respectively, suggesting this specific Nellore cattle population experienced both recent and remote autozygosity events. Segments as large as 10 Mb are traceable to inbreeding that occurred within the last five generations (Howrigan et al., 2011), and a total of 942 samples (73.7%) presented at least one homozygous fragment larger than 10 Mb. Therefore, it is likely that these long ROH are signatures of the extended use of recent popular sires.

DISTRIBUTION OF GENOME-WIDE AUTOZYGOSITY

The distributions of $G_{i,i}$ and F_{ROH} based on autosomal ROH of different minimum lengths (>0.5, >1, >2, >4, >8 or >16 Mb) are shown in **Figure 1**. Although the inclusion of the X chromosome did not cause substantial differences in the calculation of genome-wide F_{ROH} (Supplementary Figure S1), we focused on the estimates using only autosomes for the ease of comparison with other studies. The skewness of the autosomal F_{ROH} distribution increased as the minimum fragment length increased, ranging from 1.56 for $F_{ROH>0.5Mb}$ to 3.98 for $F_{ROH>16Mb}$. The number of animals with $F_{ROH} = 0$ also increased as the minimum ROH length increased, starting at 12 (0.94%) for $F_{ROH>2Mb}$ and increasing to 827 (64.71%) for $F_{ROH>16Mb}$. Under the assumption of the relationship between ROH length and age of autozygosity, these findings show that varying the minimum ROH length in the calculation of F_{ROH} can be useful to discriminate animals with recent and remote autozygosity.

As shown in **Figure 2**, the correlation between autosomal $F_{ROH>1Mb}$ and $G_{i,i}$ ($r = 0.69$) was close to the ones reported by Ferenčaković et al. (2013b) for the comparison between $F_{ROH>1Mb}$ derived from the HD panel and pedigree estimates in Brown Swiss ($r = 0.61$), Pinzgauer ($r = 0.62$), and Tyrol Gray ($r = 0.75$). Similar correlations were observed when the X chromosome was included in the analysis (Supplementary Figure S2). McQuillan et al. (2008) also reported correlations between F_{ROH} and pedigree estimates in human European populations ranging from 0.74 to 0.82. Considering that VanRaden (2008) proposed G as a proxy for a numerator relationship matrix obtained from highly reliable and recursive pedigree data, we expect that the correlations found for $G_{i,i}$ are fair approximations to the ones we would have found if complete pedigree data was available.

In the present study, correlations between F_{ROH} and $G_{i,i}$ decreased as a function of different ROH length (**Figure 2**). This may be due to the properties of the G matrix, which is based on individual loci, whereas F_{ROH} is based on chromosomal segments. Ferenčaković et al. (2013b) showed that medium density SNP panels, such as the Illumina® BovineSNP50, systematically overestimate F_{ROH} when segments shorter than 4 Mb are included in the calculations, while the Illumina® BovineHD panel is robust for the detection of shorter segments. Hence, although the inclusion of short length ROH in the calculation of F_{ROH} may be desirable for autozygosity estimates accounting for remote inbreeding, there is a compromise between SNP density, minimum ROH length and false discovery of ROH. Since the HD panel allows for the detection of short ROH, in this section we focused on the results obtained with $F_{ROH>1Mb}$ as it presented the second highest correlation with $G_{i,i}$ and is comparable with previous studies.

The minimum, average, median, and maximum autosomal $F_{ROH>1Mb}$ across all animals were 0.43, 4.79, 4.58, and 18.55%, respectively. The animal presenting the highest autozygosity value (18.55%) exhibited 69 ROH > 1 Mb encompassing 465.66 Mb of the total autosomal genome extension covered by markers (2.51 Gb), with a mean ROH length of 6.75 ± 9.20 Mb, and a maximum segment length of 43.79 Mb. The least inbred animal presented 8 ROH > 1 Mb, summing up only 10.72 Mb, with an average length of 1.34 ± 0.46 Mb and a maximum of 2.43 Mb.

The coefficient of variation (here denoted as the ratio of the standard deviation to the mean) of the $F_{ROH>1Mb}$ distribution was 37.5%, indicating moderate variability in autozygosity levels in this sample. In spite of the average genome coverage by ROH of 4.58% may seem to indicate moderate inbreeding levels for classical standards, it has to be considered that incomplete pedigree data usually fails to capture remote inbreeding, so that traditional inbreeding estimates based on pedigree are only comparable with F_{ROH} calculated over large ROH lengths, which in the present study were close to 0%.

