WILD PLANTS AS SOURCE OF NEW CROPS

EDITED BY: Petr Smýkal, Eric Von Wettberg and Thomas M. Davis

PUBLISHED IN: Frontiers in Plant Science



frontiers Research Topics



frontiers

Frontiers eBook Copyright Statement

The copyright in the text of individual articles in this eBook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this eBook is the property of Frontiers.

Each article within this eBook, and the eBook itself, are published under the most recent version of the Creative Commons CC-BY licence. The version current at the date of publication of this eBook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or eBook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714
ISBN 978-2-88966-143-5
DOI 10.3389/978-2-88966-143-5

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: frontiersin.org/about/contact

WILD PLANTS AS SOURCE OF NEW CROPS

Topic Editors:
Petr Smýkal, Palacký University, Olomouc, Czechia
Eric Von Wettberg, University of Vermont, United States
Thomas M. Davis, University of New Hampshire, United States

Citation: Smýkal, P., Von Wettberg, E., Davis, T. M., eds. (2020). Wild Plants as Source of New Crops. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88966-143-5

Table of Contents

05 ***Editorial: Wild Plants as Source of New Crops***
Eric von Wettberg, Thomas M. Davis and Petr Smýkal

08 ***Bird Dispersal as a Pre-Adaptation for Domestication in Legumes: Insights for Neo-Domestication***
Hester Brønnvik and Eric J. von Wettberg

11 ***Australian Wild Rice Populations: A Key Resource for Global Food Security***
Robert J. Henry

19 ***Inferring the Origin of Cultivated* Zizania latifolia*, an Aquatic Vegetable of a Plant-Fungus Complex in the Yangtze River Basin***
Yao Zhao, Zhiping Song, Lan Zhong, Qin Li, Jiakuan Chen and Jun Rong

32 ***Unlocking the Genetic Diversity and Population Structure of a Wild Gene Source of Wheat,* Aegilops biuncialis Vis.*, and Its Relationship With the Heading Time***
László Ivanizs, István Monostori, András Farkas, Mária Megyeri, Péter Mikó, Edina Türkösi, Eszter Gaál, Andrea Lenykó-Thegze, Kitti Szőke-Pázsi, Éva Szakács, Éva Darkó, Tibor Kiss, Andrzej Kilian and István Molnár

45 ***Domesticating* Vigna Stipulacea*: A Potential Legume Crop With Broad Resistance to Biotic Stresses***
Yu Takahashi, Hiroaki Sakai, Yuki Yoshitsu, Chiaki Muto, Toyoaki Anai, Muthaiyan Pandiyan, Natesan Senthil, Norihiko Tomooka and Ken Naito

57 ***Diversity of Naturalized Hairy Vetch (*Vicia villosa *Roth) Populations in Central Argentina as a Source of Potential Adaptive Traits for Breeding***
Juan P. Renzi, Guillermo R. Chantre, Petr Smýkal, Alejandro D. Presotto, Luciano Zubiaga, Antonio F. Garayalde and Miguel A. Cantamutto

71 ***Pod Dehiscence in Hairy Vetch (*Vicia villosa *Roth)***
Lisa Kissing Kucek, Heathcliffe Riday, Bryce P. Rufener, Allen N. Burke, Sarah Seehaver Eagen, Nancy Ehlke, Sarah Krogman, Steven B. Mirsky, Chris Reberg-Horton, Matthew R. Ryan, Sandra Wayman and Nick P.

81 ***Comparative Analysis of Early Life Stage Traits in Annual and Perennial* Phaseolus *Crops and Their Wild Relatives***
Sterling A. Herron, Matthew J. Rubin, Claudia Ciotir, Timothy E. Crews, David L. Van Tassel and Allison J. Miller

95 ***Enhancing Crop Domestication Through Genomic Selection, a Case Study of Intermediate Wheatgrass***
Jared Crain, Prabin Bajgain, James Anderson, Xiaofei Zhang, Lee DeHaan and Jesse Poland

110 ***Exploration of the Yield Potential of Mesoamerican Wild Common Beans From Contrasting Eco-Geographic Regions by Nested Recombinant Inbred Populations***
Jorge Carlos Berny Mier y Teran, Enéas R. Konzen, Antonia Palkovic, Siu M. Tsai and Paul Gepts

128 *An Updated Checklist of the Sicilian Native Edible Plants: Preserving the Traditional Ecological Knowledge of Century-Old Agro-Pastoral Landscapes*
Salvatore Pasta, Alfonso La Rosa, Giuseppe Garfì, Corrado Marcenò, Alessandro Silvestre Gristina, Francesco Carimi and Riccardo Guarino

143 *Lost and Found:* **Coffea stenophylla** *and* **C. affinis,** *the Forgotten Coffee Crop Species of West Africa*
Aaron P. Davis, Roberta Gargiulo, Michael F. Fay, Daniel Sarmu and Jeremy Haggar

161 *A Chromosome-Scale Assembly of the Garden Orach (***Atriplex hortensis** *L.) Genome Using Oxford Nanopore Sequencing*
Spencer P. Hunt, David E. Jarvis, Dallas J. Larsen, Sergei L. Mosyakin, Bozena A. Kolano, Eric W. Jackson, Sara L. Martin, Eric N. Jellen and Peter J. Maughan

181 *Seed Structural Variability and Germination Capacity in* **Passiflora edulis** *Sims f.* **edulis**
Nohra Rodríguez Castillo, Luz Marina Melgarejo and Matthew Wohlgemuth Blair

191 *Weedy Rice as a Novel Gene Resource: A Genome-Wide Association Study of Anthocyanin Biosynthesis and an Evaluation of Nutritional Quality*
Wenjia Wang, Minghui Zhao, Guangchen Zhang, Zimeng Liu, Yuchen Hua, Xingtian Jia, Jiayu Song, Dianrong Ma and Jian Sun

203 *New Food Crop Domestication in the Age of Gene Editing: Genetic, Agronomic and Cultural Change Remain Co-evolutionarily Entangled*
David L. Van Tassel, Omar Tesdell, Brandon Schlautman, Matthew J. Rubin, Lee R. DeHaan, Timothy E. Crews and Aubrey Streit Krug

219 *Genetic Analysis of Domestication Parallels in Annual and Perennial Sunflowers (***Helianthus** *spp.): Routes to Crop Development*
Sean R. Asselin, Anita L. Brûlé-Babel, David L. Van Tassel and Douglas J. Cattani

231 *Common Vetch: A Drought Tolerant, High Protein Neglected Leguminous Crop With Potential as a Sustainable Food Source*
Vy Nguyen, Samuel Riley, Stuart Nagel, Ian Fisk and Iain R. Searle

238 *Southern Species From the Biodiversity Hotspot of Central Chile: A Source of Color, Aroma, and Metabolites for Global Agriculture and Food Industry in a Scenario of Climate Change*
Luis Letelier, Carlos Gaete-Eastman, Patricio Peñailillo, María A. Moya-León and Raúl Herrera

254 *Utilizing Wild* **Cajanus platycarpus,** *a Tertiary Genepool Species for Enriching Variability in the Primary Genepool for Pigeonpea Improvement*
Shivali Sharma, Pronob J. Paul, CV Sameer Kumar and Chetna Nimje

265 *Crop Wild Relatives as Germplasm Resource for Cultivar Improvement in Mint (***Mentha** *L.)*
Kelly J. Vining, Kim E. Hummer, Nahla V. Bassil, B. Markus Lange, Colin K. Khoury and Dan Carver

frontiers in Plant Science

EDITORIAL
published: 11 September 2020
doi: 10.3389/fpls.2020.591554

Editorial: Wild Plants as Source of New Crops

Eric von Wettberg[1,2], *Thomas M. Davis*[2] *and Petr Smýkal*[3*]

[1] Plant and Soil Science and Gund Institute for the Environment, University of Vermont, Burlington, VT, United States,
[2] Department of Agriculture, Nutrition, and Food Systems, University of New Hampshire, Durham, NH, United States,
[3] Department of Botany, Faculty of Science, Palacky University, Olomouc, Czechia

Keywords: crop wild relatives, domestication, genetic diversity, neo-domestication, breeding, ethnobotany

Editorial on the Research Topic

Wild Plants as Source of New Crops

The history of agriculture can be viewed as a series of key events, such as the Neolithic Revolution, post-domestication expansion of agriculture to new regions, secondary domestications of new crops, movement over the Silk Road, the Columbian Exchange, the Industrial Revolution, the Green Revolution and even the more recent, ongoing genomic revolutions. Each of these has had positive benefits, but they have also come at a cost, including to agricultural biodiversity.

It is estimated that on the Earth there are between 300,000 and 500,000 species of higher plants, of which approximately 369,000 have been identified or described (Willis, 2017). Many species are still unknown to science, while perhaps a third is at risk of extinction (Pimm and Joppa, 2015) . The number of plant species used for food by pre-agricultural human societies is estimated to be around 7,000, but only a small fraction of the diversity of the plant kingdom has been domesticated. Our present knowledge of domesticated plants largely reflects our experience of a relatively small number of living domesticates adapted to recent, Holocene environments. The process of crop domestication was based on selection driven by human cultivation practices and agricultural environments. Approximately, 2,500 species have undergone some degree of domestication, and 250 species are considered to be fully domesticated, in the sense that their full lifecycle became dependent on human cultivation (Meyer et al., 2012; Gaut et al., 2018; Smýkal et al., 2018). Humanity relies on a small collection of crop plants such as corn, rice, wheat, soybean, and potato constituting the majority of our dietary intake. Altogether, some 10 to 50 plant species together provide about 95% of the world's caloric intake. This concentration on a few species for most food is a key element of the vulnerability of the world food supply to the impact of climate change and the outbreak of major new plant diseases.

Crop wild relatives (CWRs) remain the largest reservoir of genetic diversity for crop improvement and have been utilized for major gene disease and pest resistance, and abiotic stress tolerance (Vavilov et al., 1992; Hajjar and Hodgkin, 2007; Warschefsky et al., 2014; Castañeda-Álvarez et al., 2016; Smýkal et al., 2018; Coyne et al., 2020). However, there remains a large set of plant species from various plant families and genera which have favorable traits but have not been domesticated so far. As we have been gaining knowledge on the genomic and biological background of domestication processes, we can apply more effective selection to domesticate more wild species. Since many wild taxa are locally adapted to particular habitats and contain significant genetic diversity, this might create novel crops and help us to achieve more environmentally sustainable agriculture as we face climate change. Not all candidates for neo-domestication are CWR, although many and perhaps most will be, because the form/function of the related crop species provides a useful template to guide the neo-domestication of the CWR. On the other hand, all of the wild sources of useful genes for introgression into crop species are necessarily

OPEN ACCESS

Edited and reviewed by:
Inaki Hormaza, Institute of Subtropical and Mediterranean Horticulture La Mayora, Spain

****Correspondence:***
Petr Smýkal
petr.smykal@upol.cz

Specialty section:
This article was submitted to Plant Breeding, a section of the journal Frontiers in Plant Science

Received: *04 August 2020*
Accepted: *28 August 2020*
Published: *11 September 2020*

Citation:
von Wettberg E, Davis TM and Smýkal P (2020) Editorial: Wild Plants as Source of New Crops. Front. Plant Sci. 11:591554. doi: 10.3389/fpls.2020.591554

CWR as a prerequisite for their ability to cross with a crop relative. So, a given CWR can be a valuable source for gene introgression into its crop relative, and can at the same time be a candidate for neo-domestication.

Given our reliance on fewer and fewer crop species, as well as the need to more sustainably intensify agricultural production and to develop more nutritious crops, utilizing a greater range of plant diversity is critically important. While conventional breeding by hybridization and progeny selection augmented by marker-assisted selection are likely to be the most practicable approaches for the time-being, recent developments in genome editing technologies, genomic selection, and inoculation of plants with beneficial microbes may help accelerate this process in the future by converting the wild forms of key domestication genes to their respective domesticated forms, as reviewed by Van Tassel et al.

Several tools exist for those wishing to neo-domesticate a wild species or to make an existing crop more tractable to genetic improvement. The most obvious of newly emerging tools are gene editing approaches. Although not technically trivial, gene editing offers a neo-domesticator the opportunity to simply engineer the domestication traits of interest, such as indehiscent propagules, based on knowledge from other taxa (e.g., Ogutcen et al., 2018). However, other tools of modern genomics may be useful to those wishing to neo-domesticate a crop. These include genome sequencing and resequencing to characterize molecular variation, Genome Wide Association Studies (GWAS) to find associations between genes and traits of interest and genomic selection to improve multi-genic traits in a breeding program. We also review several core concepts in neo-domestication. Our special issue focuses nearly equally on neodomestication of wild species and harnessing an understanding of domestication traits to improve existing crop species.

This Research Topic, Wild Plants as Source of New Crops, covers the following topics:

1. De Novo Domestication of Plant Species

   One of our examples of *de novo* domestication comes from Takahashi et al., who examine neo-domestication in *Vigna stipulaceae*. The genus *Vigna*, which includes nine independent domestications, is particularly ripe for domestication. Many of the wild species, such as *V. stipulaceae*, have useful adaptation traits such as drought tolerance or salinity tolerance, making them desirable targets for neo-domestication. Renzi et al. have looked at *Vicia villosa*, a stress tolerant and rapidly growing forage. Kissing Kucek et al. addressed the shattering pod traits that makes *V. villosa* weedy and difficult to control in field settings. Selecting for non-dehiscence would help to domesticate this useful species. Common vetch has also been utilized as a drought tolerant forage (Nguyen et al.). Another take on neo-domestication is to look for traits that may pre-adapt certain lineages to domestication. Bronnvik and von Wettberg look at bird dispersal in legumes as a possible trait preadapting some legumes toward indehiscent pods.
2. Use of Crop Wild Relatives to Broaden Genetic Diversity of Crops

   Crop wild relatives are an important source of variability for existing crops and can be used more effectively when the genetic basis of domestication traits are understood. Henry reviews the diversity of wild rice relatives in Australia, while Wang et al. use GWAS analysis of weedy rice in relation to seed nutritional quality. Berny Mier y Teran et al. have examined wild relatives of common bean, *Phaseolus vulgaris*, with introgression populations. Ivanizs et al. examine the potential of *Aegilops biuncialis* for wheat improvement. Sharma et al. examine the tertiary genepool of pigeonpea, *Cajanus cajan*. Interesting re-discovery of two coffee species is reported by Davis et al.
3. Use of Perennial Species

   Crain et al. have used genomic selection for improvement of intermediate wheatgrass. Herron et al. examine early growth traits across the genus *Phaseolus*, with an emphasis on wild perennial taxa, with the aim of identifying perennial species that might be suitable targets for neo-domestication. An intriguing illustration of domestication parallels is provided by analysis of annual and perennial sunflowers (Asselin et al.). There are many domesticated tree species. They differ from typical seed propagated annual or perennial species due to their long lifespan and combination of sexual and clonal propagation. Differences in seed anatomy and germination capacity of wild and cultivated passion fruit have been studied by Castillo et al.
4. Novel Domestication Syndromes

   Zhao et al. have examined the domestication of *Zizania latifolia*, a grass with a novel domestication as a vegetable, where a symbiosis with an endosymbiotic fungus is essential to the domestication syndrome.
5. Medicinal and Horticultural Plants, Their Domestication, and Diversity

   Pasta et al. have examined the flora of Sicily for edible plants, in an effort to preserve ecological knowledge and potential medicinal value. Letelier et al. make a similar examination of the flora of Chile. A genomic approach to chromosome scale assembly of garden orach (*Atriplex hortensis*) is presented by Hunt et al. The use of CWR germplasm for cultivar improvement of the widely used herb mint (*Mentha*) is shown by Vining et al.

AUTHOR CONTRIBUTIONS

TD, EW, and PS have equally contributed to writing and final editing. All authors contributed to the article and approved the submitted version.

FUNDING

PS is supported by the Grant Agency of the Czech Republic and Palacký University grant Agency [IGA-2020_003]. EW is supported by the USDA Hatch program through the Vermont State Agricultural Experimental Station and is supported to work on diversity of mungbeans by Russian Scientific Fund Project No. 18-46-08001 on the basis of a unique scientific installation "Collectionof plant genetic resources VIR". TD is supported in part by the USDA National Institute of Food and Agriculture Hatch Project NH00678 (accession number 1019990).

REFERENCES

Castañeda-Álvarez, N. P., Khoury, C. K., Achicanoy, H. A., Bernau, V., Dempewolf, H., Eastwood, R. J., et al. (2016). Global conservation priorities for crop wild relatives. *Nat. Plants* 2, 1–6. doi: 10.1038/nplants.2016.22

Coyne, C. J., Shiv Kumar, S., von Wettberg, E. B., Marques, E., Berger, J. D., Redden, R. J., et al. (2020). Potential and limits of exploitation of crop wild relatives for pea, lentil and chickpea improvement. *Legume Sci.* 2, e36. doi: 10.1002/leg3.36

Gaut, B. S., Seymour, D. K., Liu, Q., and Zhou, Y. (2018). Demography and its effects on genomic variation in crop domestication. *Nat. Plants* 4, 512–520. doi: 10.1038/s41477-018-0210-1

Hajjar, R., and Hodgkin, T. (2007). The use of wild relatives in crop improvement: a survey of developments over the last 20 years. *Euphytica* 156, 1–13. doi: 10.1007/s10681-007-9363-0

Meyer, R. S., DuVal, A. E., and Jensen, H. R. (2012). Patterns and processes in crop domestication: an historical review and quantitative analysis of 203 global food crops. *New Phytol.* 196, 29–48. doi: 10.1111/j.1469-8137.2012.04253.x

Ogutcen, E., Pandey, A., Khan, M. K., Marques, E., Penmetsa, R. V., Kahraman, A., et al. (2018). Pod shattering: a homologous series of variation underlying domestication and an avenue for crop improvement. *Agronomy* 8, 137. doi: 10.3390/agronomy8080137

Pimm, S. L., and Joppa, L. N. (2015). How Many Plant Species are There, Where are They, and at What Rate are They Going Extinct? *Ann. Missouri Bot. Gard.* 100, 170–176. doi: 10.3417/2012018

Smýkal, P., Nelson, M. N., Berger, J. D., and Von Wettberg, E. J. (2018). The impact of genetic changes during crop domestication. *Agronomy* 8, 119. doi: 10.3390/agronomy8030026

Vavilov, N. I., Vavylov, M. I., Vavílov, N. I., and Dorofeev, V. F. (1992). *Origin and geography of cultivated plants* (Cambridge University Press).

Warschefsky, E., Penmetsa, R. V., Cook, D. R., and Von Wettberg, E. J. (2014). Back to the wilds: tapping evolutionary adaptations for resilient crops through systematic hybridization with crop wild relatives. *Am. J. Bot.* 101, 1791–1800. doi: 10.3732/ajb.1400116

K. J. Willis (Ed.) (2017). *State of the World"s Plants 2017. Report* (Kew: Royal Botanic Gardens).

Conflict of Interest: 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 © 2020 von Wettberg, Davis and Smýkal. 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) and the copyright owner(s) 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 Plant Science

MINI REVIEW
published: 15 October 2019
doi: 10.3389/fpls.2019.01293

Bird Dispersal as a Pre-Adaptation for Domestication in Legumes: Insights for Neo-Domestication

Hester Brønnvik[1†] and Eric J. von Wettberg[1,2†]*

[1] Department of Plant and Soil Science, University of Vermont, Burlington, VT, United States,
[2] Mathematical Biology and Bioinformatics Laboratory, Peter the Great St. Petersburg Polytechic University, Saint Petersburg, Russia

Keywords: neodomestication, bird dispersal, legumes (fabaceae), preadaptation, natural history, domestication syndrome, seed dispersal

OPEN ACCESS

Edited by:
Roberto Papa, Marche Polytechnic University, Italy

Reviewed by:
Giovanna Attene, University of Sassari, Italy
Michael Benjamin Kantar, University of Hawaii, United States

****Correspondence:***
Eric J. von Wettberg
ebishopv@uvm.edu

†ORCID:
orcid.org/0000-0002-2724-0317

Specialty section:
This article was submitted to Plant Breeding, a section of the journal Frontiers in Plant Science

Received: *17 July 2019*
Accepted: *18 September 2019*
Published: *15 October 2019*

Citation:
Brønnvik H and von Wettberg EJ (2019) Bird Dispersal as a Pre-Adaptation for Domestication in Legumes: Insights for Neo-Domestication. Front. Plant Sci. 10:1293. doi: 10.3389/fpls.2019.01293

The first chapter of Darwin's *On the Origin of Species* famously used examples of human selection of domesticated plants and animals to lay the groundwork for the Theory of Evolution (Darwin, 1859). Since at least the works for Darwin (1868), domestication of plants and animals have been used as major examples of strong selection radically altering the morphology, architecture, and behavior of organisms on which our contemporary society relies for food, fiber, and fuel. Consequently, it is not surprising that crop domestication remains a vibrant area of research. Despite this ongoing interest in the field of domestication, we greatly lack ecological and natural history studies of crop wild relatives in their wild settings.

One instance of the great lack of natural history of crop wild relatives is in understanding, the dispersal biology of crop wild relatives as a potential pre-adaptation for domestication. The domestication of annual crops such as cereals and grains legumes is thought to require two key traits that are part of the domestication syndrome (Hammer, 1984): seeds that lack dormancy and fruit structures that are indehiscent (Purugganan and Fuller, 2009; Ogutcen et al., 2018; Smýkal et al., 2018). Debate in the domestication literature has long been generally assumed that both of these traits are disfavored in natural populations and then speculated whether the genetic changes underlying the domestication shift arose from standing variation at low frequency in natural populations (mildly deleterious alleles) or new mutations (e.g., Morrell et al., 2012; Olsen and Wendel, 2013; Gaut et al., 2018; Hufford et al., 2019; Lye and Purugganan, 2019). However, the assumption that these traits would be disfavored in natural populations is just that and is not based on careful observation of actual dispersal or germination biology. We think it is possible that these traits may, in some, taxa have evolved long ago as part of their dispersal biology, and that this possibility has not been sufficiently studied. Birds and other animal dispersers may more effectively disperse whole fruit than individual seeds, favoring indehiscence in legumes—but we do not know this for a critical lack of natural history knowledge. If so, this is a pre-adaptation for domestication which may have pre-disposed some groups toward domestication and can be used in selected taxa for neodomesticaiton. Here, we document the limited natural history knowledge of bird dispersal in legumes that are consumed by humans and show a significant gap in our natural history knowledge of crop wild relatives.

Animal dispersal syndromes in plants are well described—for example, by the landmark review by Howe and Smallwood (1982). Birds can be particularly effective at dispersing seeds a great distance and to suitable habitats. Bird dispersal can be critical to allow offspring to escape higher disease pressure by escaping the proximity of parents. Many bird dispersal behaviors also allow directed dispersal to particular habitats where seedling establishment is more likely to be successful, whether by burying seeds through caching behavior or dropping consumed seeds with fertilizing feces. However, birds are not the only effective animal dispersers of seeds. For

TABLE 1 | Examples of bird dispersal in legumes.

Location	Plant	Disperser	Evidence	Reference
Veracruz, Tamaulipas, and Oaxaca, Mexico	*Acacia cornigera*	*Psilorhinus morio*	Germination from fecal samples	Janzen (1969)
Muak-lek, Thailand	*Acacia mangium* Willd. × *A. auriculiformis* A. Cunn. ex Benth.	"Small birds"	Observed consumption	Sornsathapornkul and Owens (1998)
Bolivian Andes	*Inga feuillei*	*Thectocercus acuticaudatus*	Observed stomatochory	Blanco et al. (2015)
Western Australia	*Medicago sp*.	*Dromaius novaehollandiae*	Found in fecal sample	Calviño-Cancela et al. (2006)
Sumba, Indonesia	*Phaseolus lunatus*	*Cacatua sulphurea citrinocristata*	Observed consumption	Hidayat (2014)
Junín, Peru	*Phaseolus vulgaris*	"Pichupa"		Tohme et al. (1995)
"Home range"	*Pueraria lobata*	"Birds"		EPPO (2007)

herbaceous plants or grasses that occur in habitats where bird dispersal may not be effective, large mammalian herbivory may be a more common form of long-distance dispersal (Janzen, 1984). Ants are also often effective dispersers, particularly in temperate environments, although they can also be significant seed predators (e.g., Hulme, 1998).

To see if animal dispersal vectors for a suite of crops are well done, we performed a literature search using the terms "bird" and "dispersal" and a list of cultivated legumes from Smýkal et al. (2014) and our own knowledge, to find published or gray-literature reports of bird dispersal of legumes in genera known to have cultivated species. In our literature search, we were only able to uncover four instances of descriptions of bird dispersal in genera with cultivated legumes (**Table 1**). This is a rather small number of reports, particularly given the very large size of both of the *Fabaceae* and the high number of cultivated species in the family. This search approach almost certainly misses reports in languages other than English but is sufficient to make our primary point very clearly, which there is an absence of important natural history work characterizing crop wild relatives.

These examples do show some important observations. In *Phaseolus*, bird dispersal is likely widespread, and likely a key component of the very large distribution of wild *Phaseolus vulgaris* (and potentially *Phaseolus lunatus,* lima bean) from Mexico to the central Andes of South America (Ariani et al., 2017). This broad distribution likely contributed to the two independent domestications of common bean (and lima bean). Although the possibility bird dispersal has been remarked upon (Gepts, pers. comm.), there are not many natural history observations of wild *Phaseolus* to determine how widespread bird dispersal is, whether different guilds of birds are responsible for it, or whether *Phaseolus* species vary in their propensity for bird dispersal.

Understanding whether birds are a disperser is a starting point for understanding dispersal syndromes in the wild. As a pre-adaptation for cultivation and domestication, the pod would need to be indehiscent. This is a trait that can be tested in a common garden, although dehiscence can be modulated by the environment, such that the humidity of the environment may affect dehiscence (e.g., Lush et al., 1980; Ogutcen et al., 2018). Consequently, dehiscence in a humid environment may not indicate dehiscence in an arid environment.

Most taxa, particularly those in parts of Africa, South, Southeast, Southwest, and East Asia, and South America, where natural history observations published in English may be particularly absent, likely are simply data deficient. Given that the majority of legumes were domesticated in Vavilovian centers of origin in these regions, certainly there is almost simply a great lack of natural history data.

There is a great need for more ecological study and natural history observation of crop wild relatives. Crop wild relatives are the most significant reservoir of adaptative variation for providing disease resistance, abiotic stress tolerance, and other important traits to cultivated species. As crop wild relatives receive almost no conservation protection in natural populations and are badly underrepresented in most Genebanks (Maxted and Kell, 2009; Warschefsky et al., 2014), natural history study of these species remains critical to their long-term conservation.

AUTHOR CONTRIBUTIONS

Both authors contributed to the idea development, literature search, and writing of this mini-review.

FUNDING

The literature review portion of this work was supported by Russian Scientific Fund Project No. 18-46-08001 on the basis of a unique scientific installation «Collection of plant genetic resources VIR», by a cooperative agreement from the United States Agency for International Development under the Feed the Future Program AID-OAA-A-14-00008 to D.R.Cook and Co-PI EW, by a grant from the US National Science Foundation Plant Genome Program under Award IOS-1339346 to D.R.Cook, and EW; US NIFA grant # 2018-67013-27619 R. Varma Penmetsa and EW. EW is further supported by the USDA Hatch program through the Vermont State Agricultural Experimental Station.

ACKNOWLEDGMENTS

The authors thank Paul Gepts and Roberto Papa for helpful conversation.

REFERENCES

Ariani, A., Berny Mier y Teran, J. C., and Gepts, P. (2017). Spatial and temporal scales of range expansion in wild Phaseolus vulgaris. *Mol. Biol. Evol.* 35 (1), p.119–131. doi: 10.1093/molbev/msx273

Blanco, G., Hiraldo, F., Rojas, A., Dénes, F. V., and Tella, J. L. (2015). Parrots as key multilinkers in ecosystem structure and functioning. *Ecol. Evol.* 5 (18), 4141–4160. doi: 10.1002/ece3.1663

Calviño-Cancela, M., Dunn, R., Van Etten, E. J., and Lamont, B. (2006). Emus as non-standard seed dispersers and their potential for long-distance dispersal. *Ecography* 29 (4), 632–640. doi: 10.1111/j.0906-7590.2006.04677.x

Darwin, C. (1859). *On the origin of species by means of natural selection*. (London: J. Murray).

Darwin, C. (1868). *2010. The variation of animals and plants under domestication,* vol. I y II. (New York: D. Appleton and Co. Cambridge University Press Digital Version). doi: 10.1017/CBO9780511709517

EPPO. (2007). Data sheets on quarantine pests. Pueraria lobata. *EPPO Bull.* 37, 230235. doi: 10.1111/j.1365-2338.2007.01113.x

Gaut, B. S., Seymour, D. K., Liu, Q., and Zhou, Y. (2018). Demography and its effects on genomic variation in crop domestication. *Nat. Plants* 4 (8), 512. doi: 10.1038/s41477-018-0210-1

Hammer, K. (1984). Das domestikationssyndrom. *Kulturpflanze* 32 (1), 11–34. doi: 10.1007/BF02098682

Hidayat, O. (2014). Komposisi, Preferensi dan Sebaran Jenis Tumbuhan Pakan Kakatua Sumba (Cacatua sulphurea citrinocristata) di Taman Nasional Laiwangi Wanggameti. *J. Penelit. Kehutanan Wallacea* 3 (1), 25–36. doi: 10.18330/jwallacea.2014.vol3iss1pp25-36

Howe, H. F., and Smallwood, J. (1982). Ecology of seed dispersal. *Annu. Rev. Ecol. Syst.* 13 (1), 201–228. doi: 10.1146/annurev.es.13.110182.001221

Hulme, P. E. (1998). Post-dispersal seed predation: consequences for plant demography and evolution. Perspectives in Plant Ecology. *Evol. Syst.* 1 (1), 32–46. doi: 10.1078/1433-8319-00050

Hufford, M. B., Berny Mier y Teran, J. C., and Gepts, P. (2019). Crop biodiversity: an unfinished magnum opus of nature. *Annu. Rev. Plant Biol.* 70, 727–751.

Janzen, D. H. (1969). Seed-eaters versus seed size, number, toxicity and dispersal. *Evolution* 23 (1), 1–27. doi: 10.1111/j.1558-5646.1969.tb03489.x

Janzen, D. H. (1984). Dispersal of small seeds by big herbivores: foliage is the fruit. *Am. Nat.* 123 (3), 338–353. doi: 10.1086/284208

Lush, W. M., Evans, L. T., and Wien, H. C. (1980). Environmental adaptation of wild and domesticated cowpeas (Vigna unguiculata (L.) Walp.). *Field Crops Res.* 3, 173–187. doi: 10.1016/0378-4290(80)90023-4

Lye, Z. N., and Purugganan, M. (2019). Copy number variation in domestication. *Trends Plant Sci.* doi: 10.1016/j.tplants.2019.01.003

Maxted, N., and Kell, S. P. (2009). *Establishment of a global network for the in situ conservation of crop wild relatives: status and needs*. (Rome, Italy: FAO Commission on Genetic Resources for Food and Agriculture).

Morrell, P. L., Buckler, E. S., and Ross-Ibarra, J. (2012). Crop genomics: advances and applications. *Nature Rev. Genet.* 13 (2), 85. doi: 10.1038/nrg3097

Ogutcen, E., Pandey, A., Khan, M., Marques, E., Penmetsa, R., Kahraman, A., et al. (2018). Pod shattering: a homologous series of variation underlying domestication and an avenue for crop improvement. *Agronomy* 8 (8), 137. doi: 10.3390/agronomy8080137

Olsen, K. M., and Wendel, J. F. (2013). A bountiful harvest: genomic insights into crop domestication phenotypes. *Annu. Rev. Plant Boil.* 64, 47–70. doi: 10.1146/annurev-arplant-050312-120048

Purugganan, M. D., and Fuller, D. Q. (2009). The nature of selection during plant domestication. *Nature* 457 (7231), 843. doi: 10.1038/nature07895

Sornsathapornkul, P., and Owens, J. H. (1998). Pollination biology in a TropicalAcaciaHybrid (*A. mangium*Willd. × *A. auriculiformis* A. Cunn. ex Benth.). *Ann. Bot.* 5, 631–645. doi: 10.1006/anbo.1998.0607

Smýkal, P., Vernoud, V., Blair, M. W., Soukup, A., and Thompson, R. D. (2014). The role of the testa during development and in establishment of dormancy of the legume seed. *Front. Plant Sci.* 5, 351. doi: 10.3389/fpls.2014.00351

Smýkal, P., Nelson, M., Berger, J., and von Wettberg, E. (2018). The impact of genetic changes during crop domestication. *Agronomy* 8 (7), 119. doi: 10.3390/agronomy8070119

Tohme, J., Toro, O. C., Vargas, J., and Debouck, D. G. (1995). Variability in Andeannuña common beans (*Phaseolus vulgaris, Fabaceae*). *Econ. Bot.* 49 (1), 78–95. doi: 10.1007/BF02862280

Warschefsky, E., Penmetsa, R. V., Cook, D. R., and von Wettberg, E. J. (2014). Back to the wilds: tapping evolutionary adaptations for resilient crops through systematic hybridization with crop wild relatives. *Am. J.Bot.* 101 (10), 1791–1800. doi: 10.3732/ajb.1400116

Conflict of Interest: 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 © 2019 Brønnvik and von Wettberg. 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) and the copyright owner(s) 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 Plant Science

REVIEW
published: 22 October 2019
doi: 10.3389/fpls.2019.01354

Australian Wild Rice Populations: A Key Resource for Global Food Security

*Robert J. Henry**

Queensland Alliance for Agriculture and Food Innovation, University of Queensland, Brisbane, QLD, Australia

Rice is one of the most important food crops contributing to the diet of large numbers of people especially in Asia. Rice (*Oryza sativa*) was domesticated in Asia many thousands of years ago and more recently independently in Africa. Wild rice populations are found around the tropical world. The extensive production of rice in many areas has displaced the wild populations that were the basis of the original domestications by humans. Recent research, reviewed here, has identified wild rice species in northern Australia that have been isolated from the impact of domestication in Asia. Wild rice populations contain novel alleles that are a source of desirable traits such as erect habit, disease resistance, large grain size, and unique starch properties. These populations include the most divergent genotypes within the primary gene pool of rice and more distant wild relatives. Genome sequencing also suggests the presence of populations that are close relatives of domesticated rice. Hybrid populations that demonstrate mechanisms of ongoing evolution of wild *Oryza* have been identified in the wild. These populations provide options for both new domestications and utilization of novel alleles to improve or adapt domesticated rice using conventional or preferably new breeding technologies. Climate change and growing food demands associated with population and economic growth are major challenges for agriculture including rice production. The availability of diverse genetic resources to support crop adaptation and new crop domestication is critical to continued production, and increased efforts to support *in situ* and *ex situ* conservation of wild *Oryza* and related species are warranted.

Keywords: wild rice, *Oryza*, domestication, genetic diversity, Australia

OPEN ACCESS

Edited by:
Eric Von Wettberg, University of Vermont, United States

Reviewed by:
Hongwei Cai, China Agricultural University (CAU), China
Fengxia Liu, China Agricultural University (CAU), China

***Correspondence:**
Robert J. Henry
robert.henry@uq.edu.au

Specialty section:
This article was submitted to Plant Breeding, a section of the journal Frontiers in Plant Science

Received: *07 August 2019*
Accepted: *01 October 2019*
Published: *22 October 2019*

Citation:
Henry RJ (2019) Australian Wild Rice Populations: A Key Resource for Global Food Security. Front. Plant Sci. 10:1354. doi: 10.3389/fpls.2019.01354

INTRODUCTION

Rice is a key food crop with ongoing need for genetic improvement to satisfy food security. The *Oryza* genus is distributed around the tropical world. Domesticated rice has been cultivated in many of the areas that would have been native habits for wild *Oryza* species. Australia is a region that has escaped from the impact of rice domestication until very recently resulting in the persistence of many extensive populations of wild *Oryza* (Henry et al., 2010) across a very large area of northern Australia.

Advances in DNA sequencing technology have made it possible to analyze the genomes of large numbers of individuals. This has had significant impact on human genetics but also on rice biology revealing much about the diversity of the populations, the relationships between

different populations, and the evolutionary events leading up to domestication by humans (Wambugu et al., 2018).

Oryza Genus

The *Oryza* genus (Mondal and Henry, 2018) is pantropical in distribution with two domesticated and 24 wild species. Rice was domesticated in Asia as *O. sativa* and in Africa as *Oryza glaberrima*. The AA genome group (*O. sativa* complex) includes the two domesticated species and six wild relatives. The genus has 11 genome types with the AA, CC, and EE genomes types found in Australia. Other members of the *Oryzeae* tribe are also found in Australia (**Table 1**). These taxa are found in northern and eastern Australia (**Figure 1**).

TABLE 1 | The *Oryzeae* tribe in Australia.

Species	Genome	Notes
Oryza meridionalis	AA	Found in Australia and extends to New Guinea
Oryza rufipogon	AA	Populations in Australia have a chloroplast genome closer to *O. meridionalis* than *O. rufipogon* from Asia and may be best considered as a distinct taxa.
Oryza officinalis	CC	Reported from only a few locations
Oryza australiensis	EE	Endemic and widespread in northern Australia
Potamophila parviflora	*not Oryza*	Found only in a few rivers on the central east coast
Leersia hexandra	*not Oryza*	Perennial from eastern and northern Australia

AA Genome Populations in Australia

The close relatives of domesticated rice, the AA genome species, represent the primary gene pool of rice. The close relatives of domesticated rice in northern Australia have been shown to be highly diverse, including at least two distinct taxa and represent a yet unexplored wider genetic resource for global rice production. Populations have been identified with morphological and genetic characteristics suggesting a hybrid origin and some degree of ongoing gene flow between these taxa that could support continued rapid evolution of the wild *Oryza* in Australia. These Australian *Oryza* could be critical in adaptation of rice to rapid climate change and changing consumer demands and preferences. Recent evidence also suggests that, despite rice being domesticated in Asia (Huang et al., 2012), the Australian populations have contributed genes to domesticated rice (Fujino et al., 2019). The AA genomes species in Australia have been generally classified as *Oryza meridionalis* or *Oryza rufipogon*.

Oryza meridionalis

Oryza meridionalis (**Figure 2**) was first described relatively recently, in 1981, being most readily distinguished from *O.*

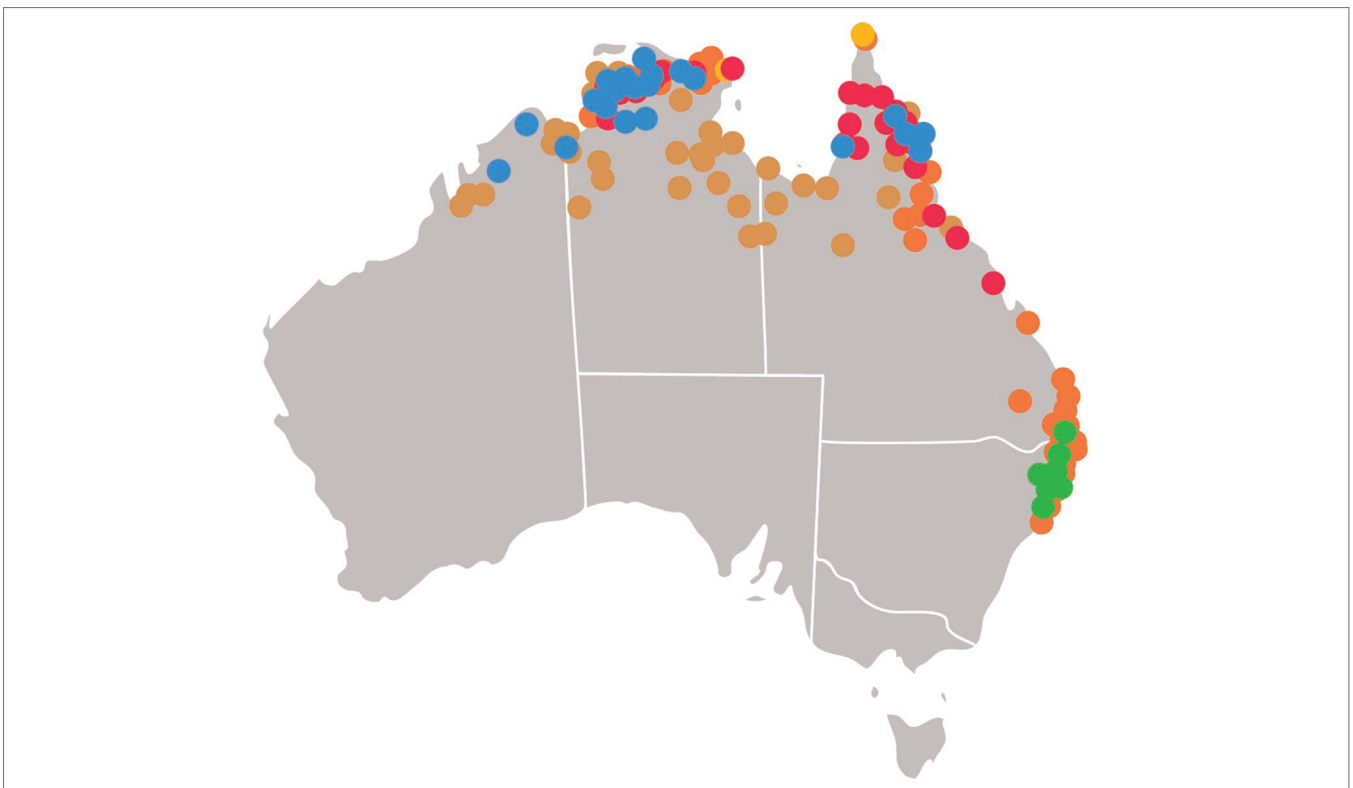


FIGURE 1 | Distribution of wild relatives of rice in Australia. Locations of collections of wild relatives of rice in Australia as detailed in the flora of Australia *(Flora of Australia* Volume 44A *Poaceae* 2, Melbourne: ABRS/CSIRO 2009). (Red) *Oryza rufipogon*, (Blue) *Oryza meridionalis*, (Yellow) *Oryza officinalis*, (Brown) *Oryza australiensis*, (Green) *Potamophila parviflora*, (Orange) *Leersia hexandra*.

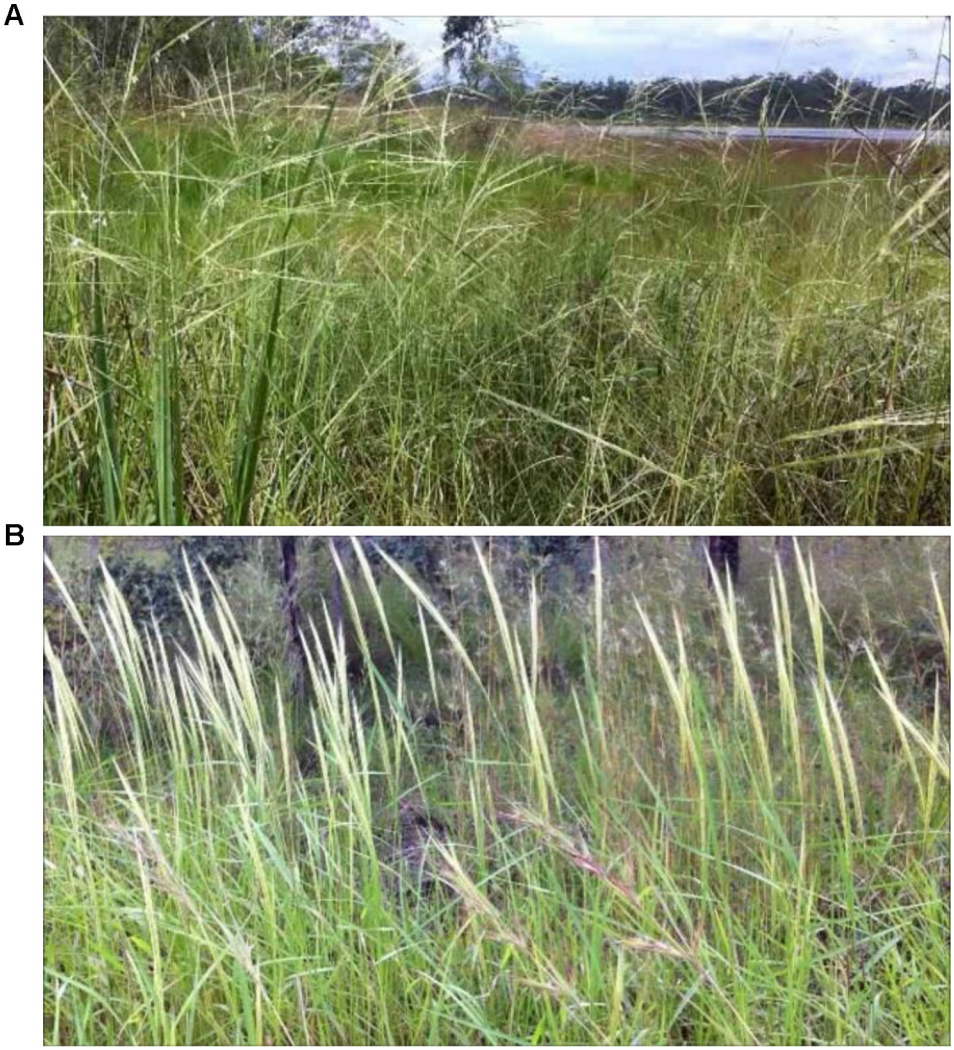


FIGURE 2 | Continued

rufipogon in the field by having smaller anthers, longer awns, and closed panicles. This species had been considered to be an annual, but more recent research has identified widespread populations of perennials in northern Queensland (Sotowa et al., 2013). The distinctness of the annual and perennial types has not been established, but both types are evidenced by the survival or otherwise, of plants from different populations, in the glasshouse after flowering (Sotowa et al., 2013).

Oryza rufipogon

O. rufipogon (**Figure 2**) was described by Griff (1851). Earlier taxonomy had included these plants in the wider group known as *Oryza perennis.* The Australian populations show significant molecular differences from Asian *O. rufipogon* and share some chloroplast sequence homology with *O. meridionalis* (Waters et al., 2012; Brozynska et al., 2014) suggesting that they might best be considered as distinct taxa. The morphological differences between Asian and Australian *O. rufipogon* are not clear with limited data from common garden experiments that would allow direct analysis of quantitative traits in the same environment. *O. rufipogon* in Australia can be most easily distinguished from *O. meridionalis* in the field by the presence of open panicle, shorter awns, and larger anthers. *O. rufipogon* and *O. meridionalis* may be confused in some Australian herbarium records with many collections being prior to the publication of *O. meridionalis*. This is further complicated by the potential for rare hybrid populations. The hybrid population has mixtures of the distinguishing traits described above. An analysis of the current distribution of these taxa and their hybrids would require sampling over a very wide area in Queensland, Northern Territory, and Western Australia.

Molecular Analysis of Australian AA Genome Species

Chloroplast Genomes

The sequence of the chloroplast genome has been widely used to determine evolutionary relationships in plants



FIGURE 2 | Australian wild rice **(A)** *Oryza.rufipogon,* **(B)** *O. meridionalis,* **(C)** *O. australiensis*, and **(D)** *Potamophila parviflora*.

(Wambugu et al., 2015). The advantages of this approach include the highly conserved nature of the chloroplast making direct comparisons of the same sequence in plants less difficult and providing a significant amount of DNA sequence to compare. Comparison of the chloroplast sequences of wild rices that are closely related to domesticated rice has defined events leading up to rice domestication.

Analysis of the chloroplasts of Australian AA genome (Nock et al., 2011) indicated that chloroplast genomes are distinct from those of other populations in Asia (Waters et al., 2012; Brozynska et al., 2014). These species have been used to develop methods for the assembly of an accurate complete chloroplast genome sequence from short read sequence data (Nock et al., 2011). These methods have used mapping to a close reference and *de*

novo assembly and rationalized the differences obtained by these two approaches (Guyeux et al., 2019).

Analysis of Australian AA genome populations confirmed the presence of two distinct chloroplast genomes that form an Australian clade that is an outgroup relative to all Asian and African populations (Moner et al., 2018). A survey of the whole chloroplast genomes of 58 genotypes of wild rice showed that the distinct Australian clade extended north to the Philippines (Moner et al., 2018). The study also confirmed the close relationships between the chloroplast sequences of *Oryza glumipatula* in South America and *Oryza longistaminata* in South Africa. This suggests inter-continental exchange of genetic material relatively recently. This exchange can be viewed as part of ongoing movement of wild rice around the tropics with birds as likely vectors. The presence of a wild rice in the Philippines with an Australian chloroplast type may be further evidence of long distance dispersal events.

Modern domesticated rices can be divided into several types including *indica*, *japonica*, and *aus* types that may have resulted from separate domestication events (Civan et al., 2015). The *indica* and *japonica* types correspond to distinct chloroplast sequences in wild populations indicating capture of these in separate domestication events. The *aus* type was found to include individuals with either wild chloroplast types (Moner et al., 2018) indicating a more complex genetic origin (Choi et al., 2017).

Nuclear Genomes

Around 3 million years ago, a divergence in the wild rice populations that are progenitors of domesticated rice and still inter-fertile with domesticated rice resulted in a lineage that became *O. meridionalis* (Moner and Henry, 2018), a species found in Australia, separating from those leading to *O. rufipogon*, the wild rice found in Asia. Analysis of the nuclear genome shows that the Australian populations of that have been considered to be *O. rufipogon* diverged from those in Asia around 1.7 million years ago (Brozynska et al., 2017). The available evidence suggests that these *O. rufipogon* like Australian populations should also be considered as a separate species. The divergence of the Asian and African wild rices in the primary gene pool of rice was shown to be more recent.

The genome of the *O. rufipogon* like Australian wild rice showed some evidence for possible introgression of chromosomal regions from wild *indica* types (Brozynska et al., 2017).

Populations with a mixture of the traits of different taxa have been identified in the wild in Australia (Moner et al., 2018). These populations have been confirmed as hybrids by examination of the chloroplast and nuclear genomes. Plant populations were identified with characteristics intermediate between those of the two AA genome taxa. The panicle varied from opened to closed and the anther length varied. The nuclear genome was divergent and, despite morphology similar to *O. meridionalis*, the chloroplast was that of the *O. rufipogon* type. Field collections of 29 populations (Moner et al., 2018) from northern Queensland have been made from the northern to the southern extremes of the current natural distribution. These collections are the subject of ongoing analysis by DNA sequencing.

Comparison with the genome of domesticated rice suggest that the Australian wild species have significant diversity, even in chromosomal regions of very low diversity ("genome deserts") in the domesticated rice gene pool (Krishnan et al., 2014), making them import sources of novel diversity for rice improvement. These regions of low diversity suggest that selective sweeps may have occurred in wild populations, possibly due to pests or diseases, and that these areas of the genome may have been further depleted of variation by bottlenecks during domestication.

Other Related Species

Other more distant relatives of rice are also found in the Australian flora. They have generally received less research attention than the Australian members of the AA genome clade. These more divergent species include *Oryza* species, *Oryza officinalis*, and *Oryza australiensis* and other species from within the tribe *Oryzeae* such as *Potamophila* (Abedinia et al., 1998) and Leersia and other species from the subfamily *Ehrhartoideae* such as *Microlaena stipoides* (Shapter et al., 2013). These more diverse species also have genes that may be useful in rice improvement with more difficulty. However, despite it being a more distant relative of rice, *O. australiensis* has already been used successful as a genetic resource in rice improvement. Techniques such as gene editing may allow more rapid deployment of useful novel alleles, from all of these wild relatives, in the domesticated gene pool.

Oryza officinalis

O. officinalis (CC genome) has been reported only in the extreme north of Australia on the mainland and on Moa Island in the Torres Strait. Collect of this species has not been reported recently, and the status of the Australian populations is uncertain. The author was unable to locate populations on Moa Island during a visit in 2016. This species is part of the CC genome group of *Oryza* species. Populations reported in Australia may not be permanently established and may be the result of occasional migration from Asia suggesting that the current presence of this species in Australia requires confirmation.

Oryza australiensis

O. australiensis (EE genome; **Figure 2**) is widespread across northern Australia. This species is usually perennial and grows in seasonally dry environments as it can survive the dry season as a rhizome. This species is often found in areas surrounding permanent water holes or lakes that support perennial AA genome *Oryza* or temporary bodies of water in depressions supporting annual AA genome *Oryza* in the wet season but also extending further away from the surface water on waterlogged soils beyond these water bodies. The plants can be very tall reaching 3 m in height but can also be observed as very small plants in environments that limit growth such as populations growing on subsoil in roadside drains (personal observations).

Potamophila parviflora

River grass *(Potamophila parviflora*; **Figure 2***)* is found in just a few rivers (Manning, Hastings, Clarence, and Richmond Rivers)

on the central east coast of Australia in northern New South Wales (Abedinia et al., 1998). It appears to grow only in the shallow running water of these rivers. This monotypic genus is probably more closely related to *Zizania* of North America than to *Oryza*. Cold adaptation and the presence of separate male and female flowers are potentially useful traits. This species grows just south of the border between Queensland and New South Wales but has not been reported in Queensland despite some searching by the author and others.

Leersia hexandra

Leersia hexandra is a perennial found in northern and eastern Australia from the Northern Territory to north-eastern New South Wales. The use of species of *Leersia* as food has not been documented. Plants have recently been collected in urban areas close to Brisbane. The exotic *Leersia oryzoides* is also found in Australia.

Microlaena stipoides

Weeping grass (*M. stipoides*) is found throughout temperate Australia demonstrating adaptation to cooler climates. Attempts have been made to domesticate this species as a cool climate dryland rice crop. Mutagenesis (Shapter et al., 2013; Malory et al., 2011) has allowed recovery of non-shattering types and this together with selection for large grain has generated genotypes with potential for grain production. This widespread species is often found in shaded environments and has been used as a lawn grass.

Useful Traits for Rice Improvement in Wild Australian Populations

The capture of genes from *O. meridionalis* has been successful because, although the F1 hybrids with *O. sativa* have pollen sterility, they retain female fertility allowing backcrossing of desirable traits into domesticated rice (Yoshimura et al., 2010). Potentially useful traits include grain yield, grain size, grain color, starch properties, cooking qualities, eating qualities, and disease resistances. The genes responsible for these traits have not been characterized to date, but the availability of genome sequence data and more intensive collections of the populations of these taxa will facilitate the identification of useful genes and alleles.

Yield

The yield of Australian wild rices in cultivation has not been evaluated. The difficulty of cultivating a species that is so prone to shattering has made this difficult. The abundant grain observed on plants in the wild suggests that good yields might be achieved if this problem was overcome. The large grain size of some Australian wild rices may be a trait that could contribute to yield. Yield evaluation in comparative trials has not been conducted for any of these species.

Grain Quality

The grains of Australian wild rice vary from medium to long grain and include some with large grain size (Tikapunya et al., 2017). Transfer of this trait to domesticated rice is in progress. The grains of the A genome species are long or medium while *O. australiensis* grains are short. The wild rice grains are all slender when compared to domesticated rice. The impact of nutrition on grain filling and grain size has not been examined for these species, and improved grin size may be possible with optimal plant nutrition.

Australian wild rice has a color that would be described as red in rice color language. The Australian populations have colors that appear to be distinct from that of Asian red rices that have been available for comparisons. This suggests that a key market option for these species is as colored rice that has not been milled.

The starch properties of rice reflect the eating and nutritional qualities of the grain (Wang et al., 2015). The Australian species have unique starch properties (Kasem et al., 2012–Tikapunya et al., 2017). All taxa (including AA genome species and *O. australiensis*) had high amylose contents and gelatinization temperatures. Pasting properties varied. Analysis of starch structure suggested the presence of shorter chain amylose in *O. australiensis* (Tikapunya et al., 2017). Much remains to be understood about the novel starch properties of these rice species and the extent of variation in wild populations. The genetic basis of the starch properties is also not determined.

Sensory evaluation of Australian wild rice has indicated acceptable eating qualities (Tikapunya et al., 2018). Australian wild rices have been compared with Asian domesticated rices (including long grain, medium grain, basmati, red basmati, and red rice) and with Canadian wild rice (*Zizania palustris*). Specific descriptors were developed for sensory evaluation of these rices. The Australian wild rice was similar to the red rice and red basmati having a mild aroma and flavor but without the lingering aftertaste. The wild rice had a firmer texture and required a longer cooking time. The overall cooking profiles, sensory, and physical attributes suggest and that Australian wild rice has potential for commercialization in the colored rice market and may be a useful genetic resource for rice breeding.

Disease Resistance

Production of domesticated Asian rice in northern Australia has been limited by a high incidence of disease relative to rice production further south in Australia. The southern production areas have been well outside the natural range of rice and rice pathogens. Rice and associated rice pathogens (Khemmuk et al., 2016) are native to northern Australia and a wide range of pathogen species are present. Much of the diversity of these organisms remains to be explored. Native Australian rices have resistance to the local strains of rice pathogens and should provide a useful source of resistance to local diseases for use in breeding rice varieties for this region. Asian domesticated rice varieties have not been developed to have resistance to the native rice pathogens of northern Australia. The wild populations are key resources for the successful establishment of rice production on the large areas of land that are suitable for production in northern Australia. Crops of domesticated rice have been devastated by disease in Northern Australia while surrounded by wild populations that appear resistant.

Human Impact on Populations

The wild populations are extensive in the extreme north but are threatened by developments including the improvements

of roads and the resulting more rapid spread of invasive weeds into new areas, associated with increased traffic. Aquatic weeds are displacing wild populations at some locations including protected areas such as national parks. Greater awareness of the presence of these populations and their potential to contribute to global food security is necessary to ensure adequate measures are taken to protect the wild populations.

The Australian landscapes in which rice is found in Australia have been inhabited by humans for a very long time. The way in which rice was used as food is not well documented. It is possible that human use has impacted on the genetics of wild rice populations in Australia. The large grain size of some Australian populations (Sotowa et al., 2013) might reflect some human selection. Shattering is a key trait associated with domestication of grains. The wild species are all extremely prone to shattering suggesting little progress toward a non-shattering domesticated type. However, recent evidence shows that domesticated populations can rapidly revert to shattering types under natural selection (Zeng et al., 2018).

Potential for New Domestications

In addition to their use in rice improvement, the Australian *Oryza* could be targeted for direct domestication as new rice types. The key trait to be overcome in domestication is probably shattering, with wild populations all demonstrating high shattering characteristics. The option of domestication needs to be considered along with the potential to source many genes for useful traits in these populations for transfer into domesticated rice. The traits that are most desirable are disease resistance, grain size, grain color, and grain quality (starch properties). The strategy of new domestication may be especially useful in developing rice varieties adapted to new climates with advancing or rapid climate change especially with the availability of extensive genomics resources for rice (Abberton et al., 2016).

CONCLUSIONS

Wild rice populations in Australia include species at varying distances from domesticated rice. These provide a rich resource for use in rice improvement. Both novel alleles and novel traits might be sourced from the Australian wild genotypes. Increased efforts to conserve these genetic resources would be justified. The current very limited *ex situ* collections in seed banks need to be supplemented by much more extensive collections. The *in situ* conservation of these species would be aided by greater awareness of the populations and their importance and may involve efforts to control weeds invading the aquatic habitats of the wild *Oryza*. Revision of the taxonomy of the A genome species is suggested by the available molecular evidence. Clarification of the taxonomy may aid conservation efforts by ensuring efforts to conserve rarer populations, including suspected hybrid populations. The growing genomic resources for the *Oryza* genus (Stein et al., 2018) would be extended usefully by more efforts on the relatively poorly characterized Australian members of this group especially as they include the more divergent members of the A genome clade.

AUTHOR CONTRIBUTIONS

This manuscript is the work of the author.

FUNDING

The Australian Research Council provided support for much of the research reviewed here.

REFERENCES

Abberton, M., Batley, J., Bentley, A., Bryant, J., Cai, H., Cockram, J., et al. (2016). Global agricultural intensification during climate change: A role for genomics. *Plant Biotechnol*. J. 14 (4), 1095–1098. doi: 10.1111/pbi.12467

Abedinia, M., Henry, R., and Clark, S. (1998). Distribution and phylogeny of Potamophila parviflora RBr a wild relative of rice from eastern Australia. *Genet. Resour. Crop Evol.* 45 (5), 399–406. doi: 10.1023/A:1008612115685

Brozynska, M., Omar, E. S., Furtado, A., Crayn, D., Simon, B., Ishikawa, R., et al. (2014). Chloroplast genome of novel rice germplasm identified in Northern Australia. *Trop. Plant Biol.* 7 (3-4), 111–120. doi: 10.1007/s12042-014-9142-8

Brozynska, M., Copetti, D., Furtado, A., Wing, R. A., Crayn, D., Fox, G., et al. (2017). Sequencing of Australian wild rice genomes reveals ancestral relationships with domesticated rice. *Plant Biotechnol. J.* 15 (6), 765–774. doi: 10.1111/pbi.12674

Choi, J. Y., Platts, A. E., Fuller, D. Q., Hsing, Y. L., Wing, R. A., and Purugganan, M. D. (2017). The rice paradox: multiple origins but single domestication in Asian rice. *Mol. Biol. Evol.* 34 (4), 969–979. doi: 10.1093/molbev/msx049

Civáň, P., Craig, H., Cox, C., and Brown, T. (2015). Three geographically separate domestications of Asian rice. *Nature Plants* 1 (11), 15164. doi: 10.1038/nplants.2015.164

Fujino, K., Hirayama, Y., Obara, M., and Ikegaya, T. (2019). Introgression of the chromosomal region with the Pi-cd locus from Oryza meridionalis into O. sativa L. during rice domestication. *Theor. Appl. Genet.* 132 (7), 1981–1990. doi: 10.1007/s00122-019-03332-1

Guyeux, C., Charr, J. C., Tran, H. T. M., Furtado, A., Henry, R. J., Crouzillat, D., et al. (2019). Evaluation of chloroplast genome annotation tools and application to analysis of the evolution of coffee species. *PLoS ONE* 14 (6), E0216347. doi: 10.1371/journal

Henry, R., Rice, N., Waters, D., Kasem, S., Ishikawa, R., Hao, Y., et al. (2010). Australian Oryza: utility and conservation. *Rice* 3 (4), 235–241. doi: 10.1007/s12284-009-9034-y

Huang, X., Kurata, N., Wei, X., Wang, Z. X., Wang, A., Zhao, Q., et al. (2012). A map of rice genome variation reveals the origin of cultivated rice. *Nature* 490 (7421), 497–501. doi: 10.1038/nature11532

Kasem, S., Waters, D. L. E., and Henry, R. J. (2012). Analysis of starch gene diversity in the wild relatives of Oryza sativa. *Trop. Plant Biol.* 5 (4), 286–308. doi: 10.1007/s12042-012-9111-z

Khemmuk, W., Shivas, R. G., Henry, R. J., and Geering, A. D. W. (2016). Fungi associated with foliar diseases of wild and cultivated rice (Oryza spp.) in northern Queensland. *Australas. Plant Pathol.* 45 (3), 297–308. doi: 10.1007/s13313-016-0418-3

Krishnan, S. G., Waters, D. L. E., and Henry, R. J. (2014). Australian wild rice reveals pre-domestication origin of polymorphism deserts in rice genome. *PloS One* 9 (6), e98843. doi: 10.1371/journal.pone.0098843

Malory, S., Shapter, F. M., Elphinstone, M. S., Chivers, I. H., and Henry, R. J. (2011). Characterizing homologues of crop domestication genes in poorly described wild relatives by high-throughput sequencing of whole genomes. *Plant Biotechnol. J.* 9 (9), 1131–1140. doi: 10.1111/j.1467-7652.2011.00640.x

Mondal, T. K., and Henry, R. J. (2018). *The Wild Oryza Genomes (Compendium of Plant Genomes)*. Cham: Springer International Publishing. doi: 10.1007/978-3-319-71997-9

Moner, A. M., and Henry, R. J. (2018). Oryza meridionalis N.Q.Ng. In: Mondal T., Henry R. (eds) *The Wild Oryza Genomes. Compendium of Plant Genomes.* Cham: Springer. 177–182. doi: 10.1007/978-3-319-71997-9_16

Moner, A. M., Furtado, A., Chivers, I., Fox, G., Crayn, D., and Henry, R. J. (2018). Diversity and evolution of rice progenitors in Australia. *Ecol. Evol.* 8 (8), 4360–4366. doi: 10.1002/ece3.3989

Moner, A. M., Furtado, A., and Henry, R. J. (2018). Chloroplast phylogeography of AA genome rice species. *Mol. Phylogenet. Evol.* doi: 10.1016/j.ympev.2018.05.002

Nock, C., Waters, D., Edwards, M., Bowen, S., Rice, N., Cordeiro, G., et al. (2011). Chloroplast genome sequences from total DNA for plant identification. *Plant Biotechnol. J.* 9 (3), 328–333. doi: 10.1111/j.1467-7652.2010.00558.x

Shapter, F. M., Henry, R. J., and Lee, L. S. (2008). Endosperm and starch granule morphology in wild cereal relatives. *Plant Genet. Resour.* 6 (2), 85–97. doi: 10.1017/S1479262108986512

Shapter, F. M., Cross, M., Ablett, G., Malory, S., Chivers, I. H., King, G. J., et al. (2013). High-throughput sequencing and mutagenesis to accelerate the domestication of microlaena stipoides as a new food crop. *PLoS One.* 8 (12), E82641. doi: 10.1371/journal.pone.0082641

Sotowa, M., Ootsuka, K., Kobayashi, Y., Yin, H., Tanaka, K., Ichitani, K., et al. (2013). Molecular relationships between Australian annual wild rice, Oryza meridionalis, and two related perennial forms. *Rice,* 6 (1), 1–19. doi: 10.1186/1939-8433-6-26

Stein, J. C., Yu, Y., Copetti, D., Zwickl, D. J., Zhang, L., Zhang, C. J., et al. (2018). Genomes of 13 domesticated and wild rice relatives highlight genetic conservation, turnover and innovation across the genus Oryza. *Nat. Genet.* 50 (2), 285–296. doi: 10.1038/s41588-018-0040-0

Tikapunya, T., Fox, G., Furtado, A., and Henry, R. (2017). Grain physical characteristic of the Australian wild rices. *Plant Genet. Resour.* 15 (5), 409–420. doi: 10.1017/S1479262116000083

Tikapunya, T., Zou, W., Yu, W., Powell, P. O., Fox, G. P., Furtado, A., et al. (2017). Molecular structures and properties of starches of Australian wild rice. *Carbohydr. Polym.* 172, 213–222. doi: 10.1016/j.carbpol.2017.05.046

Tikapunya, T., Henry, R. J., and Smyth, H. (2018). Evaluating the sensory properties of unpolished Australian wild rice. *Food Res. Int.* 103, 406–414. doi: 10.1016/j.foodres.2017.10.037

Wambugu, P. W., Brozynska, M., Furtado, A., Waters, D. L., and Henry, R. J. (2015). Relationships of wild and domesticated rices (Oryza AA genome species) based upon whole chloroplast genome sequences. *Sci. Rep.* 5, 13957. doi: 10.1038/srep13957

Wambugu, P. W., Nyamongo, D., Ndjiondjop, M. N., and Henry, R. J. (2018). Evolutionary relationships among the Oryza species. *Compend. The Wild Oryza Genomes (Compendium of Plant Genomes).* Cham: Springer International Publishing. 41–54. doi: 10.1007/978-3-319-71997-9_3

Wang, K., Wambugu, P. W., Zhang, B., Wu, A. C., Henry, R. J., and Gilbert, R. G. (2015). The biosynthesis, structure and gelatinization properties of starches from wild and cultivated African rice species (Oryza barthii and Oryza glaberrima). *Carbohydr. Polym.* 129, 92–100. doi: 10.1016/j.carbpol.2015.04.035

Waters, D., Nock, C., Ishikawa, R., Rice, N., and Henry, R. (2012). Chloroplast genome sequence confirms distinctness of Australian and Asian wild rice. *Ecol. Evol.* 2, 211– 217. doi: 10.1002/ece3.66

Yoshimura, A., Nagayama, H., Sobrizal, Kurakazu, T., Sanchez, P. L., Doi, K., et al. (2010). Introgression lines of rice (Oryza sativa L.) carrying a donor genome from the wild species, O. glumaepatula Steud. and O. meridionalis Ng. *Breed. Sci.* 60 (5), 597–603. doi: 10.1270/jsbbs.60.597

Zeng, X., Guo, Y., Xu, Q., Mascher, M., Guo, G., Li, S., et al. (2018). Origin and evolution of qingke barley in Tibet. *Nat. Commun.* 9 (1), 5433. doi: 10.1038/s41467-018-07920-5

Conflict of Interest: 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 © 2019 Henry. 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) and the copyright owner(s) 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 Plant Science

ORIGINAL RESEARCH
published: 08 November 2019
doi: 10.3389/fpls.2019.01406

Inferring the Origin of Cultivated *Zizania latifolia*, an Aquatic Vegetable of a Plant-Fungus Complex in the Yangtze River Basin

*Yao Zhao[1,2], Zhiping Song[2], Lan Zhong[3], Qin Li[1,2], Jiakuan Chen[2] and Jun Rong[1]**

[1] Jiangxi Province Key Laboratory of Watershed Ecosystem Change and Biodiversity, Center for Watershed Ecology, Institute of Life Science and School of Life Sciences, Nanchang University, Nanchang, China, [2] The Ministry of Education Key Laboratory for Biodiversity Science and Ecological Engineering, Institute of Biodiversity Science, Fudan University, Shanghai, China, [3] Institute of Vegetable, Wuhan Academy of Agriculture Science and Technology, Wuhan, China

Crop domestication is one of the essential topics in evolutionary biology. Cultivated *Zizania latifolia*, domesticated as the special form of a plant-fungus (the host *Zizania latifolia* and the endophyte *Ustilago esculenta*) complex, is a popular aquatic vegetable endemic in East Asia. The rapid domestication of cultivated *Z. latifolia* can be traced in the historical literature but still need more evidence. This study focused on deciphering the genetic relationship between wild and cultivated *Z. latifolia*, as well as the corresponding parasitic *U. esculenta*. Twelve microsatellites markers were used to study the genetic variations of 32 wild populations and 135 landraces of *Z. latifolia*. Model simulations based on approximate Bayesian computation (ABC) were then performed to hierarchically infer the population history. We also analyzed the ITS sequences of the smut fungus *U. esculenta* to reveal its genetic structure. Our results indicated a significant genetic divergence between cultivated *Z. latifolia* and its wild ancestors. The wild *Z. latifolia* populations showed significant hierarchical genetic subdivisions, which may be attributed to the joint effect of isolation by distance and hydrological unconnectedness between watersheds. Cultivated *Z. latifolia* was supposedly domesticated once in the low reaches of the Yangtze River. The genetic structure of *U. esculenta* also indicated a single domestication event, and the genetic variations in this fungus might be associated with the diversification of cultivars. These findings provided molecular evidence in accordance with the historical literature that addressed the domestication of cultivated *Z. latifolia* involved adaptive evolution driven by artificial selection in both the plant and fungus.

Keywords: domestication, *Zizania latifolia*, *Ustilago esculenta*, genetic structure, approximate Bayesian computation

OPEN ACCESS

Edited by:
Eric Von Wettberg, University of Vermont, United States

Reviewed by:
Michael Benjamin Kantar, University of Hawaii, United States
Zhe Cai, Institute of Botany, China

***Correspondence:**
Jun Rong
rong_jun@hotmail.com

Specialty section:
This article was submitted to Plant Breeding, a section of the journal Frontiers in Plant Science

Received: *24 June 2019*
Accepted: *10 October 2019*
Published: *08 November 2019*

Citation:
Zhao Y, Song Z, Zhong L, Li Q, Chen J and Rong J (2019) Inferring the Origin of Cultivated Zizania latifolia, an Aquatic Vegetable of a Plant-Fungus Complex in the Yangtze River Basin. Front. Plant Sci. 10:1406. doi: 10.3389/fpls.2019.01406

INTRODUCTION

Cultivated plants are fundamental materials for ensuring the survival and development of human beings (Hajjar and Hodgkin, 2007; Ronald, 2014; FAO, 2019). Through domestication, humans are able to change the morphological and physiological performances of wild plants. With ongoing selective pressure, wild plants would finally adapt to human-derived cultivation environments

and provide products that humans demand (Zeder et al., 2006; Gross and Olsen, 2010; Giovannoni, 2018). Plant domestication is regarded as a complex plant-animal co-evolution process, and it is an ideal model to study evolution (Purugganan and Fuller, 2009; Tufto, 2017; Whitehead et al., 2017; Zeder, 2017). As a typical scenario of recent rapid evolution, domestication has demonstrated how a species diverges from its ancestor through artificial selection (Zeder et al., 2006; Purugganan and Fuller, 2009; Diez et al., 2015; Wang et al., 2018).

For a cultivated plant, it is of great importance to decipher the puzzle of its domestication to understand the genetic basis for further improvements. Regardless of the domesticated species we study, there are several questions that need to be answered: what the ancestor is, when and where the domestication occurred, and what happened during the wild-to-cultivated transition (Doebley et al., 2006; Zeder, 2015). Benefitting from cumulative archaeological evidence and the advances of molecular biology, these questions appear to be well resolved for some major crops (Doebley et al., 2006). For example, archaeological and genetic evidence indicated that maize (*Zea mays*) had been domesticated once from its ancestor species teosinte (*Zea mexicana*) in Mexico (reviewed in van Heerwaarden et al., 2011), and common bean (*Phaseolus vulgaris*) had more than two independent domestication events (reviewed in Bitocchi et al., 2012). However, the origins of well-studied crops still may be under debate due to the uncertainty and complexity of historical human activities, such as Asian cultivated rice (*Oryza sativa*) (Ge and Sang, 2011; Huang et al., 2012; Choi and Purugganan, 2018) and a few grain crops generally thought to have originated from the Fertile Crescent (*Hordeum vulgare*, *Triticum aestivum*, and *Avena sativa*) (Fuller et al., 2012; Heun et al., 2012). Moreover, the answers to the domestication puzzles for a given species could also be blurred by post domestication migrations or genetic improvements (eg. cross-regional introduction, diversification of varieties, and hybridizations with wild progenitors) (Burger et al., 2008; Diez et al., 2015). Therefore, cultivated plants may be domesticated in their own way, which should be dissected independently.

The genus *Zizania* belongs to the rice tribe (Poaceae, Oryzeae), and it is an aquatic genus with a discontinuous distribution between eastern Asia and North America (Wen, 1999; Wu et al., 2006). Of the four species in the genus, two are field crops with their seeds harvested for food (the annual *Zizania palustris* native to North America)or enlarged young shoots as vegetable (the perennial *Zizania latifolia* native to Asia) (Oelke, 1993; Xu et al., 2010). Because the young shoots of *Z. latifolia* become swollen, soft, and edible after being infected by the obligatory smut fungus *Ustilago esculenta* and have high nutritional and economic value, it had been domesticated as a vegetable called Jiaobai which is now widely cultivated in the Yangtze River Basin (Guo et al., 2007; Ke, 2008). The cultivated *Z. latifolia* is highly domesticated. The phenotypic differences between wild and cultivated *Z. latifolia* are distinct: the domesticated plant usually has erected plant architecture, broader leaves and shortened internodes; and it could generate much larger swollen shoots containing few black spots (Guo et al., 2007). Once *Z. latifolia* is infected by *U. esculenta*, it loses the ability to undergo sexual reproduction *via* seeds because the fungus hinders the development of the flower primordium. Thus, the breeding of new accessions can only rely on the natural somatic mutants in tillers or rhizomes. In each generation, individuals with good-quality products (swollen shoots) should be screened out for clonal propagation to ensure the stability of agronomic traits of offspring (Guo et al., 2007). The plants that generate flowers or those with too many black spots in the swollen shoots are removed during harvest. Although precautions have been made, unfavorable mutants always occur in the fields, which may be attributed to the high variability of the smut fungus (Guo et al., 2007). To date, there are two main ecotypes including more than 100 local varieties with distinct phenotypic differences. One ecotype is a single-season plant that only can be harvested once each year in the fall (from August to November). This ecotype is a strictly short-day plant that can produce swollen shoots only when the days become short in fall. The other is a double-season crop that is planted in the spring and then can be harvested twice in the fall and the summer thereafter, and it is insensitive to light and endemic to the Yangtze River Basin. The single-season ecotype is taller and has fewer underground rhizomes, lower yield, and a wider distribution relative to the double-season ecotype.

Historical records can help to trace the domestication process of *Z. latifolia* in China (Ke, 2008). The oldest record showed that grains of *Z. latifolia* had been harvested to offer as tribute to the nobles in the Zhou Dynasty (from 771 to 221 BC). Until the Tang Dynasty (618–907 AD), *Z. latifolia* was famous for the tender taste of its grains. During the period between the Zhou and Tang Dynasties, the swollen shoots were rarely reported to be used for food. Only in the low reaches of the Yangtze River was it occasionally eaten by local residents. After the Tang Dynasty, people no longer harvested the grains of *Z. latifolia*. Alternatively, *Z. latifolia* was domesticated to be an aquatic vegetable (Guo et al., 2007; Chen et al., 2012; Chen et al., 2017). Although the utilization history of *Z. latifolia* has provided important clues, several blanks still remain. It was possible that *Z. latifolia* underwent pre-domestication during a long-time use as a food. In the meantime, *U. esculenta* acted as a plant disease during grain production. Then, the plant-fungus complex was domesticated. In another scenario, this vegetable was directly domesticated once or twice after the Tang Dynasty and then generated two main ecotypes with many cultivated varieties. However, these scenarios are just assumptions, and they still lack solid molecular evidence.

Cultivated *Z. latifolia* is the one and only crop that was domesticated in the form of a plant-fungus complex. Undoubtedly, its domestication was associated with the responses of the plant and the fungus to artificial selection. Thus, the cultivated *Z. latifolia* could not only provide a unique case of domestication of a plant-fungus complex but also an ideal model to investigate changes in the genetic interactions of two closely linked species under artificial selective pressures. The plant and the fungus have been independently studied. Wild populations of *Z. latifolia* are widely distributed across eastern China. This plant is of importance as a genetic resource for breeding and forage (Chen et al., 2006). Many studies have been conducted on *Z. latifolia*, including investigation of its phylogeny (Ge et al., 2002; Guo and Ge, 2005; Xu et al., 2010), ecology (Yang et al., 1999), population structure (Xu et al., 2008; Chen et al., 2012), utilization as a tertiary gene pool of rice (Liu et al., 1999), nutritional value (Zhai et al., 2001), cultivar

classification, and breeding (Guo et al., 2007). The phylogeny and cytology of *U. esculenta* have been investigated as well (Stoll et al., 2005; Zhang et al., 2012). The genetic interactions between plant and fungus were recently revealed by whole-genome sequencings and comparative transcriptome analyses (Guo et al., 2015; Wang et al., 2017; Ye et al., 2017; Zhang et al., 2017). However, the origin of cultivated *Z. latifolia* has seldom been genetically addressed and remains unknown.

This study focused on dissecting the genetic structure of *Z. latifolia* and the parasitic *U. esculenta*. Twelve microsatellites markers were used to investigate the genetic variations in *Z. latifolia* populations, and model simulations with approximate Bayesian computation (ABC) were then performed to infer the domestication history. The pattern of genetic variations for *U. esculenta* was identified based on ITS sequence analyses. By combining the genetic structures of host and endophyte, the domestication scenario of cultivated *Z. latifolia* could be unveiled.

MATERIALS AND METHODS

Plant Collection

Thirty-two natural populations of *Z. latifolia* were collected across the species distribution range in China (**Figure 1**, **Table 1**). in each population, the individuals were randomly sampled with an interval of 10 M between plants to reduce the chance of collecting samples from the same genotype. Fresh young leaves were collected from each plant and placed into plastic bags containing silica gel for dehydration. A total of 824 samples of wild *Z. latifolia* were finally obtained. The leaf samples of 135 landraces of cultivated *Z. Latifolia* (the detail information of landraces were attached in **Supplementary Table 2**) were provided by the National Aquatic Vegetable Germplasm Nursery of Chinese Academy of Agriculture Science, Wuhan.

Infected Plant Collection and Fungi Isolation

Among the 32 wild populations, we found five individuals infected by the smut fungus *U. esculenta*. A total of 37 strains of *U. esculenta* were isolated from 5 wild accessions and 32 cultivars following the method described by Zhang et al. (2012). These strains were grown on potato dextrose agar (PDA) slant at 26°C and stored in a 10°C incubator.

DNA Extraction and PCR Assays

Total genomic DNA was extracted following the protocol described by Song et al. (2006). Twelve microsatellites were selected from 100 Asian cultivated rice microsatellites (www.gramene.org) and 16 *Z. latifolia* specific microsatellites (Quan et al., 2009) (**Table S1**). The PCR products were labeled with fluorescent dyes using 5′-tagged forward primers (FAM, JOE,

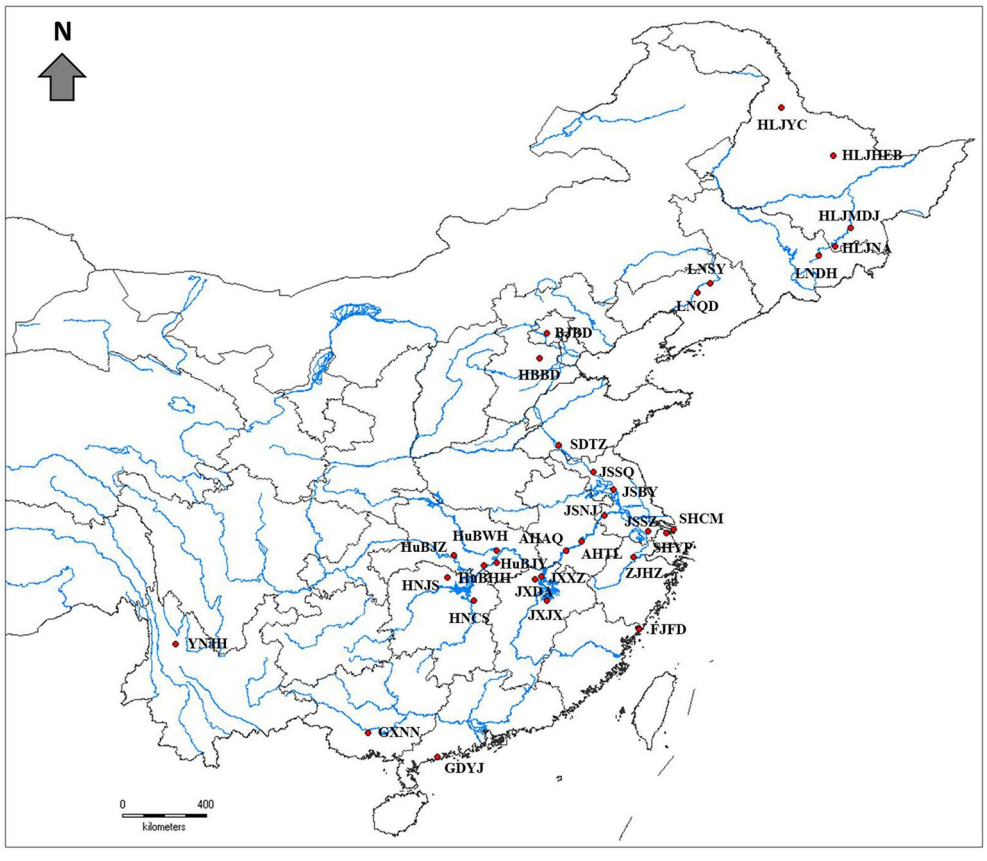


FIGURE 1 | Geographical locations of 32 wild *Zizania latifolia* populations in China (red filled circles).

TABLE 1 | Parameters of genetic diversity of *Zizania latifolia* populations and cultivars based on 12 microsatellites.

Code	Population	*N*	A_r	A_e	H_o	H_e	*F*	N_e	*Ni*	*Q (> 0.05)*	*Bottleneck*	*Habitat*
1	HLJYC	30	2.909	1.979	0.383	0.383	0.022	226				Pond
2	HLJHEB	6	3.182	2.402	0.273	0.563	**0.450**	152		0.054	S	River
3	HLJMDJ	26	2.727	1.647	0.270	0.310	**0.188**	210				Pond
4	HLJNA	30	3.273	2.078	0.415	0.457	0.069	191			Y	Lake
5	JLDH	42	3.273	1.965	0.346	0.464	**0.275**	305				River
6	LNSY	30	4.545	2.895	0.483	0.585	0.111	233				River
7	LNQD	28	4.818	2.653	0.478	0.539	0.051	276				Stream
8	BJHD	40	2.000	1.407	0.243	0.229	–0.049	256				Stream
9	HBBD	12	3.091	2.177	0.269	0.494	**0.407**	171				Marsh
10	SDTZ	34	3.727	2.140	0.395	0.472	0.108	254				Marsh
11	JSBY	39	5.727	2.893	0.567	0.588	–0.004	312				Lake
12	JSSQ	25	4.000	2.059	0.449	0.425	–0.068	255				River
13	JSNJ	14	2.818	2.116	0.483	0.487	0.013	169			S	Pond
14	JSSZ	30	4.273	1.972	0.524	0.441	–0.102	251				Lake
15	SHYP	18	1.636	1.544	0.533	0.284	**–0.816**	172			Y, S	Stream
16	SHCM	12	2.818	2.538	0.618	0.570	–0.114	150	4	0.330	Y, S	Stream
17	ZJHZ	27	3.455	1.973	0.461	0.438	0.001	192	3	0.114		Marsh
18	AHTL	14	2.091	1.705	0.571	0.371	**–0.468**	208			Y, S	Pond
19	AHAQ	33	4.545	2.493	0.498	0.546	0.058	280				Marsh
20	JXXZ	15	4.364	2.660	0.467	0.511	–0.004	187				Lake
21	JXDA	29	1.818	1.302	0.250	0.167	**–0.170**	210				Stream
22	JXJX	39	3.364	2.264	0.461	0.465	–0.039	304			Y	Stream
23	HuBWH	31	3.545	2.240	0.538	0.505	–0.098	278			Y	Lake
24	HuBJY	27	3.364	2.397	0.477	0.439	–0.133	231				Stream
25	HuBHH	32	4.727	2.545	0.542	0.546	–0.038	297				Marsh
26	HuBJZ	33	4.455	2.632	0.494	0.526	0.030	278				Stream
27	HuNJS	38	3.000	1.994	0.589	0.459	–0.222	273				Pond
28	HuNCS	35	1.455	1.450	0.434	0.229	**–0.918**	120	35	0.994	Y, S	Pond
29	FJFD	14	1.818	1.676	0.539	0.304	**–0.868**	112	14	0.987	Y, S	Stream
30	GDYJ	15	1.636	1.636	0.636	0.329	**–1.000**	150			Y, S	Stream
31	GXNN	15	1.636	1.636	0.636	0.329	**–1.000**	158			Y, S	Stream
32	YNJH	11	2.182	2.013	0.355	0.374	0.023	137	5	0.471	Y, S	Lake
Overall			3.196	2.096	0.459	0.432	–0.135					
SD			1.118	0.434	0.113	0.112	0.383					
	Cultivars	135	2.182	1.519	0.455	0.242	**–0.751**	100	—	0.997	Y	

N, sample size; A_r, allelic richness; A_e, number of effective alleles; H_o, observed heterozygosity; H_e, expected heterozygosity; F, fixation index; N_e, the effective population size; Ni, the number of individuals that genetically identical to cultivars; Q, the introgression coefficient; the significant bottleneck effect is labeled by Y (TPM model) and S (Mode shift). The F values that significantly deviated from 0 are shown in bold font.

and ROX) and GS350 as an internal size standard labeled with TAMRA. The PCR products were sequenced on an ABI 3730 (Applied Biosystems) automated sequencer. The resulting chromatograms were visualized and analyzed using GENEMAPPER v4.0 software (Applied Biosystems).

The ITS region of smut fungus was amplified using ITS1 and ITS4. PCR products were purified and sequenced on an ABI 3130XL capillary sequencer (Applied Biosystems). All ITS sequences were deposited into GeneBank, with accession numbers MK811211-MK811247.

Genetic Data Analyses

The frequencies of putative null alleles were estimated using the software FREENA (Chapuis and Estoup, 2007). Null alleles were detected in 34 of 384 population-locus combinations, and the presence of null alleles was not associated with particular populations or loci. We did not find a significant bias in the estimations of F_{st} based on the data of all loci compared to the estimate based on corrected genotype data for null alleles (*t*-test, $P = 0.419$). The parameters of population genetic variations were estimated by expected heterozygosity (H_e), observed heterozygosity (H_o) and Wright's fixation index *F* using GENALEX 6.5 (Peakall and Smouse, 2012). We also calculated the mean effective allele number (A_e) and allelic richness (A_r). Hardy–Weinberg equilibrium departures were tested using exact tests implemented in GENEPOP v4.0 for each population–locus combination. A global test across loci for departures from Hardy–Weinberg equilibrium was conducted using Fisher's method. Significant deviations were further evaluated using the sequential Bonferroni test.

BOTTLENECK (Piry et al., 1999) was used to screen each population for bottleneck signatures over the last two to four N_e generations or in a recent period. A significant heterozygote excess or deficit due to a bottleneck was assessed by a two-tailed Wilcoxon signed rank test with 1,000 iterations under the assumption of a stepwise mutation model. At the same time, the distribution of allelic frequencies was drawn to identify whether there had been a mode shift away from an L-shaped distribution, indicating recent population bottlenecks.

MIGRATE (Beerli, 2008) was used to estimate the effective population size N_e. MIGRATE uses coalescent theory to jointly

estimate the mutation scaled population size θ ($4N_e\mu$) over a long period of time (~$4N_e$ generations). Three runs were conducted using MIGRATE. First, two shorter runs (10 short chains of 10,000 sampled, 500 record and three final chains of 100,000, 5,000 recorded) were performed and used to verify that the MCMC estimated the parameters correctly. Then, a final run (10 short chains of 10,000 sampled, 500 recorded and three final chains of 500,000 sampled and 25,000 recorded) was performed, and θ values from this final run are reported. The initial run used an estimate of F_{st} as a starting parameter to calculate θ. Each subsequent run used the ML estimates from the previous run as new starting parameters.

Principal coordinate analysis (PCoA), as implemented in GENALEX 6.5, was conducted to investigate the genetic divergence between cultivated and wild *Z. latifolia* plants. The genetic relationship between the cultivated and wild plants was also estimated using STRUCTURE 2.3.4 (Pritchard et al., 2000) with $K = 2$ and following the method described by Harter et al. (2004). If the proportion of an individual's genome with ancestry (the introgression coefficient Q) from the cultivated group was >50%, we defined that finding as a sign of cultivar ferality.

The degree of population divergence was quantified using F_{st} with GENEPOP 4.0. A PCoA was performed with the wild populations only to investigate the pattern of genetic divergence within this species. The genetic structure was further explored using the Bayesian clustering algorithm implemented in STRUCTURE 2.3.4. The program was given no prior information on ancestral populations and run 10 times for each value of K ancestral populations, under the admixture model with correlated allele frequencies, using 200,000 Markov chain Monte Carlo iterations and a burn-in of 100,000 iterations. We inferred K using the *ad hoc* statistic ΔK (Evanno et al., 2005). The resulting matrices of estimated cluster membership coefficients were permuted with CLUMPP (Jakobsson and Rosenberg, 2007). The final matrix for each K value was visualized with DISTRUCT (Rosenberg, 2004).

The genetic distance (GD) matrix for all pairs of individuals and populations of wild *Z. latifolia* was estimated by using MSA 3.0 (Dieringer and Schlotterer, 2003). Unrooted neighbor-joining trees were constructed using the PHYLIP package 3.6 with 1,000 bootstraps. The relatedness between individuals and populations was visualized with the software Dendroscope 3.

Isolation by distance (IBD) among the wild populations was tested using the Mantel test implemented in GENALEX 6.5. The significance of IBD values was assessed using 9,999 permutations.

Inference of the Domestication History of Cultivated *Z. latifolia*

We tested possible models of the demographic history of wild and cultivated accessions to infer the domestication history of cultivated *Z. latifolia* using Approximate Bayesian Computing (ABC) (Beaumont et al., 2002), which allows model choice and parameter estimation without the need to calculate the likelihood function of the models. Data sets are generated under the considered models and are reduced to a set of summary statistics. Simulated data sets with the summary statistics closest to those of the observed data are used to determine the posterior distribution of the parameters and the relative posterior probability of each model. DIYABC v2.0 (Cornuet et al., 2014) was used to compare different models and infer historical parameters. According to the geographical distribution and Bayesian clustering approach implemented in STRUCTURE, the *Z. latifolia* populations could be assigned into two genetic groups, and the southern group showed watersheds-related genetic subdivisions (see the Results section for more details). To reduce the complexity of scenario settings, hierarchical methods were applied following Lombaert et al. (2011). Competing scenarios first were evaluated between the two groups to identify the one with the highest posterior probability and then repeated within the supporting group among subgroups.

We then used an ABC approach to infer demographical parameters for the best scenario. The prior parameters and designs of the scenarios are given in **Table 2** and **Figures 3** and **4**. To construct the reference table, we calculated 10,000,000 simulated data sets for each competing scenario in DIYABC. The posterior probabilities of the competing scenarios were estimated by polychotomous logistic regression based on the 1% of data sets of the simulated reference table, according to the method described in Cornuet et al. (2014).

TABLE 2 | Prior distribution of parameters for hierarchical ABC analyses and posterior parameters estimation.

Prior parameters		Value	Posterior parameters			Quantile 2.5%	Quantile 97.5%
Parameter	Distribution		Mean	Median	Mode		
Analysis 1 (2 scenarios)							
*N*s	Uniform	10–20,000	9,180	8,650	7,070	3,330	11,300
*N*1	Uniform	10–1,000	247	228	199	49.3	319
*t*1	Uniform	10–2,000	754	697	537	166	1,010
*t*2	Uniform	10–2,000	971	903	668	250	1,320
Analysis 2 (9 scenarios)							
*N*e	Uniform	10–20,000	8,040	7,840	7,630	2,660	9,750
*N*d	Uniform	10–1,000	373	340	269	54.2	505
*t*1	Uniform	10–2,000	1,600	1,400	1,030	335	2,040
*t*2	Uniform	10–2,000	1,980	1,710	1,260	508	2,470
*t*3	Uniform	10–2,000	2,550	1,930	1,330	426	3,290

The Phylogenetic Analysis of *U. esculenta* ITS Sequences

The ITS sequences were aligned with CLUSTAL W (Thompson et al., 1994) and adjusted manually. The number of polymorphic sites (S), haplotype diversity (H_d), and genetic variation measured by average pairwise differences per base pair between sequences (θ) (Nei and Li, 1979) and Watterson's estimates (θ_w) from S (Watterson, 1975), were calculated using DNASP version 4.10 (Rozas et al., 2003). Maximum parsimony (MP) searches were performed with 1,000 random taxon addition replicates followed by tree bisection-reconnection branch swapping in PAUP* 4.0b10 (Swofford, 2002). Gaps were treated as missing data. Parsimony bootstrap (PB) for the clades was examined with 1,000 bootstrap replicates using the same options as above.

RESULTS

Genetic Variations in *Z. latifolia*

We analyzed plant genotypes for 824 wild and 135 cultivated accessions. Based on the 12 SSR loci, the 959 *Z. latifolia* accessions yielded a total of 138 alleles with 11.46 alleles per locus on average. The mean effective number of alleles (A), allelic richness (A_r), observed and expected heterozygosity (H_o and H_e), and Wright's fixation index (F) are listed in **Table 1**. H_e ranged from 0.167 (JXDA) to 0.588 (JSBY) among wild populations. The genetic diversity of cultivated accessions (H_e = 0.242) was much lower than that of wild accessions (H_e = 0.432). Deviations from Hardy-Weinberg equilibrium were detected in 90 of 384 comparisons after applying the sequential Bonferroni test, which might be due to inbreeding. The Wright's fixation indices varied greatly among populations (**Table 1**). Several populations (HLJHEB, HLJMDJ, JLDH, HBBD, SHYP, AHTL, JXDA, HuNCS, FJFD, GDYJ, GXNN, and YNJH) displayed signatures of heterozygote excess or deficit.

The PCoA indicated a distinct genetic divergence between wild and cultivated accessions. However, several putative wild accessions were identified as cultivars. This identification was further supported by the result of STRUCTURE analysis (K = 2, **Figures 2A**, **B**). The ΔK analysis indicated that the wild populations of *Z. latifolia* would demonstrate fine genetic structures with an increasing K value, which were shown in **Figures 2C**, **D**. The HuNCS and FJFD populations were thoroughly composed of plants with cultivar-like genotype. In addition, another 12 accessions from 3 wild populations (SHCM, ZJHZ, and YNJH) were identified as cultivars (**Table 1**). The neighbor-joining tree showed that the cultivars were genetically close to the populations in the lower reaches of the Yangtze River (**Figure S1**).

Genetic Structure

The effective population size N_e estimated by MIGRATE ranged from 100 to 312 ($\mu = 10^{-4}$, **Table 1**). Thirteen of 32 wild populations showed the evidence for historical or recent genetic bottlenecks, suggesting that the effects of genetic drift on these populations were prominent. The cultivars also showed a strong genetic bottleneck due to domestication (**Tables 1** and **2**).

After excluding the individuals with cultivar-like genotype (SHCM, ZJHZ, HuNCS, YNJH, and FJFD), F-statistics revealed that the wild populations of *Z. latifolia* were genetically divergent (F_{st} = 0.320, Table S1). The Mantel test suggested a weak but significant sign of isolation by distance among *Z. latifolia* populations (**Figure S2**, r = 0.21, $P < 0.05$). The outcomes of a Bayesian assignment with STRUCTURE indicated that K = 2 is the best for describing the actual genetic structure of wild populations (**Figure 3A**). One group included the northern populations in the north and Northeast Plain of China, and the remaining southern populations formed another group mainly located in the Yangtze River Basin. However, the PCoA showed no distinct genetic divergence among wild populations, and the first three principal components described 28.6% of the total variances. In addition, the Bayesian assignment of the southern populations revealed watershed-related genetic subdivisions in the Yangtze River Basin: the Hongzhe-Tai Lake region (Cluster I), the Poyang Lake region (Cluster II) and the Dongting Lake region (Cluster III) (**Figure 4A**). Two southern populations located in the Pearl River Basin (GDYJ and GXNN) showed a close ancestry with the populations in the Poyang Lake region cluster (**Figure 4A**).