Compared to other cattle populations, this sample of Nellore cows presented a lower average autozygosity. For instance, Ferenčaković et al. (2013b) reported average autosomal $F_{ROH>1Mb}$ of 15.1%, 6.2%, and 6.6% for samples of the *Bos taurus* breeds Brown Swiss, Pinzgauer, and Tyrol Gray, respectively. Also, the effective population size estimated for this Nellore sample was approximately 362 animals (Supplementary Material), which is consistent with the low genome average LD reported by other studies (McKay et al., 2007; Espigolan et al., 2013; Pérez O'Brien et al., 2014) and indicative of a non-inbred population.

DISTRIBUTION OF CHROMOSOME-WISE AUTOZYGOSITY

The averages of the chromosome-wise $F_{ROH>0.5Mb}$ across samples are shown in **Figure 3**. Chromosome X exhibited a substantially higher average autozygosity when compared to the autosomes. Importantly, we found no evidence for a smaller effective population size for the X chromosome in comparison

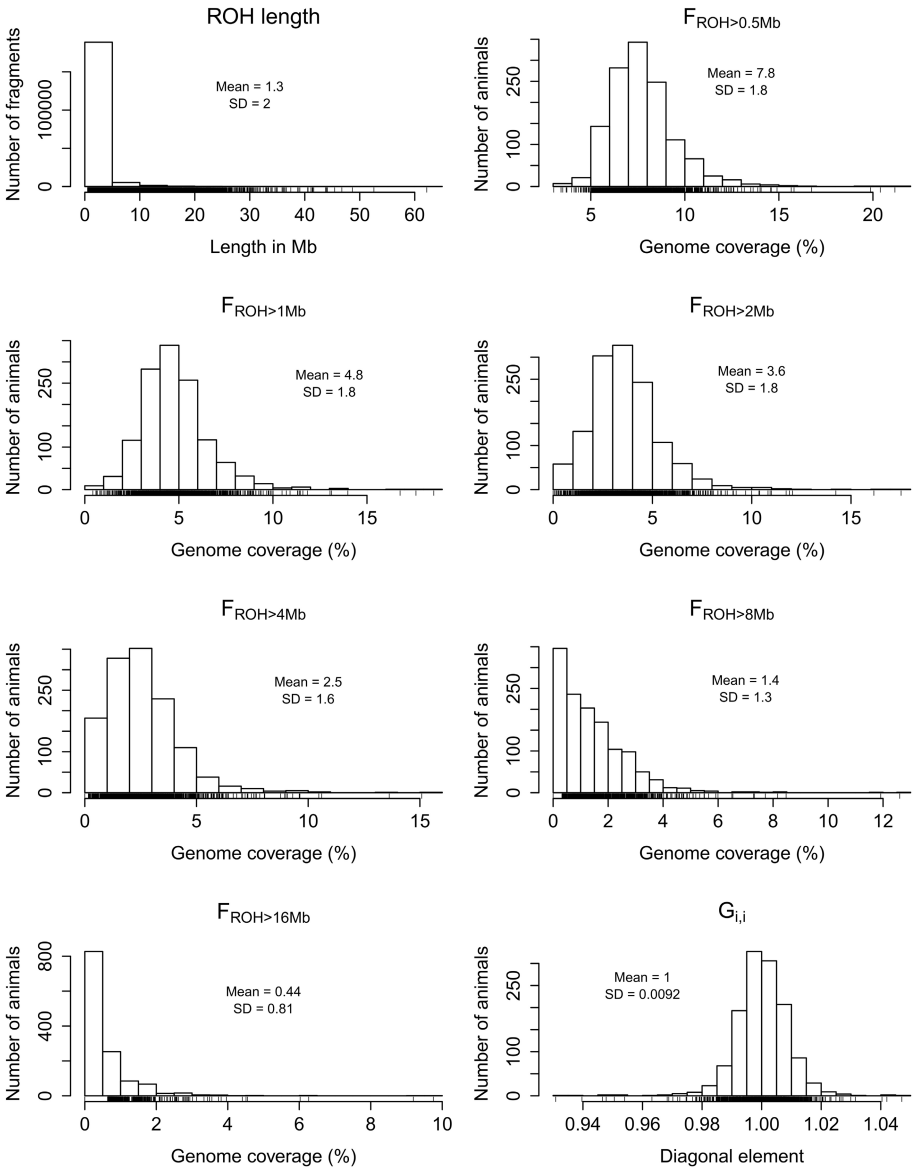


FIGURE 1 | Frequency distributions of all detected runs of homozygosity (ROH) across samples, percentage of autosomal genome coverage by ROH (F_{ROH}) of different minimum lengths (>0.5, >1, >2, >4, >8, and >16 Mb), and diagonal elements of the realized genomic relationship matrix ($G_{i,i}$).

to the autosomal genome (Supplementary Material). This may be due to the mode of inheritance of the X chromosome, which is hemizygous in the male lineage and therefore more susceptible to bottlenecks and drift even under assumptions of balanced numbers of males and females (Gottipati et al., 2011).