In particular, we tested the genetic divergence between two ecotypes of cultivars. By using the software STRUCTURE, we found no significant genetic divergence with K = 2. Instead, the PCoA revealed the genetic differences between ecotypes (**Figure S3**). The microsatellite marker ZM24 could be an effective molecular marker to distinguish two ecotypes of cultivars.

Modeling the Demography of *Z. latifolia* Domestication

Following our hierarchical ABC strategy, we first compared the two predesigned scenarios in Analysis 1. Then the posterior parameters of the best-supported scenario were estimated, and the model checking was accessed between the posterior data set and observed data (**Table 2** and **Figure S4A**). Scenario A had a higher posterior probability (0.700) (**Figure 3B**), and it indicated that the cultivars diverged from the group of southern populations. In Analysis 2, 9 competing scenarios were further investigated (**Table 2** and **Figure S4B**). Scenario 7 showed the highest posterior probability (0.295), which indicated that the Hongze-Tai Lake region cluster firstly diverged from the other clusters in the southern group 1930 generations (t3) ago. The Poyang Lake region cluster and Dongting Lake cluster then diverged 1,710 generations (t2) ago. The cultivars were single-domesticated in the Hongze-Tai Lake region 1,400 generations (t1) ago (**Figure 4B** and **Table 2**).

Sequence Variation and Phylogenetic Analysis

A total of 37 strains of *U. esculenta*, including 5 wild and 32 cultivated strains, were sequenced. The total length of the aligned sequences was 712 bp. Twenty-seven polymorphic sites were observed. Thirty-one haplotypes were found in the wild and cultivated accessions, and the haplotype diversity (H_d) was 0.991. The nucleotide diversity (θ) was 0.01, and θ_w was 0.01.

An ITS sequence of *U. alcornii* was downloaded from GeneBank and used as the outgroup taxon. Another ITS sequence of *U. esculenta* (Genebank accession number: JX219372.1) was used as

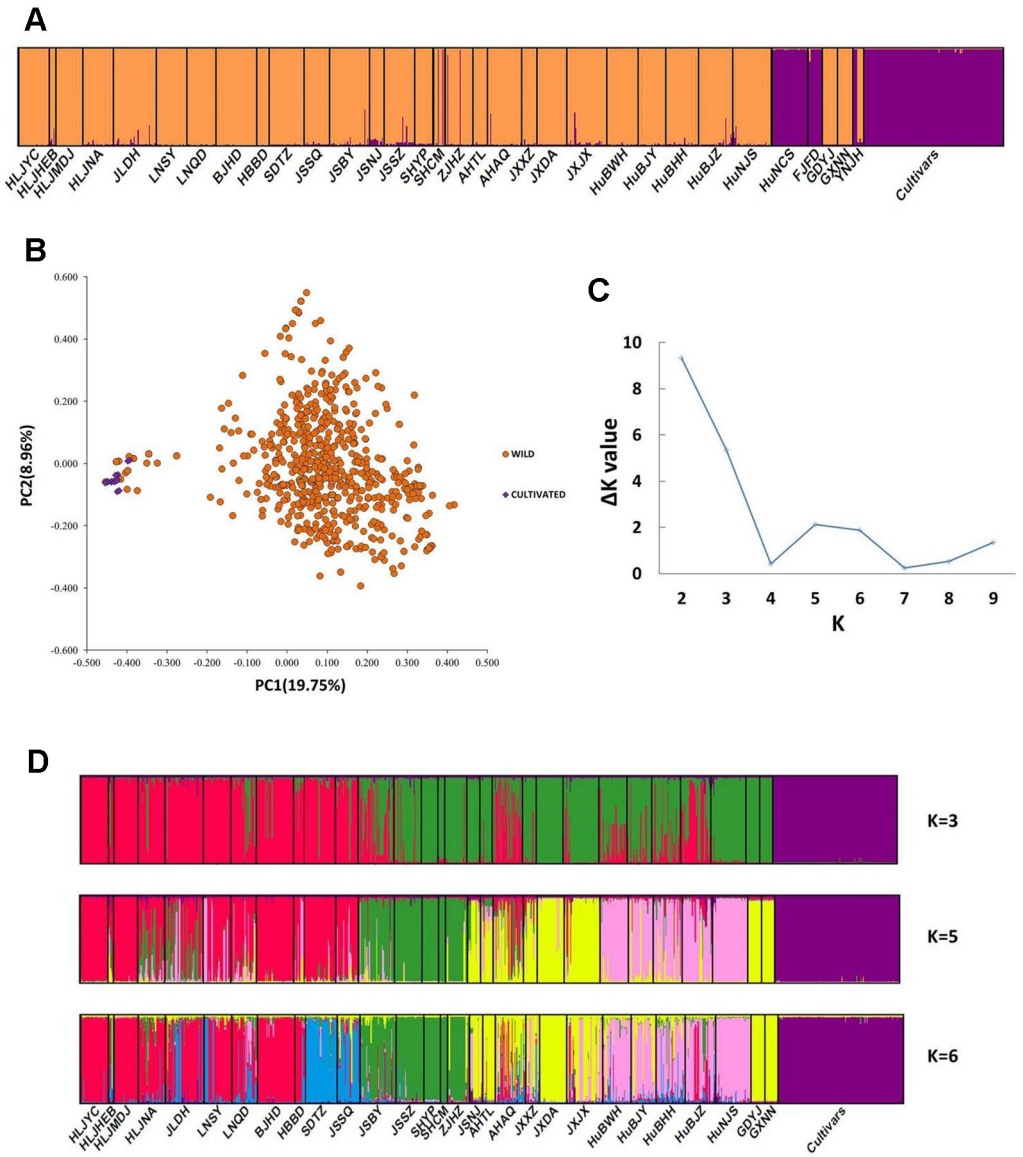


FIGURE 2 | Genetic structure in *Zizania latifolia*. **(A)** The genetic divergence between wild populations and landraces was illustrated by STRUCTURE (K = 2). **(B)** Principal Coordinate Analysis (PCoA) of 32 wild populations and cultivated accessions. **(C)** The estimations of ΔK. **(D)** Genetic subdivisions in the wild populations of *Z. latifolia* (when K = 3, 5 and 6).

the control. The MP tree is shown in **Figure 5**. The structure of the MP tree showed significant divergence between the wild fungi and those from cultivars, indicating a single domestication event.

DISCUSSION

The domestication of cultivated *Z. latifolia* is a typical example that can demonstrate how a plant-fungus complex rapidly evolved under artificial selections. Our results showed that there was prominent genetic divergence between wild populations and cultivated accessions after long-time cultivation. A single domestication event occurred in the lower reaches of the Yangtze River at the Tang Dynasty (approximately 1,400 years ago), then this vegetable was constantly cultivated and bred. The genetic divergence of the plant may be responsible for the differences between the ecotypes of cultivars, and the genetic variations in *U. esculenta* are likely associated with the diversification of cultivated varieties.

The Possible Domestication Process of *Z. latifolia*

There are distinct phenotypic differences between wild and cultivated *Z. latifolia*. Compared with its wild ancestor, the cultivated plant usually has an erected plant architecture, broader leaves, and shortened internodes (Guo et al., 2007). The cultivated

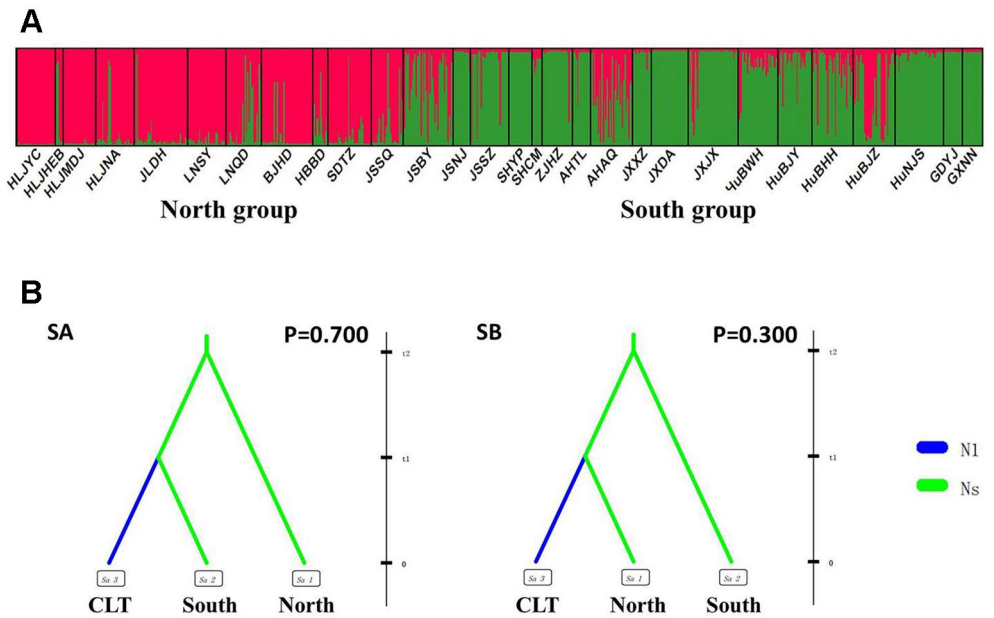


FIGURE 3 | Genetic structure and settings of two competing scenarios in the wild *Zizania latifolia* populations. **(A)** Genetic structure of wild populations illustrated by STRUCTURE. Each vertical bar represents an individual, and each color represents a cluster in the histograms of STRUCTURE. **(B)** Two competing scenarios for ABC simulations. CLT is the cultivated group, South is the south group, and North is the north group. These groups were assumed to be isolated from each other, with no exchange of migrants. In Scenario A (SA), the southern group was separated from the northern group at time scale t2, with no changes in effective population size; at time scale t1, the cultivated group was then separated from the southern group and went through a bottleneck with an effective population size change from *N*s to *N*1.

plants were reported to be vigorous in photosynthetic efficiency (Yan et al., 2013). These morphological and physiological changes together constituted the Domestication Syndrome of cultivated *Z. latifolia*, which would be referenced in distinguishing wild and cultivated plants.

We found significant genetic divergence between wild and cultivated *Z. latifolia* (**Figures 2A**, **B**). The extent of genetic differentiation between wild and cultivated plants (F_{st} = 0.288) was even greater than the differentiation among wild populations (F_{st} = 0.175). This phenomenon is usually observed in highly domesticated species such as wheat, rice or maize (Heun et al., 1997; Harter et al., 2004; Zeder et al., 2006; Zhao et al., 2013). Unlike the annual grain crops, the breeding of *Z. latifolia* relies on asexual reproduction. By utilizing natural mutants that occur in tillers or rhizomes, cultivated *Z. latifolia* was able to generate more than 100 varieties. Theoretically, when compared with annual crops bred *via* sexual reproduction, the evolutionary rate of cultivated *Z. latifolia* is relatively low. The low evolutionary rate and short domestication time (<2,000 years) should result in low genetic divergence between wild and cultivated plants, which is contrary to our finding. The unexpected high genetic divergence between wild and cultivated plants could also be attributed to the breeding system of cultivated *Z. latifolia*. Because cultivated *Z. latifolia* would occasionally degenerate to the wild phenotype, farmers had to screen out the elite plants as genets every year. Although the clonal reproduction of cultivated *Z. latifolia* greatly restricted new mutants stemmed from sexual propagation and blocked gene flows from wild accessions to cultivars, the annually strong artificial selection would somehow accumulate mutants, inducing significant genetic divergence between wild and cultivated plants.

Previous studies had inferred that the cultivated *Z. latifolia* may be domesticated in the Yangtze River Basin, but they did not provide any convincing evidence. (Xu et al., 2008; Xu et al., 2015). Using microsatellite markers and model inference, our study located the origin site of cultivated *Z. latifolia* in the areas around Hongze Lake and Tai Lake (**Figure 4**). By ABC modeling, the cultivated *Z. latifolia* diverged from ancestor populations 1,400 generations ago. Given 1 year per generation, the domestication event occurred 1,400 years ago at the Tang Dynasty. Furthermore, the genetic relationship of fungi hinted that the single-season ecotype might have been domesticated first, and then the double-season ecotype was bred from the single-season ecotype. These results were in accordance with the historical records (Ke, 2008).

Based on current molecular data, we could not be sure whether *Z. latifolia* underwent historical pre-domestication as a food crop. There is no historical record or remained landraces that can support the existence of pre-domestication. We never found the plants with any domestication traits similar to cereal crops in the wild, too. However, the vegetable cultivars of *Z. latifolia* could escape to natural habitats. We found typical escaped individuals in five provinces that were the main production areas of cultivated *Z. latifolia* (**Table 1**). Cultivated plants that cannot survive on their own in natural habitats often have maladaptive domestication traits (Campbell and Snow, 2009). Cultivated *Z. latifolia* has lost the ability to undergo sexual reproduction, which would

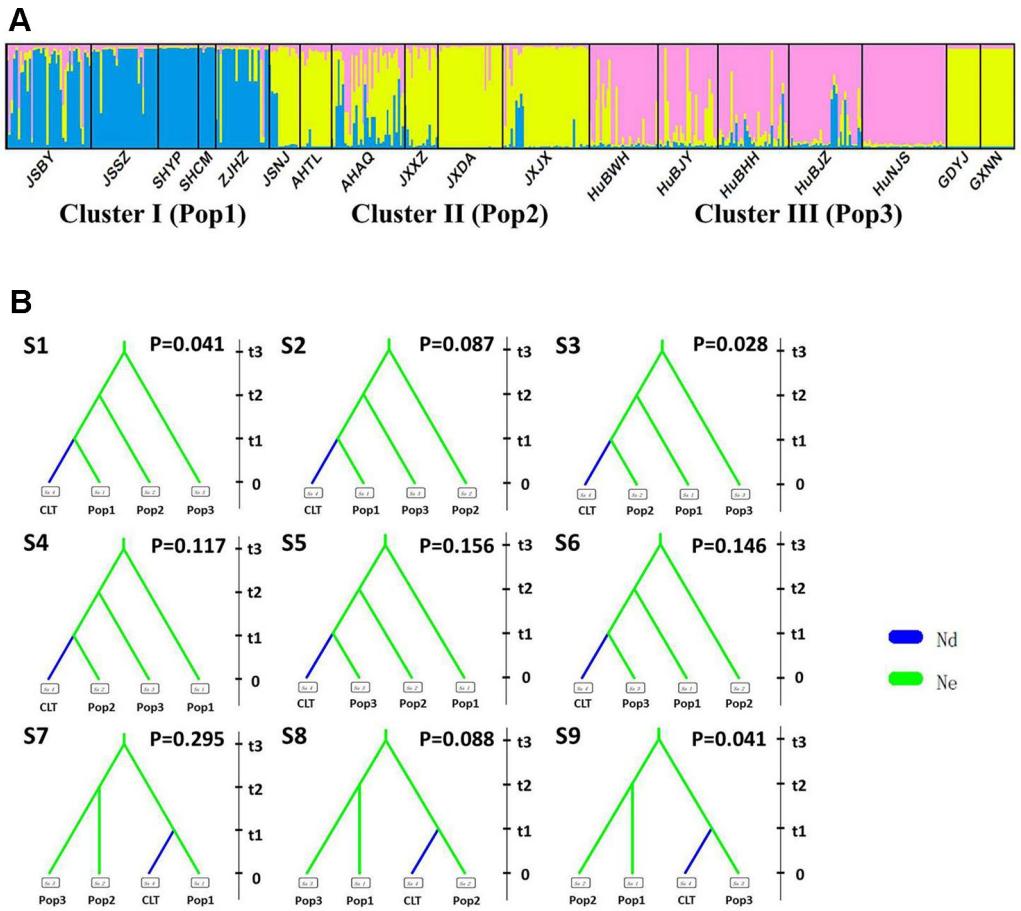


FIGURE 4 | Genetic structure and settings of nine competing scenarios in the southern genetic group of wild *Zizania latifolia*. **(A)** The subtle genetic structure in the southern genetic group of *Z. latifolia* revealed by STRUCTURE. Each vertical bar represents an individual, and each color represents a cluster in the histograms of STRUCTURE. **(B)** 9 competing scenarios for ABC simulations inferring the origin of cultivars in the south group of wild *Z. latifolia* populations. CLT is the cultivated group, Pop1 is the Hongze-Tai lake group, Pop2 is the Poyang lake group, and Pop4 is the Dongting lake group.

be unfavorable for persistence in the wild. Alternatively, feral *Z. latifolia* relies on asexual growth to reproduce in natural habitats. In fact, it often occupies the shores of ponds or streams near fields by clonal propagation *via* underground rhizomes and detached shoots, just as we had noticed in FJFD and HuNCS. Moreover, without pollen flow, it seems unlikely that wild populations would be introgressed by feral cultivars.

Genetic Diversity and Population Structure of Wild *Z. latifolia*

Xu et al. (2015) had characterized the genetic diversity of the four species in the genus *Zizania* using 3 cross-specific amplified microsatellite markers. They found that *Z. latifolia* showed a relatively low genetic diversity relative to its annual relative *Z. palustris* in North America (H_e = 0.374 vs H_e = 0.630). However, they might have underestimated the genetic diversity of both species due to insufficient genetic markers. In contrast, another study investigated the genetic variations of *Z. latifolia* restricted to the middle reaches of the Yangtze River by using 10 microsatellites and found a high level of genetic diversity (H_e = 0.532) (Chen et al., 2012). In our study, a relatively high genetic diversity (H_e = 0.497) was detected.

Most of the fixation indices (*F*) for the high-latitude populations were significantly greater than 0 (**Table 1**), indicating a heterozygous deficit that could be attributed to inbreeding or outbreeding among ramets. However, several populations in low latitudes were composed by a few dominant heterozygous genotypes, which resulted in negative *F* values. This phenomenon might be attributed to local adaption or a rapid population expansion based on a few heterozygous ancients *via* clonal growth. For example, 10 years ago, a few ancestors floating down from upstream colonized and formed the population JXDA (Zhao et al., field surveys). Previous phylogeographic analysis suggested that the genus *Zizania* originated in North America and then dispersed into eastern Asia *via* the Bering land bridge during the Tertiary (Xu et al., 2010). Hewitt (2004) had proposed that the range-wide patterns of population genetic diversity are usually shaped by past climate-driven range dynamic. In support of this view, the genealogical pattern based on the nuclear Adh1a

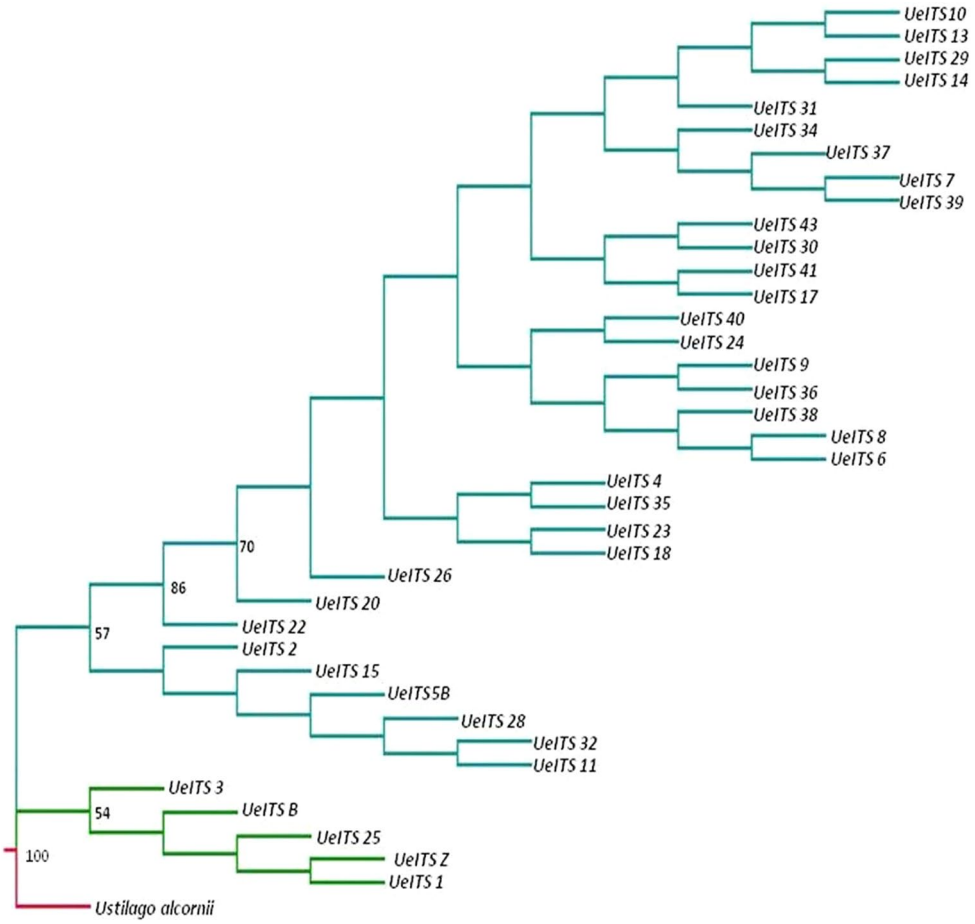


FIGURE 5 | Phylogeny tree (maximum parsimony, MP) for the 37 accessions (wild accessions in green and cultivated accessions in cyan) of *Ustilago esculenta*, *Ustilago alcornii* is the outgroup (in red). Bootstrap values calculated over 1,000 replications are given as percentages (only values > 50% were shown).

gene of *Z. latifolia* indicated that high-latitude populations harbored more abundant haplotypes (Xu et al., 2008). However, in our study, the populations in high latitudes (the northern group, H_e = 0.447) did not show a higher genetic diversity than those in low latitudes (the southern group, H_e = 0.444).

Our results and previous studies all indicated strong genetic differences among populations (**Figure 2**). The Mantel test further revealed the effect of isolation by distance (IBD, **Figure S2**). IBD is expected to be the primary factor leading to a high level of genetic divergence, due to limited gene flow and the effects of genetic drift (Slatkin, 1987). This process was supported by our study on wild *Z. latifolia* populations. The aquatic habitat *of Z. latifolia* is discrete and patchy. These isolated wetlands restrict migration between populations, and as a consequence, aquatic-living plants usually exhibit a high level of genetic divergence among populations (Barrett et al., 1993). The dispersal of seeds or detached shoots by water flow is unpredictable, and these are unlikely to disperse *via* water currents between spatially highly isolated populations. In wind-pollinated species, the pollen can be carried for long distances, but most effective pollination occurs at a local scale (Willson, 1983). Nevertheless, the isolation between wetlands might be counteracted by seasonal flooding, resulting in a temporal connection within a water system. The genetic structure of *Z. latifolia* provided evidence that supported the effects of hydrological connectivity in shaping the genetic structure of aquatic plant species. Isolation by geographical barriers and genetic similarity along rivers was found in regional populations of *Z. latifolia* (Chen et al., 2012; Chen et al., 2017). The effects of hydrological unconnectedness on genetic relationship of populations within a water system were also reported in another emergent macrophyte, *Oryza rufipogon* (Wang et al., 2008).

The genetic assignments and morphological performances of two populations (GDYJ and GXNN) at the leading-edge seemed to be rather special (**Table 1** and **Figure 2D**). On one hand, these populations are cultivated-like in phenotype, but they showed no sign of *U. esculenta* infection; on the other hand, they are genetically different from cultivated *Z. latifolia*, and have heterozygous genotypes. Their small population size and low diversity might be attributed to the "Founder effect." Bayesian assignment showed they could be grouped to the Poyang Lake

region cluster, implying a migration through the geographical barrier (Nanling Mountains).

The Association of Plant and Fungus Under Artificial Selection

As an obligate parasitic fungus, *U. esculenta* is harmful to its host under natural conditions. The fungus not only inhibits the emergence of inflorescence but also requires a large amount of carbohydrates to form the swollen shoot attractive to herbivorous insects. *U. esculenta* infection is not very common in wild populations, as we have only collected 5 samples across 32 range-wide populations. In our common garden trials, the fitness of infected wild plants is significantly lower than that of healthy ones (Zhao et al., data not shown), which implies a weak competitive ability of the infected plants.

The Red Queen hypothesis proposed that the parasite and host should keep on adapting and evolving to gain reproductive advantages (Van Valen, 1973). The antagonistic co-evolution between species would finally be terminated either by the extinction of one species or the emergence of symbiosis. The domestication of infected *Z. latifolia* seemed to accelerate the process of achieving symbiosis. On the one hand, the persistent selection force of mankind delayed the development of the fungus and extended the mycolic stage (Zhang et al., 2017). On the other hand, humans gradually screened out plants with the genotype that adapted to cultivated conditions (Yan et al., 2013). Moreover, with the efforts exerted by mankind, the distribution range of this plant-fungus complex was largely expanded.

We detected minor but significant genetic differentiation between the hosts of two ecotypes of cultivated *Z. latifolia* (**Figure S3**), which was consistent with the pattern revealed by the sequences of Adh1 (Xu et al., 2008). However, the ITS sequences of *U. esculenta* strains from cultivated Z. latifolia showed unexpectedly high levels of diversity compared with a previous study (You et al., 2011).There might be inconsistent evolutionary rates in the parasite and host. The evolution of the host determines the responses to photoperiod and temperature which resulted in the differences between two ecotypes in cultivated *Z. latifolia* (Guo et al., 2015), and the faster evolutionary rate of the parasite may lead to the diversification of varieties (**Figure 5**, also see Ye et al., 2017).

Currently, it is still difficult to dissect the relative contributions from the host and parasite in shaping Domestication Syndrome. We could only infer the changes in morphological traits during the domestication process based on the phylogeny of cultivated *U. esculenta* (**Figure 5**). First, the single-season ecotype was domesticated from a wild plant-fungus complex. This ecotype retained a few primary traits, such as light sensitivity, long internodes, and small swollen shoot. Then, artificial selections on one of the single-season varieties led to the emergence of the double-season ecotype, which was insensitive to light and could be harvested twice. The double-season ecotype was more advanced than the single-season ecotype, but it could only be cultivated in a restricted region. There is an urgent need to improve the fitness of the double-season ecotype.

DATA AVAILABILITY STATEMENT

All datasets for this study are included in the article/ **Supplementary Material**.

AUTHOR CONTRIBUTIONS

YZ, ZS, JC and JR conceived the idea and designed the research project. YZ, QL and LZ collected the data. YZ performed the data analysis. YZ drafted the initial manuscript with contribution from JR and ZS. All the authors contributed critically to the discussion and edited the manuscript before submission.

FUNDING

This study was supported by the National Natural Science Foundation of China (31600293) and the China Postdoctoral Science Foundation Grant (2015M571483).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2019.01406/full#supplementary-material

REFERENCES

Barrett, S. C. H., Eckert, C. G., and Husband, B. C. (1993). Evolutionary processes in aquatic plant populations. *Aquat. Bot.* 44 (2-3), 105–145. doi: 10.1016/0304-3770(93)90068-8

Beaumont, M.A., Zhang, W., and Balding, D.J. (2002). Approximate Bayesian computation in population genetics. *Genetics* 162, 2025–2035. doi: 10.0000/PMID12524368

Beerli, P. (2008). Migrate version 3.0: a maximum likelihood and Bayesian estimator of gene flow using the coalescent. Available from http://popgen.scs.edu/migrate.html.

Bitocchi, E., Nanni, L., Bellucci, E., Rossi, M., Giardini, A., Zeuli, P. S., et al. (2012). Mesoamerican origin of the common bean (*Phaseolus vulgaris* L.) is revealed by sequence data. *P. Natl. Acad. Sci. U. S. A.* 109 (14), E788–E796. doi: 10.1073/pnas.1108973109

Burger, J. C., Chapman, M. A., and Burke, J. M. (2008). Molecular insights into the evolution of crop plants. *Am. J. Bot.* 95 (2), 113–122. doi: 10.3732/Ajb.95.2.113

Campbell, L. G., and Snow, A. A. (2009). Can feral weeds evolve from cultivated radish (*Raphanus sativus*, Brassicaceae)? *Am. J. Bot.* 96 (2), 498–506. doi: 10.3732/ajb.0800054

Chapuis, M. P., and Estoup, A. (2007). Microsatellite null alleles and estimation of population differentiation. *Mol. Biol. Evol.* 24 (3), 621–631. doi: 10.1093/molbev/msl191

Chen, Y., Long, L. K., Lin, X. Y., Guo, W. L., and Liu, B. (2006). Isolation and characterization of a set of disease resistance-gene analogs (RGAs) from wild rice, *Zizania latifolia* Griseb. I. Introgression, copy number lability, sequence change, and DNA methylation alteration in several rice-*Zizania* introgression lines. *Genome* 49 (2), 150–158. doi: 10.1139/G05-097

Chen, Y. Y., Chu, H. J., Liu, H., and Liu, Y. L. (2012). Abundant genetic diversity of the wild rice *Zizania latifolia* in central China revealed by microsatellites. *Ann. Appl. Biol.* 161 (2), 192–201. doi: 10.1111/j.1744-7348.2012.00564.x

Chen, Y. Y., Liu, Y., Fan, X. R., Li, W., and Liu, Y. L. (2017). Landscape-scale genetic structure of widl rice *Zizania latifolia*: the roles of rivers, mountains and fragmentation. *Front. Ecol. Evol.* 5, 17. doi: 10.3389/fevo.2017.00017

Choi, J. Y., and Purugganan, M. D. (2018). Multiple origin but single domestication led to *Oryza sativa*. G3-Genes. *Genom. Genet.* 8 (3), 797–803. doi: 10.1534/g3.117.300334

Cornuet, J. M., Pudlo, P., Veyssier, J., Dehne-Garcia, A., Gautier, M., Leblois, R., et al. (2014). DIYABC v2.0: a software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. *Bioinformatics* 30 (8), 1187–1189. doi: 10.1093/bioinformatics/btt763

Dieringer, D., and Schlotterer, C. (2003). Microsatellite analyser (MSA): a platform independent analysis tool for large microsatellite data sets. *Mol. Ecol. Notes* 3 (1), 167–169. doi: 10.1046/j.1471-8286.2003.00351.x

Diez, C. M., Trujillo, I., Martinez-Urdiroz, N., Barranco, D., Rallo, L., Marfil, P., et al. (2015). Olive domestication and diversification in the Mediterranean Basin. *New Phytol.* 206 (1), 436–447. doi: 10.1111/nph.13181

Doebley, J. F., Gaut, B. S., and Smith, B. D. (2006). The molecular genetics of crop domestication. *Cell* 127 (7), 1309–1321. doi: 10.1016/j.cell.2006.12.006

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 (8), 2611–2620. doi: 10.1111/j.1365-294X.2005.02553.x

FAO (2019). *The State of the World's Biodiversity for Food and Agriculture*. Bélanger, J, and Pilling, D, editors. Rome: FAO Commission on Genetic Resources for Food and Agriculture Assessments.

Fuller, D. Q., Asouti, E., and Purugganan, M. D. (2012). Cultivation as slow evolutionary entanglement: comparative data on rate and sequence of domestication. *Veg. Hist. Archaeobot.* 21 (2), 131–145. doi: 10.1007/s00334-011-0329-8

Ge, S., Li, A., Lu, B. R., Zhang, S. Z., and Hong, D. Y. (2002). A phylogeny of the rice tribe Oryzeae (Poaceae) based on matK sequence data. *Am. J. Bot.* 89 (12), 1967–1972. doi: 10.3732/ajb.89.12.1967

Ge, S., and Sang, T. (2011). Inappropriate model rejects independent domestications of indica and japonica rice. *P. Natl. Acad. Sci. U. S. A.* 108 (39), E55–E55. doi: 10.1073/pnas.1111601108

Giovannoni, J. (2018). Tomato multiomics reveals consequences of crop domestication and improvement. *Cell* 172 (1-2), 6–8. doi: 10.1016/j.cell.2017.12.036

Gross, B. L., and Olsen, K. M. (2010). Genetic perspectives on crop domestication. *Trends Plants Sci.* 15 (9), 529–537. doi: 10.1016/j.tplants.2010.05.008

Guo, H. B., Li, S. M., Peng, J., and Ke, W. D. (2007). *Zizania latifolia* Turcz. cultivated in China. *Genet. Resour. Crop. Ev.* 54 (6), 1211–1217. doi: 10.1007/s10722-006-9102-8

Guo, L. B., Qiu, J., Han, Z. J., Ye, Z. H., Chen, C., Liu, C. J., et al. (2015). A host plant genome (*Zizania latifolia*) after a century-long endophyte infection. *Plant J.* 83 (4), 600–609. doi: 10.1111/pj.12912

Guo, Y. L., and Ge, S. (2005). Molecular phylogeny of Oryzeae (Poaceae) based on DNA sequences from chloroplast, mitochondrial, and nuclear genomes. *Am. J. Bot.* 92 (9), 1548–1558. doi: 10.3732/ajb.92.9.1548

Hajjar, R., and Hodgkin, T. (2007). The use of wild relatives in crop improvement: a survey of developments over the last 20 years. *Euphytica.* 156 (1-2), 1–13. doi: 10.1007/s10681-007-9363-0

Harter, A. V., Gardner, K. A., Falush, D., Lentz, D. L., Bye, R. A., and Rieseberg, L. H. (2004). Origin of extant domesticated sunflowers in eastern North America. *Nature* 430 (6996), 201–205. doi: 10.1038/nature02710

Heun, M., Abbo, S., Lev-Yadun, S., and Gopher, A. (2012). A critical review of the protracted domestication model for Near-Eastern founder crops: linear regression, long-distance gene flow, archaeological, and archaeobotanical evidence. *J. Exp. Bot.* 63 (12), 4333–4341. doi: 10.1093/jxb/ers162

Heun, M., SchaferPregl, R., Klawan, D., Castagna, R., Accerbi, M., Borghi, B., et al. (1997). Site of einkorn wheat domestication identified by DNA fingerprinting. *Science* 278 (5341), 1312–1314. doi: 10.1126/science.278.5341.1312

Hewitt, G. M. (2004). Biodiversity: a climate for colonization. *Heredity (Edinb)* 92 (1), 1–2. doi: 10.1038/sj.hdy.6800365

Huang, X. H., Kurata, N., Wei, X. H., Wang, Z. X., Wang, A., Zhao, Q., et al. (2012). A map of rice genome variation reveals the origin of cultivated rice. *Nature* 490 (7421), 497–501. doi: 10.1038/nature11532

Jakobsson, M., and Rosenberg, N. A. (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23 (14), 1801–1806. doi: 10.1093/bioinformatics/btm233

Ke, W. D. (2008). "Jiaobai," in *Crops and Their Wild Relatives in China*, vol. Vegetable Crops . Eds. Dong, Y. C., and Liu, X. (Beijing: China Agriculture Press), 1084–1095.

Liu, B., Liu, Z. L., and Li, X. W. (1999). Production of a highly asymmetric somatic hybrid between rice and *Zizania latifolia* (Griseb): evidence for inter-genomic exchange. *Theor. Appl. Genet.* 98 (6-7), 1099–1103. doi: 10.1007/s001220051173

Lombaert, E., Guillemaud, T., Thomas, C. E., Handley, L. J. L., Li, J., Wang, S., et al. (2011). Inferring the origin of populations introduced from a genetically structured native range by approximate Bayesian computation: case study of the invasive ladybird *Harmonia axyridis*. *Mol. Ecol.* 20 (22), 4654–4670. doi: 10.1111/j.1365-294X.2011.05322.x

Nei, M., and Li, W. H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci. U. S. A.* 76 (10), 5269–5273. doi: 10.1073/pnas.76.10.5269

Oelke, E. A. (1993). "Wild rice: domestication of a native North American genus," in *New crops*. Eds. Janick, J., and Simon,J. E. (New York: Wiley), 235–243.

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 (19), 2537–2539. doi: 10.1093/bioinformatics/bts460

Piry, S., Luikart, G., and Cornuet, J. M. (1999). BOTTLENECK: a computer program for detecting recent reductions in the effective population size using allele frequency data. *J. Hered.* 90 (4), 502–503. doi: 10.1093/jhered/90.4.502

Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155 (2), 945–959. doi: 10.1111/j.1471-8286.2004.00684.x

Purugganan, M. D., and Fuller, D. Q. (2009). The nature of selection during plant domestication. *Nature* 457 (7231), 843–848. doi: 10.1038/nature07895

Quan, Z. W., Pan, L., Ke, W. D., Liu, Y. M., and Ding, Y. (2009). Sixteen polymorphic microsatellite markers from *Zizania latifolia* Turcz. (Poaceae). *Mol. Ecol. Resour.* 9 (3), 887–889. doi: 10.1111/j.1755-0998.2008.02357.x

Raymond, M., and Rousset, F. (1995). GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *J. Hered.* 86, 248–249. doi: 10.1046/j.1420-9101.1995.8030385.x

Ronald, P. C. (2014). Lab to farm: applying research on plant genetics and genomics to crop improvement. *PloS Biol.* 12 (6), e1001878. doi: 10.1371/journal.pbio.1001878

Rosenberg, N. A. (2004). DISTRUCT: a program for the graphical display of population structure. *Mol. Ecol. Notes* 4 (1), 137–138. doi: 10.1046/j.1471-8286.2003.00566.x

Rozas, J., Sanchez-DelBarrio, J. C., Messeguer, X., and Rozas, R. (2003). DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19 (18), 2496–2497. doi: 10.1093/bioinformatics/btg359

Slatkin, M. (1987). Gene flow and the geographic structure of natural-populations. *Science* 236 (4803), 787–792. doi: 10.1126/science.3576198

Song, Z. P., Zhu, W. Y., Rong, J., Xu, X., Chen, J. K., and Lu, B. R. (2006). Evidences of introgression from cultivated rice to *Oryza rufipogon* (Poaceae) populations based on SSR fingerprinting: implications for wild rice differentiation and conservation. *Evol. Ecol.* 20 (6), 501–522. doi: 10.1007/s10682-006-9113-0

Stoll, M., Begerow, D., and Oberwinkler, F. (2005). Molecular phylogeny of *Ustilago, Sporisorium*, and related taxa based on combined analyses of rDNA sequences. *Mycol. Res.* 109, 342–356. doi: 10.1017/S0953756204002229

Swofford, D. L. (2002). *PAUP*: Phylogenetic analysis using parsimony and other methods Version 4b10*. Sunderland, USA: Sinauer Associates.

Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). Clustal-W - Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22 (22), 4673–4680. doi: 10.1093/nar/22.22.4673

Tufto, J. (2017). Domestication and fitness in the wild: a multivariate view. *Evolution* 71 (9), 2262–2270. doi: 10.1111/evo.13307

van Heerwaarden, J., Doebley, J., Briggs, W. H., Glaubitz, J. C., Goodman, M. M., Gonzalez, J. D. S., et al. (2011). Genetic signals of origin, spread, and introgression in a large sample of maize landraces. *P. Natl. Acad. Sci. U. S. A.* 108 (3), 1088–1092. doi: 10.1073/pnas.1013011108

Van Valen, L. (1973). Body size and numbers of plants and animals. *Evolution* 27 (1), 27–35. doi: 10.1111/j.1558-5646.1973.tb05914.x

Wang, M., Li, W. Z., Fang, C., Xu, F., Liu, Y. C., Wang, Z., et al. (2018). Parallel selection on a dormancy gene during domestication of crops from multiple families. *Nat. Genet.* 50 (10), 1435–1441. doi: 10.1038/s41588-018-0229-2

Wang, M. X., Zhang, H. L., Zhang, D. L., Qi, Y. W., Fan, Z. L., Li, D. Y., et al. (2008). Genetic structure of *Oryza rufipogon* Griff. in China. *Heredity* 101 (6), 527–535. doi: 10.1038/hdy.2008.61

Wang, Z. D., Yan, N., Wang, Z. H., Zhang, X. H., Zhang, J. Z., Xue, H. M., et al. (2017). RNA-seq analysis provides insight into reprogramming of culm development in *Zizania latifolia* induced by *Ustilago esculenta*. *Plant Mol. Biol.* 95 (6), 533–547. doi: 10.1007/s11103-017-0658-9

Watterson, G. A. (1975). On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* 7, 256–276. doi: 10.1016/0040-5809(75)90020-9

Wen, J. (1999). Evolution of eastern Asian and eastern North American disjunct distributions in flowering plants. *Annu. Rev. Ecol. Syst.* 30, 421–455. doi: 10.1146/annurev.ecolsys.30.1.421

Whitehead, S. R., Turcotte, M. M., and Poveda, K. (2017). Domestication impacts on plant - herbivore interactions: a meta-analysis. *Philos. T. R. Soc. B.* 372 (1712), 20160034. doi: 10.1098/rstb.2016.0034

Willson, M. F. (1983). *Plant reproductive ecology*. New York, USA: Wiley.

Wu, Z. Y., Raven, P. H., and Hong, D. Y., (2006). "Flora of China," in *Poaceae*, vol. 22. (Beijing/St Louis: Science Press/Missouri Botanical Garden).

Xiang, Q. Y., Crawford, D. J., Wolfe, A. D., Tang, Y. C., and DePamphilis, C. W. (1998). Origin and biogeography of *Aesculus* L. (Hippocastanaceae): a molecular phylogenetic perspective. *Evolution* 52 (4), 988–997. doi: 10.1111/j.1558-5646.1998.tb01828.x

Xu, X. W., Ke, W. D., Yu, X. P., Wen, J., and Ge, S. (2008). A preliminary study on population genetic structure and phylogeography of the wild and cultivated *Zizania latifolia* (Poaceae) based on Adh1a sequences. *Theor. Appl. Genet.* 116 (6), 835–843. doi: 10.1007/s00122-008-0717-3

Xu, X. W., Walters, C., Antolin, M. F., Alexander, M. L., Lutz, S., Ge, S., et al. (2010). Phylogeny and biogeography of the eastern Asian–North American disjunct wild-rice genus (*Zizania* L., Poaceae). *Mol. Phyl. Evol.* 55, 1008–1017. doi: 10.1016/j.ympev.2009.11.018

Xu, X. W., Wu, J. W., Qi, M. X., Lu, Q. X., Lee, P. F., Lutz, S., et al. (2015). Comparative phylogeography of the wild-rice genus *Zizania* (Poaceae) in eastern asia and North America. *Am. J. Bot.* 102 (2), 239–247. doi: 10.3732/ajb.1400323

Yan, N., Wang, X. Q., Xu, X. F., Guo, D. P., Wang, Z. D., Zhang, J. Z., et al. (2013). Plant growth and photosynthetic performance of *Zizania latifolia* are altered by endophytic *Ustilago esculenta* infection. *Physiol. Mol. Plant.* P. 83, 75–83. doi: 10.1016/j.pmpp.2013.05.005

Yang, Y. F., Wang, S. Z., and Lang, H. Q. (1999). Study on the structures of clone population *Zizania latifolia* in wetland of the Songnen Plain of China. *Acta Prataculturalence* 8, 66–71.

Ye, Z., Pan, Y., Zhang, Y., Cui, H., Jin, G., McHardy, A. C., et al. (2017). Comparative whole-genome analysis reveals artificial selection effects on *Ustilago esculenta* genome. *DNA Res.* 24 (6), 635–648. doi: 10.1093/dnares/dsx031

You, W. Y., Liu, Q., Zou, K. Q., Yu, X. P., Cui, W. Y., and Ye, Z. H. (2011). Morphological and molecular differences in two strains of *Ustilago esculenta.* Curr. *Microbiol.* 62, 44–54. doi: 10.1007/s00284-010-9673-7

Zeder, M. A. (2015). Core questions in domestication research. *P. Natl. Acad. Sci. U. S. A.* 112 (11), 3191–3198. doi: 10.1073/pnas.1501711112

Zeder, M. A. (2017). Domestication as a model system for the extended evolutionary synthesis. *Interface Focus* 7 (5), 20160133. doi: 10.1098/rsfs.2016.0133

Zeder, M. A., Emshwiller, E., Smith, B. D., and Bradley, D. G. (2006). Documenting domestication: the intersection of genetics and archaeology. *Trends Genet.* 22 (3), 139–155. doi: 10.1016/j.tig.2006.01.007

Zhai, C. K., Lu, C. M., Zhang, X. Q., Sun, G. J., and Lorenz, K. J. (2001). Comparative study on nutritional value of Chinese and North American wild rice. *J. Food Compos. Anal.* 14 (4), 371–382. doi: 10.1006/jfca.2000.0979

Zhang, J. Z., Chu, F. Q., Guo, D. P., Hyde, K. D., and Xie, G. L. (2012). Cytology and ultrastructure of interactions between *Ustilago esculenta* and *Zizania latifolia*. *Mycol. Prog.* 11 (2), 499–508. doi: 10.1007/s11557-011-0765-y

Zhang, Y. F., Cao, Q. C., Hu, P., Cui, H. F., Yu, X. P., and Ye, Z. H. (2017). Investigation on the differentiation of two *Ustilago esculenta* strains - implications of a relationship with the host phenotypes appearing in the fields. *BMC Microbiol.* 17, 228. doi: 10.1186/S12866-017-1138-8

Zhao, Y., Vrieling, K., Liao, H., Xiao, M., Zhu, Y., Rong, J., et al. (2013). Are habitat fragmentation, local adaptation and isolation-by-distance driving population divergence in wild rice *Oryza rufipogon*? *Mol. Ecol.* 22 (22), 5531–5547. doi: 10.1111/mec.12517

Conflict of Interest: 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 © 2019 Zhao, Song, Zhong, Li, Chen and Rong. 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) and the copyright owner(s) 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 Plant Science

ORIGINAL RESEARCH
published: 22 November 2019
doi: 10.3389/fpls.2019.01531

Unlocking the Genetic Diversity and Population Structure of a Wild Gene Source of Wheat, *Aegilops biuncialis* Vis., and Its Relationship With the Heading Time

*László Ivanizs[1†], István Monostori[1†], András Farkas[1], Mária Megyeri[1], Péter Mikó[1], Edina Türkösi[1], Eszter Gaál[1], Andrea Lenykó-Thegze[1], Kitti Szőke-Pázsi[1], Éva Szakács[1], Éva Darkó[1], Tibor Kiss[1], Andrzej Kilian[2] and István Molnár[1,3]**

[1] Agricultural Institute, Centre for Agricultural Research, Martonvásár, Hungary, [2] University of Canberra, Diversity Array Technologies, Canberra, ACT, Australia, [3] Institute of Experimental Botany, Center of the Region Haná for Biotechnological and Agricultural Research, Olomouc, Czechia

OPEN ACCESS

Edited by:
Petr Smýkal,
Palacký University, Czechia

Reviewed by:
Shigeo Takumi,
Kobe University, Japan
Ilaria Marcotuli,
University of Bari Aldo Moro, Italy

***Correspondence:**
István Molnár
molnar.istvan@agrar.mta.hu

[†]These authors share first authorship

Specialty section:
This article was submitted to Plant Breeding, a section of the journal Frontiers in Plant Science

Received: *23 August 2019*
Accepted: *01 November 2019*
Published: *22 November 2019*

Citation:
Ivanizs L, Monostori I, Farkas A, Megyeri M, Mikó P, Türkösi E, Gaál E, Lenykó-Thegze A, Szőke-Pázsi K, Szakács É, Darkó É, Kiss T, Kilian A and Molnár I (2019) Unlocking the Genetic Diversity and Population Structure of a Wild Gene Source of Wheat, Aegilops biuncialis Vis., and Its Relationship With the Heading Time. Front. Plant Sci. 10:1531. doi: 10.3389/fpls.2019.01531

Understanding the genetic diversity of *Aegilops biuncialis*, a valuable source of agronomical useful genes, may significantly facilitate the introgression breeding of wheat. The genetic diversity and population structure of 86 *Ae. biuncialis* genotypes were investigated by 32700 DArT markers with the simultaneous application of three statistical methods—neighbor-joining clustering, Principal Coordinate Analysis, and the Bayesian approach to classification. The collection of *Ae. biuncialis* accessions was divided into five groups that correlated well with their eco-geographic habitat: A (North Africa), B (mainly from Balkans), C (Kosovo and Near East), D (Turkey, Crimea, and Peloponnese), and E (Azerbaijan and the Levant region). The diversity between the *Ae. biuncialis* accessions for a phenological trait (heading time), which is of decisive importance in the adaptation of plants to different eco-geographical environments, was studied over 3 years. A comparison of the intraspecific variation in the heading time trait by means of analysis of variance and principal component analysis revealed four phenotypic categories showing association with the genetic structure and geographic distribution, except for minor differences. The detailed exploration of genetic and phenologic divergence provides an insight into the adaptation capacity of *Ae. biuncialis*, identifying promising genotypes that could be utilized for wheat improvement.

Keywords: *Aegilops biuncialis*, genetic diversity, DArTseq markers, population structure, hierarchical clustering, heading time

INTRODUCTION

The genome of wild relatives of common wheat (*Triticum aestivum* L.) can be considered as a potential reservoir of gene variants for wheat improvement (Schneider et al., 2008; Zhang et al., 2015; Kishii, 2019). Interspecific hybridization is a promising approach to enlarge the genetic diversity of cultivated bread wheat by the chromosome-mediated transfer of the wild alleles present in related species (Jauhar, 1993; Jauhar and Chibbar, 1999; Kishii, 2019). Goatgrasses (*Aegilops*),

which comprise 11 diploid, 10 tetraploid, and 2 hexaploid species, are the closest relatives of *Triticum* (van Slageren, 1994). Seven different genomes (D, S, U, C, N, M, and T) were identified in the diploid species, indicating the extreme genetic diversity of the genus. Accessions of several *Aegilops* species are highly resistant to important cereal diseases (Friebe et al., 1996; Marais et al., 2006; Bansal et al., 2017; Olivera et al., 2018), while others shows good tolerance of abiotic stresses such as salt, drought, frost, and heat stress (Rekika et al., 1997; Zaharieva et al., 2001a; Zaharieva et al., 2001b; Molnár et al., 2004; Colmer et al., 2006; Dulai et al., 2014). Some alleles associated with these agronomic traits have already been introgressed from *Aegilops* into the wheat gene pool by the development of wheat-*Aegilops* hybrids and addition or translocation lines (Schneider et al., 2008; Marais et al., 2009; Liu et al., 2011a; Liu et al., 2011b; Olson et al., 2013). However, the genetic potential of *Aegilops* is still largely underutilized. In the case of biotic stresses, the 41 resistance genes that have so far been integrated into the wheat genome originated from only 30 accessions from 12 *Aegilops* species, most of them belonging to the primary gene pool of hexaploid wheat (Zhang et al., 2015). There are numerous *Aegilops* accessions in gene banks in various parts of the world (Monneveux et al., 2000) that have not yet been utilized for wheat improvement, so their introduction into breeding programs would be desirable.

The annual allotetraploid *Aegilops biuncialis* Vis. ($2n = 4x = 28$; $U^bU^bM^bM^b$) is largely autogamous and belongs to the tertiary gene pool of bread wheat (Friebe et al., 1996). *Ae. biuncialis* is native to Mediterranean and Western Asiatic regions and populations can be found in the Aegean, Turkey, Bulgaria, Cyprus, in the western part of the Fertile Crescent, in Cis- and Transcaucasia, and in the southern parts of Russia and Ukraine (van Slageren, 1994; Kilian et al., 2011). The annual rainfall in these habitats ranges from 225–1250 mm and some of them are characterized by a dry summer season with high temperature and high irradiance (van Slageren, 1994). The wide eco-geographical distribution suggests the presence of great diversity in the phenological traits of *Ae. biuncialis*. The heading time, as one of the main phenological factors, is crucial for the ecological adaptation of plants to local conditions. Early flowering, as an avoidance mechanism, may have a major role in the adaptation of plants to a Mediterranean climate by allowing them to escape drought (Araus et al., 2004). The adaptation strategy of wild emmer wheat populations to natural habitats that are characterized by frequent high temperatures in spring includes early flowering (Peleg et al., 2005), while adaptation to habitats with high altitudes involves late heading. Nevertheless, no data have been published on the intraspecific variation of heading time in accessions representing the broad ecological adaptability of *Ae. biuncialis*, which exhibits great genetic diversity.

The genetic diversity of wild species that are gene sources for cultivated wheat is a critical component of research in evolution, population genetics, conservation and breeding (Tahernezhad et al., 2010; Wang et al., 2013; Arora et al., 2017; Edet et al., 2018; Etminan et al., 2019; Singh et al., 2019). Little has been reported about the genetic diversity of *Ae. biuncialis*. Studies based on amplified fragment length polymorphism (AFLP), sequence-specific amplified polymorphism, random amplified polymorphic DNA, and inter-simple sequence repeat molecular markers were used to reveal the genetic variability within the genome of *Ae. biuncialis* accessions (Okuno et al., 1998; Monte et al., 2001; Nagy et al., 2006; Thomas and Bebeli, 2010), but Nagy et al. (2006) concluded that the marker data of only 10 accessions cannot demonstrate the existing genetic variability in *Ae. biuncialis*, so it is essential to investigate a larger diverse collection of genotypes with high marker density for the comprehensive study of genetic diversity. The 5–10 accessions so far assessed using a larger number of loci all originated from narrow geographical regions (Greece, the Iberian Peninsula or Transcaucasia), so do not represent the wide distribution of the species and do not allow an objective picture to be obtained of the genetic diversity existing within *Ae. biuncialis* (Okuno et al., 1998; Monte et al., 2001; Thomas and Bebeli, 2010). Moreover, different parts of the genome may have undergone different evolutionary changes (Kellogg et al., 1996). In the Triticeae, structural rearrangements frequently occur in the pericentric region of the chromosomes after polyploidization (Qi et al., 2006), as also revealed in *Ae. biuncialis* accessions (Molnár et al., 2011). Additionally, the frequency of recombination and the rate of interstitial deletions and insertions of gene loci is much higher in the distal third of the chromosomes than in the proximal two-thirds (Dvorak and Akhunov, 2005). As a consequence, only part of the genetic polymorphisms present in the species can be detected if the genome is not sufficiently covered by molecular markers.

Diversity Array Technology (DArT) was originally developed as a hybridization-based microarray platform to detect polymorphism at the recognition sites of methylation-sensitive restriction enzymes (Jaccoud et al., 2001; Wenzl et al., 2004). More recently, a combination of this genome complexity reduction approach of DArT technology with next-generation sequencing technologies resulted in a sequence-independent, low-cost, high-throughput genotyping by sequencing platform allowing the simultaneous detection of several thousands of polymorphic loci spread over the genome (Wenzl et al., 2004; Tinker et al., 2009; Sansaloni et al., 2011; Kilian et al., 2012). The advanced DArTseq technology has been efficiently utilized for genotyping, genetic diversity analysis, genome-wide association studies and linkage mapping in cultivated and wild relatives of wheat such as Triticale (Badea et al., 2011), rye (Bolibok-Brągoszewska et al., 2014; Gawroński et al., 2016), barley (Comadran et al., 2011), Tibetan wild barley (Cai et al., 2013), durum wheat (Baloch et al., 2017), *Aegilops tauschii* (Kumar et al., 2015), *Triticum monococcum* (Jing et al., 2009), and hexaploid wheat cultivars (Nielsen et al., 2014; Monostori et al., 2017).

In the present work, genetic diversity and its association with the phenotypic variation in heading time were studied in a collection of 86 *Ae. biuncialis* genotypes to obtain a better understanding of the background of its genotypic and phenotypic diversification. To achieve this goal, the genetic relationships between the accessions were analyzed using three statistical methods after the DArTseq genotyping of the plants. Furthermore, the phenological variation pattern was compared with the genetic structure and eco-geographical distribution to obtain information on possible correlations between them. The comprehensive study of how changes in phenological traits relate with genetic diversity will

facilitate the utilization of the enhanced adaptability of *Ae. biuncialis*, especially in the light of climate change.

MATERIALS AND METHODS

Plant Material

Eighty-six wild *Ae. biuncialis* Vis. (2n = 4x = 28, $U^bU^bM^bM^b$) accessions, collected from 64 sites in 16 countries from Libya to Azerbaijan, were genotyped together with the Mv9kr1 wheat accession on a DArTseq® platform (**Table 1** and **Supplementary Table S1**). This population, originating from a wide range of ecological habitats, is representative of the geographical distribution of the species. The *Aegilops* genotypes were provided by the following germplasm collections: Genebank of the Agricultural Institute, Agrártudományi Kutatóközpont (ATK) (Martonvásár, Hungary), Wheat Genetics Resource Center (Kansas State University, USA), United States Department of Agriculture Agricultural Research Service (Beltsville, MD, USA), and Institute of Plant Genetics and Crop Plant Research (Gatersleben, Germany). The *A. biuncialis* accessions were maintained and grown in the Department of Plant Genetic Resources of ATK Mezőgazdasági Intézet. To avoid hybridization and intercrossing between the genotypes, the isolated accessions were multiplicated by self-fertilization.

Field Experiments

The collection of 86 *Ae. biuncialis* accessions was grown under field conditions in three consecutive seasons (2015–16, 2016–17, and 2017–18). The seeds were planted in fall in fields belonging to the Agricultural Institute of ATK (Breeders nursery, Martonvásár, Hungary, geographic coordinates: 47°19'39"N, 18°47'01"E). The soil type at each location was chernozem. Each genotype was grown in a 6 m^2 plot with 6 × 3 m rows, 50 seeds per row, and a row distance of 0.15 m, as previously described by Mikó et al. (2014). The heading time of each plot was recorded and defined as the period elapsing between January 1 and the day when 50% of the inflorescences reached the DEV59 developmental stage (Tottman, 1987).

Additionally, to monitor the stay-green ability of the plants, the leaf chlorophyll content of each *Ae. biuncialis* genotype was determined twice (on May 12 and 31, 2018) during the grain-filling period using a Soil Plant Analysis Development 502 cholorophyll meter (Minolta Camera Co., Ltd, Tokyo, Japan). At least 10 leaves were measured from randomly selected plants for each genotype on each occasion. For each leaf, the average of three SPAD readings around the midpoints of the flag leaf was taken.

TABLE 1 | Geographic distribution of the *A. biuncialis* collection maintained in Martonvásár.

Country	No. of accessions	Ratio (%)	Climate zone
Libya	4	4.65	Semi-arid Mediterranean
Bosnia and Herzegovina	4	4.65	Mediterranean and Continental
Serbia	2	2.33	Humid Continental
Macedonia	2	2.33	Humid Continental
Bulgaria	1	1.16	Humid Continental
Kosovo	1	1.16	Mediterranean and Continental
Greece	9	10.46	Mediterranean Humid Continental
Turkey	25	29.07	Mediterranean
Ukraine (Crimean)	3	3.49	Mediterranean and Continental
Cyprus	2	2.33	Mediterranean
Azerbaijan	7	8.14	Semi-arid Mediterranean Humid Continental
Iraq	1	1.16	Semi-arid
France	1	1.16	Mediterranean
Syria	11	12.79	Semi-arid Mediterranean
Jordan	6	6.98	Semi-arid Mediterranean
Israel	1	1.16	Mediterranean
Unknown	6	6.98	
Total	**86**	**100**	

The geographic origin of the Ae. biuncialis genotypes can be found in the Plant Inventory Books: USDA ARS, European Wheat Database and Genesys Accession browser. Climatic conditions at the site of origin of the Ae. biuncialis accessions, expressed according to the Köppen-Geiger classification system.

DNA Extraction and Genotyping

Genomic DNA was extracted from young leaves of 86 *Ae. biuncialis* genotypes together with hexaploid wheat (*Triticum aestivum* L.) genotype Mv9kr1 using Quick Gene-Mini80 (FujiFilm, Japan) with a QuickGene DNA tissue kit (FujiFilm, Japan) according to the manufacturer's instructions (Cseh et al., 2011).

The DNA of the *Ae. biuncialis* accessions and the Mv9kr1 wheat genotype was sent for a high-throughput genotyping commercial service to Diversity Arrays Technologies Pty. Ltd., Australia (http://www.diversityarrays.com). The "wheat DArTseq™ 1.0" genotyping by sequencing service was optimized for wheat. This method used the combination of complexity reduction methods developed initially for array-based DArT and multiplex sequencing in Illumina HiSeq2500 instrument (Sansaloni et al., 2011; Kilian et al., 2012). Briefly, DArT markers are DNA fragments obtained by the genome complexity reduction method contains the digestion of genomic DNA by *Pst*I/*Taq*I endonucleases, ligation of the genomic fragments to a *Pst*I adapter, amplification using a primer complementer with the adaptor sequence and transformation into *Escherichia coli*. DNA fragments obtained by this genome complexity reduction method were sequenced by Illumina HiSeq2500.

The markers developed by the above mentioned "wheat DArTseq™ 1.0" genotyping by sequencing platform have been included in **Supplementary Data S1**. The obtained silicoDArT marker set was filtered on the basis of individual marker-related statistics. After removing markers with inappropriate quality control parameters, including call rate <90%, reproducibility <95% and minor allele frequency < 2%, the 0/1 binary matrix of the remaining 32,700 markers was used in the subsequent analysis.

Analysis of Population Structure and Phylogenetic Relationship

In order to assess the population structure of the *Ae. biuncalis* accessions, three different statistical methods were adopted and compared. First, a clustering approach based on the Bayesian

model was applied to estimate the real number of subpopulations (K) using the admixture model of STRUCTURE software 2.3.4 with correlated allele frequencies (Pritchard et al., 2000). Three independent runs were performed for each hypothetical number of subpopulations (K) from one to eight applying a burn-in period of 100,000 iterations followed by 100,000 Markov Chain Monte Carlo iterations to obtain a precise parameter estimate. The most probable number of subpopulations was determined by means of the ΔK method using STRUCTURE HARVESTER software (Evanno et al., 2005). Each genotype was assigned to one subpopulation based on its membership probability. Principal coordinate analysis was then carried out using PAST software version 3.12 to visualize the genetic stratification within the *Ae. biuncialis* collection based on genetic correlations among individuals (Hammer et al., 2001). Thirdly, the phylogenetic relationship between the 86 accessions was estimated using PAST software with the neighbor-joining method based on marker data. The genetic distance matrix based on Jaccard's similarity coefficient was applied to construct a phylogenetic tree. The neighbor-joining dendrogram was generated using bootstrap analysis with 1,000 replicates. The wheat genotype Mv9kr1 was used as outgroup species in order to define the root position of the phylogenetic tree.

Statistical Analysis

The phenotypic values of the *Ae. biuncialis* accessions were grouped based on the subpopulations obtained in the STRUCTURE cluster analysis. The variations in heading time among the subpopulations were analyzed for 3 years using descriptive statistics: mean, standard deviation, boxplot, and histogram (Excel 2016, Microsoft Company).

One-way analysis of variance (ANOVA) (SPSS 16.0, IBM) was conducted to compare the heading time of the *Ae. biuncialis* subpopulations in each year at the $P < 0.05$ significance level. In addition, the significant differences in the heading time of each subpopulation were examined among the 3 years by means of one-way ANOVA (SPSS 16.0). Bivariate correlation of the heading times of the *Ae. biuncialis* accessions among the 3 years was analyzed pair-wise by means of Pearson coefficient (SPSS 16.0). Linear regression analysis was carried out for each subpopulation among the years.

The principal component analysis of the Statistica 6 software was used to study associations between the heading times of the *Ae. biuncialis* accessions by visualizing the grouping pattern of phenotypic variations.

The stay-green trait (delayed foliar senescence) was determined by calculating differences between the SPAD values measured on the two dates. Pearson and rank correlation (Spearman, Kendall tau) coefficients were applied to analyze the dependence and relationship between the heading time and the differences in SPAD values at the $P < 0.01$ significance level (SPSS 16.0).

RESULTS

Population Stratification

A total of 47,777 polymorphic dominant silicoDArT markers were generated from 86 *Ae. biuncialis* accessions originating from different eco-geographical habitats (**Supplementary Data S1**). The silicoDArT marker set was filtered for various quality parameters (call rate, reproducibilty, minor allele frequency) and reduced to 32,700 markers. Based on the marker data, the genetic diversity in the *Ae. biuncialis* collection was analyzed using the Bayesian clustering approach performed with STRUCTURE software. After the STRUCTURE analysis had been run for K = 1 to K = 8, the most likely number of subpopulations was estimated with the STRUCTURE Harvester software following the ΔK method. The maximum ΔK value occurred at K = 5 (**Figure 1** and **Supplementary Figure S1**). Each genotype was assigned to one of the subpopulations based on a membership probability coefficient > 0.51.

The *Ae. biuncialis* accessions were grouped into five clusters A, B, C, D, and E with 6, 14, 3, 31, and 29 accessions, respectively. Three accessions had mixed allelic patterns that could not be assigned to any of the subpopulations with a probability of more than 50% (mixed group). The accessions clustered in accordance with their origin: the accessions in cluster A originated mainly from North Africa; cluster B represented genotypes from the Central Balkans, North Greece, and West Turkey; the three genotypes in cluster C originated from different geographic regions (Balkans and Near East); cluster D contained accessions from Asia Minor, the Crimean Peninsula, and Southern Greece (Peloponnese); while cluster E comprised genotypes from the Levant region (Jordan, Syria, and South Turkey) and Azerbaijan (**Figures 2** and **3**).

Based on the membership probability of genotypes, subpopulations A, B, and C could be clearly separated, whereas D and E were less distinct due to their similar share in the genome of several accessions (**Figure 2**). Subpopulation A was differentiated and sub-divided according to the membership probability of each cluster: Three of the six accessions (AE78689, AE84388, and AE84484) segregated into a homogeneous subgroup, but the

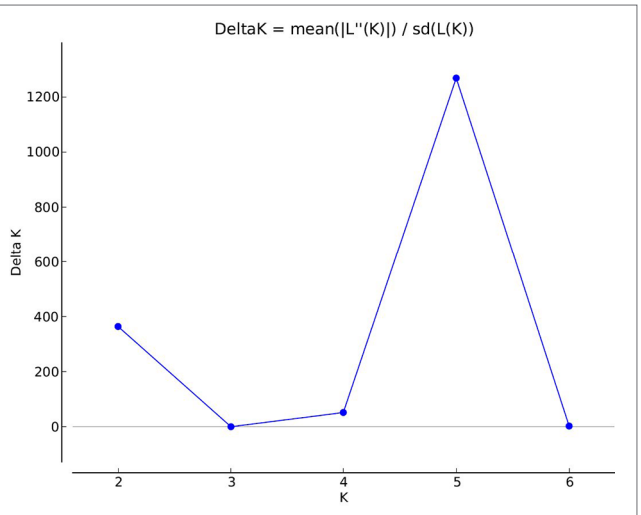


FIGURE 1 | Estimation of the genetically most probable number of *Ae. biuncialis* subpopulations based on Evanno's Delta K method (Evanno et al., 2005). The maximum value of delta K occurred at K = 5 indicates that the investigated *Ae. biuncialis* collection can be devided into five hypothetical subpopulations.

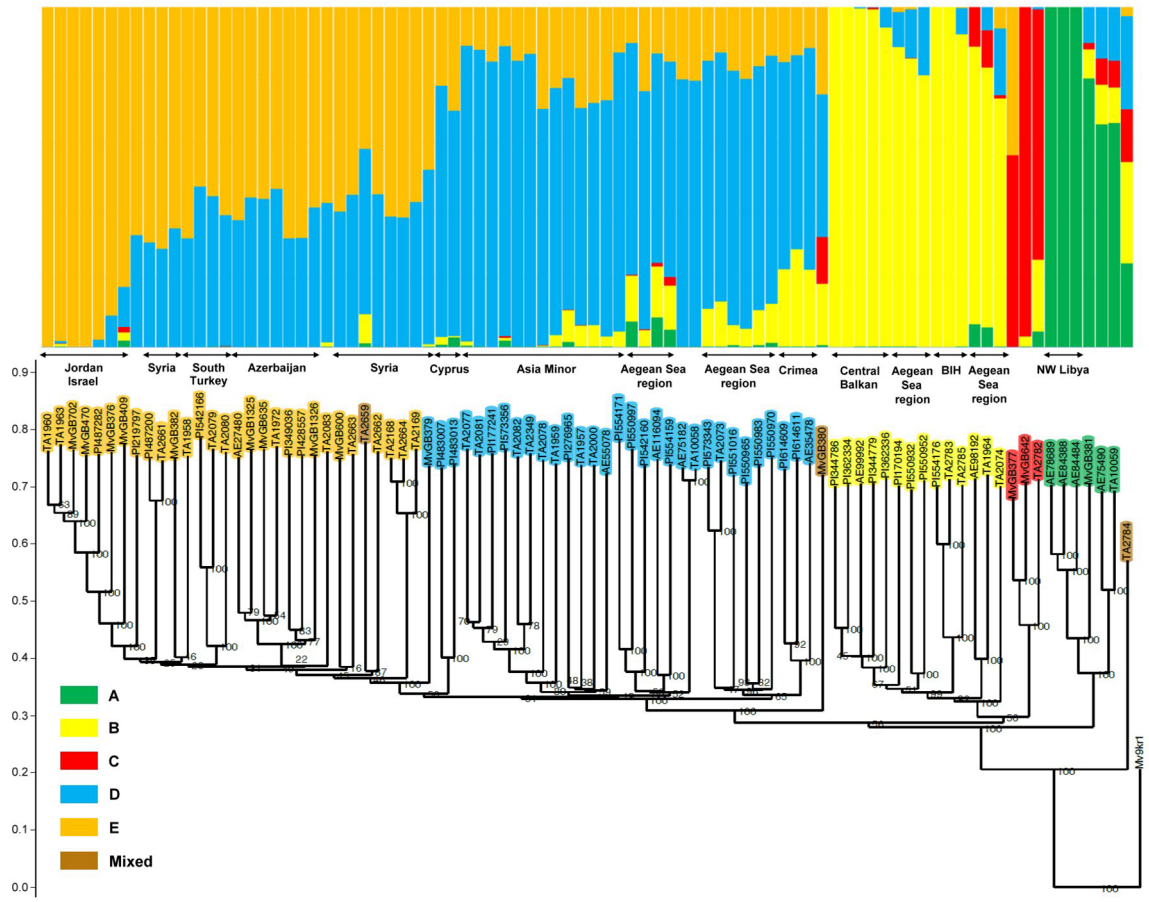


FIGURE 2 | Comparison of grouping patterns based on STRUCTURE analysis at K = 5 and the neighbor-joining dendrogram in a collection of 86 *Ae. biuncialis* accessions. The order of the genotypes on the STRUCTURE plot matches that represented on the phylogenetic tree. The wheat genotype Mv9kr1 is represented as an outgroup accession in the dendrogram. Genotypes belonging to the same clade represent similar membership probability in the relevant subpopulation and are from a common area within the larger geographic regions. BIH: Bosnia and Herzegovina.

other three formed a heterogeneous subgroup. Subpopulation A appeared in different proportions in the genome of the two subgroups. Subpopulation D contributed to the gene pool of some accessions in subpopulation B whereas the B subpopulation was also detectable in several genotypes in subpopulation D. These accessions within the two subpopulations occupy a common eco-geographic area in Greece and West Turkey (Aegean Sea region), indicating admixture events between subpopulations B and D. In subpopulation C, other subpopulations also significantly contributed to the genome of two accessions (TA2782 and MvGB377). Subpopulation B was revealed in genotype TA2782 from Kosovo, whereas subpopulation E was detected in the genome of MvGB377, originating from Jordan. The third, Syrian genotype (MvGB642) had a 97% membership probability in cluster C.

Analysis of Genetic Diversity

To refine the genetic relationship between the *Ae. biuncialis* accessions, the genetic similarity within the population was calculated based on Jaccard's coefficient. The average genetic similarity between all 86 genotypes was estimated to be 0.45 and ranged from 0.28 to 0.96, indicating that the *Ae. biuncialis* population had large genetic variability (data of the genetic similarity matrix not shown). The phylogenetic dendrogram split the *Ae. biuncialis* collection into five separate branches, which had a good fit with the subpopulations (A-E) revealed by STRUCTURE analysis (**Figure 2**). The pair-wise genetic similarity of the accessions was studied within each subpopulation, of which cluster D had the lowest average value (0.51) and the widest range (0.41–0.96), indicating its broad genetic diversity. The *Ae. biuncialis* accessions were genotyped together with wheat accession, Mv9kr1 which was used as an outgroup accession for constructing the phylogenetic dendrogram for the *Ae. biuncialis* collection.

Based on the STRUCTURE and neighbor-joining analyses results, the subpopulations could be further divided into subgroups (or intra-cluster lineage), where genotypes from common areas share similar genetic ancestry and show high genetic similarity (**Figure 2**). This intra-cluster differentiation was in agreement with the provenance of the genotypes within the larger geographic regions. Cluster B could be separated into subgroups originating mainly from Bosnia and Herzegovina, the central area of the Balkans and coast of the Aegean Sea (North

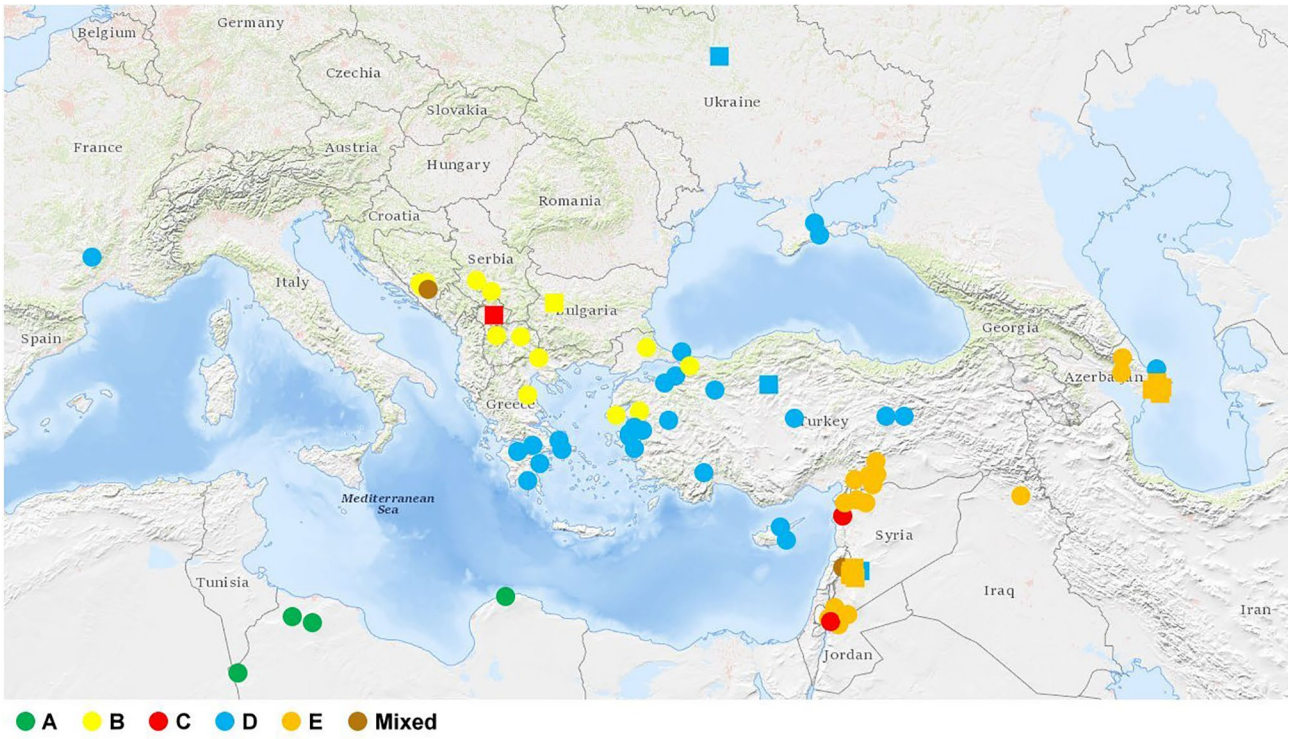


FIGURE 3 | Geographic origin of the *Ae. biuncialis* collection. Circles represent accessions with known geographic location. Where only the country of origin is known, it is indicated by a square representing the capital. There is no information about the origin of six accessions in the collection. The different colors of the symbols (circle, square) in the figure represent the clusters (or subpopulations) corresponding to the population structure obtained by STRUCTURE analysis.

Greece and Northwest Turkey). Subpopulation D was divided into clades springing from Asia Minor, Crimea and region of the Aegean See (South Greece and West Turkey). Based on their genetic similarity, three accessions from Cyprus and Syria (MvGB379, PI483007, and PI483013) could be discriminated clearly in subpopulation D as showing a close genetic relationship with subpopulation E. The latter could be partitioned into distinct subgroups representing their geographic origin, such as Jordan, Azerbaijan, Syria and the southern part of Turkey.

Principal coordinate analysis was used as an alternative way of analyzing and visualizing the population structure. The first three principal coordinates explained 38.428% of the genotypic variance (PCo1: 21.46%, PCo2: 10.81%, PCo3: 6.15%), and also discriminated the accessions of clusters A, B, D and E, confirming the STRUCTURE result (**Figure 4**, **Supplementary Figures S2** and **S3**).

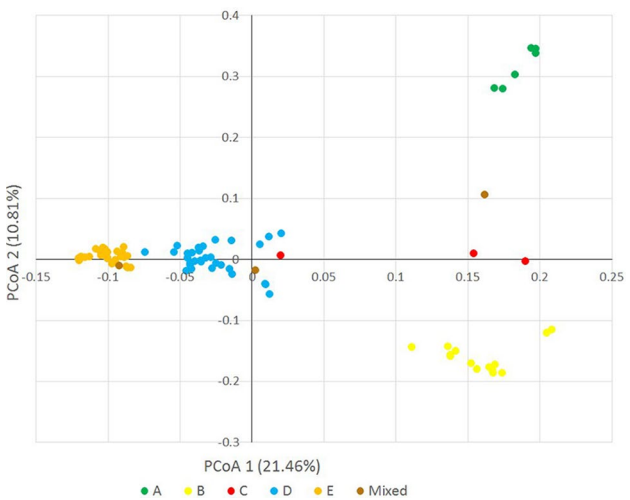


FIGURE 4 | Principal coordinate analysis on the *Ae. biuncialis* accessions based on 32,700 DArT markers. Different colors indicate different subpopulations (A–E) in the population.

Phenotypic Evaluation of Heading Time

The heading time, as a phenological trait, was grouped according to the subpopulations obtained by cluster analysis (**Supplementary Table S2**). One-way ANOVA was carried out in order to study the differences between the heading time of the five subpopulations within each year, and to analyse the differences in each subpopulation among the 3 years. Although subpopulations A and C had much smaller sample size than the others, they were not excluded from the statistical analysis.

The heading time within each subpopulation differed not significantly among 2016 and 2017, while the heading time values in the 2018 season were considerably shorter than in the other 2 years (**Figure 5** and **Supplementary Figure S4**),

which was confirmed by the correlation analysis (**Figure 6**). The difference could be explained by the very different precipitation and temperature conditions during the spring months that are critical for the flowering period (**Supplementary Figure S5**). One-way ANOVA of the heading time within 1 year pointed out significant differences between four subpopulations (A, B, D, and E) in the *Ae. biuncialis* collection (**Figure 5** and **Supplementary Figure S4**). The subpopulation C was not separated from those of subpopulations D and E; and in 2018 the subpopulation C differed not from those of A, D, and E.

To visualize the correlation between genetic diversity and phenotypic variation, principal component analysis was performed for the heading time of the *Ae. biuncialis* accessions (**Figure 7**, **Supplementary Figures S6** and **S7**). The first two components explained 97.29% of the phenotypic variance and grouped the majority of genotypes into four phenotypic categories, which corresponded with the genetic structure and geographic distribution of the population in spite of the fact that they showed not clearly differenctiation (**Figure 7**). Since the third principal component represented only 2.72% of the variation,

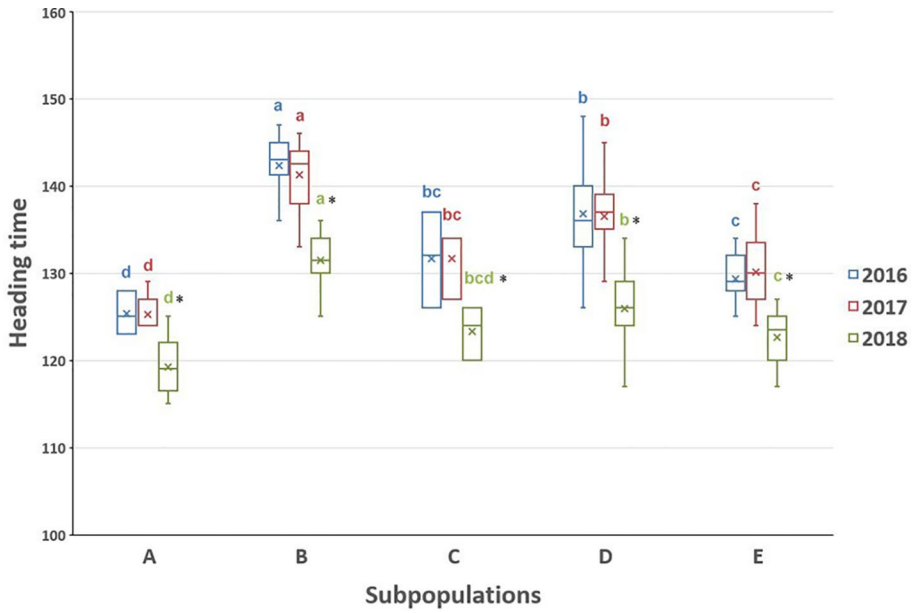


FIGURE 5 | Boxplot chart of the heading times of the *Ae. biuncialis* subpopulations **(A–E)** showing the mean, median and range of the phenotypic data in 3 years. The heading time trait was expressed in number of days elapsed from January 1 to the DEV59 developmental stage. Different letters indicate significant differences between the subpopulations within the same year at $P < 0.05$, using one-way analysis of variance. *Indicates the heading time in 2018 differs significantly from the others within relevant subpopulation at the $P < 0.05$ level.

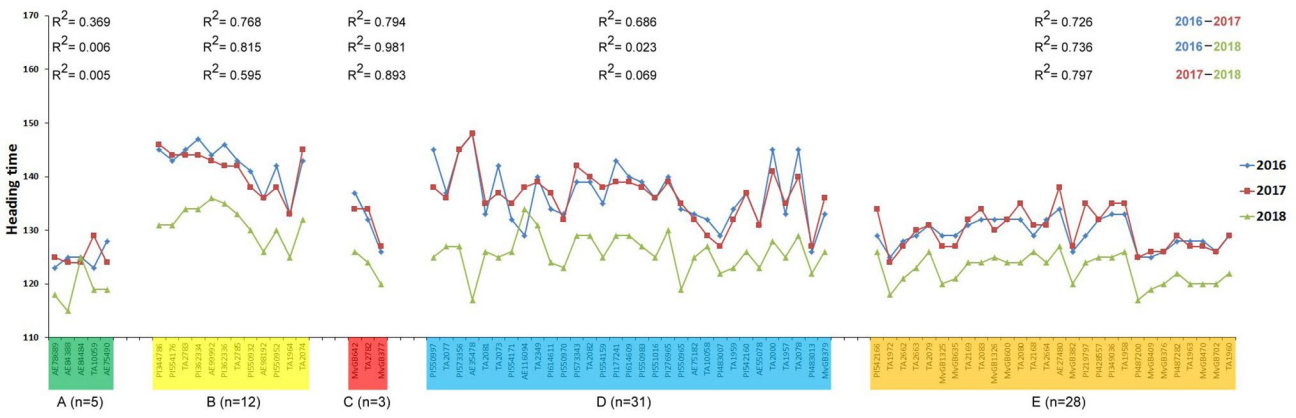


FIGURE 6 | Correlation of the heading times of *Ae. biuncialis* accessions among three different seasons. Heading time was defined as the number of days required to reach the DEV59 developmental stage from January 1. Color codes indicate different subpopulations **(A–E)** identified from STRUCTURE analysis. Number of accessions (n), clustered into one subpopulation, are represented in the figure. The R^2 values of pair-wise correlation for each subpopulation among different seasons are included in the figure.

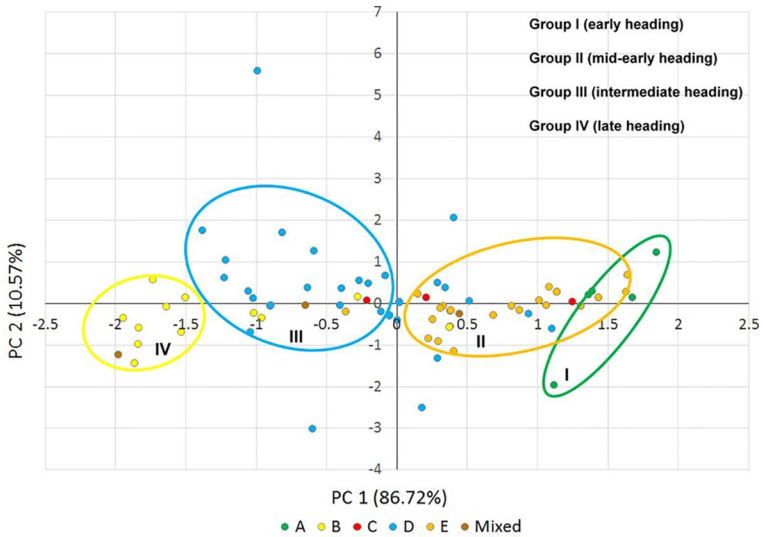


FIGURE 7 | Scatter plot of principal components 1 and 2, explaining 97.29% of the phenotypic variance in the heading time (2016, 2017, and 2018) of the *Ae. biuncialis* collection. Different colored dots indicate the *Ae. biuncialis* accessions in the subpopulations (A–E) obtained from STRUCTURE analysis. Groups I–IV indicate the phenotypic categories for heading time.

it was not taken into account in the phenotypic classification. Subpopulations A and E correlated to group I and group II, respectively, exhibiting the earlier heading phenotypes, whereas subpopulation D largely coincided with the intermediate group III. Subpopulation B corresponded the group IV, which had the latest heading date (**Figure 7**). The heading time of subpopulation C could not be classified in a separate phenotypic group, which was in accordance with the results of one-way ANOVA.

Analysis of Delayed Foliar Senescence

The relationship between delayed foliar senescence (stay-green trait of plants) and the heading time was also investigated. At the first measuring date (May 12), the average SPAD values of the genotypes ranged between 42 and 61, indicating that these plants were still green (**Supplementary Table S3**, **Supplementary Figures S8** and **S9**). In the second phase (May 31) SPAD values could not be recorded for seven *Ae. biuncialis* accessions, because the green color intensity of their leaves was below the detection level due to their advanced maturity. For the other genotypes, a significant correlation was revealed between the heading time and the difference in SPAD values based on the Pearson and rank coefficients. This revealed eight genotypes which had early heading times, but exhibited the small differences in SPAD values that are characteristic of the stay-green phenotype (**Supplementary Table S3** and **Supplementary Figure S10**).

DISCUSSION

Genetic Structure of the *Aegilops biuncialis* Collection

The present study is the first report on the use of the DArTseq platform to estimate the intraspecific genetic diversity of *Ae. biuncialis* on accessions representing the natural distribution area of the species. Three statistical methods gave consistent results for the genetic diversity and population structure: The *Ae. biuncialis* accessions clustered into five subpopulations in accordance with their place of origin. Three genotypes (MvGB380, TA2784, and TA2659) could not be classed into any of the distinct subpopulations based on their membership probability and these showed the lowest genetic similarity to the other accessions, indicating that this mixed group was the most distant genetically from the other genotypes in the population.

Based on random amplified polymorphic DNA markers, *Ae. biuncialis* accessions from Transcaucasia were classed by Okuno et al. (1998) in two clusters, which corresponded with groups from the eastern shores of the Black Sea and the western shores of the Caspian Sea, while *Ae. biuncialis* subpopulations from the central and southern parts of the Iberian Penninsula were clearly discriminated by AFLP markers that correlated with their geographical distribution (Monte et al., 2001). Based on simple sequence repeat, AFLP, and single nucleotide polymorphism markers, *Ae. tauschii* genotypes were grouped in two lineages, showed clear differentiation (Pestsova et al., 2000; Mizuno et al., 2010; Jones et al., 2013; Wang et al., 2013; Arora et al., 2017; Singh et al., 2019). Large range of genetic similarity (0.61 to 0.99) was reported for *Ae. tauschii* by Sohail et al. (2012), who used DArT markers to study the genetic relationship between 81 genotypes representing the eco-geographic distribution of the species. In the present study the *Ae. biuncialis* collection had a relatively wide genetic variability (0.28 to 0.96). When comparing the genetic diversity of various *Aegilops* species, Kilian et al. (2007) found lower genetic similarity in *Aegilops speltoides* (0.656) and *Ae. tauschii* (0.75) than in *Aegilops sharonensis* (0.84), *Aegilops searsii* (0.825), and *Aegilops longissima* (0.812). Species in the *Sitopsis* section, except for *Ae. speltoides*, have limited distribution in

the Levantine, representing a smaller geographical area in the eastern Mediterranean Basin (van Slageren, 1994; Mendlinger and Zohary, 1995; Millet, 2007; Kilian et al., 2011). The narrow geographic range of *Ae. sharonensis*, restricted to the coastal areas of Levantine, could explain the high average genetic similarity (0.82) estimated between the genotypes (Olivera and Steffenson, 2009; Olivera et al., 2010). AFLP marker analysis on a set of *Aegilops geniculata* genotypes revealed that accessions originating from the main regions of the Mediterranean Basin could be classed into seven groups (Arrigo et al., 2010). Six of the seven groups had a strong bio-geographical structure, derived from the northern parts of the Mediterranean, whereas the remaining one exhibited less genetic structuring, representing the southern Mediterranean areas, including North Africa, Levantine, and the Bosphorus Strait. In contrast, the *Ae. biuncialis* population in the southern and eastern Mediterranean regions could be divided into distinct subpopulations, where subpopulation A occupied North Africa, subpopulations E and C were characteristic of Levantine and subpopulation D was located mainly in Asia Minor. The *Ae. biuncialis* collection, fragmented in the eastern part of the Mediterranean Basin, showed a distinct phylo-geographical pattern, indicating the extremely large genetic diversity of the species.

The subpopulations of the *Ae. biuncialis* collection could be further divided into subgroups, resulting in intra-cluster differentiation, which may explain the admixture events between the genotypes (**Figure 2**). Three of the six accessions in subpopulation A had a close genetic relationship, suggesting that the higher rate of gene flow between them may have resulted in a homogeneous subgroup. They differ phenotypically from the rest of the subpopulation as they have an earlier heading date. As this homogeneous subgroup is derived from the northwest part of Libya, they were presumably isolated from the other members of the subpopulation while adapting to the arid climate. The natural distribution areas of certain accessions, representing an admixture of subpopulations B and D, overlap in the region of the Aegean Sea, suggesting that the genome shared by the two subpopulations was formed due to the intercrossing of parental genotypes derived from the different subpopulations. Three accessions (MvGB379, PI483007, and PI483013) were grouped in subpopulation D based on their membership probability, whereas they also showed a closer phylogenetic relationship with some Syrian genotypes from subpopulation E. As both subpopulations have made similar contributions to the genome of these accessions, their genetically similar germplasm and recent common ancestor suggest that the two clusters have not yet fully separated phylogenetically or that intraspecific hybridization has occurred between them. Subpopulations B and C contributed in similar proportions to the genome of the accession TA2782 originating from Kosovo, postulating an earlier admixture event between ancestral genotypes belonging to different subpopulations. Although three genotypes are not enough to represent the native distributional range of subpopulation C, the TA2782 accession seems to have developed outside this area and may have reached Kosovo as a result of migration. Genotypes AE75182 and TA10058 have close genetic similarity (0.95) in spite of being from distant regions. TA10058 originated from Azerbaijan, but was found well away from the natural distribution area of its subpupolation, so it might be an introduced accession. Herbivores, as vectors, can support the long distance dispersal of plants (Janzen, 1984; Vellend et al., 2003; Römermann et al., 2005), which is enhanced by human activities, such as pastoralism (Blondel, 2006). On the other hand, human-caused climate change may have an impact on the expansion of numerous species (Parmesan and Yohe, 2003; Thomas et al., 2004; Hickling et al., 2006; Jarvis et al., 2008; Devictor et al., 2012). The occurrence of two genotypes (PI614609 and PI614611) in the Crimean Peninsula corresponds to the prognosis that predicts a probable shift in the natural range of *Ae. biuncialis* to the Azov Sea in response to global warming (Ostrowski et al., 2016).

The large genetic diversity of *Ae. biuncialis*, enabling its adaptation to various climatic conditions (Zaharieva et al., 2004; Zaharieva and Monneveux, 2006), may serve as a rich source of potential genes to improve the adaptive capacity of cultivated wheat. It has been reported that some *Ae. biuncialis* accessions have good tolerance of heat and drought stress (Molnár et al., 2004; Dulai et al., 2014) and others are promising for resistance to rust diseases (Molnár-Láng et al., 2014; Olivera et al., 2018).

Phenotypic Diversity Pattern

Based on one-way ANOVA and correlation analysis (**Figures 5** and **6**), the heading times differed among subpopulations similarly in each year in spite of the fact that the heading times of all subpopulations was shorter in 2018 than in 2016 and 2017. The extrem weather conditions, observed in the 2018 season, can explain the low level of correlation and shorter phenotypic values. Significant differences in heading date was found between four subpopulations (A, B, D, and E), corresponding with genetic relatedness and geographic origin with the exception of subpopulation C, which could not be separated from subpopulations D and E (**Figures 5** and **7**). In subpopulation C, the variation pattern of the heading date suggested that the admixed genomes of these accessions represented a diverse gene pool, and had different geographic origin. However, subpopulation C contained relatively few accessions, so no additional conclusions can be drawn on its phenotypic diversity.

Subpopulations with different phenotype are originating from eco-geographic regions that can be characterized by different agro-climatic conditions. Subpopulation A from North Africa had early phenotype which is prevalent in the semi-arid area, whereas subpopulation B and D from the Balkans and Asia Minor exhibited later heading times, which are predominant in the Continental and Mediterranean regions, respectively. The phenological trait patterns of the subpopulations correlated largely with the climatic conditions found in the relevant geographical areas, suggesting that the heading time diversification might be an adaptive change for *Ae. biuncialis*. The broad genetic diversity and its association with adaptive changes in heading date allowed the species to spread widely. Nevertheless, genealogically-based chloroplast DNA analysis could lead to a more comprehensive identification of the ancestral and derived sublineages and a better understanding of the course of the geographic spread of *Ae. biuncialis*.

The *Aegilops triuncialis* accessions in California, which evolved due to the independent introduction of a few Eurasian genotypes (Peters et al., 1996), were grouped using microsatellite markers into three lineages, each of which had different eco-geographical distribution and showed significant differences in flowering time (Meimberg et al., 2006; Meimberg et al., 2010). However, the phenotypic diversity pattern of the invasive lineages cannot express the intraspecific variation for flowering time because the introduced accessions represent only a part of the natural distribution area of *Ae. triuncialis*.

The correlation between the population structure and flowering time diversification in a collection of 200 Ae. *tauschii* accessions was studied by Matsuoka et al. (2008) using chloroplast DNA geneological analysis. Two of the four major haplogroups (HG7 and HG16) were phenotypically heterogeneous, representing a large proportion (83.5%) of the population. When comparing the DArT marker-based genetic divergence and phenological variation pattern of *Ae. tauschii*, Matsuoka et al. (2015) found that the intraspecific lineage structure was associated with changes in the flowering trait. This finding has been supported by the present study, as the *Ae. biuncialis* subpopulations A, B, D, and E showed significant differences in the heading time. The broad intraspecific diversity of the heading phenotypes could be introduced into bread wheat by means of chromosome-mediated gene transfer in order to diversify the maturation time and prolong the harvest period.

Besides the heading time, the stay-green trait, which is associated with foliar senescence characters, was also determined in the *Ae. biuncialis* accessions (Bachmann et al., 1994; Thomas et al., 1996). Delayed senescence allows the leaves to maintain photosynthetic activity during the active grain-filling period, thus ensuring the incorporation of assimilates to the grain (Thomas and Howarth, 2000; Spano et al., 2003; Kobata et al., 2015). The stay-green phenotype is associated with the ability of plants to maintain grain development longer, resulting in better yields, especially under water-limited conditions (Borrell et al., 2000; Christopher et al., 2008). In the present experiments, seven of the 86 *Ae. biuncialis* genotypes had SPAD values below the detection limit, due to premature leaf senescence. Eight accessions that exhibited the stay-green phenotype with early flowering were identified as potential gene sources for wheat breeding programs aimed to improve drought tolerance. Terminal drought during the grain-filling period can be avoided through accelerated reproduction, while the crop yield may be increased due to the preservation of the green leaf area.

CONCLUSIONS

This paper confirmed that the DArTseq genotyping approach can be used efficiently to investigate the genetic diversity and population structure of *Ae. biuncialis*. The population structure defined using a combination of three clustering methods was in agreement with the geographic origin of the accessions. The heading trait diversity was also found to correlate largely with the genetic structure and geographic distribution. The wide-ranging genetic diversity of *Ae. biuncialis* may make it a potential genetic source for wheat improvement. The large intraspecific variation for heading time offers opportunities for the maximization of grain yield by shortening the plant life cycle, resulting in the termination of the grain-filling period before the drought season. The comparative analysis of genetic and phenological diversity patterns enables the selection of *Ae. biuncialis* genotypes adapted to a wide variety of ecological habitats, which could be used to breed new wheat cultivars able to cope with extreme climatic changes.

DATA AVAILABILITY STATEMENT

The SilicoDArT marker data of the *Ae. biuncialis* collection can be found in EBI https://www.ebi.ac.uk/biostudies/studies/S-BSST277.

AUTHOR CONTRIBUTIONS

LI participated in the analysis of the phenotypic data and wrote the manuscript. IMon performed the cluster analysis and made valuable comments on the manuscript. MM and PM maintained the germplasm and phenotyped it for the heading time. AF, ET, EG, AL-T, KS-P, and ÉS contributed to the maintenance of the plant material and the field experiments. ÉD carried out the phenotypic analysis of SPAD. TK evaluated the heading time of the genotypes. AK helped in GBS data analysis. IMol planned and managed the whole study, and supervised the manuscript preparation.