An alternative explanation is that the gene content and the sex-specific copy number of the X chromosome is under stronger selective pressure in comparison to autosomal DNA (Hammer et al., 2010; Deng et al., 2014). In both hypotheses, this higher autozygosity may reflect historical and demographical events. In the early 20th century, when more frequent importation of Nellore cattle to Brazil was initiated, the indigenous herds mainly consisted of descendants from taurine (*Bos taurus*) cattle imported since the late 15th century after the discovery of America (Ajmone-Marsan et al., 2010). In spite of the use of taurine dams for breeding during the early establishment of Nellore cattle in Brazil, the decades that followed were marked by intense backcrossing to Nellore bulls, causing most of the taurine contribution to be swept out from the Nellore autosomal genome (Utsunomiya et al., 2014). However, it is well-established that taurine mitochondrial DNA is prevalent in Nellore cattle, as it is a strict maternal contribution (Meirelles et al., 1999). Therefore, the X chromosome may have experienced a greater drift than the autosomal genome due to limited number of founders. The levels

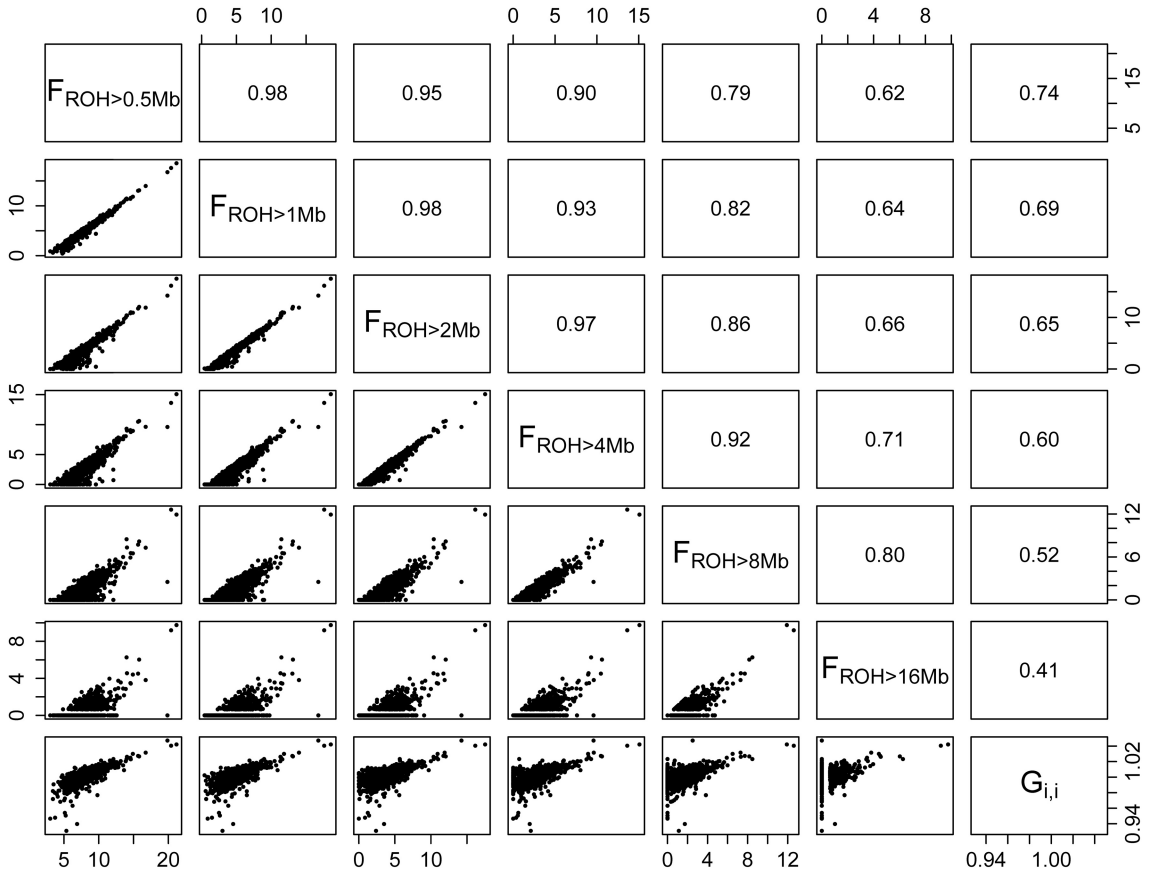


FIGURE 2 | Scatterplots (lower panel) and correlations (upper panel) of percentage of autosomal genome coverage by runs of homozygosity (F_{ROH}) of different minimum lengths (>0.5, >1, >2, >4, >8, and >16 Mb) and diagonal elements of the realized genomic relationship matrix ($G_{i,i}$). The last column of panels on the right indicates that the correlation between F_{ROH} and $G_{i,i}$ decreases as a function of minimum fragment size.

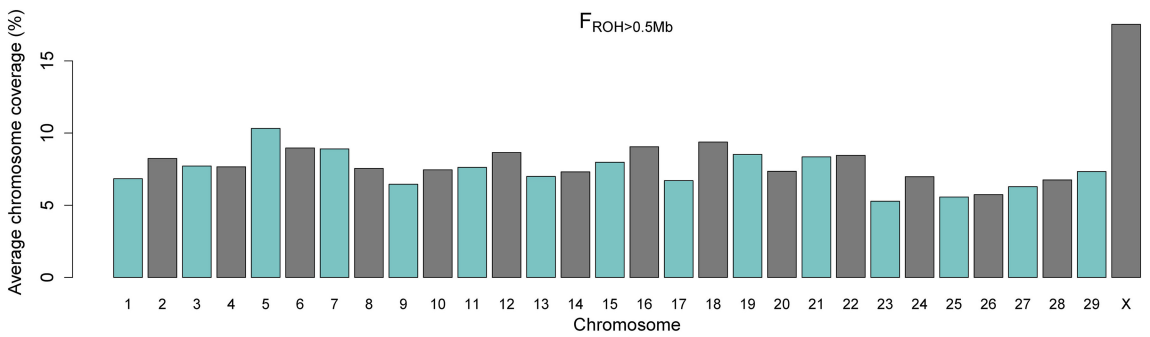


FIGURE 3 | Barplot of average percentage of chromosome coverage by runs of homozygosity (F_{ROH}) of minimum length of 0.5 Mb.

of taurine introgression still segregating in the X chromosome in this herd remain unclear.

IDENTIFICATION OF COMMON ROH

Table 1 presents the results obtained from the ROH clustering analysis. The algorithm was robust in respect to gap size between SNPs, but substantial differences were observed when the number of consecutive SNPs and the number of heterozygous genotypes were modified. Few common ROH were identified even when the minimum number of samples in the cluster was 10%, indicating that ROH distribution is not uniform across the genome. In fact, despite of the occurrence of 99.98% of the SNPs within a ROH of at least one individual, only 19.37% markers were encompassed by ROH observed in 10% or more of the samples. This finding is similar to that reported by Ferenčaković et al. (2013b), and is consistent with the stochastic nature of meiotic recombination. This

suggests that the ongoing selection for weight, carcass and reproductive traits in this population has not yet created detectable ROH-based selection signatures related to production.

The calculations of locus autozygosity were consistent with the cluster analysis using 150 SNPs and 2 heterozygous genotypes, regardless of permitted gap size (**Figure 4**). Seven distinct genomic regions, four of them on chromosome X, presented strong hotspots of autozygosity, where over half of the samples ($n = 639$) contained a ROH. The common ROH on the X chromosome are difficult to be discussed as they span several millions of bases, encompassing hundreds of genes and making functional explorations unfeasible. Besides, the assembly status of X chromosome is poorer than the autosomal ones. Hence, we focused on the three autosomal regions on chromosomes 4, 7, and 12. The three regions were relatively short, ranging from 0.73 to 1.43 Mb. For this range of ROH length, the expected number of generations since the common ancestor is estimated between 35 and 69 (Howrigan et al., 2011). Assuming a cattle generation interval of 5 years, these inbreeding events may have occurred between 175 and 345 years ago. Although this estimate does not account for birth date and overlapping generations, these remote autozygosity events are likely to predate the foundation of the Nellore breeding programs, and therefore expected to be related to natural selection, random drift or population bottlenecks.