FUNDING

This paper was financed by a Marie Curie Fellowship Grant ('AEGILWHEAT'-H2020-MSCA-IF-2016-746253) under the H2020 framework program of the European Union and by the Hungarian National Research, Development and Innovation Office—NKFIH 116277, 112226, and 119387 and by the ERDF project „Plants as a tool for sustainable global development" (no. CZ.02.1.01/0.0/0.0/16_019/0000827).

ACKNOWLEDGMENTS

Special thanks to Tamás Árendás for providing the weather data measured between 2015 and 2018. The technical assistance of Fanni Tóth, Ildikó Könyves-Lakner and Ágnes Bencze is gratefully acknowledged. Thanks are due to Barbara Hooper for revising the manuscript linguistically.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2019.01531/full#supplementary-material

REFERENCES

Araus, J. L., Slafer, G. A., Reynolds, M. P., and Royo, C., (2004). "Physiology of yield and adaptation in wheat and barley breeding," in *Physiology and Biotechnology Integration for Plant Breeding*. Eds.Nguyen, H. T., and Blum, A. (New York, USA: Marcel Dekker), 1–49. doi: 10.1201/9780203022030.ch16

Arora, S., Singh, N., Kaur, S., Bains, N. S., Uauy, C., Poland, J., et al. (2017). Genome-wide association study of grain architecture in wild wheat Aegilops tauschii. *Front. Plant Sci.* 8, 886–898. doi: 10.3389/fpls.2017.00886

Arrigo, N., Felber, F., Parisod, C., Buerki, S., Alvarez, N., David, J., et al. (2010). Origin and expansion of the allotetraploid Aegilops geniculata, a wild relative of wheat. *New Phytol.* 187, 1170–1180. doi: 10.1111/j.1469-8137.2010.03328.x

Bachmann, A., Fernández-López, J., Ginsburg, S., Thomas, H., Bouwkamp, J. C., Solomos, T., et al. (1994). Stay-green genotypes of Phaseolus vulgaris L.: chloroplast proteins and chlorophyll catabolites during foliar senescence. *New Phytol.* 126, 593–600. doi: 10.1111/j.1469-8137.1994.tb02953.x

Badea, A., Eudes, F., Salmon, D., Tuvesson, S., Vrolijk, A., Larsson, C.-T., et al. (2011). Development and assessment of DArT markers in triticale. *Theor. Appl. Genet.* 122, 1547–1560. doi: 10.1007/s00122-011-1554-3

Baloch, F. S., Alsaleh, A., Shahid, M. Q., Çiftçi, V., Sáenz de Miera, L. E., Aasim, M., et al. (2017). A whole genome DArTseq and SNP analysis for genetic diversity assessment in durum wheat from Central Fertile Crescent. *PloS One* 12, e0167821. doi: 10.1371/journal.pone.0167821

Bansal, M., Kaur, S., Dhaliwal, H. S., Bains, N. S., Bariana, H. S., Chhuneja, P., et al. (2017). Mapping of Aegilops umbellulata-derived leaf rust and stripe rust resistance loci in wheat. *Plant Pathol.* 66, 38–44. doi: 10.1111/ppa.12549

Blondel, J. (2006). The "design" of Mediterranean landscapes: a millennial story of humans and ecological systems during the historic period. *Hum. Ecol.* 34, 713–729. doi: 10.1007/s10745-006-9030-4

Bolibok-Brągoszewska, H., Targońska, M., Bolibok, L., Kilian, A., and Rakoczy-Trojanowska, M. (2014). Genome-wide characterization of genetic diversity and population structure in Secale. *BMC Plant Biol.* 14, 184–198. doi: 10.1186/1471-2229-14-184

Borrell, A. K., Hammer, G. L., and Henzell, R. G. (2000). Does maintaining green leaf area in Sorghum improve yield under drought? II. Dry matter production and yield. *Crop Sci.* 40, 1037–1048. doi: 10.2135/cropsci2000.4041037x

Cai, S., Wu, D., Jabeen, Z., Huang, Y., Huang, Y., and Zhang, G. (2013). Genome-wide association analysis of aluminum tolerance in cultivated and Tibetan wild barley. *PloS One* 8, e69776. doi: 10.1371/journal.pone.0069776

Christopher, J. T., Manschadi, A. M., Hammer, G. L., and Borrell, A. K. (2008). Developmental and physiological traits associated with high yield and stay-green phenotype in wheat. *Aust. J. Agric. Res.* 59, 354–364. doi: 10.1071/ar07193

Colmer, T. D., Flowers, T. J., and Munns, R. (2006). Use of wild relatives to improve salt tolerance in wheat. *J. Exp. Bot.* 57, 1059–1078. doi: 10.1093/jxb/erj124

Comadran, J., Russell, J. R., Booth, A., Pswarayi, A., Ceccarelli, S., Grando, S., et al. (2011). Mixed model association scans of multi-environmental trial data reveal major loci controlling yield and yield related traits in Hordeum vulgare in Mediterranean environments. *Theor. Appl. Genet.* 122, 1363–1373. doi: 10.1007/s00122-011-1537-4

Cseh, A., Kruppa, K., Molnár, I., Rakszegi, M., Doležel, J., and Molnár-Láng, M. (2011). Characterization of a new 4BS.7HL wheat–barley translocation line using GISH, FISH, and SSR markers and its effect on the β-glucan content of wheat. *Genome* 54, 795–804. doi: 10.1139/g11-044

Devictor, V., Van Swaay, C., Brereton, T., Brotons, L., Chamberlain, D., Heliölö, J., et al. (2012). Differences in the climatic debts of birds and butterflies at a continental scale. *Nat. Clim. Change* 2, 121–124. doi: 10.1038/nclimate1347

Dulai, S., Molnár, I., Szopkó, D., Darkó, É., Vojtkó, A., and Sass-Gyarmati, A. (2014).Wheat- Aegilops biuncialis amphiploids have efficient photosynthesis and biomass production during osmotic stress. *J. Plant Physiol.* 171, 509–517. doi: 10.1016/j.jplph.2013.11.015

Dvorak, J., and Akhunov, E. D. (2005). Tempos of gene locus deletions and duplications and their relationship to recombination rate during diploid and polyploid evolution in the Aegilops-Triticum alliance. *Genetics* 171, 323–332. doi: 10.1534/genetics.105.041632

Edet, O. U., Gorafi, Y. S. A., Nasuda, S., and Tsujimoto, H. (2018). DArTseq-based analysis of genomic relationships among species of tribe Triticeae. *Sci. Rep.* 8, 16397–16408. doi: 10.1038/s41598-018-34811-y

Etminan, A., Pour-Aboughadareh, A., Mehrabi, A. A., Shooshtari, L., Ahmadi-Rad, A., and Moradkhani, H. (2019). Molecular characterization of the wild relatives of wheat using CAAT-box derived polymorphism. *Plant Biosyst.* 153, 398–405. doi: 10.1080/11263504.2018.1492993

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

Friebe, B., Jiang, J., Raupp, W. J., McIntosh, R. A., and Gill, B. S. (1996). Characterization of wheat-alien translocations conferring resistance to diseases and pests: current status. *Euphytica* 91, 59–87. doi: 10.1007/BF00035277

Gawroński, P., Pawełkowicz, M., Tofil, K., Uszyński, G., Sharifova, S., Ahluwalia, S., et al. (2016). DArT markers effectively target gene space in the rye genome. *Front. Plant Sci.* 07, 1600–1613. doi: 10.3389/fpls.2016.01600

Hammer, Ø., Harper, D. A. T., and Ryan, P. D. (2001). PAST: Paleontological statistics software package for education and data analysis. *Palaeontol. Electron.* 4, 4–9. doi: 10.1016/j.bcp.2008.05.025

Hickling, R., Roy, D. B., Hill, J. K., Fox, R., and Thomas, C. D. (2006). The distributions of a wide range of taxonomic groups are expanding polewards. *Glob. Change Biol.* 12, 450–455. doi: 10.1111/j.1365-2486.2006.01116.x

Jaccoud, D., Peng, K., Feinstein, D., and Kilian, A. (2001). Diversity Arrays: a solid state technology for sequence information independent genotyping. *Nucleic Acids Res.* 29, 1–7. doi: 10.1093/nar/29.4.e25

Janzen, D. H. (1984). Dispersal of small seeds by big herbivores: foliage is the fruit. *Am. Nat.* 123, 338–353. doi: 10.1016/j.compstruc.2007.04.021

Jarvis, A., Lane, A., and Hijmans, R. J. (2008).The effect of climate change on crop wild relatives. *Agric. Ecosyst. Environ.* 126, 13–23. doi: 10.1016/j.agee.2008.01.013

Jauhar, P. P., and Chibbar, R. N. (1999). Chromosome-mediated and direct gene transfers in wheat. *Genome* 42, 570–583. doi: 10.1139/g99-045

Jauhar, P. P. (1993). "Alien gene transfer and genetic enrichment of bread wheat," in *Biodiversity and Wheat Improvement*. Ed.Damania, A. B. (Chichester, U.K:John Wiley & Sons), 103–119. Available at: https://www.cabdirect.org/cabdirect/abstract/19941605141 [Accessed July 5, 2019].

Jing, H.-C., Bayon, C., Kanyuka, K., Berry, S., Wenzl, P., and Huttner, E. (2009). DArT markers: diversity analyses, genomes comparison, mapping and integration with SSR markers in Triticum monococcum. *BMC Genomics* 10, 458–474. doi: 10.1186/1471-2164-10-458

Jones, H., Gosman, N., Horsnell, R., Rose, G. A., Everest, L. A., Bentley, A. R., et al. (2013). Strategy for exploiting exotic germplasm using genetic, morphological, and environmental diversity: the Aegilops tauschii Coss. example. *Theor. Appl. Genet.* 126, 1793–1808. doi: 10.1007/s00122-013-2093-x

Kellogg, E. A., Appels, R., and Mason-Gamer, R. J. (1996). When genes tell different stories: the diploid genera of Triticeae (Gramineae). *Syst. Bot.* 21, 321–347. doi: 10.2307/2419662

Kilian, B., Özkan, H., Deusch, O., Effgen, S., Brandolini, A., Kohl, J., et al. (2007). Independent wheat B and G genome origins in outcrossing Aegilops progenitor haplotypes. *Mol. Biol. Evol.* 24, 217–227. doi: 10.1093/molbev/msl151

Kilian, B., Mammen, K., Millet, E., Sharma, R., Graner, A., Salamini, F., et al. (2011). "Aegilops," in *Wild Crop Relatives: Genomic and Breeding Resources: Cereals*. Ed. Kole, C. (Berlin: Springer-Verlag), 1–76. doi: 10.1007/978-3-642-14871-2

Kilian, A., Wenzl, P., Huttner, E., Carling, J., Xia, L., and Blois, H. (2012). "Diversity Arrays Technology: A generic genome profiling technology on open platforms," in *Data Production and Analysis in Population Genomics*. Eds. Pompanon, F., and Bonin, A. (Totowa, NJ: Humana Press), 67–89. doi: 10.1007/978-1-61779-870-2_5

Kishii, M. (2019). An update of recent use of Aegilops species in wheat breeding. *Front. Plant Sci.* 10, 585–603. doi: 10.3389/fpls.2019.00585

Kobata, T., Shinonaga, M., Yoshida, H., Tomisaka, K., and Akai, K. (2015). Stay-green trait assessment using the leaf incubation method to examine the maintenance of assimilation rates under high temperature conditions during the grain-filling period in rice. *Plant Prod. Sci.* 18, 254–266. doi: 10.1626/pps.18.254

Kumar, A., Seetan, R., Mergoum, M., Tiwari, V. K., Iqbal, M. J., Wang, Y., et al. (2015). Radiation hybrid maps of the D-genome of Aegilops tauschii and their application in sequence assembly of large and complex plant genomes. *BMC Genomics* 16, 800–813. doi: 10.1186/s12864-015-2030-2

Liu, W., Jin, Y., Rouse, M., Friebe, B., Gill, B., and Pumphrey, M. O. (2011a). Development and characterization of wheat-Ae. searsii Robertsonian

translocations and a recombinant chromosome conferring resistance to stem rust. *Theor. Appl. Genet.* 122, 1537–1545. doi: 10.1007/s00122-011-1553-4

Liu, W., Rouse, M., Friebe, B., Jin, Y., Gill, B., and Pumphrey, M. (2011b). Discovery and molecular mapping of a new gene conferring resistance to stem rust, Sr53, derived from Aegilops geniculata and characterization of spontaneous translocation stocks with reduced alien chromatin. *Chromosom. Res.* 19, 669–682. doi: 10.1007/s10577-011-9226-3

Marais, G. F., McCallum, B., and Marais, A. S. (2006). Leaf rust and stripe rust resistance genes derived from Aegilops sharonensis. *Euphytica* 149, 373–380. doi: 10.1007/s10681-006-9092-9

Marais, F., Marais, A., McCallum, B., and Pretorius, Z. (2009). Transfer of leaf rust and stripe rust resistance genes Lr62 and Yr42 from Aegilops neglecta Req. ex Bertol. to common wheat. *Crop Sci.* 49, 871–879. doi: 10.2135/cropsci2008.06.0317

Matsuoka, Y., Takumi, S., and Kawahara, T. (2008). Flowering time diversification and dispersal in central Eurasian wild wheat Aegilops tauschii Coss.: genealogical and ecological framework. *PloS One* 3, e3138. doi: 10.1371/journal.pone.0003138

Matsuoka, Y., Takumi, S., and Kawahara, T. (2015). Intraspecific lineage divergence and its association with reproductive trait change during species range expansion in central Eurasian wild wheat Aegilops tauschii Coss. *(Poaceae). BMC Evol. Biol.* 15, 213–222. doi: 10.1186/s12862-015-0496-9

Meimberg, H., Hammond, J. I., Jorgensen, C. M., Park, T. W., Gerlach, J. D., Rice, K. J., et al. (2006). Molecular evidence for an extreme genetic bottleneck during introduction of an invading grass to California. *Biol. Invasions* 8, 1355–1366. doi: 10.1007/s10530-005-2463-7

Meimberg, H., Milan, N. F., Karatassiou, M., Espeland, E. K., McKay, J. K., and Rice, K. J. (2010). Patterns of introduction and adaptation during the invasion of Aegilops triuncialis (Poaceae) into Californian serpentine soils. *Mol. Ecol.* 19, 5308–5319. doi: 10.1111/j.1365-294X.2010.04875.x

Mendlinger, S., and Zohary, D. (1995). The extent and structure of genetic variation in species of the sitopsis group of aegilops. *Heredity (Edinb).* 74, 616–627. doi: 10.1038/hdy.1995.85

Mikó, P., Löschenberger, F., Hiltbrunner, J., Aebi, R., Megyeri, M., Kovács, G., et al. (2014). Comparison of bread wheat varieties with different breeding origin under organic and low input management. *Euphytica* 199, 69–80. doi: 10.1007/s10681-014-1171-8

Millet, E. (2007). Exploitation of Aegilops species of section Sitopsis for wheat improvement. *Isr. J. Plant Sci.* 55, 277–287. doi: 10.1560/ijps.55.3-4.277

Mizuno, N., Yamasaki, M., Matsuoka, Y., Kawahara, T., and Takumi, S. (2010). Population structure of wild wheat D-genome progenitor Aegilops tauschii Coss.: Implications for intraspecific lineage diversification and evolution of common wheat. *Mol. Ecol.* 19, 999–1013. doi: 10.1111/j.1365-294X.2010.04537.x

Molnár, I., Gáspár, L., Sárvári, É., Dulai, S., Hoffmann, B., and Molnár-Láng, M. (2004). Physiological and morphological responses to water stress in Aegilops biuncialis and Triticum aestivum genotypes. *Biol. Funct. Plant* 31, 1149–1159. doi: 10.1071/FP03143

Molnár, I., Cifuentes, M., Schneider, A., Benavente, E., and Molnár-Láng, M. (2011). Association between simple sequence repeat-rich chromosome regions and intergenomic translocation breakpoints in natural populations of allopolyploid wild wheats. *Ann. Bot.* 107, 65–76. doi: 10.1093/aob/mcq215

Molnár-Láng, M., Molnár, I., Szakács, É., Linc, G., and Bedö, Z. (2014). "Production and molecular cytogenetic identification of wheat-alien hybrids and introgression lines," in *Genomics of Plant Genetic Resources*. Eds.Tuberosa, R., Graner, A., and Frison, E. (Dordrecht:Springer Netherlands), 255–283. doi: 10.1007/978-94-007-7572-5_11

Monneveux, P., Zaharieva, M., and Rekika, D. (2000). "The utilisation of Triticum and Aegilops species for the improvement of durum wheat," in *Durum Wheat Improvement in the Mediterranean Region: New Challenges*. Eds.Royo, C., Nachit, M., Di Fonzo, N., and Araus, J. L. (CIHEAM Zaragoza), 71–81. Available at: http://om.ciheam.org/article.php?IDPDF=600008http://www.ciheam.org/%5Cnhttp://om.ciheam.org/.

Monostori, I., Szira, F., Tondelli, A., Árendás, T., Gierczik, K., Cattivelli, L., et al. (2017). Genome-wide association study and genetic diversity analysis on nitrogen use efficiency in a Central European winter wheat (Triticum aestivum L.) collection. *PloS One* 12, e0189265. doi: 10.1371/journal.pone.0189265

Monte, J. V., Nova, P. J. G., and Soler, C. (2001). AFLP-based analysis to study genetic variability and relationships in the Spanish species of the genus Aegilops. *Hereditas* 135, 233–238. doi: 10.1111/j.1601-5223.2001.00233.x

Nagy, E. D., Molnár, I., Schneider, A., Kovács, G., and Molnár-Láng, M. (2006). Characterization of chromosome-specific S-SAP markers and their use in studying genetic diversity in Aegilops species. *Genome* 49, 289–296. doi: 10.1139/g05-109

Nielsen, N. H., Backes, G., Stougaard, J., Andersen, S. U., and Jahoor, A. (2014). Genetic diversity and population structure analysis of European hexaploid bread wheat (Triticum aestivum L.) varieties. *PloS One* 9, e94000. doi: 10.1371/journal.pone.0094000

Okuno, K., Ebana, K., Noov, B., and Yoshida, H. (1998). Genetic diversity of Central Asian and north Caucasian Aegilops species as revealed by RAPD markers. *Genet. Resour. Crop Evol.* 45, 389–394. doi: 10.1023/A:1008660001263

Olivera, P. D., and Steffenson, B. J. (2009). Aegilops sharonensis: origin, genetics, diversity, and potential for wheat improvement. *Botany* 87, 740–756. doi: 10.1139/b09-040

Olivera, P. D., Anikster, Y., and Steffenson, B. J. (2010). Genetic diversity and population structure in Aegilops sharonensis. *Crop Sci.* 50, 636–648. doi: 10.2135/cropsci2009.04.0192

Olivera, P. D., Rouse, M. N., and Jin, Y. (2018). Identification of new sources of resistance to wheat stem rust in Aegilops spp. in the tertiary genepool of wheat. *Front. Plant Sci.* 9, 1719–1726. doi: 10.3389/fpls.2018.01719

Olson, E. L., Rouse, M. N., Pumphrey, M. O., Bowden, R. L., Gill, B. S., and Poland, J. A. (2013). Introgression of stem rust resistance genes SrTA10187 and SrTA10171 from Aegilops tauschii to wheat. *Theor. Appl. Genet.* 126, 2477–2484. doi: 10.1007/s00122-013-2148-z

Ostrowski, M. F., Prosperi, J. M., and David, J. (2016).Potential implications of climate change on Aegilops species distribution: sympatry of these crop wild relatives with the major European crop Triticum aestivum and conservation issues. *PloS One* 11, 1–27. doi: 10.1371/journal.pone.0153974

Parmesan, C., and Yohe, G. (2003). A globally coherent fingerprint of climate change impacts across natural systems. *Nature* 421, 37–42. doi: 10.1038/nature01286

Peleg, Z., Fahima, T., Abbo, S., Krugman, T., Nevo, E., Yakir, D., et al. (2005). Genetic diversity for drought resistance in wild emmer wheat and its ecogeographical associations. *Plant Cell Environ.* 28, 176–191. doi: 10.1111/j.1365-3040.2005.01259.x

Pestsova, E., Korzun, V., Goncharov, N. P., Hammer, K., Ganal, M. W., and Röder, M. S. (2000). Microsatellite analysis of Aegilops tauschii germplasm. *Theor. Appl. Genet.* 101, 100–106. doi: 10.1007/s001220051456

Peters, A., Johnson, D. E., and George, M. R. (1996). Barb goatgrass: a threat to California rangelands. *Rangelands* 18, 8–10.

Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959. doi: 10.1111/j.1471-8286.2007.01758.x

Qi, L., Friebe, B., and Gill, B. S. (2006). Complex genome rearrangements reveal evolutionary dynamics of pericentromeric regions in the Triticeae. *Genome* 49, 1628–1639. doi: 10.1139/g06-123

Römermann, C., Tackenberg, O., and Poschlod, P. (2005). How to predict attachment potential of seeds to sheep and cattle coat from simple morphological seed traits. *Oikos* 110, 219–230. doi: 10.1111/j.0030-1299.2005.13911.x

Rekika, D., Monneveux, P., and Havaux, M. (1997). The in vivo tolerance of photosynthetic membranes to high and low temperatures in cultivated and wild wheats of the Triticum and Aegilops genera. J. *Plant Physiol.* 150, 734–738. doi: 10.1016/S0176-1617(97)80291-X

Sansaloni, C., Petroli, C., Jaccoud, D., Carling, J., Detering, F., Grattapaglia, D., et al. (2011). Diversity Arrays Technology (DArT) and next-generation sequencing combined: genome-wide, high throughput, highly informative genotyping for molecular breeding of Eucalyptus. *BMC Proc.* 5, 54–55. doi: 10.1186/1753-6561-5-S7-P54

Schneider, A., Molnár, I., and Molnár-Láng, M. (2008). Utilisation of Aegilops (goatgrass) species to widen the genetic diversity of cultivated wheat. *Euphytica* 163, 1–19. doi: 10.1007/s10681-007-9624-y

Singh, N., Wu, S., Tiwari, V., Sehgal, S., Raupp, J., Wilson, D., et al. (2019). Genomic analysis confirms population structure and identifies inter-lineage hybrids in Aegilops tauschii. *Front. Plant Sci.* 10, 9–21. doi: 10.3389/fpls.2019.00009

Sohail, Q., Shehzad, T., Kilian, A., Eltayeb, A. E., Tanaka, H., and Tsujimoto, H. (2012). Development of diversity array technology (DArT) markers for

assessment of population structure and diversity in Aegilops tauschii. *Breed. Sci.* 62, 38–45. doi: 10.1270/jsbbs.62.38

Spano, G., Di Fonzo, N., Perrotta, C., Platani, C., Ronga, G., Lawlor, D. W., et al. (2003). Physiological characterization of "stay green" mutants in durum wheat. *J. Exp. Bot.* 54, 1415–1420. doi: 10.1093/jxb/erg150

Tahernezhad, Z., Zamani, M. J., Solouki, M., Zahravi, M., Imamjomeh, A. A., Jafaraghaei, M., et al. (2010). Genetic diversity of Iranian Aegilops tauschii Coss. using microsatellite molecular markers and morphological traits. *Mol. Biol. Rep.* 37, 3413–3420. doi: 10.1007/s11033-009-9931-6

Thomas, K. G., and Bebeli, P. J. (2010). Genetic diversity of Greek Aegilops species using different types of nuclear genome markers. *Mol. Phylogenet. Evol.* 56, 951–961. doi: 10.1016/J.YMPEV.2010.04.041

Thomas, H., and Howarth, C. J. (2000). Five ways to stay green. *J. Exp. Bot.* 51, 329–337. doi: 10.1093/jexbot/51.suppl_1.329

Thomas, H., Schellenberg, M., Vicentini, F., and Matile, P. (1996). Gregor Mendel's green and yellow pea seeds. *Bot. Acta* 109, 3–4. doi: 10.1111/j.1438-8677.1996.tb00862.x

Thomas, C. D., Cameron, A., Green, R. E., Bakkenes, M., Beaumont, L. J., Collingham, Y. C., et al. (2004). Extinction risk from climate change. *Nature* 427, 145–148. doi: 10.1038/nature02121

Tinker, N. A., Kilian, A., Wight, C. P., Heller-Uszynska, K., Wenzl, P., and Rines, H. W., et al. (2009). New DArT markers for oat provide enhanced map coverage and global germplasm characterization. *BMC Genomics* 10, 39–60. doi: 10.1186/1471-2164-10-39

Tottman, D. R. (1987). The decimal code for the growth stages of cereals, with illustrations. *Ann. Appl. Biol.* 110, 441–454. doi: 10.1111/j.1744-7348.1987.tb03275.x

van Slageren, M. W. (1994). *Wild Wheats: a Monograph of Aegilops L. and Amblyopyrum (Jaub and Spach) Eig (Poaceae)*. Wageningen: Wageningen Agricultural University Papers.

Vellend, M., Myers, J. A., Gardescu, S., Marks, P. L., et al. (2003). Dispersal of Trillium seeds by deer: implications for long-distance migration of forest herbs. *Ecology* 84, 1067–1072. doi: 10.1890/0012-9658(2003)084[1067:DOTSBD]2.0.CO;2

Wang, J., Luo, M. C., Chen, Z., You, F. M., Wei, Y., and Zheng, Y. (2013). Aegilops tauschii single nucleotide polymorphisms shed light on the origins of wheat D-genome genetic diversity and pinpoint the geographic origin of hexaploid wheat. *New Phytol.* 198, 925–937. doi: 10.1111/nph.12164

Wenzl, P., Carling, J., Kudrna, D., Jaccoud, D., Huttner, E., and Kleinhofs, A. , et al. (2004). Diversity Arrays Technology (DArT) for whole-genome profiling of barley. *Proc. Natl. Acad. Sci.* 101, 9915–9920. doi: 10.1073/pnas.0401076101

Zaharieva, M., and Monneveux, P. (2006). Spontaneous hybridization between bread wheat (Triticum aestivum L.) and its wild relatives in Europe. *Crop Sci.* 46, 512–527. doi: 10.2135/cropsci2005.0023

Zaharieva, M., Gaulin, E., Havaux, M., Acevedo, E., and Monneveux, P. (2001a). Drought and heat responses in the wild wheat relative roth. *Crop Sci.* 41, 1321–1329. doi: 10.2135/cropsci2001.4141321x

Zaharieva, M., Monneveux, P., Henry, M., Rivoal, R., Valkoun, J., and Nachit, M. M. (2001b). Evaluation of a collection of wild wheat relative Aegilops geniculata Roth and identification of potential sources for useful traits. *Euphytica* 119, 33–38. doi: 10.1023/A:1017500728227

Zaharieva, M., Prosperi, J. M., and Monneveux, P. (2004). Ecological distribution and species diversity of Aegilops L. genus in Bulgaria. *Biodivers. Conserv.* 13, 2319–2337. doi: 10.1023/B:BIOC.0000047917.11509.a2

Zhang, P., Dundas, I. S., McIntosh, R. A., Xu, S. S., Park, R. F., Gill, B. S., et al. (2015). "Wheat–Aegilops Introgressions," in *Alien Introgression in Wheat: Cytogenetics, Molecular Biology, and Genomics*. Eds. Molnár-Láng, M., Ceoloni, C., and Doležel, J. (Springer International Publishing Cham), 221–243. doi: 10.1007/978-3-319-23494-6

Conflict of Interest: 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 © 2019 Ivanizs, Monostori, Farkas, Megyeri, Mikó, Türkösi, Gaál, Lenykó-Thegze, Sző ke-Pázsi, Szakács, Darkó, Kiss, Kilian and Molnár. 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) and the copyright owner(s) 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 Plant Science

ORIGINAL RESEARCH
published: 06 December 2019
doi: 10.3389/fpls.2019.01607

Domesticating *Vigna Stipulacea*: A Potential Legume Crop With Broad Resistance to Biotic Stresses

*Yu Takahashi[1], Hiroaki Sakai[2], Yuki Yoshitsu[3], Chiaki Muto[1], Toyoaki Anai[4], Muthaiyan Pandiyan[5], Natesan Senthil[6], Norihiko Tomooka[1] and Ken Naito[1]**

[1] Genetic Resources Center, NARO, Tsukuba, Japan, [2] Advanced Analysis Center, NARO, Tsukuba, Japan, [3] Kenpoku Agricultural Institute, Iwate Agricultural Research Center, Iwate, Japan, [4] Department of Agriculture, Saga University, Saga, Japan, [5] Agricultural College and Research Institute, Tamil Nadu Agricultural University, Thanjavur, India, [6] Agricultural College and Research Institute, Tamil Nadu Agricultural University, Madurai, India

Though crossing wild relatives to modern cultivars is a usual means to introduce alleles of stress tolerance, an alternative is *de novo* domesticating wild species that are already tolerant to various kinds of stresses. As a test case, we chose *Vigna stipulacea* Kuntze, which has fast growth, short vegetative stage, and broad resistance to pests and diseases. We developed an ethyl methanesulfonate–mutagenized population and obtained three mutants with reduced seed dormancy and one with reduced pod shattering. We crossed one of the mutants of less seed dormancy to the wild type and confirmed that the phenotype was inherited in a Mendelian manner. *De novo* assembly of *V. stipulacea* genome, and the following resequencing of the F2 progenies successfully identified a Single Nucleotide Polymorphism (SNP) associated with seed dormancy. By crossing and pyramiding the mutant phenotypes, we will be able to turn *V. stipulacea* into a crop which is yet primitive but can be cultivated without pesticides.

Keywords: plant domestication, wild species, legume, *Vigna*, mutant screening, seed dormancy, pod shattering, bulked segregant analysis

OPEN ACCESS

Edited by:
Eric Von Wettberg, University of Vermont, United States

Reviewed by:
R. Varma Penmetsa, University of California, Davis, United States
Robert Henry, University of Queensland, Australia

***Correspondence:**
Ken Naito
knaito@affrc.go.jp

Specialty section:
This article was submitted to Plant Breeding, a section of the journal Frontiers in Plant Science

Received: *30 August 2019*
Accepted: *15 November 2019*
Published: *06 December 2019*

Citation:
Takahashi Y, Sakai H, Yoshitsu Y, Muto C, Anai T, Pandiyan M, Senthil N, Tomooka N and Naito K (2019) Domesticating Vigna Stipulacea: A Potential Legume Crop With Broad Resistance to Biotic Stresses. Front. Plant Sci. 10:1607. doi: 10.3389/fpls.2019.01607

INTRODUCTION

To feed the growing population in the world, we have to produce more food with less input. This is a challenging issue because of global climate change, limited water resource, and acquired resistance of pests and diseases against chemicals.

To achieve this, many scientists are now focusing on harnessing genetic diversity of genebank accessions including wild crop relatives and neglected crops (McCouch et al., 2013). One of the limitations to the above described issue is genetic vulnerability of modern cultivars. They have gone through strong bottleneck and often sacrificed resilience to biotic and abiotic stresses (Pingali, 2012). On the other hand, many wild species are well-adapted to ecological niche, which is often harsh to domesticated species (Zhang et al., 2018). Thus, utilizing the adaptability of such wild or semi-wild species will be a key to sustainable agriculture.

To fully exploit the genetic diversity of wild species, "*de novo* domestication" or "redomestication" are now proposed (Fernie and Yan, 2019). Until recently, the main idea of using wild genetic resources were to cross with cultivars to introduce resistant alleles. However, cross compatibility is often limited even between a cultivar and its close relatives. In addition, adaptation to a certain environment is often a complex trait with multiple genes involved. On the other hand,

domestication-related traits often arose with loss-of-function mutations in single loci (Doebley et al., 2006). Thus, it might be easier to introduce domestication-related mutations into wild species than to introduce adaptation-related alleles into crops.

Although the technologies of sequencing and editing genomes are expected to facilitate *de novo* domestication (Shapter et al., 2013; Li et al., 2018), we still believe that simple mutagenesis + forward phenotype screening could be the easiest way to do it. To apply CRISPR/Cas9 system, one has to sequence the whole genome and develop transformation system of the plant to be domesticated. In addition, genes involved in domestication-related traits are not well-catalogued except Solanaceae, Brassicaceae, and Poaceae (Østerberg et al., 2017). As such, mutagenesis followed by forward screening is currently the only practical approach to *de novo* domesticate most of the potentially useful wild plants, such as wild legumes.

Thus, in this study, we tried to domesticate *Vigna stipulacea* Kuntze by ethyl methanesulfonate (EMS) mutagenesis followed by phenotype screening. *V. stipulacea* inhabits mainly in South Asia, and has fast growth, short vegetative stage, and resistance to pests and diseases (Tomooka et al., 2014). The seeds are edible, and some local people cultivate it mainly as pasture but sometimes as food. However, less and less farmers use it because of the high labor caused by strong behavior of pod shattering (Tomooka et al., 2011). In addition, it retains seed dormancy, which needs to be reduced for uniform germination. To improve these traits, we screened and obtained mutants with reduced pod-shattering and with reduced seed dormancy. We also identified a SNP associated with seed dormancy in one of the obtained mutants by whole genome analyses.

MATERIALS AND METHODS

Plant Materials and Growth Condition

Table 1 summarizes the materials tested in this study. Besides *V. stipulacea*, we used two or three accessions of three domesticated species, soybean [*Glycine max* (L.) Merr.], common bean (*Phaseolus vulgaris* L.), and cowpea [*Vigna unguiculata* (L.) Walp.], to evaluate pod shattering and seed dormancy. All the materials were provided by the NARO gene bank (https://www.gene.affrc.go.jp/index_en.php). We cultivated *V. stipulacea* in a field or in a bucket with gardening soil in a greenhouse of our institute, Tsukuba, Japan (36.030577, 140.098021). We cultivated three plants of the domesticated species in a bucket with gardening soil in a greenhouse of our institute, Tsukuba, Japan.

TABLE 1 | Plant materials.

Accession no.	Scientific names	Name	Origin
JP245503	*Vigna stipulacea* Kuntze	—	India
JP31043	*Glycine max* (L.) Merr.	Williams 82	America
JP28862	*G. max* (L.) Merr.	Enrei	Japan
JP239215	*Vigna unguiculata* (L.) Walp. cv-gr. Biflora E. Westphal	—	Pakistan
JP244182	*V. unguiculata* (L.) Walp. cv-gr. Unguiculata E. Westphal	—	Pakistan
JP41232	*Phaseolus vulgaris* L.	Bayo	Peru
JP41234	*P. vulgaris* L.	Red Kidney	Peru
JP41963	*P. vulgaris* L.	Frijol Azufrado	Mexico

Development of a Mutant Population

The process of mutant screening is shown in **Figure 1**. To increase mutation density, we adopted two cycles of successive chemical mutagenesis as described by Tsuda et al. (2015). We started with 12,000 seeds of *V. stipulacea* (JP245503). We first scratched the seed coat with a knife because it is hard and waterproof, treated the scratched seeds with 0.35% EMS solution for 8 h, and thoroughly washed them with distilled water (M1 seeds). The M1 seeds were incubated in a wet condition for 3 days and those with germination were sown. The M1 plants were cultivated in our greenhouse, where the temperature was kept above >20°C. Three pods per plant were harvested from 3,000 fertile M1 plants (M2 seeds). Of the harvested seeds, four per line (12,000 in total) were scratched and treated with 0.35% EMS solution as described above (M2M1 seeds). Germinating M2M1 seeds were sown and cultivated in a field, located in Tsukuba, Japan. Three thousand fertile plants were selected and three pods per plant were harvested (M2M2 seeds). Six seeds per M2M2 line were sown in a 1 L plastic pot filled with gardening soil and cultivated in the greenhouse until three pods per line were harvested (M2M3 seeds). Of the M2M3 lines, four lines with mutant phenotypes were selected and seven seeds per line were sown in a 7 L bucket filled with the gardening soil and cultivated in the greenhouse.

Evaluation of Seed Dormancy and Pod Shattering

To evaluate pod shattering, we harvested 20 pods of each domesticated accession and three pods per M2M2 plant and counted the number of shattered pods. Before evaluation, harvested pods were left at room temperature for a month and then completely dried at 40°C for 24 h in an incubator.

To evaluate seed dormancy, 20 seeds were soaked in distilled water and incubated in a dark incubator at 25°C, and imbibed seeds were counted daily until the 3rd day, twice a week until the 14th day, and weekly until the 28th day. We replicated the measurements by seven times and performed Dunnett test to compare averages of each time point between the wild type and each mutant line.

Morphological Observation

We observed the imbibing seeds to detect the imbibition start point. We also observed pod sclerenchyma (tissue with dead cells with thickened secondary wall) with a stereoscopic microscope (ECLIPSE Ci-L, Nikon, Tokyo, Japan). Three pods were collected from each individual and were sliced by a microtome (MTH-1, Nippon Medial & Chemical Instruments Co., Ltd. Osaka, Japan), and stained with Phloroglucinol-HCL solution (1 g phloroglucin in 50 mL ethanol + 25 mL concentrated hydrochloric acid).

Whole Genome Sequencing, Assembly, and Annotation

We sequenced the whole genome of *V. stipulacea* (JP245503) with RSII sequencer (Pacific Biosciences, Menlo Park, CA), as we have done previously for azuki bean (Sakai et al., 2015). DNA

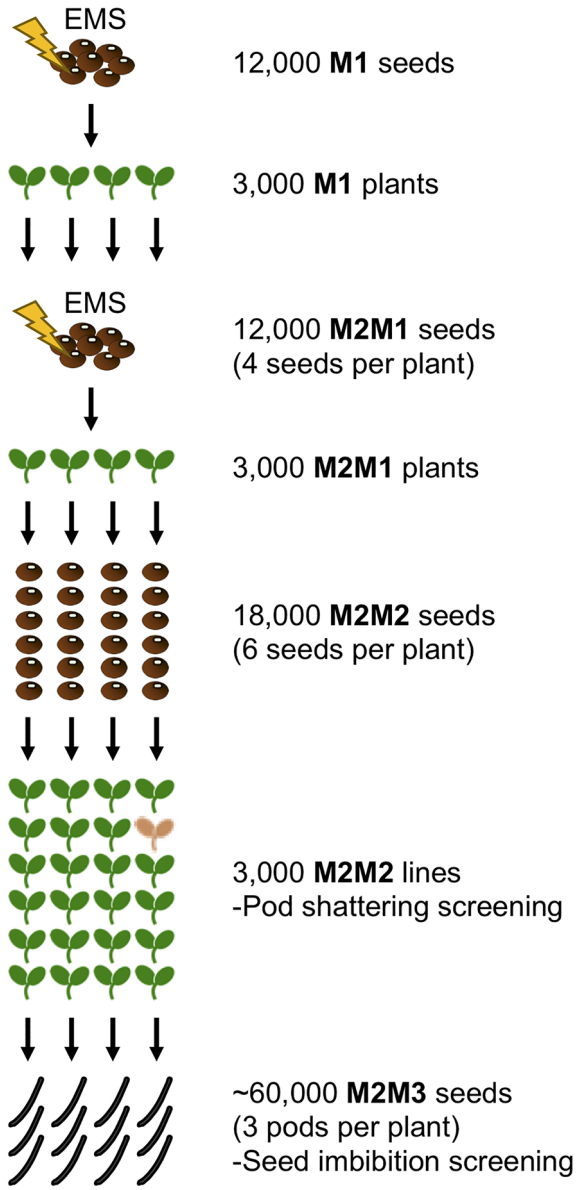


FIGURE 1 | Schematic of domesticating *V. stipulacea*. After two rounds of EMS treatment, 3,000 M2M1 plants were selected to collect M2M2 seeds (six per line). The 18,000 M2M2 plants were grown in a field, where pod shattering mutants were screened for. Three pods (~20 seeds) per plant was harvested, which were tested for seed-imbibition screening.

was isolated from 1 g of unexpanded leaves with CTAB method and purified with Genomic Tip 20/G (Qiagen K. K. Tokyo). The extracted DNA was sheared into 20 kb fragments using g-TUBE (Covaris, MA, USA) and converted into 20 kb SMRTbell template libraries. The library was size-selected for a lower cutoff of 10 kb with BluePippin (Sage Science, MA, USA). Sequencing was performed on the PacBio RS II using P5 polymerase binding and C3 sequencing kits with 360 min acquisition. In total, 52 SMRT cells were used to obtain ~19.6 Gb of subreads.

In total, ~3.3 million PacBio reads were used for *de novo* assembly with Celera Assembler 8.3rc1 (asmOvlErrorRate = 0.1, asmUtgErrorRate = 0.06, asmCgwErrorRate = 0.1, asmCnsErrorRate = 0.1, asmObtErrorRate = 0.08, utgGraphErrorRate = 0.05, utgMergeErrorRate = 0.05) (Berlin et al., 2015). About 25.2x of the longest error-corrected and trimmed reads were assembled to contigs. Redundant contigs were discarded by conducting all-to-all BLASTN searches. The non-redundant contigs were polished by PacBio subreads by using Quiver in SMRT Analysis v2.2.0 (Pacific Biosciences of California, Inc.) and then further polished by Illumina short-reads using BWA 0.7.9a (Li and Durbin, 2009), Samtools 0.1.19 (Li et al., 2009), Picard 1.94 (http://picard.sourceforge.net/), and GATK 3.3 (McKenna et al., 2010). The polished contigs were scaffolded by Reference-Assisted Chromosome Assembly (RACA) program v.0.9.1.1 (Kim et al., 2013b) using the genome sequences of *Vigna angularis* and *P. vulgaris* as the reference and outgroup species, respectively.

Repetitive sequences in the genome assembly were predicted using Censor (Kohany et al., 2006) with a composite library consisting of *de novo* created library constructed by RepeatModeler 1.0.8 (http://www.repeatmasker.org) and the MIPS Repeat Element Database ver. 9.3 (Nussbaumer et al., 2013).

Ab initio gene prediction was done by BRAKER version 1.6 (Hoff et al., 2016) with RNA-Seq data. Besides, gene structures were predicted by genome-guided and *de novo* RNA-Seq data assembly approaches using TopHat 2.1.0 (Kim et al., 2013a), Cufflinks 2.2.1 (Trapnell et al., 2010), Trinity 2.1.1 (Grabherr et al., 2011), and PASA pipeline 2.0.2 (Haas et al., 2003; Rhind et al., 2011). Open reading frames were predicted by Transdecoder 2.0.1 and Trinotate 2.0.2 (Bryant et al., 2017). Protein sequences of the *G. max* (Wm82.a2.v1), *P. vulgaris* (v.1.0), *Medicago truncatula* Gaertn. (Mt4.0v1), and *V. angularis* (Willd.) Ohwi & H.Ohashi (VANGULARIS_V1.A1) were downloaded from Phytozome (*G. max*, *P. vulgaris*, and *M. truncatula*) (Goodstein et al., 2012) and *Vig*GS (*V. angularis*) (Sakai et al., 2016) and mapped to the genome assembly by Exonerate 2.2.0 (Slater and Birney, 2005). The *ab initio* gene models and transcript and protein alignments were then combined by EvidenceModeler 1.1.1 (Haas et al., 2008) and the predicted gene models were updated by PASA. Gene models with extremely long introns and those merged by PASA were manually curated. BUSCO v3 (Waterhouse et al., 2017) was used to evaluate protein sequences of annotated genes.

Bulked Segregant Analysis

We performed MutMap (Abe et al., 2012) to isolate the candidate gene for the mutant phenotype. We crossed one of the mutants of increased seed imbibition (*isi1*) to the wild type (JP245503) and obtained 280 F2 seeds. We cultivated all of them, extracted DNA from an unexpanded leaf of each F2 plants by CTAB method, and obtained F3 seeds. To evaluate seed dormancy, 20 seeds of each F3 line was soaked with distilled water for 24 h to test imbibition. DNA of 59 F2 plants with mutant phenotype (100% seed imbibition) and DNA of 221 F2 plants with wild type phenotype (0% imbibition) were pooled to make the MT pool and the WT pool, respectively, and were sequenced with Illumina HiSeq X (Illumina Co. Ltd, San Diego, USA).

The obtained sequences were mapped to the draft genome sequences with bwa-0.7.17 (Li and Durbin 2009) and formatted with samtools-1.9 (Li et al., 2009) as described in the distributed manuals. SNPs were called with bcftools-1.9, and the DP4 values were used to calculate the SNP-index. If a SNP site is responsible for the mutant phenotype, the locus must be fixed with the alternative allele [a] in the MT pool ([aa] shares 100%). On the other hand, in the WT pool, the expected ratio of the reference allele [A] and the alternative [a] is 2:1, because the WT pool [AA]:[Aa]:[aa] should be 1:2:0. Thus, at the responsible locus, the expected values of SNP-index are 1.0 in the MT pool and 0.33 in the WT pool. Thus, after calculating SNP-index, we screened all the SNP sites for those that met the criterion above.

Genotyping by Amplicon Sequencing

We genotyped 280 F2 plants by Sanger sequencing for two SNPs whose SNP-index was 1.0 in the MutMap analysis. Two primer pairs were designed to amplify a 232 bp genomic region including 6,804,429 nt (5′-gagggaatacgaagagtttaaggtt-3′ and 5′-ttgaaaaccaggtcttttctctcta-3′) and to amplify a 243 bp genomic region including 7,009,873 nt (5′-acagagcaaaagattaaacgagaga-3′ and 5′-aaagccgcttcctagtccttac-3′) on scf0015. For Sanger sequencing, we amplified the template DNA with AmpliTaq Gold 360 Master Mix (Thermo Fisher Scientific K. K., Tokyo), performed sequencing reaction with BigDye Terminator v3.1 (Thermo Fisher Scientific K. K., Tokyo), and sequenced with ABI Genetic Analyzer 3130xl (Thermo Fisher Scientific K. K., Tokyo), according to the provider's protocol.

RESULTS

Phenotypes of Seed Dormancy and Pod Shattering in Domesticated Legumes

Though many studies have investigated seed dormancy in legume crops, there are still some arguments on the sites of water entry (Smýkal et al., 2014). Thus, we evaluated seed imbibition and pod shattering in soybean, common bean, and cowpea.

As a result, domesticated species showed variation in both imbibition start sites and time to imbibe (**Figure 2**). Among the domesticated accessions tested, soybean imbibed within a few minutes and was the quickest. In soybean, the whole testa seemed permeable. The second quickest was the cowpea "JP244182," where the testa was partially permeable and imbibition was observed in an hour. The common bean "JP41232" showed imbibition in 2 to 3 h, which started at the site of micropyle. It took 4 to 5 h for the common bean "JP41234" and the cowpea "JP239215," where imbibition started at the site of lens. All the seeds were fully imbibed within 48 h.

We also observed abscission layers between the valves of seed pods and the sclerenchyma on the endocarp, because pod shattering is dependent on these tissues (Murgia et al., 2017). In soybean, as described by Dong et al. (2014), abscission layer was not completely formed at the fiber cap cells (**Figure 3**). Abscission layer was even less completely formed in the common bean accession JP41963 whereas it was fully formed in all other accessions (**Figure 3**). Pod sclerenchyma was thicker in the soybean accessions and the cowpea accession JP244182, whereas it was thinner in common bean accessions (**Figure 3**).

Mutant Screening

To obtain mutants with reduced seed dormancy and reduced pod shattering, we treated 12,000 seeds of *V. stipulacea* with EMS and developed 3,000 M2M2 plants. Of the 3,000, we found one line which exhibited reduced pod shattering and designated it as *reduced pod shattering1* (*rps1*) (**Figure 4**). We also screened M2M3 seeds for seed imbibition and obtained three lines which exhibited increased seed imbibition and designated them as *increased seed imbibition1* (*isi1*), *increased seed imbibition2* (*isi2*), and *increased seed imbibition3* (*isi3*), respectively (**Figure 5**). Of them, the *isi1* seeds had cracks in the hilum, and the *isi3* seeds showed reduced pigmentation in the seed coat (**Figure 5**).

The mode of imbibition also varied across the mutants (**Figure 5**). In the *isi1*, the water entry was not necessarily through the cracks in the hilum but through the whole seed coat. Even though we sealed the hilum with glue, imbibition was quickly initiated [see the "*isi1* (sealed hilum)" in Figure 5]. In *isi2*, the imbibition was initiated at the lens. In *isi3*, the imbibition started near the hilum, but not necessarily through the micropyle or the lens.

Mutant Phenotypes Regarding Seed Dormancy

To elucidate the extent of seed dormancy in the mutant lines, we put the seeds on wet filter paper and evaluated the rate of imbibed seeds (**Figure 6**). We also tested the seeds of 1 and 6 months old to elucidate the effect of duration time after harvesting. When we did the experiment with the wild type seeds, the imbibition rate was zero for at least 4 weeks in 1-month-old seeds and was 11% in the 6-month-old seeds (**Figure 6**).

When we tested the mutant seeds of 1 month old, the imbibition was quick in *isi1* and *isi2*, whereas it was slow in *isi3*. In *isi1*, all the seeds fully imbibed within 1 h (**Figure 6**). In *isi2*, 85% of the seeds imbibed within a week, and almost 100% did so within 4 weeks. In contrast, the imbibition rate of the *isi3* seeds was only 16% in 4 weeks, which was not significantly different from the wild type. The imbibition rate of the reduced shattering mutant *rps1* was also higher but not significantly different compared to the wild type (**Figure 6**).

When we tested the mutant seeds of 6 months old, the overall imbibition rate was increased compared to the experiment with 1-month-old seeds (**Figure 6**). In *isi1*, the seeds also fully imbibed within 1 h. In *isi2*, the imbibition rate was already significantly different from the wild type in 1 day, more than 90% of the seeds imbibed within a week, and reached plateau (~95%) in two weeks. In *isi3*, 20% of the seeds imbibed in 3 days, 36% in a week, and 71% in 4 weeks. However, this line exhibited a large standard deviation between the replicates and was not significantly different before 2 weeks (**Figure 6**). In *rps1*, ~20% of the seeds imbibed in 4 weeks, which was twice as high as the wild type but not significantly different from the wild type was not significant.

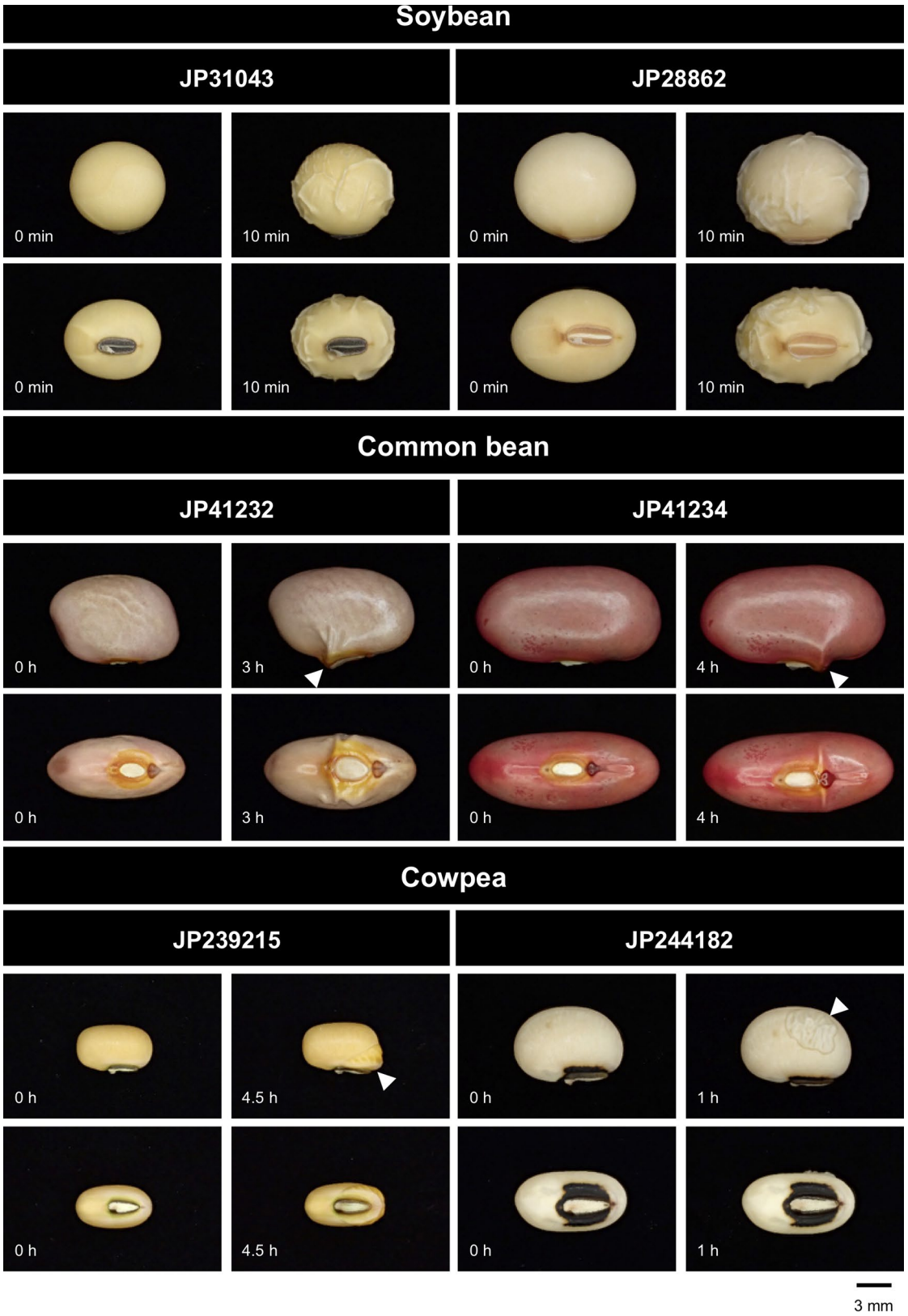


FIGURE 2 | Water entry sites in domesticated legumes. The entry sites (indicated by arrowheads) varied across accessions, but did not vary within accessions.

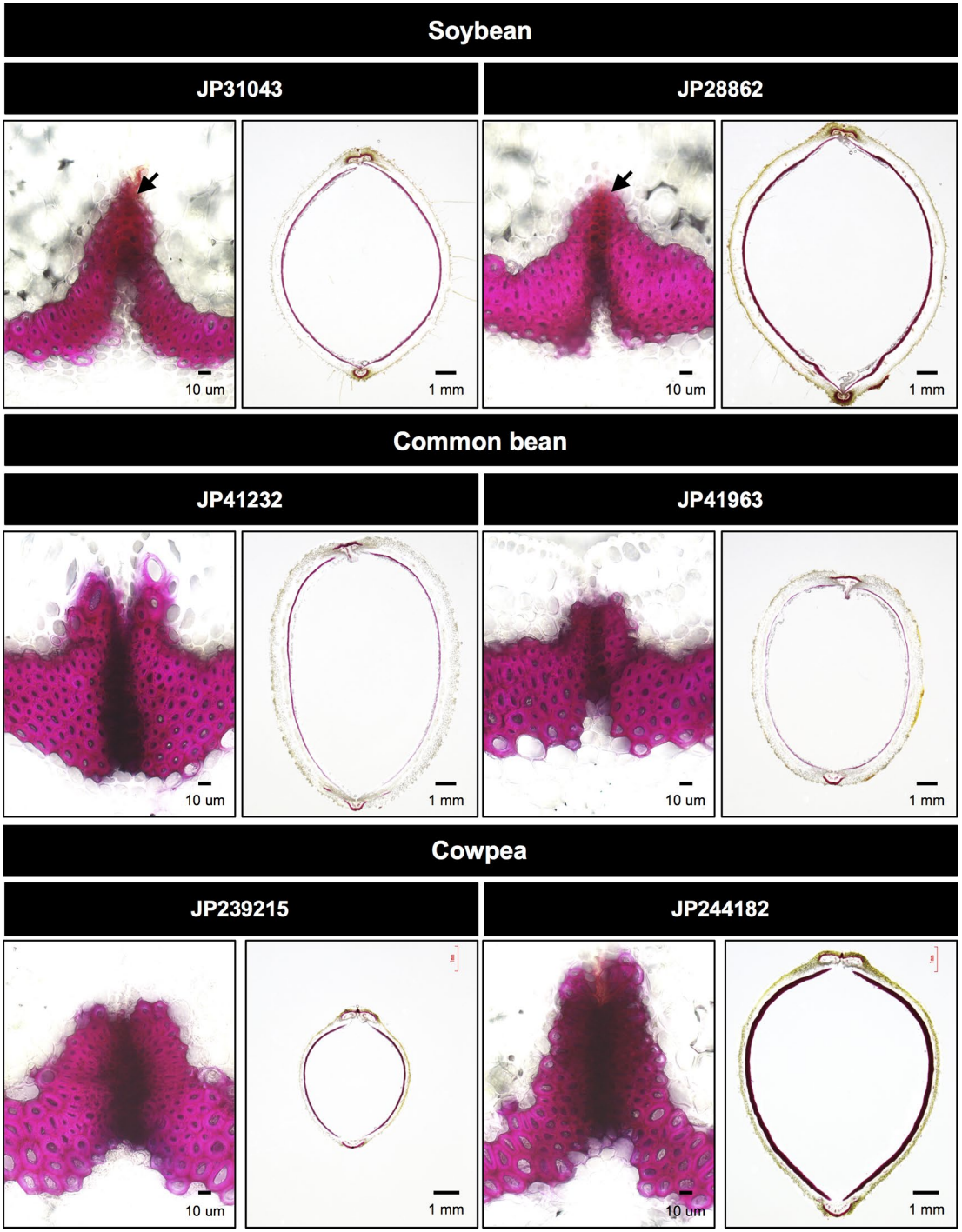


FIGURE 3 | Cross-sections of seed pods in domesticated accessions. The left panel of each accession is a close-up of abscission zone. Phloroglucinol-HCl stains pod sclerenchyma with bright red while it stains abscission zone with dark red. Soybean accessions have brightly-stained sites at the tip of fiber cap cells (arrows), which indicates abscission layer is not completely formed.

Mutant Phenotypes Regarding Pod Shattering

To evaluate pod shattering in the mutant lines, we calculated the rate of shattering of the harvested pods which were completely dried in the incubator. Whereas the shattering rate was 100% in the wild type, it was 0% in the *rps1* mutant (**Table 2**). The *rps1* mutant also showed a reduced twisting of the seed pod. The number of twists/cm in the pods was 0.371 ± 0.018 in the *rps1*

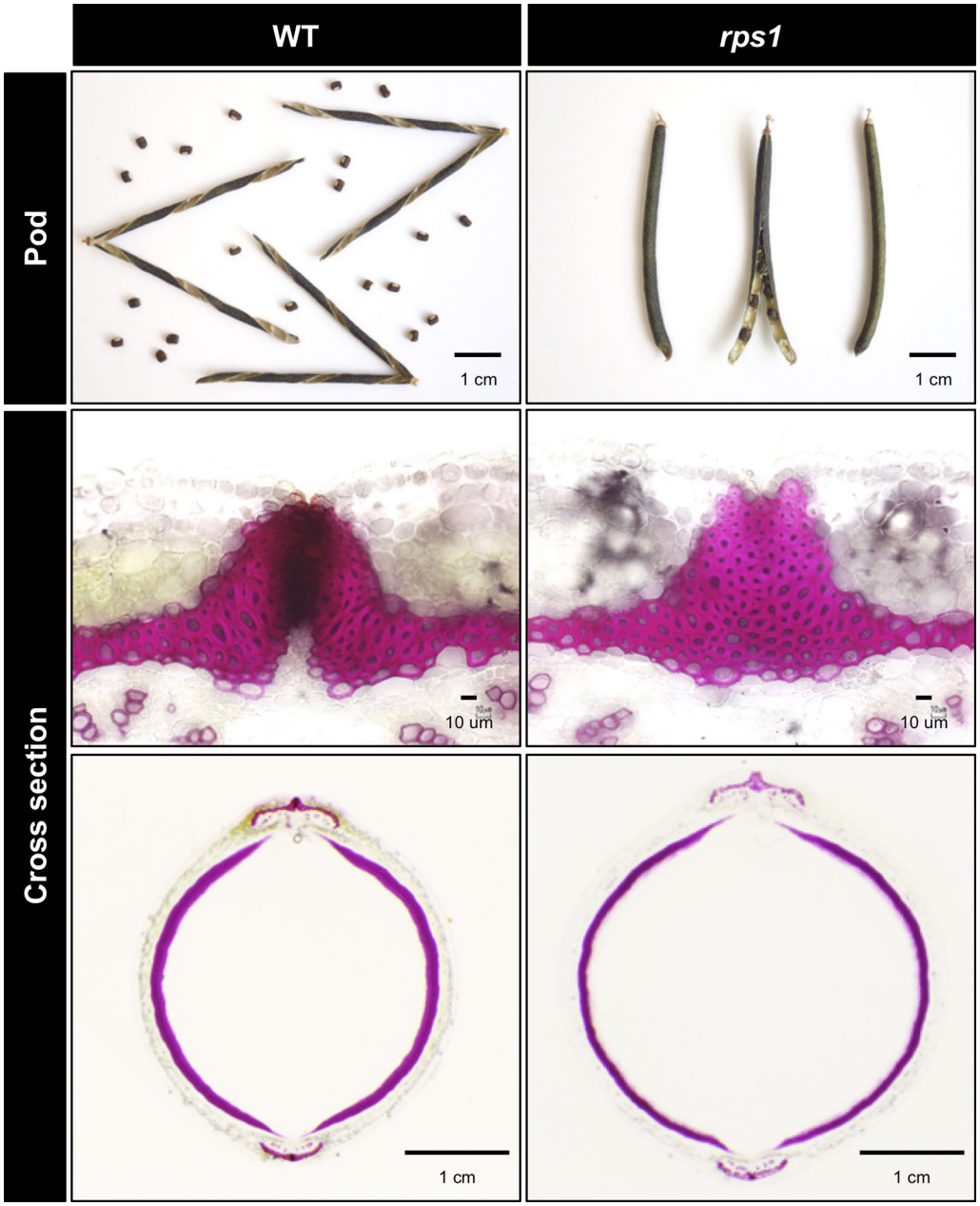


FIGURE 4 | Non-shattering mutant of *V. stipulacea*. Pod shattering was strongly suppressed in the *rps1* mutant, where the abscission layer between the valves were not formed at all, whereas the thickness of sclerenchyma did not greatly change.

mutant, which was less than half of the wild type (0.866 ± 0.022) (**Table 2**).

Interestingly, the *isi1*, one of the mutants of seed imbibition, also exhibited slightly reduced shattering rate (73.99 ± 18.53%) and number of twists/cm (0.579 ± 0.093) (**Table 2**). Other mutants were also slightly reduced in number of twists/cm, but their shattering rate was 100% (**Table 2**).

We also observed cross-sections of seed pods and found the *rps1* mutant did not form abscission layer between the valves at all (**Figure 4**).

Whole Genome Sequence and Annotation of *V. stipulacea*

To build a reference sequence of *V. stipulacea,* we sequenced the genomic DNA with a PacBio sequencer, assembled, and annotated. We obtained 19.6 Gbp of subreads using 52 SMRT cells. Genome size of the *V. stipulacea* was estimated to be ~445.1 Mbp based on the k-mer frequency distribution obtained from 10.3 Gbp of the Illumina short-reads. The assembled contigs showed N50 length of ~1.9 Mbp and mean length of ~169.8 kbp and covered ~387.7 Mbp (87.9%) of the estimated genome size

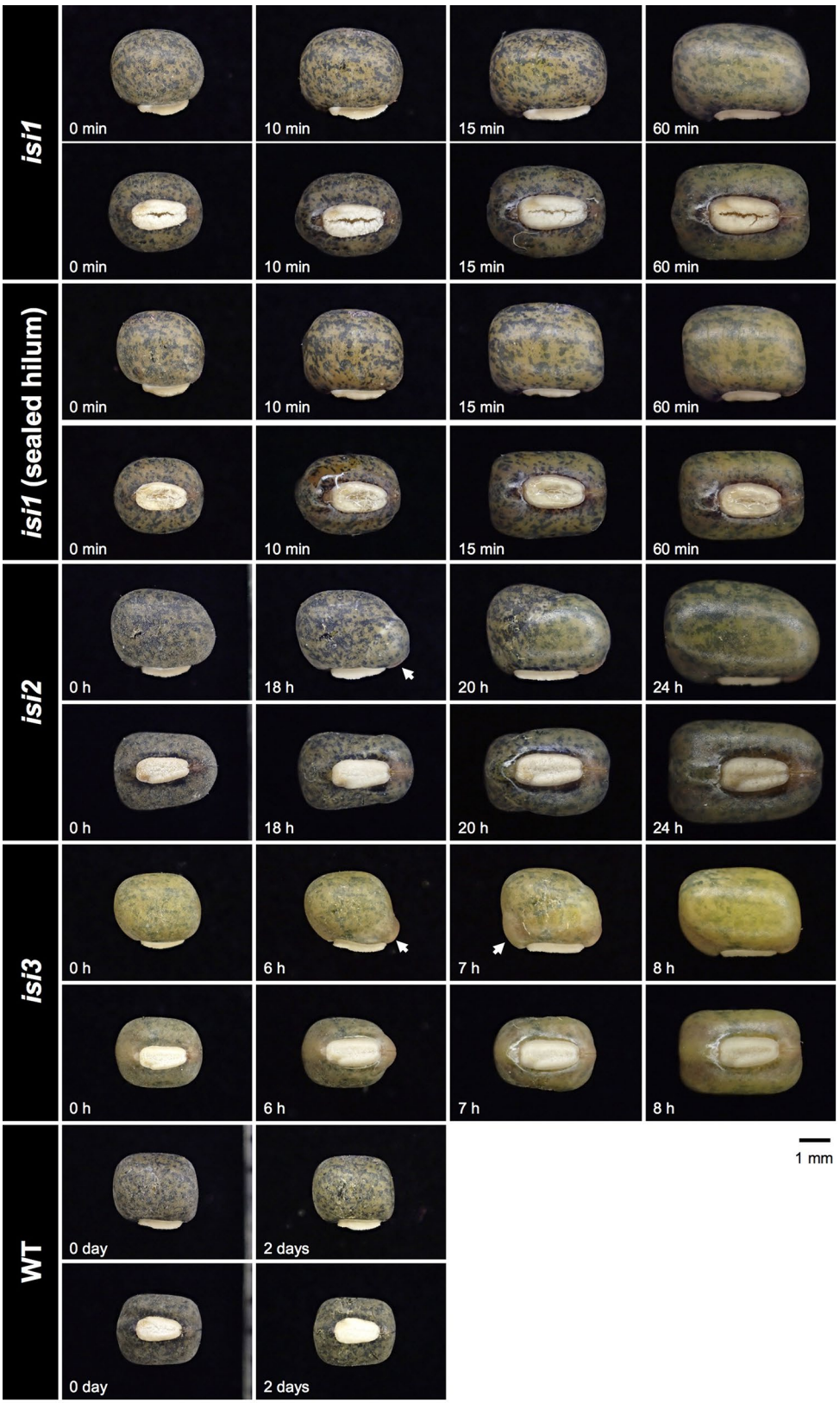


FIGURE 5 | Increased seed imbibition mutants of *V. stipulacea.* Time after watering is indicated at the bottom-left of each photo. Arrows indicate where water entry was initiated.

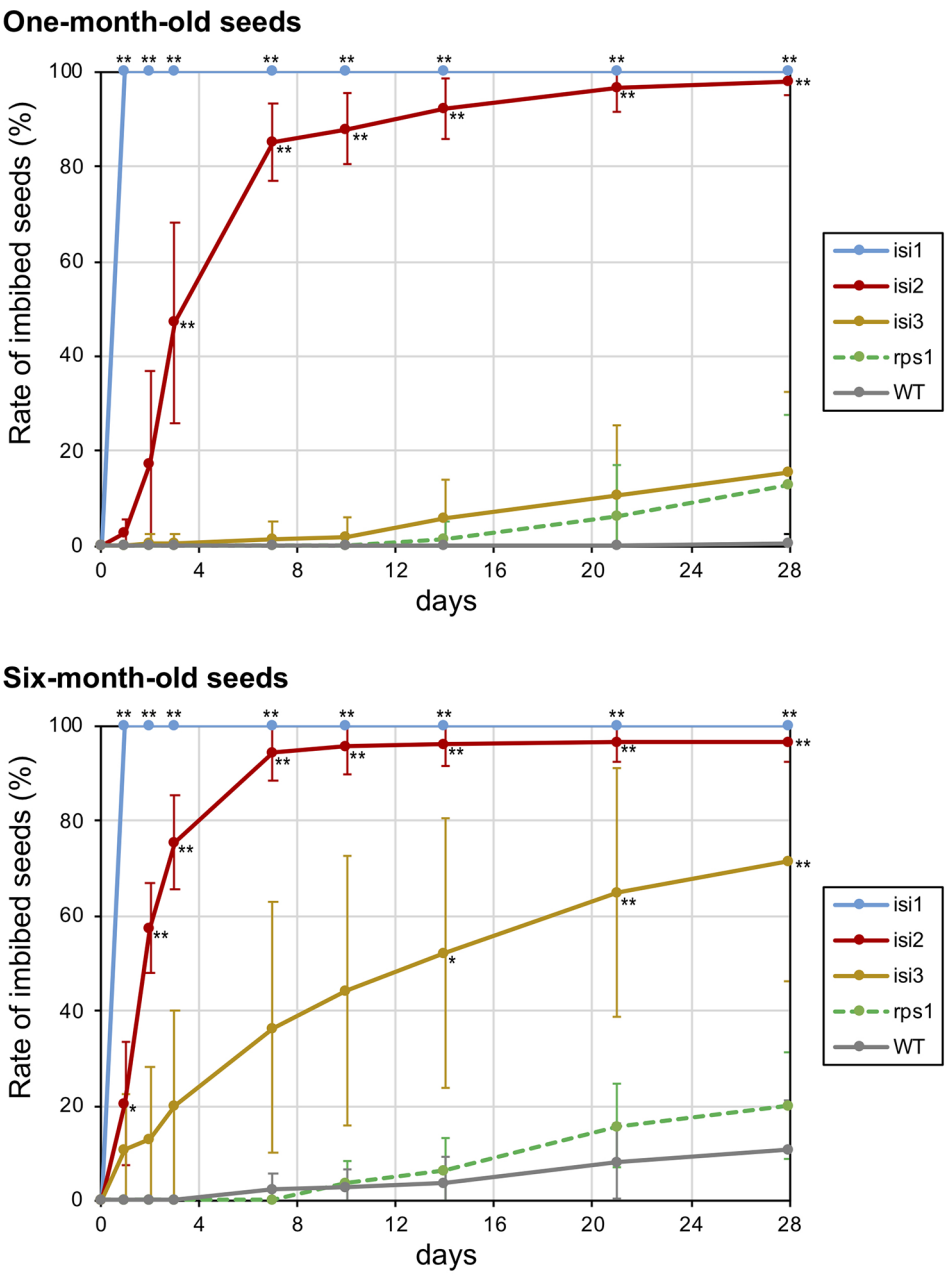


FIGURE 6 | Rate of imbibed seeds in the mutant plants over time. Twenty seeds of 1 and 6 months old were soaked in distilled water and number of imbibed seeds were manually counted twice a week for 4 weeks. The error bars indicate standard deviation of replicated evaluations (n = 7). Asterisks indicate that the mean values are significantly different from the wild type (** for $p < 0.01$ and * for $p < 0.05$).

TABLE 2 | Phenotypic data of mutant lines.

Names	Symbol	No. twist/cm ± SD	Pod shattering rate (%) ± SD
increased seed imbibition 1	*isi1*	0.579 ± 0.093	73.99 ± 18.53
increased seed imbibition 2	*isi2*	0.615 ± 0.149	100 ± 0
increased seed imbibition 3	*isi3*	0.712 ± 0.066	100 ± 0
reduced pod shattering 1	*rps1*	0.371 ± 0.018	0 ± 0
Wild type (JP245503)	WT	0.866 ± 0.022	100 ± 0

(**Table 3**). Out of the 2,102 scaffolds, 52 scaffolds were comprised of 233 contigs scaffolded by RACA.

By combining the *ab initio* gene models and transcript and protein alignments, 26,038 protein coding genes were predicted on the assembled genome (**Table 3**). We assessed the completeness of the gene set by BUSCO and found that 95.8% of the BUSCO gene models were completely detected (**Table 3**). The BUSCO metrics was almost the same as that assessed for

TABLE 3 | Stats of the assembled genome sequence of *V. stipulacea*.

Statistic items	Stats
Estimated genome size (Mbp)	441.5
Total contig length (bp)	387,726,606
No. contigs	2,283
N50 contig (bp)	1,936,224
Mean contig length (bp)	169,832
Total scaffold length (bp)	387,906,601
No. total scaffolds	2,102
N50 scaffold (bp)	8,789,545
No. protein coding genes	26,038
Mean protein length	419.6
Complete BUSCOs	1,379 (95.8%)
Complete and single-copy BUSCOs	1,274 (88.5%)
Complete and duplicated BUSCOs	105 (7.3%)
Fragmented BUSCOs	14 (1.0%)
Missing BUSCOs	47 (3.2%)

the recently published high quality genome sequence of cowpea (Lonardi et al., 2019), suggesting that our genome assembly covered nearly complete gene space of the *V. stipulacea* genome.

Mapping a Locus for *isi1* Phenotype

To identify the responsible genes for the mutant phenotype, we performed MutMap analysis. Of the 280 F2 plants derived from *isi1* x WT, 221 showed the wild type phenotype and 59 showed the mutant phenotype, which was close to the expected 3:1 ratio (p = 0.13). We sequenced the genomic DNA of the wild type plants (WT pool) and the mutants (MT pool) and identified 33,936 SNPs across the whole genome. We then calculated SNP-index of each SNP and screened for those where it was 1.0 in MT pool and 0.3 ± 0.1 in WT pool.

As a result, we found only two on scf0015, a C to T substitution at 6,804,429 nt and a G to A at 7,009,873 nt. The C to T SNP was present in the 8th exon of Vigst.0015s042600.01, which encoded CELLULOSE SYNTHASES A 7 (CesA7) protein, and turned the Glu codon into a STOP codon. The G to A SNP was present in Vigst.0015s044500.01, which encoded a nucleoporin-like protein, but was synonymous.

For further mapping, we looked for recombinants between the two candidate SNPs by direct sequencing of the SNP-containing regions in all the F2 plants. As a result, one F2 plant, which showed the wild type phenotype, was heterozygous at Vigst.0015s042600.01 but was fixed with the mutant A allele at Vigst.0015s044500.01 (see #130 in **Supplementary Table 1**). Another F2 plant, which showed the mutant phenotype, was fixed with the alternative T allele at Vigst.0015s042600.01 but was heterozygous at Vigst.0015s044500.01 (see #251 in **Supplementary Table 1**). The result was further confirmed by genotyping and phenotyping F3 plants derived from the F2 plants with recombination between the two candidate SNPs (#130, #251, and #275 in **Supplementary Table 1**). As shown in **Supplementary Table 2**, the phenotype was completely linked with the SNP at scf0015_6804429 on CesA7 gene but not with the other.

DISCUSSION

In this study we demonstrated a practical approach to *de novo* domesticate a wild legume species *V. stipulacea* by chemically-induced mutations. We successfully obtained three independent mutations in seed dormancy and one in pod shattering. In addition, we identified a SNP in the candidate gene, which encoded CELLULOSE SYNTHASE A 7, in one of the three mutants of reduced seed dormancy.

We demonstrated that forward screening following mutagenesis was practical enough to screen for mutants in domestication-related traits. This could be easily predicted from the fact that legume crops have reduced seed dormancy and pod shattering in various manners (**Figures 2** and 3). This indicates that many genetic loci are involved in these traits, and loss-of-function mutation in any of them could cause such phenotypes. This is why we identified three mutants for seed dormancy and one for pod shattering from only 3,000 M2M2 lines (18,000 plants) (Figure 1). We also note that we obtained the mutants within 3 years since we initiated the first mutagenesis.

Of the three mutants for seed dormancy, *isi1* showed the severest phenotype and imbibition was always completed within hours (**Figures 5** and **6**). This was because of water-permeability in the seed coat, which were probably caused by the loss-of-function mutation in CesA7. CesA7 is involved in cellulose synthesis and its malfunction might have disrupted development of hilum and seed coat. The phenotype of cracked hilum had not been observed in other legume crops and makes the seed almost completely non-dormant. However, the seeds might not be suitable for long-term storage, because the cotyledons are exposed to the air and could be easily oxidized. In addition, *isi1* showed slightly lower rate of pod shattering. This could be a pleiotropic effect of the mutation, because cellulose fibers are also important for pod shattering (Suanum et al., 2016). However, we cannot exclude the possibility that *isi1* contains other mutations involved in pod shattering.

Compared to *isi1*, *isi2*, and *isi3* showed milder phenotypes (**Figure 6**). Because *isi2* seeds imbibe through lens, the mutation disturbed the same pathway as in the common bean "JP41234" or the cowpea "JP239215" (**Figure 2**). However, it took much longer time for *isi2* to complete imbibition than for those cultivars. This might be because the domesticated accessions have accumulated multiple mutations involved in seed coat permeability (Jang et al., 2015; Sun et al., 2015). As for *isi3*, the seed coat color was retarded but imbibition did not always initiate in the seed coat. It sometimes did through the lens and sometimes through the micropyle. Given the water entry sites are basically stable in the cultivated accessions, the *isi3* mutant might be a novel phenotype of seed dormancy. Or, it is also possible that *isi3* have multiple mutations involved in seed imbibition, given we are currently aware of only the seed color in this mutant. If so, such mutant alleles (except the one involved in seed pigmentation) could

be segregating and lead to variation of water entry sites and imbibition rate within the mutant line (**Figure 6**). Though *isi2* and *isi3* have only partial effect on seed imbibition, double-mutant of both might be able to complete imbibition within a week with smaller effect on long-term storability.

The *rps1* mutant almost completely lost the pod shattering behavior because of suppressed formation of the abscission layer between the valves (**Figure 4**). This phenotype was similar to the effect of the domestication-type allele of SHAT1-5 in soybean (**Figure 2**) (Dong et al., 2014). Thus, the mutation in *rps1* might be in a gene involved in the SHAT1-5 pathway. The responsible gene for *rps1* phenotype might be useful for improving shattering problem in other legumes because *rps1* phenotype was severer than soybean SHAT1-5. On the other hand, however, severe disruption in development of abscission layer could increase labor to thresh. In addition, though not significant, we repeatedly observed that the *rps1* mutant exhibited slightly increased seed imbibition compared to the wild type (**Figure 6**). Such pleiotropy, unless it has other mutations involved in seed dormancy, might be because secondary wall thickening plays important roles in shattering behavior in seed pod (Suanum et al., 2016; Parker et al., 2019; Rau et al., 2019; Takahashi et al., 2019) and water permeability in seed coat (Smýkal et al., 2014; Hradilová et al., 2019).

Last but not least, the technology of long-read sequencing (Eid et al., 2009) enabled us to obtain high-quality genome assembly of *V. stipulacea*. Although we did not use further scaffolding techniques such as optical mapping (Lam et al., 2012) or Hi-C (Burton et al., 2013), it achieved the N50 of nearly 2 Mbp and the BUSCO score of 96.8 (**Table 3**). The high-quality reference genome should be the reason for that we successfully identified the candidate SNP by a single resequencing analysis.

As described above, here we report that we have achieved the first step of domesticating the wild legume *V. stipulacea*. As reported in Tomooka et al. (2011), this legume showed resistance to broad range of pests and diseases in Tamil Nadu, India, and does not need pesticides to spray. Broad range of pests and diseases resistance of *V. stipulacea* was confirmed during the mutant screening procedure conducted in our experimental field in Tsukuba, Japan. Thus, if we cross and pyramid the mutant phenotypes obtained in this study, we will be able to make *V. stipulacea* easier to cultivate and easier to harvest, and we believe it will be a step forward to accomplish low input agriculture. By screening for more useful traits from more mutagenized population, we hope this kind of "domesticated wild plants" will be popular and prevalent.

DATA AVAILABILITY STATEMENT

The raw and assembled sequence data (DRA009127) are all available from DNA Data Bank of Japan (https://www.ddbj.nig.ac.jp/dra/index.html) or *Vig*GS (https://viggs.dna.affrc.go.jp/).

AUTHOR CONTRIBUTIONS

NT and KN provided the idea of the study. YT, MP, NS, NT and KN planned the study. YT, CM, TA, NT and KN cultivated plants and collected data. YT, YY, HS and KN analyzed data. YT, HS, MP, NS, NT and KN wrote the paper.

FUNDING

This study was supported by JSPS KAKENHI Grant Number 13J09808, 19KT0016, and 26850006. It was also partially supported by Research Supporting Program of the Advanced Analysis Center, National Agriculture and Food Research Organization (NARO) and the Genebank Project, NARO.

ACKNOWLEDGMENTS

We are grateful for Ms. Motoyoshi, Ms. Asano and Ms. Yamamoto for taking good care of plant materials. We also appreciate for the support provided by Research Supporting Program of the Advanced Analysis Center, National Agriculture and Food Research Organization (NARO) and the Genebank Project, NARO.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2019.01607/full#supplementary-material

REFERENCES

Østerberg, J. T., Xiang, W., Olsen, L. I., Edenbrandt, A. K., Vedel, S. E., Chistiansen, A., et al. (2017). Accelerating the domestication of new crops. *Trends Plant Sci.* 22, 373–384. doi: 10.1016/j.tplants.2017.01.004

Abe, A., Kosugi, S., Yoshida, K., Natsume, S., Takagi, H., Kanzaki, H., et al. (2012). Genome sequencing reveals agronomically important loci in rice using MutMap. *Nat. Biotechnol.* 30, 174–178. doi: 10.1038/nbt2095

Berlin, K., Koren, S., Chin, C. S., Drake, J. P., Landolin, J. M., and Phillippy, A. M. (2015). Assembling large genomes with single-molecule sequencing and locality-sensitive hashing. *Nat. Biotechnol.* 33, 623–630. doi: 10.1038/nbt3238

Bryant, D. M., Johnson, K., Ditommaso, T., Tickle, T., Couger, M. B., Payzin-Dogru, D., et al. (2017). A tissue-mapped axolotl de novo transcriptome enables identification of limb regeneration factors. *Cell Rep.* 18, 762–776. doi: 10.1016/j.celrep.2016.12.063

Burton, J. N., Ady, A., Patwardhan, R. P., Qiu, R., Kitzman, J. O., and Shendure, J. (2013). Chromosome-scale scaffolding of *de novo* genome assemblies based on chromatin interactions. *Nat. Biotechnol.* 31, 1119–1125. doi: 10.1038/nbt2727

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

Dong, Y., Yang, X., Liu, J., Wang, B. H., Liu, B. L., and Wang, Y. Z. (2014). Pod shattering resistance associated with domestication is mediated by NAC gene in soybean. *Nat. Commun.* 5, 3352. doi: 10.1038/ncomms4352

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., et al. (2009). Real-time DNA sequencing from single polymerase molecules. *Science* 323, 133–138. doi: 10.1126/science.1162986

Fernie, A. R., and Yan, J. (2019). De novo domestication: an alternative route toward new crops for the future. *Mol. Plant* 12, 615–631. doi: 10.1016/j.molp.2019.03.016

Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., et al. (2012). Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* 40, D1178–D1186. doi: 10.1093/nar/gkr944

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29, 644–652. doi: 10.1038/nbt1883

Haas, B. J., Delcher, A. L., Mount, S. M., Wortman, J. R., Smith, R. K. Jr, Hannick, L. I., et al. (2003). Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res.* 31, 5654–5666. doi: 10.1093/nar/gkg770

Haas, B. J., Salzberg, S. L., Zhu, W., Pertea, M., Allen, J. E., Orvis, J., et al. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* 9, R7. doi: 10.1186/gb-2008-9-1-r7

Hoff, K. J., Lange, S., Lomsadze, A., Borodovsky, M., and Stanke, M. (2016). BRAKER1: Unsupervised RNA-Seq-Based Genome Annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics* 32, 767–769. doi: 10.1093/bioinformatics/btv661

Hradilová, I., Duchoslav, M., Brus, J., Pechanec, V., Hýbl, M., Kopecký, P., et al. (2019). Variation in wild pea (*Pisum sativum* subsp. *elatius*) seed dormancy and its relationship to the environment and seed coat traits. *PeerJ* 7, e6263. doi: 10.7717/peerj6263

Jang, S. J., Sato, M., Sato, K., Jitsuyama, Y., Fujino, K., Mori, H., et al. (2015). A Single-Nucleotide Polymorphism in an Endo-1,4-β-Glucanase Gene Controls Seed Coat Permeability in Soybean. *PloS One* 10, e0128527. doi: 10.1371/journal.pone.0128527

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S. L. (2013a). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 14, R36. doi: 10.1186/gb-2013-14-4-r36

Kim, J., Larkin, D. M., Cai, Q., Asan, Zhang, Y., Ge, R. L., et al. (2013b). Reference-assisted chromosome assembly. *Proc. Natl. Acad. Sci. U.S.A.* 110, 1785–1790. doi: 10.1073/pnas.1220349110

Kohany, O., Gentles, A. J., Hankus, L., and Jurka, J. (2006). Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. *BMC Bioinf.* 7, 474. doi: 10.1186/1471-2105-7-474

Lam, E. T., Hastie, A., Lin, C., Ehrlich, D., Das, S. K., Austin, M. D., et al. (2012). Genome mapping on nanochannel arrays for structural variation analysis and sequence assembly. *Nat. Biotechnol.* 30, 771–776. doi: dio: 10.1038/nbt.2303

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

Li, T., Yang, X., Yu, Y., Si, X., Zhai, X., Zhang, H., et al. (2018). Domestication of wild tomato is accelerated by genome editing. *Nat. Biotechnol.* 36, 1160–1163. doi: 10.1038/nbt4273

Lonardi, S., Munoz-Amatriain, M., Liang, Q., Shu, S., Wanamaker, S. I., Lo, S., et al. (2019). The genome of cowpea (*Vigna unguiculata* [L.] Walp.). *Plant J.* 98, 767–782. doi: 10.1111/tpj.14349

McCouch, S., Baute, G. J., Bradeen, J., Bramel, P., Bretting, P. K., Buckler, E., et al. (2013). Agriculture: Feeding the future. *Nature* 499, 23–24. doi: 10.1038/499023a

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

Murgia, M. L., Attene, G., Rodriguez, M., Bitocchi, E., Bellucci, E., et al. (2017). A comprehensive phenotypic investigation of the "pod-shattering syndrome" in common bean. *Front. Plant Sci.* 8, 251. doi: 10.3389/fpls.2017.00251

Nussbaumer, T., Martis, M. M., Roessner, S. K., Pfeifer, M., Bader, K. C., Sharma, S., et al. (2013). MIPS PlantsDB: a database framework for comparative plant genome research. *Nucleic Acids Res.* 41, D1144–D1151. doi: 10.1093/nar/gks1153

Parker, T. A., Mier y Teran, J. C. B., Palkovic, A., Jernstedt, J., and Gepts., P. (2019). Genetic control of pod dehiscence in domesticated common bean: associations with range expansion and local aridity conditions. *biorxiv*, 517516. doi: 10.1101/517516

Pingali, P. L. (2012). Green revolution: impacts, limits and the path ahead. *Proc. Natl. Acad. Sci. U.S.A.* 109, 12302–12308. doi: 10.1073/pnas.0912953109

Rau, D., Murgia, M. L., Rodriguez, M., Bitocchi, E., Bellucci, E., Fois, D., et al. (2019). Genomic dissection of pod shattering in common bean: mutations at non-orthologous loci at the basis of convergent phenotypic evolution under domestication of leguminous species. *Plant J.* 97, 693–714. doi: 10.1111/tpj.14155

Rhind, N., Chen, Z., Yassour, M., Thompson, D. A., Haas, B. J., Habib, N., et al. (2011). Comparative functional genomics of the fission yeasts. *Science* 332, 930–936. doi: 10.1126/science.1203357

Sakai, H., Naito, K., Ogiso-Tanaka, E., Takahashi, Y., Iseki, K., Muto, C., et al. (2015). The power of single molecule real-time sequencing technology in the *de novo* assembly of a eukaryotic genome. *Sci. Rep.* 5, 16780. doi: 10.1038/srep16780

Sakai, H., Naito, K., Takahashi., Y., Sato., T., Yamamoto., T., Muto., et al. (2016). The Vigna Genome Server, 'VigGS': a genomic knowledge base of the genus vigna based on high-quality, annotated genome sequence of the azuki bean, vigna angularis (Willd) Ohwi & Ohashi. *Plant Cell Physiol.* 57, e2. doi: 10.1093/pcp/pcv189

Shapter, F. M., Cross, M., Ablett, G., Malory, S., Chivers, I. H., King, G. J., et al. (2013). High-throughput sequencing and mutagenesis to accelerate the domestication of *Microlaena stipoides* as a new food crop. *PloS One* 8, e82641. doi: 10.1371/journal.pone.0082641

Slater, G. S., and Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. *BMC Bioinf.* 6, 31. doi: 10.1186/1471-2105-6-31

Smýkal, P., Vernoud, V., Blair, M. W., Soukup, A., and Thompson, R. D. (2014). The role of the testa during development and in establishment of dormancy of the legume seed. *Front. Plant Sci.* 5, 351. doi: 10.3389/fpls.2014.00351

Suanum, W., Somta, P., Kongjaimun, A., Yimram, T., Kaga, A., Tomooka, N., et al. (2016). Co-localization of QTLs for pod fiber content and pod shattering in F2 and backcross populations between yardlong bean and wild cowpea. *Mol. Breed.* 36, 80. doi: 10.1007/s11032-016-0505-8

Sun, L., Miao, Z., Cai, C., Zhang, D., Zhao, M., Wu, Y., et al. (2015). GmHs1-1, encoding a calcineurin-like protein, controls hard-seededness in soybean. *Nat. Genet.* 47, 939–943. doi: 10.1038/ng3339

Takahashi, Y., Kongjaimun, A., Muto, C., Kobayashi, Y., Kumagai, M., Sakai, H. et al. (2019) Genetic factor for twisting legume pods identified by fine-mapping of shattering-related traits in azuki bean and yard-long bean. *bioRxiv* doi: 10.1101/774844

Tomooka, N., Pandiyan, M., and Senthil, N. (2011). Conservation of leguminous crops and their wild relatives in Tamil Nadu, India. *Annu. Rep. Explor. Introduction Plant Genet. Resour.* 27, 111–127.

Tomooka, N., Naito, K., Kaga, A., Sakai, H., Isemura, T., Ogiso-Tanaka, E., et al. (2014). Evolution, domestication and neo-domestication of the genus Vigna. *Plant Genet. Resour.* 12, S168–S171. doi: 10.1017/S1479262114000483

Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., Van Baren, M. J., et al. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28, 511–515. doi: 10.1038/nbt1621

Tsuda, M., Kaga, A., Anai, T., Shimizu, T., Sayama, T., Takagi, K., et al. (2015). Construction of a high-density mutant library in soybean and development of a mutant retrieval method using amplicon sequencing. *BMC Genomics* 16, 1014. doi: 10.1186/s12864-015-2079-y

Waterhouse, R. M., Seppey, M., Simão, F. A., Manni, M., Ioannidis, P., Klioutchnikov, G., et al. (2017). BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol. Biol. Evol.* 35, 543–548. doi: 10.1093/molbev/msx319

Zhang, H., Li, Y., and Zhu, J. K. (2018). Developing naturally stress-resistant crops for as sutaibable agriculture. *Nat. Plants* 4, 989–996. doi: 10.1038/s41477-018-0309-4

Conflict of Interest: 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 © 2019 Takahashi, Sakai, Yoshitsu, Muto, Anai, Pandiyan, Senthil, Tomooka and Naito. 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) and the copyright owner(s) 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 Plant Science

ORIGINAL RESEARCH
published: 28 February 2020
doi: 10.3389/fpls.2020.00189

Diversity of Naturalized Hairy Vetch (*Vicia villosa* Roth) Populations in Central Argentina as a Source of Potential Adaptive Traits for Breeding

Juan P. Renzi[1,2*], *Guillermo R. Chantre*[2,3], *Petr Smýkal*[4], *Alejandro D. Presotto*[2,3], *Luciano Zubiaga*[1], *Antonio F. Garayalde*[5] *and Miguel A. Cantamutto*[1,2,3]

[1] EEA H. Ascasubi Instituto Nacional de Tecnología Agropecuaria, Buenos Aires, Argentina, [2] Departamento de Agronomía, Universidad Nacional del Sur, Bahía Blanca, Argentina, [3] Centro de Recursos Naturales Renovables de la Zona Semiárida (CERZOS), CONICET, Bahía Blanca, Argentina, [4] Department of Botany, Palacký University in Olomouc, Olomouc, Czechia, [5] Departamento de Matemática, Universidad Nacional del Sur, Bahía Blanca, Argentina

OPEN ACCESS

Edited by:
Laurent Gentzbittel, National Polytechnic Institute of Toulouse, France

Reviewed by:
Hamid Khazaei, University of Saskatchewan, Canada
Fred Stoddard, University of Helsinki, Finland

***Correspondence:**
Juan P. Renzi
renzipugni.juan@inta.gob.ar

Specialty section:
This article was submitted to Plant Breeding, a section of the journal Frontiers in Plant Science

Received: *07 October 2019*
Accepted: *07 February 2020*
Published: *28 February 2020*

Citation:
Renzi JP, Chantre GR, Smýkal P, Presotto AD, Zubiaga L, Garayalde AF and Cantamutto MA (2020) Diversity of Naturalized Hairy Vetch (Vicia villosa Roth) Populations in Central Argentina as a Source of Potential Adaptive Traits for Breeding. Front. Plant Sci. 11:189. doi: 10.3389/fpls.2020.00189

Hairy vetch (*Vicia villosa* ssp. *villosa* Roth) is native of Europe and Western Asia and it is the second most cultivated vetch worldwide. Hairy vetch is used as forage species in semiarid environments and as a legume cover crop in sub-humid and humid regions. Being an incompletely domesticated species, hairy vetch can form spontaneous populations in a new environment. These populations might contain novel and adaptive traits valuable for breeding. Niche occupancy based on geographic occurrence and environmental data of naturalized populations in central Argentina showed that these populations were distributed mainly on disturbed areas with coarse soil texture and alkaline-type soils. Low rainfall and warm temperatures during pre- and post-seed dispersal explained the potential distribution under sub-humid and semiarid conditions from Pampa and Espinal ecoregions. Conversely, local adaptation along environmental gradients did not drive the divergence among recently established Argentinian (AR) populations. The highest genetic diversity revealed by microsatellite analysis was observed within accessions (72%) while no clear separation was detected between AR and European (EU) genotypes, although naturalized AR populations showed strong differentiation with the wild EU accessions. Common garden experiments were conducted in 2014–16 in order to evaluate populations' germination, flowering, and biomass traits. European cultivars were characterized by low physical seed dormancy (PY), while naturalized AR accessions showed higher winter biomass production. Detected variation in the quantitative assessment of populations could be useful for selection in breeding for traits that convey favorable functions within specific contexts.

Keywords: *Vicia villosa* genotypes, naturalized population, niche-modeling, genetic resource, phenotypic characterization, microsatellites

INTRODUCTION

The *Vicia* genus, of the Fabaceae family, includes several winter annual legumes, generically grouped as "vetches." Within this complex, *Vicia villosa* ssp. *villosa* Roth, commonly known as hairy vetch (HV), is a relevant member. It is native in Europe and West Asia, being introduced as a crop or weed worldwide to temperate climate regions. Hairy vetch is considered a cosmopolitan species due to its high capacity to naturalize under different conditions. It is present in the flora of both South and North Americas, including Argentina (Van de Wouw et al., 2001; Bryant and Hughes, 2011).

Hairy vetch is the second most important vetch in agricultural systems worldwide (Francis et al., 1999). Generally, it is grown for forage, consumed under direct or indirect grazing, or for green manure. In conservation agriculture, the use of HV as cover crop is increasing. Hairy vetch displays high tolerance to biotic and abiotic stresses (Francis et al., 1999). It is one of the recommended cover crop in organic or conservation farming, mainly because it enhances soil nitrogen content by biological fixation (Vanzolini, 2011). Due to its valuable traits, HV could help to improve soil structure, reduce soil erosion, and enhance weed suppression (Clark, 2007; Wayman et al., 2016; Frasier et al., 2017). Typically, HV produces between 2.6 and 6.2 ton ha^{-1} of above-ground dry biomass (Lawson et al., 2015; Mirsky et al., 2017; Ackroyd et al., 2019).

Hairy vetch shows the capacity to form spontaneous populations in ruderal habitats of cultivated areas (Aarssen et al., 1986; Renzi and Cantamutto, 2013). Under natural conditions, the ability to regenerate populations from the soil seed bank is associated with primary combinational seed dormancy (i.e., physical plus physiological dormancy, PY+PD) (Renzi et al., 2014). These naturalized populations can be useful as a genetic resource for breeding. However, despite the high potential agronomic value, HV is an incompletely domesticated species and only a few improved varieties exist. Likewise, HV cover crops are often unreliable in terms of establishment, performance and biomass production (Wilke and Snapp, 2008; Aapresid, 2018). Altogether these drawbacks frequently limit HV adoption by farmers (Maul et al., 2011). Studies concerning geographic distribution and climatic requirements of this species are scarce (Aarssen et al., 1986). The study of the ecological niches of natural HV populations would improve our understanding of its potential adaptation to different environmental factors.

The most important breeding goals of HV include high early vigor, high winter, and spring biomass production and low level of seed dormancy (mainly due to the physical component of primary dormancy). Rapid growth under low temperatures is especially important when HV is used to produce biomass at the end of winter (i.e., cover crop). The time-window for HV growth control (by desiccation or mechanical methods) during early spring, depends on the planting date of the subsequent summer crop. As the growth rate of HV accelerates with the spring advance the adjustment of the control intervention is critical to determine the cover crop performance. Moreover, HV spring biomass production largely determines the amount of nitrogen supplied to subsequent cash crops (Vanzolini, 2011). An additional challenge of HV is seed dormancy control, which can lead to incomplete emergence after seeding (Jacobsen et al., 2010; Maul et al., 2011). On the contrary, in agroecosystems of semiarid regions, HV natural reseeding capacity is a desirable trait reducing establishment costs (Renzi et al., 2017; Renzi et al., 2019), especially when used as forage crop by livestock farmers.