The most autozygous locus was found at chromosome 7:51605639-53035752. This region was previously reported in genome-wide scans for signatures of selection in cattle through the comparison of *Bos taurus* and *indicus* breeds via F_{ST} analysis (Bovine HapMap Consortium, 2009; Porto-Neto et al., 2013) and was detected as a ROH hotspot in an analysis of three taurine and indicine breeds each (Sölkner et al., 2014). This region has been implicated in the control of parasitemia in cattle infected by *Trypanosoma congolense* (Hanotte et al., 2003), and is orthologous to the human chromosome segment 5q31-q33, known as the Th2 cytokine gene cluster, which has been shown to be implicated in the control of allergy and resilience against infectious diseases such as malaria (Garcia et al., 1998; Rihet et al., 1998; Flori et al., 2003; Hernandez-Valladares et al., 2004) and leishmaniasis (Jeronimo et al., 2007). The region also flanks *SPOCK1*, a candidate gene for puberty both in humans (Liu et al., 2009) and cattle (Fortes et al., 2010). Although fertility and resistance to infectious diseases are candidate biological drivers of this ROH hotspot, the gene and the phenotype underlying this putative selection signature are unknown.

The common ROH at 12:28433881-29743057 identified in the present study also overlaps a common ROH hotspot (Sölkner et al., 2014) and a region of divergent selection between *Bos taurus* and *Bos indicus* cattle (Gautier et al., 2009; Porto-Neto et al., 2013), and the segment encompasses the human ortholog *BRCA2*, involved in Fanconi anemia in humans (Howlett et al., 2002). A signature of selection nearby the 4:46384250-47113352 region detected here has also been reported by Gautier and Naves (2011), but the genes involved and the selective pressure remain uncharacterized.

Table 1 | Detection of common runs of homozygosity according to different number of consecutive SNPs, percentage of animals, gap size, and number of heterozygous genotypes.

Gap size	30 SNPs				150 SNPs				Heterozygotes
	10%	20%	25%	50%	10%	20%	25%	50%	
100 Kb	437	106	57	9	186	29	12	1	0
	471	288	183	29	365	91	47	7	2
500 Kb	479	126	76	13	193	32	14	1	0
	768	334	214	45	375	96	50	7	2

CONCLUSIONS

We used high-density SNP genotypes to successfully characterize autozygosity in Nellore cows under artificial selection for reproductive, carcass and weight traits. We have shown that, although the massive use of relatively few sires and artificial insemination has generated long stretches of homozygous haplotypes in the genomes of over 70% of these animals, inbreeding levels were considerably low in this population. We also found few genomic regions with high homozygosity across individuals, suggesting that the ongoing selection for reproductive, weight and carcass traits in this population is not very intensive or too recent to have left selection signatures in the form of ROH. Furthermore, the current common breeding practices of avoiding inbreeding in the mating schemes are antagonistic to additive trait selection,

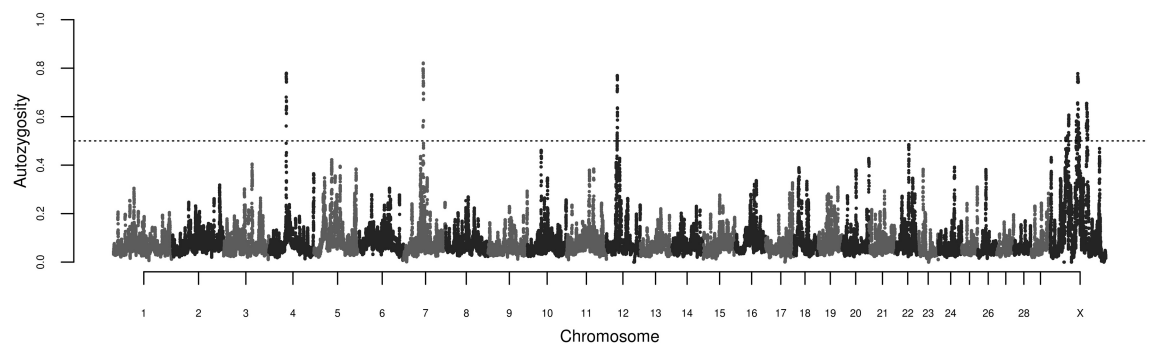


FIGURE 4 | Manhattan plot of genome-wide locus autozygosity in Nellore cows. The dashed line represents the 50% threshold.

making it hard to maintain ROH signatures in the herds. The three candidate regions under selection identified herein were likely to be contributions from remote ancestors, predating the foundation of the Nellore breeding programs. The selective pressure effects and the genes involved in these regions should be subject of future investigation.

ACKNOWLEDGMENTS

We thank Guilherme Penteado Coelho Filho and Daniel Biluca for technical assistance in sample acquisition. We also thank to Fernando Sebastian Baldi Rey for the manuscript revision and pertinent suggestions. This research was supported by: National Counsel of Technological and Scientific Development (CNPq - http://www.cnpq.br/) (process 560922/2010-8 and 483590/2010-0); and São Paulo Research Foundation (FAPESP - http://www.fapesp.br/) (process 2013/15869-2 and 2014/01095-8). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. Mention of trade name proprietary product or specified equipment in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the authors or their respective institutions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://www.frontiersin.org/journal/10.3389/fgene.2015.00005/abstract

REFERENCES

Ajmone-Marsan, P., Garcia, J. F., Lenstra, J. A., and The Globaldiv Consortium (2010). On the origin of cattle: how aurochs became cattle and colonized the world. *Evol. Anthropol.* 19, 148–157. doi: 10.1002/evan.20267

Bjelland, D. W., Weigel, K. A., Vukasinovic, N., and Nkrumah, J. D. (2013). Evaluation of inbreeding depression in Holstein cattle using whole-genome SNP markers and alternative measures of genomic inbreeding. *J. Dairy Sci.* 96, 4697–4706. doi: 10.3168/jds.2012-6435

Bovine HapMap Consortium (2009). Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Science* 324, 528–532. doi: 10.1126/science.1167936

Curik, I., Ferenčaković, M., and Sölkner, J. (2014). Inbreeding and runs of homozygosity: a possible solution to an old problem. *Livest. Sci.* 166, 26–34. doi: 10.1016/j.livsci.2014.05.034

Deng, X., Berletch, J. B., Nguyen, D. K., and Disteche, C. M. (2014). X chromosome regulation: diverse patterns in development, tissues and disease. *Nat. Rev. Genet.* 15, 367–378. doi: 10.1038/nrg3687

Espigolan, R., Baldi, F., Boligon, A. A., Souza, F. R., Gordo, D. G., Tonussi, R. L., et al. (2013). Study of whole genome linkage disequilibrium in Nellore cattle. *BMC Genomics*. 14:305. doi: 10.1186/1471-2164-14-305

Falconer, D. S., and Mackay, T. F. C. (1996). *Introduction to Quantitative Genetics*, 4 Edn. (Harlow, UK: Addison-Wesley Longman), 464.

Ferenčaković, M., Hamziæ, E., Gredler, B., Solberg, T. R., Klemetsdal, G., Curik, I., et al. (2013a). Estimates of autozygosity derived from runs of homozygosity: empirical evidence from selected cattle populations. *J. Anim. Breed. Genet.* 130, 286–293. doi: 10.1111/jbg.12012

Ferenčaković, M., Sölkner, J., and Curik, I. (2013b). Estimating autozygosity from high-throughput information: effects of SNP density and genotyping errors. *Genet. Sel. Evol.* 45:42. doi: 10.1186/1297-9686-45-42

Flori, L., Kumulungui, B., Aucan, C., Esnault, C., Traore, A. S., Fumoux, F., et al. (2003). Linkage and association between *Plasmodium falciparum* blood infection levels and chromosome 5q31-q33. *Genes Immun.* 4, 265–268. doi: 10.1038/sj.gene.6363960

Fortes, M. R. S., Reverter, A., Zhang, Y., Collis, E., Nagaraj, S. H., Jonsson, N. N., et al. (2010). Association weight matrix for the genetic dissection of puberty in beef cattle. *Proc. Natl. Acad. Sci. U.S.A.* 107, 13642–136427. doi: 10.1073/pnas.1002044107

Garcia, A., Marquet, S., Bucheton, B., Hillaire, D., Cot, M., Fievet, N., et al. (1998). Linkage analysis of blood *Plasmodium falciparum* levels: interest of the 5q31-q33 chromosome region. *Am. J. Trop. Med. Hyg.* 58, 705–709.

Gautier, M., Flori, L., Riebler, A., Jaffrézic, F., Laloé, D., Gut, I., et al. (2009). A whole genome Bayesian scan for adaptive genetic divergence in West African cattle. *BMC Genomics*. 10:550. doi: 10.1186/1471-2164-10-550

Gautier, M., and Naves, M. (2011). Footprints of selection in the ancestral admixture of a New World Creole cattle breed. *Mol. Ecol.* 20, 3128–3143. doi: 10.1111/j.1365-294X.2011.05163.x

Gibson, J., Morton, N. E., and Collins, A. (2006). Extended tracts of homozygosity in outbred human populations. *Hum. Mol. Genet.* 15, 789–795. doi: 10.1093/hmg/ddi493

Gottipati, S., Arbiza, L., Siepel, A., Clark, A. G., and Keinan, A. (2011). Analyses of X-linked and autosomal genetic variation in population-scale whole genome sequencing. *Nat. Genet.* 43, 741–743. doi: 10.1038/ng.877