HV was introduced in Argentina more than a century ago (Manganaro, 1919). Since then, several naturalized populations have been established in ruderal habitats surrounding agricultural areas. These populations are considered an unexplored genetic resource for breeding. However, to confirm their potential value as a genetic resource, it is imperative to collect and characterize such material, by comparison, to currently registered cultivated accessions.

The objectives of this study are: i) to describe the natural habitats of naturalized HV populations from Argentina, ii) to assess the phenotypic variability of naturalized populations compared with a set of 41 introduced accessions of HV (including wild and cultivars), and iii) to study the genetic structure using simple sequence repeat (SSR) markers.

MATERIALS AND METHODS

Ecological Characterization

Naturalized Populations

The study area comprised nine provinces: Buenos Aires, La Pampa, Río Negro, Neuquén, Mendoza, Córdoba, San Luis, Santa Fe, and Entre Ríos, belonging to three eco-regions: Pampa, Espinal, and Shrubs of Plateau and Plains (**Figure 1A**). Three exploration trips were accomplished during December 2013-2015, covering a total of 21.400 km. The survey on HV populations was based on specialized systematic bibliography (Burkart, 1952), and voucher specimens deposited at Instituto de Botánica Darwinion (http://www2.darwin.edu.ar) and Museo de La Plata (http://www.museo.fcnym.unlp.edu.ar). To be considered, studied HV populations must be observed for at least two different years at the same locality, and they must contain more than 50 individuals. Recorded information of collection site consisted of i) ecological region, ii) latitude, longitude and altitude, iii) environment (soil and climate) and plant community (dominance of co-occurring plant species) characterized by family, iv) life cycle and origin (Marzocca, 1994). Global positioning system (GPS) coordinates of 63 naturalized populations were recorded (**Table S1**).

Environmental Variables

The WorldClim (http://worldclim.org) version 2.0 database was used to extract information about the climate (period 1970–2000). Data were extracted using DIVA-GIS software from ESRI grids with a spatial resolution of 30 arc-seconds (~ 1 km) in the WGS-84 (EPSG: 4326). Bioclimatic variables (BIO1–BIO19) were derived from monthly temperature and rainfall values (Fick and Hijmans, 2017). To avoid over-parameterization, 10 bioclimatic variables were selected to represent annual trends and extreme conditions of temperature and precipitation: annual

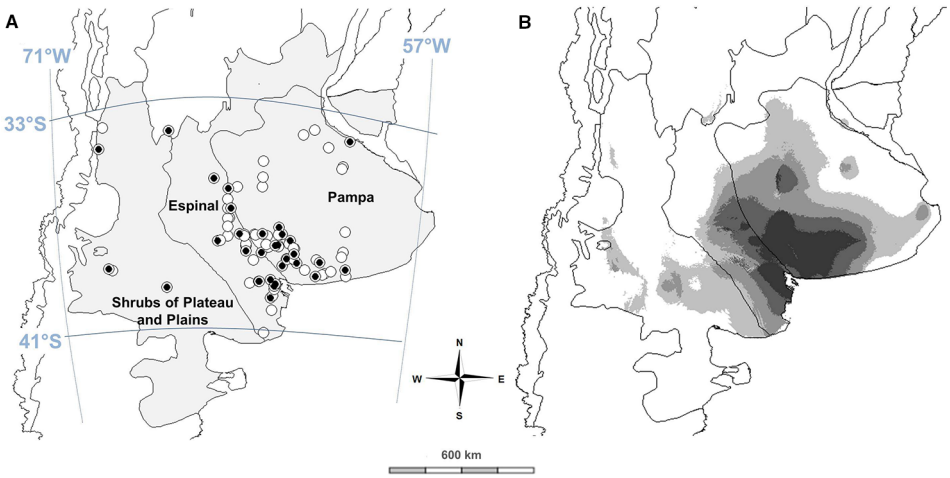


FIGURE 1 | Spatial distribution and seed collection sites (circles) of naturalized hairy vetch populations in Argentina. Gray area shows studied regions and black circles indicate the populations used for phenotypic characterization in the common garden **(A)**. Predicted potential distribution of HV populations in central Argentina based on climatic niche modeling results **(B)**. Lighter colors correspond to lower probabilities of occurrence while darker colors correspond to higher probabilities of occurrence (created with MaxEnt 3.4.1k).

mean temperature (BIO1), maximum temperature of the warmest month (BIO5), minimum temperature of the coldest month (BIO6), mean temperature of warmest quarter (BIO10), mean temperature of coldest quarter (BIO11), annual precipitation (BIO12), precipitation of the wettest month (BIO13), precipitation of the driest month (BIO14), precipitation of wettest quarter (BIO16), and precipitation of driest quarter (BIO17) (Hijmans et al., 2005). In addition to the bioclimatic variables, edaphic variables related to soil texture, pH, and bulk density were obtained from soil databases (FAO/IIASA/ISSCAS/JRC, 2012) using WGS84 and spatial resolution of 30 arc-seconds. For details, including basic descriptive statistics of each environmental variable see **Table S1**.

Soil samples (at depth of 0–15 cm) were collected at each site to assess the data obtained from soil databases. Soil samples were air-dried and sieved to < 2 mm. pH was measured using a glass electrode pH-meter (soil: water, 1:2.5). Texture analysis (% clay, silt, sand) on HCl and H_2O_2 treated and chemically [0.05 M $(NaPO_3)_6$ and 0.15 M Na_2CO_3] dispersed samples was carried out by a combination of sieving and pipette methods (Cantamutto et al., 2008; Santos et al., 2017). Pearson's Correlation Coefficient was performed between soil databases and samples using the InfoStat software (Di Rienzo et al., 2013).

Niche Analysis

Ecological niche models were constructed using the geographic locations of HV naturalized populations. Maxent (version 279 3.4.1, Phillips et al., 2018) at default conditions, a maximum-entropy based machine learning method was used for modeling purposes. Maxent showed better performance than other methods when samples sizes are small (Hernandez et al., 2006) and it estimates the potential niche instead of the realized distribution of the modeled entity (Phillips et al., 2006; Hradilová et al., 2019). As environmental predictors, bioclimatic and soil variables at a resolution of 2.5 arc minutes were used. Logistic output with suitability values ranging from 0 (unsuitable habitat) to 1 (optimal habitat) was used. Occurrence points (75%) were used to calibrate the model. The remaining (25%) occurrence points were used for model evaluation. Model strength was quantified using the area under the curve (AUC) of the receiver operator generated within Maxent (Zhu et al., 2017; Phillips et al., 2018).

Phenotypic Characterization

Plant Material

In order to determine the biodiversity of studied naturalized populations (**Figure 1A**, **Table S1**), we compared accessions originating from Argentina (AR) and Europe (EU). Twenty-nine naturalized populations were evaluated in a common garden (2014 and/or 2016) experiment (**Table 1**). They were selected based on the amount of available collected seed stock (> 30 g) and wide distribution (**Figure 1A**). Cultivated germplasm from AR consisted of 10 varieties (landraces) maintained by farmers (**Table 1**) and two registered cultivars (Tolse F.C.A and Ascasubi INTA) (www.inase.gov.ar).

Wild (n = 5) and cultivated germplasm (n = 24) of HV from EU was represented by 29 accessions (**Table 1**). Origin and accession name of each cultivar were provided by the Research Institute of Crop Production (CRI) of the Czech Republic (**Table 1**; for more information see https://grinczech.vurv.cz/gringlobal/search.aspx) (Renzi et al., 2016).

Common Garden

Plants (**Table 1**) were cultivated at the Experimental Agricultural Station (EEA) of Hilario Ascasubi (Buenos Aires, Argentina; 62°

TABLE 1 | Country, improvement status (cultivar, wild, and naturalized), and name of hairy vetch accessions included in the phenotypic and genotypic studies.

Country of origin	N	Accession/locality† (name)	Evaluation**
Cultivar			
Hungary	1	Kartali	a
	2	Simabuekkoeny	a
Poland	3	Rea	a
	4	Rod MPI	a
	5	Sielecka	a,b,c
Turkey	6	617 81	a,b
	7	Capello	a,b
Bulgaria	8	266 99	a,b,c
Czech Republic	9	Arida	a
	10	HS 1884	a
	11	Modra	a,b
	12	nsl. Dobrenice	a
	13	Troubsko	a,b,c
	14	Viola	a
France	15	Savane	a,b
Germany	16	Ebsdorfer	a
	17	Oregon	a,b
	18	Polyp	a
	19	SAM 21	a
	20	Welta	a
Russia	21	Pridesnjanskaja	a
	22	Stavcanka	a,b,c
	23	Stenskaja 24	a
Serbia	24	Sarajevo	a
Wild			
Czech Republic	1	Pouzdrany†	a
	2	Zavojno jezero I†	a
Serbia	3	Petrovo Selo†	a,b
	4	Radenka II †	a,b,c
	5	Senokos†	a
Naturalized			
Argentina	1	Algarrobo†	a,b
	2	Bordenave†	a
	3	C. Dorrego†	b
	4	C. Pringles†	b
	5	C. Suarez†	a
	6	Chelforó†	b
	7	Doblas†	a
	8	E. Martini†	b
	9	E. S. Pablo†	a
	10	G. Acha†	b
	11	Guaminí†	a,b
	12	Guatraché	b
	13	I. Rico†	b
	14	Médanos†	a
	15	Ombucta†	a
	16	P. Carretas†	b
	17	P. Luro†	b
	18	Pasman†	a
	19	Pigue†	a,b
	20	Rancul†	a
	21	Rivera†	b
	22	S. Luis†	b
	23	S. Pedro†	a,b,c
	24	Saldungaray†	a,b,c
	25	T. Arroyos†	b
	26	T. Origone†	a
	27	T. Picos†	a,b
	28	Tratayen†	b
	29	Winifreda†	a
Cultivar*			
Argentina	1	Algarrobo†	a
	2	Ascasubi INTA	a,b,c
	3	Bordenave†	a
	4	Buratovich†	a
	5	Carhué†	a
	6	G. Chavez†	a
	7	Guatraché†	a
	8	M. Juárez†	a
	9	Oriente†	a
	10	Pergamino†	a
	11	T. Origone†	a
	12	Tolse F.C.A	a,b,c

Include VNS, variety not stated.

***Phenotypic common garden evaluation in 2014 (a) and 2016 (b), and genotypic analysis with SSR (c).*

†Correspond to name of a locality.

37′W, 39°23′S) during 2014 and 2016 growing seasons. The predominant climate in this location is semiarid-temperate with 489 mm mean annual precipitation and 14.8°C mean annual temperature (EEA H. Ascasubi, 1966–2018). The soil was an entic haplustoll, sandy loam, slightly alkaline (pH ≈ 7.5), high in phosphorus (P) content (≈ 22 ppm P Bray & Kurtz) and low organic matter content (≈ 1.6%) at 20 cm (Renzi et al., 2016). Weather data from each year were registered at the nearby meteorological station (less than 500 m) (http://inta.gob.ar/documentos/informes-meteorologicos).

The accessions were arranged in row plots in a randomized complete block design, with three replications. Each experimental unit consisted of a row of 2.50 meters sown with 30 seeds on 10th April 2014 and 27th April 2016. Original collected seeds were used in 2014 and 2016 experiments. Seeds were inoculated with commercial inoculum (*Rhizobium leguminosarum* bv *viciae*) immediately before sowing (Deaker et al., 2004).

To determine the number of days from sowing to 50% flowering, the growing stage was recorded twice a week. After 50% flowering, leaf length (mm), leaflets per leaf, and the number of flowers per raceme were measured on 10 randomly selected individual stems in each plot. For foliar observations, the leaf of the third upper node of the stem and the basal leaflet of the leaf were chosen. Above-ground total dry matter was measured at end of winter (mid-September) and late spring (mid-December) by cutting plant shoots at ground level in a 0.50 m row in each plot. The biomass of each accession, in each year, was expressed in relation to the average biomass per year. Maturity was defined by the presence of 75% ripe pods, approximately at the beginning of summer. Seeds from mature pods were immediately harvested and threshed by hand, on 21st December 2014 and 26th December 2016, for physical (PY) testing. Moisture content at harvest was ≤ 14% (Renzi and Cantamutto, 2013).

Statistical Analysis

Analyses of variance (ANOVA) considering a randomized complete block, between improvement status (EU Cultivar, Wild, AR Cultivar, Naturalized; **Table 1**) and between accessions for each improvement status, were performed using InfoStat software. Accessions and improvement status means were compared by Fisher's least significant difference test. Correlations between quantitative traits were calculated using Pearson's correlation coefficient. Canonical variate analysis (CVA) was performed with all phenotypic traits based on Euclidean distance through InfoStat software.

Physical Dormancy Testing

After the harvest from the common garden experiment, seeds were cleaned and seed weight was estimated, in a sample of 100 seeds in 2014 (n = 1) or 50 seeds in 2016 (n = 3). PY seeds (i.e., "hard" or impermeable) were determined by an imbibition test performed at 20 ± 2°C for 3 days (Baskin and Baskin, 2014). Intact non-germinated seeds of each replication were placed on moist filter paper in Petri dishes and watered daily with tap water. Imbibed seeds showed a visible change in its size/volume ratio, and were easily distinguished from non-imbibed ones (Renzi et al., 2014). Seed viability of non-germinated seeds was assessed by slicing longitudinally with razor and immersion in a 0.5% (wt/vol) tetrazolium chloride (2,3,5-triphenyltetrazolium chloride) (Sigma-Aldrich) solution for 24 h at 30°C in the dark (ISTA, 2019). Seeds with pink or red-stained embryos were considered viable. The total number of viable seeds consisted of germinated plus stained one (Renzi et al., 2016).

For all accessions, PY break dynamics as a function of storage time (38 days) and temperature (20°C) under wet conditions were analyzed using the area under the curve (AUC) calculated by GraphPad Prism Software (GraphPad, San Diego, California, USA). Where AUC = 1 indicates seed without PY (Initial non-PY seeds = 100%) and AUC = 0 indicate PY seed (final non-PY seeds = 0%) (Renzi et al., 2016). Accessions were grouped by improvement status and further compared by Fisher's least significant difference test using InfoStat software.

Relationships Among Phenotypic Traits, Geography, and Environment

To assess relationships between phenotypic traits in 16 and 20 naturalized populations of Argentina (2014 and 2016) and both geographic and environmental distances, six matrices were prepared and examined using the Mantel test (Smouse et al., 1986). The physical distance between naturalized populations was estimated using geographic distance (GGD) for latitude (x)/ longitude (y) values: $GGD = \sqrt{(xi - xj)^2 + (yi - yj)^2}$ (Peakall and Smouse, 2006; Garayalde et al., 2011). The geographic matrix contained pairwise geographical distances while phenotypic distance was calculated as Euclidean distances between populations. All environmental variables were standardized and were calculated using soil and climatic variables ("environmental"). The significance of the normalized Mantel coefficient was calculated using a two-tailed Monte Carlo permutation test with 1,000 permutations using InfoStat software.

Genotypic Characterization

Plant Sampling and DNA Isolation

Three representative naturalized populations (n = 3) were contrasted with AR cultivar (n = 2), EU cultivar (n=4), and wild accession (n = 1) (**Table 1**). Seeds were sown in a common garden at the Experimental Agricultural Station (EEA) of Hilario Ascasubi. Leaf material from 10 randomly selected plants from each accession was collected at the vegetative stage (August 2016). Genomic DNA was extracted using a modified cetrimonium bromide (CTAB) method (Hoisington et al., 1994) from leaf tissue dried on silica gel.

Microsatellite Markers

The five most polymorphic SSR markers were chosen (**Table S2**) from set of 36 simple sequence repeat (SSR) markers developed by Raveendar et al. (2015) for common vetch (*Vicia sativa* subsp. *sativa*) (Chung et al., 2013) and being applicable for HV genotyping analysis. Amplification reactions were performed in 17 µl volumes containing: 0.25 U of Taq DNA polymerase (Invitrogen), 1 mM MgCl2, 1.1 pmol of primers, 1 mM of each deoxynucleoside triphosphate (dNTP), and 30 ng of genomic DNA template. The optimum annealing temperature was determined for each primer set: KF008505 (55°C), KF008507 (59°C), KF008512 (59°C), KF008526 (59°C), and KF008536 (60° C). Amplifications were initially checked on 1.5% agarose gels. PCR products were analyzed on 6% denaturing polyacrylamide gel, 1×TBE electrophoresis buffer at 60 W for 75 min and the bands were visualized by silver staining and scanned (modified from Tang et al., 2003 and Garayalde et al., 2011). The size of each SSR allele was estimated using a 100 bp molecular weight marker. Each DNA fragment was considered as an allele of a single co-dominant locus.

Genetic Data Scoring and Genetic Variability

The amplified SSR loci were scored for 10 accessions. Homozygous and heterozygous genotypes were inferred from the band patterns and allele frequencies (*pi*) calculated accordingly. The absence of band (null allele) was scored as missing data. Mean expected heterozygosity values (*He*) and the percentage of polymorphic loci (*P%*) were calculated: $HE = 1 - \sum pi^2$; $P\,\% = \frac{Lp}{LT} \times 100$, where *pi* is the frequency of the *i*th allele, *Lp* is the number of polymorphic loci, and *LT* is the total number of loci. Hardy–Weinberg equilibrium was tested using chi-squared test $X^2 = x^2 = \sum_{i=1}^{k} \frac{(Oi - Ei)^2}{Ei}$, where O_i is the observed number of individuals of the *i*th genotype, E_i is the expected number under equilibrium hypothesis, and *K* is the total number of genotypes. Degrees of freedom for the chi-squared test were calculated as d.f. = [Na(Na–1)]/2, where Na is the number of alleles at the locus (following Garayalde et al., 2011).

The calculation of genetic distances (GD) followed the method of Peakall et al. (1995) and Smouse and Peakall (1999). For the analysis of a SSR single-locus, the first step involves the calculation of the vector by additive genotype

scoring convention per individuals. Subsequently, the squared distance (d^2) between any two genotypes is one-half the Euclidean distance between their respective pair of vectors as follows: $d_{ij}^2 = \frac{1}{2}\sum_{k=1}^{k}(y_{ik} - y_{jk})^2$, where i and j are the genotypes and k is the scoring character. Squared distances range from 0, when individuals share the same alleles, to 4 when individuals are homozygous for different alleles. Genetic distance matrices for each locus were summed across loci under the assumption of independence. At population level, a $Ø_{PT}$ (analogue of F_{ST}) obtained from analysis of molecular variance (AMOVA) was used as an estimate of population genetic differentiation with SSR markers. Principal coordinate analyses (PCO) were performed on GD matrices. The correlation between genetic distance and phenotypic matrix was analyzed by the Mantel test (see Garayalde et al., 2011).

Analysis of Molecular Variance

The individual pairwise GD matrices were subjected to AMOVA. Total genetic variation was partitioned into three levels: within and between accessions and between origin (AR and EU). Variation was summarized both as the proportion of the total variance and as Ø-statistics (Excoffier et al., 1992). Genetic variability measures, distance metrics, PCO analysis, correlation analysis, and AMOVA were analyzed using GenAlEx 6 (Peakall and Smouse, 2006).

RESULTS

Ecological Niche of Naturalized Populations

Naturalized populations of hairy vetch were found in the three monitored regions (**Figure 1A**), corresponding to Pampa, Espinal, and Shrubs of Plateau and Plains. In the central temperate region of Argentina, annual rainfall varies from semi-arid to arid conditions with only 200 mm, to sub-humid and humid environments with approximately 1,000 mm. The vegetation changes in all three regions, from arid steppes in the west (Shrubs of Plateau and Plains) to grass steppes without woody species in the east (Pampa). The Espinal is an intermediate savannah, with grasses and scarce xeric trees, mainly of the *Prosopis* genus.

The proposed niche modeling explained most of the variation in HV geographical distribution. The area under the receiver operating curve (AUC) score of MaxEnt models, both training and test AUC values, were 0.957 and 0.956, respectively, indicating that most climatically suitable areas predicted by MaxEnt were highly correlated with the occurrence of natural HV populations. The distribution was significantly affected by precipitation amount of the driest quarter (BIO17), max temperature of the warmest month (BIO5), annual mean temperature (BIO1), and clay content in the soil surface (t_clay, **Table 2**). The main suitable habitats for HV are distributed in the southeast of Espinal and southwest of the Pampa region, characterized by sub-humid and semiarid temperate climates, with warm-dry summers and cold-wet winters (**Figure 1B**).

Plant communities associated with HV comprised of 63 species (**Table S1**). Most frequent species were cosmopolitan weeds, including *Avena fatua*, *Cynodon dactylon, Sorghum halepense* (Poaceae), *Carduus* sp., and *Centaurea solstitialis* (Asteraceae) and natural communities of the perennial pasture *Festuca arundinacea* (Poaceae). **Table 3** presents the life cycles and origins of the 20 species frequently associated with naturalized HV populations in the explored region, considered as the dominant community species. Exotic species represent 85% of the co-occurring vegetation.

Plant Growth and Phenotypic Variability

Registered rainfall in EEA Hilario Ascasubi during HV growing season (from April to December) was 50% higher in 2014 (498 mm) and 21% lower (264 mm) in 2016, compared to historical long-term means (331 mm). Mean daily air temperature values were slightly higher in 2014 (13.5°C) compared to 2016 (12.8°C) growing season. All AR accessions performed well, except for Tolse F.C.A in 2014. However, nine out of the twenty-four EU

TABLE 2 | Contribution (%) of the bioclimatic and soil variables in the MaxEnt models, and suitable habitats of naturalized hairy vetch populations from Argentina.

Environmental variables	Contribution (%)	Suitable habitats Mean (range)	Unit
Precipitation of driest quarter	27.2	25.5 (16.6–39.3)	mm month^{-1}
Max temperature of warmest month	16.2	25.9 (24.1–27.7)	°C month^{-1}
Annual mean temperature	15.2	12.3 (11.0–13.6)	°C month^{-1}
Clay content in the surface soil	11.5	28.3 (8.0–38.0)	%
Annual precipitation	5.0	632 (4,007–812)	mm year^{-1}
Soil pH	4.7	6.6 (6.1–9.7)	
Sand content in the surface soil	4.6	37.1 (23.0–84.0)	%
Min temperature of coldest month	4.3	1.5 (0.3–2.5)	°C month^{-1}
Mean temperature of coldest quarter	3.6	6.6 (5.7–7.3)	°C month^{-1}
Precipitation of driest month	2.9	20.5 (9.0–27.0)	mm month^{-1}
Silt content in the surface soil	2.6	34.6 (8.0–49.0)	%
Mean temperature of warmest quarter	0.9	18.3 (16.8–19.8)	°C month^{-1}
Precipitation of wettest quarter	0.8	71.7 (44.6–91.7)	mm month^{-1}
Precipitation of wettest month	0.4	84.4 (51.0–103.0)	mm month^{-1}
Soil bulk density	0.1	1.3 (1.2–1.6)	kg dm^{-3}

TABLE 3 | Dominant community species associated with hairy vetch populations in central Argentina.

Species	Family	Cycle	Origin	Frequency
Avena fatua L.	Poaceae	A	E	0.50
Carduus nutans L. and *C. acanthoides* L.	Asteraceae	A	E	0.45
Cynodon dactylon (L.) Pers.	Poaceae	P	E	0.42
Centaurea solstitialis L.	Asteraceae	A	E	0.40
Sorghum halepense (L.) Pers.	Poaceae	P	E	0.38
Festuca arundinacea Schreb.	Poaceae	P	E	0.38
Stipa ambigua Speg. *and A. caudate* Trin.	Asteraceae	P	Na	0.37
Taraxacum campylodes G.E. Haglund	Asteraceae	P	E	0.33
Diplotaxis tenuifolia (L.) DC.	Brassicaceae	P	E	0.29
Melilotus albus Medik.	Fabaceae	A	E	0.28
Medicago lupulina L.	Fabaceae	A-P	E	0.28
Ammi majus L.	Umbelliferae	A	E	0.24
Lolium multiflorum Lam.	Poaceae	A	E	0.22
Dactylis glomerata L.	Poaceae	P	E	0.22
Rapistrum rugosum (L.) All.	Brassicaceae	A	E	0.22
Eragrostis curvula (Schrad.) Nees	Poaceae	P	E	0.20
Erigeron bonariensis L.	Asteraceae	A	Na	0.20
Plantago lanceolata L.	Plantaginaceae	P	E	0.19
Medicago sativa L.	Fabaceae	P	E	0.15
Bromus catharticus Vahl	Poaceae	B	Na	0.14

Life cycle: A, annual; P, perennial; B, biannual. Origin: Na, native; E, exotic.

tested cultivars (nr. 1, 7, 9, 10, 12, 13, 15, 16 and 19; **Table 1**) and two out of the five wild populations (1 and 2) did not produce pods during 2014.

Using canonical discriminant analysis with the phenotypic traits, the accessions were grouped in three clusters (**Figure 2**). These corresponded to improvement status and origin. Cluster 1 consisted predominantly of wild populations, cluster 2 consisted accessions of AR origin, and in cluster 3 were EU cultivars.

AR accessions showed higher winter biomass production compared to EU cultivars and wild genotypes (**Table 4**).

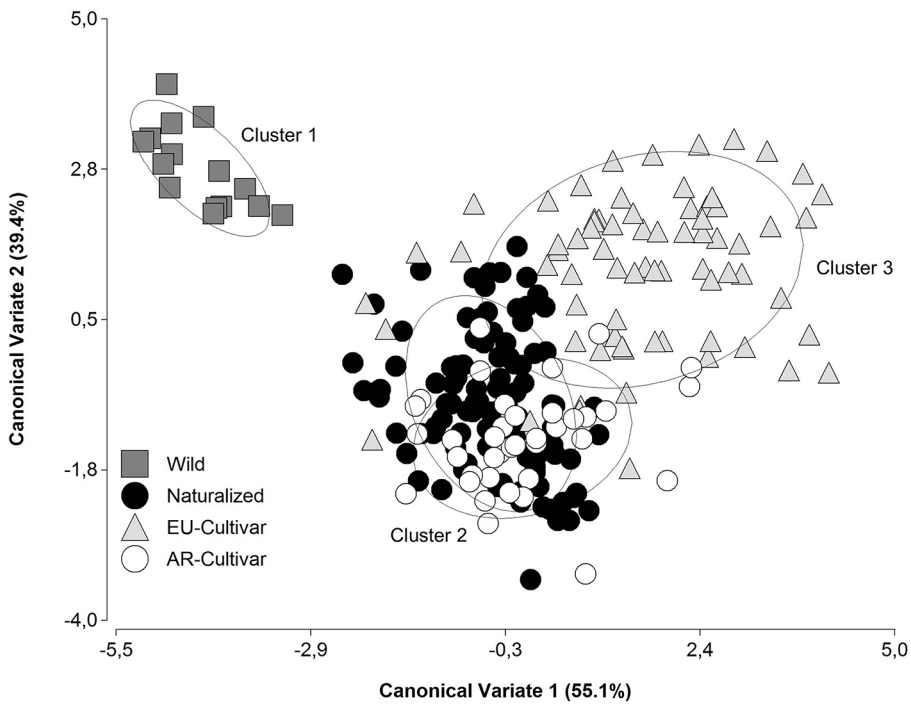


FIGURE 2 | Scatterplot of the four improvement status groupings on the two canonical discriminant functions based on phenotypic traits in a common garden during 2014 and 2016.

TABLE 4 | Phenotypic variability of hairy vetch (HV) for each improvement status and origin (means and range) grown in a common garden during 2014 and 2016 in Experimental Agricultural Station (EEA) Hilario Ascasubi.

Improvement status	Origin	n	Relative biomass†					Days to			Seed	
			Winter	Spring	Leaf length	Leaflet number	Flower number	Flowering	Maturity	Physical dormancy	Viability	Weight
					(mm)	(n)				AUC	(%)	(mg)
Cultivar	EU	24										
Mean			0.69^{B}	0.94^{B}	57	15.3^{B}	26^{B}	195^{AB}	245^{AB}	0.72^{D}	97^{A}	33^{B}
Range			(0.10–1.84)	(0.35–1.59)	(35–83)	(12–19)	(10–40)	(145–203)	(204–255)	(0.15–0.99)	(84–100)	(21–42)
Within status ($p < 0.05$)			**	**	**	*	**	**	**	**	**	**
Cultivar	ARG	12										
Mean			1.47^{D}	0.99^{BC}	59	14.7^{A}	23^{A}	197^{BC}	248^{B}	0.48^{C}	98^{AB}	32^{B}
Range			(0.72–2.39)	(0.37–1.63)	(22–81)	(10–19)	(13–28)	(169–202)	(221–255)	(0.11–0.88)	(88–100)	(27–41)
Within status ($p < 0.05$)			NS	**	**	*	NS	**	**	**	*	**
Naturalized	ARG	30										
Mean			1.27^{C}	1.06^{C}	58	15.2^{B}	25^{B}	193^{A}	241^{A}	0.39^{B}	99^{B}	31^{B}
Range			(0.26–2.35)	(0.48–1.89)	(32–99)	(13–19)	(18–32)	(178–202)	(221–260)	(0.08–0.66)	(88–100)	(22–40)
Within status ($p < 0.05$)			**	**	**	**	NS	**	**	**	*	**
Wild	EU	5										
Mean			0.20^{A}	0.77^{A}	51	17.2^{C}	21^{A}	200^{C}	246^{AB}	0.13^{A}	96^{A}	15^{A}
Range			(0.01–0.59)	(0.23–1.07)	(37–77)	(15–20)	(15–33)	(188–213)	(233–255)	(0.00–0.23)	(76–100)	(10–21)
Within status ($p < 0.05$)			*	**	*	**	NS	**	NS	NS	**	*
Between status ($p < 0.05$)			**	**	NS	**	**	*	**	**	**	**
LSD			0.15	0.08		0.5	2	4	4	0.06	1	2
Status * Year			NS	NS		**	NS	NS	**	NS	NS	NS

For mean values, different letters indicate significant differences among status (Fisher's LSD test, $\alpha = 0.05$).

*'**' Indicates significance at $p < 0.01$ and '*' at $p < 0.05$.*

†Relative to mean biomass per year (winter 222 and 168 g m^{-1} and spring 1,174 and 887 g m^{-1} in 2014 and 2016, respectively).

NS, not significant.

Biomass in spring was less variable between improvement status. Over the total data set, HV winter biomass was negatively correlated to the number of leaflets per leaf (r = -0.20; P < 0.01), and the spring biomass was positively correlated with the leaf length (r = 0.74; P < 0.001). There was strong positive correlation between winter and spring biomass (r = 0.40; P < 0.001). Variations in the days to flowering among improvement status were narrow (≤ 7 days) but statistically significant within status (**Table 4**).

Seed viability was over 80% in all cases. The area under the curve (AUC), showing the following dormancy gradient rank: EU cultivars < AR cultivars < naturalized < wild genotype. No significant interaction between improvement status x year was found in the AUC (**Table 4**). Wild genotypes had a smaller seed weight (**Table 4**).

Figure 3 shows the relationship between the main evaluated traits (winter biomass and seed dormancy), in order to identify the more suitable genotypes for breeding programs. Accessions number 9, 12, 19, 21, 26 of naturalized populations and 2, 5, 10, 12 of AR cultivars showed better potential for winter biomass with high PY. While the genotypes number 3, 6, 8 for AR cultivars, 2 for naturalized populations and 6 for EU cultivar showed better potential for both winter biomass and low PY.

Genotypic Characterization

The five selected polymorphic SSR loci produced altogether 24 alleles in 100 tested individuals. The mean number of alleles per locus (N) was 4.8 ± 0.9 ranging from 3 to 8 among the five loci. The mean expected heterozygosity (He) was 0.667 ± 0.03.

Differences in variability were observed between improvement status, being higher in EU cultivars (He = 0.64 ± 0.03; N = 4.4 ± 1.0) and naturalized populations (He = 0.63 ± 0.05; N = 4.6 ± 0.9) compared to AR cultivars (He = 0.57 ± 0.06; N = 3.8 ± 0.9) and wild (He = 0.30 ± 0.09; N = 2.2 ± 0.4) genotypes. The lower value of variability in wild ones might be due to the smaller number of analyzed individuals (n = 10). Two private alleles were found only in naturalized populations. Equilibrium tests were significant in the 68% of cases, indicating non-random mating within populations (**Table S3**). AMOVA showed a significant differentiation between improvement status, which explained around 19.1% of the variance. Naturalized populations differed from the wild accession (a genetic differentiation of 46.7%) and also from EU (8.3%) and AR (14.3%) cultivars. Lower differentiation was found between AR and EU cultivars (4.3%) (**Figure 4**).

AMOVA for the total marker data set is shown in **Table 5**. Genetic diversity was high within accessions. Between-accession estimated variance was significant and around 25% of the total variation. A small but significant portion of variance (3%) was found attributable to differences between AR and EU genotypes.

Relationships Among Distance Measures

Environmental distance matrices were significantly correlated with geographic distance matrices (Mantel test; environment: r_{2014} = 0.61, P < 0.01; r_{2016} = 0.78, P < 0.01), suggesting that environmental (climatic + soil) conditions diverge with increasing geographic distance. However, geographic (r_{2014} = 0.19, P = 0.16; r_{2016} = 0.10, P = 0.20) nor environmental (r_{2014} =

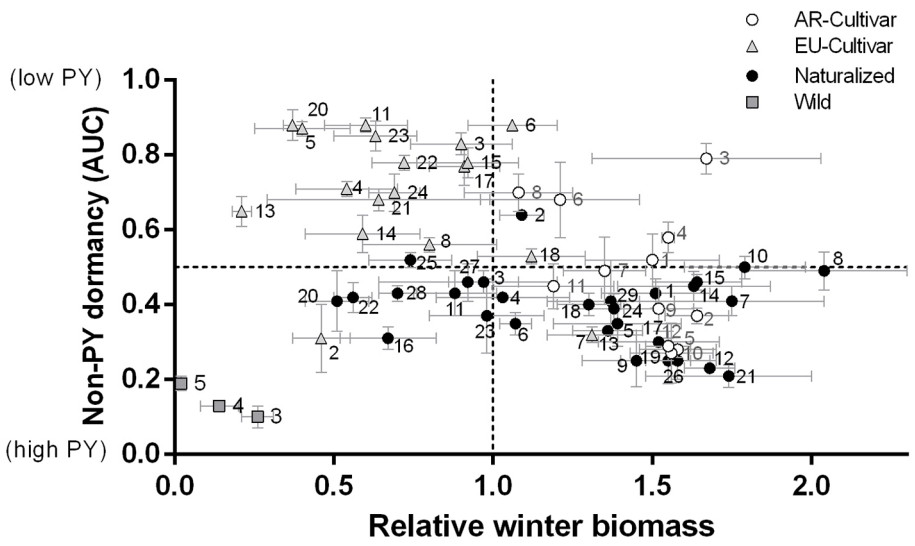


FIGURE 3 | Relationship between relative winter biomass and physical (PY) dormancy [area under the curve (AUC)] for each genotype (mean and standard error) evaluated in common garden during 2014 and 2016. For number references see **Table 1**.

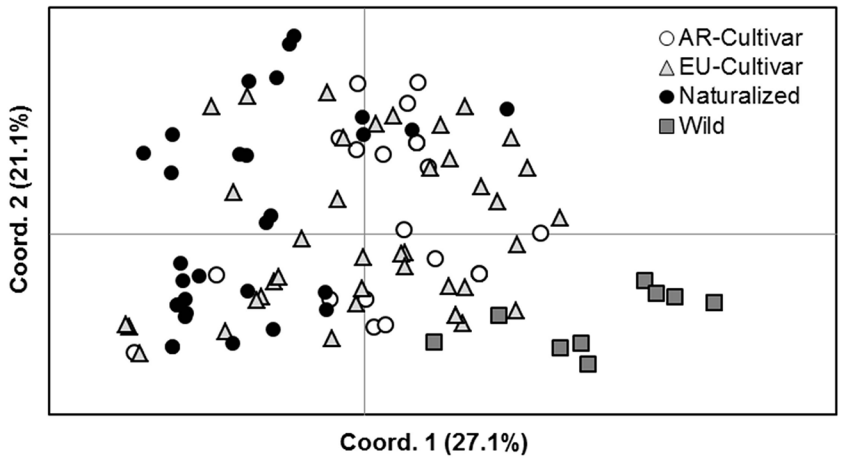


FIGURE 4 | Principal coordinate analyses (PCO) plots based on the individual simple sequence repeat (SSR) distance matrix.

TABLE 5 | Analysis of molecular variance (AMOVA) and sources of variation for hairy vetch accessions.

Source of variation	df	SS*	MS**	% variation	*p-value*
Among origin	1	24,097	24,097	3	<0,02
Among accessions	8	136,098	17,012	25	<0,01
Within accession	90	338,600	3,762	72	<0,01
Total	99	498,795		100	

**SS, sum of squares.*
***MS, mean sum of squares.*

-0.04, P = 0.54; r_{2016} = 0.15, P = 0.10) distances were not significantly correlated with phenotypic distance.

Genetic and phenotypic distance matrices of naturalized populations showed a positive but non-significant statistical correlation (Mantel test; r = 0.19, P = 0.21).

DISCUSSION

Ecological Niche

Argentinian hairy vetch populations occur on a transitional zone between two defined ecological regions, Pampas with sub-humid climate and Espinal with semiarid conditions (**Figure 1**). HV was used as a forage species in Buenos Aires province before 1900 (Manganaro, 1919). Thereafter, it probably escaped from cultivation (Aarssen et al., 1986) and became naturalized. As HV natural seed dispersal potential is very limited (Jannink et al., 1997), seed spillage due to handling and transportation was probably the most important distribution method. Human-mediated dispersal is the most likely explanation of establishment and subsequent naturalization of HV into further suitable habitats (Horvitz et al., 2017; Pascher et al., 2017).

After its introduction in Argentina, HV spread over the areas which met the appropriate conditions, following a patchy distribution. HV showed adaptation in broad geographic (33–41°S, latitude, 60–66°W longitude) and climatic range (400–800 mm rainfall; 11–13.6°C annual mean temperature; **Table 2**). Duke (1981) mentioned that HV is well adapted to a greater range of annual mean temperatures between 4.3–21°C. In this study, low rainfall and warm temperatures during summer months explained HV potential distribution of natural populations (**Table 2**). HV was generally associated with neutral-alkaline (range 6.1–9.7 pH), sandy or sandy loam soils. However, it could occur on most soil types with sufficient drainage capacity (Duke, 1981). Clark (2007) stated that HV preferred neutral (pH 6.0–7.0) soils with tolerance to alkalinity (Duke, 1981). Conversely, low pH (< 6.2) can decrease the rate growth, nodulation, and nitrogen fixation (Aarssen et al., 1986).

Fitness for the Ecological Niche

The ability of HV to produce PY+PD dormant seeds and their subsequent germination and emergence are important factors that influence natural population dynamics and persistence (Kimball et al., 2010). During the period of seed formation, dry and warm conditions could shorten the species life cycle due to rapid thermal-time accumulation (Petraityte et al., 2007) as well as favor a decrease on seed moisture content. In HV, the acquisition of PY is initiated only when the moisture content

of the seeds is ≤14% (Hyde, 1954). Furthermore, HV is a cross-pollinated species where bees play an important role (Zhang and Mosjidis, 1995; Renzi et al., 2017), thus dry and warm weather is favorable for the activity of pollinating insects (Petraityte et al., 2007; Al-Ghzawi et al., 2009).

In the humid central region of Argentina, the spread of naturalized populations of HV would be limited by a negative combination of two main factors. First, the abundant rainfall stimulates the virulence of foliar fungal diseases (e.g., *Ramularia sphaeroidea* Sacc. and *Ascochyta viciae* Lib.), which reduce photosynthetic leaf area thus limiting seed formation (Petraityte et al., 2007; Renzi and Cantamutto, 2013). In addition, high humidity conditions enhance HV indeterminate growth, non-uniform maturity and extended growing season favoring a biennial behavior (Duke, 1981). These consequently limit the seed formation, drying, and acquisition of PY, and can cause a sharp reduction of seed bank persistence and consequent natural regeneration, similar to observations of Toser and Ooi (2014) in *Acacia saligna*.

After the seed dispersal, warm summer temperatures are required for seed dormancy release (Renzi et al., 2014; Renzi et al., 2016). Seed dormancy acquisition, release, and seedling emergence requirements are important fitness traits determining ecological niches of HV (**Figure 1B**). These adaptive traits seem to have evolved in Mediterranean-like environments, where hot and dry summer conditions regulate seed dormancy alleviation while cool and wet winters contribute to enhance vegetative growth providing safe-sites for seedling recruitment (Van Assche and Vandelook, 2010; Picciau et al., 2019).

Phenotypic Variability

Two cultivars of woolly-pod vetch (Capello and Tolse F.C.A) were found among AR and EU cultivars. This subspecies (*V. v.* ssp *varia*) is characterized by shorter leaves (29.8 ± 6 *vs*. 47.8 ± 9.3 mm in HV, $P < 0.01$), fewer leaflets (12.3 ± 1.9 *vs*. 15.2 ± 1.1, $P < 0.01$); and flowers (20.1 ± 7.7 *vs*. 26.9 ± 3.4, $P < 0.01$), early flowering and maturity (more than 2 week before, $P < 0.01$), higher winter (**Figure 3**) but lower spring biomass in relation to HV (≈ 60%, $P < 0.01$). On the other hand, Jannink et al. (1997) found that accessions of HV were more winter-hardy than woolly-pod vetch, and late flowering may be positively genetically correlated with winter hardiness (Maul et al., 2011). The susceptibility of HV to low-temperature increases with more advanced phenological stages (Brandsaeter et al., 2002) while slow growth during winter can be an adaptation attribute to avoid frost damage (Loi et al., 1993). Prompt maturity is a desirable trait for earlier biomass production as well as N accumulation in regions with a shorter growing season. Also, flowering timing can greatly influence the capacity for weed-suppression as a cover crop (Mischler et al., 2010; Maul et al., 2011). Therefore, woolly-pod vetch could be a source of desirable genes for breeding program seeking early flowering cultivars for cover crop usage in mild winter zones. Notably, most of the genotypes characterized as early flowering by Maul et al. (2011) corresponded to woolly-pod vetch (https://npgsweb.ars-grin.gov/gringlobal/search.aspx).

Quantitative traits evaluated among accessions resembled a continuous probability distribution, although the difference between improvement status was statistically significant (**Table 4**). AR accessions showed higher winter biomass accumulation compared to EU, probably due to a greater adaptation to Argentinian ecological conditions.

Seed dormancy is largely genetically determined but also depends on the environmental conditions experienced by the mother plant (maternal effect) and the subsequent degree of seed dehydration (Hudson et al., 2015; Finch-Savage and Footitt, 2017). HV seeds were collected from mature pods with seed moisture content less than 14% (determined as a critical value for PY acquisition, Hyde, 1954), and the environmental effect between years was not detected on AUC of PY (**Table 4**). PY was variable among genotypes and could act mainly as an adaptive trait (Hudson et al., 2015; Long et al., 2015). EU cultivars had lower PY values compared to AR cultivars and naturalized accessions. PY is a highly heritable trait (Hudson et al., 2015) thus it could be useful germplasm for both breeding and artificial (or natural) selection for higher or lower levels of dormancy (Lacerda et al., 2007). It is probable that observed differences between accessions could be explained by genetic adaptations of HV to the local environment (Baskin and Baskin, 2014) as shown in pea (Hradilová et al., 2019) or by selection for improved genotypes (Fuller and Allaby, 2009; Kluyver et al., 2013; Renzi et al., 2016).

Observed large variability within available germplasm could be used by breeders to select parental accessions for HV improvement breeding program that maximizes the winter-spring biomass with low PY for cover crops (Wilke and Snapp, 2008; Wayman et al., 2016), or with high PY for ´ley farming´ systems (Loi et al., 2005; Renzi et al., 2017; Renzi et al., 2018). No significant correspondence was found between the geographic distance matrix and the phenotypic distance matrix ($P > 0.15$) among the naturalized AR populations and these results differ from *Medicago polymorpha* L., in which a correspondence between collection site and phenotypic traits (Loi et al., 1993; Helliwell et al., 2018) was observed.

Among the measured traits there was a significant correlation between winter and spring biomass. The latter was also highly correlated with the length of the leaf as described in *Vicia sativa* (De La Rosa et al., 2002). Leaf size is related to the photosynthetic rates which in turn affect growth and could be potentially maximized with water availability (Carlson et al., 2016). Thus, leaf size would be an indirect selection trait for biomass that improves breeding program efficiency.

Genotypic Variability

There is scarce information on genetic diversity of HV and therefore the knowledge of genetic variability is useful for other studies. Our data showed low genetic differentiation between AR and EU cultivars, but strong differentiation between wild EU and

naturalized AR populations. Similarly to other outcrossing species, variation between accessions was small in comparison with the variance found within populations (**Table 5**), which is expected for an obligate cross-pollinated species (Hamrick and Godt, 1989). Maul et al. (2011) reported similar results, with 93% of genetic diversity within populations.

The Hardy-Weinberg equilibrium was not stable within populations (**Table S3**) and this can be attributed to mutations, natural selection, non-random mating, genetic drift, and gene flow. As mentioned above, a lack of geographical signature in the pattern of population variation, as occurs with other allogamous species (Garayalde et al., 2011), can also be explained by human activities on seed dispersal and genetic drift (Knapp and Rice, 1998). Cultivated populations are open-pollinated and highly heterogeneous and would be subject to natural selection and genetic drift throughout the cycles (Wiering et al., 2018). These results are consistent with observed in the phenotypic traits.

CONCLUSIONS

This study increases the understanding of the value of naturalized hairy vetch populations in agroecosystems of Argentina. Naturalized populations showed good soil adaptation in disturbed areas and neutral response to alkaline soil niches from central Argentina. Low rainfall and warm temperatures during pre- and post-dispersal seem to explain and regulate the potential distribution of HV populations. Within this ecological context, dry and warm climate may be considered as favorable environmental conditions to increase seed dormancy and timing of germination-triggering. Considering HV genetic variability and agro-ecological adaptation, naturalized populations could be considered as a source of potential adaptive traits for breeding. The AR germplasm constitutes an important reservoir of genes for high winter and spring biomass production. On the other hand, high levels of innate seed dormancy of HV accessions from Argentina reduce its possible use as a cover crop. In this sense, dedicated crosses with more domesticated EU cultivars will serve to reduce the seed dormancy.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/ **Supplementary Material**.

AUTHOR CONTRIBUTIONS

JR and MC conceived the topic. JR and LZ performed the experiments. JR and AG analyzed all statistical data. JR, GC, and AP wrote the manuscript. PS and MC revised the manuscript.

FUNDING

This work was supported by the Instituto Nacional de Tecnología Agropecuaria (PE-E6-I146). Agencia Nacional de Promoción Científica y Tecnológica. MINCyT (PICT-2017-0473 and PICT-2016-1575) and Universidad Nacional del Sur (PGI 24/A223 and PGI 24/A225).

ACKNOWLEDGMENTS

The authors would like to acknowledge V. Holubec of the Department of Gene Bank of the Crop Research Institute (CRI), for providing the cultivars of hairy vetch used in this study, and T. Vymyslický of the Research Institute for Fodder Crops, for performing taxonomic identification of the Serbian accessions. We would like to thank the personnel at CRF-INIA, in particular to M.T. Marcos Prado for help in the niche analysis. Appreciation is extended to O. Reinoso, J. Castillo and M. Bruna, agricultural specialists at the EEA Hilario Ascasubi, for assistance in establishing the field experiments.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.00189/full#supplementary-material

REFERENCES

Aapresid (2018). Cultivos de cobertura en Argentina. Qué se está haciendo y qué falta? https://www.aapresid.org.ar/rem/wp-content/uploads/sites/3/2018/03/Analisis-encuesta-sobre-CC-web.pdf.

Aarssen, L. W., Hall, I. V., and Jensen, K. I. N. (1986). The biology of Canadian weed: *Vicia angustifolia* L., *V. cracca* L., *V. sativa* L., *V. tetrasperma* (L.) Schreb. and *V. villosa* Roth. *Can. J. Plant Sci.* 66, 711–737. doi: 10.4141/cjps86-092

Ackroyd, V. J., Cavigelli, M. A., Spargo, J. T., Davis, B., Garst, G., and Mirsky, S. B. (2019). Legume cover crops reduce poultry litter application requirements in organic systems. *Agron. J.* 111, 1–9. doi: 10.2134/agronj2018.09.0622

Al-Ghzawi, A. A., Samarah, N., Zaitoun, S., and Alqudah, A. (2009). Impact of bee pollinators on seed set and yield of *V. villosa* spp. *dasycarpa* (Leguminosae) grown under semiarid conditions. *Ital. J. Anim. Sci.* 8, 65–74. doi: 10.4081/ijas.2009.65

Baskin, C. C., and Baskin, J. M. (2014). *Seeds: Ecology, biogeography, and evolution of dormancy and germination. 2nd ed.* (San Diego: Academic Press).

Brandsaeter, L. O., Olsmo, A., Tronsmo, A. M., and Fykse, H. (2002). Freezing resistance of winter annual and biennial legumes at different developmental stages. *Crop Sci.* 42, 437–443. doi: 10.2135/cropsci2002.4370

Bryant, J. A., and Hughes, S. G. (2011). "Vicia," in *Wild crop relatives: Genomic and breeding resources legume crops and forage*. Ed. C. Kole (Berlin, Heidelberg: Springer), 273–289.

Burkart, A. (1952). *Las leguminosas Argentinas: Silvestres y cultivadas. 2nd ed.* (Argentina. Acme Agency: Buenos Aires).

Cantamutto, M., Poverene, M., and Peinemann, N. (2008). Multi-scale analysis of two annual Helianthus species naturalization in Argentina. *Agric. Ecosyst. Environ.* 123, 69–74. doi: 10.1016/j.agee.2007.04.005

Carlson, J. E., Adams, C. A., and Holsinger, K. E. (2016). Intraspecific variation in stomatal traits, leaf traits and physiology reflects adaptation along aridity gradients

in a South African shrub. *Ann. Bot.* 117, 195–207. doi: 10.1093/aob/mcv146

Chung, J. W., Kim, T. S., Suresh, S., Lee, S. Y., and Cho, G. T. (2013). Development of 65 novel polymorphic cdna-ssr markers in common vetch (*Vicia sativa* subsp *sativa*) using next generation sequencing. *Molecules* 18, 8376–8392. doi: 10.3390/molecules18078376

Clark, A. (2007). *Managing Cover Crops Profitably. 3rd ed. Handbook series* (College Park, MD: SARE).

De La Rosa, L., Lázaro, A., and Varela, F. (2002). *Utilización en mejora de la variabilidad morfo-agronómica de Vicia sativa L* (Almería, España: Congreso de Mejora Genética de Plantas), 319–324.

Deaker, R., Roughley, R. J., and Kennedy, I. R. (2004). Legume seed inoculation technology – a review. *Soil Biol. Biochem.* 36, 1275–1288. doi: 10.1016/j.soilbio.2004.04.009

Di Rienzo, J. A., Casanoves, F., Balzarini, M. G., Gonzalez, L., Tablada, M., and Robledo, C. W. (2013). *InfoStat version 2013 Grupo InfoStat, FCA* (Argentina: Universidad Nacional de Córdoba).

Duke, J. A. (1981). *Handbook of legumes of world economic importance* (New York and London: Plenum Press), 345 pp.

Excoffier, L., Smouse, P. E., and Quattro, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131, 479–491.

FAO/IIASA/ISSCAS/JRC. (2012). *Harmonized World Soil Database (version 1.2)* (Rome, Italy and IIASA, Laxenburg, Austria: FAO).

Fick, S. E., and Hijmans, R. J. (2017). WorldClim 2: new 1km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* 37, 4302–4315. doi: 10.1002/joc.5086

Finch-Savage, W. E., and Footitt, S. (2017). Seed dormancy cycling and the regulation of dormancy mechanisms to time germination in variable field environments. *J. Exp. Bot.* 68, 843–856. doi: 10.1093/jxb/erw477

Francis, C. M., Enneking, D., and Abd El Moneim, A. (1999). "When and where will vetches have an impact as grain legumes?" in *Linking research and marketing opportunities for pulses in the 21st Century. Proceedings of the Third International Food Legume Research Conference" Adelaide, S. Aust., 1997. Current Plant Science and Biotechnology in Agriculture*. Vol. 34. Ed. R. Knight (Dordrecht, The Netherlands: Kluwer Academic Publishers), 671–683.

Frasier, I., Noellemeyer, E., Amiotti, N., and Quiroga, A. (2017). Vetch-rye biculture is a sustainable alternative for enhanced nitrogen availability and low leaching losses in a no-till cover crop system. *Field Crops Res.* 214, 104–112. doi: 10.1016/j.fcr.2017.08.016

Fuller, D. Q., and Allaby, R. (2009). "Seed Dispersal and Crop Domestication: shattering, germination and seasonality in evolution under cultivation," in *Fruit Development and Seed Dispersal*. Ed. L. Ostergaard (Blackwell: Oxford), 238–295.

Garayalde, A. F., Poverene, M., Cantamutto, M., and Carrera, A. D. (2011). Wild sunflower diversity in Argentina revealed by ISSR and SSR markers: an approach for conservation and breeding programmes. *Ann. Appl. Biol.* 158, 305–317. doi: 10.1111/j.1744-7348.2011.00465.x

Hamrick, J. L., and Godt, M. J. W. (1989). "Allozyme diversity in plant species," in *Plant Population Genetics, Breeding and Genetic Resources*. Eds. A. H. D. Brown, M. T. Clegg, A. L. Kahler and B. S. Weir (Sunderland, MA, USA: Sinauer Associates), 43–63.

Helliwell, E., Faber-Hammond, J., Lopez, Z. C., Garoutte, A., von Wettberg, E., Friesen, M. L., et al. (2018). Rapid establishment of a flowering cline in *Medicago polymorpha* after invasion of North America. *Mol. Ecol.* 27, 4758–4774. doi: 10.1111/mec.14898

Hernandez, P. A., Graham, C. H., Master, L. L., and Albert, D. L. (2006). The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography* 29, 773–785. doi: 10.1111/j.0906-7590.2006.04700.x

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., and Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* 25, 1965–1978. doi: 10.1002/joc.1276

Hoisington, D., Khairallah, M., and González de León, D. (1994). *Laboratory Protocols: CIMMYT Applied Molecular Genetics Laboratory. 2nd edn.* (México, DF, Mexico: CIMMYT).

Horvitz, N., Wang, R., Wan, F.-H., and Nathan, R. (2017). Pervasive human-mediated large-scale invasion: analysis of spread patterns and their underlying mechanisms in 17 of China's worst invasive plants. *J. Ecol.* 105, 85–94. doi: 10.1111/1365-2745.12692

Hradilová, I., Duchoslav, M., Brus, J., Pechanec, V., Hýbl, M., Kopecký, P., et al. (2019). Variation in wild pea (Pisum sativum subsp. elatius) seed dormancy and its relationship to the environment and seed coat traits. *PeerJ* 7, e6263. doi: 10.7717/peerj.6263

Hudson, A. R., Ayre, D. J., and Ooi, M. K. J. (2015). Physical dormancy in a changing climate. *Seed Sci. Res.* 25, 66–81. doi: 10.1017/S0960258514000403

Hyde, E. O. C. (1954). The function of the hilum in some Papilionaceae in relation to the ripening of the seed and the permeability of the testa. *Ann. Bot.* 11, 241–256. doi: 10.1093/oxfordjournals.aob.a083393

International Seed Testing Association (2019). *International Rules for Seed Testing* (Zürich: ISTA), 300. doi: 10.15258/istarules.2019.F

Jacobsen, K. L., Gallagher, R. S., Burnham, M., Bradley, B. B., Larson, Z. M., Walker, C. W., et al. (2010). Mitigation of seed germination impediments in hairy vetch. *Agr. J.* 102, 1346–1351. doi: 10.2134/agronj2010.0002n

Jannink, J. L., Merrick, L. C., Liebman, M., Dyck, E. A., and Corson, S. (1997). Management and winter hardiness of hairy vetch in Maine. *Tech. Bull.* 167, 35. Maine Agricultural and Forest Experiment Station.

Kimball, S., Angert, A. L., Huxman, T. E., and Venable, D. E. (2010). Contemporary climate change in the Sonoran Desert favors cold-adapted species. *Glob Chang Biol.* 16, 1555–1565. doi: 10.1111/j.1365-2486.2009.02106.x

Kluyver, T. A., Charles, M., Jones, G., Rees, M., and Osborne, L. P. (2013). Did greater burial depth increase the seed size of domesticated legumes? *J. Exp. Bot.* 64, 4101–4108. doi: 10.1093/jxb/ert304

Knapp, E. E., and Rice, K. J. (1998). Comparison of isozymes and quantitative traits for evaluating patters of genetic variation in purple needlegrass (*Nassella pulchra*). *Conserv. Biol.* 12, 1031–1041. doi: 10.1046/j.1523-1739.1998.97123.x

Lacerda, D. R., Lemos Filho, J. P., Goulart, M. F., and Ribeiro, R. A. (2007). Seed-dormancy variation in natural populations of two tropical leguminous tree species: Senna multijuga (Caesalpinoideae) and Plathymenia reticulata (Mimosoideae). *Seed Sci. Res.* 14, 127–135. doi: 10.1079/SSR2004162

Lawson, A., Cogger, C., Bary, A., and Fortuna, A.-M. (2015). Influence of seeding ratio, planting date, and termination date on rye-hairy vetch cover crop mixture performance under organic management. *PloS One* 10 (6), e0129597. doi: 10.1371/journal.pone.0129597

Loi, A., Howieson, J. G., Cocks, P. S., and Caredda, S. (1993). The adaptation of *Medicago polimorpha* to a range of edaphic and environmental conditions: effect of temperature on growth, and acidity stress on nodulation and nod gene induction. *Aust. J. Exp. Agr.* 33, 25–30. doi: 10.1071/EA9930025

Loi, A., Howieson, J. G., Nutt, B. J., and Carr, S. J. (2005). A second generation of annual pasture legumes and their potential for for inclusion in Mediterranean-type farming systems. *Aust. J. Exp. Agr.* 45, 289–299. doi: 10.1071/EA03134

Long, R. L., Gorecki, M. J., Renton, M., Scott, J. K., Colville, L., Goggin, D. E., et al. (2015). The ecophysiology of seed persistence: a mechanistic view of the journey to germination or demise. *Biol. Rev.* 90, 31–59. doi: 10.1111/brv.12095

Manganaro, A. (1919). "Leguminosas bonaerenses," in *Anales de la Sociedad Científica Argentina*. Eds. E. Carrete and C. Lizer (Buenos Aires), 77–264.

Marzocca, A. (1994). *Guía descriptiva de Malezas en el Cono Sur* (Buenos Aires, Argentina: INTA).

Maul, J., Mirsky, S., Emche, S., and Devine, T. (2011). Evaluating a germplasm collection of the cover crop hairy vetch for use in sustainable farming systems. *Crop Sci.* 51, 2615–2625. doi: 10.2135/cropsci2010.09.0561

Mirsky, S., Ackroyd, V. J., Cordeau, S., Curran, W. S., Hashemi, M., Reberg-Horton, S. C., et al. (2017). Hairy vetch biomass across the Eastern United States: effects of latitude, seeding rate and date, and termination timing. *Agron. J.* 109, 1–10. doi: 10.2134/agronj2016.09.0556

Mischler, R., Duiker, S. W., Curran, W., and Wilson, D. (2010). Hairy vetch management for no-till organic corn production. *Agron. J.* 102, 355–362. doi: 10.2134/agronj2009.0183

Pascher, K., Hainz-Renetzeder, C., Gollmann, G., and Schneeweiss, G. M. (2017). Spillage of viable seeds of oilseed rape along transportation routes: ecological risk assessment and perspectives on management efforts. *Front. Ecol. Evol.* 5, 104. doi: 10.3389/fevo.2017.00104

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

Peakall, R., Smouse, P. E., and Huff, D. R. (1995). Evolutionary implications of allozyme and RAPD variation in diploid populations of buffalograss Buchloe

dactyloides. *(Nutt. Engelm.). Mol. Ecol.* 4, 135–147. doi: 10.1111/j.1365-294X.1995.tb00203.x

Petraityte, N., Sliesaravicius, A., and Dastikaite, A. (2007). Potential reproduction and real seed productivity of *Vicia villosa* L. *Biologija* 53, 48–51.

Phillips, S. J., Anderson, R. P., and Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecol. Model* 190, 231–259. doi: 10.1016/j.ecolmodel.2005.03.026

Phillips, S. J., Dudík, M., and Schapire, R. E. (2018). Maxent software for modeling species niches and distributions (Version 3.4.1). http://biodiversityinformatics.amnh.org/opensource/maxent/. Accessed on 2018-10-05.

Picciau, R., Pritchard, H. W., Mattana, E., and Bacchetta, G. (2019). Thermal thresholds for seed germination in Mediterranean species are higher in mountain compared with lowland areas. *Seed Sci. Res.* 29, 44–54. doi: 10.1017/S0960258518000399

Raveendar, S., Lee, G. A., Jeon, Y. A., Lee, Y. J., Lee, J. R., Cho, G. T., et al. (2015). Cross-amplification of *Vicia sativa* subsp. *sativa* microsatellites across 22 other *Vicia* species. *Molecules* 20, 1543–1550. doi: 10.3390/molecules20011543

Renzi, J. P., and Cantamutto, M. A. (2013). *Vicias: Bases agronómicas para el manejo en la Región Pampeana (Vicias: Agronomic bases for management in the Pampas)* (Buenos Aires, Argentina: Ediciones INTA), 299.

Renzi, J. P., Chantre, G. R., and Cantamutto, M. A. (2014). Development of a thermal-time model for combinational dormancy release of hairy vetch (*Vicia villosa* ssp. *villosa*). *Crop Pasture Sci.* 65, 470–478. doi: 10.1071/CP13430

Renzi, J. P., Chantre, G. R., and Cantamutto, M. A. (2016). Effect of water availability and seed source on physical dormancy break of *Vicia villosa* ssp. *Villosa. Seed Sci. Res.* 26, 254–263. doi: 10.1017/S096025851600012X

Renzi, J. P., Chantre, G. R., and Cantamutto, M. A. (2017). Self-regeneration of hairy vetch (*Vicia villosa* Roth) as affected by seedling density and soil tillage method in a semi-arid agroecosystem. *Grass Forage Sci.* 72, 535–544. doi: 10.1111/gfs.12255

Renzi, J. P., Chantre, G., and Cantamutto, M. A. (2018). *Vicia villosa* ssp. *villosa* Roth field emergence model in a semiarid agroecosystem. *Grass Forage Sci.* 73, 146–158. doi: 10.1111/gfs.12295

Renzi, J. P., Chantre, G. R., González-Andújar, J. L., and Cantamutto, M. A. (2019). Development and validation of a simulation model for hairy vetch (*Vicia villosa* Roth) self-regeneration under different crop rotations. *Field Crops Res.* 235, 79–86. doi: 10.1016/j.fcr.2019.02.020

Santos, D. J., Wilson, M. G., and Ostinelli, M. M. (2017). *Metodología de muestreo de suelos y ensayos a campo* (Entre Ríos, Argentina: Ediciones INTA).

Smouse, P. E., and Peakall, R. (1999). Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure. *Heredity* 82, 561–573. doi: 10.1038/sj.hdy.6885180

Smouse, P. E., Long, J. C., and Sokal, R. R. (1986). Multiple regression and correlation extensions of the Mantel test of matrix correspondence. *Syst. Zool.* 35, 627–632. doi: 10.2307/2413122

Tang, S., Kishore, V. K., and Knapp, S. J. (2003). PCR-multiplexes for a genome-wide framework of simple sequence repeat marker loci in cultivated sunflower. *Theor. Appl. Genet.* 107, 6–19. doi: 10.1007/s00122-003-1233-0

Toser, M. G., and Ooi, M. K. J. (2014). Humidity-regulated dormancy onset in the Fabaceae: a conceptual model and its ecological implications for the Australian wattle *Acacia saligna*. *Ann. Bot.* 114, 579–590. doi: 10.1093/aob/mcu144

Van Assche, J. A., and Vandelook, F. (2010). Combinational dormancy in winter annual Fabaceae. *Seed Sci. Res.* 20, 237–242. doi: 10.1017/S0960258510000218

Van de Wouw, M., Enneking, D., Robertson, L. D., and Maxted, N. (2001). "Vetches (*Vicia* L.)," in *Plant genetic resources of legumes in the Mediterranean*. Eds. N. Maxted and S. J. Bennett (Dordrecht, The Netherlands: Kluwer), 132–157.

Vanzolini, J. I. (2011). *La vicia villosa como cultivo de cobertura: efectos de corto plazo sobre el suelo y la productividad del maíz bajo riego en el Valle Bonaerense del río Colorado* (Universidad Nacional del Sur, Buenos Aires, Argentina: MScThesis).

Wayman, S., Kissing Kucek, L., Mirsky, S. B., Ackroyd, V., Cordeau, S., and Ryan, M. R. (2016). Organic and conventional farmers differ in their perspectives on cover crop use and breeding. *Renew Agr. Food Syst.* 32, 376–385. doi: 10.1017/S1742170516000338

Wiering, N. P., Flavin, C., Sheaffer, C. C., Heineck, G. C., Sadok, W., and Ehlke, N. J. (2018). Winter hardiness and freezing tolerance in a hairy vetch collection. *Crop Sci.* 58, 1594–1604. doi: 10.2135/cropsci2017.12.0748

Wilke, B. J., and Snapp, S. S. (2008). Winter cover crops for local ecosystems: Linking plant traits and ecosystem function. *J. Sci. Food Agric.* 88, 551–557. doi: 10.1002/jsfa.3149

Zhang, X., and Mosjidis, J. A. (1995). Breeding systems of several *Vicia* species. *Crop Sci.* 35, 1200–1202. doi: 10.2135/cropsci1995.0011183X003500040049x

Zhu, G., Li, H., and Zhao, L. (2017). Incorporating anthropogenic variables into ecological niche modeling to predict areas of invasion of Popillia japonica. *J. Pest Sci.* 90, 151–160. doi: 10.1007/s10340-016-0780-5

Conflict of Interest: 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 © 2020 Renzi, Chantre, Smýkal, Presotto, Zubiaga, Garayalde and Cantamutto. 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) and the copyright owner(s) 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 Plant Science

ORIGINAL RESEARCH
published: 03 March 2020
doi: 10.3389/fpls.2020.00082

Pod Dehiscence in Hairy Vetch (*Vicia villosa* Roth)

Lisa Kissing Kucek[1], Heathcliffe Riday[1], Bryce P. Rufener[1], Allen N. Burke[2], Sarah Seehaver Eagen[3], Nancy Ehlke[4], Sarah Krogman[5], Steven B. Mirsky[2], Chris Reberg-Horton[3], Matthew R. Ryan[6], Sandra Wayman[6] and Nick P. Wiering[4]*

[1] Dairy Forage Research Center, USDA-ARS, Madison, WI, United States, [2] Sustainable Agricultural Systems Laboratory, Beltsville Agricultural Research Center, USDA-ARS, Beltsville, MD, United States, [3] Crop and Soil Science, North Carolina State University, Raleigh, NC, United States, [4] Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN, United States, [5] Noble Research Institute, Ardmore, OK, United States, [6] School of Integrated Plant Science, Cornell University, Ithaca, NY, United States

Hairy vetch, *Vicia villosa* (Roth), is a cover crop that does not exhibit a typical domestication syndrome. Pod dehiscence reduces seed yield and creates weed problems for subsequent crops. Breeding efforts aim to reduce pod dehiscence in hairy vetch. To characterize pod dehiscence in the species, we quantified visual dehiscence and force required to cause dehiscence among 606 genotypes grown among seven environments of the United States. To identify potential secondary selection traits, we correlated pod dehiscence with various morphological pod characteristics and field measurements. Genotypes of hairy vetch exhibited wide variation in pod dehiscence, from completely indehiscent to completely dehiscent ratings. Mean force to dehiscence also varied widely, from 0.279 to 8.97 N among genotypes. No morphological traits were consistently correlated with pod dehiscence among environments where plants were grown. Results indicated that visual ratings of dehiscence would efficiently screen against genotypes with high pod dehiscence early in the breeding process. Force to dehiscence may be necessary to identify the indehiscent genotypes during advanced stages of selection.

Keywords: pod dehiscence, germplasm characterization, vetch, domestication syndrome, phenotyping, legumes (Fabaceae)

OPEN ACCESS

Edited by:
Petr Smýkal, Palacký University, Olomouc, Czechia

Reviewed by:
Tomas Vymyslicky, Agricultural Research, Ltd., Czechia
Paul Gepts, University of California, Davis, United States

***Correspondence:**
Heathcliffe Riday
Heathcliffe.Riday@usda.gov

Specialty section:
This article was submitted to Plant Breeding, a section of the journal Frontiers in Plant Science

Received: *30 September 2019*
Accepted: *21 January 2020*
Published: *03 March 2020*

Citation:
Kissing Kucek L, Riday H, Rufener BP, Burke AN, Eagen SS, Ehlke N, Krogman S, Mirsky SB, Reberg-Horton C, Ryan MR, Wayman S and Wiering NP (2020) Pod Dehiscence in Hairy Vetch (Vicia villosa Roth). Front. Plant Sci. 11:82. doi: 10.3389/fpls.2020.00082

INTRODUCTION

Hairy vetch, *Vicia villosa* (Roth), is an outcrossing diploid legume (2n = 14; Chooi, 1971; Yeater et al., 2004). Commonly used as a green manure cover crop (SARE et al., 2015; CTIC et al., 2016; USDA-NASS, 2019), the species excels in winter hardiness (Brandsæter et al., 2002) and nitrogen supply to subsequent cash crops (Parr et al., 2000). With prevalent pod dehiscence (PD) and seed dormancy, hairy vetch does not exhibit a typical domestication syndrome (Meyer and Purugganan, 2013; Abbo et al., 2014). PD raises seed costs and causes weediness in fields, reducing utilization of this cover crop. Breeding efforts are underway to reduce or eliminate PD in hairy vetch. The goal of such efforts is to increase cover crop use and improve soil and water conservation.

Few published studies have evaluated PD in the genus *Vicia*. Renzi et al. (2017) documented 15% to 46% PD in one Argentinian landrace evaluated at one location over two years, but no studies have

evaluated PD of hairy vetch among diverse germplasm or growing conditions. In common vetch (*Vicia sativa* L.), PD varied widely (3% to 96%) among diverse lines (Abd El-Moneim, 1993; Dong et al., 2017). Common vetch lines differing in PD exhibited 22 differentially expressed unigenes (Dong et al., 2017).

In other members of the Fabeae tribe, domestication has successfully eliminated PD (e.g. *Pisum* sp.) or reduced PD to very low levels relative to wild types (e.g. *Lens* sp.; Abbo et al., 2014; see review by Ogutcen et al., 2018). PD was controlled by one to three dominant loci in lentil (*Lens* sp.; Ladizinsky, 1979; Ladizinsky, 1985; Fratini et al., 2007) and one to two dominant loci in pea (*Pisum* sp.; Blixt, 1972; Weeden et al., 2002; Weeden, 2007). PD has been more extensively studied in the Phaseoleae tribe of Fabaceae. In soybean (*Glycine max* L. Merr.), the transcription factor *SHATI-5* and the gene *Pod dehiscence 1* (*Pdh1*) meditate and control PD. In common bean (*Phaseolus vulgaris* L.), various QTL have been identified among bean races (Parker et al., 2019), the most documented being the *Stringless* (*St*) gene in snap beans (Koinange et al., 1996).

PD is influenced by environmental conditions, length of pod drying, and handling methods postharvest (see reviews by Ogutcen et al., 2018 and Zhang et al., 2018). With varying maturity timings, diverse genotypes can be exposed to differing weather conditions during pod development. Consequently, genotype by environment interactions (Tukamuhabwa et al., 2002; Helms, 2011) can cloud genetic effects. More controlled measurements of PD, such as oven drying of pods to standardize moisture and/or applying force to a pod to induce dehiscence were more associated with genetic effects than measuring PD under field conditions (Weeden, 2007; Dong et al., 2016; Murgia et al., 2017; Parker et al., 2019). Such methods, particularly measuring the force needed to induce dehiscence, demand substantial phenotyping time and specialty equipment. Identification of traits that are easier to measure and highly correlated with PD could improve breeding efficiency (Falconer, 1981). Such secondary selection traits could accelerate improvement of hairy vetch.

The objectives of this research were to characterize PD among diverse genotypes of hairy vetch, and to identify the most efficient phenotyping methods for PD. First, we assessed PD among 606 genotypes of hairy vetch in seven environments of the United States. Second, we correlated potential secondary selection traits with measures of PD including visual PD, force to induce dehiscence, and pod spiraling.

MATERIALS AND METHODS

Data Collection

The germplasm for our genetic analysis of PD originated from an existing hairy vetch breeding program. For further details on source germplasm, nursery design, field data collection, and management see Kissing Kucek et al. (2019). In the late summer and early fall of 2017, seven nursery environments (Clayton, NC; Goldsboro, NC; Beltsville, MD; Ithaca, NY; Varna, NY; Prairie Du Sac, WI; and St. Paul, MN) planted half-sib seed from 92 maternal plants selected in the summer of 2017, in addition to 12 check lines (**Table S1**). Fifteen lines were replicated at all environments, including three maternal lines from the 2017 nurseries, four commercially available cultivars, one breeding line, and seven PI accessions. The remaining 77 lines at each site were half-sibling progeny of maternal plants selected from the breeding program in 2017. The seven environments planted a total of 14,304 genotypes of hairy vetch.

Individual plants were screened for visual vigor on a scale from zero to nine in the fall and spring (see Kissing Kucek et al., 2019 for details). When 50% of lines had begun to flower in each nursery, collaborators recorded maturity of each individual plant in the field using the Kalu and Fick (1981). In the spring of 2018, each site selected the top 53 to 122 hairy vetch plants based on winter survival, flowering time, and fall and spring vigor. The selected individuals cross pollinated at each site. Sites recorded the dry weight of selected plants at pod maturity, except Clayton, NC and Prairie du Sac, WI. All sites measured the seed yield of individual plants.

Collaborators harvested a subset of pods from a total of 606 selected plants for PD evaluation. These 606 plants are subsequently referred to as "genotypes." Collaborators at each site were instructed to collect a subset of around 50 mature pods from the selected maternal plant. If pods appeared to contain few seeds per pod, more pods were collected per plant, to obtain enough seed for a separate analysis on germination. **Figure S1** shows the distribution of pods collected per maternal line per site.

Pods were stored in paper bags under low humidity (< 35% R.H.) conditions at 20°C for five months. A calibration experiment determined that drying pods for least 24 h at 30°C was necessary to stabilize the PD trait and reach critical pod moisture in *Vicia villosa* (see supplementary material and **Figure S2**). Consequently, pods were dried above 30°C for three days to stabilize moisture and PD prior to analysis. Maternal lines were evaluated for metrics of PD and morphology (**Table 1**) in a completely randomized order.

For each pod per maternal line, visual dehiscence was scored on a 0–3 scale, with zero indicating a fully intact pod (no openings along sutures), one indicating one suture was partially opened (one side of pod), two indicating two sutures were partially opened (both sides of pod), and three indicating that the pod had fully opened.

Pods with a zero rating (unopened) were further tested with an MTS Insight ® 1kN (MTS Systems Corporation) machine for measuring cracking force on unopened pods. The pod was placed horizontally on a bottom compression platen, similar to the methods of Dong et al. (2016), and crushed with a top platen until the machine detected peak force required for suture rupture. The evaluation used a 100 N load cell and TestWorks ® [v.4.11B] (MTS Systems Corporation) set to 95% break sensitivity, 0.1 N break threshold, and 25 Hz data acquisition rate. Maternal lines were only evaluated if they had at least three unopened pods available for testing.

TABLE 1 | Metrics of pod dehiscence evaluation, including photos of pods exemplifying the maximum and minimum of rating scales.

Type	Metric	Rating Scale		
Dehiscence Metric	Visual Dehiscence	0 = fully intact pod (no openings along sutures), 1 = one suture was partially opened (one side of pod), 2 = two sutures were partially opened (both sides of pod), 3 = pod had opened in full or partially	0 	3 
	Force to Dehiscence	peak force (N) applied to create a break in the pod suture		
	Spiraling	0 = no spiraling of pod, 1 = some spiraling of pod, 2 = strong spiraling of pod	0 	2 
Morphology Metric	Corrugation	0 = smooth on the surface of the pod, 1 = intermediate, 2 = ridged on the surface of the pod - seeds evident	0 	2 
	Fracture	0 = nonlinear fracturing of pod wall when broken, 1 = linear fracturing of pod wall when broken	0 	1 
	Flexibility	1 = flexible, 0 = sturdy		
	Pith Tissue	0 = absence of pith tissue inside of pod, 1 = presence of pith tissue inside of pod	0 	1 

After pods were hand-threshed between two ribbed rubber surfaces, the degree of spiraling was rated by the tightness of the curls formed in the carpels. A zero was given to pods with no curling of the carpels, one for intermediate curling, and two for very tight curling of the carpels.

Pods from each maternal line were further rated for morphological features. Traits included corrugation, fracture of the wall, flexibility, and pith tissue (**Table 1**). Prior to threshing, corrugation was visually rated by the compression of the pod wall on the seeds inside the pod. Pods were rated zero if no signs of indents formed by the seeds in the pod wall, one if some indenting was present, and two if the pod wall had significant indentation formed by the seeds. After threshing, pod walls were rated for fracture structure by hand bending the carpel until a

break formed. The break would be categorized as zero for pods that formed a jagged fracture, and one if the break followed a straight path. Flexibility was tested in a similar method to fracture. Pods received a zero if they displayed resistance when bending the carpel, and one if the carpel bent with little resistance. After threshing, pods were observed for the presence or absence of a spongy white pith tissue inside of the pod. Pods were given values of zero if there was little to no pith tissue, and one if pith tissue was prevalent.

To verify that differences in pod moisture did not impact PD, a subset of lines diverging for visual PD and force to PD were analyzed for percent moisture. Three lines each from high and low PD categories from each site were included, for a total of 42 lines based on visual dehiscence scores and 40 lines based on force to PD. Some lines overlapped, leaving a total of 76 lines analyzed. Threshed pods, without seeds, were weighed, then dried at 105°C for 24 h and reweighed.

Statistical Analysis

Analyses investigated the relationships between all metrics of PD (visual dehiscence, force to dehiscence, and pod spiraling), pod morphology (corrugation, fracture, flexibility, pith tissue), field measurements (fall vigor, spring vigor, maturity timing, plant weight, and seed yield), and pod moisture. Mean values per line were used for visual dehiscence and force to dehiscence. Lines with fewer than four pods observed were removed from analyses. One extreme outlier for force to dehiscence (see **Figures 1B** and **2B**) was also removed prior to analysis.

Line mean visual dehiscence, force to dehiscence, maturity timing, and pod moisture were continuous variables. Corrugation

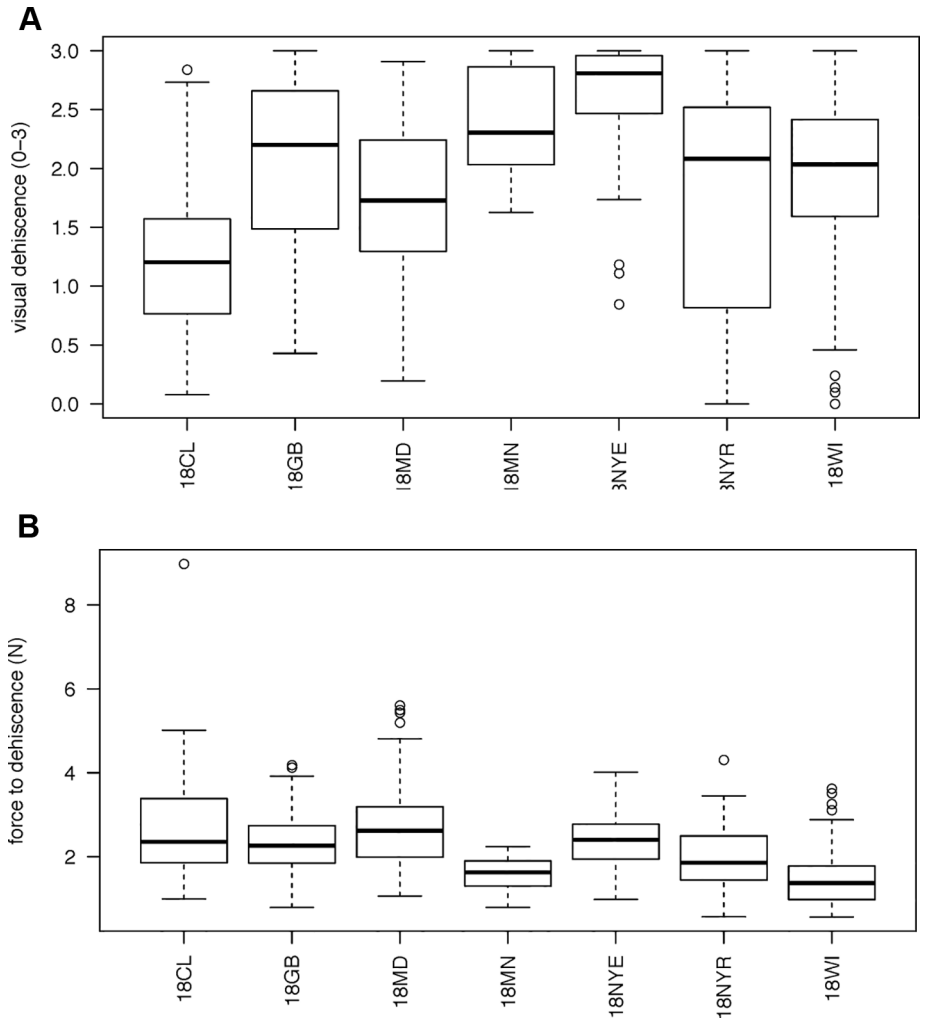


FIGURE 1 | Boxplots of variation in vetch genotypes for **(A)** mean visual dehiscence **(B)** force necessary to cause dehiscence. Environments are listed as 18CL (Clayton, NC), 18GB (Goldsboro, NC), 18MD (Beltsville, MD), 18MN (St. Paul, MN), 18NYE (Ithaca, NY), 18NYR (Varna, NY), 18WI (Prairie Du Sac, WI). There is a strong outlier for force to dehiscence at 18CL. Visual dehiscence was scored on a 0–3 scale, with "0" indicating a pod fully intact pod (no openings along sutures), "1" indicating one suture was partially opened (one side of pod), "2" indicating two sutures were partially opened (both sides of pod), and "3" indicating that the pod had opened fully or partially.

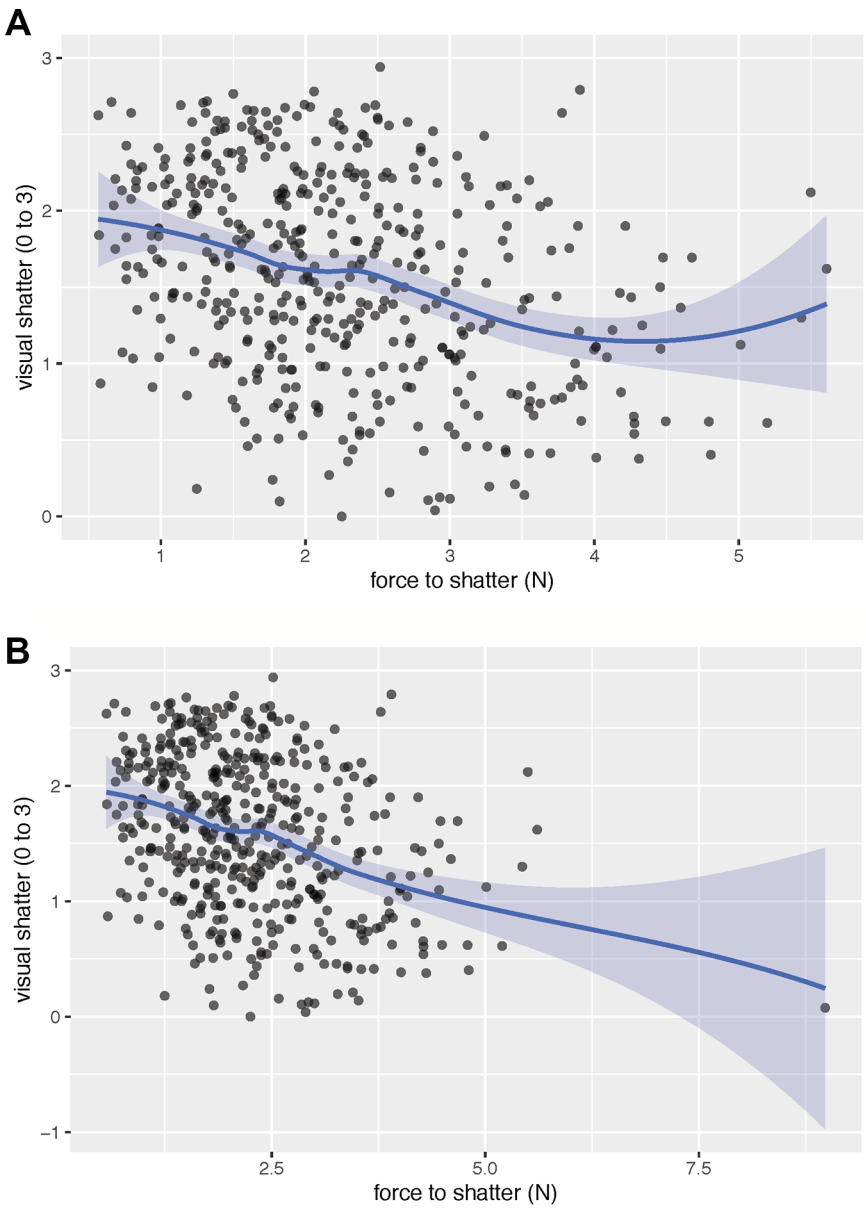


FIGURE 2 | The fitted spline shows the relationship between visual rating of dehiscence and the force required to cause dehiscence, excluding **(A)** and including **(B)** an extreme outlier for force to dehiscence. When excluding the outlier, there was a moderate linear relationship (r = –0.33) between visual dehiscence and force to dehiscence. The linear relationship broke down at low levels of visual dehiscence and high force required to cause pod dehiscence. Both traits may be necessary to identify lines most resistant to dehiscence. Visual dehiscence was scored on a 0–3 scale, with "0" indicating a pod fully intact pod (no openings along sutures), "1" indicating one suture was partially opened (one side of pod), "2" indicating two sutures were partially opened (both sides of pod), and "3" indicating that the pod had opened fully or partially.

and pod spiraling were ordinal variables artificially created to represent underlying continuous normal distributions, while fracture, flexibility, and pith tissue were dichotomous variables created to represent underlying continuous normal distributions. Maximum likelihood estimates and standard errors of correlations were computed using Pearson's product-moment correlation coefficient (r) for continuous vs. continuous metrics, polyserial maximum likelihood estimates (ρ) for continuous vs. dichotomous or ordinal metrics, and polychoric maximum likelihood estimates (ρ) for dichotomous or ordinal vs. dichotomous or ordinal metrics using the 'polycor' package (Fox, 2016) in R [version 3.6.0] (Olsson, 1979; Olsson and Drasgow, 1982; R Core Team, 2019). To visualize potential nonlinear relationships between visual dehiscence and force to PD, smoothing splines were generated using the 'loess' function in R [version 3.6.0] (R Core Team, 2019) and plotted using 'ggplot2' (version 3.0.0) (Wickham, 2016).

Relationships were tested for difference from zero using Pearson's product-moment correlation for continuous vs. continuous metrics, Kruskal-Wallis one-way analysis of variance for continuous vs. dichotomous or ordinal metrics with nonnormal distribution, one-way analysis of variance (ANOVA) for continuous vs. dichotomous or ordinal metrics with normal distribution, and chi-squared test of independence for dichotomous or ordinal vs. dichotomous or ordinal metrics.

For each combination of metrics, we tested whether correlations varied among environments. Using the 'psych' package [v.1.8.12] (Revelle, 2019) in R [version 3.6.0] (R Core Team, 2019), independent correlations among environments for each trait combination were z-transformed. The difference between the z-transformed correlations was divided by the standard error of the difference of the z scores. All analyses were conducted in R [version 3.6.0] (R Core Team, 2019), using a significance threshold of $P < 0.05$.

To further understand the contribution of pod morphology and field maturity to PD, visual dehiscence and force to dehiscence were regressed onto pod morphology and maturity timing using the model below. All effects were treated as random to determine variance contribution of each effect. Variances were estimated using the 'lme4' package [version1.1-10] (Bates et al., 2015) in R [version 3.6.0] (R Core Team, 2019).

$$Y_{ijklmn} = \mu + \alpha_i + \beta_j + \gamma_k + \delta_l + \zeta_m + \eta_n + \epsilon_{ijk}$$

Y_{ijklmn}: visual or force to dehiscence of a pod from environment i, corrugation j, fracture k, flexibility l, pith tissue m, and flowering maturity n

μ: grand mean of visual dehiscence or force to dehiscence

α_i: random effect of environment i

β_j: random effect of corrugation j

γ_k: random covariate of fracture k

δ_l: random covariate of flexibity l

ζ_m: random covariate of pith tissue m

η_n: random covariate of flowering maturity n

ϵ_{ijklmn}: error term

RESULTS

606 genotypes varied from completely indehiscent to completely dehiscent (**Figure 1A**). Twenty genotypes, sourced from five environments, created pods that all opened (mean visual score = 3). In two genotypes, sourced at Prairie du Sac, WI and Varna, NY, all pods were fully closed (mean visual score = 0).

A total of 458 genotypes had at least three unopened pods available for force testing. Mean force to dehiscence varied from 0.279 to 8.97 N among maternal lines (**Figure 1B**). One outlier genotype was evident from the Clayton, NC site. Force to dehiscence for this line was four times the grand mean. Nine of ten pods evaluated for this line were in the 99th percentile of force to dehiscence measured in the study, requiring between 6.2 and 12 N to break the sutures. This line also exhibited the lowest visual dehiscence from its growing location, indicating its promise as a genetic oddity for PD.

Pod moisture varied from 6.7% to 9.3% in a subset of 76 divergent lines for PD. Pod moisture was not significantly related to visual dehiscence, force to dehiscence, flowering maturity, nor pod morphology traits (**Table S2**). Pod moisture showed a weak inverse correlation with spiraling (ρ = -0.25, p = 0.02048). As a result of low influence of moisture in the extreme phenotype subset, pod moisture was not included in further analyses, nor evaluated on the remaining 535 lines.

Corrugation of the pods and presence of pith tissue on the interior of the pod were only common at the three southern nurseries, and rare in the northern nurseries. Northern nurseries were subsequently removed from corrugation correlation calculations. Pod walls at one site (Ithaca, NY) nearly all fractured linearly, and therefore, pods from Ithaca, NY were removed from correlation calculations for fracture. Flowering maturity did not strongly differ at Beltsville, MD. Consequently, correlations with flowering maturity did not include Beltsville, MD.

All metrics of PD were significantly related to one another (**Table 2**). Across environments, spiraling was highly correlated with visual dehiscence (ρ = 0.64 to 0.87) and moderately correlated to force to dehiscence (ρ = −0.17 to −0.42). Correlations between visual and force dehiscence were moderate (r = −0.14 to −0.43). Visual dehiscence exhibited a linear relationship with force to dehiscence, except for pods resisting dehiscence at high levels of force (> 4 N; **Figure 2**).

PD was influenced by the environment where the pods developed. For the response of force to dehiscence, environment explained more variance (13.27%) than any pod morphology metrics (0% to 8.51%) or field maturity traits (0%, **Table 3**). Visual dehiscence was also influenced by environment (7.63% of variance), but to a lesser degree than force to dehiscence.

Corrugation, pith tissue, fracturing structure, and flexibility of the pod tended to be related to metrics of PD. However, low frequency of corrugation and pith tissue at the northern environments limited the inference of these morphological correlations to only three southern environments. Pod corrugation had the highest correlation with PD among the studied morphology metrics. Higher levels of corrugation were associated with lower visual dehiscence (ρ = −0.33 to −0.54), and spiraling (ρ = −0.36 to −0.76) at all environments. Corrugation also accounted for 46% of variance in the random effects model with the response of visual dehiscence (**Table 3**), and 3.9% of variance with the response of force to dehiscence.

The presence of pith tissue in the pod was associated with more force required to split the pod (ρ = 0.12 to 0.48) and reduced spiraling (ρ = −0.16 to −0.27) among environments. In the random effects model with the response of force to dehiscence, pith tissue accounted for the most variance among pod morphology metrics. However, the variance explained by pith tissue was small (8.5%, **Table 3**).

The fracturing structure of the pod wall was related to spiraling at all environments (ρ = 0.23 to 0.52), and visual dehiscence at five of six environments (ρ = 0.23 to 0.46). Pod flexibility was moderately related to spiraling (ρ = −0.14 to −0.58), and to visual dehiscence at five of six environments (ρ = −0.27 to −0.63). In the random effects model with the

TABLE 2 | Strength, significance, and consistency of association between pod dehiscence, morphology, and field data.

	Visual Dehiscence	**Force to Dehiscence**	**Spiraling**	**Corrugation**	**Fracture**	**Flexibility**	**Pith Tissue**	**Flowering Maturity**	**Fall Vigor**	**Spring Vigor**	**Plant Weight**	**Seed Weight**
Visual Dehiscence		Pearson	Polyserial/ Kruskal-Wallis	Polyserial/ Kruskal-Wallis	Polyserial/ Kruskal-Wallis	Polyserial/ Kruskal-Wallis	Polyserial/ Kruskal-Wallis	Pearson	Pearson	Pearson	Pearson	Pearson
Force to Dehiscence	−0.33 ***		Polyserial/ Kruskal-Wallis	Polyserial/ Kruskal-Wallis	Polyserial/ Kruskal-Wallis	Polyserial/ Kruskal-Wallis	Polyserial/ Kruskal-Wallis	Pearson	Pearson	Pearson	Pearson	Pearson
Spiraling	0.79 ***	−0.47 ***		Polychoric/χ^2	Polychoric/χ^2	Polychoric/χ^2	Polychoric/χ^2	Polyserial/ Kruskal-Wallis	Polyserial/ Kruskal-Wallis	Polyserial/ Kruskal-Wallis	Polyserial/ Kruskal-Wallis	Polyserial/ Kruskal-Wallis
Corrugation	−0.56 ¥ ***	0.35 ¥ ***	−0.73 ¥ ***		Polychoric/χ^2	Polychoric/χ^2	Polychoric/χ^2	Polyserial/ Kruskal-Wallis	Polyserial/ Kruskal-Wallis	Polyserial/ Kruskal-Wallis	Polyserial/ Kruskal-Wallis	Polyserial/ Kruskal-Wallis
Fracture	0.39 ***	−0.041 NS	0.47 ***	−0.045 ¥ NS		Polychoric/χ^2	Polychoric/χ^2	Polyserial/ Kruskal-Wallis	Polyserial/ Kruskal-Wallis	Polyserial/ Kruskal-Wallis	Polyserial/ Kruskal-Wallis	Polyserial/ Kruskal-Wallis
Flexibility	−0.44 ***	0.023 NS	−0.51 ***	0.26 ¥ ***	−0.61 ***		Polychoric/χ^2	Polyserial/ Kruskal-Wallis	Polyserial/ Kruskal-Wallis	Polyserial/ Kruskal-Wallis	Polyserial/ Kruskal-Wallis	Polyserial/ Kruskal-Wallis
Pith Tissue	−0.12 ¥ *	0.44 ***	−0.39 ¥ ***	0.35 ¥ ***	0.15 ¥ *	0.010 ¥ NS		Polyserial/ Kruskal-Wallis	Polyserial/ Kruskal-Wallis	Polyserial/ Kruskal-Wallis	Polyserial/ Kruskal-Wallis	Polyserial/ Kruskal-Wallis
Flowering Maturity	−0.11 **	−0.09 *	−0.01 NS	0.06 † NS	0.073 NS	0.10 *	−0.10 †		Pearson	Pearson	Pearson	Pearson
Fall Vigor	−0.02	0.03	0.03	−0.02 ¥	0.00	−0.08	0.06 ¥	0.08		Pearson	Pearson	Pearson
Spring Vigor	−0.13 **	0.22 ***	−0.30 ***	0.18 ¥ **	−0.06	−0.05	0.40 ¥ ***	−0.14 **	0.28 ***		Pearson	Pearson
Plant Weight	−0.06	0.07 ¥	−0.14 ***	0.05 †	−0.10	−0.03	0.28 † ***	−0.44 ¥ ***	−0.12 **	0.28 ***		Pearson
Seed Weight	0.12 **	−0.37 ***	0.38 ***	−0.30 ¥ ***	0.13 *	−0.14 *	−0.37 ¥ ***	0.20 ***	0.13 **	−0.18 ***	0.29 ***	

, **, and * indicate significance of correlations at the 0.05, 0.01, and 0.001 probability level, respectively.*

NS indicates that correlations are not significant at the $\alpha < 0.05$ threshold.

† indicates only one observed paired environment comparison.

¥ indicates only three observed paired environment comparisons.

Numbers show Pearson's product-moment correlation coefficient for continuous vs. continuous metrics, polyserial maximum likelihood estimates for dichotomous vs. ordinal or dichotomous metrics, and polychoric maximum likelihood estimates for dichotomous vs. dichotomous or ordinal metrics across environments. Significance values were calculated through Pearson's product-moment correlation for continuous vs. continuous metrics, chi-squared test of independence for dichotomous or ordinal vs. dichotomous or ordinal metrics, and Kruskal-Wallis one-way analysis of variance for continuous vs. dichotomous or ordinal metrics. Shading indicates consistency of the correlation among environments. Dark grey indicates significant differences in correlations among environments, with change in direction; light grey indicates significant differences in correlations among environments, but no change in direction; and white indicates no significant differences among environments).

TABLE 3 | Variance contribution of environment, pod morphology, and flowering maturity on metrics of dehiscence.