Hammer, M. F., Woerner, A. E., Mendez, F. L., Watkins, J. C., Cox, M. P., and Wall, J. D. (2010). The ratio of human X chromosome to autosome diversity is positively correlated with genetic distance from genes. *Nat. Genet.* 42, 830–831. doi: 10.1038/ng.651

Hanotte, O., Ronin, Y., Agaba, M., Nilsson, P., Gelhaus, A., Horstmann, R., et al. (2003). Mapping of quantitative trait loci controlling trypanotolerance in a cross of tolerant West African N'Dama and susceptible East African Boran cattle. *Proc. Natl. Acad. Sci. U.S.A.* 100, 7443–7448. doi: 10.1073/pnas.1232392100

Hernandez-Valladares, M., Rihet, P., ole-MoiYoi, O. K., and Iraqi, F. A. (2004). Mapping of a new quantitative trait locus for resistance to malaria in mice by a comparative mapping approach with human chromosome 5q31-q33. *Immunogenetics* 56, 115–117. doi: 10.1007/s00251-004-0667-0

Howlett, N. G., Taniguchi, T., Olson, S., Cox, B., Waisfisz, Q., De Die-Smulders, C., et al. (2002). Biallelic inactivation of *BRCA2* in Fanconi anemia. *Science* 297, 606–609. doi: 10.1126/science.1073834

Howrigan, D. P., Simonson, M. A., and Keller, M. C. (2011). Detecting autozygosity through runs of homozygosity: a comparison of three autozygosity detection algorithms. *BMC Genomics* 12:460. doi: 10.1186/1471-2164-12-460

Jeronimo, S. M., Holst, A. K., Jamieson, S. E., Francis, R., Martins, D. R., Bezerra, F. L., et al. (2007). Genes at human chromosome 5q31.1 regulate delayed-type hypersensitivity responses associated with *Leishmania chagasi* infection. *Genes Immun.* 8, 539–551. doi: 10.1038/sj.gene.6364422

Keller, M. C., Visscher, P. M., and Goddard, M. E. (2011). Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. *Genetics* 189, 237–249. doi: 10.1534/genetics.111.130922

Kim, E. S., Cole, J. B., Huson, H., Wiggans, G. R., Van Tassell, C. P., Crooker, B. A., et al. (2013). Effect of artificial selection on runs of homozygosity in u.s. Holstein cattle. *PLoS ONE* 8:e80813. doi: 10.1371/journal.pone.0080813

Lencz, T., Lambert, C., DeRosse, P., Burdick, K. E., Morgan, V., Kane, J. M., et al. (2007). Runs of homozygosity reveal highly penetrant recessive loci in schizophrenia. *Proc. Natl. Acad. Sci. U.S.A.* 104, 19942–19947. doi: 10.1073/pnas.0710021104

Leroy, G. (2014). Inbreeding depression in livestock species: review and meta-analysis. *Anim. Genet.* 45, 618–628. doi: 10.1111/age.12178

Liu, Y. Z., Guo, Y. F., Wang, L., Tan, L. J., Liu, X. G., Pei, Y. F., et al. (2009). Genome-wide association analyses identify SPOCK as a key novel gene underlying age at menarche. *PLoS Genet.* 5:e1000420. doi: 10.1371/journal.pgen.1000420

MacArthur, D. G., Balasubramanian, S., Frankish, A., Huang, N., Morris, J., Walter, K., et al. (2012). A systematic survey of loss-of-function variants in human protein-coding genes. *Science* 335, 823. doi: 10.1126/science.1215040

McKay, S. D., Schnabel, R. D., Murdoch, B. M., Matukumalli, L. K., Aerts, J., Coppieters, W., et al. (2007). Whole genome linkage disequilibrium maps in cattle. *BMC Genet.* 8:74. doi: 10.1186/1471-2156-8-74

McQuillan, R., Leutenegger, A. L., Abdel-Rahman, R., Franklin, C. S., Pericic, M., Barac-Lauc, L., et al. (2008). Runs of homozygosity in European populations. *Am. J. Hum. Genet.* 83, 359–372. doi: 10.1016/j.ajhg.2008.08.007

Meirelles, F. V., Rosa, A. J. M., Lobo, R. B., Garcia, J. M., Smith, L. C., and Duarte, F. A. M. (1999). Is the American zebu really *Bos indicus*? *Genet. Mol. Biol.* 22, 543–546. doi: 10.1590/S1415-47571999000400013