Visual Dehiscence			
Groups	Variance	Standard Deviation	% of Total Variance
Environment	0.068	0.26	7.63%
Corrugation	0.407	0.64	45.70%
Fracture	0.040	0.20	4.49%
Flexibility	0.021	0.15	2.41%
Pith Tissue	0.000	0.00	0.00%
Flowering Maturity	0.004	0.06	0.45%
Residual	0.350	0.59	39.32%
Force to Dehiscence			
Groups	Variance	Standard Deviation	% of Total Variance
Environment	0.120	0.35	13.27%
Corrugation	0.035	0.19	3.85%
Fracture	0.000	0.00	0.00%
Flexibility	0.000	0.00	0.00%
Pith Tissue	0.077	0.28	8.51%
Flowering Maturity	0.000	0.00	0.00%
Residual	0.673	0.82	74.37%

Visual dehiscence was most explained by corrugation, while force to dehiscence was explained by pith tissue, followed by corrugation.

response of visual dehiscence, fracturing structure and flexibility accounted for a small portion of variance, at 5.5% and 2.2%, respectively (**Table 3**).

Some measures of field performance were associated with PD. Spring vigor showed a weak correlation with visual dehiscence ($r = -0.01$ to -0.28), spiraling ($\rho = -0.04$ to -0.27), and force to dehiscence ($r = -0.08$ to 0.23), but the relationship varied by environment. Seed yield was correlated with spiraling at most sites ($r = 0.08$ to 0.56), force to dehiscence ($r = 0.01$ to -0.24) at some sites, and visual dehiscence at the southern sites ($r = 0.13$ to 0.47).

DISCUSSION

Our results indicated the potential to select for indehiscence in hairy vetch. Wide variation in visual and force to dehiscence existed among diverse genotypes. More importantly for selection, multiple lines exhibited indehiscence or very low levels of dehiscence.

This dataset also demonstrated environmental influence on PD, which is well documented in other species (see reviews by Ogutcen et al., 2018 and Zhang et al., 2018). Growing environment contributed substantial amounts of variance for visual dehiscence and force to dehiscence (**Table 3**). Moreover, correlations between metrics of PD, pod morphology, and flowering maturity significantly differed among environments (**Table 2**). To separate genetic effects from environmental influences and interactions, PD studies should utilize multiple environments. Secondary selection traits to speed phenotyping would need to consistently correlate among diverse environments within a breeding program region of interest.

Spiraling was highly correlated with PD and was a high-throughput measurement, requiring only 15 seconds per sample. Visual dehiscence provided higher resolution in PD than spiraling and was moderately time intensive, requiring 5 min to rate per line, at 50 pods evaluated per line. Force to dehiscence was the most involved measurement, requiring specialty equipment, a trained operator, and 18.5 min of evaluation time per line, with five pods evaluated per line. Although visual dehiscence and spiraling may be adequate to identify strongly dehiscent lines, force to dehiscence may be useful for identifying extreme lines most resistant to dehiscence. For initial screenings of dehiscence, spiraling could identify the genotypes most susceptible to PD at low cost. Visual dehiscence would be useful in early and middle stages of selection to eliminate moderately dehiscent lines. Once mean visual dehiscence levels become low (< 1) in a breeding population, force to dehiscence measurements would likely be necessary to further advance gains in selection (**Figure 2**).

Although pod morphology metrics were fast to measure (15 seconds per line), none were strongly related to PD among environments. Pod corrugation was moderately correlated with all measures of PD, and explained a large portion of variance for visual dehiscence. Pith tissue was moderately correlated with force to dehiscence and pod spiraling. However, pod corrugation and pith tissue did not commonly appear at the three environments in the northern United States. Consequently, pod corrugation and pith tissue would not be useful PD secondary selection traits for breeding programs including cold temperate climates.

Further study is needed to understand the physiology of pith tissue in hairy vetch pods. The pith tissue created a foam-like structure that seemed to inhibit compression force from breaking a pod, hence the trait's contribution to force to dehiscence (**Table 3**). However, the pith tissue may not be genetic resistance to PD, but rather a plant response to an environmental threat (e.g. a pathogen). To separate out environmental effects from true PD, the trait of pith tissue could serve as a covariate when analyzing force to dehiscence.

The fracture structure of the pod wall was moderately related to spiraling, and with visual dehiscence at some environments. The linear fracture morphology described in our paper likely relates to the alignment of pod wall fibers at an angle to pod sutures, which can cause spiraling of the carpel (Funatsuki et al., 2014). As the evaluation of spiraling required equal time to measure as fracture, and spiraling was more correlated to other metrics of PD, we see little utility for a rating of fracture.

Some traits showed inconsistent correlation with PD metrics among environments, such as pod flexibility. Such traits would not be reliable secondary selection traits for PD.

Pod moisture was not correlated with visual dehiscence or force to dehiscence. Consequently, pods in our study had likely reached the critical pod moisture required for PD. The weak correlation between spiraling and pod moisture could indicate that some samples were above the critical pod moisture threshold for PD. Moisture contents in our evaluation (6.7% to 9.3%) were below the critical pod moisture (10.1% to 10.4%) associated with PD in soybean (Zhang et al., 2018). However, these moisture contents were above the stable moisture found in common vetch (5%) (Dong et al., 2016). **Supplementary Figure 2** shows the results of PD after various pod drying times, heat conditions, and pod moistures to identify critical pod moisture in hairy vetch.

Flowering timing of lines were not strongly correlated with any measures of PD. In other species, genotypes with earlier flowering timing have exhibited more PD, as they were exposed for more time to heat and drying forces that can cause rupture of the dehiscence zone (Zhang et al., 2018). In our dataset, the stabilization of pod moisture *via* drying may have reduced the influence of maturity timing on PD.

Selecting for pod indehiscence may conflict with other field traits of interest in *Vicia villosa*. Lines with high spring vigor, a trait desired by growers (Wayman et al., 2016), also tended to have low PD, indicating the potential to select for both desired traits. However, there was a tradeoff between PD and seed yield in some environments. Selection for PD should closely monitor seed yield, to ensure lines developed for low PD also produce adequate yield for seed growers.

AUTHOR'S NOTE

Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation, to any qualified researcher.

AUTHOR CONTRIBUTIONS

LK and HR designed and analysed the experiment. LK wrote results. BR, AB, NE, SK, SM, CR-H, MR, SE, SW, and NW acquired and assisted with interpretation of the data and critical revision for intellectual content. All authors provided approval of publication and agree to be accountable for all aspects of the work.

FUNDING

Funding provided by National Institute for Food and Agriculture, Organic Research Extension Initiative Grant number 2015-51300-24192 and National Institute for Food and Agriculture Grant number 2018-67013-27570.

ACKNOWLEDGMENTS

The authors thank Rebecca Heidelberger for technical assistance in the field.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.00082/full#supplementary-material

REFERENCES

Abbo, S., Pinhasi van-Oss, R., Gopher, A., Saranga, Y., Ofner, I., and Peleg, Z. (2014). Plant domestication versus crop evolution: a conceptual framework for cereals and grain legumes. *Trends Plant Sci.* 19, 351–360. doi: 10.1016/j.tplants.2013.12.002

Abd El-Moneim, A. M. (1993). Selection for non-shattering common vetch, *Vicia sativa* L. *Plant Breed.* 110, 168–171. doi: 10.1111/j.1439-0523.1993.tb01231.x

Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Software* 67, 1–48. doi: 10.18637/jss.v067.i01

Blixt, S. (1972). Mutation genetics in pisum. *Agric. Hortic. Genet.* 30, 1–293.

Brandsæter, L. O., Olsmo, A., Tronsmo, A. M., and Fykse, H. (2002). Freezing resistance of winter annual and biennial legumes at different developmental stages. *Crop Sci.* 42, 437–443. doi: 10.2135/cropsci2002.0437

Chooi, W. (1971). Variation in nuclear DNA content in the genus vicia. *Genetics* 68, 195–211.

CTIC, SARE and ASTA. (2016). Annual report 2015-2016 cover crop survey. (West Lafayette, IN: Conservation Technology Information Center) doi: 10.1002/ejoc.201200111

Dong, R., Dong, D., Shao, K., Zhou, Q., Chai, X., Dong, Y., et al. (2016). The digital evaluation of the cracking force of common vetch with the different characteristics of pod shattering. *Pratacultural Sci.* 33, 2511–2517. doi: 10.11829/j.issn.1001-0629.2015-0704

Dong, R., Dong, D., Luo, D., Zhou, Q., Chai, X., Zhang, J., et al. (2017). Transcriptome analyses reveal candidate pod shattering-associated genes involved in the pod ventral sutures of common vetch (*Vicia sativa* L.). *Front. Plant Sci.* 8, 1–14. doi: 10.3389/fpls.2017.00649

Falconer, D. S. (1981). *Introduction to Quantitative Genetics. 2nd ed.* (London: Longmann Group).

Fox, J. (2016). Package "polycor:" polychoric and polyserial correlations. R package version 0.7–9. Available at: http://cran.r-project.org/package=polycor.

Fratini, R., Durán, Y., García, P., and Pérez De La Vega, M. (2007). Identification of quantitative trait loci (QTL) for plant structure, growth habit and yield in lentil. *Spanish J. Agric. Res.* 5, 348–356. doi: 10.5424/sjar/2007053-255

Funatsuki, H., Suzuki, M., Hirose, A., Inaba, H., Yamada, T., Hajika, M., et al. (2014). Molecular basis of a shattering resistance boosting global dissemination of soybean. *Proc. Natl. Acad. Sci. U. S. A.* 111, 17797–17802. doi: 10.1073/pnas.1417282111

Helms, T. C. (2011). Greenhouse and field evaluation of pod dehiscence in soybean. *Can. J. Plant Sci.* 74, 699–701. doi: 10.4141/cjps94-126

Kalu, B. A., and Fick, G. W. (1981). Quantifying morphological development of alfalfa for studies of herbage quality. *Crop Sci.* 21, 267–271. doi: 10.2135/cropsci1981.0011183X002100020016x

Kissing Kucek, L., Riday, H., Ehlke, N., Reberg-Horton, C., Maul, J., Mirsky, S. B., et al. (2019). Environmental influences on the relationship between fall and spring vigor in hairy vetch. *Crop Sci.* 59, 2443–2454. doi: 10.2135/cropsci2018.09.0569

Koinange, E. M. K., Singh, S. P., and Gepts, P. (1996). Genetic control of the domestication syndrome in common bean. *Crop Sci.* 36, 1037–1045. doi: 10.2135/cropsci1996.0011183X003600040037x

Ladizinsky, G. (1979). The genetics of several morphological traits in the lentil. *J. Hered.* 70, 135–137. doi: 10.1093/oxfordjournals.jhered.a109209

Ladizinsky, G. (1985). The genetics of hard seed coat in the genus Lens. *Euphytica* 34, 539–543. doi: 10.1007/BF00022952

Meyer, R. S., and Purugganan, M. D. (2013). Evolution of crop species: genetics of domestication and diversification. *Nat. Rev. Genet.* 14, 840–852. doi: 10.1038/nrg3605

Murgia, M. L., Attene, G., Rodriguez, M., Bitocchi, E., Bellucci, E., Fois, D., et al. (2017). A comprehensive phenotypic investigation of the "pod-shattering syndrome" in common bean. *Front. Plant Sci.* 8, 1–15. doi: 10.3389/fpls.2017.00251

Ogutcen, E., Pandey, A., Khan, M. K., Kahraman, A., Penmetsa, R. V., Marques, E., et al. (2018). Pod shattering: a homologous series of variation underlying domestication and an avenue for crop improvement. *Agronomy* 8, 137. doi: 10.3390/agronomy8080137

Olsson, U., and Drasgow, F. (1982). The polyserial correlation coefficient. *Psychometrika* 47, 337–347. doi: 10.1007/bf02294164

Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika1* 44, 443–460. doi: 10.1007/bfb0067701

Parker, T. A., Berny, J. C., Palkovic, A., and Jernstedt, J. (2019). Genetic control of pod dehiscence in domesticated common bean : associations with range expansion and local aridity conditions. *biorxiv*, 517516. doi: 10.1101/517516

Parr, M., Grossman, J. M., Brinton, C., and Crozier, C. (2000). Nitrogen delivery from legume cover crops in no-till organic corn production. *Agronomy J.* 103, 1578–1590. doi: 10.2134/agronj2011.0007

R Core Team (2019). R: a language and environment for statistical computing organization. (Vienna, Austria: R Foundation for Statistical Computing). Available at: http://www.r-project.org/.

Renzi, J. P., Chantre, G. R., and Cantamutto, M. A. (2017). Self-regeneration of hairy vetch (*Vicia villosa* Roth) as affected by seedling density and soil tillage method in a semi-arid agroecosystem. *Grass Forage Sci.* 72, 524–533. doi: 10.1111/gfs.12255

Revelle, W. (2019). Package "psych.". Available at: https://personality-project.org/r/psych.

SARE, CTIC and ASTA (2015). Cover crop survey annual report 2014-2015. (West Lafayette, IN: Conservation Technology Information Center). doi: 10.1002/ejoc.201200111

Tukamuhabwa, P., Dasheill, K. E., Rubaihayo, P., and Nabasirye, M. (2002). Determination of field yield loss and effect of environment on pod shattering in soybean. *Afr. Crop Sci. J.* 10, 203–209.

USDA-NASS (2019). *2017 Census of Agriculture* (Washington, DC: United States Department of Agriculture). Available at: https://www.nass.usda.gov/Publications/AgCensus/2017/index.php.

Wayman, S., Kissing Kucek, L., Mirsky, S. B., Ackroyd, V., Cordeau, S., and Ryan, M. R. (2016). Organic and conventional farmers differ in their perspectives on cover crop use and breeding. *Renew. Agric. Food Syst.*, 32, 376–386. Available at http://www.journals.cambridge.org/abstract_S1742170516000338.

Weeden, N. F., Brauner, S., and Przyborowski, J. (2002). Genetic analysis of pod dehiscence in pea (*Pisum Sativum* L.). *Cell. Mol. Biol. Lett.* 7, 657–663.

Weeden, N. F. (2007). Genetic changes accompanying the domestication of *Pisum sativum*: Is there a common genetic basis to the "domestication syndrome" for legumes? *Ann. Bot.* 100, 1017–1025. doi: 10.1093/aob/mcm122

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis* (New York, NY: Springer-Verlag).

Yeater, K. M., Bollero, G. A., Bullock, D. G., and Rayburn, A. L. (2004). Flow cytometric analyses for ploidy level differentiation of 45 hairy vetch accessions. *Ann. Appl. Biol.* 145, 123–127. doi: 10.1111/j.1744-7348.2004.tb00367.x

Zhang, Q., Tu, B., Liu, C., and Liu, X. (2018). Pod anatomy, morphology and dehiscing forces in pod dehiscence of soybean (Glycine max (L.) Merrill). *Flora Morphol. Distrib. Funct. Ecol. Plants* 248, 48–53. doi: 10.1016/j.flora.2018.08.014

Conflict of Interest: 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 © 2020 Kissing Kucek, Riday, Rufener, Burke, Eagen, Ehlke, Krogman, Mirsky, Reberg-Horton, Ryan, Wayman and Wiering. 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) and the copyright owner(s) 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 Plant Science

ORIGINAL RESEARCH
published: 10 March 2020
doi: 10.3389/fpls.2020.00034

Comparative Analysis of Early Life Stage Traits in Annual and Perennial *Phaseolus* Crops and Their Wild Relatives

Sterling A. Herron[1,2*], *Matthew J. Rubin*[2], *Claudia Ciotir*[1†], *Timothy E. Crews*[3], *David L. Van Tassel*[3] *and Allison J. Miller*[1,2*]

[1] Department of Biology, Saint Louis University, St. Louis, MO, United States, [2] Donald Danforth Plant Science Center, St. Louis, MO, United States, [3] The Land Institute, Salina, KS, United States

OPEN ACCESS

Edited by:
Petr Smýkal,
Palacký University, Czechia

Reviewed by:
Steven B. Cannon,
United States Department of Agriculture, United States
Péter Török,
University of Debrecen, Hungary

***Correspondence:**
Sterling A. Herron
sterling.herron@slu.edu
Allison J. Miller
allison.j.miller@slu.edu

†Present address:
Claudia Ciotir,
Department of Evolutionary and Environmental Biology and Institute of Evolution, University of Haifa, Haifa, Israel

Specialty section:
This article was submitted to Plant Breeding, a section of the journal Frontiers in Plant Science

Received: 25 October 2019
Accepted: 13 January 2020
Published: 10 March 2020

Citation:
Herron SA, Rubin MJ, Ciotir C, Crews TE, Van Tassel DL and Miller AJ (2020) Comparative Analysis of Early Life Stage Traits in Annual and Perennial Phaseolus Crops and Their Wild Relatives. Front. Plant Sci. 11:34. doi: 10.3389/fpls.2020.00034

Herbaceous perennial species are receiving increased attention for their potential to provide both edible products and ecosystem services in agricultural systems. Many legumes (Fabaceae Lindl.) are of special interest due to nitrogen fixation carried out by bacteria in their roots and their production of protein-rich, edible seeds. However, herbaceous perennial legumes have yet to enter widespread use as pulse crops, and the response of wild, herbaceous perennial species to artificial selection for increased seed yield remains under investigation. Here we compare cultivated and wild accessions of congeneric annual and herbaceous perennial legume species to investigate associations of lifespan and cultivation with early life stage traits including seed size, germination, and first year vegetative growth patterns, and to assess variation and covariation in these traits. We use "cultivated" to describe accessions with a history of human planting and use, which encompasses a continuum of domestication. Analyses focused on three annual and four perennial species of the economically important genus *Phaseolus*. We found a significant association of both lifespan and cultivation status with seed size (weight, two-dimensional lateral area, length), node number, and most biomass traits (with cultivation alone showing additional significant associations). Wild annual and perennial accessions primarily showed only slight differences in trait values. Relative to wild forms, both cultivated annual and cultivated perennial accessions exhibited greater seed size and larger overall vegetative size, with cultivated perennials showing greater mean trait differences relative to wild accessions than cultivated annuals. Germination proportion was significantly lower in cultivated relative to wild annual accessions, while no significant difference was observed between cultivated and wild perennial germination. Regardless of lifespan and cultivation status, seed size traits were positively correlated with most vegetative traits, and all biomass traits examined here were positively correlated. This study highlights some fundamental similarities and differences between annual and herbaceous perennial legumes and provides insights into how perennial legumes might respond to artificial selection compared to annual species.

Keywords: perennial grain, *Phaseolus*, Fabaceae, legume, pulse, crop wild relative, domestication

INTRODUCTION

Life history theory traditionally categorizes plants as annuals, typified by fast growth and high reproductive effort, and perennials, typified by long-term survival and delayed reproduction (Cole, 1954; Harper, 1977). However, exceptions to this general trend occur; patterns of plant resource allocation exist along a gradient, with some perennial species showing fast growth and high reproductive output (e.g., Verboom et al., 2004; González-Paleo and Ravetta, 2015). Exploration of this life history diversity is important in plant breeding, as wild, herbaceous perennial species are increasingly considered as sources of novel genetic variation through crosses and as entirely new domesticates targeted for seed products. Many major crops have perennial wild relatives that remain uncharacterized despite their potential.

Nearly all grain and legume crops grown for human consumption are annual plants that complete their life cycle in a single year, or are perennial species cultivated as annuals (Van Tassel et al., 2010). Although thousands of herbaceous perennial species exist within major crop families (e.g., Ciotir et al., 2016; Ciotir et al., 2019), annual plant species were likely selected during the early stages of domestication due to pre-existing agriculturally favorable traits, e.g., high reproductive yield in a single season and accelerated germination and flowering (Van Tassel et al., 2010). Over time, artificial selection has led to exceptional gains in reproductive output in annual crops, particularly in the last century (Mann, 1997); however, cultivation intensity and other agronomic practices have resulted in widespread soil loss (FAO, 2019). Current research focuses in part on the development of crops to support the ecological intensification of agriculture, which aims to achieve both high yields and ecosystem services, such as soil and water retention (Ryan et al., 2018).

The use of herbaceous perennial species as seed crops has been proposed as one method of achieving ecological intensification (Jordan and Warner, 2010; Ryan et al., 2018). Longer-lived species have deep, persistent root systems that mitigate erosion and enhance nutrient uptake, they produce perennating shoots which reduce yearly planting costs, and they have a longer photosynthetically active growth period, allowing for high biomass production each year (Cox et al., 2006; Crews et al., 2018). However, only a few perennial seed crops (principally cereals, oilseeds, and pulses) have entered the domestication process, and we know relatively little about how artificial selection for increased seed production will impact the rest of the perennial plant.

Artificial selection is the human-mediated evolutionary process that leads to changes in plant traits over the course of generations. This process happens as a result of selective cultivation, the act of preferentially planting individuals with desired features. Over time cultivated populations evolve in response to artificial selection, leading to domestication: the evolution of morphological and genetic changes in cultivated populations relative to their wild progenitors (Harlan, 1995; Meyer et al., 2012). Because domestication is an ongoing, evolutionary process, a continuum of cultivated populations exist in species undergoing domestication, ranging from cultivated populations which display little or no differences relative to their wild (uncultivated) progenitors, to highly modified elite breeding lines, such contemporary maize, which differs dramatically from its closest wild relatives (Harris, 1989; Bharucha and Pretty, 2010; Breseghello and Coelho, 2013). To determine the extent to which domestication has occurred, the precise identity of the wild ancestor of a crop and the exact cultivated populations derived from it are required. Because these data are not available for all species examined here, we use the term "cultivated" to refer to any accession that has a history of cultivation, acknowledging that this encompasses a broad spectrum of phenotypic and genetic change under artificial selection.

The "domestication syndrome" describes a common suite of trait changes seen across different species in response to artificial selection for seed and/or fruit production (Hammer, 1984; Harlan, 1992). In annual species, the domestication syndrome typically includes loss of seed dormancy, higher germination, loss of shattering, greater seed size or number, and erect, determinate growth, among many others (Harlan et al., 1973; Olsen and Wendel, 2013; Abbo et al., 2014). In contrast to annuals, the domestication syndrome of woody perennials, such as fruit and nut trees, is characterized by an extended juvenile phase, outcrossing mating system, and often clonal propagation; consequently, woody perennials have typically undergone fewer cycles of sexual selection under domestication and retain a greater proportion of wild genetic diversity relative to annual domesticates (Zohary and Spiegel-Roy, 1975; Miller and Gross, 2011; Gaut et al., 2015). Herbaceous perennial plants cultivated for both edible seeds and sustained perennation were only recently targeted for selection (Suneson et al., 1963; DeHaan et al., 2016; Kane et al., 2016; Kantar et al., 2016; Crews and Cattani, 2018), and it is unclear if they will follow an evolutionary trajectory similar to domesticated annual and woody perennial species, or if they will show a unique domestication syndrome.

The agricultural context provides an entirely new adaptive landscape that may allow novel combinations of traits in herbaceous perennial species (e.g., high reproductive output and longevity), combinations that are often unfavorable in natural environments (Crews and DeHaan, 2015; Cox et al., 2018). Ongoing work seeks to understand how artificial selection for increased seed yield in herbaceous perennial species might impact vegetative traits and the capacity for perennation more generally. One hypothesis is that perennial seed crops are constrained by a vegetative-reproductive trade-off, where high reproductive allocation and sufficient storage allocation for perennation cannot coexist (Van Tassel et al., 2010). In other words, it may be possible for artificial selection to drive increases in seed yield in wild, herbaceous perennial species, but those increases may cause losses in allocation to vegetative and perennating structures, resulting in a shift from perenniality to annuality (Denison, 2012; Smaje, 2015). Some studies have supported such a trade-off (e.g., González-Paleo et al., 2016; Vico et al., 2016; Cattani, 2017; Pastor-Pastor et al., 2018). An alternative hypothesis is that reproductive yield and vegetative

biomass may be selected for in concert, leading to sustained perennation. Concomitant perennation and high seed yield have been observed in some perennial cereals (Sacks et al., 2007; Jaikumar et al., 2012; Culman et al., 2013; Huang et al., 2018).

Herbaceous perennial seed crop studies can benefit from including comparisons with closely related annual domesticates, to clarify if domestication responses are dependent upon lifespan. Vico et al. (2016), in a meta-analysis of 67 annual-perennial pair studies across nine plant families, found greater reproductive allocation in annual crops and greater root allocation in perennial crops, which was the same pattern found in unselected wild groups. Since a majority of the studies in this meta-analysis were from the grass family, such annual-perennial meta-analyses may be augmented by empirical research on specific lineages of plants, to determine if phylogenetically focused trends are similar to the broader patterns observed, as well as to allow a more precise biological interpretation.

Here we focus on the legume family (Fabaceae), which includes 19,500+ species of which more than 30% are predominantly herbaceous perennials (Ciotir et al., 2019). Fabaceae is the second most economically important plant family after the grasses (Poaceae), with 41 domesticated species dating back to the first agricultural systems, and 1,000+ species cultivated for various purposes across the world (Harlan, 1992; Lewis et al., 2005; Hammer and Khoshbakht, 2015). To date, few herbaceous perennial legume species have been thoroughly assessed for agriculturally relevant traits such as seed size, germination rate, time to maturity, root/shoot allocation, and reproductive yield. Characterizing these and similar traits in herbaceous perennial crops and their wild relatives will be critical to assess what genetic variation is available for crop improvement through introgression and *de novo* domestication (Schlautman et al., 2018; Smýkal et al., 2018).

While there are known consistent differences in growth and resource allocation in some groups of annual and perennial species, gaps remain in our knowledge about how herbaceous perennials respond to artificial selection relative to their annual congeners in many plant lineages. Here we explore life history differences in members of the common bean genus *Phaseolus* by addressing the following questions: 1) How do *wild* annual and perennial *Phaseolus* species from multiple geographic origins allocate resources to seed production and vegetative growth? 2) What is the phenotypic signature of artificial selection in *cultivated* annual and perennial *Phaseolus* species? 3) What covariation exists between seed and adult vegetative growth traits, and is it consistent across lifespans and between cultivated and wild forms in *Phaseolus*? We address these questions by examining seed size, germination, and vegetative growth allocation among three annual and four perennial *Phaseolus* species (**Figure 1**). Through this work, we hope to contribute to ongoing efforts characterizing life history differences in closely related annuals and perennials within Fabaceae and shed light on how these lifespan groups may change with artificial selection.

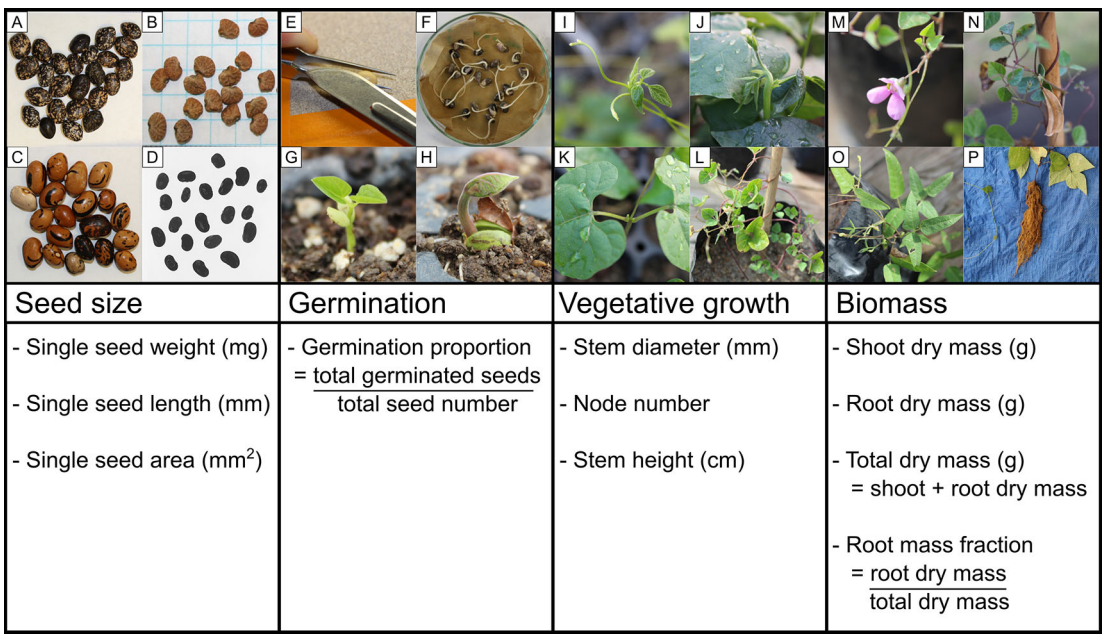


FIGURE 1 | Schematic of the four developmental stages analyzed in this study (seed size, germination, early vegetative growth, and biomass harvest) and examples of phenotypic diversity across the *Phaseolus* species studied: seeds of **(A)** *Phaseolus acutifolius*, **(B)** *P. angustissimus*, and **(C)** *P. coccineus*; **(D)** *P. coccineus* seeds prepared for analysis in ImageJ; **(E)** scarification by nicking the seed coat; **(F)** *P. vulgaris* germinants; **(G)** epigeal germination of *P. filiformis*; **(H)** hypogeal germination of *P. coccineus*; **(I, J)** *Phaseolus* shoot apex at which stem height was measured to; **(K)** *Phaseolus* first node (unifoliate) below which stem diameter was measured; **(L)** fully grown *P. filiformis* shoot with developed nodes (unfolded leaves); **(M, N)** *P. filiformis* flower and ripe fruit; **(O)** *P. acutifolius* shoot biomass; **(P)** *P. coccineus* root biomass. Photo credit: SH.

METHODS

Plant Material

Three annual *Phaseolus* species (*P. acutifolius, P. filiformis, P. vulgaris*; 58 accessions) and four perennial *Phaseolus* species (*P. angustissimus, P. coccineus, P. dumosus, P. maculatus*; 66 accessions) were included in this study (**Tables 1** and **2**). Species were chosen based on phylogenetic proximity and similar habitat types. The perennials *P. coccineus* and *P. dumosus* and annuals *P. acutifolius* and *P. vulgaris* are in the Vulgaris clade, and the perennial *P. angustissimus* and annual *P. filiformis* are in the Filiformis clade. Perennial *P. maculatus* is the only species in this study from the Polystachios clade (Delgado-Salinas et al., 2006). Geographically, our sampling of *Phaseolus* includes arid-adapted species of the Sonoran Desert (*P. acutifolius*, *P. angustissimus*, *P. filiformis*, and *P. maculatus*) and more tropically distributed species (primarily Mesoamerica: *P. coccineus*, *P. dumosus*, and *P. vulgaris*, with some South American accessions of the latter; **Supplementary Table S4**).

In total, we obtained seeds from 124 accessions from the United States Department of Agriculture's National Plant Germplasm System (Western Regional PI Station, Pullman, WA, stored at -18°C) in spring 2016, which were stored in a desiccator at 4°C, 33-50% relative humidity. All seeds were derived from plants regrown from the original collection material at the germplasm facility, but there were nevertheless possible genotype by environment and maternal effects that cannot be resolved here. 2,759 seeds were germinated and a subset of these grown from July to September 2016. The seed age for each accession, i.e., the length of time they were in frozen storage, ranged from one year to greater than 46 years (coded as 46).

Lifespan and Cultivation Status Assignment

We classified species in terms of the predominant lifespan observed in wild populations from their native range. In some cases, this differed from lifespan assignment in the USDA accession description, in which case it was confirmed by

TABLE 1 | Summary of sampling for seed size and germination traits for *Phaseolus*, by species and cultivation status.

Lifespan	Species†	Status	Seed size		Germination	
			Accessions	Seeds	Accessions	Seeds
annual	*P. acutifolius*	cultivated	17	25 (20-26)	20	25 (24-26)
		wild	6	25 (25)	10	24 (23-25)
	P. filiformis	wild	7	25 (25-26)	7	24 (23-25)
	P. vulgaris	cultivated	6	22 (21-24)	6	21 (20-23)
		wild	15	23 (20-30)	15	23 (20-30)
perennial	*P. angustissimus*	wild	2	23 (23)	2	23 (23)
	P. coccineus	cultivated	42	20 (19-25)	40	20 (19-25)
		wild	9	25 (20-42)	9	25 (19-41)
	P. dumosus	cultivated	8	25 (20-26)	8	24 (20-25)
		wild	2	26 (25-26)	2	23 (20-25)
	P. maculatus	wild	3	22 (21-25)	3	21 (21)

† Some species included subspecies designations: P. acutifolius (wild: 3 accessions of var. tenuifolius; cultivated: 2 accessions of var. acutifolius); P. coccineus (wild: 2 accessions of var. coccineus); P. vulgaris (wild: 8 accessions of var. aborigineus); P. maculatus (all accessions of ssp. ritensis). P. vulgaris includes wild accessions from both the Mesoamerican and South American ranges.

'Accessions' refers to the total number of accessions used for each species/cultivation status combination. 'Seeds' refers to the average seed number among accessions studied for seed size and germination (range of seed number in parentheses). Seed size collectively includes seeds at least measured for length and area, as well as seed weight in all but a few cases.

TABLE 2 | Summary of sampling for early vegetative growth and biomass traits for *Phaseolus*, by species and cultivation status.

Lifespan	Species	Status	Stem diameter		Node number		Stem height		Shoot dry mass		Root dry mass	
			Acc.	Plants	Acc.	Plants	Acc.	Plants	Acc.	Plants	Acc.	Plants
annual	*P. acutifolius*	cultivated	3	11 (9-12)	3	11 (9-12)	8	10 (2-25)	9	6 (1-11)	9	1 (1-2)
		wild	5	16 (10-20)	5	15 (6-21)	7	16 (7-24)	6	16 (10-21)	6	4 (2-7)
	P. filiformis	wild	7	14 (10-19)	7	14 (10-19)	7	14 (10-19)	7	7 (2-12)	7	2 (1-5)
	P. vulgaris	cultivated	5	7 (3-14)	5	7 (3-14)	2	5 (4-6)	4	7 (2-12)	4	2 (1-2)
		wild	12	13 (2-22)	12	13 (2-22)	9	13 (2-22)	12	9 (2-15)	12	2 (1-4)
perennial	*P. angustissimus*	wild	2	7 (6-7)	2	7 (6-7)	2	7 (6-8)	2	2 (2)	1	2
	P. coccineus	cultivated	10	9 (2-11)	10	9 (2-11)	13	9 (2-16)	14	7 (1-14)	12	2 (1-4)
		wild	5	9 (2-15)	6	8 (1-15)	4	12 (9-15)	5	7 (1-11)	5	2 (1-3)
	P. dumosus	cultivated	1	2	1	2	3	3 (1-5)	2	3 (2-4)	2	1 (1)
	P. maculatus	wild	3	8 (2-18)	3	8 (2-17)	2	10 (2-18)	3	6 (3-11)	3	2 (1-3)

"Acc." refers to the total number of accessions used for each species/cultivation status combination. "Plants" refers to the average plant number among accessions studied for that trait (range of plant number in parentheses).

extensive literature review; this occurred for *P. coccineus*, *P. dumosus*, and *P. filiformis*. Wild *P. coccineus* is a vigorous, perennial, indeterminate vine with an extensive root system; perenniality is also maintained in many cultivated forms (Delgado-Salinas, 1988; Smartt, 1988; Debouck, 1992; Freytag and Debouck, 2002*). P. dumosus*, a hybrid of *P. coccineus* and *P. vulgaris*, is also perennial, although it is less frost tolerant than *P. coccineus* (Smartt, 1988; Schmit and Debouck, 1991; Debouck, 1992; Freytag and Debouck, 2002; Mina-Vargas et al., 2016). Lastly, *P. filiformis* is an ephemeral, annual vine primarily found in the Sonoran Desert, which can survive up to seven months in favorable conditions (Buhrow, 1983; Nabhan and Felger, 1985; Freytag and Debouck, 2002). In addition, one accession of *P. maculatus* (PI 494138) was labeled as annual, although the species in general and and other accessions of this species are classified as perennial (Freytag and Debouck, 2002).

Cultivation status was taken directly from the USDA's description, with the descriptors "cultivated," "cultivar," and "landrace" all categorized as "cultivated" for the purpose of this data set. We use the umbrella term "cultivated" rather than "domesticated," since we do not have the data to determine the extent to which phenotypic and genetic change has occurred from the original wild population selected upon (see Introduction). All *Phaseolus* species are native to the Americas and were originally cultivated in either Mesoamerica or South America with some later selection occurring in Eurasia (Bitocchi et al., 2017). Cultivated *Phaseolus* species included here were first domesticated at least 1000 years before present, with most being domesticated much earlier (Kaplan and Lynch, 1999). Species-specific details on geographic origin and domestication are available in **Supplementary Table S4**.

Traits Measured

Seed Size

A total of 51 annual accessions (1,227 seeds) and 66 perennial accessions (1,436 seeds) were analyzed for size traits (**Figure 1**; **Table 1**). Seeds from each accession were weighed in bulk to the nearest mg and mean single seed weight was estimated by dividing the bulk weight by the total number of seeds for that accession. We imaged all accessions on a light table with a fixed camera at a resolution of 640 × 360 or 1349 × 748 pixels (differences accounted for in linear models), with all seeds oriented on their lateral side. From these two-dimensional images, we used ImageJ (Rasband, 1997-2016) to measure mean single seed length and area.

Germination

58 annual accessions (1,390 seeds) and 64 perennial accessions (1,369 seeds) were monitored for germination (sampling differences from seed size are due to a few germinated accessions not being analyzed for seed size and vice versa, and some seed loss in the germination procedure). Number of germinated seeds was monitored for each accession and used to calculate germination proportion (**Figure 1**; **Table 1**). Seeds were germinated on RO-water dampened germination paper or with a dampened cotton ball in petri dishes, following 1) sterilization by soaking in 1% bleach for 2 minutes and rinsing, 2) scarification by nicking the seed coat with a scalpel (to break physical dormancy, i.e., a water-impermeable seed coat), and 3) soaking for an average of 23 hours (range: 13-32 hours) by submersion in RO water. The soaking start date ranged from June 26 to July 8, 2016 (one accession soaked on June 19); this was considered time point 0 (i.e., the sowing date) for germination counts, since all necessary resources were available for the seeds to germinate. The germination apparatus was placed onto a 24°C heat mat in 24-hour dark conditions (except for germinant counts and planting; López Herrera et al., 2001). Germination was defined as an extension of the radicle past the seed coat. In rare cases where the seed coat was lost or very hard, it was defined as a distinct vigorous movement of the radicle away from the seed or a distinct pushing of the seed coat away from the seed, respectively. Petri dishes were treated with Banrot 40WP fungicide solution (prepared according to the product label). Fungus-infected, potentially salvageable seeds were soaked in 1% to 2% bleach and rinsed with RO water. Germinated seeds were also scored for seed quality following any potential damage from pre-germination treatments (0-2, with 0 being no damage and 2 being the highest damage), and seeds which were compromised due to procedural damage or had prematurely germinated in storage were removed from the analysis. Germination was monitored prior to planting; accessions with remaining seeds were checked once more 7–10 days after planting for any new germinants. Any variation in germination protocol was noted and addressed in statistical models, as well as the covariates seed quality and seed age (years of frozen storage at the germplasm center since the last seed increase). Subsets of individuals from the original number germinated were chosen for further growth measurement based on the presence of ≥10 vigorous individuals, if the accession was from the native range of the species (preferred), and if the accession was from a duplicate geographic location (removed).

Vegetative Growth Measurements

495 annual individuals (40 accessions) and 224 perennial individuals (29 accessions) of *Phaseolus* were transplanted to a greenhouse and measured for at least one vegetative trait (details below; **Figure 1**; **Table 2**). Seedlings were initially planted on July 13–14, 2016 in a mixture of unsterilized local riverine soil (Smoky Hill River, Salina, KS: 38.765948 N, -97.574213 W), sand, and potting soil (PRO-MIX) in small trays until they could be planted in 8" tall x 4" wide bag pots after 2-3 weeks. Initial planting date in small trays was used as the baseline for future measurements (days after planting, DAP), since despite different sowing dates, they were developmentally similar upon planting. Bag pots were filled with a mixture of the same riverine soil and coarse sand, to mimic field soil while also maximizing drainage. All *Phaseolus* were twining and were trained up four-foot bamboo poles. Plants did not receive any rhizobial inoculant treatment. Plants were initially bottom-watered twice daily (10AM, 6PM) for 20 minutes, and a shade cloth was incorporated in the greenhouse. After measurement of early vegetative growth (before biomass), some modifications were made. On August 18, 2016, watering was changed to once for 10

minutes every two days, and the shade cloth was removed on August 23, 2016. 90 of the most vigorous *Phaseolus* plants were moved from the greenhouse to the outdoors on August 25-26, 2016 to expose them to a more light-intense, natural environment. At this time, individual plants in both the greenhouse and outdoors were randomized to reduce spatial bias. All growth analyses were conducted at the research facilities of The Land Institute (Salina, KS).

At 19-23 DAP (25-40 days after sowing), plants were measured for early vegetative growth traits, which included stem diameter below the first node, total developed node number counted from the unifoliate node to the last node with an unfolded leaf, and stem height from ground to shoot apex on the tallest main stem, with twining stems being uncoiled from their poles and straightened as far as possible without damaging the plant. Plants were checked for reproductive status before biomass harvest. At 68-75 DAP (74-93 days after sowing), a random subset of plants was harvested for shoot and root (washed) biomass (**Figure 1**; **Table 2**), which was dried at a minimum of 37°C for at least 24 hours. Biomass was weighed on a precision or analytical scale depending on plant size. Root mass fraction was calculated on an individual plant basis from biomass measurements (root dry mass/total dry mass). Variation in growth conditions (greenhouse or outdoors), plant health (ordinal rating: 0,1,2; unhealthy, moderate health, healthy), and reproductive status (ordinal rating: 0,1,2,3; no reproduction, budding, flowering, fruiting) were recorded for individual plants and accounted for in statistical analyses (see below).

Statistical Analyses

In order to assess associations of lifespan and cultivation status with trait variation, we used linear models and *post hoc* comparative analyses on a set of mean values for each trait from each accession (**Table 3**). Associations of lifespan, cultivation (nested within lifespan), and species (nested within lifespan and cultivation), in addition to any relevant covariates for the focal trait, were tested using linear models. All potentially confounding factors were included in the original model and were then dropped sequentially if found to be nonsignificant. The base model for all main analyses was: *trait = lifespan + lifespan/cultivation + lifespan/cultivation/species*. Analyses were calculated for accession-level means for all traits and covariates. Due to concerns about lifespan lability of *P. dumosus*, linear models for all traits were checked with this species included and then with the species removed from the dataset. Accessions with any uncertainty associated with their lifespan or cultivation status were also dropped from the model and checked in the same manner; these included PI 494138 (*P. maculatus*; called annual while usually perennial—see Lifespan Assignment) and PI 390770 (*P. vulgaris*, noted as "wild or naturalized" in the USDA description). Pairwise comparisons of lifespan and cultivation effects for each trait were evaluated with *post-hoc* Tukey HSD tests (**Supplementary Table S1**).

We ran separate linear models and Tukey HSD tests for *Phaseolus* wild accessions to detect any phenotypic signatures of geographic origin, divided broadly into desert-adapted and tropical-adapted species (**Supplementary Tables S2** and **S3**). The base model for the geographic analyses was: *trait = geography + lifespan + geography×lifespan + geography/lifespan/species*, including any relevant covariates. Cultivated accessions were not included in this model due to potentially confounding effects of artificial selection. See **Supplementary Table S4** for the assignment of geographic origin categories to species.

To assess trait covariation within the whole dataset, Pearson product-moment correlations were performed on all pairwise combinations of the 11 measured traits using mean accession-level data, to determine the magnitude and direction of their relationships. Pearson correlations were also run on the

TABLE 3 | Results of linear models for all traits in the total *Phaseolus* dataset.

	Trait	Lifespan	Cultivation	Species			
(a)	Seed weight	$F_1 = 99.07^{***}$	$F_2 = 38.11^{***}$	$F_7 = 1.28$			
	Seed length†	$F_1 = 162.81^{***}$	$F_2 = 69.20^{***}$	$F_7 = 3.59^{**}$			
	Seed area†	$F_1 = 124.61^{***}$	$F_2 = 41.76^{***}$	$F_7 = 1.47$			
	Trait	**Lifespan**	**Cultivation**	**Species**	**Seed age**	**Soak time**	**Seed quality**
(b)	Germination proportion	$F_1 = 3.10$	$F_2 = 3.15^{*}$	$F_7 = 1.84$	$F_1 = 2.43$	$F_1 = 1.72$	$F_1 = 13.69^{***}$
	Trait	**Lifespan**	**Cultivation**	**Species**	**Health**		
(c)	Stem diameter	$F_1 = 4.03$	$F_2 = 36.82^{***}$	$F_6 = 3.44^{**}$	$F_1 = 0.24$		
	Node number	$F_1 = 20.00^{***}$	$F_2 = 7.13^{**}$	$F_6 = 4.99^{***}$	$F_1 = 17.53^{***}$		
	Stem height	$F_1 = 1.94$	$F_2 = 2.22$	$F_6 = 1.51$	$F_1 = 10.99^{**}$		
	Trait	**Lifespan**	**Cultivation**	**Species**	**Health**	**Reproductive status**	**Outdoor proportion**
(d)	Shoot dry mass	$F_1 = 21.25^{***}$	$F_2 = 18.71^{***}$	$F_6 = 2.43^{*}$	$F_1 = 2.08$	$F_1 = 1.80$	$F_1 = 5.00^{*}$
	Root dry mass	$F_1 = 11.80^{**}$	$F_2 = 10.53^{***}$	$F_6 = 1.37$	$F_1 = 8.35^{**}$	$F_1 = 3.57$	$F_1 = 0.69$
	Total dry mass	$F_1 = 14.44^{***}$	$F_2 = 14.47^{***}$	$F_6 = 2.02$	$F_1 = 7.33^{**}$	$F_1 = 3.49$	$F_1 = 0.43$
	Root mass fraction	$F_1 = 3.33$	$F_2 = 4.92^{*}$	$F_{6\ =}\ 4.89^{***}$	$F_1 = 13.89^{***}$	$F_1 = 0.09$	$F_1 = 4.15^{*}$

*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

†Differences in image resolution (640x360 or 1349×748 pixels) were accounted for in this model: F values were nonsignificant (0.0566 for seed length and 0.8301 for seed area).

Letters denote separate models with different covariates, while the main effects are the same for all traits: (a) seed size traits, (b) germination proportion, (c) early vegetative growth traits, and (d) biomass traits. See the Methods for explanations of each covariate.

following subsets of the data to qualitatively assess any trait covariation differences: annual, perennial, cultivated, and wild (**Supplementary Figure S1**). Each subset of the data was only restrictive in regard to one criterion, e.g., the annual subset contained data for both cultivated and wild annual accessions. Statistical analyses and figure generation were performed in R v. 3.6.1 (R Core Team, 2019).

All accession-level and individual plant data can be found in **Supplemental Table S5**. All individual seed size data may be found in **Supplemental Table S6**.

RESULTS

We investigated annual and perennial *Phaseolus* species for potential differences and covariation in seed size, germination, and vegetative growth. We found that wild accessions of annual species showed nonsignificantly greater germination and vegetative trait values than wild perennial accessions, but did not show greater seed size traits (**Figure 2**). Cultivated accessions of both annual and perennial species had greater trait values compared to wild accessions for almost all traits measured, with

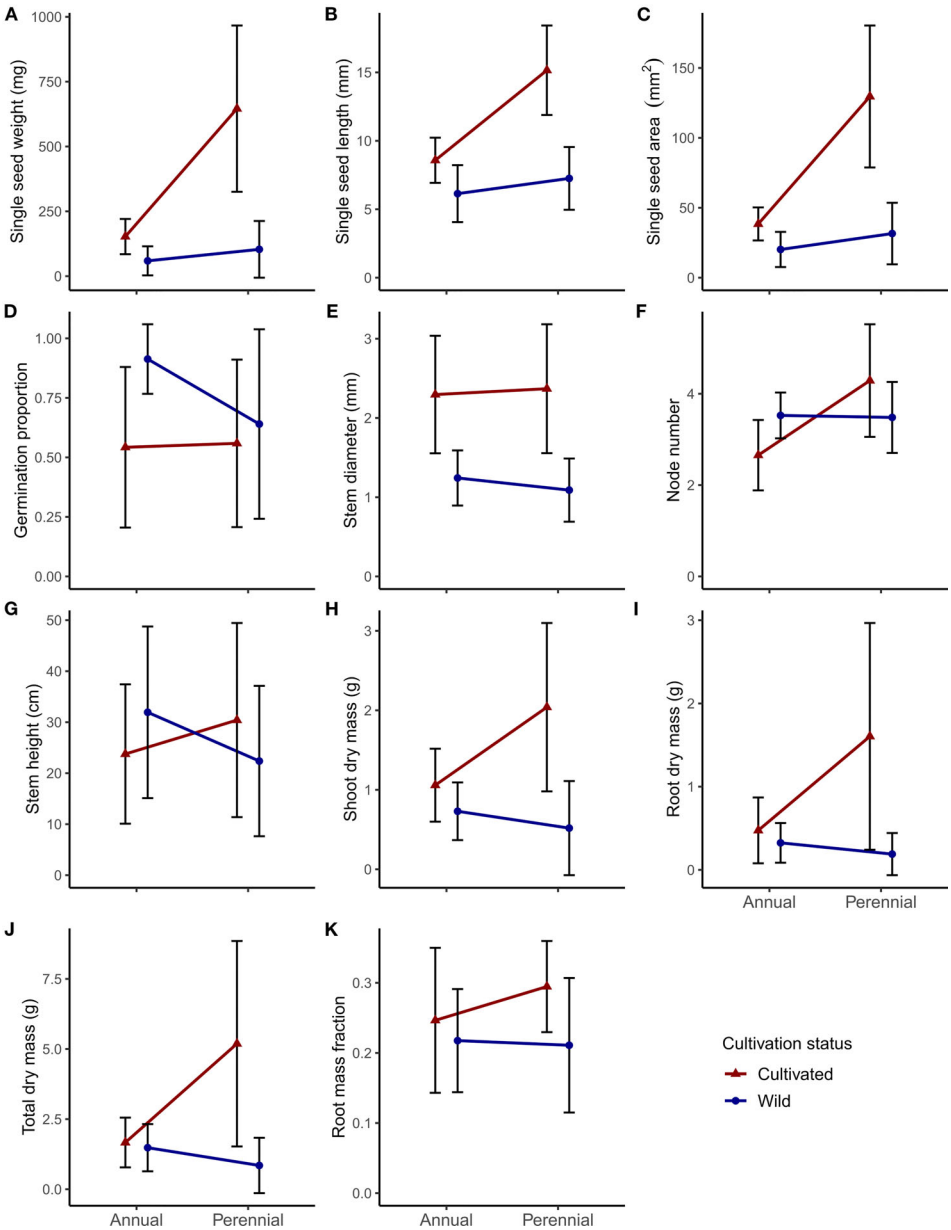


FIGURE 2 | Panel of phenotypic differences between lifespan (annual or perennial) and cultivation status (cultivated or wild) for the 11 focal traits in *Phaseolus*: **(A–C)** seed size traits; **(D)** germination proportion; **(E–G)** early vegetative growth traits; **(H–K)** shoot and root biomass traits. Central points represent means of all accessions for that category; error bars represent one standard deviation.

greater mean increases observed in perennial species (**Figure 2**). Lastly, seed and vegetative traits were significantly positively correlated, with some variation in this trend within subsets of the data (**Figure 3**; **Supplementary Figure S1**).

Trait Differences in Wild Annual vs. Perennial *Phaseolus* Accessions

Although nonsignificant, wild perennial *Phaseolus* mean seed weight was nearly twice that of wild annuals (59 mg annual vs. 104 mg perennial; **Figure 2A**; see **Supplementary Table S1** for all mean values, standard deviation, and Tukey test significance). Wild annual germination proportion was nonsignificantly higher than that of wild perennials (0.91 annual vs. 0.64 perennial; **Figure 2D**; **Supplementary Table S1**). Wild annual *Phaseolus* had similar to nonsignificantly larger vegetative trait values compared to wild perennials, with the largest relative differences seen in root dry mass (0.32 g annual vs. 0.19 g perennial) and total dry mass (1.48 g annual vs. 0.85 g perennial; **Figures 2I, J**; **Supplementary Table S1**). Mean root mass fraction was nearly equivalent for wild annuals (0.22) and wild perennials (0.21; **Figure 2K**; **Supplementary Table S1**).

In our main linear models, lifespan explained a significant amount of the variation seen in all seed size traits, node number, and most biomass traits (except root mass fraction; **Table 3**). Seed age and soak time were not significant predictors of germination; seed quality had significance at $P < 0.001$ (**Table 3**). Plant health was significant (at least at $P < 0.01$) in the linear models for all vegetative growth and biomass traits except stem diameter and shoot dry mass (**Table 3**). Reproductive status was not significant for any biomass trait; outdoor proportion was significant at $P < 0.05$ for shoot dry mass and root mass fraction (**Table 3**). The removal of *P. dumosus* and the accessions PI 494138 (*P. maculatus*) and PI 390770 (*P. vulgaris*; see Methods) changed some linear model results; here we note traits for which any variable's significance was lost or gained in the main model, which occurred for some vegetative traits. The exclusion of *P. dumosus* resulted in lifespan becoming significant for stem diameter ($P < 0.05$), in species becoming nonsignificant for shoot dry mass, and in outdoor proportion becoming nonsignificant for root mass fraction. Similarly, the exclusion of PI 494138 resulted in lifespan becoming significant for stem diameter ($P < 0.05$) but also for root mass fraction ($P < 0.05$). The exclusion of PI 390770 did not change significance in any of our models.

While our data do not allow precise interpretation of geographic effects, linear models including geographic origin as a main effect along with lifespan found that geographic origin explained a significant portion of the variation seen in all seed

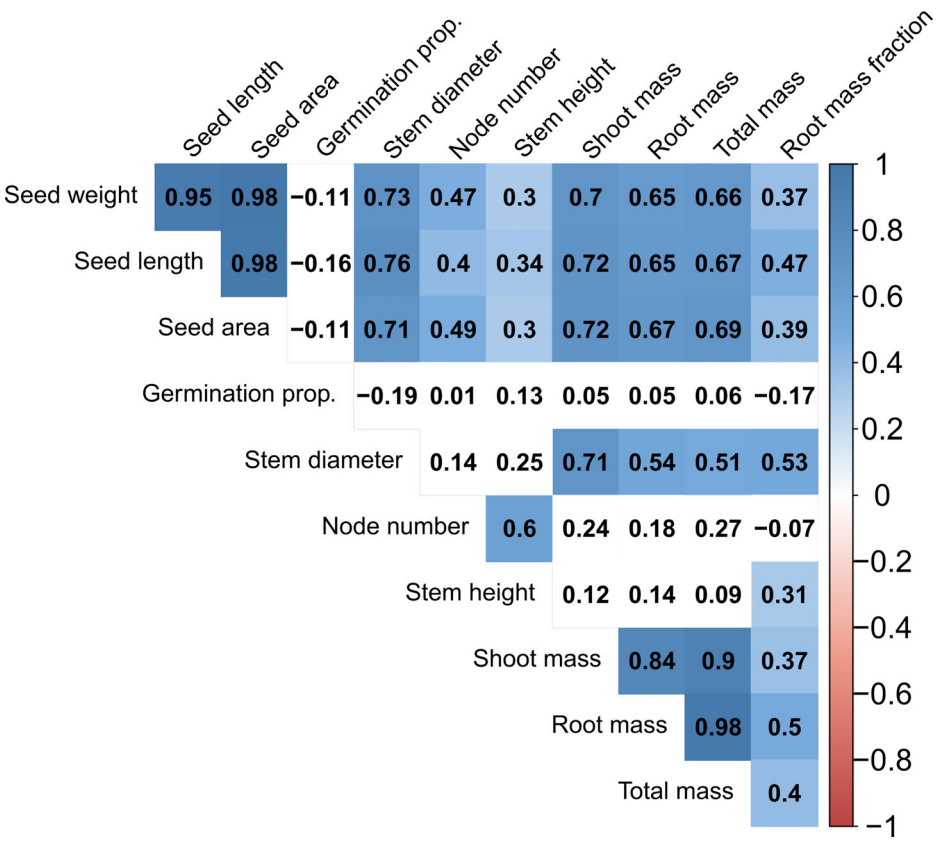


FIGURE 3 | Correlation diagram of all traits for all accessions in the *Phaseolus* dataset. Numbers in boxes represent the Pearson correlation coefficient. Blue and red colors indicate significant positive and negative correlations (at $P < 0.05$), respectively; absence of color indicates lack of significance.

size traits and most vegetative traits for wild *Phaseolus* accessions (**Supplementary Table S2**). In general, tropically distributed *Phaseolus* accessions had larger seed and vegetative growth traits compared to desert species (**Supplementary Table S3**). Mean germination proportion was relatively high (> 0.85) for all *Phaseolus* groups except tropical perennials (0.48, although tropical perennials also had the greatest standard deviation; **Supplementary Table S3**).

Trait Differences in Cultivated vs. Wild *Phaseolus* Accessions

Cultivated annual and perennial *Phaseolus* accessions showed generally larger seed and vegetative size characteristics compared to wild relatives (**Figures 2**). Cultivation differences in seed size were only significant for seed length in annual *Phaseolus*, although cultivated annual seed weight (153 mg) was nearly three times larger than wild annuals (59 mg) (**Figure 2A**; **Supplementary Table S1**). Cultivated perennial *Phaseolus* had significantly greater seed size in all traits, with seed weight over six times larger in cultivated perennials (646 mg) than wild perennials (104 mg) (**Figure 2A**; **Supplementary Table S1**). Germination proportion was significantly lower in cultivated annual *Phaseolus* (0.54) relative to wild annuals (0.91), while cultivated perennial germination proportion was only slightly lower (0.56) than wild perennials (0.64) (**Figure 2D**; **Supplementary Table S1**).

Cultivated perennial *Phaseolus* accessions tended to have significantly larger vegetative features compared to their wild relatives, whereas cultivated annual *Phaseolus* accessions usually displayed nonsignificantly larger vegetative features compared to their wild relatives (**Figures 2E–K**; **Supplementary Table S1**). Stem diameter was however significantly greater in both cultivated annual and perennial *Phaseolus* (**Figure 2E**; **Supplementary Table S1**). Cultivated annual *Phaseolus* showed nonsignificantly lower values in node number and stem height compared to wild annuals, while cultivated perennial *Phaseolus* showed nonsignificantly larger values in both traits compared to wild perennials (**Figures 2F, G**; **Supplementary Table S1**). Both cultivated annual and perennial *Phaseolus* had greater dry biomass trait values than their wild relatives, but for cultivated annual *Phaseolus*, the only significantly larger biomass value was shoot dry mass (1.06 g cultivated vs. 0.73 g wild; **Figure 2H**; **Supplementary Table S1**). In contrast, all biomass trait values were significantly greater in cultivated relative to wild perennial *Phaseolus* (except root mass fraction), with a greater than six-fold larger root dry mass value (1.60 g cultivated vs. 0.19 g wild; **Figure 2I**; **Supplementary Table S1**). In our linear models, cultivation explained a significant amount of variation for all traits except stem height (**Table 3**).

Phenotypic Covariation Across All Traits

Trait covariation was predominantly positive within the total dataset (**Figure 3**) and each subset of the data (**Supplementary Figures S1A-D**). Considering the entire dataset, all seed dimensions (weight, length, area) were nearly perfectly correlated (R^2 = 0.95-0.98; **Figure 3**). Seed traits were also significantly positively correlated with all vegetative growth traits, having the highest correlation with stem diameter and most biomass traits (R^2 = 0.65-0.76; **Figure 3**). Stem diameter had positive albeit non-significant correlations with node number and stem height, and significant positive correlations with all biomass traits. Node number and stem height had nonsignificant positive correlations with most biomass traits (**Figure 3**). Biomass traits including shoot dry mass, root dry mass, total dry mass, and root mass fraction were significantly positively correlated with one another, and there was a tight relationship between shoot and root dry mass (R^2 = 0.84; **Figure 3**). Germination proportion was the only trait in the total dataset to have no significant correlations and to have more than two negative correlations with other traits (**Figure 3**).

Subsets of our data also displayed similarly positive trait correlations with some exceptions. Due to sample size, each subset was restricted in regard to only one criterion (lifespan or cultivation status), e.g., the perennial subset contained both cultivated and wild perennial accessions. Notably, the annual subset showed negative correlations between node number and most traits (significant for seed size traits, stem diameter, and shoot dry mass), as well as a significant negative correlation of germination proportion with seed weight and stem diameter (**Supplementary Figure S1A**). In contrast, the perennial, cultivated, and wild subsets respectively tended to show positive correlations for the same trait pairings which were negatively correlated in annuals, although germination proportion and node number showed at least one negative correlation in all subsets (**Supplementary Figure S1B-D**). Significant positive seed size to vegetative trait correlations tended to be stronger and more common in the perennial and cultivated subsets than the annual and wild subsets (**Supplementary Figures S1**).

DISCUSSION

This study examined lifespan and cultivation effects among annual and herbaceous perennial legume species in the economically important genus *Phaseolus*, with the aim of identifying common trends and potential covariation among traits. Consistent trends included greater seed and plant size in cultivated relative to wild accessions for both annual and perennial species, and positive correlations among most seed and vegetative traits for annual and perennial species in the wild and under cultivation.

Wild Annual and Perennial *Phaseolus* Accessions Show Only Slight Differences in Trait Values

Wild perennial *Phaseolus* accessions in this study exhibited somewhat greater seed size trait values relative to wild annual accessions, which is consistent with the general life history expectation that later successional species produce larger seeds that are better able to compete for resources (Thompson and Hodkinson, 1998). This trend was consistent for *Phaseolus* species

native to the desert and the tropics (**Supplementary Table S3**). Tropical *Phaseolus* species had higher mean seed size than desert *Phaseolus*, which is consistent with the broad pattern of increasing seed size at lower latitudes (Moles et al., 2007; **Supplementary Table S3**). The large range of phenotypic variation observed in annual and perennial *Phaseolus* (**Figures 2A–K**) is also consistent with the large distributions of several species, such as *P. vulgaris* and *P. coccineus,* which inhabit environments ranging from very arid to very humid (Dohle et al., 2019).

While few studies have compared germination in wild *Phaseolus* species, Bayuelo-Jiménez et al. (2002) reported a similarly high germination proportion (0.85+) for scarified wild annual and perennial *Phaseolus* accessions, including the desert species *P. angustissimus* and *P. filiformis*. Although wild annuals showed higher mean germination than wild perennials considering the whole dataset, desert perennial accessions all reached 100% germination, suggesting that the broader trend is more applicable to the tropical species studied here (**Supplementary Table S3**). Our lowest wild mean germination was observed in the tropical perennials (*P. coccineus* and *P. dumosus*; **Supplementary Table S3**), which may reflect the trend that seed dormancy is less common in tropical herbaceous species in regions of higher rainfall (Baskin and Baskin, 2014).

While germination showed consistent differences between wild annuals and perennials, we cannot make ecologically robust conclusions here, since we effectively removed the legumes' primary form of physical dormancy through scarification, and each accession was stored frozen for different lengths of time at USDA facilities, which can have diverse effects on species' germination biology (Walters et al., 2005). Our results may be more reflective of the viability of seeds after long-term storage. In this light, our findings are consistent with life history predictions that annual species will maintain a complex, long-lived seed bank in which dormancy breaking and germination is staggered over time to ensure offspring survival in variable environments (Venable and Lawlor, 1980; Gremer et al., 2016). Perennials may be less selectively constrained by this pressure due to the parent plant's persistent survival in more stable environments (Cohen, 1966; Thompson et al., 1998).

The lack of significant differences in vegetative traits between wild annual and perennial *Phaseolus* suggests that the traits studied here do not diverge at this growth stage among different *Phaseolus* life history strategies in nature. Wild perennials' slightly lower mean vegetative growth could be due to a slower growth rate and the early stage of growth at which the traits were measured (three to eleven weeks after planting), since some perennials have been shown to be able to achieve a higher total biomass than related annuals when the entire growing season is considered (Dohleman and Long, 2009). Previous studies have also found that vegetative growth, shoot biomass, and root biomass are similar between closely related annuals and perennials, up to 40 days of growth, before their resource allocation patterns diverge (De Souza and Vieira Da Silva, 1987; Garnier, 1992). The significance of plant health in vegetative linear models for stem height and node number may reflect a greater sensitivity of response in stem length traits to environmental stressors compared to stem diameter; the health effect on root dry mass and derived traits (total dry mass and root mass fraction) may be due to a lower sample size for these traits compared to shoot dry mass alone (**Tables 2** and **3**).

Cultivated *Phaseolus* Accessions Show Greater Seed and Plant Size Relative to Wild Accessions

Phaseolus species showed larger seed and vegetative trait values in cultivated relative to wild accessions, consistent with domestication syndrome expectations; however, there were some notable differences in this trend between the annual and perennial groups. Consistent with previous studies in *Phaseolus* (Smartt, 1988; Koinange et al., 1996; Aragao et al., 2011), all cultivated annuals and perennials exhibited greater seed size relative to wild accessions. Perennial *Phaseolus* showed a particularly large difference in cultivated vs. wild seed size, as well as substantial variation in this trait, suggesting a large amount of genetic diversity. Much of the perennial seed size variation stemmed from *P. coccineus,* which is primarily outcrossing and more genetically diverse than other *Phaseolus* crops (Bitocchi et al., 2017; Guerra-García et al., 2017). Our seed size observations in USDA accessions are consistent with an analysis of wild and domesticated *Phaseolus* accessions from the International Center for Tropical Agriculture (CIAT), with the exception that they observed larger ranges of seed weight in cultivated annual species (*P. acutifolius*, *P. vulgaris*; Chacón-Sánchez, 2018). They similarly found that the perennial *P. coccineus* had the largest range in seed size in both cultivated and wild accessions, that *P. coccineus* showed the largest mean seed weight increase with cultivation, and that the perennial *P. dumosus* also showed some of the largest cultivated seed weights (Chacón-Sánchez, 2018).

Cultivation effects on *Phaseolus* germination may reflect different annual-perennial seed dormancy strategies. Domestication in annual common bean (*P. vulgaris*) and other seed crops has been known to reduce seed coat thickness and seed dormancy (Koinange et al., 1996; Fuller and Allaby, 2009), which could expose cultivated accessions to premature water imbibition and seed mortality in storage (López Herrera et al., 2001). Comparably low germination proportion in cultivated and wild perennial *Phaseolus* could be due to a less selective pressure for seed dormancy (and therefore lower seed longevity) in wild perennials relative to wild annuals, resulting in less change in germination proportion with domestication. Since our accessions were of various ages and geographic origin, more precise studies on fresh seedstock are necessary to confirm germination differences with cultivation for both annual and perennial species.

Vegetative size was generally larger in cultivated relative to wild annuals and perennials, although there were some notable exceptions. Mean node number and stem height were lower in cultivated annual *Phaseolus* and higher in cultivated perennial *Phaseolus* relative to wild accessions. Similarly, node number and height have been found to be lower in cultivated relative to wild annual *P. vulgaris* (Koinange et al., 1996; Berny Mier y Teran et al., 2018). This suggests that on average cultivated annual

Phaseolus exhibit lower degrees of stem growth, which may be the result of direct selection for determinate growth or an indirect consequence of increased harvest index, i.e., biomass allocated to reproductive structures at the expense of vegetative growth. This is in spite of all of the cultivated accessions in this study retaining their twining habit. In contrast, cultivated perennial *Phaseolus* exhibited consistently higher values for vegetative growth relative to wild perennials, possibly indicating a persistence of indeterminate growth in cultivation (Smartt, 1988), allowing for simultaneous selection on increased vegetative growth and reproduction (but see Delgado-Salinas, 1988). Both cultivated annual and cultivated perennial accessions displayed higher shoot and root dry mass relative to wild accessions, in agreement with the findings of Berny Mier y Teran et al. (2018) in *P. vulgaris*; this coincided with higher values in stem diameter for cultivated accessions of both lifespans, in agreement with Smartt (1988). Under cultivation, higher shoot and root biomass relative to wild accessions have also been observed in 30 diverse annual crop species, attributed to greater seed size and leaf area allowing enhanced downstream effects on growth (Milla et al., 2014; Milla and Matesanz, 2017). Our study suggests that some cultivated perennial species in *Phaseolus* exhibit similar size increases.

Root mass fraction, a resource-conservative trait, exhibited higher values in cultivated relative to wild accessions for both lifespans; this contradicts the expectation that plants become more resource-acquisitive with domestication (McKey et al., 2012; Vilela and González-Paleo, 2015; Pastor-Pastor et al., 2018). Higher root mass fraction was also observed in the annuals *Phaseolus vulgaris* (Berny Mier y Teran et al., 2018) and *Pisum sativum* (Weeden, 2007) relative to wild forms. This difference could be a byproduct of a more general higher allocation to vegetative growth, allowing greater amounts of photosynthate to be allocated to roots (Weeden, 2007). In summary, while many vegetative traits exhibit higher values in cultivated relative to wild accessions for both annual and perennial *Phaseolus* species, some traits may exhibit divergent lifespan patterns. It remains to be determined if this is due to lifespan *per se* or each species' unique history of artificial selection.

Trait Covariation Is Significantly Positive in Most Cases

Positive correlations among traits observed in our dataset suggest that some suites of traits are synergistic, including seed size, some early growth traits, and biomass allocation (**Figure 3**). Positive seed-vegetative size correlations have also been found in other herbaceous plant systems (Geber, 1990; Kleyer et al., 2019), as well as across vascular plants more generally (Díaz et al., 2015). The nearly perfect correlation among seed size parameters (weight, area, and length) suggests that both two-dimensional seed features and mass are jointly selected upon and can be used interchangeably for these species. Significant positive correlations between seed size and vegetative biomass are consistent with the fundamental constraint of plant size on seed size, i.e., small plants cannot produce very large seeds (Venable and Rees, 2009). Furthermore, positive correlations between stem diameter and seed size support the hypothesis that seed size is biomechanically limited by the size of the subtending branch (Aarssen, 2005; Venable and Rees, 2009). Seed size alone may be less reflective of whole-plant reproductive allocation (Vico et al., 2016) and more so of a tolerance-fecundity tradeoff, where larger seeds are produced in fewer numbers but have greater competitive ability in stressful conditions (Muller-Landau, 2010). Nevertheless, the perennial grain crop *Thinopyrum intermedium* showed a positive relationship of total biomass and single seed mass with total reproductive yield across three years of growth, suggesting that these traits are positively correlated in some perennial species (Cattani and Asselin, 2018; also see Kleyer et al., 2019). Lastly, significant positive associations between all dry biomass traits do not support aboveground-belowground tradeoffs; rather, the data suggest that larger plants require greater resource acquisition with proportionally larger roots. This is consistent with a previous study reporting positive correlations among shoot dry mass, root dry mass, and root mass fraction in *Phaseolus vulgaris* (Berny Mier y Teran et al., 2018).

There were a few notable exceptions to the general trend of positive correlations. There were unexpectedly no significant correlations between seed size and germination proportion considering the entire dataset, with only slight negative correlations detected. There is a theoretical expectation that germination dormancy and seed bank persistence are negatively related to seed size, since dormancy and seed size entail different strategies of surviving in different environmental conditions, i.e., small, dormant seeds dominate seasonal environments and large, non-dormant seeds are typical of aseasonal environments (Baskin and Baskin, 2014; Rubio de Casas et al., 2017). The annual and wild subsets showed greater negative seed size-germination correlations (significant for germination vs. seed weight for annuals), contrasting with positive correlations in the perennial and cultivated subsets, suggesting some association of lifespan and cultivation with this trait relationship (**Supplementary Figure S1**). The removal of the physical dormancy barrier through scarification must also be considered, although we may still expect differences in seed longevity in storage proportional to seed size.

There was also a general lack of significant covariation between early vegetative stem growth traits (stem height and node number) and most biomass traits (**Figure 3**). This is in contrast to a previous study reporting that stem height is a highly interconnected hub trait in herbaceous perennial species (Kleyer et al., 2019), although we find many of the same positive associations between vegetative and seed size traits. The significant negative correlations detected in annual species between seed size traits and node number may reflect tradeoffs between seed size and stem length allocation due to the biomechanical restraints of bearing larger seeds on a longer stem (Kleyer et al., 2019) or may be the result of reduced stem growth during artificial selection (see above discussion), although this correlation was positive in all other data subsets (**Supplementary Figure S1**).

Overall, positive correlations among traits measured here suggests that future breeding efforts targeting greater seed size

in perennials may be possible without concomitant reductions in vegetative allocation. Consistently high correlations may also allow for measurement of simpler traits (e.g., stem diameter, shoot dry mass) as proxies for traits that are more difficult to assess (e.g., root dry mass, total yield).

CONCLUSION

For this set of *Phaseolus* species, we found that cultivation is associated with an increase in seed size and overall vegetative size in both annuals and perennials, and that seed size and most vegetative traits were positively correlated in both cultivated and wild annuals and perennials. Traits are shaped and predicted by more specific factors than lifespan alone, including habitat differences, evolutionary history, and which specific organs were targeted during artificial selection. Resolving these factors will require a focused sample of close sister species from the same native range, or intraspecific ecotypes of differing lifespan. Also, all perennials in our study were analyzed in their first year of growth, while their overall life history strategy is contingent upon their survival across multiple years. Thus, this study may also be augmented by multiyear assessments of reproductive and vegetative trait variation in perennial individuals. Such studies will advance our basic knowledge of life history evolution and inform plant breeding as we determine viable methods of ecological intensification. Conclusively, we highlight here common features in cultivated annual and perennial species compared to their wild relatives, and we observed few tradeoffs among seed and vegetative traits in the first year of growth. This offers insight into how perennial legume crops may respond to artificial selection relative to their annual relatives, and it suggests that many traits of interest may be selected for in concert.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/ **Supplementary Material**.

AUTHOR CONTRIBUTIONS

AM and SH designed the project. SH implemented the project and wrote the manuscript, with significant input from AM and MR. MR assisted in statistical analysis and figure generation. CC, TC, and DT provided extensive technical assistance, helpful feedback on the basis and design of the study, and revisions to the manuscript. All authors read and approved the final version of the manuscript.

FUNDING

This research was funded by the Perennial Agriculture Project (Malone Family Land Preservation Foundation and The Land Institute). SH is supported by a graduate assistantship from Saint Louis University. MR is supported by the Donald Danforth Plant Science Center and the Perennial Agriculture Project.

ACKNOWLEDGMENTS

The Land Institute provided research tools and space for this project. Seeds were provided by the United States Department of Agriculture Western Regional PI Station (Pullman, WA). We are especially grateful to Daniel Debouck and Colin Khoury for their extensive feedback on *Phaseolus* species' biology and lifespan assignment. We thank the following Land Institute researchers and associates for valuable input in experimental design and intellectual contributions: Lee R. DeHaan, Matthew Newell, Damian A. Ravetta, Alejandra Vilela, and Shuwen Wang. We extend a special thanks to The Land Institute interns, technicians, and associates who assisted in experimental set-up, tending plants, and measuring traits: Mindelena Adams, Sheila Cox, Eliot Cusick, Tiffany Durr, Nick Feijen, Katherine Fortin, Maya Kathrineberg, Laura Kemp, Ron Kinkelar, Jordan Lowry, Maged Nosshi, Diego Sanchez, Codie Van de Meter, and Eline Van de Ven. Brandon Schlautman and REU student Dahlia Martinez also assisted in phenotyping and plant caretaking. We are grateful to the Miller Lab Group for valuable comments on previous versions of this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.00034/full#supplementary-material

REFERENCES

Aarssen, L. W. (2005). Why don't bigger plants have proportionately bigger seeds? *Oikos* 111, 199–207. doi: 10.1111/j.0030-1299.2005.14206.x

Abbo, S., Pinhasi, R., Gopher, A., Saranga, Y., Ofner, I., and Peleg, Z. (2014). Plant domestication versus crop evolution: a conceptual framework for cereals and grain legumes. *Trends Plant Sci.* 19, 351–360. doi: 10.1016/j.tplants.2013.12.002

Aragao, F. J., Brondani, R. P., and Burle, M. L. (2011). "*Phaseolus.*" in *Wild Crop Relatives: Genomic and Breeding Resources. Legume Crops and Forages*, *1st ed.* Ed. C. Kole (Springer-Verlag Berlin Heidelberg), 223–236. doi: 10.1007/978-3-642-14387-8

Baskin, C. C., and Baskin, J. M. (2014). *Seeds: Ecology, Biogeography, and Evolution of Dormancy and Germination.* (San Diego, CA: Academic Press), 1600.

Bayuelo-Jiménez, J. S., Craig, R., and Lynch, J. P. (2002). Salinity tolerance of *Phaseolus* species during germination and early seedling growth. *Crop Sci.* 42, 1584–1594. doi: 10.2135/cropsci2002.1584

Berny Mier y Teran, J. C., Konzen, E. R., Medina, V., Palkovic, A., Ariani, A., Tsai, S. M., et al. (2018). Root and shoot variation in relation to potential intermittent drought adaptation of Mesoamerican wild common bean (*Phaseolus vulgaris* L.). *Ann. Bot.* 20, mcy221. doi: 10.1093/aob/mcy221

Bharucha, Z., and Pretty, J. (2010). The roles and values of wild foods in agricultural systems. *Philos. Trans. R. Soc. B.: Biol. Sci.* 365, 2913–2926. doi: 10.1098/rstb.2010.0123

Bitocchi, E., Rau, D., Bellucci, E., Rodriguez, M., Murgia, M. L., Gioia, T., et al. (2017). Beans (*Phaseolus* ssp.) as a model for understanding crop evolution. *Front. Plant Sci.* 8, 722. doi: 10.3389/fpls.2017.00722

Breseghello, F., and Coelho, A. S. G. (2013). Traditional and modern plant breeding methods with examples in rice (Oryza sativa L.). *J. Agric. Food Chem.* 61, 8277–8286. doi: 10.1021/jf305531j

Buhrow, R. (1983). The wild beans of southwestern North America. *Desert Plants* 5 (2), 67–88.

Cattani, D. J., and Asselin, S. R. (2018). Has selection for grain yield altered intermediate wheatgrass? *Sustainability* 10, 688. doi: 10.3390/su10030688

Cattani, D. J. (2017). Selection of a perennial grain for seed productivity across years: intermediate wheatgrass as a test species. *Can. J. Plant Sci.* 524, 516–524. doi: 10.1139/cjps-2016-0280

Chacón-Sánchez, M. I. (2018). "The domestication syndrome in *Phaseolus* crop plants: A review of two key domestication traits," in *Origin and Evolution of Biodiversity*. Ed. P. Pontarotti (Cham, Switzerland: Springer), 37–59. doi: 10.1007/978-3-319-95954-2

Ciotir, C., Townesmith, A., Applequist, W., Herron, S., Van Tassel, D., DeHaan, L., et al. (2016). Global Inventory and Systematic Evaluation of Perennial Grain, Legume, and Oilseed Species for Pre-breeding and Domestication. [http://www.tropicos.org/Project/PAPGI: month, 2016]. Missouri Botanical Garden, St. Louis, Missouri, USA.

Ciotir, C., Applequist, W., Crews, T. E., Cristea, N., DeHaan, L. R., Frawley, E., et al. (2019). Building a botanical foundation for perennial agriculture: Global inventory of wild, perennial herbaceous Fabaceae species. *Plants People Planet* 1, 375–386. doi: 10.1002/ppp3.37

Cohen, D. (1966). Optimizing reproduction in a randomly varying environment. *J. Theor. Biol.* 12, 119–129. doi: 10.1016/0022-5193(66)90188-3

Cole, L. C. (1954). The population consequences of life history phenomena. *Q. Rev. Biol.* 29, 103–137. doi: 10.1086/400074

Cox, T. S., Glover, J. D., Van Tassel, D. L., Cox, C. M., and DeHaan, L. R. (2006). Prospects for developing perennial grain crops. *BioScience* 56, 649–659. doi: 10.1641/0006-3568(2006)56

Cox, S., Nabukalu, P., Paterson, A. H., Kong, W., and Nakasagga, S. (2018). Development of perennial grain sorghum. *Sustainability* 10, 172. doi: 10.3390/su10010172

Crews, T. E., and Cattani, D. J. (2018). Strategies, advances, and challenges in breeding perennial grain crops. *Sustainability* 10, 2192. doi: 10.3390/su10072192

Crews, T. E., and DeHaan, L. R. (2015). The strong perennial vision: a response. *Agroecol. Sustain. Food Syst.* 39, 500–515. doi: 10.1080/21683565.2015.1008777

Crews, T. E., Carton, W., and Olsson, L. (2018). Is the future of agriculture perennial? Imperatives and opportunities to reinvent agriculture by shifting from annual monocultures to perennial polycultures. *Global Sustainability* 1, e11. doi: 10.1017/sus.2018.11

Culman, S. W., Snapp, S. S., Ollenburger, M., Basso, B., and DeHaan, L. R. (2013). Soil and water quality rapidly responds to the perennial grain Kernza wheatgrass. *Agron. J.* 105, 735–744. doi: 10.2134/agronj2012.0273

Díaz, S., Kattge, J., Cornelissen, J. H. C., Wright, I. J., Lavorel, S., Dray, S., et al. (2015). The global spectrum of plant form and function. *Nature* 529, 167–171. doi: 10.1038/nature16489

De Souza, J. G., and Vieira Da Silva, J. (1987). Partitioning of carbohydrates in annual and perennial cotton (*Gossypium hirsutum* L.). *J. Exp. Bot.* 38, 1211–1218. doi: 10.1093/jxb/38.7.1211

Debouck, D. G. (1992). "Frijoles, *Phaseolus* spp," in *Cultivos marginados: otra perspectiva de 1492*. Eds. E. Hernández Bermejo and J. León (Rome, Italy: Food and Agriculture Organization of the United Nations), 45–60.

DeHaan, L. R., Van Tassel, D. L., Anderson, J. A., Asselin, S. R., Barnes, R., Baute, G. J., et al. (2016). A pipeline strategy for grain crop domestication. *Crop Sci.* 56, 1–14. doi: 10.2135/cropsci2015.06.0356

Delgado-Salinas, A., Bibler, R., and Lavin, M. (2006). Phylogeny of the genus *Phaseolus* (Leguminosae): A recent diversification in an ancient landscape. *Syst. Bot.* 31, 779–791. doi: 10.1600/036364406779695960

Delgado-Salinas, A. (1988). "Variation, taxonomy, domestication, and germplasm potentialities in *Phaseolus coccineus*.," in *Genetic Resources of Phaseolus Beans*. Ed. P. Gepts (Norwell, MA: Kluwer Academic Publishers), 441–463. doi: 10.1007/978-94-009-2786-5_18

Denison, R. F. (2012). "What won't work: Misguided mimicry of natural ecosystems," in *Darwinian Agriculture: How Understanding Evolution Can Improve Agriculture* (Princeton, NJ: Princeton University Press), 95–119.

Dohle, S., Carlos, J., Mier, B., Egan, A., Kisha, T., and Khoury, C. K. (2019). "Wild Beans (*Phaseolus* L.) of North America," in *North American Crop Wild Relatives*, vol. 2 . Eds. S. L. Greene, K. A. Williams, C. K. Khoury, M. B. Kantar and L. F. Marek (Cham, Switzerland: Springer), 99–127. doi: 10.1007/978-3-319-97121-6

Dohleman, F. G., and Long, S. P. (2009). More productive than maize in the Midwest: How does *Miscanthus* do it? *Plant Physiol.* 150, 2104–2115. doi: 10.1104/pp.109.139162

FAO (Food and Agriculture Organization of the United Nations) (2019). *Outcome document of the Global Symposium on Soil Erosion*. Rome, Italy.

Freytag, G. F., and Debouck, D. G. (2002). *Taxonomy, distribution, and ecology of the genus Phaseolus (Leguminosae-Papilionoideae) in North America, Mexico, and Central America* (Fort Worth, TX: Botanical Research Institute of Texas (BRIT)).

Fuller, D. Q., and Allaby, R. (2009). Seed dispersal and crop domestication: Shattering, germination and seasonality in evolution under cultivation. *Annu. Plant Rev.* 38, 238–295. doi: 10.1002/9781444314557.ch7

Garnier, E. (1992). Growth analysis of congeneric annual and perennial grass species. *J. Ecol.* 80, 665–675. doi: 10.2307/2260858

Gaut, B. S., Díez, C. M., and Morrell, P. L. (2015). Genomics and the contrasting dynamics of annual and perennial domestication. *Trends Genet.* 31, 709–719. doi: 10.1016/j.tig.2015.10.002

Geber, M. A. (1990). The cost of meristem limitation in *Polygonum arenastrum*: Negative genetic correlations between fecundity and growth. *Evolution* 44, 799–819. doi: 10.1111/j.1558-5646.1990.tb03806.x

González-Paleo, L., and Ravetta, D. A. (2015). Carbon acquisition strategies uncoupled from predictions derived from species life-cycle. *Flora* 212, 1–9. doi: 10.1016/j.flora.2015.02.004

González-Paleo, L., Vilela, A. E., and Ravetta, D. A. (2016). Back to perennials: Does selection enhance tradeoffs between yield and longevity? *Ind. Crops Products* 91, 272–278. doi: 10.1016/j.indcrop.2016.07.018

Gremer, J. R., Kimball, S., and Venable, D. L. (2016). Within-and among-year germination in Sonoran Desert winter annuals: bet hedging and predictive germination in a variable environment. *Ecol. Lett.* 19, 1209–1218. doi: 10.1111/ele.12655

Guerra-García, A., Suárez-Atilano, M., Mastretta-Yanes, A., Delgado-Salinas, A., and Piñero, D. (2017). Domestication genomics of the open-pollinated scarlet runner bean (*Phaseolus coccineus* L.). *Front. Plant Sci.* 8, 1891. doi: 10.3389/fpls.2017.01891

Hammer, K., and Khoshbakht, K. (2015). A domestication assessment of the big five plant families. *Genet. Resour. Crop Evol.* 62, 665–689. doi: 10.1007/s10722-014-0186-2

Hammer, K. (1984). Das Domestikationssyndrom. *Kulturpflanze* 32, 11–34. doi: 10.1007/BF02098682

Harlan, J. R., de Wet, J. M. J., and Price, E. G. (1973). Comparative evolution of cereals. *Evolution* 27, 311–325. doi: 10.2307/2406971

Harlan, J. R. (1992). *Crops & man* (Madison, WI, USA: American Society of Agronomy).

Harlan, J. R. (1995). *The Living Fields: Our Agricultural Heritage* (Cambridge, UK: Cambridge University Press).

Harper, J. L. (1977). *Population Biology of Plants* (University of California: Academic Press).

Harris, D. R. (1989). "An evolutionary continuum of people–plant interaction," in *Foraging and farming: the evolution of plant exploitation*. Eds. D. R. Harris and G. C. Hillman (London, UK: Unwin Hyman Ltd.), 11–26. doi: 10.4324/9781315746425

Huang, G., Qin, S., Zhang, S., Cai, X., Wu, S., Dao, J., et al. (2018). Performance, economics and potential impact of perennial rice PR23 relative to annual rice cultivars at multiple locations in Yunnan Province of China. *Sustainability* 10, 1086. doi: 10.3390/su10041086

Jaikumar, N. S., Snapp, S. S., Murphy, K., and Jones, S. S. (2012). Agronomic assessment of perennial wheat and perennial rye as cereal crops. *Agron. J.* 104, 1716–1726. doi: 10.2134/agronj2012.0291

Jordan, N., and Warner, K. D. (2010). Enhancing the multifunctionality of US agriculture. *BioScience* 60, 60–66. doi: 10.1525/bio.2009.60.1.10

Kane, D. A., Rogé, P., and Snapp, S. S. (2016). A systematic review of perennial staple crops literature using topic modeling and bibliometric analysis. *PloS One* 11, e0155788. doi: 10.1371/journal.pone.0155788

Kantar, M., Tyl, C., Dorn, K., Zhang, X., Jungers, J., Kaser, J. M., et al. (2016). Perennial grain and oilseed crops. *Annu. Rev. Plant Biol.* 67, 1–27. doi: 10.1146/annurev-arplant-043015-112311

Kaplan, L., and Lynch, T. F. (1999). *Phaseolus* (Fabaceae) in archaeology: AMS radiocarbon dates and their significance for pre-Colombian agriculture. *Econ. Bot.* 53 (3), 261–272. doi: 10.1007/BF02866636

Kleyer, M., Trinogga, J., Cebrián, M. A., Ejrnæs, R., Trenkamp, A., Fløjgaard, C., et al. (2019). Trait correlation network analysis identifies biomass allocation traits and stem specific length as hub traits in herbaceous perennial plants. *J. Ecol.* 107, 829–842. doi: 10.1111/1365-2745.13066

Koinange, E. M. K., Singh, S. P., and Gepts, P. (1996). Genetic control of the domestication syndrome in common bean. *Crop Sci.* 36, 1037–1045. doi: 10.2135/cropsci1996.0011183X003600040037x

López Herrera, M., Aguirre Rivera, J. R., Trejo, C., and Peña-Valdivia, C. B. (2001). Differences in seed germination of wild and domesticated common bean (*Phaseolus vulgaris* L.) in response to storage. *South Afr. J. Bot.* 67, 620–628. doi: 10.1016/S0254-6299(15)31192-3

Lewis, G., Schrire, B., Mackinder, B., and Lock, M. (2005). *Legumes of the World.* (Surrey, UK: Royal Botanic Garden, Kew).

Mann, C. (1997). Reseeding the green revolution. *Science* 277, 1038–1043. doi: 10.1126/science.277.5329.1038

McKey, D. B., Elias, M., Pujol, B., and Duputié, A. (2012). "Ecological approaches to crop domestication," in *Biodiversity in Agriculture: Domestication, Evolution, and Sustainability*. Eds. P. Gepts, T. R. Famula and R. L. Bettinger, et al (Cambridge, UK: Cambridge University Press), 377–406. doi: 10.1017/CBO9781139019514.023

Meyer, R. S., DuVal, A. E., and Jensen, H. R. (2012). Patterns and processes in crop domestication: an historical review and quantitative analysis of 203 global food crops. *New Phytol.* 196, 29–48. doi: 10.1111/j.1469-8137.2012.04253.x

Milla, R., and Matesanz, S. (2017). Growing larger with domestication: a matter of physiology, morphology or allocation? *Plant Biol.* 19, 475–483. doi: 10.1111/plb.12545

Milla, R., Morente-Lopez, J., Alonso-Rodrigo, J., Martin-Robles, N., and Chapin, F. (2014). Shifts and disruptions in resource-use trait syndromes during the evolution of herbaceous crops. *Proc. R. Soc B.* 281, 20141429. doi: 10.1098/rspb.2014.1429

Miller, A. J., and Gross, B. L. (2011). From forest to field: perennial fruit crop domestication. *Am. J. Bot.* 98, 1389–1414. doi: 10.3732/ajb.1000522

Mina-Vargas, A. M., McKeown, P. C., Flanagan, N. S., Debouck, D. G., Kilian, A., Hodkinson, T. R., et al. (2016). Origin of year-long bean (*Phaseolus dumosus* Macfady, Fabaceae) from reticulated hybridization events between multiple *Phaseolus* species. *Ann. Bot.* 118, 957–969. doi: 10.1093/aob/mcw138

Moles, A. T., Ackerly, D. D., Tweddle, J. C., Dickie, J. B., Smith, R., Leishman, M. R., et al. (2007). Global patterns in seed size. *Global Ecol. Biogeogr.* 16, 109–116. doi: 10.1111/j.1466-822x.2006.00259.x

Muller-Landau, H. C. (2010). The tolerance – fecundity trade-off and the maintenance of diversity in seed size. *Proc. Natl. Acad. Sci.* 107, 4242–4247. doi: 10.1073/pnas.0911637107

Nabhan, G. P., and Felger, R. S. (1985).Wild desert relatives of crops: their direct uses as food, in: *Plants for arid lands: proceedings of the Kew International Conference on Economic Plants for Arid Lands held in the Jodrell Laboratory, Royal Botanical Gardens*, London/Boston. pp. 19–33, In Allen & Unwin (Eds.). doi: 10.1007/978-94-011-6830-4_3

Olsen, K. M., and Wendel, J. F. (2013). Crop plants as models for understanding plant adaptation and diversification. *Front. Plant Sci.* 4, 290. doi: 10.3389/fpls.2013.00290

Pastor-Pastor, A., Vilela, A. E., and González-Paleo, L. (2018). The root of the problem of perennials domestication: is selection for yield changing key root system traits required for ecological sustainability? *Plant Soil* 435, 161–174. doi: 10.1007/s11104-018-3885-1

R Core Team (2019). *R: A language and environment for statistical computing* (Vienna, Austria: R Foundation for Statistical Computing). URL https://www.R-project.org/.

Rasband, W. S. (1997-2016). ImageJ. U. S. National Institutes of Health, Bethesda, Maryland, USA, https://imagej.nih.gov/ij/.

Rubio de Casas, R., Willis, C. G., Pearse, W. D., Baskin, C. C., Baskin, J. M., and Cavender-Bares, J. (2017). Global biogeography of seed dormancy is determined by seasonality and seed size: a case study in the legumes. *New Phytol.* 214, 1527–1536. doi: 10.1111/nph.14498

Ryan, M. R., Crews, T. E., Culman, S. W., Dehaan, L. R., Hayes, R. C., Jungers, J. M., et al. (2018). Managing for multifunctionality in perennial grain crops. *BioScience* 68, 294–304. doi: 10.1093/biosci/biy014

Sacks, E. J., Dhanapala, M. P., Sta. Cruz, M. T., and Sallan, R. (2007). Clonal performance of perennial *Oryza sativa*/*O. rufipogon* selections and their combining ability with *O. sativa* cultivars for survival, stolon production and yield. *Field Crops Res.* 100, 155–167. doi: 10.1016/j.fcr.2006.06.003

Schlautman, B., Barriball, S., Ciotir, C., Herron, S. A., and Miller, A. J. (2018). Perennial grain legume domestication phase I: Criteria for candidate species selection. *Sustainability* 10, 730. doi: 10.3390/su10030730

Schmit, V., and Debouck, D. G. (1991). Observations on the origin of *Phaseolus polyanthus* Greenman. *Econ. Bot.* 45 (3), 345–364. doi: 10.1007/BF02887077

Smýkal, P., Nelson, M. N., Berger, J. D., and von Wettberg, E. J. (2018). The impact of genetic changes during crop domestication. *Agronomy* 8, 119. doi: 10.3390/agronomy8070119

Smaje, C. (2015). The strong perennial vision: a critical review. *Agroecol. Sustain. Food Syst.* 39, 471–499. doi: 10.1080/21683565.2015.1007200

Smartt, J. (1988). "Morphological, physiological, and biochemical changes in *Phaseolus* beans under domestication," in *Genetic Resources of Phaseolus Beans*. Ed. P. Gepts (Norwell, MA: Kluwer Academic Publishers), 143–161. (Ed.),. doi: 10.1007/978-94-009-2786-5_8

Suneson, C. A., Sharkawy, E., and Hall, W. E. (1963). Progress in 25 years of perennial wheat development. *Crop Sci.* 3, 437–439. doi: 10.2135/cropsci1963.0011183x000300050021x

Thompson, K., and Hodkinson, D. J. (1998). Seed mass, habitat and life history: A re-analysis of Salisbury, (1942, 1974). *New Phytol.* 138 (1), 163–167. doi: 10.1046/j.1469-8137.1998.00886.x

Thompson, K., Bakker, J. P., Bekker, R. M., and Hodgson, J. G. (1998). Ecological correlates of seed persistence in soil in the north-west European flora. *J. Ecol.* 86 (1), 163–169. doi: 10.1046/j.1365-2745.1998.00240.x

Van Tassel, D. L., Dehaan, L. R., and Cox, T. S. (2010). Missing domesticated plant forms: can artificial selection fill the gap? *Evol. Appl.* 3, 434–452. doi: 10.1111/j.1752-4571.2010.00132.x

Venable, D. L., and Lawlor, L. (1980). Delayed germination and dispersal in desert annuals: Escape in space and time. *Oecologia* 46, 272–282. doi: 10.1007/BF00540137

Venable, D. L., and Rees, M. (2009). The scaling of seed size. *J. Ecol.* 97, 27–31. doi: 10.1111/j.1365-2745.2008.01461.x

Verboom, G. A., Linder, H. P., and Stock, W. D. (2004). Testing the adaptive nature of radiation: Growth form and life history divergence in the African grass genus *Ehrharta* (Poaceae: Ehrhartoideae). *Am. J. Bot.* 91, 1364–1370. doi: 10.3732/ajb.91.9.1364

Vico, G., Manzoni, S., Nkurunziza, L., Murphy, K., and Weih, M. (2016). Trade-offs between seed output and life span - a quantitative comparison of traits between annual and perennial congeneric species. *New Phytol.* 209, 104–114.

Vilela, A. E., and González-Paleo, L. (2015). Changes in resource-use strategy and phenotypic plasticity associated with selection for yield in wild species native to arid environments. *J. Arid Environ.* 113, 51–58. doi: 10.1111/nph.13574

Walters, C., Wheeler, L. M., and Grotenhuis, J. M. (2005). Longevity of seeds stored in a genebank: species characteristics. *Seed Sci. Res.* 15, 1–20. doi: 10.1079/SSR2004195

Weeden, N. F. (2007). Genetic changes accompanying the domestication of *Pisum sativum*: Is there a common genetic basis to the 'domestication syndrome' for legumes? *Ann. Bot.* 100, 1017–1025. doi: 10.1093/aob/mcm122

Zohary, D., and Spiegel-Roy, P. (1975). Beginnings of fruit growing in the Old World. *Science* 187, 319–327. doi: 10.1126/science.187.4174.319

Conflict of Interest: 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 © 2020 Herron, Rubin, Ciotir, Crews, Van Tassel and Miller. 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) and the copyright owner(s) 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 Plant Science

ORIGINAL RESEARCH
published: 24 March 2020
doi: 10.3389/fpls.2020.00319

Enhancing Crop Domestication Through Genomic Selection, a Case Study of Intermediate Wheatgrass

Jared Crain[1], Prabin Bajgain[2], James Anderson[2], Xiaofei Zhang[3], Lee DeHaan[4] and Jesse Poland[1]**

[1] Department of Plant Pathology, Kansas State University, Manhattan, KS, United States, [2] Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN, United States, [3] The Alliance of Bioversity International and International Center for Tropical Agriculture, Cali, Colombia, [4] The Land Institute, Salina, KS, United States

Perennial grains could simultaneously provide food for humans and a host of ecosystem services, including reduced erosion, minimized nitrate leaching, and increased carbon capture. Yet most of the world's food and feed is supplied by annual grains. Efforts to domesticate intermediate wheatgrass (*Thinopyrumn intermedium*, IWG) as a perennial grain crop have been ongoing since the 1980's. Currently, there are several breeding programs within North America and Europe working toward developing IWG into a viable crop. As new breeding efforts are established to provide a widely adapted crop, questions of how genomic and phenotypic data can be used among sites and breeding programs have emerged. Utilizing five cycles of breeding data that span 8 years and two breeding programs, University of Minnesota, St. Paul, MN, and The Land Institute, Salina, KS, we developed genomic selection (GS) models to predict IWG traits. Seven traits were evaluated with free-threshing seed, seed mass, and non-shattering being considered domestication traits while agronomic traits included spike yield, spikelets per inflorescence, plant height, and spike length. We used 6,199 genets – unique, heterozygous, individual plants – that had been profiled with genotyping-by-sequencing, resulting in 23,495 SNP markers to develop GS models. Within cycles, the predictive ability of GS was high, ranging from 0.11 to 0.97. Across-cycle predictions were generally much lower, ranging from −0.22 to 0.76. The prediction ability for domestication traits was higher than agronomic traits, with non-shattering and free threshing prediction abilities ranging from 0.27 to 0.75 whereas spike yield had prediction abilities ranging from −0.22 to 0.26. These results suggest that progress to reduce shattering and increase the percent free-threshing grain can be made irrespective of the location and breeding program. While site-specific programs may be required for agronomic traits, synergies can be achieved in rapidly improving key domestication traits for IWG. As other species are targeted for domestication, these results will aid in rapidly domesticating new crops.

Keywords: intermediate wheatgrass, genomic selection, multi-environment, domestication, perennial crops, shared data resources

OPEN ACCESS

Edited by:
Eric Von Wettberg, University of Vermont, United States

Reviewed by:
Ken Naito, National Agriculture and Food Research Organization (NARO), Japan
Steven B. Cannon, Agricultural Research Service, United States Department of Agriculture, United States

****Correspondence:***
Lee DeHaan
dehaan@landinstitute.org
Jesse Poland
jpoland@ksu.edu

Specialty section:
This article was submitted to Plant Breeding, a section of the journal Frontiers in Plant Science

Received: *06 January 2020*
Accepted: *04 March 2020*
Published: *24 March 2020*

Citation:
Crain J, Bajgain P, Anderson J, Zhang X, DeHaan L and Poland J (2020) Enhancing Crop Domestication Through Genomic Selection, a Case Study of Intermediate Wheatgrass. Front. Plant Sci. 11:319. doi: 10.3389/fpls.2020.00319

Abbreviations: BLUPs, best linear unbiased predictors; GBS, genotyping-by-sequencing; GS, genomic selection; IWG, intermediate wheatgrass; PCA, principal component analysis; QTL, quantitative trait loci; SNP, single nucleotide polymorphism; TLI, The Land Institute.

INTRODUCTION

Currently, 80% of the world's calories are provided by annual crops (Pimentel et al., 2012), with only three crops, maize (*Zea mays*), wheat (*Triticum aestivum*), and rice (*Oryza sativa*), providing nearly 60% of human calorie consumption (fao.org). Additionally, 70% of arable land is planted to annual crops (Cox et al., 2010; Pimentel et al., 2012) that are resource-intensive and can result in environmental degradation (Power, 2010; Crews et al., 2018). Perennial grain crops could provide abundant ecosystem services while simultaneously providing food, feed, and fuel for the global population. To date, there are no widely planted perennial grain crops, but recent research has resulted in large scale evaluations of perennial rice (Huang et al., 2018) and perennial versions of several other crops including wheat, sorghum (*Sorghum bicolor*), sunflower (*Helianthus*), and pulses are in development (Batello et al., 2013).

One species showing promise for domestication and wide-scale production is intermediate wheatgrass (*Thinopyrum intermedium,* IWG). Intermediate wheatgrass is native to Eastern Europe and the Mediterranean region (Tsvelev, 1983) and was introduced into the United States for erosion control and forage purposes in 1932 (Vogel and Jensen, 2001). This species was selected for domestication as a grain crop in the 1980's from an evaluation of nearly 100 perennial grasses based on its seed size, vigorous growth habit, and potential for mechanical harvest, among other desirable characteristics, at the Rodale Institute, Kutztown, PA (Wagoner, 1990). In the early 2000's after two cycles of selection at the USDA's Big Flats Plant Materials Center, Corning, NY, breeding efforts shifted to The Land Institute (TLI), Salina, KS (Zhang et al., 2016). Since 2003, nine cycles of selection have been completed at TLI. Interest in IWG has led to the development of several other breeding programs, including the University of Minnesota (UMN) and University of Manitoba in 2011 using material from the third cycle of selection from TLI (TLI-C3) (Zhang et al., 2016). Products made from IWG grain are being sold under the trade name Kernza in limited markets (DeHaan and Ismail, 2017).

Along with intensive breeding effort, IWG has also been evaluated for a host of ecosystem services. Research has shown that IWG can reduce soil nitrate leaching by 86% or more compared to annual wheat crop systems (Culman et al., 2013). Jungers et al. (2019) found that nitrate leaching under perennial grasses, including IWG, were one to two orders of magnitude less than annual maize. IWG has also been reported to have 15 times more root growth and nearly two times the above-ground biomass of annual wheat (Sprunger et al., 2018), which should translate into greater below-ground carbon storage rates. Research also indicates that perennial landscapes have significantly increased and diverse microbial communities, allowing for greater food web complexity and increased nutrient cycling capacity (Culman et al., 2010; Pugliese et al., 2019). While IWG has the potential to provide both food and ecosystem services, factors such as grain yield and ability to mechanically harvest must be improved to an economically viable level for farmers to adopt this new crop.

Breeding new crops from wild species requires domestication, which often utilizes rare allelic mutations to facilitate the development of crops. One common domestication trait has been the prevention of shattering, which enables mechanical harvest. Numerous domestication events have been recorded in barley, rice, and sorghum (Østerberg et al., 2017), with reduction of shattering a hallmark of domestication as the plant becomes more dependent on humans for seed dispersal (Purugganan and Fuller, 2009). Other key traits that have evolved through domestication include larger seed size, free threshing seeds, and an increase in percent seed set (Harlan et al., 1973). Within IWG breeding, key domestication traits being targeted are greater percent of free threshing seeds, reduction in seed shattering, and increased seed mass.

Early work in domestication architecture through quantitative trait loci (QTL) often suggested single or a few genes with large effects (Koinange et al., 1996; Olsen and Wendel, 2013), which would allow for more efficient selection than selection on numerous loci with small effects (Falconer and Mackay, 1996). As molecular tools and studies have improved, there has been increasing evidence that many domestication traits are controlled by numerous loci with small effects. While the exact number of domestication genes is unknown (Meyer and Purugganan, 2013), in maize one study has identified nearly 500 genomic regions that had been under selection for domestication features (Hufford et al., 2012).

While original IWG breeding work utilized recurrent phenotypic selection, modern genetic tools have provided breeders with new options. One of the most promising genetic tools for breeding is genomic selection (GS). Proposed by Meuwissen et al. (2001), GS functions by having dense marker coverage of the entire genome so that each QTL is in linkage disequilibrium with a marker (Goddard and Hayes, 2007). Using a population that has been both phenotyped and genotyped, a model can be developed to predict the phenotype of individuals that have only been genotyped. GS has been shown to increase the rate of genetic gain in animal and plant breeding (Bernardo and Yu, 2007; García-Ruiz et al., 2016; Crossa et al., 2017). Given the predicted benefits of GS, Zhang et al. (2016) evaluated the potential of GS in IWG for the UMN breeding program, and currently, TLI is primarily using GS within their IWG breeding program.

While multiple locations are breeding IWG, there has been limited integration of information between breeding programs. The opportunity to utilize molecular tools like GS across wide environments could open new potentials for faster genetic improvement, specifically by increasing the training population size (VanRaden et al., 2009), integrating more genotypes (Knapp and Bridges, 1990), and taking advantage of correlated environments (Spindel and McCouch, 2016). Genomic selection has been used to improve a variety of polygenic agronomic traits including yield, quality and disease resistance (Rutkoski et al., 2014; Battenfield et al., 2016; Guzman et al., 2016). For crop wild relatives undergoing domestication, there has been less work on the ability of GS to improve key domestication

traits such as shattering and free threshing. Work by Zhang et al. (2016) suggested that GS could be used to improve free threshing in IWG.

Applying GS to IWG across multiple environments could be a very cost-effective and efficient method to increase IWG breeding gains, but even within annual crops there is limited information about multi-site or multi-environment GS compared to single site GS studies. Lopez-Cruz et al. (2015) found that using marker-by-environment interactions resulted in a greater prediction accuracy than using within-environment models. A reaction norm model was used to generate prediction accuracies up to 0.4 in wheat in different environments throughout Kansas (Jarquín et al., 2017). In barley (*Hordeum vulgare*), a multi-environment GS model was shown to increase prediction accuracy 11% over single-environment analysis (Oakey et al., 2016). Resende et al. (2012) found that prediction accuracies in loblolly pine (*Pinus taeda*) were relatively consistent across environments as long as the environments were within the same breeding zone. However, in both of these examples many of the lines had true replication, whereas the IWG programs usually have single genotypes due to the challenges of cloning large numbers of individuals. As IWG breeding expands, the ability to combine data across multiple locations and breeding programs with differing, unreplicated, germplasm could be beneficial to increasing the rate of genetic gain.

Given the need for new crops and the challenges associated with developing perennial crops, this study focused on (1) How data from diverse sites and breeding programs could be combined to improve prediction abilities of models for enhanced selection decisions, (2) The ability of GS to accurately predict traits across a range of environments and traits, with emphasis on differences between domestication and agronomic traits, and (3) How insights gained from IWG breeding could be applied to other potential new crops undergoing domestication.

MATERIALS AND METHODS

Plant Material and Field Establishment

Using terminology consistent with Zhang et al. (2016), we refer to a genet as a unique individual plant with its own genetic makeup. The genets used for this study consisted of the TLI Cycles 6, 7, and 8 (TLI-C6, TLI-C7, TLI-C8) and UMN Cycles 1 and 2 (UMN-C1, UMN-C2) breeding programs. The IWG TLI-C6 consisted of 3,658 genets from 674 full-sib families grown in one site location at Salina, KS (38.7684° N, 97.5664° W) between 2015 and 2017. Genets were established in the fall of 2015 with 91 cm between rows and 61 cm between columns, and phenotypic evaluations were conducted in 2016 and 2017. DeHaan et al. (2018) provide additional details about the TLI-C6 population. TLI-C7 was formed from random intermating between selected TLI-C6 genets. Genomic selection was used in the TLI-C7 generation, and a training population consisting of 1,179 genets from approximately 4,000 genotyped genets, was planted in the fall of 2017. TLI-C8 genets were progeny from selected TLI-C7 individuals and consisted of 988 selected, training population genets from approximately 3,500 genotyped genets, with field planting occurring in the fall of 2018. Both TLI-C7 and C8 were divided into two groups with approximately half of each cycle being planted in an irrigated field, and the other half in a non-irrigated field, providing two contrasting environments for evaluation.

UMN-C1 consisted of 2,560 genets from 66 half-sib families from TLI-C3 material. Genets were established in the field, St. Paul, MN (44.9906° N, 93.1799° W) in the fall of 2011 with field observations in 2012 and 2013. Additional information about the UMN-C1 population can be found in Zhang et al. (2016). The UMN-C2 training population consisted of 372 genets that were established in the fall of 2014 with observations in 2015 and 2016. UMN-C2 was obtained from open-pollination of 48 genets selected from the UMN-C1 population with the best agronomic performance. UMN-C2 consisted of 1,656 genets, but phenotypic observations were only recorded for 372 genets, the training population for GS within the UMN breeding program. In both cycles, genets were planted in a single replication at a distance of 1 m rows and columns, 67 kg ha^{-1} of N was applied in April of each year. Weed control in the plant nurseries was primarily done manually with a one-time application of herbicide Dual II Magnum (S-Metolachlor 82.4%, Syngenta) in April at a rate of 1.2 L ha^{-1}. Experimental genets were surrounded on all sides with IWG plants. While each program is selecting genets for its respective growing region, all original UMN material, i.e., UMN-C1, came from TLI-C3, providing a common genetic link between the programs. All genets were evaluated as single plants with no replication.

Field Evaluations

Field evaluations were completed for several key domestication and agronomic traits including: plant height, spikelets per inflorescence, spike length, spike yield, shattering, seed mass, and free-threshing. Plant height was measured after plants reached physiological maturity and was measured from the ground to the tip of the tallest spike. Shattering was measured on a five point scale, with 0 representing no shattering and 4 representing over 50% shattering by visual observation (DeHaan et al., 2018). Spike length was measured from the peduncle to the tip of the spike, and spikelets per inflorescence represented the average number of spikelets per head. Not all traits were measured for each year and genet, resulting in an unbalanced data set. Of key domestication traits, shattering was the only trait not observed in UMN-C1 and seed mass was not available for UMN-C2; all other traits were recorded in all cycles. In addition, minor differences in data collection between programs were noted. For UMN-C1 free threshing was measured on a four-point categorical scale, while for other years free threshing was estimated on a 0–100 percentage scale. The four-point scale was translated to match the percentage scale. For TLI cycles, spike yield was the mass of clean seed from one head, whereas in UMN cycles spike yield was estimated by weighing the entire seed head. Trait data was measured for 2 years with each year being considered a separate trait, with the

exception of TLI-C8 with first year phenotypic data being recorded in 2019.

Genotyping and Bioinformatic Methods

All genets were profiled using genotyping-by-sequencing following protocols of Poland et al. (2012) using a two enzyme restriction digest with *Pst*I and *Msp*I. Libraries were prepared by multi-plexing 192 samples per GBS library, and all GBS libraries were sequenced on Illumina HiSeq 2500. Single nucleotide polymorphisms (SNPs) were called using the GBS pipeline in Trait Analysis by aSSociation, Evolution, and Linkage (TASSEL) software version 5.2 (Glaubitz et al., 2014) in association with the IWG reference genome (access provided by the *Thinopyrum intermedium* Genome Sequencing Consortium.[1]

Initial SNP discovery resulted in identifying 126,138 SNPs. To identify a final data set, filtering was completed using the following criteria, (1) minor allele frequency greater than 0.01, (2) each SNP was called in 30% or more of the individuals, (3) GBS tags uniquely aligned (one location) to the reference genome to prevent aligning to orthologous sequences, (4) only biallelic SNPs were retained, (5) a minimum read depth of four tags per individual were required to call a homozygote. Using a custom Perl script, homozygotes that had less than four reads per site were set to missing. Heterozygotes were called with a minimum of two contrasting tags. Additionally, any genet that had more than 95% missing SNPs calls was discarded from the analysis, resulting in a final data set of 23,495 SNP loci and 6,199 genets. Any missing genotype calls in the final data set were imputed using Beagle version 4.1 using the default settings (Browning and Browning, 2016).

The STRUCTURE program (Pritchard et al., 2000) was used to evaluate population structure among the 6,199 genets. A subset of 8,011 markers that had minor allele frequency greater than 0.05 and were present in more than 50% of the individuals were used to evaluate population structure. A total of 10 subgroups (K = 1–10) were evaluated using the admixture model with 100,000 reps and the first 25,000 as burn-in. Ten replicates of each value of K were assessed, with Structure Harvester (Earl and vonHoldt, 2012) used to determine the optimal number of K. CLUMPP (version 1.1.2) (Jakobsson and Rosenberg, 2007) was used to evaluate K = 1 and K = 2 through graphically assigning individuals to a cluster. In addition to STRUCTURE, principal component analysis (PCA) was performed on the imputed marker matrix in R (R Core Team, 2017). The PCA results were used to subset genets into two similarity groups based on breeding programs.

Statistical Analysis

A mixed linear model using ASREML version 4.1 (Gilmour et al., 2015) was fit to the data to develop best linear unbiased predictors (BLUPs) for each genet in each cycle. The model consisted of a two-step model, where each cycle was analyzed separately (Piepho et al., 2012), and BLUPs were then combined for GS. The model accounted for the genetic relationships between genets using the realized additive genomic relationship matrix and spatial location by fitting a separate row and column autoregressive order 1 (AR1 × AR1) residual structure for each site. The general form of the mixed model is (Isik et al., 2017):

$$\boldsymbol{y} = \boldsymbol{Xb} + \boldsymbol{Zu} + \boldsymbol{e} \quad (1)$$

where $\boldsymbol{y}$ is a vector of observed phenotypes, $\mathbf{X}$ and $\mathbf{Z}$ are design matrices for fixed and random effects, respectively, $\boldsymbol{b}$ and $\boldsymbol{u}$ are vectors of coefficients for fixed and random effects, and $\boldsymbol{e}$ is a vector of random residuals. The vector $\boldsymbol{y}$ is assumed to be distributed normally with mean $\boldsymbol{Xb}$ and variance $\mathbf{V}$, $\boldsymbol{y} \sim N(\boldsymbol{Xb}, \boldsymbol{V})$. The total variance, $\mathbf{V}$, is defined as $\mathbf{V} = \binom{u}{e} = \binom{G\ 0}{0\ R}$. The $\mathbf{G}$ structure accounts for the variation between genets using the realized additive genomic relationship matrix and is defined as $\mathbf{G} = \sigma_A^2\mathbf{K}$ where σ_A^2 is the additive genetic variance and $\mathbf{K}$ is the realized additive genomic relationship matrix. $\mathbf{K}$ is computed as $\theta\mathbf{MM}$' where $\mathbf{M}$ is a matrix with n individuals and m columns of markers and θ is a proportionality constant (Endelman and Jannink, 2012). The genomic relationship matrix was computed using the function *A.mat* in *rrblup* (Endelman, 2011) R package using the methods of Endelman and Jannink (2012). The $\mathbf{R}$ structure accounts for residual variation using the row-column design for each cycle. The $\mathbf{R}$ for each site was defined as $\mathbf{R} = \sigma_e^2\Sigma_c(\rho_c) \otimes \sigma_e^2\Sigma_r(\rho_r)$, fitting an AR1 row and AR1 column effect with an independent error variance for each site. A total of seven sites were fit, as TLI-C7 and TLI-C8 each had two separate locations, whereas all other cycles were grown in one location. Σ is an identity matrix with dimensions equal to the number of rows or columns (Σ_r, Σ_c) respectively and ρ is the correlation parameter between rows and columns, respectively. A minimum of 350 observations were recorded from each cycle for use in GS models after adjusting phenotypic data for genetic relationship and spatial location in the field.

Genomic Selection

Using the five cycles of data, GS models using the genomic best linear unbiased predictor (GBLUP) were developed to assess prediction ability. Within each cycle, a fivefold cross-validation method was repeated 100 times. For each iteration of the cross-validation, we randomly sampled all of the genets that were in a given cycle, splitting the genets into a training population (80% of genets) and a prediction population (20% of genets). The GS model was fit with the training population using *rrBLUP kin.blup* function (Endelman, 2011), with predictions then being made on the prediction population. The GS model has the form (Endelman, 2011):

$$\boldsymbol{y} = \boldsymbol{Wg} + \boldsymbol{e} \quad (2)$$

where $\boldsymbol{y}$ is a vector of observations (phenotypic BLUPs, section Statistical Analysis), $\boldsymbol{W}$ is a design matrix relating genets to observations, $\boldsymbol{g}$ is a vector of genotypic values, and $\boldsymbol{e}$ is a vector of random residuals. The vector of genotypic values, $\boldsymbol{g}$, is distributed as $g \sim N(0, \mathbf{K}\sigma_g^2)$, where $\mathbf{K}$ is the realized additive relationship matrix and σ_g^2 is the additive genotypic variance.

For each iteration, a random sampling without replacement was used to divide the training and prediction populations. Additionally, the random sampling did not prevent full or

[1]https://phytozome-next.jgi.doe.gov/info/Tintermedium_v2_1

half-siblings from being both in the training and prediction populations, potentially upwardly biasing predictions. Predictive ability was assessed using Pearson correlation between the predicted value (genomic BLUP, GBLUP) and the BLUP for the respective phenotype. From the GS model, variance components were extracted to calculate genomic heritability using the genetic variance and residual error variance using the formula (Endelman and Jannink, 2012):

$$h^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2} \quad (3)$$

where h^2 is narrow-sense heritability, σ_a^2 is genetic variance, and $\sigma_a^2 + \sigma_e^2$ is the sum of genetic and residual variance representing total phenotypic variance.

To evaluate multi-environment predictions, each cycle was used as the training population to predict all other cycles. In this method, each cycle was fit as the training population, and then all other cycle genets were predicted. Using BLUPs for observed traits, accuracy was considered the correlation between the phenotypes and the GBLUPs, with the 95% confidence intervals for the correlation computed using the *psychometric* R package (Fletcher, 2010). Along with predicting all other sites from each site, a model was evaluated with a leave-one-out strategy, where the training population consisted of four cycles, and the final cycle was predicted from the combined training population.

Two other models were developed with the goal of identifying the best ways to use the data sets to increase genetic gain. A subset of data was made using the results of the PCA analysis to create two similar groups, UMN-PCA and TLI-PCA. These models used the 2nd principal component to divide UMN and TLI material (**Figure 1**), with training data only consisting of genets within a respective group. In addition, to developing training sets by genetic similarity, each individual breeding program was used as a prediction set to predict all other cycles. The multi-environment models, where one cycle was predicted from all others, were ran again using these two data subsets to evaluate the effect of using more related training data sets in the prediction model. A minimum of 100 genets were required to be in the training set to make predictions for each model.

RESULTS

Phenotypic Evaluations

We analyzed 8 years of breeding program field trials representing two independent breeding programs and five cycles of selection. Across all sites, several traits were measured, including the key domestication traits of free threshing and shattering and agronomic traits like spike yield and seed mass (**Table 1**). For all of these traits, a large range in observations were observed in all cycles. For example, individuals in most cycles ranged from no shattering to maximum shattering. For agronomic traits, a two or threefold range was present for spike length and spikelets per infloresence (**Table 1**).

Population Structure

We implemented a Bayesian cluster method to estimate population structure. While all genets were derived from TLI breeding material, this study evaluated five cycles of selection at different locations, times, and generations from the base population, allowing for potential population structure. Results from this analysis suggested that there was no population grouping of genets. Further analysis using PCA confirmed minimal population structure as the first principal component contained 3% of the variation and the first 10 components only accounted for 13% of the total variation. There was minor clustering among cycles (**Figure 1**), with the second principal component partially separating the UMN material and later TLI-C6 and C7 breeding programs.

Genomic Selection Models

Within-Cycle Predictive Ability

To evaluate the potential of GS to increase the rate of genetic gain in IWG breeding, we fit several GS models to the phenotypic BLUPs. To determine predictive ability of GS, we fit a random fivefold cross-validation model to each cycle and trait individually. Using 100 iterations, within-cycle prediction ability, correlation between predicted value and the phenotypic BLUP, ranged from 0.11 to 0.97 (**Table 2**). Within cycles, prediction abilities were generally high, with a trend that free threshing percent, seed mass and shattering had higher average within-site prediction compared to agronomic traits like spike yield, plant height, and spikelets per inflorescence.

Across-Cycle Predictive Ability

After confirming that GS could accurately predict traits within cycles, we fit GS models to predict across cycles. For each trait, all cycles were used individually as the training population, and then all other cycles were predicted from the chosen training population. This resulted in predicting each cycle from four different cycles. Across all traits, prediction ability ranged from −0.22 to 0.76, but there were striking differences between traits. For key domestication traits there was relatively high predictive ability with seed shattering in a range of 0.50–0.74, and free threshing had a range of 0.27–0.75. In comparison the agronomic trait of spike yield had a much lower range from -0.22 to 0.26 (**Figure 2**). These traits represent a general trend that was seen among all traits and years, allowing further discussion to be defined to domestication and agronomic traits. All other traits are provided in **Supplementary Figure S1**. Additionally for a trait with high and low predictive ability, scatter plots of predicted versus observed values are provided in **Supplementary Figure S2**.

To further investigate the validity of the across-environment GS results, we developed GS models that used all cycle data except for the prediction set. This resulted in a larger training population which could increase GS accuracy. Prediction accuracy based on all other sites ranged from 0.35 to 0.77 for domestication traits (**Figure 3**). Agronomic traits such as spike yield ranged from -0.10 to 0.37 (**Table 3**). The predictions from this leave-one-out strategy were paired with the genomic heritability that was calculated from the GS models. Plotting these two values showed

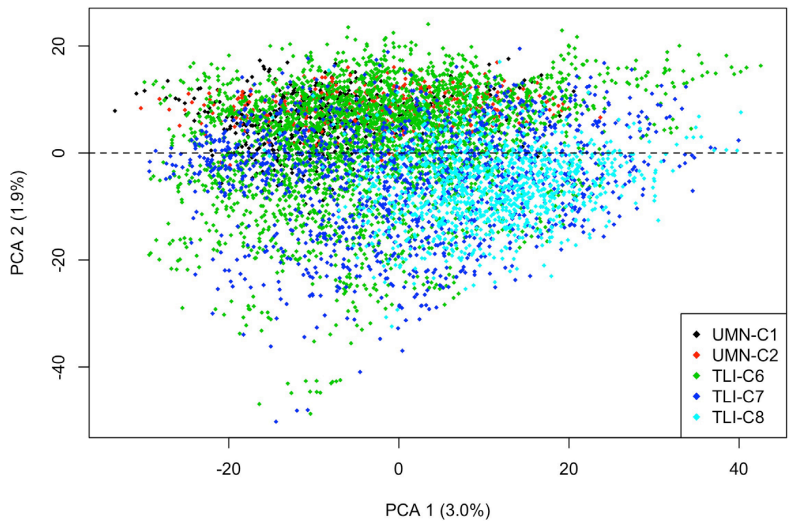


FIGURE 1 | Scatterplot of the first two principal component axis for intermediate wheatgrass genets, made from principal component analysis on the marker matrix, *n* = 6,199 genets, markers = 23,495. Each point is an individual genet that is color coded by cycle, with the 2nd principal component providing separation between the UMN and TLI breeding programs at the dashed line. Total variance explained by each principal component is listed on the axis.

TABLE 1 | Range of phenotypic observations collected for five cycles of intermediate wheatgrass breeding trials from the University of Minnesota and The Land Institute.

		Free threshing %		Plant height (cm)		Seed Mass (mg)		Shattering 0–4		Spikelets per Inflorescence		Spike length (cm)		Spike yield (g)	
Cycle	**Year**	**n**	**Range**	**n**	**Range**	**n**	**Range**	**n**	**Range**	**n**	**Range**	**n**	**Range**	**n**	**Range**
UMN-C1	2012	494	3–75†	494	61–170	494	3.7–13.4			494	9–32	494	17–38	494	0.4–2.2‡
	2013	477	3–75†	491	76–185	474	3.7–13.9			484	10–32	485	17–36	484	0.1–2.0‡
UMN-C2	2015	372	0–95	372	70–182			372	0–4			372	13–41	372	0.4–2.0‡
	2016	356	0–80	368	104–190			360	0–4			367	19–37	366	0.4–1.7‡
TLI-C6	2016	2496	0–100	2482	30–180	2494	0.8–13.7	2507	0–4	2507	9–43	2506	13–60	2508	0.0–0.8
	2017	1691	0–100	1278	80–170	1690	2.5–13.3	1719	0–5	1714	12–39	1707	12–53	1723	0.0–0.7
TLI-C7	2018	851	0–100	1179	35–160	848	4.0–14.6	1140	0–4	1139	8–32	1141	0–52	1140	0.0–0.9
	2019	1164	0–100	1147	80–190	1162	2.9–17.2	1167	0–4	1168	11–32			1168	0.0–1.2
TLI-C8	2019	872	4–100	961	40–140	867	4.8–16.3	873	0–4	873	11–31			870	0.0–1.1

Range of phenotypic observations and the number of individuals (n) for each phenotype are displayed. †*Trait was measured on a five-point categorical scale and converted to percentages.* ‡*Spike yield for UMN sites measured as entire inflorescence, not just clean seed.*

TABLE 2 | Within-site fivefold cross-validation genomic selection predictions for intermediate wheatgrass traits.

		Free Threshing%		Plant Height		Seed Mass		Shattering		Spikelets per Inflorescence		Spike Length		Spike yield	
Training Cycle	**Year**	***r***	**sd**	***r***	**sd**	***r***	**sd**	***r***	**sd**	***r***	**sd**	***r***	**sd**	***r***	**sd**
UMN-C1	2012	0.85	0.03	0.76	0.05	0.79	0.04			0.8	0.04	0.7	0.05	0.76	0.04
	2013	0.85	0.03	0.72	0.05	0.76	0.04			0.77	0.04	0.74	0.04	0.79	0.03
UMN-C2	2014	0.72	0.05	0.73	0.05			0.82	0.03			0.86	0.03	0.76	0.04
	2015	0.85	0.03	0.76	0.04			0.79	0.04			0.71	0.05	0.68	0.06
TLI-C6	2016	0.96	0	0.89	0.01	0.95	0	0.97	0	0.93	0.01	0.9	0.01	0.94	0.01
	2017	0.93	0.01	0.81	0.02	0.93	0.01	0.95	0.01	0.92	0.01	0.91	0.01	0.91	0.01
TLI-C7	2018	0.35	0.06	0.92	0.01	0.87	0.02	0.91	0.01	0.92	0.01	0.88	0.02	0.91	0.02
	2019	0.91	0.01	0.88	0.01	0.92	0.01	0.93	0.01	0.92	0.01			0.9	0.01
TLI-C8	2019	0.79	0.03	0.84	0.02	0.78	0.03	0.85	0.02	0.11	0.07			0.84	0.02

Prediction abilities are reported as correlation (r) between predicted value and phenotypic best linear unbiased estimator (BLUP), along with the standard deviation (sd) of 100 random iterations.

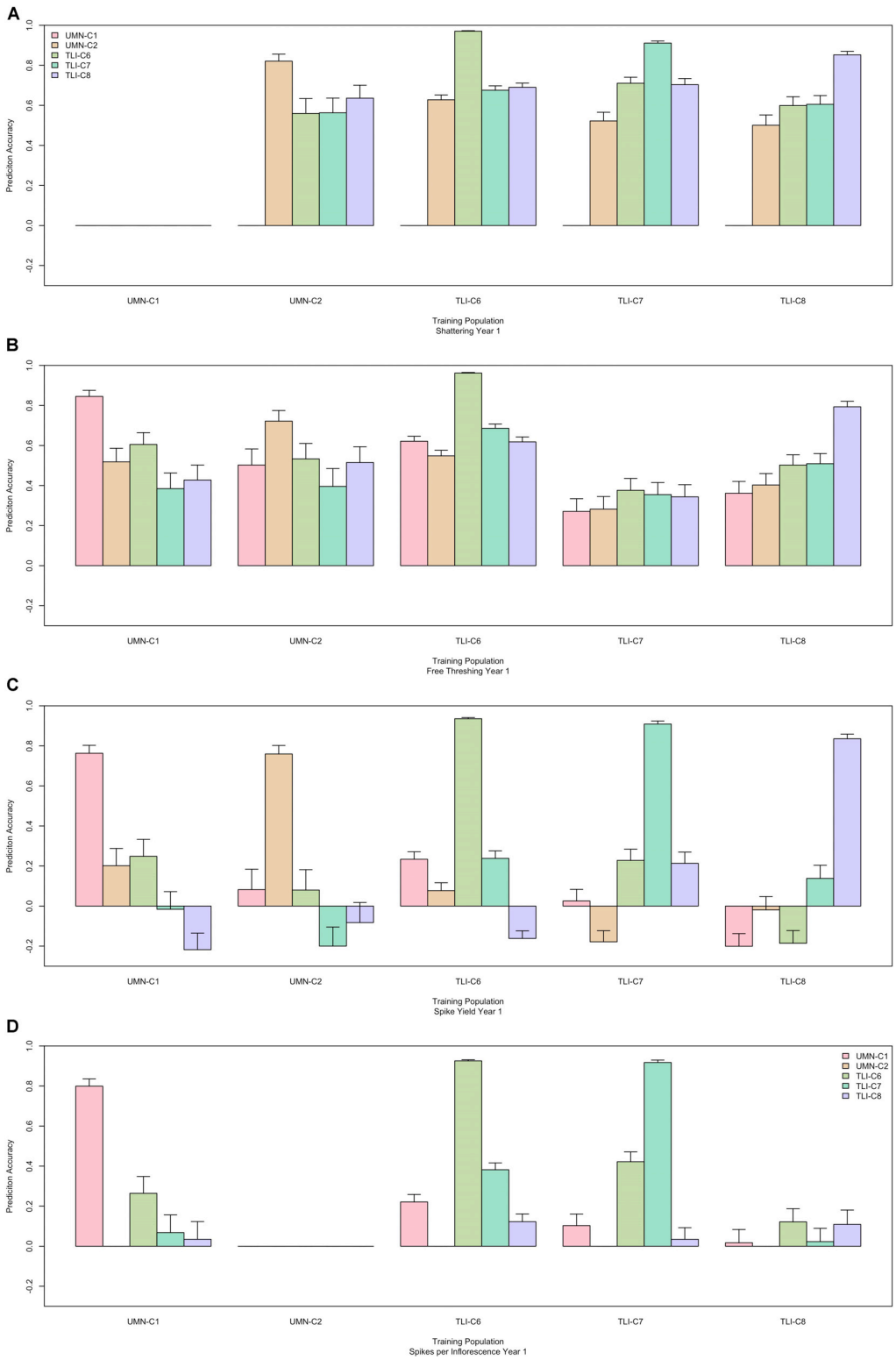


FIGURE 2 | Performance of genomic selection (GS) across five cycles. Each panel represents one trait, shattering **(A)**, free threshing **(B)**, spike yield **(C)**, and spikelets per inflorescence **(D)**. The x-axis is the cycle that was used as the prediction population. Colored bars represent the prediction ability for each of the four other cycles, where each cycle forms the training population. For comparison, the fivefold cross-validation within cycle is represented for each training and prediction cycle, which usually provides the highest predictive ability. The y-axis is the prediction ability which is the correlation between the GS predicted value and the phenotypic best linear unbiased predictor (BLUP). Error bars represent the 95% confidence interval for the correlation value.

a significant relationship between these variables ($p < 0.001$, **Figure 4**). Key domestication traits of shattering, free threshing, and seed mass showed high heritability across cycle predictions. In comparison, spike yield, spikes per inflorescence, and plant height had lower heritability estimates and prediction accuracies.

Optimizing GS Prediction and Training Set

Finally, in an effort to determine ideal GS training populations and enhance GS results, we used two different sub-setting methods. The first subset utilized results from the PCA decomposition of the genomic marker matrix to develop two subpopulations based on relatedness, **Figure 1**. The second sub-setting method used each individual breeding program as a unique training population. Using these data sets, we evaluated the same across-environment GS models, with the GS training population being more closely related to the prediction population. The GS model using all cycles in a leave-one-cycle-out method, with all other cycles in the training population (**Figure 3**), was used as the reference. A model was declared better than the reference if the 95% confidence intervals were non-overlapping. We tested five different training populations for each of 55 cycle/trait combinations. The top performing model for each combination is listed in **Table 4**. Overall, there was much inconsistency between the best performing model and each cycle/trait combination (**Figure 5** and **Supplementary Figure S3**). However, using the leave-one-out as a reference resulted in the best performing model 62% of the time (34 of 55 combinations).

DISCUSSION

Combining Data Resources

The affordability of next-generation sequencing provides many opportunities for breeding that were previously unavailable. Particularly for programs that are implementing GS, there is an opportunity to leverage data across breeding programs and identify synergistic opportunities. This is particularly the case for minor and emerging crops. We were able to combine five cycles representing nearly a decade of breeding progress for IWG in the Central USA. Across the two programs, many key traits were measured each year, but there were often minor differences in trait measurement, specifically scoring of free-threshing and total spike yield between the TLI and UMN programs. While our results did not show any marked difference in these traits, i.e., consistent free-threshing prediction and

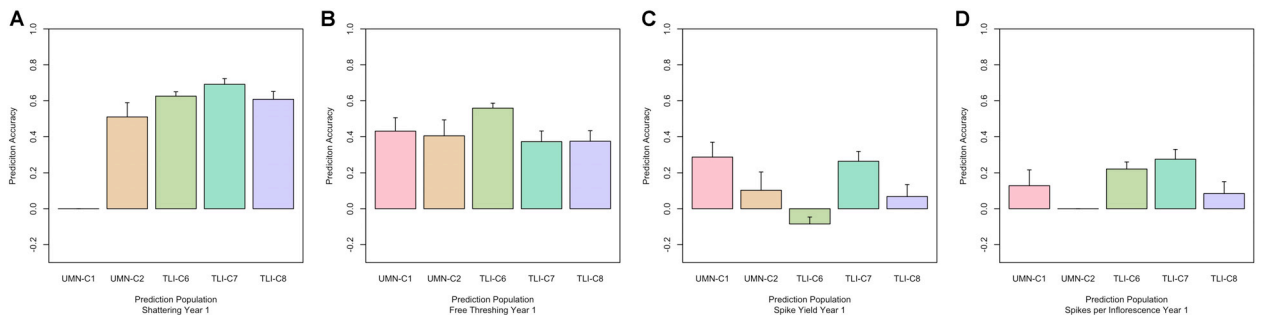


FIGURE 3 | Genomic selection (GS) performance for shattering, free threshing, spike yield, and spikelets per inflorescence, **(A–D)**, respectively. Within each panel the x-axis is grouped by cycle name. Predictions were made by leaving out the named cycle and predicting that cycle from all other data. The prediction ability is the correlation between the predicted GS value and the phenotypic best linear unbiased predictor (BLUP), with standard error bars representing the 95% confidence interval.

TABLE 3 | Genomic selection prediction abilities of intermediate wheatgrass traits across sites.

		Free Threshing%		Plant Height		Seed Mass		Shattering		Spikelets per Inflorescence		Spike Length		Spike Yield	
Prediction Site	**Year**	***r***	**CI**	***r***	**CI**	***r***	**CI**	***r***	**CI**	***r***	**CI**	***r***	**CI**	***r***	**CI**
UMN-C1	2012	0.43	0.07	0.25	0.08	0.54	0.06			0.13	0.09	0.22	0.08	0.29	0.08
	2013	0.58	0.06	0.32	0.08	0.42	0.07			0.10	0.09	0.42	0.07	0.37	0.07
UMN-C2	2014	0.40	0.08	0.42	0.08			0.51	0.07			0.20	0.10	0.10	0.10
	2015	0.42	0.08	0.48	0.08			0.61	0.06			0.46	0.08	0.12	0.10
TLI-C6	2016	0.56	0.03	0.02	0.04	0.51	0.03	0.63	0.02	0.22	0.04	0.34	0.03	−0.08	0.04
	2017	0.71	0.02	0.22	0.05	0.67	0.03	0.64	0.03	0.39	0.04	0.42	0.04	−0.10	0.05
TLI-C7	2018	0.37	0.06	0.45	0.04	0.53	0.05	0.69	0.03	0.27	0.05	0.48	0.04	0.26	0.05
	2019	0.77	0.02	0.28	0.05	0.59	0.04	0.69	0.03	0.48	0.04			0.05	0.06
TLI-C8	2019	0.37	0.06	0.29	0.06	0.35	0.06	0.61	0.04	0.08	0.07			0.07	0.07

Prediction population was one cycle, with the training population comprising all other cycles. Predictive ability is reported as correlation between predicted value and phenotypic best linear unbiased predictor (BLUP) with ± range for the 95% confidence interval for correlation.

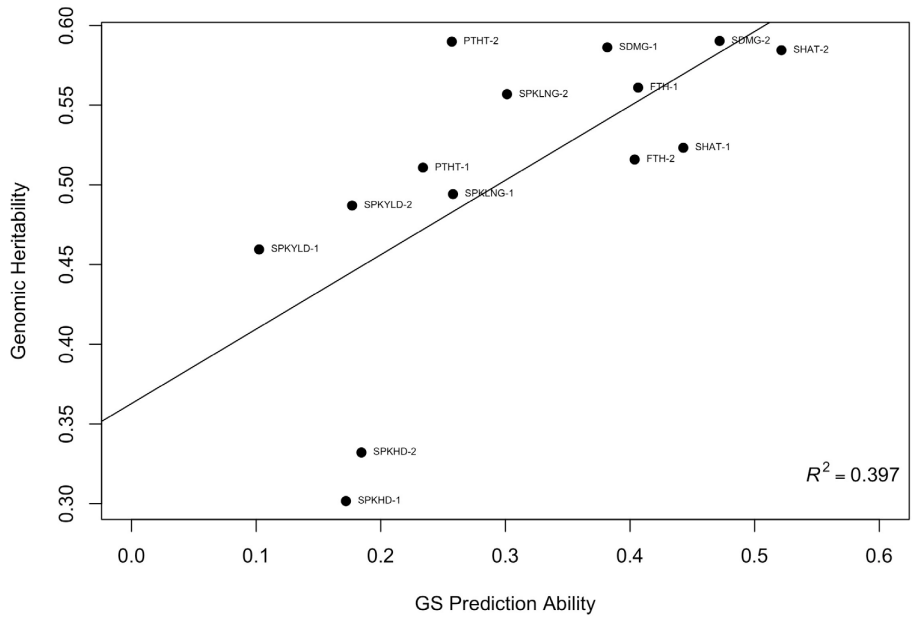


FIGURE 4 | Scatter plot of genomic selection predictive ability and genomic heritability for 12 traits. For each point the trait name is provided with 1 or 2 representing year of observation where SPKYLD is spike yield; PTHT, plant height; SPKLNG, spike length; SDMG, seed mass; SPKHD, spikelets per inflorescence; SHAT, shattering; FTH, free threshing.

TABLE 4 | Highest performing genomic selection (GS) model for each trait/cycle combination across five breeding cycles representing two different breeding programs.

Prediction site	Year	Free threshing	Plant height	Seed mass	Shattering	Spikelets per inflorescence	Spike length	Spike yield
UMN-C1	2012	MN†	LOO	TLI-PCA		LOO	KS	LOO
	2013	LOO	LOO	LOO		TLI-PCA	LOO	LOO
UMN-C2	2014	MN	MN		LOO		KS	LOO
	2015	LOO	LOO		LOO		LOO	MN
TLI-C6	2016	LOO	KS	LOO	UMN-PCA	LOO	LOO	MN
	2017	LOO	LOO	LOO	LOO	LOO	LOO	MN
TLI-C7	2018	LOO	LOO	TLI-PCA	LOO	LOO	LOO	TLI-PCA
	2019	LOO	LOO	KS	TLI-PCA	LOO		KS
TLI-C8	2019	UMN-PCA	LOO	LOO	UMN-PCA	UMN-PCA		UMN-PCA

Predictive ability was assessed as the correlation between the GS predicted value and the phenotypic best linear unbiased predictor (BLUP). Models differed with respect to the training population used to develop the model. The leave one-out, LOO, model was used as the reference model and only if a model exceeded the 95% confidence interval of the LOO model was it considered superior. †Models are: LOO, leave-one-out, prediction cycle is left out of the training set, and all other cycles are used to train the model. MN and KS are breeding-program specific where only genets from Minnesota (or Kansas) are used to predict each cycle. For TLI-C6 2016 plant height, KS training population would consist of TLI-C7 and TLI-C8, with TLI-C6 as the prediction population. UMN-PCA and TLI-PCA are where the training population is made from PCA analysis of the marker matrix, with UMN-PCA encompassing most UMN lines and some of TLI that were more similar to UMN material than the TLI subset.

inconsistent spike yield across other cycles, it is unknown if more consistent data collection would result in higher predictive ability within this data set. As other breeding programs are established trait standardization using crop ontology (Shrestha et al., 2010) could greatly increase the inter-operability of experimental data.

Genomic Selection Accuracy and Analysis

Within-Cycle Predictive Ability

Using data generated from the field trials and next generation sequencing, we evaluated the potential of GS to predict trait values across geographically distant IWG breeding programs. First, within-cycle predictions were generated to verify GS could appropriately predict trait values (**Table 2**). These cross-validation predictions were the highest GS predictive abilities achieved because the training sets were highly related and the training and test sets were grown in the same environment, minimizing any genotype by environment interactions (Desta and Ortiz, 2014; Zhang et al., 2016). These prediction abilities provide a potential maximum value that could be achieved utilizing the current markers and phenotypes within the study. Additionally, these predictions show that within breeding programs, GS could be an effective way to enhance genetic gain in IWG.

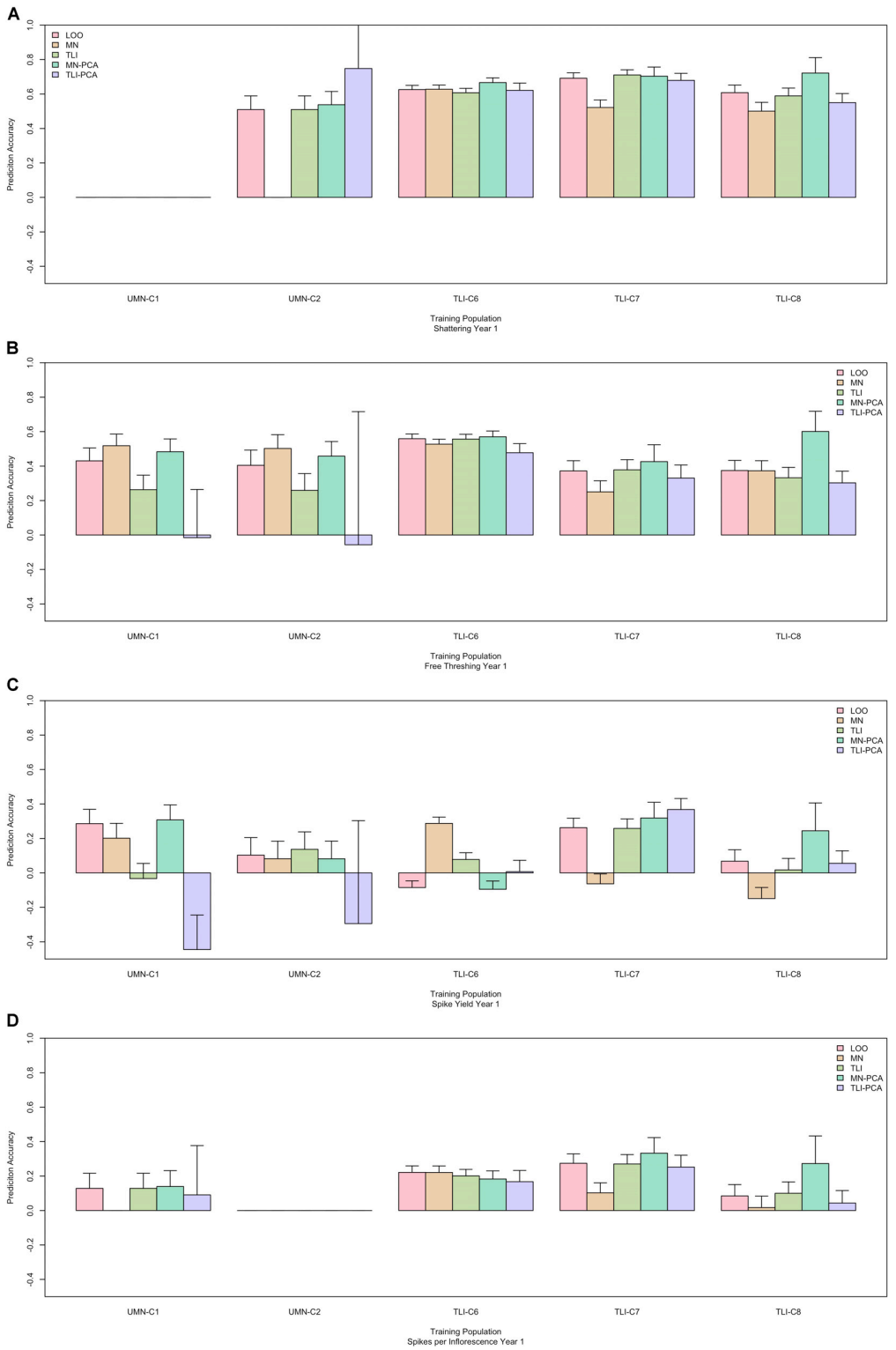


FIGURE 5 | Performance of genomic selection (GS) across five cycles with different training populations. Each panel represents one trait, shattering **(A)**, free threshing **(B)**, spike yield **(C)**, and spikelets per inflorescence **(D)**. Within each panel the x-axis is grouped by cycle name. Predictions were made by: LOO, leave one out where all data other than the predicted cycle were used in the training population. MN or TLI where only data from each separate breeding program, Minnesota or Kansas respectively, were used as the training population. MN-PCA or TLI-PCA where principal component analysis (PCA) was used to cluster genets within breeding programs, MN or TLI, and form the training populations. The prediction ability is the correlation between the predicted GS value and the phenotypic best linear unbiased predictor (BLUP), with standard error bars representing the 95% confidence interval.

Across-Cycle Predictive Ability

After evaluating within-site GS prediction, across-site predictions were generated for all cycles. As the relatedness and environments changed, a decrease in GS predictive ability was observed. Within these evaluations, two general trends emerged. For key domestication traits such as shattering and free-threshing, GS predictions were relatively high and constant across environment (**Figure 2**). For agronomic and yield related traits, the results were inconsistent, with some sites even producing negative prediction abilities. This suggests that certain traits may be more amenable to multi-environment GS than other traits.

To further investigate this trend, we examined the genomic heritability from the GS models. Plotting genomic heritability and predictive ability (**Figure 4**) suggests that domestication traits may exhibit lower genotype-by-environment interaction than agronomic traits. Additionally, the resulting prediction abilities of various traits were reaching the level of the trait heritability.

Domestication traits were highly predictive across environments, possibly indicating that these traits are not as influenced by environment as other traits. Within wheat, there are several well-known genes that control these traits. Free threshing in wheat is determined by a recessive mutation in the *Tg* (*tenacious glume*) locus and the dominant mutation of the *Q* gene. The *Tg* loci has been reported to explain up to 44% of the variation in threshability, and at least five other quantitative trait loci (QTLs) for threshability have been observed in wheat (Jantasuriyarat et al., 2004). Within IWG, recent research by Larson et al. (2019) found that QTL markers explained up to 46% of variation for free threshing across two locations. The *Br* (*brittle rachis*) locus controls shattering in wheat with two dominant genes and is homoeologous to the *Btr* loci in barley (Nalam et al., 2006). Traits such as free threshing and shattering that may have larger-effect QTL could be both increasing GS predictions as well as maintaining predictive ability across environments.

For agronomic traits, many more QTL of much smaller size have been reported. Bajgain et al. (2019) identified over 154 QTL for seven agronomic traits in IWG with the largest QTL effect sizes explaining only 4% of the phenotypic variation. Larson et al. (2019) found 12 QTL that explained up to 27% of the variation of spike yield in a biparental population grown in five environments. As the number of QTL increase and their size decreases, adequately accounting for their effects across environments may be more challenging. Simulation studies have shown that as heritability decreases GS accuracies are lowered (Iwata and Jannink, 2011). Other research has indicated that GS accuracy diminishes as the number of QTL increases (Shengqiang et al., 2009).

Optimizing GS Prediction

Complementary to evaluating how traits may respond to GS, we also examined how the training population could be optimized to achieve the best results when combining data across breeding programs. While all germplasm originated from TLI material, UMN-C1 was only a subset of the entire TLI program and UMN-C2 was selected for MN conditions, which are different than KS. Additionally, from the founding lines (TLI-C3) two and five generations of selection had occurred for UMN and TLI respectively, allowing for potential population divergence.

We evaluated models using a leave-one-out approach for all cycles, which should result in the largest training population available for GS prediction. This leave-one-out strategy insured that the models were not biased by the size or the relationship of the training population (Desta and Ortiz, 2014) in comparison to GS prediction made from individual cycles. Additionally, we used PCA to develop a subset of data more related to each breeding program to ensure any large population structure differences did not influence GS prediction (Norman et al., 2018). Finally, data predictions were also developed using data specific to each breeding program.

The results from these models were inconsistent, with the leave-one-out model performing as well as or better the majority of the time. Often breeding program-specific or PCA-specific subsets performed well, but there was no clear pattern to this performance (**Table 4**). For example, the optimized training set using PCA for UMN provided the best prediction for TLI-C8 free threshing, whereas the TLI-PCA optimized training set provided the best prediction for UMN-C1 seed mass. In this case the training sets had optimal performance in data sets for which they were not specifically optimized. While developing highly optimized prediction sets has been shown to increase prediction accuracies (Isidro et al., 2015; Rutkoski et al., 2015) we did not note this in this data set. This could result from the large amount of genetic variance compared to domesticated crops. While future breeding efforts may be enhanced by optimizing the training set, these data suggest that increasing the training population, i.e., leave-one-out, is generally more useful than optimizing relatedness to prediction candidates.

Implication for Future Development

As concerted efforts to develop new crops through domestication of crop wild relatives continues for food security and environmental benefits (Glover et al., 2010; Mayes et al., 2012), we have evaluated approaches for genomics-assisted breeding of neo-domesticated crops with insights into maximizing genetic gains. While plant breeding is both expensive and time consuming (Crews and DeHaan, 2015; DeHaan et al., 2016), genomic technologies provide a way to accelerate compared to phenotypic selection (Varshney et al., 2012; Unamba et al., 2015). Next-generation sequencing coupled with powerful tools such as GS and genome wide association studies could allow for significantly improving agronomic and domestication traits in short periods of time, especially in non-model plants. Within the TLI-IWG breeding program, GS has reduced the breeding cycle time from 2 years to 1 year, which should effectively double the rate of genetic gains if the predictability is roughly equivalent to the narrow-sense heritability. Additionally, the genetic resources generated can be used to better understand the genetic architecture of important agronomic and domestication traits (examples include Bajgain et al., 2019; Larson et al., 2019).

These results show that as plant species undergo early domestication, collaboration will accelerate progress, i.e., not

every breeding program will have to solve the same domestication problems and that progress can be made across programs. As domestication traits are fixed, breeding programs can work toward developing adapted lines for targeted growing regions. DeHaan et al. (2016) suggest a pipeline strategy for new crop domestication where many candidates are tested and attrition occurs as information about candidates are gained. Cooperative efforts in early breeding stages, along with applied genomics, should result in more quickly advancing and developing promising species into commercially viable crops.

These data provide several potential use cases for breeding programs. If a program is beginning, there appears to be little downside in utilizing training data sets from across programs. As programs mature and have sufficient, data from multiple years and locations, GS models can be developed within programs. This could be especially important for agronomic traits such as spike yield as combining data across programs could result in negative predictions (**Figure 5**). However, when looking at GS models using program-specific data, the GS predictions were always positive, so program-specific models may be the most conservative way to insure genetic gains. For domestication traits, predictions were usually similar regardless of the training population, suggesting minimal benefit to pooling multiple locations.

Our results show that GS can be a powerful tool in breeding programs, yet GS is not a single, stand-alone solution for quickly developing new crops. While we envision GS improving with larger data sets and new statistical model development, multi-environment predictions are extremely complex. To fully leverage genomic resources, GS should be integrated with phenomic and environmental data. High-throughput phenotyping is an emerging field that is providing dense phenomic measurements (White et al., 2012; Araus et al., 2018) that have been shown to increase GS model accuracy (Rutkoski et al., 2016; Crain et al., 2018). A further complement to better predict how the environment influences phenotype will include incorporating crop models to better understand plant development within a range of environments (i.e., review of crop models in wheat by Chenu et al., 2017). Future advances in these areas as well as incorporating them into unified prediction models will allow scientist to drive genetic gain in novel crops across a range of environments.

CONCLUSION

Domesticating crop wild relatives is a challenging and time consuming task (Cox et al., 2002; DeHaan et al., 2014). Previous research at TLI has shown that a 77% increase in seed yield was achieved in two cycles of selection, however, to reach yields of annual wheat another 20 years of sustained breeding gains would be required with even longer time intervals to achieve similar seed mass to wheat (DeHaan et al., 2014).

Perennial grains derived from the domestication of wild species hold much promise for environmental and human benefit. To achieve these benefits, specific traits of wild species will need to be modified. Within IWG, free-threshing and non-shattering seed types are two key domestication traits that must be improved for wide-scale adoption. In addition, the economic yield of IWG must be sufficient to incentivize the transition to new crops. Along with fixing key traits for domestication, breeding efforts should also ensure that crops are broadly adapted (DeHaan et al., 2016).

The ability to use molecular tools such as GS, combined with modern breeding methodologies, may allow perennial crops and crop wild relatives to compress the 10,000 year selection history of many annual crops into a few decades. While GS predictions for agronomic traits like spike yield were low between breeding sites and environments, significant synergies could be achieved by utilizing collective information about domestication traits. While site-specific or regional programs will be necessary to breed for the best locally adapted genets, progress made toward improving key domestication traits could be shared among all programs. This is especially important for resource-limited programs that are domesticating new crops, allowing improvement for traits that are less environmentally influenced and are essential for domestication. Early domestication work could be carried out by a single program or shared among programs with each program phenotyping a few lines in diverse locations to quickly and efficiently improve key traits. As more programs are initialized for the breeding of IWG, they will be able to identify germplasm that has key domestication traits and be able to focus breeding efforts toward achieving higher site-specific agronomic performance.

DATA AVAILABILITY STATEMENT

The genotypic datasets analyzed for this study have been placed in the NCBI Sequence Read Archive (SRA) (https://www.ncbi.nlm.nih.gov/bioproject/) BioProject accession numbers PRJNA563706, PRJNA609095, and PRJNA608473. All phenotypic data and scripts for data analysis have been placed in the Dryad Digital Repository (https://doi.org/10.5061/dryad.3j9kd51d9).

AUTHOR CONTRIBUTIONS

JC, LD, and PB conceived experimental ideas and methods. LD, JA, XZ, and PB conducted all field evaluations. LD, JP, PB, and XZ performed DNA extraction and genotyping. JC completed data analysis and wrote the manuscript. All authors read, reviewed, and approved the final manuscript.

FUNDING

This work was funded in part by the Perennial Agriculture Project, in conjunction with the Malone Family Land Preservation Foundation and The Land Institute.

ACKNOWLEDGMENTS

We acknowledge the excellent field and laboratory assistance of Shuangye Wu, Marty Christians, Brett Heim, and Professor Donald Wyse. The *Thinopyrum intermedium* Genome Sequencing Consortium provided pre-publication access to the IWG genome sequence. Computational work was completed on the Beocat Research Cluster at Kansas State University, which is funded in part by NSF grants CNS-1006860, EPS-1006860, and EPS-0919443.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.00319/full#supplementary-material

FIGURE S1 | Performance of genomic selection (GS) across five cycles. Each panel represents one trait, A-N. The x-axis is the cycle that was used as the prediction population. Colored bars represent the prediction ability for training populations for other cycles. For comparison, the within cycle fivefold cross-validation is represented for each training and prediction site, which usually provides the highest predictive ability. The y-axis is the prediction ability which is the correlation between the GS predicted value and the phenotypic best linear unbiased predictor (BLUP). Error bars represent the 95% confidence interval for the correlation value.

FIGURE S2 | Relationship of genomic selection (GS) predicted values (x-axis) and observed values (y-axis) for two traits, including line of best fit. In panel A, a free threshing (n = 2,496), a trait with high predictive ability is shown, while panel B represents a trait with low predictive ability, spike yield (n = 2,508). The training population was TLI-C7 and the prediction population was TLI-C6.

FIGURE S3 | Genomic selection (GS) performance across five cycles where each panel represents one trait A-N. Within each panel the x-axis is grouped by the cycle of data that was predicted, with different training populations represented by colored bars. Leave-one-out (LOO) was a training population where the cycle of interest was left out and all other cycles were used to predict the cycle. Breeding program-specific training populations were developed for the Minnesota (MN) and Kansas (KS) breeding programs. Finally, UMN-PCA and TLI-PCA are training populations that were developed using principal component analysis (PCA) of the marker matrix. UMN-PCA is a training population more closely related to UMN genets while TLI-PCA is more closely related to TLI genets. The prediction ability is the correlation between the predicted GS value and the phenotypic best linear unbiased predictor (BLUP), with standard error bars representing the 95% confidence interval.

REFERENCES

Araus, J. L., Kefauver, S. C., Zaman-Allah, M., Olsen, M. S., and Cairns, J. E. (2018). Translating high-throughput phenotyping into genetic gain. *Trends Plant Sci.* 23, 451–466. doi: 10.1016/j.tplants.2018.02.001

Bajgain, P., Zhang, X., and Anderson, J. A. (2019).). Genome-wide association study of yield component traits in intermediate wheatgrass and implications in genomic selection and breeding. *Genes| Genomes| Genetics* 9, 2429–2439. doi: 10.1534/g3.119.400073

Batello, C., Wade, L., Cox, S., Pogna, N., Bozzini, A., and Choptiany, J. (eds) (2013). "Perennial crops for food security," in *Proceedings of the FAO Expert Workshop*, Rome.

Battenfield, S. D., Guzmán, C., Gaynor, R. C., Singh, R. P., Peña, R. J., Dreisigacker, S., et al. (2016). Genomic selection for processing and end-use quality traits in the CIMMYT spring bread wheat breeding program. *Plant Genome* 9, 1–12. doi: 10.3835/plantgenome2016.01.0005

Bernardo, R., and Yu, J. (2007). Prospects for genomewide selection for quantitative traits in maize. *Crop Sci.* 47, 1082–1090. doi: 10.2135/cropsci2006.11.0690

Browning, B. L., and Browning, S. R. (2016). Genotype imputation with millions of reference samples. *Am. J. Hum. Genet.* 98, 116–126. doi: 10.1016/j.ajhg.2015.11.020

Chenu, K., Porter, J. R., Martre, P., Basso, B., Chapman, S. C., Ewert, F., et al. (2017). Contribution of crop models to adaptation in wheat. *Trends Plant Sci.* 22, 472–490. doi: 10.1016/j.tplants.2017.02.003

Cox, T. S., Bender, M., Picone, C., Van Tassel, D. L., Holland, J. B., Brummer, E. C., et al. (2002). Breeding perennial grain crops. *CRC. Crit. Rev. Plant Sci.* 21, 59–91. doi: 10.1080/0735-260291044188

Cox, T. S., Van Tassel, D. L., Cox, C. M., and Dehaan, L. R. (2010). Progress in breeding perennial grains. *Crop Pasture Sci.* 61, 513–521. doi: 10.1071/CP09201

Crain, J., Mondal, S., Rutkoski, J., Singh, R. P., and Poland, J. (2018). Combining high-throughput phenotyping and genomic information to increase prediction and selection accuracy in wheat breeding. *Plant Genome* 11, 1–14. doi: 10.3835/plantgenome2017.05.0043

Crews, T. E., Carton, W., and Olsson, L. (2018). Is the future of agriculture perennial? Imperatives and opportunities to reinvent agriculture by shifting from annual monocultures to perennial polycultures. *Glob. Sustain.* 1, 1–18. doi: 10.1017/sus.2018.11

Crews, T. E., and DeHaan, L. R. (2015). The strong perennial vision: a response. *Agroecol. Sustain. Food Syst.* 39, 500–515. doi: 10.1080/21683565.2015.1008777

Crossa, J., Pérez-Rodríguez, P., Cuevas, J., Montesinos-López, O., Jarquín, D., de los Campos, G., et al. (2017). Genomic selection in plant breeding: methods, models, and perspectives. *Trends Plant Sci.* 22, 961–975. doi: 10.1016/j.tplants.2017.08.011

Culman, S. W., DuPont, S. T., Glover, J. D., Buckley, D. H., Fick, G. W., Ferris, H., et al. (2010). Long-term impacts of high-input annual cropping and unfertilized perennial grass production on soil properties and belowground food webs in Kansas. *USA. Agric. Ecosyst. Environ.* 137, 13–24. doi: 10.1016/j.agee.2009.11.008

Culman, S. W., Snapp, S. S., Ollenburger, M., Basso, B., and DeHaan, L. R. (2013). Soil and water quality rapidly responds to the perennial grain Kernza wheatgrass. *Agron. J.* 105, 735–744. doi: 10.2134/agronj2012.0273

DeHaan, L., Christians, M., Crain, J., and Poland, J. (2018). Development and evolution of an intermediate wheatgrass domestication program. *Sustainability* 10:1499. doi: 10.3390/su10051499

DeHaan, L. R., and Ismail, B. P. (2017). Perennial cereals provide ecosystem benefits. *Cereal Foods World* 62, 278–281. doi: 10.1094/CFW-62-6-0278

DeHaan, L. R., Van Tassel, D. L., Anderson, J. A., Asselin, S. R., Barnes, R., Baute, G. J., et al. (2016). A pipeline strategy for grain crop domestication. *Crop Sci.* 56, 917–930. doi: 10.2135/cropsci2015.06.0356

DeHaan, L. R., Wang, S., Larson, S. R., Cattani, D. J., Zhang, X., and Kantarski, T. (2014). "Current efforts to develop perennial wheat and domesticate *Thinopyrum intermedium* as a perennial grain," in *Perennial Crops for Food Security Proceedings of the FAO Expert Workshop, 28–30 Aug. 2013*, eds C. Batello, L. Wade, S. Cox, N. Pogna, A. Bozzini, and J. Choptiany (Rome), 72–89.

Desta, Z. A., and Ortiz, R. (2014). Genomic selection: genome-wide prediction in plant improvement. *Trends Plant Sci.* 19, 592–601. doi: 10.1016/j.tplants.2014.05.006

Earl, D. A., and vonHoldt, B. M. (2012). Structure harvester: a website and program for visualizing structure output and implementing the Evanno method. *Conserv. Genet. Resour.* 4, 359–361. doi: 10.1007/s12686-011-9548-7

Endelman, J. B. (2011). Ridge regression and other kernels for genomic selection with R package rrBLUP. *Plant Genome J.* 4, 250–255. doi: 10.3835/plantgenome2011.08.0024

Endelman, J. B., and Jannink, J.-L. (2012). Shrinkage estimation of the realized relationship matrix. *Genes| Genomes| Genetics* 2, 1405–1413. doi: 10.1534/g3.112.004259

Falconer, D. S., and Mackay, T. F. (1996). *Introduction to Quantitative Genetics 4th Edn*. Essex: Longman.

Fletcher, T. D. (2010). *psychometric: Applied Psychometric Theory*. Available at: https://cran.r-project.org/web/packages/psychometric/index.html (accessed March 11, 2020).

García-Ruiz, A., Cole, J. B., VanRaden, P. M., Wiggans, G. R., Ruiz-López, F. J., and Van Tassell, C. P. (2016). Changes in genetic selection differentials and generation intervals in US Holstein dairy cattle as a result of genomic selection. *Proc. Natl. Acad. Sci. U.S.A.* 113, E3995–E4004. doi: 10.1073/pnas.1519061113

Gilmour, A. R., Gogel, B. J., Cullis, B. R., Welham, S. J., and Thompson, R. (2015). *ASReml User Guide Release 4.1 Functional Specification*. Hemel Hempstead: VSN International Ltd.

Glaubitz, J. C. J., Casstevens, T. M. T., Lu, F., Harriman, J., Elshire, R. R. J., Sun, Q., et al. (2014). TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. *PLoS One* 9:e90346. doi: 10.1371/journal.pone.0090346

Glover, J. D., Reganold, J. P., Bell, L. W., Borevitz, J., Brummer, E. C., Buckler, E. S., et al. (2010). Increased food and ecosystem security via perennial grains. *Science* 328, 1638–1639. doi: 10.1126/science.1188761

Goddard, M. E., and Hayes, B. J. (2007). Genomic selection. *J. Anim. Breed. Genet.* 124, 323–330. doi: 10.1111/j.1439-0388.2007.00702.x

Guzman, C., Peña, R. J., Singh, R., Autrique, E., Dreisigacker, S., Crossa, J., et al. (2016). Wheat quality improvement at CIMMYT and the use of genomic selection on it. *Appl. Transl. Genomics* 11, 0–5. doi: 10.1016/j.atg.2016.10.004

Harlan, J. R., de Wet, J. M. J., and Price, E. G. (1973). Comparative evolution of cereals. *Evolution* 27, 311–325. doi: 10.2307/2406971

Huang, G., Qin, S., Zhang, S., Cai, X., Wu, S., Dao, J., et al. (2018). Performance, economics and potential impact of perennial rice PR23 relative to annual rice cultivars at multiple locations in Yunnan Province of China. *Sustain* 10, 1–18. doi: 10.3390/su10041086

Hufford, M. B., Xu, X., Van Heerwaarden, J., Pyhäjärvi, T., Chia, J. M., Cartwright, R. A., et al. (2012). Comparative population genomics of maize domestication and improvement. *Nat. Genet.* 44, 808–811. doi: 10.1038/ng.2309

Isidro, J., Jannink, J. L., Akdemir, D., Poland, J., Heslot, N., and Sorrells, M. E. (2015). Training set optimization under population structure in genomic selection. *Theor. Appl. Genet.* 128, 145–158. doi: 10.1007/s00122-014-2418-4

Isik, F., Holland, J., and Maltecca, C. (2017). *Genetic Data Analysis for Plant and animal Breeding*. Berlin: Springer.

Iwata, H., and Jannink, J. L. (2011). Accuracy of genomic selection prediction in barley breeding programs: a simulation study based on the real single nucleotide polymorphism data of barley breeding lines. *Crop Sci.* 51, 1915–1927. doi: 10.2135/cropsci2010.12.0732

Jakobsson, M., and Rosenberg, N. A. (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23, 1801–1806. doi: 10.1093/bioinformatics/btm233

Jantasuriyarat, C., Vales, M. I., Watson, C. J. W., and Riera-Lizarazu, O. (2004). Identification and mapping of genetic loci affecting the free-threshing habit and spike compactness in wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* 108, 261–273. doi: 10.1007/s00122-003-1432-8

Jarquín, D., Lemes da Silva, C., Gaynor, R. C., Poland, J., Fritz, A., Howard, R., et al. (2017). Increasing genomic-enabled prediction accuracy by modeling genotype × environment interactions in kansas wheat. *Plant Genome* 10, 1–15. doi: 10.3835/plantgenome2016.12.0130

Jungers, J. M., DeHaan, L. H., Mulla, D. J., Sheaffer, C. C., and Wyse, D. L. (2019). Reduced nitrate leaching in a perennial grain crop compared to maize in the Upper Midwest. *USA. Agric. Ecosyst. Environ.* 272, 63–73. doi: 10.1016/j.agee.2018.11.007

Knapp, S. J., and Bridges, W. C. (1990). Using molecular markers to estimate quantitative trait locus parameters: power and genetic variances for unreplicated and replicated progeny. *Genetics* 126, 769–777.

Koinange, E. M. K., Singh, S. P., and Gepts, P. (1996). Genetic control of the domestication syndrome in common bean. *Crop Sci.* 36, 1037–1045. doi: 10.2135/cropsci1996.0011183x003600040037x

Larson, S., DeHaan, L., Poland, J., Zhang, X., Dorn, K., Kantarski, T., et al. (2019). Genome mapping of quantitative trait loci (QTL) controlling domestication traits of intermediate wheatgrass (*Thinopyrum intermedium*). *Theor. Appl. Genet.* 132, 2325–2351. doi: 10.1007/s00122-019-03357-6

Lopez-Cruz, M., Crossa, J., Bonnett, D., Dreisigacker, S., Poland, J., Jannink, J.-L., et al. (2015). Increased prediction accuracy in wheat breeding trials using a marker × environment interaction genomic selection model. *G3 (Bethesda)* 5, 569–582. doi: 10.1534/g3.114.016097

Mayes, S., Massawe, F. J., Alderson, P. G., Roberts, J. A., Azam-Ali, S. N., and Hermann, M. (2012). The potential for underutilized crops to improve security of food production. *J. Exp. Bot.* 63, 1075–1079. doi: 10.1093/jxb/err396

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, R. S., and Purugganan, M. D. (2013). Evolution of crop species: genetics of domestication and diversification. *Nat. Rev. Genet.* 14, 840–852. doi: 10.1038/nrg3605

Nalam, V. J., Vales, M. I., Watson, C. J. W., Kianian, S. F., and Riera-Lizarazu, O. (2006). Map-based analysis of genes affecting the brittle rachis character in tetraploid wheat (*Triticum turgidum* L.). *Theor. Appl. Genet.* 112, 373–381. doi: 10.1007/s00122-005-0140-y

Norman, A., Taylor, J., Edwards, J., and Kuchel, H. (2018). Optimising genomic selection in wheat: effect of marker density, population size and population structure on prediction accuracy. *G3 Genes, Genomes, Genet.* 8, 2889–2899. doi: 10.1534/g3.118.200311

Oakey, H., Cullis, B., Thompson, R., Comadran, J., Halpin, C., and Waugh, R. (2016). Genomic selection in multi-environment crop trials. *G3 (Bethesda)* 6, 1–34. doi: 10.1534/g3.116.027524

Olsen, K. M., and Wendel, J. F. (2013). A bountiful harvest: genomic insights into crop domestication phenotypes. *Annu. Rev. Plant Biol.* 64, 47–70. doi: 10.1146/annurev-arplant-050312-120048

Østerberg, J. T., Xiang, W., Olsen, L. I., Edenbrandt, A. K., Vedel, S. E., Christiansen, A., et al. (2017). Accelerating the domestication of new crops: feasibility and approaches. *Trends Plant Sci.* 22, 373–384. doi: 10.1016/j.tplants.2017.01.004

Piepho, H. P., Möhring, J., Schulz-Streeck, T., and Ogutu, J. O. (2012). A stage-wise approach for the analysis of multi-environment trials. *Biometrical J.* 54, 844–860. doi: 10.1002/bimj.201100219

Pimentel, D., Cerasale, D., Stanley, R. C., Perlman, R., Newman, E. M., Brent, L. C., et al. (2012). Annual vs. perennial grain production. *Agric. Ecosyst. Environ.* 161, 1–9. doi: 10.1016/j.agee.2012.05.025

Poland, J. A., Brown, P. J., Sorrells, M. E., and Jannink, J. L. (2012). Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS One* 7:e32253. doi: 10.1371/journal.pone.0032253

Power, A. G. (2010). Ecosystem services and agriculture: tradeoffs and synergies. *Philos. Trans. R. Soc. B Biol. Sci.* 365, 2959–2971. doi: 10.1098/rstb.2010.0143

Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959. doi: 10.1111/j.1471-8286.2007.01758.x

Pugliese, J. Y., Culman, S. W., and Sprunger, C. D. (2019). Harvesting forage of the perennial grain crop Kernza (*Thinopyrum intermedium*) increases root biomass and soil nitrogen cycling. *Plant Soil* 437, 241–254. doi: 10.1007/s11104-019-03974-6

Purugganan, M. D., and Fuller, D. Q. (2009). The nature of selection during plant domestication. *Nature* 457, 843–848. doi: 10.1038/nature07895

R Core Team (2017). *R: a Language and Environment for Statistical Computing*. Available at: https://www.r-project.org/ (accessed March 11, 2020).

Resende, M. F. R., Muñoz, P., Acosta, J. J., Peter, G. F., Davis, J. M., Grattapaglia, D., et al. (2012). Accelerating the domestication of trees using genomic selection: accuracy of prediction models across ages and environments (vol 193, 617, 2012). *New Phytol.* 193:1099. doi: 10.1111/j.1469-8137.2011.04048.x

Rutkoski, J., Poland, J., Mondal, S., Autrique, E., González Párez, L., Crossa, J. J., et al. (2016). Canopy temperature and vegetation indices from high-throughput phenotyping improve accuracy of pedigree and genomic selection for grain yield in wheat. *Genes| Genomes| Genetics* 6, 2799–2808. doi: 10.1534/g3.116.032888

Rutkoski, J., Singh, R. P., Huerta-Espino, J., Bhavani, S., Poland, J., Jannink, J. L., et al. (2015). Efficient use of historical data for genomic selection: a case study of stem rust resistance in wheat. *Plant Genome* 8, 1–10. doi: 10.3835/plantgenome2014.09.0046

Rutkoski, J. E., Poland, J. A., Singh, R. P., Huerta-espino, J., Bhavani, S., Barbier, H., et al. (2014). Genomic selection for quantitative adult plant stem rust resistance in wheat. *Plant Genome J.* 7, 1–10. doi: 10.3835/plantgenome2014.02.0006

Shengqiang, Z., Dekkers, J. C. M., Fernando, R. L., and Jannink, J. L. (2009). Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: a barley case study. *Genetics* 182, 355–364. doi: 10.1534/genetics.108.098277

Shrestha, R., Arnaud, E., Mauleon, R., Senger, M., Davenport, G. F., Hancock, D., et al. (2010). Multifunctional crop trait ontology for breeders' data: field book, annotation, data discovery and semantic enrichment of the literature. *AoB Plants* 2010:plq008. doi: 10.1093/aobpla/plq008

Spindel, J. E., and McCouch, S. R. (2016). Viewpoints when more is better?: how data sharing would accelerate genomic selection of crop plants. *New Phytol.* 212, 814–826. doi: 10.1111/nph.14174

Sprunger, C. D., Culman, S. W., Robertson, G. P., and Snapp, S. S. (2018). Perennial grain on a Midwest Alfisol shows no sign of early soil carbon gain. *Renew. Agric. Food Syst.* 33, 360–372. doi: 10.1017/S1742170517000138

Tsvelev, N. (1983). *Grasses of the Soviet Union*. New Delhi: Oxonian Press Pvt. Ltd.

Unamba, C. I. N., Nag, A., and Sharma, R. K. (2015). Next generation sequencing technologies: the doorway to the unexplored genomics of non-model plants. *Front. Plant Sci.* 6:1074. doi: 10.3389/fpls.2015.01074

VanRaden, P. M. M., Van Tassell, C. P. P., Wiggans, G. R. R., Sonstegard, T. S. S., Schnabel, R. D. D., Taylor, J. F. F., et al. (2009). Invited review: reliability of genomic predictions for North American Holstein bulls. *J. Dairy Sci.* 92, 16–24. doi: 10.3168/jds.2008-1514

Varshney, R. K., Ribaut, J. M., Buckler, E. S., Tuberosa, R., Rafalski, J. A., and Langridge, P. (2012). Can genomics boost productivity of orphan crops? *Nat. Biotechnol.* 30, 1172–1176. doi: 10.1038/nbt.2440

Vogel, K. P., and Jensen, K. J. (2001). Adaptation of perennial triticeae to the eastern central great plains. *J. Range Manag.* 54, 674–679. doi: 10.2307/4003670

Wagoner, P. (1990). Perennial grain new use for intermediate wheatgrass. *J. Soil Water Conserv.* 45, 81–82.

White, J. W., Andrade-sanchez, P., Gore, M. A., Bronson, K. F., Coffelt, T. A., Conley, M. M., et al. (2012). Field-based phenomics for plant genetics research. *F. Crop. Res.* 133, 101–112. doi: 10.1016/j.fcr.2012.04.003

Zhang, X., Sallam, A., Gao, L., Kantarski, T., Poland, J., DeHaan, L. R., et al. (2016). Establishment and optimization of genomic selection to accelerate the domestication and improvement of intermediate wheatgrass. *Plant Genome* 9, 1–18. doi: 10.3835/plantgenome2015.07.0059

Conflict of Interest: 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 © 2020 Crain, Bajgain, Anderson, Zhang, DeHaan and Poland. 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) and the copyright owner(s) 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 Plant Science

ORIGINAL RESEARCH
published: 03 April 2020
doi: 10.3389/fpls.2020.00346

Exploration of the Yield Potential of Mesoamerican Wild Common Beans From Contrasting Eco-Geographic Regions by Nested Recombinant Inbred Populations

Jorge Carlos Berny Mier y Teran[1†], Enéas R. Konzen[2†‡], Antonia Palkovic[1], Siu M. Tsai[2] and Paul Gepts[1†]*

[1] Department of Plant Sciences, University of California, Davis, Davis, CA, United States, [2] Cell and Molecular Biology Laboratory, Centro de Energia Nuclear na Agricultura, Universidade de São Paulo, Piracicaba, Brazil

OPEN ACCESS

Edited by:
Petr Smıkal,
Palackı University, Czechia

Reviewed by:
Kirstin E. Bett,
University of Saskatchewan, Canada
Juan M. Osorno,
North Dakota State University,
United States

***Correspondence:**
Paul Gepts
plgepts@ucdavis.edu

†ORCID:
Jorge Carlos Berny Mier y Teran
orcid.org/0000-0003-3709-9131
Enéas R. Konzen
orcid.org/0000-0001-5176-7410
Paul Gepts
orcid.org/0000-0002-1056-4665

†Present address:
Enéas R. Konzen,
Universidade Federal do Rio Grande do Sul, Campus Litoral Norte, Imbei, Brazil

Specialty section:
This article was submitted to Plant Breeding, a section of the journal Frontiers in Plant Science

Received: 31 October 2019
Accepted: 09 March 2020
Published: 03 April 2020

Citation:
Berny Mier y Teran JC, Konzen ER, Palkovic A, Tsai SM and Gepts P (2020) Exploration of the Yield Potential of Mesoamerican Wild Common Beans From Contrasting Eco-Geographic Regions by Nested Recombinant Inbred Populations. Front. Plant Sci. 11:346. doi: 10.3389/fpls.2020.00346

Genetic analyses and utilization of wild genetic variation for crop improvement in common bean (*Phaseolus vulgaris* L.) have been hampered by yield evaluation difficulties, identification of advantageous variation, and linkage drag. The lack of adaptation to cultivation conditions and the existence of highly structured populations make association mapping of diversity panels not optimal. Joint linkage mapping of nested populations avoids the later constraint, while populations crossed with a common domesticated parent allow the evaluation of wild variation within a more adapted background. Three domesticated by wild backcrossed-inbred-line populations (BC_1S_4) were developed using three wild accessions representing the full range of rainfall of the Mesoamerican wild bean distribution crossed to the elite drought tolerant domesticated parent SEA 5. These populations were evaluated under field conditions in three environments, two fully irrigated trials in two seasons and a simulated terminal drought in the second season. The goal was to test if these populations responded differently to drought stress and contained progenies with higher yield than SEA 5, not only under drought but also under water-watered conditions. Results revealed that the two populations derived from wild parents of the lower rainfall regions produced lines with higher yield compared to the domesticated parent in the three environments, i.e., both in the drought-stressed environment and in the well-watered treatments. Several progeny lines produced yields, which on average over the three environments were 20% higher than the SEA 5 yield. Twenty QTLs for yield were identified in 13 unique regions on eight of the 11 chromosomes of common bean. Five of these regions showed at least one wild allele that increased yield over the domesticated parent. The variation explained by these QTLs ranged from 0.6 to 5.4% of the total variation and the additive effects

ranged from −164 to 277 kg ha^{-1}, with evidence suggesting allelic series for some QTLs. Our results underscore the potential of wild variation, especially from drought-stressed regions, for bean crop improvement as well the identification of regions for efficient marker-assisted introgression.

Keywords: common bean, crop wild relative, eco-geographic adaptation, nested backcrossed inbred populations, quantitative trait loci, yield

INTRODUCTION

Among pulses, common bean (*Phaseolus vulgaris* L.; $2n = 2x = 22$) plays an important nutritional and economical role (Broughton et al., 2003; Gepts et al., 2008). The yields of pulses are usually lower than those of cereals, mainly because their production is located in more marginal cultivation niches, produce more energy-dense seeds and the cost of association with nitrogen fixing rhizobia (Sinclair and Vadez, 2012). Production is also constrained by biotic and abiotic factors, drought being one of the main causes of yield reduction and crop failure in beans (Singh, 2001; Beebe et al., 2013; Ramirez-Cabral et al., 2016). Furthermore, drought severity is likely to increase due to the effects of climate change (Prudhomme et al., 2014). Increasing crop yield and resilience is an essential goal of crop breeding and cultivar development, as well as a direct advantage to farmers and, ultimately, to consumers (Kissoudis et al., 2016). Several strategies to improve yield include maximizing nitrogen fixation, photosynthesis and partitioning to grain, as well as minimizing water deficit impacts (Monteith and Moss, 2006; Zhu et al., 2010; Sinclair and Vadez, 2012).

Common bean is one of the five domesticated species among ∼80 wild *Phaseolus* species, all originating in the American continent (Freytag and Debouck, 2002; Delgado-Salinas et al., 2006). Wild common bean originated in Mexico and was dispersed by at least two long-distance movements to Ecuador and northern Peru (some 500,000 years ago) and the southern Andes (southern Peru, Bolivia, and northwestern Argentina; some 100,000 years ago) leading to two additional gene pools in the Andes (Bitocchi et al., 2012; Rendón-Anaya et al., 2017; Ariani et al., 2018; Gepts, 2019). Common bean was domesticated independently within two of the three gene pools (Southern Andes and Mesoamerica), followed by divergence into six genetically distinct races (Kwak and Gepts, 2009). Mesoamerican beans were presumably domesticated in Western Mexico while Andean beans were domesticated in northern Argentina and southern Bolivia (Kwak et al., 2009; Rodriguez et al., 2016). Within the domesticated Mesoamerican gene pool, the larger-seeded races 'Jalisco' and 'Durango' are distributed in the sub-humid and semi-arid highlands, respectively, in Central and Northern Mexico, while the small-seeded race Mesoamerica is distributed in the lowlands from southern Mexico to northeast Brazil (Singh et al., 1991). Within each race, there are thousands of landrace types, as well as fewer, but established, commercial market classes (Gentry, 1969; Singh et al., 1991; Moghaddam et al., 2016). In common bean breeding, the variation included in the breeding programs is mostly constrained within market classes and within races, as the inheritance of color, size, and shape of seeds and plant architecture are highly polygenic and dispersed throughout the genome (Park et al., 2000; McClean et al., 2002; Checa and Blair, 2008; Schmutz et al., 2014). Introgressions between market classes or gene pools have focused mostly on transferring disease resistance, modifying growth habit, and introducing drought resistance (Kelly, 2001; Acosta-Gallegos et al., 2007; Beebe, 2012; Porch et al., 2013; Dohle et al., 2019).

The use of wild variation has been even more limited. Wild variation has been identified as a source of resistance to bruchids (Kornegay et al., 1993; Osborn et al., 2003, 2006), common bacterial blight (Beaver et al., 2012), web blight (Beaver et al., 2012) and white mold (Mkwaila et al., 2011). However, it is possible that beneficial wild variation for highly quantitative important traits, like grain yield and drought adaptation, is not present the domesticated forms due to early genetic bottlenecks (Gepts et al., 1986; Tanksley and McCouch, 1997; Acosta-Gallegos et al., 2007). For example, within the Mesoamerican gene pool, a single domestication event originated in only one of the three genetically and geographically distinct groups of wild common beans (Kwak et al., 2009; Ariani et al., 2018). Although domesticated forms expanded to areas of those non-domesticated, wild groups, and can outcross with the wild relatives, gene flow has been found highly asymmetrical, introgressing more regions from the domesticated to the wild types (Papa and Gepts, 2003). Therefore, diversity unique to the non-domesticated groups might not be present in the domesticates. This is supported by a strong genetic diversity bottleneck, especially during the Mesoamerican bean domestication (Gepts et al., 1986; Sonnante et al., 1994; Schmutz et al., 2014). Previous efforts in genetic analyses in common bean using wild by domesticated crosses include: (a) two populations with the same Andean domesticated cultivar (ICA Cerinza) crossed to a wild type from Colombia (Blair et al., 2006) and to one from northern Mexico (Blair and Izquierdo, 2012); (b) a population resulting from a cross of Midas (an Andean snap bean), and a wild type from central Mexico G12873 (Koinange et al., 1996); and (c) a population involving Peruvian accessions both originating in the Andean gene pool (Singh et al., 2019). In the first two efforts, genomic regions carrying a positive-effect allele for yield from the wild parent were detected in the wild accession from a high rainfall area in Colombia (Blair et al., 2006), but not in the accession from a low rainfall region in northern Mexico (Blair and Izquierdo, 2012).

In the present investigation, the development of three backcrossed-inbred-line populations in a nested design is described, as is their evaluation under drought pressure in field conditions followed by a QTL analysis of grain yield.

The populations result from the cross between three wild Mesoamerican accessions originating from areas with different levels of precipitation/evapotranspiration to the same elite breeding line, also of Mesoamerican origin. The nested design allows the sampling of more diversity than single biparental populations, increases the power and mapping resolution, but most importantly allows the testing of genetic effects among accessions in similar genetic backgrounds (Yu et al., 2008; Garin et al., 2017). Our two-fold hypothesis is that: (1) The three wild accessions have different adaptations to rainfall/evapotranspiration regimes; and (2) Some of these contrasting adaptation mechanisms were not included into the domesticated gene pool during domestication.

MATERIALS AND METHODS

Parental Materials and Population Development

Three populations were established, which resulted from a cross of the Mesoamerican elite breeding line SEA 5 with three wild accessions from the Mesoamerican gene pool. SEA 5 (PI 613166; Number and DOI of the CIAT gene bank: G51502, doi: 10.18730/PHA81) was obtained at the International Center for Tropical Agriculture (CIAT, Cali, Colombia) from an interracial (Durango × Mesoamerica) double cross (BAT 477/San Cristobal 83//Guanajuato 31/Rio Tibagi) within the Mesoamerican gene pool (Singh et al., 2001). SEA 5 was selected for high productivity under drought, has an indeterminate inclined growth habit IIb, photoperiod neutrality, resistance to Fusarium root rot and Bean Common Mosaic Virus (*I* gene), but shows susceptibility to anthracnose, common bacterial blight and rust (Singh et al., 2001; Terán and Singh, 2002). SEA 5 develops a vigorous and deep root system (Polania et al., 2017), has a high capacity of photoassimilate remobilization (Polania et al., 2016; Rao et al., 2017), and a high capacity for nitrogen fixation under drought (Devi et al., 2012).

Three wild accessions within the Mesoamerican wild gene pool (Ariani et al., 2018; Berny Mier y Teran et al., 2019a) were chosen to maximize genetic variation and annual precipitation of their collection site. From North to South and along a gradient of increasing temperature and rainfall, the wild accessions included PI 319441 (G10022, DOI: 10.18730/JRH59), collected in the state of Durango, Mexico (104.47°N, 24.24°W, mean annual temperature: 18°C, mean annual precipitation: 588 mm), PI 417653 (G12910, DOI: 10.18730/PGEK4) from Guanajuato, Mexico (101.72°W, 20.62°N, 19.4°C, 733 mm) and PI 343950 from Huehuetenango, Guatemala (91.82°N, 15.68°W, 23.5°C, 1600 mm). The seeds were obtained from the National Plant Germplasm System (NPGS) of the USDA at the Western Regional Plant Introduction Station in Pullman, WA, United States. Climatic conditions of each collection site were extracted from the WorldClim database[1] (Hijmans et al., 2005).

The three populations (henceforth called dw319441, dw417653, and dw343950) were developed in an identical way as follows: (1) SEA 5 was crossed to the wild accession using SEA 5 as the male parent; (2) The F_1 plant was then used as the male parent crossed back to SEA 5 plants to obtain at least 250 BC_1 seeds; (3) Each BC_1 seed was scarified before planting and allowed to self in the greenhouse; (4) $BC_1S_{1:2}$ families were grown in the field in the summer months (June–September, short night/long day photoperiod) planting ∼50 seeds per family, to allow selection against photoperiod sensitivity. Seeds from a randomly chosen plant from each family were harvested; (5) The plants were grown in the greenhouse for two more cycles using single-seed-descent to obtain 220, 237, and 238 BC_1S_4 plants from the dw319441, dw417653, and dw343950 populations, respectively.

Trial Design and Phenotyping

The populations were evaluated under field conditions in 2014 and 2015 at the Plant Sciences Field Facility on the University of California, Davis, campus (38.53°N, 121.78°W). The soil type of the site belongs to the Yolo series, a member of fine-silty loam, mixed, non-acid, thermic Mollic Xerofluvents, considered well-drained, with slow to medium runoff and moderate permeability[2]. The seeding was carried out on the 8th of June in 2014 and on the 9th of June in 2015. The plants were harvested on the 22–26th of September in 2014 and the 14–25th of September in 2015. In 2014, the populations were evaluated only under full irrigation, using furrow water delivery as needed, in four irrigations. In 2015, a terminal drought water regime was added. Terminal drought was simulated by withdrawing irrigation in the last two of the four irrigations of the full irrigated treatment. Plants received water only from irrigations as there were no rain events during the experiments. The agricultural management was according to standard practices in California (Long et al., 2010).

In both years, the RILs were planted in un-replicated fashion and SEA 5 in replicated fashion (eight replicates) according to a modified augmented design (Lin and Poushinsky, 1985). The field was divided in blocks and planted with a main check in the middle of each block, as well as two to three secondary checks randomly distributed within each block. In 2014, 30 blocks of 3 plots × 9 plots were used and, in 2015, 21 blocks of 5 plots × 5 plots per treatment were used. UCD 9634, a pink-seeded breeding line, was included in both years as the main check due its high yield and stability. In 2014, Tio Canela 75 (Rosas et al., 2010), Matterhorn (Kelly et al., 1999), and Flor de Mayo Eugenia (Acosta Gallegos et al., 2010) were included as secondary checks. In 2015, a small, black-seeded line L88-63 (Frahm et al., 2004) was added as a check. The experimental unit was a plot of 60 plants grown in a row of 6 m-long rows and 0.76 m between rows (density of 131,578 plants per hectare). In 2014, 230, 238, and 237 progeny lines of the dw319441, dw417653, and dw343950 populations, respectively, were grown.

There was a bimodal distribution in days to flowering (**Figure 1**) with a smaller set of lines that showed late flowering. These later genotypes were not used for further evaluation and genotyping, as a wide variation in phenology can be a

[1] www.worldclim.org

[2] https://soilseries.sc.egov.usda.gov

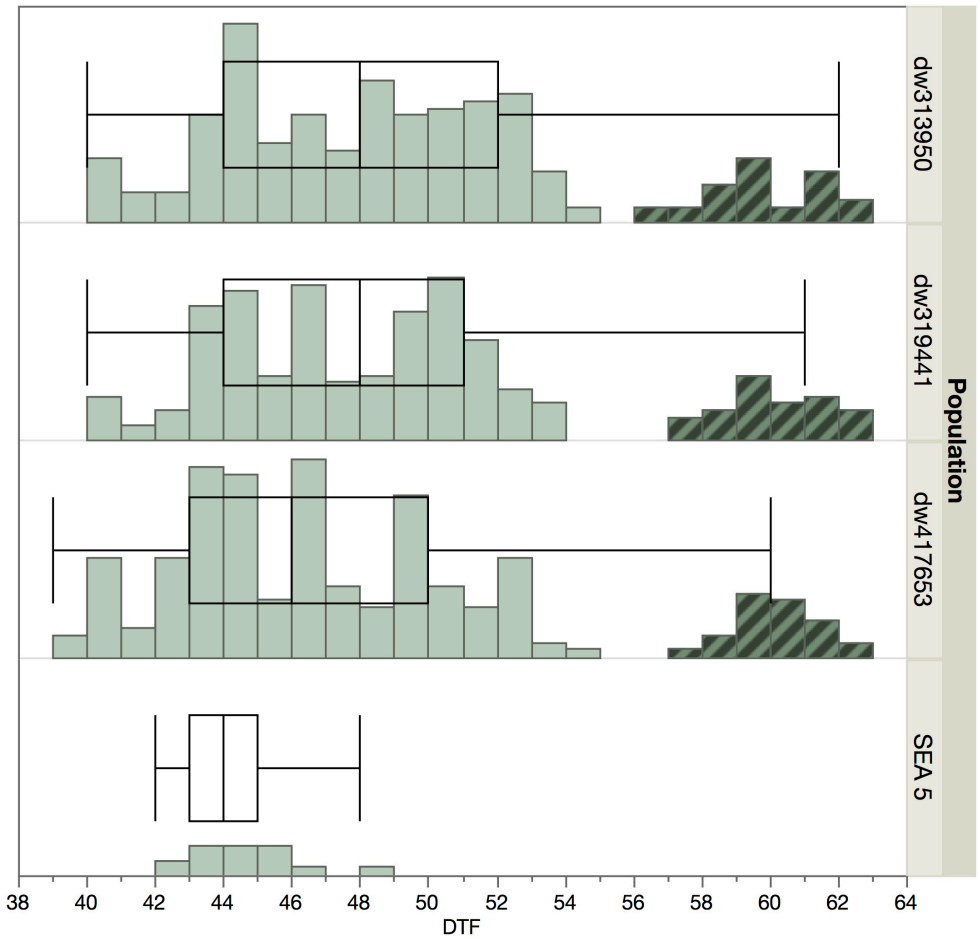


FIGURE 1 | Distribution of flowering time among populations and SEA 5 in 2014. The highlighted accessions in the later part of the distribution were not planted in 2015 to minimize the effect of phenology.

confounding effect (Pinto et al., 2010), especially in the case of late flowering and maturity, which are negatively correlated with yield in beans (Klaedtke et al., 2012; Berny Mier y Teran et al., 2019b). The final analyses used 171, 170, and 165 lines from the dw319441, dw417653, and dw343950 populations, respectively.

Grain yield was measured per experimental plot and extrapolated to kilograms per hectare. Flowering time was taken when at least 50% of the plot had an open flower. Seed weight was evaluated in random subsample of 100 seeds.