Pérez O'Brien, A. M., Mészáros, G., Utsunomiya, Y. T., Sonstegard, T. S., Garcia, F. J., Van Tassell, C. P., et al. (2014). Linkage disequilibrium levels in *Bos indicus* and *Bos taurus* cattle using medium and high density SNP chip data and different minor allele frequency distributions. *Livest. Sci.* 166, 121–132. doi: 10.1016/j.livsci.2014.05.007

Porto-Neto, L. R., Sonstegard, T. S., Liu, G. E., Bickhart, D. M., Da Silva, M. V., Machado, M. A., et al. (2013). Genomic divergence of zebu and taurine cattle identified through high-density SNP genotyping. *BMC Genomics* 14:876. doi: 10.1186/1471-2164-14-876

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795

Purfield, D. C., Berry, D. P., McParland, S., and Bradley, D. G. (2012). Runs of homozygosity and population history in cattle. *BMC Genet.* 13:70. doi: 10.1186/1471-2156-13-70

Rihet, P., Traor,eì, Y., Abel, L., Aucan, C., Traoreì-Leroux, T., and Fumoux, F. (1998). Malaria in humans: *Plasmodium falciparum* blood infection levels are linked to chromosome 5q31-q33. *Am. J. Hum. Genet.* 63, 498–505. doi: 10.1086/301967

Sölkner, J., Ferenčaković, M., Karimi, Z., Perez O'Brien, A. M., Mészáros, G., Eaglen, S., et al. (2014). "Extremely non-uniform: patterns of runs of homozygosity in bovine populations," in *Proceedings of the 10th World Congress of Genetics Applied to Livestock Production* (Vancouver, BC).

Szpiech, Z. A., Xu, J., Pemberton, T. J., Peng, W., Zollner, S., Rosenberg, N. A., et al. (2013). Long runs of homozygosity are enriched for deleterious variation. *Am. J. Hum. Genet.* 93, 90–102. doi: 10.1016/j.ajhg.2013.05.003

Utsunomiya, Y. T., Bomba, L., Lucente, G., Colli, L., Negrini, R., Lenstra, J. A., et al. (2014). Revisiting AFLP fingerprinting for an unbiased assessment of genetic structure and differentiation of taurine and zebu cattle. *BMC Genet.* 15:47. doi: 10.1186/1471-2156-15-47

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91, 4414–4423. doi: 10.3168/jds.2007-0980

VanRaden, P. M., Olson, K. M., Wiggans, G. R., Cole, J. B., and Tooker, M. E. (2011). Genomic inbreeding and relationships among Holsteins, Jerseys, and Brown Swiss. *J. Dairy Sci.* 94, 5673–5682. doi: 10.3168/jds.2011-4500

Xue, Y., Chen, Y., Ayub, Q., Huang, N., Ball, E. V., Mort, M., et al. (2012). Deleterious- and disease-allele prevalence in healthy individuals: insights from current predictions, mutation databases, and population-scale resequencing. *Am. J. Hum. Genet.* 91, 1022–1032. doi: 10.1016/j.ajhg.2012.10.015

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 31 October 2014; paper pending published: 03 December 2014; accepted: 07 January 2015; published online: 29 January 2015.

Citation: Zavarez LB, Utsunomiya YT, Carmo AS, Neves HHR, Carvalheiro R, Ferenčaković M, Pérez O'Brien AM, Curik I, Cole JB, Van Tassell CP, da Silva MVGB, Sonstegard TS, Sölkner J and Garcia JF (2015) Assessment of autozygosity in Nellore cows (Bos indicus) through high-density SNP genotypes. Front. Genet. ***6****:5. doi: 10.3389/fgene.2015.00005*

This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics.

Copyright © 2015 Zavarez, Utsunomiya, Carmo, Neves, Carvalheiro, Ferenčaković, Pérez O'Brien, Curik, Cole, Van Tassell, da Silva, Sonstegard, Sölkner and Garcia. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

ADVANTAGES OF PUBLISHING IN FRONTIERS

FAST PUBLICATION
Average 90 days from submission to publication

COLLABORATIVE PEER-REVIEW
Designed to be rigorous – yet also collaborative, fair and constructive

RESEARCH NETWORK
Our network increases readership for your article

OPEN ACCESS
Articles are free to read, for greatest visibility

TRANSPARENT
Editors and reviewers acknowledged by name on published articles

GLOBAL SPREAD
Six million monthly page views worldwide

COPYRIGHT TO AUTHORS
No limit to article distribution and re-use

IMPACT METRICS
Advanced metrics track your article's impact

SUPPORT
By our Swiss-based editorial team

frontiers

EPFL Innovation Park · Building I · 1015 Lausanne · Switzerland
T +41 21 510 17 00 · info@frontiersin.org · frontiersin.org