Statistical Analyses

To adjust for within-field spatial variation, a two-dimensional tensor product was used, i.e., the penalized splines (P-splines) method with the package SpATS (Rodríguez-Álvarez et al., 2018) in R (R Development Core Team, 2018). Briefly, the method fits the mixed model $y = X\beta + X_s\beta_s + Z_s s + Z_u u + Z_g g + e$, where β and β_S are vectors that include the intercept and check cultivar effect as fixed term, $X_S$is the design matrix, $X_s\beta_s$ and $Z_s s$ are fixed and random components of the mixed model, respectively, s is the vector of random spatial effects, u is a random row and column effect sub-vector accounting for discontinuous variation, and g is the random genotypic effect (Malosetti et al., 2017; Rodríguez-Álvarez et al., 2018). The genotype was considered as a random effect in both years and irrigation treatment as a fixed effect in the second year. An analysis of variance was carried out as a linear mixed model using population, environment, and their interaction as fixed effects, and genotype within population as random effects. The evaluations in 2014 under full irrigation, in 2015 under terminal drought and full irrigation were considered as separate environments. The statistical analyses were performed with JMP (SAS Institute Inc., 2016). The Drought Susceptibility Index (DSI) calculated from the two irrigation treatments in 2015 as $(1 - Y_{ds}/Y_{ns})/1 - (X_{ds}/X_{ns})$, where Y_{ds} and Y_{ns} are yields in drought stress and no stress environments, respectively, and X_{ds} and X_{ns} are the overall yield under drought stress and no stress treatments, respectively (Fischer and Maurer, 1978; Beebe et al., 2013).

Genotyping

The BC-RILs, parents, and F_1 hybrids were genotyped with the BARCBean6K_3 BeadChip platform of 5,398 SNP markers (Song

et al., 2015) at the Genome Center at the University of California-Davis. After filtering in GenomeStudio Module v1.8.4 (Illumina Inc., San Diego, CA, United States), SNP calling was performed with the software's cluster algorithm, with subsequent manual adjustments and a quality control with a 0.15 Gencall score cutoff. The markers were filtered for less than 10% missing data and polymorphism between parents, verified with the F_1 hybrids.

Map Construction

Linkage maps were constructed for each population in R (R Development Core Team, 2018) using the package asMAP (Taylor and Butler, 2017). Genetic distances were determined using the Kosambi function (Kosambi, 1944). The recombination fraction was estimated with qtl (Wu et al., 2003). A consensus map, combining results for the three individual maps, was developed with LPmerge (Endelman and Plomion, 2014), with the linkage maps of the three populations constructed without first filtering for co-located markers, to maximize the markers shared among populations. LPMerge uses linear modeling to keep the maximum interval with the lowest root mean squared error applying weights to population size (Endelman and Plomion, 2014). When there were co-localized markers in the consensus map, only one marker per bin was kept. Chromosomes were numbered Pv01 to Pv11 matching the standard numbering in *P. vulgaris* (Pedrosa-Harand et al., 2008).

QTL Mapping

The QTL analysis was performed with the NAM function of QTL IciMapping version 4.1 (Meng et al., 2015). The NAM module implements joint inclusive composite interval mapping (JICIM), uses generalized linear models with population and marker by population interaction as fixed effects through stepwise regression for marker selection and subsequent interval mapping with adjusted phenotypes from the selected markers outside the current marker interval (Li et al., 2007, 2011; Wang et al., 2016). A probability of 0.01 and a step of 1 cM was used for the stepwise regression and the significance. The LOD threshold was calculated by 2,000 permutations in each environment at a significance level of 0.05. Furthermore, a QTL analysis was also conducted for the average yield across the three environments in 2014 and 2015 and the drought susceptibility index (DSI).

Data Availability

The marker segregation and phenotypic evaluation data have been deposited in the Dryad public database: https://datadryad.org/stash/dataset/doi:10.25338/B8FW3M.

RESULTS

Sources of Variation

The effects of population, environment, and their interaction were highly significant (**Table 1**). Nevertheless, the environmental effect was larger (F = 1,672, $P < 0.001$) than both population and their interaction (F = 5.7 and 3.6, at $P < 0.01$, respectively). Across environments, the yield of the dw417653 population was the highest, being significantly higher (113 kg, +11%) than the yield of the dw343950 population, whereas the dw319441 population was not statistically different from other two populations. Across populations, the yield in 2014 was significantly lower than both water regimes in 2015, with a reduction of 56 and 47%, relative to drought and well-watered treatments, respectively. There was a significant yield reduction in 2015 because of terminal drought, with 17% lower yield relative to the irrigated treatment.

TABLE 1 | Descriptive statistics and analysis of variance for yield (kg ha^{-1}) for three backcrossed inbred lines and the recurrent parent evaluated in three environments.

RIL	Mean	1082.3
	Min	469.1
	Max	1987.2
SEA 5	Mean	1575.7
F-values	Population	5.7**[1]
	Environment	1672.5***
	PxE	3.6**
Population mean	dw319441	1084.1 AB
	dw417653	1137.1 A
	dw343950	1024.1 B
Environmental means	2014C[3]	624.5 C[2]
	2015C	1434.0 A
	2015D	1186.7 B

*[1]Significance codes: ***<0.001, **< 0.01. [2]Levels not connected with the same letter are significantly different according to the Tukey-Kramer HSD test (P < 0.05). [3]Treatments: full irrigation (C), terminal drought (D).*

Comparison of Populations Across Environments

Although the population by environment effect was significant (**Table 1**), the ranks of the performance of the populations were similar between environments. Across environments, the dw417653 population produced higher yield than the dw319441 population, which, in turn, was higher than the dw343950 population. However, population dw417653 was significantly different from population dw343950 only in 2015 (**Figure 2**). In all environments, the average of the population yields was lower than that of SEA 5. Across environments, SEA 5 produced 45% more yield than the average of the RILs (**Table 1**).

For breeding purposes, some RILs showed transgressive segregation for higher yield compared to the average of SEA 5 plots in 2015, although none were higher than the highest yielding SEA 5 plot in 2014 (**Figure 2**). However, in 2015 in both treatments, some RILs yielded more than the highest plot of SEA 5. Furthermore, there were high-performing RILs from each population in both years and treatments (**Figure 2**). Two populations, dw417653 and dw319441, included progeny lines with yields that were significantly higher (+20% on average) – across the three environments – than that of the domesticated control SEA 5 (**Table 2**). These lines, nevertheless, had similar number of days to flowering (42 to 54 days vs. 47 for SEA 5) and only slightly smaller seeds than SEA 5) (15–20 g/100 seeds vs. 24 g/100 seeds for SEA 5; **Table 2**).

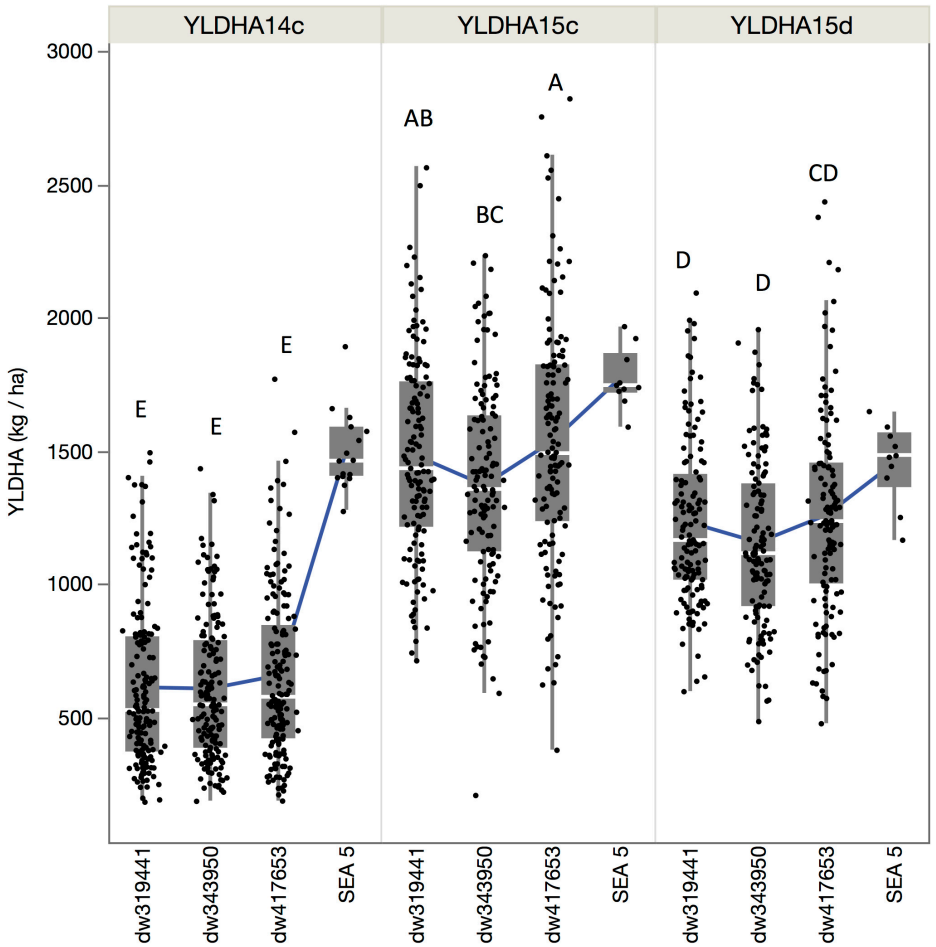


FIGURE 2 | Quantile boxplot and overall mean (green line) of the yield of the three populations and recurrent parent for the three environments: fully irrigated in 2014 (YLDHA14c), fully irrigated in 2015 (YLDHA15c), and terminal drought in 2015 (YLDHA15c). Averages not connected with the same letter are significantly different according to the Tukey-Kramer HSD test ($P < 0.05$).

Correlations Among Traits and Distribution

The correlation of yield between the 2014 and 2015 well-watered treatments and the 2015 terminal drought treatment was relatively low, with an R of 0.51 and 0.5, respectively (**Figure 3**), but the correlation between drought and well-watered conditions in 2015 was higher ($R = 0.88$). The overall distributions of yield among the RILs in the first year and in 2015 under drought were not normally distributed ($P < 0.001$, Shapiro–Wilks test) but were skewed toward the low- yielding side of the distribution, while the distribution of yield in well-watered plots in 2015 was normally distributed ($P = 0.17$).

Molecular Linkage Map and QTL Analyses

Polymorphism, Recombination Rate and Allele Frequency

There were 1554, 1404, and 1499 polymorphic markers between SEA 5 and PI 343950, PI 417563, and PI 419441, respectively. Jointly, there were 2,201 polymorphic markers between the wild accessions and SEA 5 and 1,858 markers shared among the three wild accessions. The trends in recombination rate were similar across populations. Higher recombination rates were observed in the distal parts of all the chromosomes, except on chromosome Pv06 where higher recombination was observed from the middle to the distal portion and chromosome Pv09 where higher recombination was located in the second and fourth quarter of the chromosome (**Figure 4**).

The wild allele frequency was variable between and within chromosomes. After one backcross to the domesticated parent, one would expect an average frequency of 0.25 of the wild parent. The average frequency was 0.23 across chromosomes, with a range of 0.17 to 0.4, in Pv01 and Pv02, respectively. The populations had similar frequencies across chromosomes; however, within chromosomes there were differences among populations. For example, in Pv02, compared to the other two populations, the dw417653 population showed higher wild allele frequency in the central region but very low frequencies at the chromosomal ends. In Pv10, the P319441 population had

TABLE 2 | Progeny, backcross-inbred lines with highest yields (kg ha^{-1}) across the three environments [fully irrigated in 2014 (YLDHA14), fully irrigated in 2015 (YLDHA15c), and terminal drought in 2015 (YLDHA15d)].

Population	Lines	100SW[1]	DTF[2]	Average over the three environments		2014 (well-watered)		2015 (well-watered)		2015 (drought-stressed)	
				YLDHAave	Rank	YLDHA14	Rank	YLDHA15c	Rank	YLDHA15d	Rank
dw417653	p53-36	20	46	**1987**	1	1286	5	**2824**	1	**2379**	2
	p53-235	20	42	**1889**	2	**1572**	1	**2610**	3	**1956**	5
	p53-114	20	43	**1866**	3	1463	3	**2527**	5	**2063**	3
dw319441	p41-262	15	48	**1856**	4	1461	4	**2566**	4	**1993**	4
dw417653	p53-65	18	54	**1856**	5	824	6	**2756**	2	**2437**	1
Control	SEA 5	24	47	1576	6	1500	2	1773	6	1455	6
Difference between best progeny lines and SEA 5 yields				+20%		−12%		+50%		+49%	

Yield values in bold represent yields that are significantly higher than those of the control. [1] *100SW, 100-seed weight (g).* [2] *DTF, number of days to flowering.*

higher wild allele frequency than the other populations. In some chromosome regions, all three populations showed a lower than expected wild allele frequency, like in the distal region of Pv01. There were two regions with markedly low wild allele frequency, the region mentioned before in Pv02 and another region in Pv07, both at the end of the chromosomes.

Consensus Molecular Linkage Map

The consensus map was built with 721 markers and had a genetic length of 925 cM (**Table 3**). Per chromosome, there were on average 66 markers, an average length of 84 cM, and an average spacing of 1.3 cM. The average maximum spacing was 9 cM, while the largest interval was 14 cM in Pv01. Through a comparison of the range of the markers on each chromosome and their physical position according to the G19833 reference genome version 2.1 (Schmutz et al., 2014), the genetic map spanned 510,318,067 bp, which represents 99% of the sequenced genome. The coverage ranged from 96.5% of Pv06 to 99.7% of Pv08.

Identification of Additive QTLs and Their Distribution on the Molecular Linkage Map

QTL analyses were performed for grain yield in each environment, consisting of 2 years grown in full irrigation, terminal drought in the second year, and the average yield across environments. The same analysis was conducted for the DSI calculated from the drought and well-watered conditions in 2015 (**Table 4** and **Figure 5**). Significance thresholds for LOD scores, calculated by permutations for each environment, were 3.7 (environment 2014), 3.6 (2015c), and 3.6 (2015d). LOD scores for averages across environments had a threshold of 4.7. The threshold LOD score for DSI was 12.1. Twenty QTLs had LOD scores about the respective thresholds. They were distributed on eight chromosomes. These included three QTLs in 2014, five in 2015c, six in 2015d, five for the average across environments, and one for DSI.

The magnitude of these QTLs ranged from 0.6% to 5.4% with an outlier at 38% (for DSI on chromosome 9, observed in the dw319441 and dw343950 populations). The allelic effects of the chromosome regions marked by these QTLs varied considerably among the three accessions between negative and positive values. The most negative value (−164.8 kg ha^{-1}) was for yield QTL Pv07.55 in 2015, inherited from PI 343950. The largest positive allelic effect (277.4 kg ha^{-1}) was observed in well-watered 2015 conditions and was inherited from PI 417653. Overall, the average allele effect of PI 343950 was the lowest in all environments and in the mean across environment (**Figure 6** and **Table 4**). In contrast, the allele effect of PI 319441 was higher than that of PI 417653 in 2014 and 2015, both under full irrigation, while the average allele effect of PI 417653 was higher than that of PI 319441 under drought in 2015. In addition, the allele effects across genotypes were relatively higher under drought than under full irrigation in 2015. The average QTL significance interval was about 0.8 Mbp, ranging between 0.07 Mbp (on Pv08) and 2.18 Mbp (on Pv01). Many but not all QTL intervals appeared to be located toward the extremities of the chromosomes (**Table 4**).

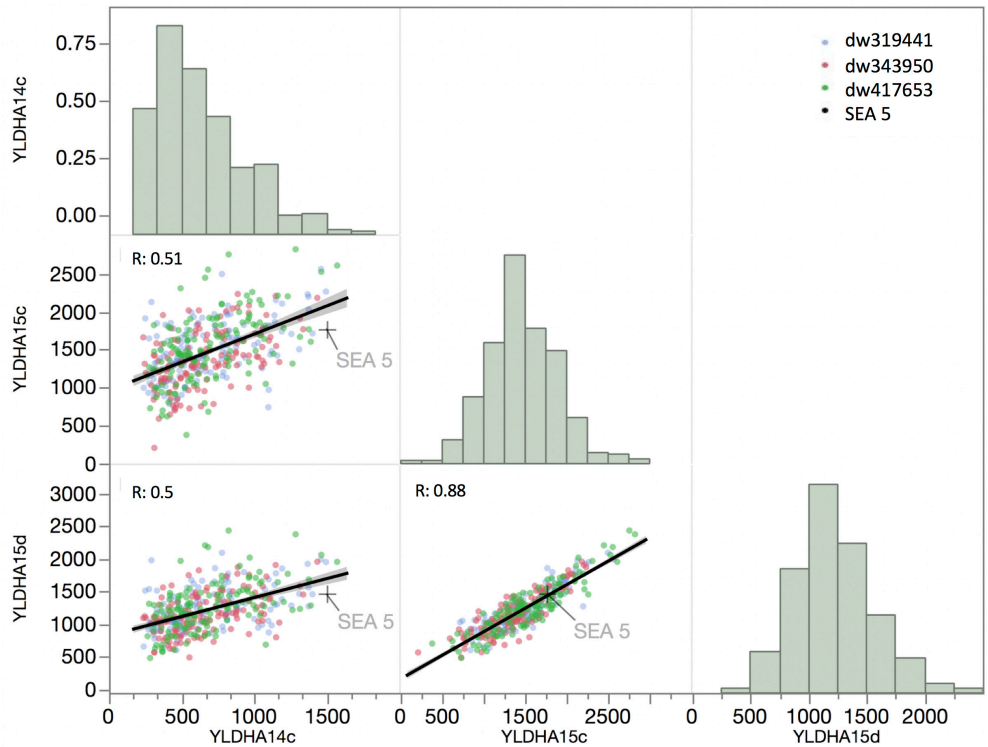


FIGURE 3 | Matrix of trait frequency distribution (diagonal), joint distribution and correlation coefficient (lower triangle) of yield evaluated in 2014 under full irrigation (YLDAH14c) and in 2015 under full irrigation and terminal drought (YLDAH15c and YLDAH15d, respectively). Traits. The mean of SEA 5 is showed as plus (+) sign.

DISCUSSION

Population Development and Segregation Distortion

The main objective of the investigation was to survey eco-geographic adaptive variation in wild beans as a source of novel alleles to increase productivity, as well as to test if these alleles have differential expression under drought constraints. A resource commonly used for genetic studies are diversity panels, which allow a large sample of the variation to be tested and higher genetic resolution to be obtained. Prior to this study, a panel of wild accessions of the Mesoamerican gene pool was evaluated in a greenhouse setting revealing phenotypic variation in root and shoot traits, as well as genomic regions controlling these traits (Berny Mier y Teran et al., 2019a). However, these studies are limited by the confounding effects of population structure and relatedness, as well as low power of detection of rare alleles (Anderson et al., 2011; Bazakos et al., 2017). Population structure is even more geographically constrained in wild bean populations than in its domesticated forms, as dispersal and intercrossing between wild populations are limited compared with domesticated populations (Papa and Gepts, 2003; Zizumbo-Villarreal et al., 2005). In addition, some wild forms, including in common bean, are not adapted to cultivated conditions, due to their profuse and extended climbing growth habit and photoperiod sensitivity (Acosta-Gallegos et al., 2007).

Multiparent populations have several advantages over biparental populations for genetic studies: an increase in allelic diversity, sample size at a specific locus (increasing the power of detection), and mapping resolution (Garin et al., 2017; Holland, 2015). Various schemes for multiparent populations have been proposed; they are mostly divided into two groups: intermating of many parental lines and interconnected biparental populations (Fouilloux and Bannerot, 1988; Cavanagh et al., 2008; Buckler et al., 2009; Holland, 2015). Here, three populations were developed from domesticated by wild crosses using an elite domesticated breeding parent, selected for drought tolerance, as the common parent and three wild accessions from a range of rainfall conditions from the driest part of the distribution in Durango, northern Mexico, to a high-rainfall region in highland Guatemala. A domesticated genotype as the common parent allows for comparison of QTLs contributed by the different wild genotype relative to each other and to those of the domesticated parent. Domesticated × wild nested populations have been developed in maize (Liu et al., 2016) and barley (Schnaithmann et al., 2014; Maurer et al., 2015; Nice et al., 2016). In all cases, the populations were developed after one to four backcrosses to the domesticated parent.

Inbred-backcrossed lines facilitate QTL detection (Kaeppler, 1997) as the lines are more homogeneous and their interaction with other traits can be defined more precisely. For example, the benefit of earliness as an escape for terminal drought can be better assessed in this type of population after measuring

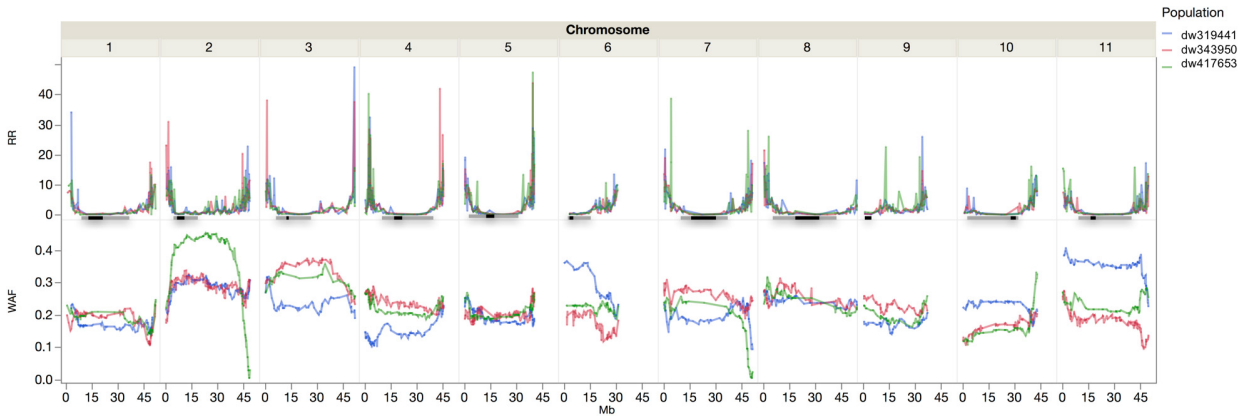


FIGURE 4 | Recombination rate (RR, cM Mb^{-1}) and wild allele frequency (WAF) of the three wild by domesticated populations. The horizontal lines represent the approximate location of the pericentromere (gray) and centromere (black) according to Schmutz et al. (2014).

earliness and yield under different irrigation regimes (Beebe et al., 2014; Polania et al., 2017). In addition, if superior lines are identified in domesticated × wild inbred backcross populations, few or no additional backcrosses are needed for breeding use (Tanksley and Nelson, 1996). However, if a trait is controlled epistatically by one or more loci without independent additive effect in the wild, it will be more difficult to identify such epistatic interactions in backcross populations (Tanksley and Nelson, 1996; Kaeppler, 1997; Johnson and Gepts, 2002). Only one backcross generation was used to limit the loss of detection power of the effect of wild alleles as every added backcross decreases the number of alleles from the wild parent and, hence, the power of QTL detection (Kaeppler, 1997).

Nested populations have been developed and used to study an array of traits in maize (Buckler et al., 2009; Wu et al., 2016; Xiao et al., 2016), soybean (Song et al., 2017), rice (Fragoso et al., 2017), barley (Saade et al., 2016), wheat (Bajgain et al., 2016), and common bean (Hoyos-Villegas et al., 2016), among others. While wild by domesticated nested populations exist in maize (Liu et al., 2016) and barley (Schnaithmann et al., 2014; Maurer et al., 2015; Nice et al., 2017), ours is the first domesticated by wild nested population in common bean. Besides the beneficial alleles for yield, these three populations could be of great use as breeding material and for future evaluations as the domesticated parent and the wild accessions can be polymorphic for other potentially useful traits. In evaluations of wild germplasm, PI 319441 was found to have high sulfur amino acid content in the seed and a high degree of protein hydrolysis after cooking (Montoya et al., 2008), a high content in polyphenols (Espinosa-Alonso et al., 2006), a high iron concentration (Blair et al., 2013) and, thus, could be used as a source for improved nutrition. PI 417653 had a high root efficiency ratio (total P content: root area) in low nitrogen conditions (Araújo et al., 1998), a high level of resistance to cucumber mosaic virus (Griffiths, 2002), and medium tolerance to salinity during early vegetative growth (Bayuelo-Jiménez et al., 2002).

The average wild allele frequency in our populations was 0.23, very close to the expected frequency of 0.25 in a biparental population with a single backcross. Allele frequencies significantly different from the expected frequency can be due to genetic mechanisms of segregation distortion, selection against photoperiod sensitivity and late flowering, or unintended selection. Processes that lead to segregation distortion include gametic incompatibility, genetic load, and asymmetric allelic inheritance in heterozygotes (Bodénès et al., 2016; Lyttle, 1991) besides the selection applied to the populations during their development. Genomic regions of low wild allele frequency were found in various chromosomes. In Pv01, an area of low frequency in all three populations was found around the 48 Mb position. The low frequency of wild alleles is likely due to the presence of a photoperiod sensitivity gene, *Ppd*, identified in this region (Koinange et al., 1996; Kwak et al., 2008; Weller et al., 2019) and selected against during population development (**Figure 1**). Segregation distortion in this region has also been found in two biparental wild by domesticated common bean populations (Blair et al., 2006; Blair and Izquierdo, 2012).

At least two other photoperiod loci have been identified, but have not been mapped as yet: the locus *Hr*, which is recessive and hypostatic to *Ppd* (Gu et al., 1998; Kwak et al., 2008) and the locus *Tip*, which is also recessive and increases earliness at cooler temperatures in long daylength (White et al., 1996). *Tip* might be allelic to either *Ppd* or *Hr* (White et al., 1996). Other regions controlling quantitatively photoperiod sensitivity have been located on Pv03 and Pv04 (Wallach et al., 2018).

There were other regions almost devoid of wild alleles at the end of chromosome Pv02 in one population (dw417653) and at the end of Pv07 in two populations (dw319441 and dw417653). Distortion in the latter region was also observed by Blair et al. (2006). In contrast, some regions in Pv02 and Pv11 showed a wild allele frequency of 0.4, that is, higher than the expected wild allele frequency after a single backcross. Nevertheless, although QTL analyses assume low segregation distortion, including distorted markers does not necessarily increase false positives or bias the effect and position, especially in large populations (Xu, 2008).

TABLE 3 | SNP distribution among the linkage groups/chromosomes of the consensus map.

Chromo-some	Number of markers	Length of linkage groups	Average spacing	Maximum spacing	Physical range		Physical length	Recombination distance	Genome coverage
		cM	Marker cM^{-1}	cM	bp	bp	bp	Kb/cM	%
1	52	98.0	1.9	13.6	497,265	52,165,138	51,667,873	527.5	99.0%
2	111	102.1	0.9	6.0	195,928	49,033,712	48,837,784	478.4	99.6%
3	58	95.8	1.7	11.4	61,877	52,141,173	52,079,296	543.9	99.6%
4	74	80.6	1.1	5.6	90,666	45,793,192	45,702,526	567.0	99.4%
5	75	87.0	1.2	10.2	89,353	40,526,191	40,436,838	465.0	99.1%
6	48	61.6	1.3	6.8	521,607	31,379,071	30,857,464	501.1	96.5%
7	65	92.3	1.4	7.4	61,520	51,592,532	51,531,012	558.1	99.6%
8	82	96.4	1.2	9.5	159,002	59,613,870	59,454,868	616.9	99.7%
9	52	67.1	1.3	5.6	582,540	37,374,817	36,792,277	548.5	98.2%
10	44	57.4	1.3	9.3	190,638	43,194,406	43,003,768	749.7	99.4%
11	60	86.6	1.4	13.4	51,396	50,005,757	49,954,361	576.8	99.2%
Average	65.5	84.1	1.3	9.0				557.5	99.0%
Total	721	924.7					510,318,067		

Field Evaluation

The three populations were evaluated under field conditions in 2 years under optimal irrigation conditions. In the second year, a terminal drought stress was imposed by withdrawing the final two irrigations. Thus, the three populations were tested in a total of three environments. The analysis of variance showed a significant effect of population, environment, and their interaction, for yield production. The environmental effect was larger than the population and interaction effects. The overall yield in 2014 was almost half that of 2015 and the correlation between treatments was higher between drought and full irrigation in 2015, than between full irrigation in 2014 and full irrigation in 2015 and drought in 2015. This might be explained by the effect of hot weather experienced during flowering in 2014. In addition, terminal drought resulted in a marked 17% yield reduction relative to the irrigated treatment. Nevertheless, the ranks of the populations within environments were similar. The dw417653 population was higher-yielding than the dw319441 populations, which, in turn, was higher-yielding than the dw343950 population.

The wild parent of the latter population originated in the warmest and wettest climate (23.5°C, 1600 mm) of the three wild parental lines, suggesting that variation for increased productivity can be found in drier and cooler areas. However, the wild parent of population dw417653, which had the highest yield, originated in a slightly warmer (19.4°C vs. 18°C) and wetter (733 mm vs. 588 mm) environment as the second-ranked population (PI 319441). Thus, there may not be a linear relationship between the aridity of the environment of origin and the ability to increase yields in domesticated × wild crosses. Other factors may play a role in addition to aridity adaptation, such genetic distance between the wild accession and the domesticated gene pool. PI 417653 (G12910) has been implicated in the Mesoamerican domestication of common bean (Kwak et al., 2009).

Map Development and Joint Linkage QTL Analysis

A consensus map from the three populations was built, which was adequately dense for joint linkage analysis and subsequent QTL mapping (**Table 3**). Furthermore, the map covered 99% of the G19833 reference genome (Schmutz et al., 2014). Through joint linkage mapping, 20 QTLs for grain yield were identified in the individual environments (year and irrigation regime), the overall grain yield across environments, and the drought susceptibility index calculated from the drought effect in the second year. Among the 20 QTLs, there were 13 non-overlapping QTLs. Two QTLs on chromosome Pv01 at 76 and 81 cM were tightly linked genetically and physically and had similar additive effect patterns among wild alleles. It is possible, therefore, that they are the same QTL. On the same chromosome, a QTL was mapped at the 72 cM position; however, the pattern of additive effects was reversed from that at the 76 and 81 cM positions, suggesting a different QTL. On Pv07, QTLs at the 78 and 87 cM positions had different additive effect patterns for two of three parental loci suggesting these two QTLs are distinct.

TABLE 4 | Summary of the QTL analysis for yield (kg ha^{-1}) and drought susceptibility index (DSI) evaluated in three environments (2014 and 2015 under full irrigation, and 2015 under terminal drought).

Chromo-some	Position (cM)	Year[1]	Treatment	LOD score	PVE[2]	Additive effect (kg ha^{-1}) relative to SEA 5			Confidence interval				Potentially overlapping QTLs[3]
						PI 319441	PI 417653	PI 343950	Marker		Position (Mbp)		
1	45	2015	Well-watered	3.8	1.8	−8.3	−135.5	−36.5	ss715641107	ss715647281	36.46	38.63	
1	72	2015	Drought	4.6	1.0	99.4	0.0	111.0	ss715646589	ss715645924	48.52	48.90	Trapp et al. (2016)
1	76	2015	Well-watered	5.7	4.6	54.3	190.3	−59.7	ss715645914	ss715645869	48.99	49.48	
1	81	2015	Drought	6.0	4.8	−47.0	277.4	−7.4	ss715645252	ss715645251	50.22	50.30	
3	72	Mean		6.1	3.5	87.2	0.0	62.4	ss715639323	ss715639320	47.53	47.95	
3	72	2015	Well-watered	5.2	1.7	105.2	0.0	96.0	ss715639323	ss715639320	47.53	47.95	
4	33	Mean		4.8	1.9	−51.0	−44.5	−61.1	ss715650237	ss715639357	7.97	9.30	Blair et al. (2006)
4	33	2014	Well-watered	3.9	3.5	−85.1	−62.3	−56.6	ss715650237	ss715639357	7.97	9.30	
5	68	2015	Well-watered	4.5	1.3	−105.4	−60.4	−65.7	ss715645447	ss715645308	38.52	38.72	Trapp et al. (2016)
7	55	2015	Well-watered	8.0	2.6	−46.3	−12.7	−164.8	ss715640487	ss715647648	39.30	39.94	Hoyos-Villegas et al. (2016)
8	54	Mean		7.6	5.4	−75.4	−110.0	−73.5	ss715649162	ss715649744	48.44	48.51	
8	54	2014	Well-watered	4.6	4.7	−19.7	−96.7	−81.6	ss715649162	ss715649744	48.44	48.51	
8	78	2015	Drought	6.7	2.4	111.0	198.6	118.7	ss715646502	ss715646515	57.32	57.55	
8	87	2015	Drought	7.9	1.9	−33.2	115.2	−123.3	ss715646762	ss715646767	58.78	58.94	
9	33	2015	DSI	20.9	38.0	0.6	0.0	0.4	ss715645146	ss715639258	23.96	25.49	Mukeshimana et al. (2014), Berny Mier y Teran et al. (2019b)
9	33	Mean		6.1	3.2	−65.5	0.0	−80.1	ss715645146	ss715639258	23.96	25.49	
9	33	2014	Well-watered	4.1	4.1	−68.8	0.0	−90.9	ss715645146	ss715639258	23.96	25.49	
10	46	Mean		5.1	2.8	17.2	−52.5	−71.1	ss715645524	ss715645496	41.21	42.15	Hoyos-Villegas et al. (2016)
10	46	2015	Well-watered	3.6	3.2	116.4	−46.7	−102.0	ss715645524	ss715645496	41.21	42.15	
10	46	2015	Drought	3.8	0.6	10.0	−25.6	−100.5	ss715645524	ss715645496	41.21	42.15	

[1]Mean: average of the three treatments. [2]Percentage of variation explained. [3]For further explanations, see the section "Discussion."

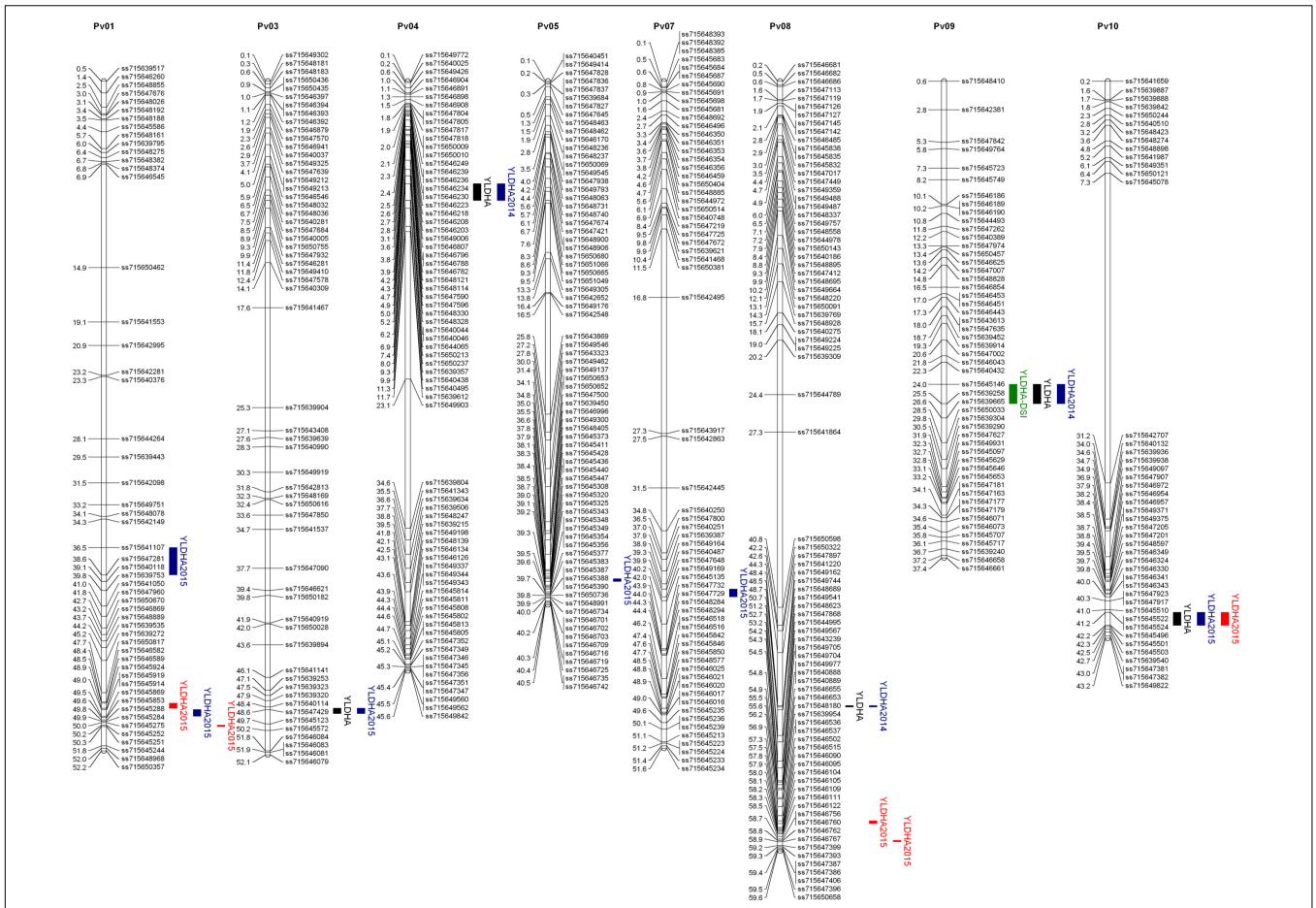


FIGURE 5 | Joint chromosomal linkage map and QTLs for grain yield (YLDHA, kg ha^{-1}) and drought susceptibility index (DSI). The bars in blue represent the environments under full irrigation, red for terminal drought, black for the mean yield, and green for DSI. The position in the chromosome are the physical positions of the markers in Mbp and plotted with MapChart (Voorrips, 2002).

From the 20 QTLs, three were detected in 2014, five in 2015 under full irrigation, six in 2015 under terminal drought, five in the overall average across environments and one for DSI. From these, four QTLs were unique to 2015C and three to 2015D. The variation explained by the QTLs ranged from 0.6 to 5.4%, expected for a highly polygenic trait such as yield (Johnson and Gepts, 2002; Blair et al., 2006). For all QTLs that were expressed in different environments, the additive effects were of similar sign, suggesting that there were no tradeoffs between years or treatments. Although for most QTLs the effect of the wild alleles showed the same sign, for some, e.g., Pv08.87, the allele of PI 417653 had a positive effect (115 kg ha^{-1}) while the alleles of PI 343950 and PI 319441 had a negative effect (−123 kg ha^{-1} and −33 kg ha^{-1}, respectively). This observation suggests that this QTL consists of an allelic series at one locus or represents several, linked loci (Buckler et al., 2009). The detection of allelic series is one of the advantages of nested populations compared with bi-allelic variation in genome-wide association of diversity panels or other multiparent populations (Brachi et al., 2011).

Furthermore, within and across environments, the alleles of PI 343950, the wild population from the wettest location, had the lowest average additive effect. Conversely, the average allelic effects of the accessions from the drier habitats were higher: PI 319441 alleles in both 2014 and 2015, both under full irrigation, whereas PI 417653 alleles were higher under drought in 2015. The overall allele effect of PI 319441 was higher than that of PI 417653 in 2014 and 2015, both under full irrigation, while the average allele effect of PI 417653 was higher than that of PI 319441 under drought in 2015. This suggest that drought tolerance could be a driver of local adaptation. Some traits, like deeper rooting, water use efficiency, earliness in flowering and maturity (Berny Mier y Teran et al., 2019a), might be also beneficial under cultivation if growing seasons are shorter. Berny Mier y Teran et al. (2019a) observed that PI 417653, the wild accession with the strongest allelic effects under drought, has deeper roots and the fastest early growth (Days to the V3 stage) compared to the other two wild accessions. On the contrary, high rainfall conditions could increase selection pressure toward more vigorous vegetative growth and higher disease resistance.

There are two previously published wild by domesticated yield QTL analyses. Blair et al. (2006) evaluated a cross between a domesticated Andean cultivar (ICA Cerinza) and a wild type

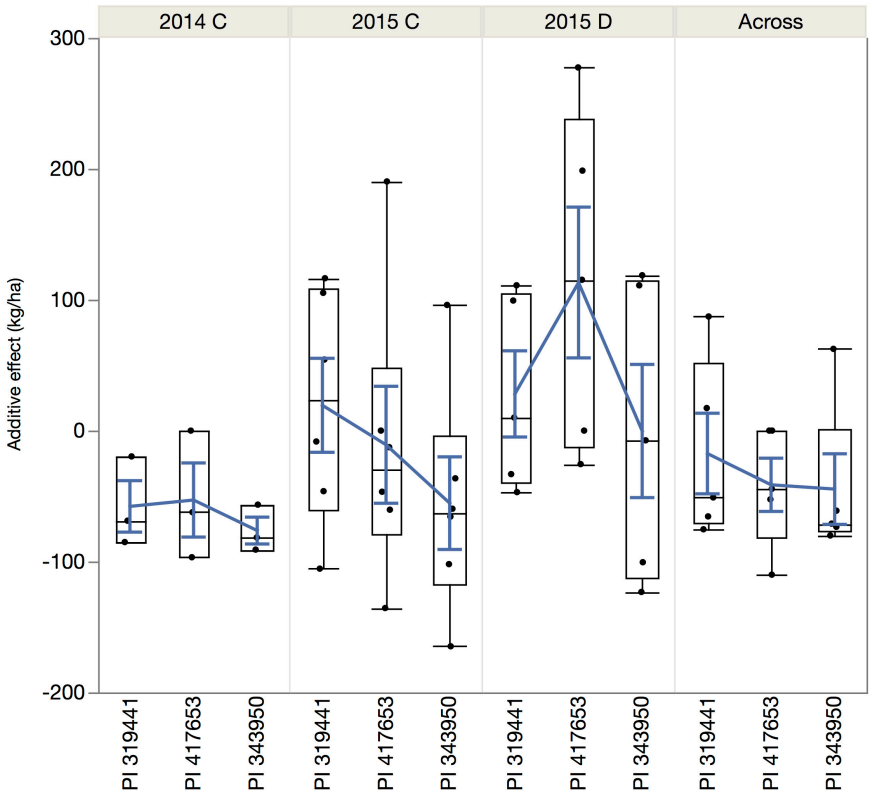


FIGURE 6 | Quantile boxplot and overall mean (blue line) of the additive effect of the alleles of the wild accessions in comparison to the reference allele from SEA 5, the shared domesticated parent.

from Guatemala (G24404). They found nine QTLs for grain yield, with four of them having the wild allele increasing the trait. A QTL on Pv04 overlapped with our findings, which was detected in 2014 and as average across the three environments (**Table 4**). Their confidence interval was 0.4 to 9.5 Mbp while the QTL identified in this study was located within a 7.9 to 9.3 Mbp interval (**Table 4**). However, the additive effect of the G24404 accession was positive while the alleles from the three wild sources in our study had a negative effect. It is possible that the allele from the Andean gene pool in ICA Cerinza had a relatively smaller effect than that of SEA 5 and the wild accessions. Blair and Izquierdo (2012) evaluated a population of the same domesticated cultivar (ICA Cerinza) crossed to G10022 (PI 319441), one of the parents of the nested populations studied here. They found one QTL on Pv05, with the domesticated allele having the positive effect. This QTL did not overlap with Pv05.68 of the current investigation. By comparing the current results with our previous evaluation of a panel of wild germplasm (Berny Mier y Teran et al., 2019a), the QTL in Pv10 located in the interval of 41.2 to 42.1 Mb was situated near a SNP at 38.3 Mb associated with total biomass in the wild panel. This QTL showed a positive additive effect on yield resulting from one parental allele (PI 417653) but a negative effect from the other two parents, suggesting that this genomic location might be involved in local adaptation within the wild germplasm.

Two domesticated by domesticated mapping populations have been developed using SEA 5 (Briñez et al., 2017; Mukeshimana et al., 2014). The latter found three QTLs for yield, with one QTL at Pv09 (SY9.1, confidence interval of 25.1–27.1 Mb) overlapping with our findings (Mukeshimana et al., 2014). The allele of SEA 5 had a negative effect compared to a parental line of the Andean gene pool, while SEA 5 had a positive effect when compared to the wild types in the present study. Briñez et al. (2017) developed a population based on SEA 5 crossed to an Andean accession. They found two QTLs for grain yield, which did not overlap with our findings. Within other published QTL analyses for yield, our findings overlapped with QTLs found by Trapp et al. (2016) in Pv01 (SY1.1) and Pv05 (SY5.1), by Hoyos-Villegas et al. (2016) in Pv07 (SY7.4) and Pv10 (SY10.1), and by Berny Mier y Teran et al. (2019b) in Pv09. The latter QTL also overlapped with the SY9.1 QTL found by Mukeshimana et al. (2014).

From the 13 non-overlapping QTLs in this study, five had at least one wild allele with a significant positive additive effect. The allele of PI 417653 at Pv01.81 had the largest effect (277 kg ha^{-1}), detected under drought in 2015. PI 417653 was the only parent that had effects close to zero, in three of the 13 QTLs. This wild parent is part of the putative Mesoamerican domestication center of common bean in west-central Mexico and is, therefore, the closest genetically to the domesticated parent of the three wild populations used in this study (Kwak et al., 2009). This

observation suggests that the ancestor of this wild population contributed yield alleles to the Mesoamerican domesticated gene pool or that it could harbor an introgression from the domesticated gene pool (Papa and Gepts, 2003; Papa et al., 2005).

Putative Candidate Genes for QTLs

The QTLs detected in this study encompass genomic regions with high number of genes, in general. For example, the statistical significance region of the QTL at Pv09.33, which was detected in the 2014 well-watered treatment and for DSI, ranges from 24.43 to 25.98 Mb in version 2.1 of the common bean reference genome deposited in Phytozome[3]. Within this region, 105 distinct gene models have been identified. Their annotation (obtained from PFAM[4] and PANTHER[5]; see **Supplementary Table S1**) reveals genes implicated in a variety of cellular processes and molecular functions, such as response to stress, oxidative response, signal transduction, protein ubiquitination, chromosome modification, histone modification, metal ion binding, and RNA processing. For example, within this QTL region, Phvul.009G164600.1 was annotated as a serine carboxypeptidase, which was described as involved in oxidative stress in rice (Liu et al., 2008). Phvul.009G169400.2 is related to callose synthase, involved in callose deposition, a functional category that was also described by Recchia et al. (2018), in an experiment with BAT477 under drought and well-watered conditions, in the presence or absence of arbuscular mycorrhizal fungi. Moreover, this QTL region also contained members of a leucin-rich repeat family. This group of genes belongs a larger protein family of receptor-like kinases, playing important roles in stress resistance (Ye et al., 2017), a category also found by Recchia et al. (2018) in common bean. However, because the genomic significance region of a low-heritability trait like yield is so large, an exhaustive list of gene models for each QTL identified in this study becomes quite long as illustrated in **Supplementary Table S1**. It is difficult to focus on likely candidate gene models without further experimentation, which falls beyond the scope of this work.

An alternative approach is to map genes with a putative role in drought-tolerance and examine to what extent they co-segregate with QTLs. Genome-wide categorizations have been published for a few families in common bean, such as aquaporins (AQP) (Ariani and Gepts, 2015) and Dehydration Responsive Element-Binding (DREB) genes (Cortés et al., 2012; Konzen et al., 2019). AQPs play important roles as water channel proteins in plants. The AQP gene PvPIP1;1 is located near a QTL detected under drought in 2015 on chromosome Pv01, with a distance of approximately 120 Kbp from marker ss715645251. Two transcription factors belonging to the DREB gene family, traditionally characterized as involved in abiotic stress responses such as drought, were located within a specific QTL. Phvul.001G136100 is located within the QTL at Pv01.45, a QTL detected in the 2015 well-watered treatment and Phvul.003G241700 is located at Pv03.72, a QTL detected in the well-watered treatment in 2015. Both genes were previously categorized as DREB2 genes, which are normally involved in responses to abiotic stresses such as caused by water deficit (Konzen et al., 2019).

[3]https://phytozome.jgi.doe.gov/pz/portal.html

[4]https://pfam.xfam.org/

[5]http://pantherdb.org/

CONCLUSION

Our original hypothesis was that wild *P. vulgaris* populations from drier areas would be better sources of yield-enhancing genes in a domesticated line than those from more humid areas. We showed here that this is indeed the case. However, we also showed that these same wild populations from drier areas also increased yields under well-watered conditions. On average, the superior progeny lines from arid wild beans increased yield by around 20%. Taken together, various wild genomic regions were identified that had positive effects on yield under well-watered and drought-stress conditions. Our results have the potential to make future introgressions assisted with markers faster and more efficient. The alleles with positive (and negative) effects help explain the transgressive segregation found in this study and underscore the potential of wild variation to improve the productivity of domesticated beans. Future work is needed to validate the QTLs with a positive effect on yield introgressed in different domesticated genetic (Des Marais et al., 2013; Shen et al., 2018). Variation within some QTLs in the magnitude of the effect or sign of the effect suggest allelic series associated with a range of phenotypic variation. This variation could be the basis to local adaptation (Buckler et al., 2009; Kronholm et al., 2012). In addition, further sampling of the wild variation is warranted (Stich, 2009), as well as evaluation in wetter and more humid conditions than in the field site in California.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in Dryad https://doi.org/10.25338/B8FW3M.

AUTHOR CONTRIBUTIONS

JB and PG designed the experiments. JB, AP, and EK carried out the field trial and genotyping. JB analyzed the data and drafted the manuscript. JB, EK, AP, ST, and PG contributed to and edited the manuscript.

FUNDING

Funding was provided by the Agriculture and Food Research Initiative (AFRI) Competitive Grant (2013-67013-21224) from the USDA National Institute of Food and Agriculture (NIFA).

ACKNOWLEDGMENTS

We would like to thank the crew at the Plant Sciences Field Facility for assistance in the field experiments. Thanks

to Laura Gaminño, Jamily Ramos de Lima, Arthur Martins Almeida Bernardeli, Poliana Silva Rezende, Bruno Lima Martins, Guilherme Coelho Portilho, Higor Da Costa Ximenes de Souza, Mayara Rocha, Adam Yang, Ninh Khuu and the rest of the Gepts Lab for the help during planting, harvest, and sample processing.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.00346/full#supplementary-material

REFERENCES

Acosta Gallegos, J. A., Jiménez Hernández, Y., Sánchez García, B. M., Mendoza Hernández, F. M., Herrera Hernández, M. G., Salinas Pérez, R. A., et al. (2010). Flor de Mayo Eugenia, nueva variedad de frijol para riego y temporal en el centro de México. *Rev. Mex. Ciencias Agrícolas* 1, 751–757.

Acosta-Gallegos, J. A., Kelly, J. D., and Gepts, P. (2007). Prebreeding in common bean and use of genetic diversity from wild germplasm. *Crop Sci.* 47, S–44. doi: 10.2135/cropsci2007.04.0008IPBS

Anderson, J. T., Willis, J. H., and Mitchell-Olds, T. (2011). Evolutionary genetics of plant adaptation. *Trends Genet.* 27, 258–266. doi: 10.1016/j.tig.2011.04.001

Araújo, A. P., Teixeira, M. G., and de Almeida, D. L. (1998). Variability of traits associated with phosphorus efficiency in wild and cultivated genotypes of common bean. *Plant Soil* 203, 173–182. doi: 10.1023/A:1004398930218

Ariani, A., and Gepts, P. (2015). Genome-wide identification and characterization of aquaporin gene family in common bean (*Phaseolus vulgaris* L.). *Mol. Genet. Genomics* 290, 1771–1785. doi: 10.1007/s00438-015-1038-2

Ariani, A., Berny Mier Teran, J. C., and Gepts, P. (2018). Spatial and temporal scales of range expansion in wild Phaseolus vulgaris. *Mol. Biol. Evol.* 35, 119–131. doi: 10.1093/molbev/msx273

Bajgain, P., Rouse, M. N., Tsilo, T. J., Macharia, G. K., Bhavani, S., Jin, Y., et al. (2016). Nested association mapping of stem rust resistance in wheat using genotyping by sequencing. *PLoS ONE* 11:e0155760. doi: 10.1371/journal.pone.0155760

Bayuelo-Jiménez, J. S., Debouck, D. G., and Lynch, J. P. (2002). Salinity tolerance in Phaseolus species during early vegetative growth. *Crop Sci.* 42, 2184–2192. doi: 10.2135/cropsci2002.2184

Bazakos, C., Hanemian, M., Trontin, C., Jiménez-Gómez, J. M., and Loudet, O. (2017). New strategies and tools in quantitative genetics: how to go from the phenotype to the genotype. *Annu. Rev. Plant Biol.* 68, 435–455. doi: 10.1146/annurev-arplant-042916-040820

Beaver, J. S., Zapata, M., Alameda, M., Porch, T. G., and Rosas, J. C. (2012). Registration of PR0401-259 and PR0650-31 dry bean germplasm lines. *J. Plant Regist.* 6, 81. doi: 10.3198/jpr2011.05.0283crg

Beebe, S. (2012). Common bean breeding in the tropics. *Plant Breed. Rev. Vol.* 36, 357–426. doi: 10.1002/9781118358566

Beebe, S. E., Rao, I. M., Blair, M. W., and Acosta-Gallegos, J. A. (2013). Phenotyping common beans for adaptation to drought. *Front. Physiol.* 4:35. doi: 10.3389/fphys.2013.00035

Beebe, S. E., Rao, I. M., Devi, M. J., and Polania, J. (2014). Common beans, biodiversity, and multiple stresses: challenges of drought resistance in tropical soils. *Crop Pasture Sci.* 65:667. doi: 10.1071/CP13303

Berny Mier y Teran, J. C., Konzen, E. R., Medina, V., Palkovic, A., Ariani, A., Tsai, S. M., et al. (2019a). Root and shoot variation in relation to potential intermittent drought adaptation of Mesoamerican wild common bean (Phaseolus vulgaris L.). *Ann. Bot.* 124, 917–932. doi: 10.1093/aob/mcy221

Berny Mier y Teran, J. C., Konzen, E. R., Palkovic, A., Tsai, S. M., Rao, I. M., Beebe, S., et al. (2019b). Effect of drought stress on the genetic architecture of photosynthate allocation and remobilization in pods of common bean (*Phaseolus vulgaris* L.), a key species for food security. *BMC Plant Biol.* 19:171. doi: 10.1186/s12870-019-1774-2

Bitocchi, E., Nanni, L., Bellucci, E., Rossi, M., Giardini, A., Zeuli, P. S., et al. (2012). Mesoamerican origin of the common bean (*Phaseolus vulgaris* L.) is revealed by sequence data. *Proc. Natl. Acad. Sci. U.S.A.* 109, E788–E796. doi: 10.1073/pnas.1108973109

Blair, M. W., Iriarte, G., and Beebe, S. (2006). QTL analysis of yield traits in an advanced backcross population derived from a cultivated Andean x wild common bean (*Phaseolus vulgaris* L.) cross. *Theor. Appl. Genet.* 112, 1149–1163. doi: 10.1007/s00122-006-0217-2

Blair, M. W., and Izquierdo, P. (2012). Use of the advanced backcross-QTL method to transfer seed mineral accumulation nutrition traits from wild to Andean cultivated common beans. *Theor. Appl. Genet.* 125, 1015–1031. doi: 10.1007/s00122-012-1891-x

Blair, M. W., Izquierdo, P., Astudillo, C., and Grusak, M. A. (2013). A legume biofortification quandary: variability and genetic control of seed coat micronutrient accumulation in common beans. *Front. Plant Sci.* 4:275. doi: 10.3389/fpls.2013.00275

Bodénès, C., Chancerel, E., Ehrenmann, F., Kremer, A., and Plomion, C. (2016). High-density linkage mapping and distribution of segregation distortion regions in the oak genome. *DNA Res.* 23, 115–124. doi: 10.1093/dnares/dsw001

Brachi, B., Morris, G. P., and Borevitz, J. O. (2011). Genome-wide association studies in plants: the missing heritability is in the field. *Genome Biol.* 12:232. doi: 10.1186/gb-2011-12-10-232

Briñez, B., Perseguini, J. M. K. C., Rosa, J. S., Bassi, D., Gonçalves, J. G. R., Almeida, C., et al. (2017). Mapping QTLs for drought tolerance in a SEA 5 x AND 277 common bean cross with SSRs and SNP markers. *Genet. Mol. Biol.* 40, 813–823. doi: 10.1590/1678-4685-gmb-2016-0222

Broughton, W. J., Hernández, G., Blair, M., Beebe, S., Gepts, P., and Vanderleyden, J. (2003). Beans (Phaseolus spp.) – model food legumes. *Plant Soil* 252, 55–128. doi: 10.1023/A:1024146710611

Buckler, E. S., Holland, J. B., Bradbury, P. J., Acharya, C. B., Brown, P. J., Browne, C., et al. (2009). The genetic architecture of maize flowering time. *Science (80-.)* 325, 714–718. doi: 10.1126/science.1174276

Cavanagh, C., Morell, M., Mackay, I., and Powell, W. (2008). From mutations to MAGIC: resources for gene discovery, validation and delivery in crop plants. *Curr. Opin. Plant Biol.* 11, 215–221. doi: 10.1016/j.pbi.2008.01.002

Checa, O. E., and Blair, M. W. (2008). Mapping QTL for climbing ability and component traits in common bean (*Phaseolus vulgaris* L.). *Mol. Breed.* 22, 201–215. doi: 10.1007/s11032-008-9167-5

Cortés, A. J., This, D., Chavarro, C., Madriñán, S., and Blair, M. W. (2012). Nucleotide diversity patterns at the drought-related DREB2 encoding genes in wild and cultivated common bean (*Phaseolus vulgaris* L.). *Theor. Appl. Genet.* 125, 1069–1085. doi: 10.1007/s00122-012-1896-5

Delgado-Salinas, A., Bibler, R., and Lavin, M. (2006). Phylogeny of the genus *Phaseolus* (Leguminosae): a recent diversification in an ancient landscape. *Syst. Bot.* 31, 779–791. doi: 10.1600/036364406779695960

Des Marais, D. L., Hernandez, K. M., and Juenger, T. E. (2013). Genotype-by-environment interaction and plasticity: exploring genomic responses of plants to the abiotic environment. *Annu. Rev. Ecol. Evol. Syst.* 44, 5–29. doi: 10.1146/annurev-ecolsys-110512-135806

Devi, M. J., Sinclair, T. R., Beebe, S. E., and Rao, I. M. (2012). Comparison of common bean (*Phaseolus vulgaris* L.) genotypes for nitrogen fixation tolerance to soil drying. *Plant Soil* 364, 29–37. doi: 10.1007/s11104-012-1330-4

Dohle, S., Berny Mier y Teran, J. C., Egan, A., Kisha, T., and Khoury, C. K. (2019). "Wild beans (*Phaseolus* L.) of North America," in *North American Crop Wild Relatives: Important Species*, Vol. 2, eds S. L. Greene, K. A. Williams, C. K. Khoury, M. B. Kantar, and L. F. Marek (Cham: Springer International Publishing), 99–127. doi: 10.1007/978-3-319-97121-6_4

Endelman, J. B., and Plomion, C. (2014). LPmerge: an R package for merging genetic maps by linear programming. *Bioinformatics* 30, 1623–1624. doi: 10.1093/bioinformatics/btu091

Espinosa-Alonso, L. G., Lygin, A., Widholm, J. M., Valverde, M. E., and Paredes-Lopez, O. (2006). Polyphenols in wild and weedy Mexican common beans (*Phaseolus vulgaris* L.). *J. Agric. Food Chem.* 54, 4436–4444. doi: 10.1021/jf060185e

Fischer, R., and Maurer, R. (1978). Drought resistance in spring wheat cultivars. I. Grain yield responses. *Aust. J. Agric. Res.* 29:897. doi: 10.1071/AR9780897

Fouilloux, G., and Bannerot, H. (1988). "Selection methods in the common bean (*Phaseolus vulgaris*)," in *Genetic Resources of Phaseolus Beans: Their Maintenance, Domestication, Evolution and Utilization*, ed. P. L. Gepts (Berlin: Springer), 503–542. doi: 10.1007/978-94-009-2786-5_21

Fragoso, C. A., Moreno, M., Wang, Z., Heffelfinger, C., Arbelaez, L. J., Aguirre, J. A., et al. (2017). Genetic architecture of a rice nested association mapping population. *G3* 7, 1913–1926. doi: 10.1534/g3.117.041608

Frahm, M., Rosas, J., Mayek-Pérez, N., López-Salinas, E., Acosta-Gallegos, J., and Kelly, J. (2004). Breeding beans for resistance to terminal drought in the Lowland tropics. *Euphytica* 136, 223–232. doi: 10.1023/B:euph.0000030671.03694.bb

Freytag, G. F., and Debouck, D. G. (2002). *Taxonomy, Distribution, and Ecology of the Genus Phaseolus (Leguminosae-Papilionodeae) in North America, Mexico and Central America. Taxonomía, distribución y ecología del género Phaseolus (Leguminosae-Papilionodeae) en Norteamérica, México y Centroaméric*. Dalles, TX: SIDA, Bot. Misc.

Garin, V., Wimmer, V., Mezmouk, S., Malosetti, M., and van Eeuwijk, F. (2017). How do the type of QTL effect and the form of the residual term influence QTL detection in multi-parent populations? A case study in the maize EU-NAM population. *Theor. Appl. Genet.* 130, 1753–1764. doi: 10.1007/s00122-017-2923-3

Gentry, H. S. (1969). Origin of the common bean, *Phaseolus vulgaris*. *Econ. Bot.* 23, 55–69. doi: 10.1007/BF02862972

Gepts, P. (2019). Origin and evolution of common bean: past events and recent trends. *HortScience* 33, 1124–1130. doi: 10.21273/hortsci.33.7.1124

Gepts, P., Aragão, F. J. L., Barros, E., de Blair, M. W., Brondani, R., Broughton, W., et al. (2008). "Genomics of phaseolus beans, a major source of dietary protein and micronutrients in the tropics," in *Genomics of Tropical Crop Plants*, eds P.H. Moore and R. Ming (New York, NY: Springer New York), 113–143. doi: 10.1007/978-0-387-71219-2_5

Gepts, P., Osborn, T. C., Rashka, K., and Bliss, F. A. (1986). Phaseolin-protein variability in wild forms and landraces of the common bean (*Phaseolus vulgaris*): evidence for multiple centers of domestication. *Econ Bot.* 40, 451–458. doi: 10.1007/BF02859659

Griffiths, P. D. (2002). *Evaluation of Phaseolus germplasm to Identify Resistance to Cucumber Mosaic Virus. New York State IPM Progr*. Available at: https://hdl.handle.net/1813/45914.

Gu, W., Zhu, J., Wallace, D. H., Singh, S. P., and Weeden, N. F. (1998). Analysis of genes controlling photoperiod sensitivity in common bean using DNA markers. *Euphytica* 102, 125–132. doi: 10.1023/A:1018340514388

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., and Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* 25, 1965–1978. doi: 10.1002/joc.1276

Holland, J. B. (2015). MAGIC maize: a new resource for plant genetics. *Genome Biol.* 16:163. doi: 10.1186/s13059-015-0713-2

Hoyos-Villegas, V., Song, Q., Wright, E. M., Beebe, S. E., and Kelly, J. D. (2016). Joint linkage QTL mapping for yield and agronomic traits in a composite map of three common bean RIL populations. *Crop Sci.* 56, 2546–2563. doi: 10.2135/cropsci2016.01.0063

Johnson, W. C., and Gepts, P. (2002). The role of epistasis in controlling seed yield and other agronomic traits in an Andean× Mesoamerican cross of common bean (*Phaseolus vulgaris* L.). *Euphytica* 125, 69–79.

Kaeppler, S. M. (1997). Power analysis for quantitative trait locus mapping in populations derived by multiple backcrosses. *Theor. Appl. Genet.* 95, 618–621. doi: 10.1007/s001220050603

Kelly, J. D. (2001). "Remaking bean plant architecture for efficient production," in *Advances in Agronomy*, ed. D. L. Sparks (Amsterdam: Elsevier), 109–143. doi: 10.1016/S0065-2113(01)71013-9

Kelly, J. D., Hosfield, G. L., Varner, G. V., Uebersax, M. A., and Taylor, J. (1999). Registration of 'Matterhorn' great northern bean. *Crop Sci.* 39:589. doi: 10.2135/cropsci1999.0011183X003900020058x

Kissoudis, C., Sunarti, S., van de Wiel, C., Visser, R. G. F., van der Linden, C. G., and Bai, Y. (2016). Responses to combined abiotic and biotic stress in tomato are governed by stress intensity and resistance mechanism. *J. Exp. Bot.* 67, 5119–5132. doi: 10.1093/jxb/erw285

Klaedtke, S. M., Cajiao, C., Grajales, M., Polania, J., Borrero, G., Guerrero, A., et al. (2012). Photosynthate remobilization capacity from drought-adapted common bean (*Phaseolus vulgaris* L.) lines can improve yield potential of interspecific populations within the secondary gene pool. *J. Plant Breed. Crop Sci.* 4, 49–61. doi: 10.5897/JPBCS11.087

Koinange, E. M. K., Singh, S. P., and Gepts, P. (1996). Genetic control of the domestication syndrome in common bean. *Crop Sci.* 36, 1037–1045. doi: 10.2135/cropsci1996.0011183X003600040037x

Konzen, E. R., Recchia, G. H., Cassieri, F., Caldas, D. G. G., Berny Mier y Teran, J. C., Gepts, P., et al. (2019). DREB genes from common bean (Phaseolus vulgaris L.) show broad to specific abiotic stress responses and distinct levels of nucleotide diversity. *Int. J. Genomics* 2019, 28.

Kornegay, J., Cardona, C., and Posso, C. E. (1993). Inheritance of resistance to Mexican bean weevil in common bean, determined by bioassay and biochemical tests. *Crop Sci.* 33:589. doi: 10.2135/cropsci1993.0011183x003300030034x

Kosambi, D. (1944). The estimation of map distances from recombination values. *Ann Eugen* 12, 125–130. doi: 10.1111/j.1469-1809.1943.tb02321.x

Kronholm, I., Picó, F. X., Alonso-Blanco, C., Goudet, J., and de Meaux, J. (2012). Genetic basis of adaptation in Arabidopsis thaliana: local adaptation at the seed dormancy QTL DOG1. *Evolution (N.Y.)* 66, 2287–2302. doi: 10.1111/j.1558-5646.2012.01590.x

Kwak, M., and Gepts, P. (2009). Structure of genetic diversity in the two major gene pools of common bean (*Phaseolus vulgaris* L., Fabaceae). *Theor. Appl. Genet.* 118, 979–992. doi: 10.1007/s00122-008-0955-4

Kwak, M., Kami, J. A., and Gepts, P. (2009). The putative mesoamerican domestication center of *Phaseolus vulgaris* is located in the lerma-santiago basin of Mexico. *Crop Sci.* 49, 554–563. doi: 10.2135/cropsci2008.07.0421

Kwak, M., Velasco, D., and Gepts, P. (2008). Mapping homologous sequences for determinacy and photoperiod sensitivity in common bean (*Phaseolus vulgaris*). *J. Hered.* 99, 283–291. doi: 10.1093/jhered/esn005

Li, H., Bradbury, P., Ersoz, E., Buckler, E. S., and Wang, J. (2011). Joint QTL linkage mapping for multiple-cross mating design sharing one common parent. *PLoS ONE* 6:e17573. doi: 10.1371/journal.pone.0017573

Li, H., Ye, G., and Wang, J. (2007). A modified algorithm for the improvement of composite interval mapping. *Genetics* 175, 361–374. doi: 10.1534/genetics.106.066811

Lin, C. S., and Poushinsky, G. (1985). A modified augmented design (Type 2) for rectangular plots. *Can. J. Plant Sci.* 65, 743–749. doi: 10.4141/cjps85-094

Liu, H., Wang, X., Zhang, H., Yang, Y., Ge, X., and Song, F. (2008). A rice serine carboxypeptidase-like gene OsBISCPL1 is involved in regulation of defense responses against biotic and oxidative stress. *Gene* 420, 57–65. doi: 10.1016/j.gene.2008.05.006

Liu, Z., Cook, J., Melia-Hancock, S., Guill, K., Bottoms, C., Garcia, A., et al. (2016). Expanding maize genetic resources with predomestication alleles: maize–teosinte introgression populations. *Plant Genome* 9. doi: 10.3835/plantgenome2015.07.0053

Long, R., Temple, S., Schmiere, J., Canevari, M., and Meyers, R. D. (2010). *Common dry Bean Production in California, Davis*. Available at: https://anrcatalog.ucanr.edu/pdf/8402.pdf

Lyttle, T. (1991). Segregation distorters. *Annu. Rev. Genet.* 25, 511–557. doi: 10.1146/annurev.genet.25.1.511

Malosetti, M., van Eeuwijk, F. A., Jordan, D. R., Velazco, J. G., Eilers, P. H. C., Boer, M. P., et al. (2017). Modelling spatial trends in sorghum breeding field trials using a two-dimensional P-spline mixed model. *Theor. Appl. Genet.* 130, 1375–1392. doi: 10.1007/s00122-017-2894-4

Maurer, A., Draba, V., Jiang, Y., Schnaithmann, F., Sharma, R., Schumann, E., et al. (2015). Modelling the genetic architecture of flowering time control in barley through nested association mapping. *BMC Genomics* 16:290. doi: 10.1186/s12864-015-1459-7

McClean, P. E., Lee, R. K., Otto, C., Gepts, P., and Bassett, M. J. (2002). Molecular and phenotypic mapping of genes controlling seed coat pattern and color in common bean (Phaseolus vulgaris L.). *J. Hered.* 93, 148–152. doi: 10.1093/jhered/93.2.148

Meng, L., Li, H., Zhang, L., and Wang, J. (2015). QTL IciMapping: integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations. *Crop J.* 3, 269–283. doi: 10.1016/j.cj.2015.01.001

Mkwaila, W., Terpstra, K. A., Ender, M., and Kelly, J. D. (2011). Identification of QTL for agronomic traits and resistance to white mold in wild and landrace germplasm of common bean. *Plant Breed.* 130, 665–672. doi: 10.1111/j.1439-0523.2011.01876.x

Moghaddam, S. M., Mamidi, S., Osorno, J. M., Lee, R., Brick, M., Kelly, J., et al. (2016). Genome-wide association study identifies candidate loci underlying agronomic traits in a Middle American diversity panel of common bean. *Plant Genome* 9. doi: 10.3835/plantgenome2016.02.0012

Monteith, J. L., and Moss, C. J. (2006). Climate and the efficiency of crop production in Britain. *Philos. Trans. R. Soc. B Biol. Sci.* 281, 277–294. doi: 10.1098/rstb.1977.0140

Montoya, C. A., Leterme, P., Victoria, N. F., Toro, O., Souffrant, W. B., Beebe, S., et al. (2008). Susceptibility of phaseolin to in vitro proteolysis is highly variable across common bean varieties (*Phaseolus vulgaris*). *J. Agric. Food Chem.* 56, 2183–2191. doi: 10.1021/jf072576e

Mukeshimana, G., Butare, L., Cregan, P. B., Blair, M. W., and Kelly, J. D. (2014). Quantitative trait loci associated with drought tolerance in common bean. *Crop Sci.* 54, 923–938. doi: 10.2135/cropsci2013.06.0427

Nice, L. M., Steffenson, B. J., Blake, T. K., Horsley, R. D., Smith, K. P., and Muehlbauer, G. J. (2017). Mapping agronomic traits in a wild barley advanced backcross–nested association mapping population. *Crop Sci.* 57, 1199–1210. doi: 10.2135/cropsci2016.10.0850

Nice, L. M., Steffenson, B. J., Brown-Guedira, G. L., Akhunov, E. D., Liu, C., Kono, T. J. Y., et al. (2016). Development and genetic characterization of an advanced backcross-nested association mapping (AB-NAM) population of wild × cultivated barley. *Genetics* 203, 1453–1467. doi: 10.1534/genetics.116.190736

Osborn, T. C., Alexander, D. C., Sun, S. S. M., Cardona, C., and Bliss, F. A. (2006). Insecticidal activity and lectin homology of arcelin seed protein. *Science* 240, 207–210. doi: 10.1126/science.240.4849.207

Osborn, T. C., Hartweck, L. M., Harmsen, R. H., Vogelzang, R. D., Kmiecik, K. A., and Bliss, F. A. (2003). Registration of genetic stocks with altered seed protein compositions. *Crop Sci.* 43, 1570. doi: 10.2135/cropsci2003.1570

Papa, R., Acosta-Gallegos, J. A., Delgado-Salinas, A., and Gepts, P. (2005). A genome-wide analysis of differentiation between wild and domesticated *Phaseolus vulgaris* from Mesoamerica. *Theor. Appl. Genet.* 111, 1147–1158. doi: 10.1007/s00122-005-0045-9

Papa, R., and Gepts, P. (2003). Asymmetry of gene flow and differential geographical structure of molecular diversity in wild and domesticated common bean (*Phaseolus vulgaris* L.) from Mesoamerica. *Theor. Appl. Genet.* 106, 239–250. doi: 10.1007/s00122-002-1085-z

Park, S. O., Coyne, D. P., Jung, G., Skroch, P. W., Arnaud-Santana, E., Steadman, J. R., et al. (2000). Mapping of QTL for seed size and shape traits in common bean. *J. Am. Soc. Hort. Sci.* 125, 466–475. doi: 10.21273/JASHS.125.4.466

Pedrosa-Harand, A., Porch, T., and Gepts, P. (2008). Standard nomenclature for common bean chromosomes and linkage groups. *Annu. Rep. Bean Improv. Coop.* 51:106.

Pinto, R. S., Reynolds, M. P., Mathews, K. L., McIntyre, C. L., Olivares-Villegas, J.-J., and Chapman, S. C. (2010). Heat and drought adaptive QTL in a wheat population designed to minimize confounding agronomic effects. *Theor. Appl. Genet.* 121, 1001–1021. doi: 10.1007/s00122-010-1351-4

Polania, J., Rao, I., Cajiao, C., Grajales, M., Rivera, M., Velasquez, C., et al. (2017). Shoot and root traits contribute to drought resistance in recombinant inbred lines of MD 23-24 x SEA 5 of common bean. *Front. Plant Sci.* 8:296. doi: 10.3389/fpls.2017.00296

Polania, J., Rao, I. M., Cajiao, C., Rivera, M., Raatz, B., and Beebe, S. (2016). Physiological traits associated with drought resistance in Andean and Mesoamerican genotypes of common bean (*Phaseolus vulgaris* L.). *Euphytica* 210, 17–29. doi: 10.1007/s10681-016-1691-5

Porch, T., Beaver, J., Debouck, D., Jackson, S., Kelly, J., and Dempewolf, H. (2013). Use of wild relatives and closely related species to adapt common bean to climate change. *Agronomy* 3, 433–461. doi: 10.3390/agronomy3020433

Prudhomme, C., Giuntoli, I., Robinson, E. L., Clark, D. B., Arnell, N. W., Dankers, R., et al. (2014). Hydrological droughts in the 21st century, hotspots and uncertainties from a global multimodel ensemble experiment. *Proc. Natl. Acad. Sci. U.S.A.* 111, 3262–3267. doi: 10.1073/pnas.1222473110

R Development Core Team (2018). *R: A Language and Environment for Statistical Computing*. Vienna: R Development Core Team. doi: 10.1007/978-3-540-74686-7

Ramirez-Cabral, N. Y. Z., Kumar, L., and Taylor, S. (2016). Crop niche modeling projects major shifts in common bean growing areas. *Agric. For. Meteorol.* 21, 102–113. doi: 10.1016/j.agrformet.2015.12.002

Rao, I. M., Beebe, S. E., Polania, J., Grajales, M., Cajiao, C., Ricaurte, J., et al. (2017). Evidence for genotypic differences among elite lines of common bean in the ability to remobilize photosynthate to increase yield under drought. *J. Agric. Sci.* 155, 857–875. doi: 10.1017/S0021859616000915

Recchia, G. H., Konzen, E. R., Cassieri, F., Caldas, D. G., and Tsai, S. M. (2018). Arbuscular mycorrhizal symbiosis leads to differential regulation of drought-responsive genes in tissue-specific root cells of common bean. *Front. Microbiol.* 9:1339. doi: 10.3389/fmicb.2018.01339

Rendón-Anaya, M., Montero-Vargas, J. M., Saburido-Álvarez, S., Vlasova, A., Capella-Gutierrez, S., Ordaz-Ortiz, J. J., et al. (2017). Genomic history of the origin and domestication of common bean unveils its closest sister species. *Genome Biol.* 18:60. doi: 10.1186/s13059-017-1190-6

Rodriguez, M., Rau, D., Bitocchi, E., Bellucci, E., Biagetti, E., Carboni, A., et al. (2016). Landscape genetics, adaptive diversity and population structure in *Phaseolus vulgaris*. *New Phytol.* 209, 1781–1794. doi: 10.1111/nph.13713

Rodríguez-Álvarez, M. X., Boer, M. P., van Eeuwijk, F. A., and Eilers, P. H. C. (2018). Correcting for spatial heterogeneity in plant breeding experiments with P-splines. *Spat. Stat.* 23, 52–71. doi: 10.1016/j.spasta.2017.10.003

Rosas, J. C., Varela, O. I., and Beaver, J. S. (2010). Registration of 'Tío Canela-75' small red bean (race Mesoamerica). *Crop Sci.* 37:1391. doi: 10.2135/cropsci1997.0011183x003700040080x

Saade, S., Maurer, A., Shahid, M., Oakey, H., Schmöckel, S. M., Negrão, S., et al. (2016). Yield-related salinity tolerance traits identified in a nested association mapping (NAM) population of wild barley. *Sci. Rep.* 6:32586.

Sas Institute Inc (2016). *JMP®*. Available at: https://www.jmp.com/en_us/home.html.

Schmutz, J., McClean, P. E., Mamidi, S., Wu, G. A., Cannon, S. B., Grimwood, J., et al. (2014). A reference genome for common bean and genome-wide analysis of dual domestications. *Nat. Genet.* 46, 707–713. doi: 10.1038/ng.3008

Schnaithmann, F., Kopahnke, D., and Pillen, K. (2014). A first step toward the development of a barley NAM population and its utilization to detect QTLs conferring leaf rust seedling resistance. *Theor. Appl. Genet.* 127, 1513–1525. doi: 10.1007/s00122-014-2315-x

Shen, L., Wang, C., Fu, Y., Wang, J., Liu, Q., Zhang, X., et al. (2018). QTL editing confers opposing yield performance in different rice varieties. *J. Integr. Plant Biol.* 60, 89–93. doi: 10.1111/jipb.12501

Sinclair, T. R., and Vadez, V. (2012). The future of grain legumes in cropping systems. *Crop Pasture Sci.* 63, 501–512.

Singh, J., Gezan, S., and Vallejos, C. E. (2019). Developmental pleiotropy shaped the roots of the domesticated common bean *Phaseolus vulgaris* L. *Plant Physiol.* 180, 1467–1479. doi: 10.1104/pp.18.01509

Singh, S. P. (2001). Broadening the genetic base of common bean cultivars: a review. *Crop Sci.* 41, 1659–1675. doi: 10.2135/cropsci2001.1659

Singh, S. P., Gepts, P., and Debouck, D. G. (1991). Races of common bean (*Phaseolus vulgaris*, Fabaceae). *Econ. Bot.* 45, 379–396. doi: 10.1007/BF02887079

Singh, S. P., Terán, H., and Gutiérrez, J. A. (2001). Registration of SEA 5 and SEA 13 drought tolerant dry bean germplasm. *Crop Sci.* 41, 276–277. doi: 10.2135/cropsci2001.411276x

Song, Q., Jia, G., Hyten, D. L., Jenkins, J., Hwang, E.-Y., Schroeder, S. G., et al. (2015). SNP assay development for linkage map construction, anchoring whole-genome sequence, and other genetic and genomic applications in common bean. *G3* 5, 2285–2290. doi: 10.1534/g3.115.020594

Song, Q., Yan, L., Quigley, C., Jordan, B. D., Fickus, E., Schroeder, S., et al. (2017). Genetic characterization of the soybean nested association mapping population. *Plant Genome* 10. doi: 10.3835/plantgenome2016.10.0109

Sonnante, G., Stockton, T., Nodari, R. O., Becerra Velásquez, V. L., and Gepts, P. (1994). Evolution of genetic diversity during the domestication of common-bean (*Phaseolus vulgaris* L.). *Theor. Appl. Genet.* 89, 629–635. doi: 10.1007/BF00222458

Stich, B. (2009). Comparison of mating designs for establishing nested association mapping populations in maize and *Arabidopsis thaliana*. *Genetics* 183, 1525–1534. doi: 10.1534/genetics.109.108449

Tanksley, S. D., and McCouch, S. R. (1997). Seed banks and molecular maps: unlocking genetic potential from the wild. *Science* 277, 1063–1066. doi: 10.1126/science.277.5329.1063

Tanksley, S. D., and Nelson, J. C. (1996). Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from

unadapted germplasm into elite breeding lines. *Theor. Appl. Genet.* 92, 191–203. doi: 10.1007/BF00223376

Taylor, J., and Butler, D. (2017). R Package ASMap: efficient genetic linkage map construction and diagnosis. *J. Stat. Softw.* 79. doi: 10.18637/jss.v079.i06

Terán, H., and Singh, S. P. (2002). Selection for drought resistance in early generations of common bean populations. *Can. J. Plant Sci.* 82, 491–497. doi: 10.4141/P01-134

Trapp, J. J., Urrea, C. A., Zhou, J., Khot, L. R., Sankaran, S., and Miklas, P. N. (2016). Selective phenotyping traits related to multiple stress and drought response in dry bean. *Crop Sci.* 56, 1460–1472. doi: 10.2135/cropsci2015.05.0281

Voorrips, R. E. (2002). MapChart: software for the graphical presentation of linkage maps and QTL. *J. Hered.* 93, 77–78. doi: 10.1093/jhered/93.1.77

Wallach, D., Hwang, C., Correll, M. J., Jones, J. W., Boote, K., Hoogenboom, G., et al. (2018). A dynamic model with QTL covariables for predicting flowering time of common bean (*Phaseolus vulgaris*) genotypes. *Eur. J. Agron.* 101, 200–209. doi: 10.1016/j.eja.2018.10.003

Wang, J., Li, H., Zhang, L., Li, C., and Meng, L. (2016). *Users Manual of QTL IciMapping. Beijing, China: Institute of Crop Science, Chinese Academy of Agricultural Sciences; Crop Research Iinformatics Lab and International Maize and Wheat Improvement Center (CIMMYT)*. Available at: http://www.isbreeding.net.

Weller, J. L., Vander Schoor, J. K., Perez-Wright, E. C., Hecht, V., González, A. M., Capel, C., et al. (2019). Parallel origins of photoperiod adaptation following dual domestications of common bean. *J. Exp. Bot.* 70, 1209–1219. doi: 10.1093/jxb/ery455

White, J. W., Kornegay, J., and Cajiao, C. (1996). Inheritance of temperature sensitivity of the photoperiod response in common bean (*Phaseolus vulgaris* L.). *Euphytica* 91, 5–8. doi: 10.1007/BF00035270

Wu, H., Churchill, G. A., Broman, K. W., and Sen, S. (2003). R/qtl: QTL mapping in experimental crosses. *Bioinformatics* 19, 889–890. doi: 10.1093/bioinformatics/btg112

Wu, X., Li, Y., Shi, Y., Song, Y., Zhang, D., Li, C., et al. (2016). Joint-linkage mapping and GWAS reveal extensive genetic loci that regulate male inflorescence size in maize. *Plant Biotechnol. J.* 14, 1551–1562. doi: 10.1111/pbi.12519

Xiao, Y., Tong, H., Yang, X., Xu, S., Pan, Q., Qiao, F., et al. (2016). Genome-wide dissection of the maize ear genetic architecture using multiple populations. *New Phytol.* 210, 1095–1106. doi: 10.1111/nph.13814

Xu, S. (2008). Quantitative trait locus mapping can benefit from segregation distortion. *Genetics* 180, 2201–2208. doi: 10.1534/genetics.108.090688

Ye, Y., Ding, Y., Jiang, Q., Wang, F. J., Sun, J. W., and Zhu, C. (2017). The role of receptor-like protein kinases (RLKs) in abiotic stress response in plants. *Plant Cell Rep.* 36, 1–8. doi: 10.1007/s00299-016-2084-x

Yu, J., Holland, J. B., McMullen, M. D., and Buckler, E. S. (2008). Genetic design and statistical power of nested association mapping in maize. *Genetics* 178, 539–551. doi: 10.1534/genetics.107.074245

Zhu, X.-G., Long, S. P., and Ort, D. R. (2010). Improving photosynthetic efficiency for greater yield. *Annu. Rev. Plant Biol.* 61, 235–261. doi: 10.1146/annurev-arplant-042809-112206

Zizumbo-Villarreal, D., Colunga-GarcíaMarín, P., Payró de la Cruz, E., Delgado-Valerio, P., and Gepts, P. (2005). Population structure and evolutionary dynamics of wild–weedy–domesticated complexes of common bean in a Mesoamerican region. *Crop Sci.* 35, 1073–1083. doi: 10.2135/cropsci2004.0340

Conflict of Interest: 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 © 2020 Berny Mier y Teran, Konzen, Palkovic, Tsai and Gepts. 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) and the copyright owner(s) 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 Plant Science

ORIGINAL RESEARCH
published: 29 April 2020
doi: 10.3389/fpls.2020.00388

An Updated Checklist of the Sicilian Native Edible Plants: Preserving the Traditional Ecological Knowledge of Century-Old Agro-Pastoral Landscapes

Salvatore Pasta[1], Alfonso La Rosa[2], Giuseppe Garfì[1], Corrado Marcenò[3], Alessandro Silvestre Gristina[1], Francesco Carimi[1] and Riccardo Guarino[4]*

[1] Institute of Biosciences and Bioresources (IBBR), National Research Council of Italy (CNR), Palermo, Italy, [2] Cooperativa Silene, Palermo, Italy, [3] Department of Botany and Zoology, Faculty of Science, Masaryk University, Brno, Czechia, [4] Dipartimento STeBiCeF, Sezione Botanica, University of Palermo, Palermo, Italy

The traditional use of native wild food plants (NWFP) may represent a valuable supplementary food source for the present and future generations. In Sicily, the use of wild plants in the human diet dates back to very ancient times and still plays an important role in some rural communities. Moreover, in this regard, the natural and cultural inheritance of this island is wealthy and diversified for several reasons. First, Sicily hosts a rich vascular flora, with 3,000 native and 350 endemic plants. Second, due to its central position in the Mediterranean, the island has acted as a veritable melting pot for the ethnobotanical knowledge of the rural communities of the entire basin. We reviewed all the available literature and, starting from such omnicomprehensive checklist, partially improved thanks to the data issuing from recent field investigations, we critically revised the whole species list, basing our review on field data issuing from interviews and on our expert knowledge. As a result, we provide a substantially updated list of 292 NWFP growing on the island. Further 34 species, reported as NWFP on previous papers were discarded because they are not native to Sicily, while 45 species were listed separately because their identity, occurrence and local use as food is doubtful and needs to be further investigated. Moreover, we tried to shed light on the ecology (growth form and preferential habitat) of the Sicilian NWFP, with special focus on crop wild relatives (CWR). Our preliminary ecological analyses point out that a high percentage of these plants are linked with the so-called 'cultural' landscapes, patchy semi-natural environments rich in ecotones, leading to the conclusion that the maintenance of century-old agro-pastoral practices may represent an effective way to preserve the local heritage of edible plants. Our study allowed to identify as much as 102 taxa of agronomic interest which could be tested as novel crops in order to face ongoing global changes and to comply with sustainable agriculture policies. Among them, 39 taxa show promising traits in terms of tolerance to one or more environmental stress factors, while 55 more are considered CWR and/or can be easily cultivated and/or show high productivity/yield potential.

Keywords: ethnobotany, agro-pastoral landscapes, sustainable agriculture, TEK (traditional environmental knowledge), Ellenberg Indicator Values (EIV)

OPEN ACCESS

Edited by:
Petr Smýkal,
Palacký University Olomouc, Czechia

Reviewed by:
Hanno Schaefer,
Technical University of Munich, Germany
Rosario Schicchi,
University of Palermo, Italy

****Correspondence:***
Francesco Carimi
francesco.carimi@ibbr.cnr.it

Specialty section:
This article was submitted to Plant Breeding, a section of the journal Frontiers in Plant Science

Received: *31 October 2019*
Accepted: *18 March 2020*
Published: *29 April 2020*

Citation:
Pasta S, La Rosa A, Garfì G, Marcenò C, Gristina AS, Carimi F and Guarino R (2020) An Updated Checklist of the Sicilian Native Edible Plants: Preserving the Traditional Ecological Knowledge of Century-Old Agro-Pastoral Landscapes. Front. Plant Sci. 11:388. doi: 10.3389/fpls.2020.00388

INTRODUCTION

Sicily is the largest Mediterranean island (**Figure 1**), with an extension of c. 25,400 km^2. Its territory is predominantly hilly or mountainous, with less than 20% of its surface below 300 m a.s.l. The geographical position of Sicily, its complex geological history (Carbone et al., 2018) and the sharp topographic, edaphic and climatic contrasts make this island one of the most heterogeneous Mediterranean territories. Moreover, Sicily and its satellite islands belong to the Tyrrhenian area, one of the main hot-spots of plant diversity in the whole Mediterranean basin (Médail and Quézel, 1999). Its rich flora counts about 3,000 native plant taxa, around 12% of which are endemics (Guarino and Pasta, 2018).

Thanks to its location, Sicily has also been a major pathway of human migration, acting as a cultural, economic and ethnic crossing point and melting pot. Human presence in Sicily knows no pauses since 14–13 thousand years ago (Mannino et al., 2012), and with no doubt during this long time-lapse entire plant assemblages have been profoundly shaped by hunters-gatherers, and subsequently even wiped out by the early onset of agro-silvopastoral practices (Leighton, 1999; Tinner et al., 2016).

The vegetation of the island shows almost everywhere the traces of long-lasting land exploitation; not surprisingly, anthropogenic plant communities (*Artemisietea vulgaris*, *Chenopodietea*, *Papaveretea rhoeadis*, *Parietarietea judaicae*, *Poetea bulbosae*, and *Polygono-Poetea*, etc.) characterize almost 50% of the island's vegetation (Guarino and Pasta, 2017). Furthermore, hard-wheat crop fields currently occupy a large portion of the island's territory, and other traditional forms of dry-land farming (olive, almond, carob tree, pistachio, ash-tree, hazelnut, and chestnut groves) still characterize part of the Sicilian rural landscape. Along with the disappearance of traditional practices (e.g., seasonal transhumance) and crafts (e.g., charcoal burners, pipe manufacturers, cork and sumac gatherers, miners, millers, etc.), also many man-made landscapes and habitats, such as dry-stone terraces, fruit orchards, dry-land groves and fallows are fading or have already disappeared. Abandonment and land-use changes are responsible for the fast collapse of local natural and cultivated plant richness (La Mantia et al., 2011; Albrecht et al., 2016). Currently, those traditional landscapes and many others suffer from abandonment or, worse, undergo deep transformations due to agricultural intensification and urban sprawl. Intensive cultivations already cover around 25% of the island's surface, and they are still expanding. Two blatant examples are the ongoing replacement of *Citrus* orchards with greenhouses, while intensive vineyards are gradually substituting dry groves. Mechanized agricultural practices and the massive input of chemical fertilizers and pesticides select the weeds, to the detriment of Mediterranean plants and archaeophytes and favoring plenty of non-native tropical and subtropical species, which also take advantage of high nutrient- and water input (Guarino et al., 2015).

In respect to this scenario and considering the challenges related to future global climate changes, Sicilian native wild food plants (from now on NWFP) could play a significant role with respect to the fulfillment of both sustainable agriculture and dietary supplement. In addition to raising awareness on the importance of Mediterranean diet (Willett, 2006; Gamboni et al., 2012), there are many other good reasons to go on with the investigation and valuation of the ethnobotanical knowledge on NWFP throughout the Mediterranean area (Guijt, 1998). In fact, during the last 20 years, an increasing number of scientific papers (e.g., Grivetti and Ogle, 2000; Trichopoulou et al., 2000; Simopoulos, 2004; Pieroni and Price, 2005; Schaffer et al., 2005; Bharucha and Pretty, 2010; Ranfa et al., 2014; Disciglio et al., 2017; Renna, 2017) highlighted the strong correlation between wild food consumption and health.

An increasing amount of field data highlights the fact that the Mediterranean countries were not only important as a target of the spread of the so-called 'Neolithic revolution' but played an active role in the early cultivation of the wild ecotypes of several woody species, such as *Olea europaea* (Diez et al., 2015), *Ceratonia siliqua* (Viruel et al., 2019), and *Vitis vinifera* (De Michele et al., 2019), as well as of many vegetables (e.g., *Allium* spp., *Brassica* spp., *Cynara* spp., etc.) and cereal crops (e.g., *Avena* spp., *Hordeum* spp., and *Triticum* spp.), even far before than expected (Snir et al., 2015). The current change of paradigm about the history and geography of early agriculture underlines the importance of studying, conserving and effectively exploiting the germplasm of Mediterranean CWR.

In this paper, we aimed to critically review the so far available lists of traditional Sicilian NWFP, and to provide a first evaluation of the ecological requirements of these plants, paying special attention to the natural and semi-natural ecosystems where they grow and stressing the importance of managing and maintaining them in view of the ongoing global changes. These ecosystems represent important potential reservoirs of food because they host most of the NWFP and many CWR. Additionally, we evaluated some ecological features of NWFP that could be retained as traits of interest for the detection of novel crops and/or the improvement of the existing ones. These plants may help to face the future challenges of Mediterranean agriculture due to global change, such as extreme heat events, water shortage, salinization and adaptation to nutrient-poor soils.

MATERIALS AND METHODS

Data Set

The list of the Sicilian NWFP provided by Pasta et al. (2011), based on a critical review of the whole regional ethnobotanical literature concerning edible plants up to 2010, was improved and updated by consulting all the most recent and authoritative works on this topic, i.e., Arcidiacono (2016), Schicchi and Geraci (2016), Morreale (2018), and Geraci et al. (2018). Coria (1989), Lentini and Mazzola (1998), Arcidiacono et al. (2007, 2010), as well as all the Sicilian papers published after 2010, i.e., Aleo et al. (2013), Tuttolomondo et al. (2014a,b), Mazzola et al. (2015), Signorello (2015), Licata et al. (2016), Quave and Saitta (2016), Schicchi and Geraci (2016), Cucinotta and Pieroni (2018), Gargano et al. (2018), and Monello (2018) were consulted, too.

The works of Ghirardini et al. (2007) and Guarrera and Savo (2016) were used as basic reference for the whole Italian territory. The assessment of the native status of the Sicilian edible plants

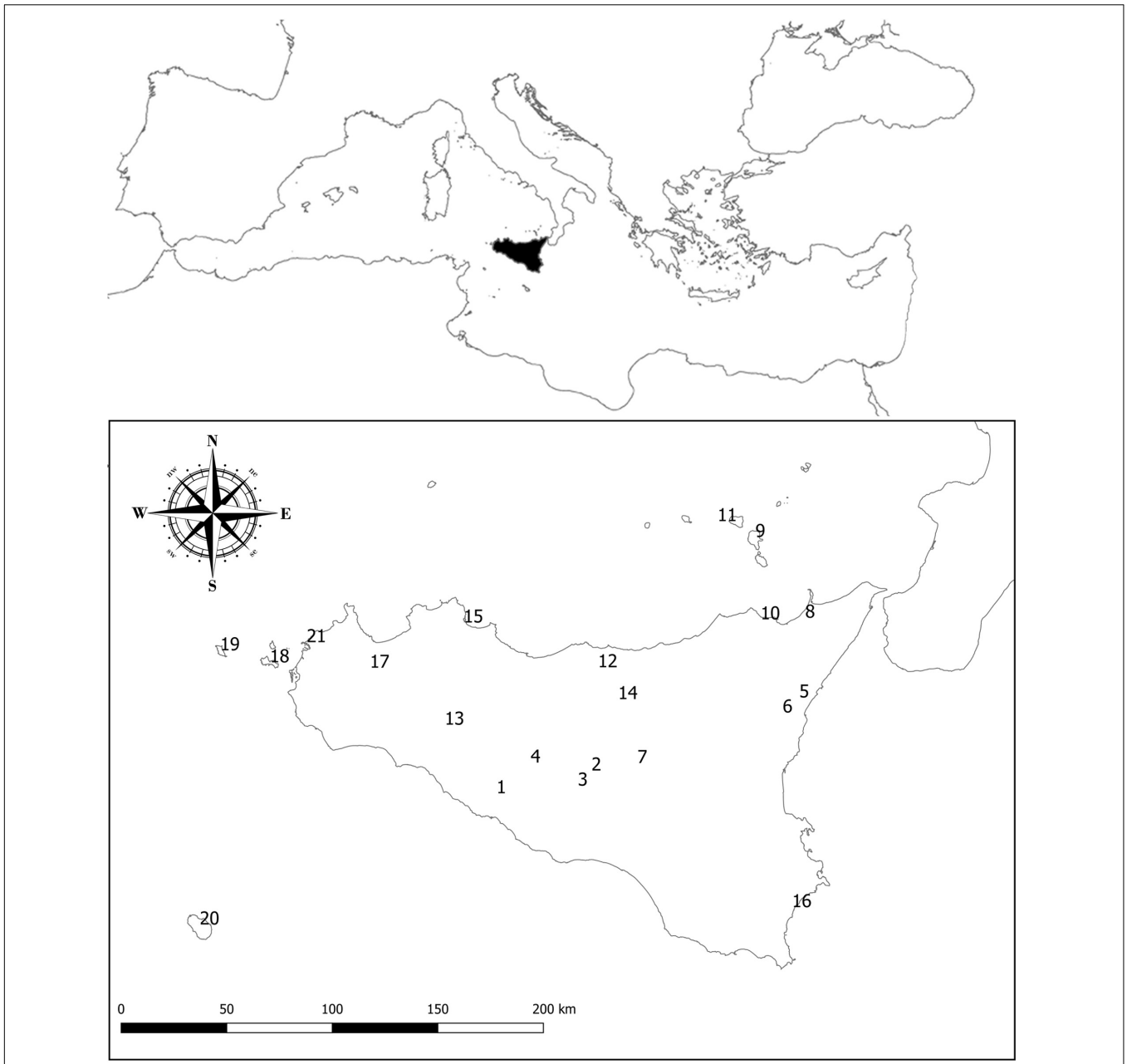


FIGURE 1 | Sicily (black) and its position in the Mediterranean Basin. The numbers refer to the localities where interviews were carried out in order to perform an additional assessment of the harvest frequency of Sicilian NWFP and to update the available knowledge on their use (see **Supplementary Table S1**), i.e., Joppolo Giancaxio (1), Caltanissetta (2), San Cataldo (3), Sutera (4), Piedimonte Etneo (5), Sant'Alfio (6), Pergusa (7), Barcellona Pozzo di Gotto (8), Lipari Island (9), Patti (10), Salina Island (11), Castelbuono (12), Chiusa Sclafani (13), Gangi (14), Palermo (15), Avola (16), Calatafimi (17), Favignana Island (18), Marettimo Island (19), Pantelleria Island (20), and Trapani (21).

was mostly based on Galasso et al. (2018), and non-native food plants were not taken into account for further elaborations. For each considered taxon, we reported the scientific name according to the second edition of the Flora of Italy (Pignatti et al., 2017–2019), the synonym(s) encountered in the consulted literature on NWFP, and the plant family according to Chase et al. (2016), while their growth form was assessed according to Pignatti et al. (2017–2019). We decided to skip the analysis of the life forms and chorotypes of the Sicilian NWFP and we provided just basic information concerning other interesting ethnobotanical issues, such as edible part(s) and traditional use(s) as food, because all these topics have been already treated in the most recent and comprehensive papers concerning Sicilian edible plants (Lentini and Venza, 2007; Pasta et al., 2011; Geraci et al., 2018). Basic information on the potential risks induced by toxic or poisonous compounds contained in the Sicilian NWFP is provided, too.

Additionally, a preliminary semi-quantitative evaluation of the importance of Sicilian NWFP to local harvesters has been performed by comparing the data reported by Geraci et al. (2018) with those issuing from the interviews carried out across the whole Sicilian territory between 2016 and 2019 (**Figure 1**).

Finally, we annotated whether the listed NWFP were considered CWR by Heywood and Zohary (1995) or maybe deemed CWR following the definition: 'A crop wild relative is a wild plant taxon that has an indirect use derived from its relatively close genetic relationship to a crop' (Maxted et al., 2006).

Ecological Assessment

In order to outline the ecological preferences of the Sicilian NWFP, the altitudinal range and the Ellenberg Indicator Values (EIVs) concerning each species, extracted from the database of Pignatti et al. (2017–2019), were adopted.

The preferential habitat types of the Sicilian NWFP were assessed following the list of diagnostic species of phytosociological classes of the European vascular plant communities (Mucina et al., 2016, Appendix S6). The phytosociological classes were associated with the EUNIS habitats recorded from Sicily in Carta della Natura (ISPRA, 2012) according to the crosswalk proposed by Angelini et al. (2009). In order to obtain more clues about the correlation between NWFP and different habitat types, we also used the habitats assigned in the database of Pignatti et al. (2017–2019), which recognizes only 24 habitat types belonging to three main ecological groups, i.e., 'terrestrial,' 'aquatic,' and 'anthropogenic.' In both cases, all the habitats where a given species may occur were assigned according to the fuzzy logic, with a numerical score ranging from 0 to 1. In this way, the fuzzy value was equal to 1 for the stenoecious species, whereas for the dioecious species the fuzzy values resulted to be >0 and <1 depending on the number of habitats assigned to a species.

We used the non-metric multidimensional scaling (NMDS) analysis to summarize the similarity in habitat composition among NWFP, using Bray–Curtis distance measure to calculate the NWFP distance. This analysis was performed by CANOCO 5 (ter Braak and Smilauer, 2012).

We avoided to correlate the values concerning the relative abundance of edible plants basing on the habitat coverage reported in the already mentioned Sicilian map of EUNIS habitats because we considered that any attempt to do this without a statistically significant amount of plot data may lead to unreliable and even misleading results. In fact, a plant may be very rare or very frequent in a given habitat regardless to the area of occupancy of the habitat itself in the island.

RESULTS

Inventory of Sicilian NWFP

A critical and updated inventory is provided and counts a total of 292 Sicilian NWFP (**Supplementary Table S1**). The taxa are listed in alphabetical order according to their scientific name. Synonyms, plant family, growth form, EIVs, altitudinal range, preferential habitat types, and CWR status are reported. Basic information on edible part(s), raw vs. cooked consumption, potential risks due to toxic or poisonous compounds, importance to local harvesters are provided, too.

According to the data issuing from our field investigations and those published by Geraci et al. (2018), the most commonly harvested Sicilian NWFP are *Asparagus acutifolius*, *Beta vulgaris* subsp. *maritima*, *Borago officinalis*, *Brassica rapa* subsp. *campestris*, *Cichorium intybus*, *Foeniculum vulgare s.l.*, *Sonchus oleraceus* and *Sonchus tenerrimus*, while the less commonly harvested are *Centranthus ruber*, *Narcissus tazetta s.l.*, *Papaver setigerum*, *Rorippa sylvestris*, *Rumex crispus*, and *Tordylium apulum*.

Our inventory includes 28 species which should probably be considered as archaeophytes in Sicily, but they are not mentioned in the checklist of the vascular alien flora of Italy (Galasso et al., 2018), and their regional status remains uncertain. The scientific name of these plant taxa is followed by an asterisk in the **Supplementary Table S1**. Moreover, many species commonly cultivated in Sicily and sometimes spreading from cultivations into semi-natural habitats were also taken into account, because palaeobotanical records testify that they occurred in Sicily throughout the Holocene. Hence, their native status cannot be discarded, even if their current distribution does not mirror the distribution range of their original wild populations. This is the case of *Castanea sativa*, *Celtis australis*, *Ceratonia siliqua*, *Corylus avellana*, *Ficus carica*, *Laurus nobilis*, *Mespilus germanica*, *Olea europaea*, *Pinus pinea*, *Sorbus domestica*, and *Vitis vinifera*.

On the contrary, a set of 34 plant species, despite being mentioned in the Sicilian ethnobotanical literature, were not considered in our analyses because they are not native (23 archaeophytes and 11 neophytes, see **Supplementary Table S2**). We also treated as archaeophytes several species which are native to other Italian regions but have certainly been introduced by man in Sicily. This is the case of *Asparagus officinalis*, *Rhus coriaria*, *Ruscus hypophyllum*, and *Salvia officinalis*, never spreading far from rural and disturbed areas, *Origanum onites*, whose Sicilian distribution range is restricted to the ruins of the ancient Syracuse, and *Cercis siliquastrum*, recently experiencing some successful colonization even in semi-natural habitats, but always next to the urban areas where it has been introduced.

The list of Sicilian NWFP was revised taking into account also the misleading information contained in some previous studies. In fact, due to identification errors, to the misinterpretation of Sicilian vernacular plant names and/or to the inappropriate inclusion of congeneric species, some authors reported plants which, either: (1) do not occur in Sicily or at least not in the study areas the papers were focused on; (2) have an extremely narrow niche and distribution range, so they could not be considered of any interest for local rural communities; (3) have been assumed to be food plants because they belong to the same genus of some NWFP, but their local exploitation needs to be better documented; (4) sometimes are used in Sicily as healing plants but cannot be used raw as food because they contain stinky, disgusting, toxic or even lethal compounds.

As a result of all the considerations above, 45 wrong, questionable and doubtful records were excluded from our NWFP inventory. These are listed in **Supplementary Table S3**,

along with the reasons for the rejection and the source(s) of the wrong/doubtful records.

Indeed, numerous reliable sources attest that Sicilian people harvest and eat at least 30 potentially toxic and 11 poisonous plant species. The Apiaceae (11 taxa) and Polygonaceae (6) account for more than 40% of them. Other families rich in dangerous food plants are Amaryllidaceae, Asteraceae, Fabaceae, and Ranunculaceae (3 taxa each), and Asphodelaceae, Iridaceae, and Solanaceae (2).

Diversity and Ecology of Sicilian NWFP

The 292 taxa listed in **Supplementary Table S1** belong to 51 plant families; **Figure 2** shows the families including more than 5 NWFP. Around 2/3 of the Sicilian NWFP belong to six families only, i.e., Asteraceae (80, i.e., 27.4% of the whole list), Brassicaceae (39, 13.4%), Lamiaceae (18, 6.2%), Fabaceae (18, 6.1%), Apiaceae (16, 5.5%), and Rosaceae (16, 5.5%). The number of species and genera mirrors the overall taxonomic richness of the main NWFP families, except for Fabaceae, which showed rather a low percentage of edible species despite being one of the richest families of the Sicilian vascular flora. None of the Sicilian NWFP belongs to Poaceae, even if this family is the second richest of the Sicilian vascular flora, counting 293 different species on the Island.

If we focus on the five growth forms which count more than 10 Sicilian NWFP, the rosulate edible herbs result proportionally more represented than in the whole regional flora and the same pattern has been observed for climbing herbs and lianas (**Figure 3**). The percentage of scapose species among wild food plants is slightly higher, whereas that of caespitose and geophytes slightly lower than that of the whole Sicilian flora.

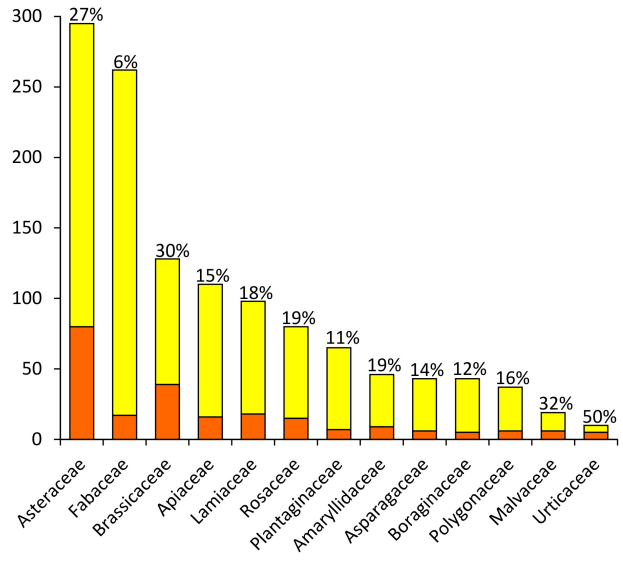


FIGURE 2 | Plant families including more than 5 NWFP in Sicily. Orange: number of NWFP; Yellow: number of non-edible plants. The percentage of NWFP per family is reported on the top of each column.

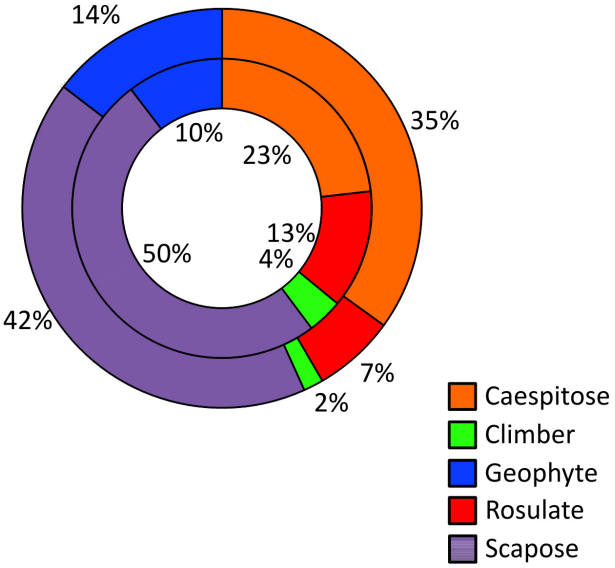


FIGURE 3 | Frequency of the five commonest growth forms among NWFP (inner circle) and in the entire Sicilian vascular flora (outer circle).

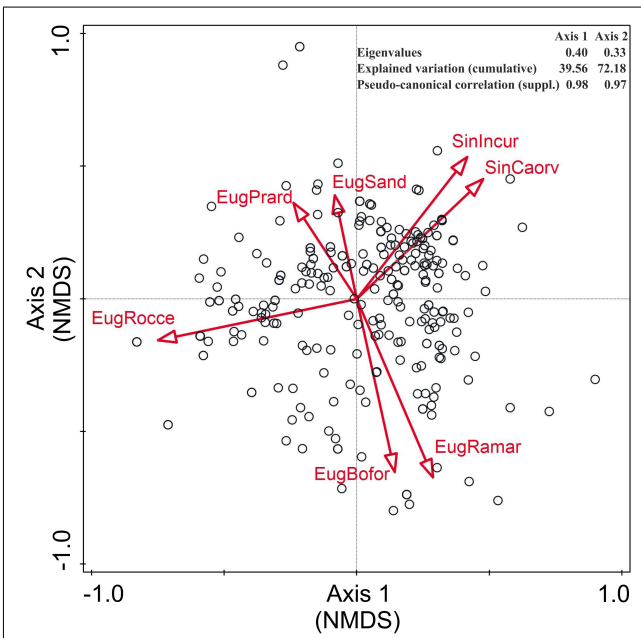


FIGURE 4 | Non-metric multidimensional scaling (NMDS) ordination plot of the NWFP of Sicily with respect to the habitats reported in the Flora of Italy (Pignatti et al., 2017–2019). Only the habitats explaining >3% of the total variance are displayed. EugBofor, woodlands and forests; EugPrard, dry grasslands; EugRamar, edges, clearings, deciduous shrubberies; EugRocce, rocks, barks, small outcrops; EugSand, sands; SinCaorv, crops, vegetable gardens, orchards, vineyards, olive groves; and SinIncur, fallows, ruderal and urban habitats.

The relationship between the NWFP and habitat types is shown in the NMDS ordination plots (**Figure 4** and **Supplementary Figure S1**). In both of the considered habitat

classifications, most species are concentrated near the center of the graphics, proving to be rather dioecious, with a clear preference for anthropogenic habitats. The first two axes of the NMDS explain 80.7% (EUNIS) and 72.18% (Flora of Italy) of the total variance, and in both cases, they appear to be related to a gradient of anthropogenic disturbance and a gradient of edaphic humidity, respectively.

Non-metric multidimensional scaling ordination plots show the prevalence of ruderal-nitrophilous species linked to anthropogenic vegetation and man-made habitats and ecotones, which represent the core of local edible plants. Yet, three further species groups cluster rather clearly: (1) the fruit-bearing woody species linked to forest and shrubland communities, (2) a group rich in chamaephytes and hemicryptophytes adapted to the harsh conditions of bare and nutrient-poor soils mostly occurring in rocky habitats, and (3) a group including many species adapted to sandy or salty soils.

Good Performers and Stress-Tolerant Sicilian NWFP

As far as EIVs are concerned, some significant differences in the *U*-, *R*-, and *N*-values of the Sicilian NWFP with respect to the whole regional flora (data not shown) have been detected. In particular, even if the mean values are comparable, the EIVs assigned to the NWFP tend to be less dispersed than those of the regional flora. Based on the EIVs, the Sicilian NWFP displaying the maximum tolerance to heat, drought, soil salinity and adaptation to nutrient-poor soils are reported in **Table 1**. This list includes 39 plant species having good performances under harsh environmental conditions, being able to overcome at least one out of four severe stress factors, i.e., two edaphic (low soil nutrient content and high soil salinity) ad two climatic (lack of water and extreme thermic events). In particular, *Capparis* species can be considered well adapted to harsh climatic and edaphic conditions (Mercati et al., 2019), with *Capparis sicula* showing a high ecological plasticity, as it can survive both in poorly aerated, salty and clayey soils as well as in secondary habitats such as roadsides (Gristina et al., 2014), while *C. spinosa* mainly colonizes coastal rocky cliffs. Wild cabbages (*Brassica* spp.) can thrive on nutrient-poor soils and tolerate intense heat and water stress growing on base-rich cliffs (from the coastline up to 1,000–1,200 m a.s.l.), while *Diplotaxis crassifolia* and *Thymbra capitata* can stand the enduring summer drought stress typical to Mediterranean grasslands and shrublands. Many NWFP belonging to Asteraceae, Brassicaceae, Lamiaceae, Valerianaceae tolerate very low soil nutrient availability and water stress but cannot stand thermal stress. *Cakile maritima*, *Crithmum maritimum*, *Echinophora spinosa*, and *Juncus acutus*, all growing in coastal habitats, are well adapted to salt-rich soils, the last two being resistant to nutrient-poor soils and only the latter also tolerant to water stress.

Sicilian NWFP and CWR

Based on literature reviews, we found that as much as 55 Sicilian NWFP (i.e., 18.8% of the total) are CWR (**Table 2**). Many of these plant taxa show high interest for their promising agronomic traits and/or for their high stress-tolerance. The richest families in CWR are Brassicaceae (9 taxa), Rosaceae (8), Asteraceae (7), Lamiaceae (6), Fabaceae (5), Apiaceae, and Asparagaceae (4), Amaryllidaceae and Capparaceae (2). With four taxa, the genera *Asparagus* and *Brassica* result to be the richest ones, followed by *Allium*, *Capparis*, *Cichorium*, *Mentha*, *Rubus*, *Salvia*, and *Sorbus* (2 taxa each).

To the family Asteraceae belong the wild progenitors of globe artichoke (*Cynara scolymus* L.), lettuce (*Lactuca serriola* L.) and chicories (*Cichorium intybus* and *Cichorium pumilum*), while the family Brassicaceae includes different species of wild cabbages (*Brassica incana*, *B. nigra*, *B. rupestris* and *B. tournefortii*) and wild rockets (*Eruca sativa*). Lamiaceae count many relatives of common spices such as mints (*Mentha spicata* and *Mentha suaveolens*), oregano (*Origanum vulgare* subsp. *viridulum*), rosemary (*Rosmarinus officinalis*), sages (*Salvia fruticosa* subsp. *thomasii* and *Salvia sclarea*). Other noteworthy CWR are the wild fennel (*Foeniculum vulgare*), wild strawberry (*Fragaria vesca*), wild garlic (*Allium ampeloprasum* and *Allium commutatum*) and wild capers (*Capparis sicula* and *C. spinosa*).

DISCUSSION

A Matter of Natural and Land-Use History

The landscapes of Sicily are the result of anthropogenic disturbances occurring over several millennia. Through the centuries, rural communities have managed their environment and farmed the land in their natural way, creating a rich diversity of landscapes, choral representation of the historical identity of the territory and human cultural heritage (Guarino and Pasta, 2017, and references therein). One of the first effects of human land use was an increased fire frequency: wildfires were the easiest way to obtain grass-dominated terrains, which could be used as rangelands or, eventually, cultivated. The prevalence of NWFP in semi-natural and anthropogenic habitats could be seen as the result of a process of selection and adaptation lasting since 10,000 years, at least. The plants growing in habitats of this kind are providing most of the harvest, because it is easy to identify them and because they are usually found in dense and homogeneous populations. For instance, the relatively high percentage of rosulate plants among NWFP could mirror their ecological adaptation to grazing disturbance. This growth form enables to reduce the damages (namely defoliation) caused by the bites and by the trampling activity of domestic herbivores (Xu et al., 2013, and references therein). Most of the edible climbers, vines and lianas (e.g., *Clematis*, *Lathyrus*, *Rubia*, *Smilax*, *Tamus*, and *Vicia*) are instead common in ecotones, which have been created and shaped by humans.

As far as we know, no literature data is explaining why rosette-bearing and scapose (erect) taxa are more common than caespitose among wild food plants, whereas the relatively low percentage of edible geophytes is likely to depend on the high frequency of poisonous species within this group. As for EIVs, the values of 'U' (edaphic humidity) of the Sicilian edible plants appear to be relatively low. This pattern matches with

TABLE 1 | An overview on the Sicilian NWFP which show the best performance with respect to four major stress factors, as suggested by the numerical values of the selected EIVs, i.e., T (temperature, range 1–12), U (edaphic humidity, range 1–11), N (nutrients, range 1–9), and S (salinity, range 0–3).

Scientific name according to Pignatti et al. (2017–2019)	**Plant Family (Chase et al., 2016)**	***T***	***U***	***N***	***S***
Taxa which are well adapted to climatic and edaphic stress					
Capparis sicula Veill.	Capparaceae	10	2	**1**	**3**
Capparis spinosa L.	Capparaceae	10	2	**1**	2
Lotus cytisoides L.	Fabaceae	10	**1**	**1**	2
Taxa which are well adapted to nutrient-poor soils and tolerate thermic and water stress					
Astragalus boeticus L.	Fabaceae	11	**1**	**1**	0
Biscutella maritima Ten.	Brassicaceae	10	2	**1**	0
Brassica incana Ten.	Brassicaceae	10	2	**1**	0
Brassica rupestris Raf.	Brassicaceae	10	2	**1**	0
Brassica tournefortii Gouan	Brassicaceae	**12**	2	**1**	0
Carlina gummifera (L.) Less.	Asteraceae	11	**1**	**1**	0
Carlina sicula Ten.	Asteraceae	11	**1**	**1**	0
Chamaerops humilis L.	Arecaceae	10	**1**	**1**	0
Diplotaxis crassifolia (Raf.) DC.	Brassicaceae	11	2	**1**	0
Plantago serraria L.	Plantaginaceae	10	2	**1**	0
Rumex bucephalophorus L.	Polygonaceae	**12**	2	**1**	0
Thymbra capitata (L.) Cav.	Lamiaceae	10	2	**1**	0
Taxa which are well adapted to nutrient-poor and salt-rich soils and tolerant to water stress					
Crithmum maritimum L.	Apiaceae	8	**1**	**1**	**3**
Taxa which are adapted to nutrient-poor and salt-rich soils					
Echinophora spinosa L.	Apiaceae	8	4	**1**	2
Taxa which are well adapted to nutrient-poor soils and tolerant to water stress					
Centranthus ruber (L.) DC.	Valerianaceae	8	2	**1**	0
Fedia graciliflora Fisch. & C.A. Mey.	Valerianaceae	9	2	**1**	0
Hyoseris radiata L.	Asteraceae	8	2	**1**	0
Hyoseris scabra L.	Asteraceae	9	**1**	**1**	0
Hypochoeris cretensis (L.) Chaub. & Bory	Asteraceae	8	2	**1**	0
Hypochaeris glabra L.	Asteraceae	8	2	**1**	0
Hypochaeris laevigata (L.) Ces., Pass. & Gibelli	Asteraceae	8	2	**1**	0
Hypochaeris radicata L.	Asteraceae	8	2	**1**	0
Lavandula stoechas L.	Lamiaceae	9	2	**1**	0
Lobularia maritima (L.) Desv.	Brassicaceae	9	2	**1**	0
Micromeria juliana (L.) Benth. ex Rchb.	Lamiaceae	8	2	**1**	0
Papaver setigerum DC.	Papaveraceae	8	2	**1**	0
Phagnalon saxatile (L.) Cass.	Asteraceae	9	2	**1**	0
Rosmarinus officinalis L.	Lamiaceae	8	2	**1**	0
Thymus spinulosus Ten.	Lamiaceae	8	2	**1**	0
Tolpis umbellata Bertol.	Asteraceae	9	2	**1**	0
Tolpis virgata (Desf.) Bertol.	Asteraceae	8	2	**1**	0
Valerianella eriocarpa Desv.	Valerianaceae	9	2	**1**	0
Taxa which are well adapted to nutrient-poor soils					
Plantago lagopus L.	Plantaginaceae	9	3	**1**	0
Reseda alba L. s.l.	Resedaceae	8	3	**1**	0
Taxa which are adapted to salt-rich soils					
Cakile maritima Scop.	Brassicaceae	8	6	8	**3**
Juncus acutus L.	Juncaceae	8	8	3	**3**
Number of taxa showing the maximum value of tolerance to the selected stress factors		**2**	**7**	**37**	**4**

Best responses to one or more stress factors are in bold.

TABLE 2 | Families of Sicilian NWFP – including CWR and/or stress-tolerant taxa – to be tested in order to detect and develop new crops for their high potential as resources for agronomic, genetic, pharmaceutical, and nutritional purposes.

Family of NWFP of agronomic interest	Number of sp./subsp.	Number of genera	Number of CWR	CWR	Other taxa to be tested for the development of new crops	Other taxa to be tested for the development of new stress-tolerant crops
Amaranthaceae	2	2	1	*Beta vulgaris* L. subsp. *maritima* (L.) Arcang.	*Chenopodium album* L.	
Amaryllidaceae	4	1	2	*Allium ampeloprasum* L. *Allium commutatum* Guss.	*Allium ursinum* L. *Allium triquetrum* L.	
Apiaceae	6	6	4	*Apium graveolens* L. *Daucus carota* L. *s.l.* *Foeniculum vulgare* Mill. *s.l.* *Helosciadium nodiflorum* (L.) W.D.J.Koch subsp. *nodiflorum*		*Crithmum maritimum* L. *Echinophora spinosa* L.
Asparagaceae	6	3	4	*Asparagus acutifolius* L. *Asparagus albus* L. *Asparagus aphyllus* L. *Asparagus horridus* L.	*Muscari comosum* (L.) Mill. *Ruscus aculeatus* L.	
Asteraceae	29	19	7	*Cichorium intybus* L. *Cichorium pumilum* Jacq. *Cynara cardunculus* L. *s.l.* *Lactuca serriola* L. *Matricaria chamomilla* L. *Scorzonera undulata* Vahl subsp. *deliciosa* (DC.) Maire *Tragopogon porrifolius* L.	*Hedypnois cretica* (L.) Dum.Cours. *Hedypnois rhagadioloides* (L.) F.W.Schmidt *Helminthotheca echioides* (L.) J.Holub *Leontodon siculus* (Guss.) Nyman *Leontodon tuberosus* L. *Reichardia picroides* (L.) Roth *Rhagadiolus stellatus* (L.) Gaertn. *Scorzoneroides cichoracea* (Ten.) Greuter *Scorzoneroides muelleri* (Sch.Bip.) Greuter & Talavera subsp. *muelleri* *Sonchus oleraceus* L. *Sonchus tenerrimus* L. *Urospermum picroides* (L.) F.W.Schmidt	*Carlina sicula* Ten. *Hyoseris radiata* L. *Hyoseris scabra* L. *Hypochaeris cretensis* (L.) Chaub. & Bory *Hypochaeris glabra* L. *Hypochaeris laevigata* (L.) Ces., Pass. & Gibelli *Hypochaeris radicata* L. *Phagnalon saxatile* (L.) Cass. *Tolpis umbellata* Bertol. *Tolpis virgata* (Desf.) Bertol. *s.l.*
Betulaceae	1	1	1	*Corylus avellana* L.		
Brassicaceae	17	11	9	*Barbarea bracteosa* Guss. *Brassica incana* Ten. *Brassica nigra* (L.) Koch *Brassica rapa* L. subsp. *campestris* (L.) Clapham *Brassica rupestris* Raf. *s.l.* *Eruca sativa* Mill. *Nasturtium officinale* R.Br. *Raphanus raphanistrum* L. *s.l.* *Sinapis alba* L. *s.l.*	*Erucastrum virgatum* C.Presl *Rapistrum rugosum* (L.) All. *Diplotaxis erucoides* (L.) DC. *Diplotaxis muralis* (L.) DC. *Diplotaxis tenuifolia* (L.) DC.	*Cakile maritima* Scop. *Lobularia maritima* (L.) Desv. *Diplotaxis crassifolia* (Raf.) DC.
Capparaceae	2	1	2	*Capparis sicula* Veill. *Capparis spinosa* L. subsp. *rupestris* (Sm.) Nyman		
Ericaceae	1	1	0		*Arbutus unedo* L.	
Fabaceae	4	4	4	*Ceratonia siliqua* L. *Glycyrrhiza glabra* L. *Lathyrus cicera* L. *Pisum sativum* L. subsp. *biflorum* (Raf.) Soldano *Vicia sativa* L.		
Fagaceae	1	1	1	*Castanea sativa* Mill.		
Iridaceae	1	1	1	*Crocus longiflorus* Raf.		

(Continued)

TABLE 2 | Continued

Family of NWFP of agronomic interest	Number of sp./subsp.	Number of genera	Number of CWR	CWR	Other taxa to be tested for the development of new crops	Other taxa to be tested for the development of new stress-tolerant crops
Lamiaceae	11	9	6	*Mentha spicata* L. *Mentha suaveolens* Ehrh. *Origanum vulgare* L. subsp. *viridulum* (Martrin-Donos) Nyman *Rosmarinus officinalis* L. *Salvia fruticosa* Mill. subsp. *thomasii* (Lacaita) Brullo, Guglielmo, Pavone, and Terrasi *Salvia sclarea* L.	*Clinopodium nepeta* (L.) Kuntze	*Lavandula stoechas* L. *Micromeria juliana* (L.) Benth. ex Rchb. *Thymbra capitata* (L.) Cav. *Thymus spinulosus* Ten.
Lauraceae	1	1	1	*Laurus nobilis* L.		
Moraceae	1	1	1	*Ficus carica* L. *s.l.*		
Oleaceae	1	1	1	*Olea europaea* L. var. *sylvestris* (Mill.) Lehr.		
Rosaceae	8	6	8	*Crataegus monogyna* Jacq. *Fragaria vesca* L. *Prunus spinosa* L. *Pyrus spinosa* Forssk. *Rubus idaeus* L. *Rubus ulmifolius* Schott *Sorbus aucuparia* L. subsp. *praemorsa* (Guss.) Nyman *Sorbus domestica* L.		
Urticaceae	4	1	0		*Urtica dioica* L. *Urtica membranacea* Poir. *Urtica pilulifera* L. *Urtica urens* L.	
Valerianaceae	1	1	1	*Valerianella locusta* (L.) Laterr.		
Vitaceae	1	1	1	*Vitis vinifera* L. subsp. *sylvestris* (C.C.Gmel.) Hegi		
Total	102	72	55			

the fact that basing on the bioclimatic classification of Sicily (Bazan et al., 2015), most of the Sicilian NWFP thrive under thermo-mediterranean bioclimatic conditions, characterized by long-lasting summer drought stress. The values of 'R' and 'N' indicate that many NWFP prefer base- and nutrient-rich soils, typical to the Sicilian hilly landscapes prone to human (namely agro-pastoral) practices. Interestingly, the high rate of ruderal-nitrophilous species among the edible plants appreciated for their fleshy stem or foliage was confirmed in the study carried out in northern Croatia by Vitasović Kosić et al. (2017), who also used EIVs to investigate the ecology of local wild food plants. Moreover, the altitudinal distribution pattern of the Sicilian NWFP (data not shown) suggest that they are slightly more concentrated in the lowlands and hilly areas of the island, where most of the permanent settlements and arable lands are concentrated since millennia.

Non-metric multidimensional scaling provides interesting clues on the spatial and temporal distribution of the food resources afforded by NWFP. For instance, most of the woody species are linked to forests, riverine plant communities and shrublands. They produce large amounts of fruits or berries that can be eaten raw, and represent an important food resource between the end of spring and the end of summer when all the annual and most of the perennial herbaceous NWFP are not available anymore.

Most of the edible Lamiaceae group together and are well adapted to nutrient-poor and rocky habitats (**Figure 4** and **Supplementary Figure S1**), where they cope with water and nutrient shortage by producing many aromatic compounds that play the twofold action of helping them to save water and to inhibit predation by herbivores. Other stress-tolerant plants typical to sandy and/or salty soils form another cluster rich in species adapted to high salt input and long-lasting water shortage.

Under an ecological perspective, nutrient-poor habitats promote selection for traits allowing efficient resource conservation, while nutrient-rich environments select for species with acquisitive trait profiles (Reich, 2014). Domesticated plants, and in particular herbaceous crops, share few common traits and represent a small portion of the phenotypic spectrum displayed by their wild ancestors: the majority of those living on nutrient-rich soils bear soft, large and short-lived leaves, are fast-growing and proficient competitors, with high leaf nitrogen concentration and tall canopies (Freschet et al., 2015; Milla et al., 2018). As cultivation generally involved higher and more regular nutrient and water supply rates, humans probably focused their interest on NWFP bearing resource-acquisition

trait profiles (Bogaard et al., 2013; Araus et al., 2014). Some of them, like the ancestors of several cereal crops, were chosen because they showed a rapid shift of their functional traits under domestication, being able to invest more energy on leaf biomass and height growth than other wild grasses that were used by hunter-gatherers, but were never domesticated (Cunniff et al., 2014). Many other cultivated plants descend from wild ancestors which were pre-adapted for cultivation thanks to several favoring traits (Triboullois et al., 2015; Martin-Robles et al., 2019). These considerations fit perfectly with the fact that most of the Sicilian NWFP, commonly used as vegetables and concentrated near the center of **Figure 4** and **Supplementary Figure S1**, share many morphological and ecological traits: they colonize nutrient-rich soils and ecotones, they bear large and tender leaves, they grow very fast, they are good competitors.

The massive number of edible Asteraceae (subfam. Cichorioideae) is a common pattern in Mediterranean countries (Della et al., 2006; Leonti et al., 2006; Rivera et al., 2006; Hadjichambis et al., 2008). Interestingly, Cichorioideae appear to be strongly associated with pastoral activities since ancient times (Florenzano et al., 2015). Also, Brassicaceae are an essential component in wild harvesting in south Italian regions (Biscotti et al., 2018). These plants have probably been gathered since the very first stages of agriculture by Neolithic farmers, as they behaved as weeds.

All these ecological clues lead to the same conclusion: the majority of the Sicilian NWFP may be considered as 'old companions' of local human communities. Early Sicilians probably discovered very soon that many 'proto-weeds' colonizing the open spaces created after the disruption of pristine woodlands were useful and tasty, as it has been recently shown also for Israel (Snir et al., 2015). Also, the available palaeoecological studies (Sadori et al., 2008; Stika et al., 2008; Noti et al., 2009; Tinner et al., 2009) suggest that from the beginning of the Holocene (∼10 Ka) onward, human impact has been an important – if not the main – factor inducing the final opening of the Sicilian landscape. With their activities (burning, clearing, cutting, farming, plowing, etc.) local Neolithic communities not only fostered the success of many non-native pioneer light-demanding plants inadvertently introduced with crop species (the so-called 'archaeophytes') but also ended the shaping of the regional natural and semi-natural landscape, giving rise to a complex mosaic of prevalently open habitats with scattered nuclei of woodlands, shrublands and species-rich garrigues and grasslands (Guarino et al., 2005; Guarino, 2006; Brullo and Guarino, 2007).

The Importance of a Multidisciplinary Approach

An unexpected result of our investigation concerned the detection of 46 wrong, questionable and doubtful plants that cannot be included among the Sicilian NWFP.

There are two distinct groups of experts coping with useful plants, plant uses and plant names: from one side the 'practitioners,' such as farmers, shepherds, artisans, cooks, forest workers, fishermen, hunters, etc., from the other side the 'scholars,' like ethnobotanists, ethno-anthropologists, philologists, historians, biochemists, veterinaries, etc. Until recent past, these two groups have been hardly sharing their knowledge. This is the main reason why we frequently find plant misidentifications in the papers written by non-botanists, while 'pure' botanists often fail to spell correctly, to trace the root of vernacular names and/or to report adequately the uses of the plants they study. Such 'communication gaps' may cause risky misidentifications due to the transfer of erroneous information concerning the alimentary use and therapeutic properties of some poisonous plants (i.e., Corsi and Pagni, 1979; Provitina, 1991; Atzei, 2003; Giardina et al., 2007; Łuczaj et al., 2012; see **Supplementary Table S3**).

The use of toxic and poisonous plants is a common pattern worldwide, especially were seasonality strongly affects food availability, and local human communities learned to exploit even the less attracting resources. This fact is not only a consequence of famine but also a question of timing and a result of century-old folk knowledge: in fact, some plants may contain a much lower concentration of poisonous compounds in some periods of the year and/or in some specific organs. Moreover, the fact that some communities use a species does not necessarily mean it is safe to eat it. Hence, detailed data on the gathering and cooking procedures are of paramount importance, and such information should be taken into account before discarding any poisonous plant from the list of edible ones: this is the case of several species which are cooked before being consumed as vegetables (like the tender shoots of *Asphodeline lutea*, *Carlina gummifera*, *Clematis vitalba*, and *Tamus communis*) or to prepare jams (such as the fruits of *Rubia peregrina*).

The data reported by Geraci et al. (2018) are based on interviews carried out throughout the main island of Sicily and only focused on the plants and plant portions consumed as vegetables, while c. 25% of our interviews were carried out on the circum-Sicilian islets and regarded the whole spectrum of NWFP, including 90 aromatic and fruit-bearing plant species. The same value of harvest frequency was assigned to 42 out of the 202 plants (20.8%) of the common pool, while for 49 taxa (24.2%) the evaluation was consistently different (2–4 points of the adopted scale). Most of the differences between the two assessments may depend on the way (total number, geographical distribution and scope) the interviews were carried out by the two research groups.

Our results point out the need of making further efforts in order to retrieve and homogenize the information already available in the gray ethnobotanical literature, mostly published in Italian on regional and local papers which are often hard to find and to consult. To overcome errors and to improve the quality and effectiveness of their researches, ethnobotanists, ethnologists, and all the skilled persons involved in wild plant harvesting should start a tighter collaboration and launch an ambitious multidisciplinary research program focused on Sicilian traditional environmental knowledge (TEK: Heckler, 2012) and edible plants' use.

In order to avoid the transfer of erroneous information, in our opinion, the international scientific community should perform a more severe review of the data concerning the human diet.

As a rule of thumb, papers concerning edible plants should be not only written but also reviewed (and eventually rejected) by skilled local botanists in order to avoid that wrong information is published on scientific journals and reaches (and eventually kills) uninformed readers.

Useful, Yet at the Brink of Oblivion

Taking into account the worrying forecasts regarding food security for the forthcoming years due to climate warming and global change (Ford-Lloyd et al., 2011), during the last decades an increasing effort has been addressed on the detection and conservation of CWR (Valdés et al., 1997; Heywood et al., 2007; Maxted et al., 2010; Dempewolf et al., 2014). Regional and national inventories have been strongly encouraged, and seed banks are currently created and implemented (Maxted et al., 2007, 2012; Vincent et al., 2013). These efforts aim at preserving wild food plants, restoring old varieties, finding out new technologies to enhance their use and new solutions to manage modern agricultural systems (Hajjar and Hodgkin, 2007) more sustainably (Caneva et al., 2013). In this framework, we are convinced that in-depth ecological insights on NWFP may innovate and increase the spectrum of cultivated vegetables.

In the perspective of implementing sustainable agriculture under the current global change scenario, each of the 39 plants reported in **Table 1**, showing the most promising traits in terms of tolerance to several stress factors (water shortage, high temperatures, and edaphic constraints), could be cultivated for experimental purposes. Also, given that a considerable number of Sicilian NWFP can be considered CWR, they could be tested in *ad hoc* genetic improvement programs of traditional crops, like it has been already done for *Allium* (Odeny and Narina, 2011), *Asparagus* (Falavigna et al., 2008; Kanno and Yokoyama, 2011), *Beta* (McGrath et al., 2011), *Daucus* (Grzebelus et al., 2011; Iorizzo et al., 2013), *Lactuca* (Davey and Anthony, 2011), *Vicia* (Bryant and Hughes, 2011), and for many Brassicaceae (*Brassica*: Branca and Cartea, 2011; *Diplotaxis*: Pignone and Martínez-Laborde, 2011; *Eruca*: Pignone and Gómez-Campo, 2011) and Rosaceae (Fragaria: Hummer et al., 2011; *Malus*: Ignatov and Bodishevskaya, 2011; *Prunus*: Potter, 2011; *Pyrus*: Bell and Itai, 2011; *Rubus*: Graham and Woodhead, 2011). Moreover, further research should be focused on many species-rich and promising families and genera of NWFP (**Table 2**) whose members could be used in domestication programs aiming at developing new crops: i.e., new fragrances such as thymes (e.g., *Thymbra capitata* and *Thymus spinulosus*) and other Lamiaceae such as *Clinopodium nepeta*, *Lavandula stoechas*, and *Micromeria juliana*, new vegetables such as several wild species of *Asparagus* and wild rockets (e.g., *Diplotaxis* spp., *Erucastrum virgatum*).

Any future breeding activity concerning the most promising NWFP should start from a better understanding of their ecological traits, preferring those species that may provide additional ecosystem services and belong to genera that proved to be prone to a sensitive shift of above- and below-ground biomass allocation during the domestication process (Denison, 2012; Milla et al., 2015, 2017).

Despite their potential and prominent interest, especially for future generations, most of Sicilian NWFP seem doomed to oblivion. Due to ongoing social and economic changes in the Mediterranean area, the vast majority of the last custodians of ethnobotanical knowledge are more than 70 years old. As a consequence of their aging, not only the traditional knowledge but also entire cultural landscapes – shaped by men for millennia and providing the most suitable habitat for plenty of NWFP – is deemed to be lost within the next 10–20 years. We will lose most of the diversity of NWFP if we will stop eating them. Interviews to last living memories of the past cultural heritage (De Gregorio, 2008; Sottile and Genchi, 2010) are urgently needed before traditional knowledge fades forever (Quave and Saitta, 2016; Cucinotta and Pieroni, 2018), in addition to the collection and the conservation of Sicilian wild and cultivated germplasm (Hammer and Perrino, 1995; La Mantia and Pasta, 2005; Portis et al., 2005; Hammer and Laghetti, 2006; Schicchi et al., 2008; Forconi and Guidi, 2013; Marino et al., 2013).

CONCLUSION

In this paper, we provide some preliminary clues about the ecology of all Sicilian NWFP, but knowledge on this topic needs to be improved, in order to detect any discrepancy between their real frequency, their altitudinal range, their distribution pattern and their local use as a food resource. To do that, we still need to analyze the data of all the published vegetation surveys carried out in Sicily and to cover the ethnobotanical knowledge gaps in some sectors of the island.

Tightly connected with the traditional agro-pastoral practices and landscapes, many NWFP are becoming rarer and rarer as a consequence of the ongoing processes of land-use change and abandonment. The only way to maintain both TEK and NWFP is to envisage concrete and shared measures aiming at promoting the self-sustainment of traditional agro-silvopastoral practices at the European, the national and the regional level.

Humans have shaped natural landscapes worldwide since prehistoric times. However, they often succeeded to modify ecosystem services and functioning without destroying them (Kareiva et al., 2007; Willis et al., 2007). Instead of trying to come back to nature, which may result a hard and even anachronistic target (Willis and Birks, 2006), we should try to inherit and replicate the past practices combining extensive and sustainable land use with the conservation of species diversity and ecosystem services (Plieninger et al., 2006; Guarino and Pignatti, 2010). In the end, TEK in general, and NWFP use in particular, issue from a wise and dynamic combination of silvopastoral and crop farming activities, and land use practices which allowed people and landscape to co-evolve and co-occur over millennia (Mercuri et al., 2019).

We should try to learn more about the best solutions adopted by mankind throughout history, focusing on the co-evolution between weeds, plant harvesters and farmers. At the same time, we should make more efforts in order to detect and value the TEK of the last Mediterranean bio-cultural refugia (Barthel et al., 2013).

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

SP, FC, and RG conceived and supervised the project. AL carried out interviews across the whole Sicilian territory and built up the revised checklist of Sicilian NWFP with the help of SP and RG. SP, CM, RG, FC, and AG analyzed and interpreted the data. SP, RG, AG, GG, CM, and FC wrote the first draft. All authors made a substantial, direct and intellectual contribution to this work. All authors approved the final version of the manuscript.

FUNDING

This research was supported by Regione Siciliana (PO FESR 2014/2020, Azione 1.1.5 -Project: Sicily Seeds: Metodologie e tecnologie innovative per il recupero, la moltiplicazione, la valorizzazione e l'utilizzo di piante spontanee commestibili della flora siciliana, Grant No. 08PA6201000062). CM was supported by the grant no. 19-28491X of the Czech Science Foundation.

ACKNOWLEDGMENTS

The kind support of Francesca La Bella (IBBR-CNR, Palermo), Tommaso La Mantia and Giovanna Sala (Department of Agricultural, Food and Forest Sciences, University of Palermo) during our bibliographic research was much appreciated. We are grateful to Giovanni Salerno (Department of Environmental Biology, University of Rome III) for his advice about the best way to treat the doubtful records concerning potentially dangerous species reported as edible by other authors coping with Sicilian wild food plants.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.00388/full#supplementary-material

REFERENCES

Albrecht, H., Cambecèdes, J., Lang, M., and Wagner, M. (2016). Management options for the conservation of rare arable plants in Europe. *Bot. Lett.* 163, 389–415. doi: 10.1080/23818107.2016.1237886

Aleo, M., Cambria, S., and Bazan, G. (2013). Tradizioni etnofarmacobotaniche in alcune comunità rurali dei Monti di Trapani (Sicilia occidentale). *Quad. Bot. Ambientale Appl.* 24, 27–48.

Angelini, P., Bianco, P., Cardillo, A., Francescato, C., and Oriolo, G. (2009). *Gli Habitat in Carta Della Natura*. ISPRA, Manuali e Linee Guida 49/2009: Roma.

Araus, J. L., Ferrio, J. P., Voltas, J., Aguilera, M., and Buxó, R. (2014). Agronomic conditions and crop evolution in ancient Near East agriculture. *Nat. Commun.* 5:3953. doi: 10.1038/ncomms4953

Arcidiacono, S. (2016). *Etnobotanica Etnea. Le Piante Selvatiche e l'uomo*. Palermo: Danaus, 148.

Arcidiacono, S., Costa, R., Marletta, G., Pavone, P., and Napoli, M. (2010). Usi popolari delle piante selvatiche nel territorio di Villarosa (EN – Sicilia Centrale). *Quad. Bot. Ambientale Appl.* 21, 95–118.

Arcidiacono, S., Napoli, M., Oddo, G., and Pavone, P. (2007). Piante selvatiche d'uso popolare nei territori di alcara li fusi e militello rosmarino (Messina, N-E Sicilia). *Quad. Bot. Ambientale Appl.* 18, 105–146.

Atzei, D. (2003). *Le Piante Nella Tradizione Popolare Della Sardegna*. Sassari: C. Delfino Editore.

Barthel, S., Crumley, C. L., and Svedin, U. (2013). Biocultural refugia: combating the erosion of diversity in landscapes of food production. *Ecol. Soc.* 18:71. doi: 10.5751/ES-06207-180471

Bazan, G., Marino, P., Guarino, R., Domina, G., and Schicchi, R. (2015). Bioclimatology and vegetation series in sicily: a geostatistical approach. *Ann. Bot. Fenn.* 52, 1–18. doi: 10.5735/085.052.0202

Bell, R. L., and Itai, A. (2011). "Pyrus," in *Wild Crop Relatives: Genomic and Breeding Resources: Temperate Fruits*, ed. C. Kole (Berlin: Springer), 147–177.

Bharucha, Z., and Pretty, J. (2010). The roles and values of wild foods in agricultural systems. *Philos. Trans. R. Soc. B* 365, 2913–2926. doi: 10.1098/rstb.2010.0123

Biscotti, N., Bonsanto, D., and Del Viscio, G. (2018). The traditional food use of wild vegetables in Apulia (Italy) in the light of Italian ethnobotanical literature. *Italian Bot.* 5, 1–24. doi: 10.3897/italianbotanist.5.22297

Bogaard, A., Fraser, R., Heaton, T. H. E., Wallace, M., Vaiglova, P., Charles, M., et al. (2013). Crop manuring and intensive land management by Europe's first farmers. *Proc. Natl. Acad. Sci. U.S.A.* 110, 12589–12594. doi: 10.1073/pnas.1305918110

Branca, F., and Cartea, E. (2011). "Brassica," in *Wild Crop Relatives: Genomic and Breeding Resources: Oilseeds*, ed. C. Kole (Berlin: Springer), 17–36.

Brullo, S., and Guarino, R. (2007). The Mediterranean weedy vegetation and its origin. *Ann. Bot.* 7, 101–110.

Bryant, J. A., and Hughes, S. G. (2011). "Vicia," in *Wild Crop Relatives: Genomic and Breeding Resources: Legume Crops and Forages*, ed. C. Kole (Berlin: Springer), 273–290.

Caneva, G., Pieroni, A., and Guarrera, P. M. (2013). *Etnobotanica. Conservazione di un Patrimonio Culturale Come Risorsa per uno Sviluppo Sostenibile*. Bari: Edipuglia.

Carbone, S., Di Stefano, P., Monaco, C., and Sulli, A. (2018). *Sicilia e Isole Minori: Guide Geologiche Regionali*. Bologna: Società Geologica Italiana.

Chase, M. W., Christenhusz, M. J. M., Fay, M. F., Byng, J. W., Judd, W. S., Soltis, D. E., et al. (2016). An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG IV. *Bot. J. Linn. Soc.* 181, 1–20. doi: 10.1111/boj.12385

Coria, G. (1989). *La Magia Verde. Piante Fiori ed Erbe Negli Iblei*. Palermo: Publisicula, 164.

Corsi, G., and Pagni, A. M. (1979). Studi sulla flora e vegetazione del Monte Pisano (Toscana Nord-Occidentale). V. Le piante spontanee nell'alimentazione popolare. *Atti. Soc. Tosc. Sci. Tat. Mem. B* 86, 79–101. doi: 10.1080/00837792.1980.10670205

Cucinotta, F., and Pieroni, A. (2018). 'If you want to get married, you have to collect virdura': the vanishing custom of gathering and cooking wild food plants on Vulcano, Aeolian Islands, Sicily. *Food Cult. Soc.* 21, 539–567. doi: 10.1080/15528014.2018.1481263

Cunniff, J., Wilkinson, S., Charles, M., Jones, G., Rees, M., and Osborne, C. P. (2014). Functional traits differ between cereal crop progenitors and other wild grasses gathered in the Neolithic fertile Crescents. *PLoS One* 9:e87586. doi: 10.1371/journal.pone.0087586

Davey, M. R., and Anthony, P. (2011). "Lactuca," in *Wild Crop Relatives: Genomic and Breeding Resources: Vegetables*, ed. C. Kole (Berlin: Springer), 115–128.

De Gregorio, N. (2008). *Cibo e Parole di una Comunità di Montagna. A Cammarata con il Questionario dell'ALS:Materiali e ricerche dell'ALS n. 20*. Palermo: Centro di studi filologici e linguistici siciliani, 173.

De Michele, R., La Bella, F., Gristina, A. S., Fontana, F., Pacifico, D., Garfi, G., et al. (2019). Phylogenetic relationship among wild and cultivated grapevine in Sicily:

a hotspot in the middle of the Mediterranean Basin. *Front. Plant Sci.* 10:1506. doi: 10.3389/fpls.2019.01506

Della, A., Paraskeva-Hadjichambi, D., and Hadjichambis, A. C. (2006). An ethnobotanical survey of wild edible plants of Paphos and Larnaca countryside of Cyprus. *J. Ethnobiol. Ethnomed.* 2:34. doi: 10.1186/1746-4269-2-34

Dempewolf, H., Eastwood, R. J., Guarino, L., Khoury, C. K., Müller, J. V., and Toll, J. (2014). Adapting agriculture to climate change: a global initiative to collect, conserve, and use crop wild relatives. *Agroecol. Sust. Food* 38, 369–377. doi: 10.1080/21683565.2013.870629

Denison, R. (2012). *Darwinian Agriculture. How Understanding Evolution Can Improve Agriculture*. Princeton: Princeton University Press.

Diez, C. M., Trujillo, I., Urdiroz, N. M., Barranco, D., Rallo, L., Marfil, P., et al. (2015). Olive domestication and diversification in the Mediterranean Basin. *New Phytol.* 206, 436–447. doi: 10.1111/nph.13181

Disciglio, G., Tarantino, A., Frabboni, L., Gagliardi, A., Giuliani, M. M., Tarantino, E., et al. (2017). Qualitative characterization of cultivated and wild edible plants: mineral elements, phenols content and antioxidant capacity. *Ital. J. Agron.* 12:1036. doi: 10.4081/ija.2017.1036

Falavigna, A., Alberti, P., Casali, P. E., Toppino, L., Huaisong, W., and Mennella, G. (2008). Interspecific hybridization for asparagus breeding in Italy. *Acta Hortic.* 776, 291–297. doi: 10.17660/ActaHortic.2008.776.37

Florenzano, A., Marignani, M., Rosati, L., Fascetti, S., and Mercuri, A. M. (2015). Are Cichorieae an indicator of open habitats and pastoralism in current and past vegetation studies? *Plant Biosyst.* 149, 154–165. doi: 10.1080/11263504.2014.998311

Forconi, V., and Guidi, S. (2013). *Frutti Dimenticati e Biodiversità Recuperata. Il Germoplasma Frutticolo e Viticolo Delle Agricolture Tradizionali Italiane. Casi studio: Isole Della Sicilia, Lombardia*. Ispra: Quaderni Natura e Biodiversità, 136.

Ford-Lloyd, B. V., Schmidt, M., Armstrong, S. J., Barazani, O., Engels, J., Hadas, R., et al. (2011). Crop wild relatives: undervalued, underutilized and under threat? *BioScience* 61, 559–565. doi: 10.1525/bio.2011.61.7.10

Freschet, G. T., Swart, E. M., and Cornelissen, J. H. C. (2015). Integrated plant phenotypic response to contrasting above- and below-ground resources: key roles of specific leaf area and root mass fraction. *New Phytol.* 206, 1247–1260. doi: 10.1111/nph.13352

Galasso, G., Conti, F., Peruzzi, L., Ardenghi, N. M. G., Banfi, E., Celesti-Grapow, L., et al. (2018). An updated checklist of the vascular flora alien to Italy. *Plant Biosyst.* 152, 556–592. doi: 10.1080/11263504.2018.1441197

Gamboni, M., Carimi, F., and Migliorini, P. (2012). "Mediterranean diet: an integrated view," in *Sustainable Diets and Biodiversity: Directions and Solutions for Policy, Research and Action*, eds B. Burlingame and S. Dernini (Rome: Food and Agriculture Organization and Bioversity International), 262–273.

Gargano, M. L., Venturella, G., Lazzara, S., Lo Nardo, R., and Saporita, P. (2018). Ethnobotanical knowledge in some rural communities of northern Sicily (Palermo, Italy). *Atti. Soc. Tosc. Sci. Nat. Mem. B* 125, 31–40. doi: 10.2424/ASTSN.M.2017.28

Geraci, A., Amato, F., Di Noto, G., Bazan, G., and Schicchi, R. (2018). The wild taxa utilized as vegetables in Sicily (Italy): a traditional component of the Mediterranean diet. *J. Ethnobiol. Ethnomed.* 14:14. doi: 10.1186/s13002-018-0215-x

Ghirardini, M. P., Carli, M., Del Vecchio, N., Rovati, A., Cova, O., Valigi, F., et al. (2007). The importance of a taste. A comparative study on wild food plant consumption in twenty-one local communities in Italy. *J. Ethnobiol. Ethnomed.* 3:22. doi: 10.1186/1746-4269-3-22

Giardina, G., Raimondo, F. M., and Spadaro, V. (2007). A catalogue of plants growing in Sicily. *Bocconea* 20:582.

Graham, J., and Woodhead, M. (2011). "Rubus," in *Wild Crop Relatives: Genomic and Breeding Resources: Temperate Fruits*, ed. C. Kole (Berlin: Springer), 179–196.

Gristina, A. S., Fici, S., Siragusa, M., Fontana, I., Garf, G., and Carimi, F. (2014). Hybridization in *Capparis spinosa* L.: molecular and morphological evidence from a Mediterranean island complex. *Flora* 209, 733–741. doi: 10.1016/j.flora.2014.09.002

Grivetti, L. E., and Ogle, B. M. (2000). Value of traditional foods in meeting macro- and micronutrient needs: the wild plant connection. *Nutr. Res. Rev.* 13, 31–46. doi: 10.1079/095442200108728990

Grzebelus, D., Baranski, R., Spalik, K., Allender, C., and Simon, P. W. (2011). "Daucus," in *Wild Crop Relatives: Genomic and Breeding Resources: Vegetables*, ed. C. Kole (Berlin: Springer), 91–113.

Guarino, R. (2006). On the origin and evolution of the Mediterranean dry grasslands. *Ber. Reinhold Tüxen Ges.* 18, 195–206. doi: 10.1111/j.1365-294X.2008.03836.x

Guarino, R., Cutaia, F., Giacopelli, A. L., Menegoni, P., Pelagallo, F., Trotta, C., et al. (2015). Disintegration of Italian rural landscapes to international environmental agreements. *Int. Environ. Agreem.* 17, 161–172. doi: 10.1007/s10784-015-9310-9

Guarino, R., Giusso del Galdo, G., and Pignatti, S. (2005). The Mediterranean dwarf shrubs: origin and adaptive radiation. *Ann. Bot.* 5, 93–101.

Guarino, R., and Pasta, S. (2017). "Botanical Excursions in Central and Western Sicily," in *Proceedings of the Field Guide for the 60th IAVS Symposium* (Palermo: Palermo University Press).

Guarino, R., and Pasta, S. (2018). Sicily: the island that didn't know to be an archipelago. *Ber. Reinhold Tüxen Ges.* 30, 133–148.

Guarino, R., and Pignatti, S. (2010). Diversitas and biodiversity: the roots of a 21st century myth. *Rendic. Lincei. Sci. Fis. Nat.* 20, 351–357. doi: 10.1007/s12210-010-0104-2

Guarrera, M. (1994). *Il Patrimonio Etnobotanico del Lazio*. Roma: Lazio Regione, Univ.La Sapienza-Dep. Biologia vegetale,Tip. Tipar.

Guarrera, P. M., and Savo, V. (2016). Wild food plants used in traditional vegetable mixtures in Italy. *J. Ethnopharmacol.* 185, 202–234. doi: 10.1016/j.jep.2016.02.050

Guijt, I. (1998). "Valuing wild plants with economics and participatory methods: an overview of the Hidden Harvest methodology," in *Plants for Food and Medicine*, eds H. D. V. Prendergast, N. L. Etkin, D. R. Harris, and P. J. Houghton (Richmond: Royal Botanic Gardens, Kew), 223–235.

Hadjichambis, A. C. H., Paraskeva-Hadjichambi, D., Della, A., Giusti, M. E., De Pasquale, C., Lenzarini, C., et al. (2008). Wild and semi-domesticated food plant consumption in seven circum-Mediterranean areas. *Intern. J. Food Sci. Nutr.* 59, 383–414. doi: 10.1080/09637480701566495

Hajjar, R., and Hodgkin, T. (2007). The use of wild relatives in crop improvement: a survey of developments over the last 20 years. *Euphytica* 156, 1–13. doi: 10.1007/s10681-007-9363-0

Hammer, K., and Laghetti, G. (2006). *Le Piccole Isole Rurali Italiane*. Bari: Istituto di Genetica Vegetale C.N.R.

Hammer, K., and Perrino, P. (1995). Plant genetic resources in South Italy and Sicily – studies towards in situ and on farm conservation. *Pl. Gen. Res. Newsl.* 103, 19–23.

Heckler, S. (2012). *Landscape, Process and Power: Re-Evaluating Traditional Environmental Knowledge. Studies in Environmental Anthropology and Ethnobiology*. New York, NY: Berghahn Books.

Heywood, V. H., Casas, A., Ford-Lloyd, B., Kell, S., and Maxted, N. (2007). Conservation and sustainable use of crop wild relatives. *Agr. Ecosyst. Environ.* 121, 245–255. doi: 10.1016/j.agee.2006.12.014

Heywood, V. H., and Zohary, D. (1995). A catalogue of the wild relatives of cultivated plants native to Europe. An enumeration of the wild genetic resources of native European plants that are grown in Europe for food, forage, ornament, timber and other purposes. *Fl. Medit.* 5, 375–415.

Hummer, K. E., Bassil, N., and Njuguna, W. (2011). "Fragaria," in *Wild Crop Relatives: Genomic and Breeding Resources: Temperate Fruits*, ed. C. Kole (Berlin: Springer), 17–44.

Ignatov, A., and Bodishevskaya, A. (2011). "Malus," in *Wild Crop Relatives: Genomic and Breeding Resources: Temperate Fruits*, ed. C. Kole (Berlin: Springer), 45–64.

Iorizzo, M., Senalik, D. A., Ellison, S. L., Grzebelus, D., Cavagnaro, P. F., Allender, C., et al. (2013). Genetic structure and domestication of carrot (*Daucus carota* subsp. sativus) (Apiaceae). *Am. J. Bot.* 100, 930–938. doi: 10.3732/ajb.1300055

ISPRA (2012). *Carta Degli Habitat Alla Scala 1:50.000 Della Sicilia. Sistema Informativo di Carta della Natura*. Available online at: http://www.isprambiente.gov.it/it/servizi-per-lambiente/sistema-carta-della-natura/carta-della-natura-alla-scala-1-50.000/sicilia (accessed June 11, 2019).

Kanno, A., and Yokoyama, J. (2011). "Asparagus: vegetables," in *Wild Crop Relatives: Genomic and Breeding Resources*, ed. C. Kole (Berlin: Springer), 23–42.

Kareiva, P., Watts, S., McDonald, R., and Boucher, T. (2007). Domesticated nature: shaping landscapes and ecosystems for human welfare. *Science* 316, 1866–1869. doi: 10.1126/science.1140170

La Mantia, T., Carimi, F., Di Lorenzo, R., and Pasta, S. (2011). The agricultural heritage of Lampedusa (Pelagie Archipelago, South Italy) and its key role for cultivar and wildlife conservation. *Ital. J. Agron.* 6:e17. doi: 10.4081/ija.2011.e17

La Mantia, T., and Pasta, S. (2005). "The Sicilian phanerophytes: still a noteworthy patrimony, soon a lost resource?," in *Proceedings of the Monitoring and Indicators of Forest Biodiversity in Europe – from Ideas to Operationality. I.U.F.R.O. Conference, Florence, E.F.I. Proceedings n. 51*, ed. M. Marchetti (Florence: EFI), 515–526.

Leighton, R. (1999). *Sicily Before History: An Archaeological Survey from the Palaeolithic to the Iron Age*. London: Duckworth Publishers.

Lentini, F., and Mazzola, P. (1998). "Le piante utilizzate nella tradizione locale," in *Il Popolamento Vegetale Della Riserva Dello Zingaro (Sicilia)*, eds F. M. Raimondo and R. Schicchi (Palermo: Azienda Foreste Demaniali Regione Siciliana), 83–134.

Lentini, F., and Venza, F. (2007). Wild food plants of popular use in Sicily. *J. Ethnobiol. Ethnomed.* 3:15. doi: 10.1186/1746-4269-3-15

Leonti, M., Nebel, S., Rivera, D., and Heinrich, M. (2006). Wild gathered food plants in the European Mediterranean: a comparative analysis. *Econ. Bot.* 60, 130–142. doi: 10.1663/0013-0001(2006)60[130:wgfpit]2.0.co;2

Licata, M., Tuttolomondo, T., Leto, C., Virga, G., Bonsangue, G., Cammalleri, I., et al. (2016). A survey of wild species for food use in Sicily (Italy): results of a 3-year study in four regional parks. *J. Ethnobiol. Ethnomed.* 12:12. doi: 10.1186/s13002-015-0074-7

Łuczaj, Ł., Pieroni, A., Tardío, J., Pardo de Santayana, M., Sõukand, R., Svanberg, I., et al. (2012). Wild food plant use in 21st century Europe: the disappearance of old traditions and the search for new cuisines involving wild edibles. *Acta Soc. Bot. Poloniae* 81, 359–370. doi: 10.5586/asbp.2012.031

Mannino, M. A., Catalano, G., Talamo, S., Mannino, G., Di Salvo, R., Catalano, G., et al. (2012). Origin and diet of the prehistoric hunter-gatherers on the Mediterranean Island of Favignana (Egadi Islands, Sicily). *PLoS One* 7:e49802. doi: 10.1371/journal.pone.0049802

Marino, P., Schicchi, R., Barone, E., Raimondo, F. M., and Domina, G. (2013). First results on the phenotypic analysis of wild and cultivated species of *Pyrus* in Sicily. *Flora Medit.* 23, 237–243. doi: 10.7320/FlMedit23.237

Martin-Robles, N., Morente-López, J., Freschet, G. T., Poorter, H., Roumet, C., and Milla, R. (2019). Root traits of herbaceous crops: pre-adaptation to cultivation or evolution under domestication? *Funct. Ecol.* 33, 273–285. doi: 10.1111/1365-2435.13231

Maxted, N., Ford-Lloyd, B. V., Jury, S., Kell, S., and Scholten, M. (2006). Towards a definition of a crop wild relative. *Biodivers. Conserv.* 15, 2673–2685. doi: 10.1007/s10531-005-5409-6

Maxted, N., Kell, S., Ford-Lloyd, B., Dulloo, E., and Toledo, Á (2012). Toward the systematic conservation of global crop wild relative diversity. *Crop Sci.* 52, 774–785. doi: 10.2135/cropsci2011.08.0415

Maxted, N., Kell, S., Toledo, Á, Dulloo, E., Heywood, V., Hodgkin, T., et al. (2010). A global approach to crop wild relative conservation: securing the gene pool for food and agriculture. *Kew Bull.* 65, 561–576. doi: 10.1007/s12225-011-9253-4

Maxted, N., Scholten, M., Codd, R., and Ford-Lloyd, B. (2007). Creation and use of a national inventory of crop wild relatives. *Biol. Cons.* 140, 142–159. doi: 10.1016/j.biocon.2007.08.006

Mazzola, P., Nibali Lupica, S., and Bartolotta, L. (2015). Contributo alla conoscenza delle piante alimurgiche del Parco dei Nebrodi: le specie utilizzate nella tradizione popolare di Caronia (Messina). *Quad. Bot. Ambientale Appl.* 26, 3–10.

McGrath, J. M., Panella, L., and Frese, L. (2011). "Beta," in *Wild Crop Relatives: Genomic and Breeding Resources: Industrial Crops*, ed. C. Kole (Berlin: Springer), 1–28.

Médail, F., and Quézel, P. (1999). Biodiversity hotspots in the Mediterranean Basin: setting global conservation priorities. *Conserv. Biol.* 13, 1510–1513. doi: 10.1046/j.1523-1739.1999.98467.x

Mercati, F., Fontana, I., Gristina, A. S., Martorana, A., El Nagar, M., De Michele, R., et al. (2019). Transcriptome analysis and codominant markers development in caper, a drought tolerant orphan crop with medicinal value. *Sci. Rep.* 9:10411. doi: 10.1038/s41598-019-46613-x

Mercuri, A. M., Florenzano, A., Burjachs, F., Giardini, M., Kouli, K., Masi, A., et al. (2019). From influence to impact: the multifunctional land-use in Mediterranean prehistory emerging from palynology of archaeological sites (8.0-2.8 ka BP). *Holocene* 29, 830–846. doi: 10.1177/0959683619826631

Milla, R., Bastida, J. M., Turcotte, M. M., Jones, G., Violle, C., Osborne, C. P., et al. (2018). Phylogenetic patterns and phenotypic profiles of the species of plants and mammals farmed for food. *Nat. Ecol. Evol.* 2, 1808–1817. doi: 10.1038/s41559-018-0690-4

Milla, R., García-Palacios, P., and Matesanz, S. (2017). Looking at past domestication to secure ecosystem services of future croplands. *J. Ecol.* 105, 885–889. doi: 10.1111/1365-2745.12790

Milla, R., Osborne, C. P., Turcotte, M. M., and Violle, C. (2015). Plant domestication through an ecological lens. *Trends Ecol. Evol.* 30, 463–469. doi: 10.1016/j.tree.2015.06.006

Monello, N. (2018). *Antologia. Monti Iblei, Avola, Sant'Elia*. Sicily: Sebastiano Monieri Editore, 460.

Morreale, F. (2018). *Piante Spontanee Alimentari in Sicilia. Guida di fitoalimurgia*. Siracusa: Casa Editrice Natura Sicula.

Mucina, L., Buültmann, H., Dierssen, K., Theurillat, J.-P., Raus, T., Andraž Èarni, A., et al. (2016). Vegetation of Europe: hierarchical floristic classification system of vascular plant, bryophyte, lichen, and algal communities. *Appl. Veg. Sci.* 19(Suppl. 1), 3–264. doi: 10.1111/avsc.12257

Noti, R., van Leeuwen, J., Colombaroli, D., Vescovi, E., Pasta, S., La Mantia, T., et al. (2009). Mid- and Late-Holocene vegetation and fire history of Biviere di Gela, a coastal lake in southern Sicily. *Ital. Veg. Hist. Archaeobot.* 18, 371–387. doi: 10.1007/s00334-009-0211-0

Odeny, D. A., and Narina, S. S. (2011). "Allium," in *Wild Crop Relatives: Genomic and Breeding Resources: Vegetables*, ed. C. Kole (Berlin: Springer), 1–10.

Pasta, S., Garfì, G., La Bella, F., Rühl, J., and Carimi, F. (2011). "An overview on the human exploitation of Sicilian native edible plants," in *Wild Plants: Identification, Uses and Conservation*, ed. E. R. Davis (Hauppauge, NY: Nova Science Publishers Press), 195–268.

Pieroni, A., and Price, L. (2005). *Eating and Healing: Traditional Food as Medicine*. New York, NY: Food Products Press.

Pignatti, S., Guarino, R., and La Rosa, M. (2017–2019). *Flora d'Italia*, 2nd Edn, Vol. 1-4. Bologna: Edagricole.

Pignone, D., and Gómez-Campo, C. (2011). "Eruca," in *Wild Crop Relatives: Genomic and Breeding Resources: Oilseeds*, ed. C. Kole (Berlin: Springer), 149–160.

Pignone, D., and Martínez-Laborde, J. B. (2011). "Diplotaxis," in *Wild Crop Relatives: Genomic and Breeding Resources: Oilseeds*, ed. C. Kole (Berlin: Springer), 137–147. doi: 10.1007/978-3-642-14871-2_7

Plieninger, T., Höchtl, F., and Spek, T. (2006). Traditional land-use and nature conservation in European rural landscapes. *Environ. Sci. Policy* 9, 317–321. doi: 10.1016/j.envsci.2006.03.001

Portis, E., Mauromicale, G., Bardi, L., Mauro, M., and Lanteri, S. (2005). Population structure and genetic variation in autochthnous globe artichoke germplasm from Sicily Island. *Plant Sci.* 168, 1591–1598. doi: 10.1016/j.plantsci.2005.02.009

Potter, D. (2011). "Prunus," in *Wild Crop Relatives: Genomic and Breeding Resources: Oilseeds: Temperate Fruits*, ed. C. Kole (Berlin: Springer), 129–145.

Provitina, F. M. (1991). *Flora sicula. Dizionario trilingue illustrato*. Palermo: Kefagrafica, 415.

Quave, C. L., and Saitta, A. (2016). Forty-five years later: the shifting dynamic of traditional ecological knowledge on Pantelleria Island. *Ital. Econ. Bot.* 70, 380–393. doi: 10.1007/s12231-016-9363-x

Ranfa, A., Maurizi, A., Romano, B., and Bodesmo, M. (2014). The importance of traditional uses and nutraceutical aspects of some edible wild plants in human nutrition: the case of Umbria (central Italy). *Plant Biosyst.* 148, 297–306. doi: 10.1080/11263504.2013.770805

Reich, P. B. (2014). The world-wide "fast-slow" plant economics spectrum: a traits manifesto. *J. Ecol.* 102, 275–301. doi: 10.1111/1365-2745.12211

Renna, M. (2017). Wild edible plants as a source of mineral elements in the daily diet. *Progr. Nutr.* 19, 219–222. doi: 10.23751/pn.v19i2.5662

Rivera, D., Obón, C., Heinrich, M., Inocencio, C., Verde, A., and Fajardo, J. (2006). "Gathered mediterranean food plants. Ethnobotanical investigations and historical development," in *Local Mediterranean Food Plants and*

Nutraceuticals. Forum Nutr, eds M. Heinrich, W. E. Müller, and C. Galli (Basel: Karger), 18–74. doi: 10.1159/000095207

Sadori, L., Zanchetta, G., and Giardini, M. (2008). Last glacial to Holocene palaeoenvironmental evolution at Lago di Pergusa (Sicily, southern Italy) as inferred by pollen, microcharcoal, and stable isotopes. *Quat. Internat.* 181, 4–14. doi: 10.1016/j.quaint.2007.02.024

Schaffer, S., Schmitt-Schillig, S., Müller, W. E., and Eckert, G. P. (2005). Antioxidant properties of Mediterranean food plant extracts: geographical differences. *J. Physiol. Pharmacol.* 56, 115–124.

Schicchi, R., and Geraci, A. (2016). *Verdure Spontanee di Sicilia. Guida al Riconoscimento, alla Raccolta e alla Preparazione*. Palermo: IDIMED.

Schicchi, R., Geraci, A., and Marino, P. (2008). Genetic diversity on *Pyrus* L. (Rosaceae) in Sicily. *Bocconea* 23, 141–146.

Signorello, P. (2015). *Piante Spontanee Siciliane Mangerecce E Velenose*. Acireale: Stampa Massimino, 191.

Simopoulos, A. P. (2004). Omega-3 fatty acids and antioxidants in edible wild plants. *Biol. Res.* 37, 263–277. doi: 10.4067/s0716-97602004000200013

Snir, A., Nadel, D., Groman-Yaroslavski, I., Melamed, Y., Sternberg, M., Bar-Yosef, O., et al. (2015). The origin of cultivation and proto-weeds, long before Neolithic farming. *PLoS One* 10:e0131422. doi: 10.1371/journal.pone.0131422

Sottile, R., and Genchi, M. (2010). *Lessico Della Cultura Dialettale Delle Madonie. 1. L'alimentazione. Centro di Studi Filologici e Linguistici Siciliani-Dipartimento di Scienze Filologiche e Linguistiche*. Palermo: L'ALS per la scuola e il territorio.

Stika, H.-P., Heiss, A. G., and Zach, B. (2008). Plant remains from the early Iron Age in western Sicily: differences in subsistence strategies of Greek and Elymian sites. *Veg. Hist. Archaeobot.* 17, 139–148. doi: 10.1007/s00334-008-0171-9

ter Braak, C. J. F., and Smilauer, P. (2012). *CANOCO Reference Manual and User's Guide: Software for Ordination (Version 5.0)*. Ithaca, NY: Microcomputer Power.

Tinner, W., van Leeuwen, J. F. N., Colombaroli, D., Vescovi, E., van der Knaap, P. W. O., Henne, P. D., et al. (2009). Holocene environmental and climatic changes at Gorgo Basso, a coastal lake in southern Sicily. *Italy. Quat. Sci. Rev.* 28, 1498–1510. doi: 10.1016/j.quascirev.2009.02.001

Tinner, W., Vescovi, E., van Leeuwen, J. F. N., Colombaroli, D., Henne, P. D., Kaltenrieder, P., et al. (2016). Holocene vegetation and fire history of the mountains of Northern Sicily (Italy). *Veg. Hist. Archeobot.* 25, 499–519. doi: 10.1007/s00334-016-0569-8

Triboullois, H., Fort, F., Cruz, P., Charles, R., Flores, O., Garnier, E., et al. (2015). A functional characterisation of a wide range of cover crop species: growth and nitrogen acquisition rates, leaf traits and ecological strategies. *PLoS One* 10:e0122156. doi: 10.1371/journal.pone.0122156

Trichopoulou, A., Vasilopoulou, E., Holiman, P., Chamalides, C., Foufa, E., Kaloudis, T., et al. (2000). Nutritional composition and flavonoid content of edible wild greens and green pies: a potential rich sources of antioxidant nutrients in the Mediterranean diet. *Food Chem.* 70, 319–323. doi: 10.1016/S0308-8146(00)00091-1

Tuttolomondo, T., Licata, M., Leto, C., Bonsangue, G., Gargano, M. L., Venturella, G., et al. (2014a). Popular uses of wild plant species for medicinal purposes in the Nebrodi regional park (North-Eastern Sicily, Italy). *J. Ethnopharmacol.* 157, 21–37. doi: 10.1016/j.jep.2014.08.039

Tuttolomondo, T., Licata, M., Leto, C., Gargano, M. L., Venturella, G., and La Bella, S. (2014b). Plant genetic resources and traditional knowledge on medicinal use of wild shrub and herbaceous plant species in the Etna regional park (Eastern Sicily, Italy). *J. Ethnopharmacol.* 155, 1362–1381. doi: 10.1016/j.jep.2014.07.043

Valdés, B., Heywood, V. H., Raimondo, F. M., and Zohary, D. (1997). Proceedings of the workshops on 'Conservation of the wild relatives of European cultivated plants'. *Bocconea* 7, 1-479.

Vincent, H., Wiersema, J., Kell, S., Fielder, H., Dobbie, S., Castañeda-Álvarez, N. P., et al. (2013). A prioritized crop wild relative inventory to help underpin global food security. *Biol. Cons.* 167, 265–275. doi: 10.1016/j.biocon.2013.08.011

Viruel, J., Galliot, N., Pironon, S., Feliner, G., Suc, J.-P., Lakhal, M. F., et al. (2019). Astrong east-west Mediterranean divergence supports a new phylogeographic history of the carob tree (*Ceratonia siliqua*, Leguminosae) and multiple domestications from native populations. *J. Biogeogr.* 47, 460–471. doi: 10.1111/jbi.13726

Vitasović Kosić, I., Juračak, J., and Łuczaj, Ł (2017). Using Ellenberg-Pignatti values to estimate habitat preferences of wild food and medicinal plants: an example from northeastern Istria (Croatia). *J. Ethnobiol. Ethnomed.* 13, 2–19. doi: 10.1186/s13002-017-0159-6

Willett, W. C. (2006). The Mediterranean diet: science and practice. *Public Health Nutr.* 9, 105–110. doi: 10.1079/phn2005931

Willis, K. J., Araújo, M. B., Bennett, K. D., Figueroa-Rangel, B., Froyd, C. A., and Myers, N. (2007). How can a knowledge of the past help to conserve the future? Biodiversity conservation and the relevance of long-term ecological studies. *Philos. Trans. R. Soc. B* 362, 175–187. doi: 10.1098/rstb.2006.1977

Willis, K. J., and Birks, H. J. B. (2006). What is natural? The need for a long-term perspective in biodiversity conservation. *Science* 314, 1261–1265. doi: 10.1126/science.1122667

Xu, L., Freitas, S. M. A., Yu, F.-H., Dong, M., Anten, N. P. R., and Werger, M. J. A. (2013). Effects of trampling on morphological and mechanical traits of dryland shrub species do not depend on water availability. *PLoS One* 8:e53021. doi: 10.1371/journal.pone.0053021

Conflict of Interest: 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 © 2020 Pasta, La Rosa, Garfì, Marcenò, Gristina, Carimi and Guarino. 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) and the copyright owner(s) 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 Plant Science

ORIGINAL RESEARCH
published: 19 May 2020
doi: 10.3389/fpls.2020.00616

Lost and Found: *Coffea stenophylla* and *C. affinis*, the Forgotten Coffee Crop Species of West Africa

Aaron P. Davis[1], Roberta Gargiulo[1], Michael F. Fay[1], Daniel Sarmu[2] and Jeremy Haggar[3]**

[1] Royal Botanic Gardens, Kew, Richmond, United Kingdom, [2] Welthungerhilfe, Freetown, Sierra Leone, [3] Department of Agriculture, Health and Environment, Faculty of Engineering and Science, Natural Resources Institute, University of Greenwich, Medway, United Kingdom

Coffea arabica (Arabica) and *C. canephora* (robusta) almost entirely dominate global coffee production. Various challenges at the production (farm) level, including the increasing prevalence and severity of disease and pests and climate change, indicate that the coffee crop portfolio needs to be substantially diversified in order to ensure resilience and sustainability. In this study, we use a multidisciplinary approach (herbarium and literature review, fieldwork and DNA sequencing) to elucidate the identity, whereabouts, and potential attributes, of two poorly known coffee crop species: *C. affinis* and *C. stenophylla*. We show that despite widespread (albeit small-scale) use as a coffee crop species across Upper West Africa and further afield more than 100 years ago, these species are now extremely rare in the wild and are not being farmed. Fieldwork enabled us to rediscover *C. stenophylla* in Sierra Leone, which previously had not been recorded in the wild there since 1954. We confirm that *C. stenophylla* is an indigenous species in Guinea, Sierra Leone, and Ivory Coast. *Coffea affinis* was discovered in the wild in Sierra Leone for the first time, having previously been found only in Guinea and Ivory Coast. Prior to our rediscovery, *C. affinis* was last seen in the wild in 1941, although sampling of an unidentified herbarium specimen reveals that it was collected in Guinea-Conakry in 2015. DNA sequencing using plastid and ITS markers was used to: (1) confirm the identity of museum and field collected samples of *C. stenophylla*; (2) identify new accessions of *C. affinis*; (3) refute hybrid status for *C. affinis*; (4) identify accessions confused with *C. affinis*; (5) show that *C. affinis* and *C. stenophylla* are closely related, and possibly a single species; (6) substantiate the hybrid *C. stenophylla* × *C. liberica*; (7) demonstrate the use of plastid and nuclear markers as a simple means of identifying F1 and early-generation interspecific hybrids in *Coffea*; (8) infer that *C. liberica* is not monophyletic; and (9) show that hybridization is possible across all the major groups of key Africa *Coffea* species (Coffee Crop Wild Relative Priority Groups I and II). *Coffea affinis* and *C. stenophylla* may possess useful traits for coffee crop plant development, including taste differentiation, disease resistance, and climate resilience. These attributes would be best accessed via breeding programs, although the species may have niche-market potential via minimal domestication.

Keywords: agronomy, climate change, coffee, West Africa, crop wild relatives (CWRs), DNA, Sierra Leone, speciality coffee

OPEN ACCESS

Edited by:
Petr Smýkal, Palackı University Olomouc, Czechia

Reviewed by:
Alan C. Andrade, Brazilian Agricultural Research Corporation (EMBRAPA), Brazil
Benoit Bertrand, Centre de Coopération Internationale en Recherche Agronomique pour le Développement (CIRAD), France

****Correspondence:***
Aaron P. Davis
a.davis@kew.org
Jeremy Haggar
J.P.Haggar@greenwich.ac.uk

Specialty section:
This article was submitted to Plant Breeding, a section of the journal Frontiers in Plant Science

Received: *19 December 2019*
Accepted: *21 April 2020*
Published: *19 May 2020*

Citation:
Davis AP, Gargiulo R, Fay MF, Sarmu D and Haggar J (2020) Lost and Found: Coffea stenophylla and C. affinis, the Forgotten Coffee Crop Species of West Africa. Front. Plant Sci. 11:616. doi: 10.3389/fpls.2020.00616

INTRODUCTION

Coffee is a globally significant crop that supports a multibillion-dollar global industry (International Coffee Organization (ICO), 2019), over a lengthy value chain from farmer to consumer. Coffee farming alone involves the farming activities of around 100 million people worldwide (Vega et al., 2003). Two species dominate global coffee production: Arabica (*Coffea arabica*) and robusta (*C. canephora*), providing c. 60% and c. 40% of traded coffee, respectively (International Coffee Organization (ICO), 2019). Liberica coffee is cultivated worldwide in small quantities, and is insignificant in terms of global trade, although production in the Philippines and Malaysia can be substantial. Aside from *C. arabica*, *C. canephora*, and *C. liberica*, there are another 121 coffee species known to science (Davis et al., 2006, 2011, 2019). Some of these are used to make the beverage coffee, such as *C. congensis*, *C. eugenioides*, and *C. racemosa*, some have been used in breeding programs, and others have been used as high performing pest and diseases resistant rootstocks (Davis et al., 2019). Many more are used on a small, local scale, or are harvested directly from the wild, in Africa, Madagascar, and Asia. In previous centuries, and particularly at the end of the 1800s and early 1900s, there was considerable interest in, and use of, a range of beverage producing coffee species, more than there is today (Davis et al., 2019). It is also of note that since those times, a substantial proportion of the world's coffee species diversity has been discovered and named by science, particularly from the 1960s onward (Bridson, 1988; Davis et al., 2006; Davis and Rakotonasolo, 2009). The decline in the interest in these 'other' coffee species has been largely due to the overwhelming success of robusta coffee, which was itself transformed from a wild plant, and minor African crop species, to a major global commodity in around 150 years (Davis et al., 2019). Robusta gained market share against Arabica from the early 1900s onward due to its resistance to coffee leaf rust (CLR; *Hemileia vastatrix*) (Wrigley, 1988), a broader agroecological envelope (Davis et al., 2006), higher productivity (Wellman, 1961; Wrigley, 1988), lower purchase price (International Coffee Organization (ICO), 2019), and other specific attributes (Davis et al., 2019). Recently, however, there has been renewed interest in underutilized, forgotten, and little known coffee species, both cultivated and wild, due to their potential to counter specific pests and diseases, and provide resilience in an era of accelerated climate change (Davis et al., 2019). There is also an increasing curiosity in lesser known coffee species from the specialty coffee sector, in its quest to discover new and differentiated sensory experiences in coffee.

Among those of particular interest are two West African species: *C. stenophylla* and *C. affinis*, mainly due to historical reports of a superior taste, particularly for *C. stenophylla* (Cheney, 1925) but also *C. affinis* (De Wildeman, 1904). Given that these two species occur in Upper West Africa at relatively low elevations (see below) there may also be the potential for climate resilience. Both species fall within Coffee Crop Wild Relative Priority Group II, which includes species closely related to the main crop species, for which gene transfer to the crop is proven or assumed (with low to high post-crossing fertility rates) (Davis et al., 2019). Priority Group II includes all African species, apart from the main coffee crop species and their progenitors (*C. arabica*, *C. canephora*, *C. liberica*, and *C. eugenioides*: Priority Group I) and African species of Priority Group III. Priority Group III includes all the short-styled *Coffea* species (previously assigned to the genus *Psilanthus*) from Africa, Asia and Australasia, and all Madagascan species and Mascarene species (Davis et al., 2019).

Our recent knowledge of *C. stenophylla* and *C. affinis* is principally limited to germplasm surveys. *Coffea stenophylla* is recorded as a living plant in several (*ex situ*) coffee research collections (Anthony et al., 2007; Engelmann et al., 2007; Bramel et al., 2017); *C. affinis* is included in the most recent of these reviews (Bramel et al., 2017) but only as an entry based on our knowledge of accepted coffee species (Govaerts et al., 2019). Contemporary evaluations of coffee species diversity (Davis et al., 2006, 2011, 2019; Maurin et al., 2007; Hamon et al., 2017) clearly show that our knowledge of *C. stenophylla* and *C. affinis* is inadequate. Initial review of literature for *C. affinis* showed almost no extra knowledge of this species has been gained since 1937 (Portères, 1937a), with the exception of work in Ivory Coast and Guinea in the 1980s (Berthaud, 1983, 1986; Le Pierrès et al., 1989). It is imperative that we improve our knowledge of these two species, both in cultivation (including any commercial production) and in the wild.

In this study, our main objectives were to elucidate: the current cultivated and wild status of *C. stenophylla* and *C. affinis;* the taxonomic identity and systematic position of the poorly known *C. affinis*; and to assemble available information on crop plant attributes. To achieve these objectives we undertook: (1) a literature review; (2) a survey of herbarium and economic botany collections; (3) field surveys in Sierra Leone, visiting farms, research stations and natural forest locations; and (4) DNA sequencing of recently collected material, historical samples (herbarium and economic botany collection samples), known interspecies hybrids, and their analysis incorporating a reference set of previously published *Coffea* sequences.

MATERIALS AND METHODS

Literature Review

We examined all key literature pertaining to *C. affinis* and *C. stenophylla*. Knowledge of *ex situ* cultivation in research collections was gleaned from published works (Anthony, 1982, 1992; Anthony et al., 2007; Engelmann et al., 2007; Bramel et al., 2017; Davis et al., 2019), supported by personal observations (A. Davis, J. Haggar) and personal communication.

Review of Herbarium Collections and Economic Botany Collections

Herbarium specimens are well suited to this type of study because they are verifiable in space (location), time (date) and form (species identity), and are often accompanied by additional information on the herbarium label (e.g., ecology, elevation, geology, and uses). We consulted herbarium specimen records from nine herbaria (BM, BR, K, MO, P, UPS, WAG) including

those in Sierra Leone (SL, FBC). Herbarium codes follow standard abbreviations (Holmgren et al., 1990; Thiers, 2019). The specimen data was disaggregated into unique records and duplicate specimens. Unique records comprise the combination of collector's name and number (e.g., *Chillou 2381*) or collector's name and date (e.g., Cope *s.n.*, 7 iii 1912); *s.n.* is an abbreviation for *sine numero*, and lowercase Roman numerals to represent the month). Duplicate specimens possess the same unique identifier, i.e., they are from the same plant or possibly nearby individuals, but are found on separate herbarium sheets; these may either be found in the same herbarium or across two or more herbaria.

Fieldwork in Sierra Leone

During 2014 and 2016 we made a request for samples of *C. affinis* and *C. stenophylla* and any atypical coffee morphotypes, from NGOs and farmer associations representing 10,000 coffee farmers across Kenema, Kailahun, and Kono Districts, which represent the major coffee producing region of Sierra Leone. We also visited the Sierra Leone Agricultural Research Institute (SLARI) research collection at Pendembu, Kailahun District (**Table 1**), to sample putative examples of *C. stenophylla* and *C. affinis*. In addition, 50 A4 posters showing the most obvious morphological differences (leaf shape and size) between the two cultivated coffee species, robusta coffee (*C. canephora*) and Liberica (*C. liberica*) and *C. stenophylla* were printed and distributed to district agriculture offices with coffee farming communities in southern Sierra Leone, between Freetown and Kenema. The aim was to provide an additional means of identifying farms that might be cultivating *C. stenophylla* or *C. affinis*. Visits to sites where *C. stenophylla* had been recorded in cultivation in northern Sierra Leone (based on the herbarium survey) were visited in 2017. In December 2018, we followed up on the poster survey, by visiting five farms that had stated cultivation of *C. stenophylla*. On the same trip, we visited the last known (1954) forest sites for *C. stenophylla* in the Kasewe Hills (Southern Province), and several possible locations: around Freetown (Western Area), near Moyamba Junction (Southern Province), and the forest area of Kambui Hills, adjacent to Kenema (Eastern Province). Follow up visits to the Kambui Hills were made throughout 2019 and in early 2020.

Assembly of DNA Reference Collection

The most taxonomically comprehensive DNA dataset for wild coffee species is for plastid (*trnL–F* intron, *trnL–F* intergenic spacer (IGS), *rpl16* intron and *accD–psaI* IGS) and the internal transcribed spacer (ITS) region of nuclear rDNA (ITS 1/5.8S/ITS 2) (Maurin et al., 2007; Davis et al., 2011). These markers have the ability to distinguish between African coffee species, and identify recently formed hybrids via differential inheritance of plastid and nuclear genomes (Maurin et al., 2007). Thirty-one accessions (**Tables 1**, **2**) were sequenced with the four markers: 14 collected by us in Sierra Leone; nine reference samples from the museum collections of RBG Kew (K), including three for *C. stenophylla*, four for *C. affinis* (including two farmed accessions), and four other coffee species; five samples associated with the production of the artificial hybrid *C. arabica* × *C. racemosa* (Medina Filho et al., 1977a,b); and one unpublished sequence of *C. zanguebariae* (**Table 2**). The five museum collections of *C. stenophylla* and *C. affinis* were selected to represent authentic material, i.e., that being cultivated at the end of the 19th and beginning of the 20th centuries (respectively) in Sierra Leone. The published reference sequence dataset (Maurin et al., 2007; Davis et al., 2011) included a single verified example of *C. stenophylla*. Two known interspecific hybrids, as identified in a previous study using the same markers (Maurin et al., 2007), were included in the sampling: *C. arabica* (*C. canephora* × *C. eugenioides*) and *C. liberica* × *C. eugenioides* [originally accessioned (Maurin et al., 2007) as *C. heterocalyx*]. Initial analyses were conducted using the study species (see above) and a global data set of African, Madagascar, Mascarene, and Asian coffee species (Maurin et al., 2007; Davis et al., 2011). Following this analysis and confirming general placement of accessions, this was reduced to African taxa, excluding short-styled *Coffea* species (former *Psilanthus*), equating to Coffee Crop Wild Relative (Priority) Groups I and II (Davis et al., 2019).

DNA Extraction, Sequencing, and Data Analysis

Total DNA was extracted from silica dried leaves, fresh seeds, and seeds from herbarium specimens and other archival material (**Tables 1**, **2**) using a modified CTAB method (Doyle and Doyle, 1987) and purified using the QIAquick PCR purification kit (QIAGEN). Genetic variation among the accessions was assessed by employing four regions: nuclear internal transcribed spacers (ITS1 and ITS2), plastid *trnL–trnF* (*trnL* intron and *trnL–trnF* intergenic spacer), *rpl16* intron and *accD–psaI* intergenic spacer. Amplifications were carried out following the protocol of Maurin et al. (2007); PCR products were purified using QIAquick PCR purification kit (QIAGEN) and sequenced following the methods employed by Maurin et al. (2007). Capillary electrophoresis was conducted on an ABI3730 DNA Analyzer (Applied Biosystems). Sequencing results were inspected in GENEIOUS v. 8.1.7 (Kearse et al., 2012). Newly sequenced accessions and unpublished sequences held at RBG Kew were referenced against GenBank accessions of *Coffea* species (Maurin et al., 2007; Davis et al., 2011). The sequences were aligned using MUSCLE (Edgar, 2004), as implemented in GENEIOUS. Gaps were treated as missing data and ambiguities were scored with IUPAC ambiguity codes. The model of character evolution was assessed in jModelTest v. 2.1.10 (Posada, 2008). Relationships among the taxa were reconstructed in MrBayes v. 3.2.7a (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003) as implemented on the CIPRES Scientific Gateway v3.3; *C. rhamnifolia* was used as the outgroup. Analyses were conducted separately for the ITS and the plastid DNA datasets and regions difficult to align were excluded from the phylogenetic analysis. We also conducted a separate analysis on the concatenated ITS/plastid matrix for the Upper Guinea (UG) clade (including *C. togoensis*, *C. affinis*, and C. *stenophylla*), with *C. canephora* and *C. liberica* as outgroups. MCMC sampling was performed with two runs and four chains for 2×10^7 generations, with a sampling

TABLE 1 | List of material examined, with origin, source of material, and (DNA) identification.

Accession of *Coffea* as received	Origin	Material	Name on DNA trees, and final identification
Sierra Leone cultivated accessions			
C. ?*affinis* [1a] (purple fruits, and purple tinge to stem and leaves)	Sierra Leone. Kangama, Gorama, Kono.	Leaf	*C. canephora* (1) SL cult.
C. sp. [2-16 Office]	Sierra Leone, Kenema	Leaf	*C. canephora* (2) SL cult.
C. sp. [1-16 Office]	Sierra Leone, Kenema	Leaf	*C. canephora* (3) SL cult.
C. liberica (leaves, 30 cm long, and almost as broad) [12a]	Sierra Leone, CEPAH, Kono	Leaf	*C. liberica* (1) SL cult.
C. sp. [14-16, Site 1]	Sierra Leone, East Fiama, Kono	Leaf	*C. liberica* (2) SL cult.
C. sp. [16-16 Site 3]	Sierra Leone, Fiama, Kono	Leaf	*C. liberica* (3) SL cult.
C. sp. [18-16, Site 1]	Sierra Leone, Lei Chiefdom, Kono	Leaf	*C. liberica* (4) SL cult.
C. arabica Pendembu, SLARI collection [10a]	Sierra Leone, Pendembu, Kailahun, SLARI collection	Leaf	*C. stenophylla* × *C. liberica* (1) SL cult.
C. affinis [8]	Sierra Leone, Pendembu, Kailahun, SLARI collection	Leaf	*C. stenophylla* × *C. liberica* (2) SL cult.
C. affinis [9]	Sierra Leone, Pendembu, Kailahun, SLARI collection	Leaf	*C. stenophylla* × *C. liberica* (3) SL cult.
Sierra Leone wild accessions			
C. affinis	Sierra Leone, Kambui Hills, 2020, *Sarmus.n.* (K)	Seed	*C. affinis* (1) SL
C. affinis	Sierra Leone, Kambui Hills, 2020, *Sarmus.n.* (K)	Leaf	*C. affinis* (2) SL
C. stenophylla	Sierra Leone, Kasewe Hills	Leaf	*C. stenophylla* (5) SL
C. stenophylla	Sierra Leone, Kambui Hills	Leaf	*C. stenophylla* (6) SL
Kew (K) accessions (museum)			
C. sp. unknown	Guinea-Conakry, Coyah Prefecture, Saliya, 2015, *Couch 757*(K)		*C. affinis* (3) Guinea
C. liberica	Uganda, cultivated.	Seed	*C. liberica* (5) Uganda cult.
C. affinis	Sierra Leone, Plantation, Sierra Leone, 1902, *Cope* s.n. (K)	Seed	*C.* sp. (1) SL cult.
C. affinis	Sierra Leone, Plantation, Sierra Leone, 1902, *Cope* s.n. (K)	Seed	*C.* sp. (2) SL cult.
C. stenophylla [1]	Sierra Leone, Botanical Station, 15 Jan 1896	Seed	*C. stenophylla* (1) SL cult.
C. stenophylla [2]	Sierra Leone, 1856 (Highland coffee). No. 4	Seed	*C. stenophylla* (2) SL cult.
C. stenophylla [4]	Sierra Leone, cult. Brazil	Seed	*C. stenophylla* (7) Brazil cult.
C. stenophylla [3]	Sierra Leone, cult. lowland grounds, 1856, No. 3	Seed	No DNA data
Brazil Accessions			
C. arabica ('Typica')	Brazil (cultivated), originally from Ethiopia	Leaf	*C. arabica* (2) cult.
C. arabica × *C. racemosa* [Clone 3]	Brazil (cultivated man made hybrid)	Leaf	*C. arabica* × *C. racemosa* (1) Brazil cult.
C. arabica × *C. racemosa* [Clone 12]	Brazil (cultivated, artificial hybrid)	Leaf	*C. arabica* × *C. racemosa* (2) Brazil cult.
C. arabica × *C. racemosa* × *C. arabica* [IAC1196]	Brazil (cultivated, artificial hybrid)	Leaf	*C. racemosa* × *C. arabica* (backcross) Brazil cult.
C. racemosa (putative hybrid parent)	Brazil (cultivated), originally from East Africa	Leaf	*C. racemosa* (1) Brazil cult.
C. stenophylla	Brazil (cultivated), originally from Upper West Africa	Lead	*C. stenophylla* (7) Brazil cult.
Other material			
C. liberica	DR Congo ex Belgian Botanic Gardens (BR) # 19370045 (BR)	Leaf	*C. liberica* (9) DRC

*The taxa in the furthest right-hand column are those used on the phylogenetic trees (**Figures 4, 5**); to find the accession identifier (as received and as discussed in the text) use left-hand column. For previously sequenced accessions included in **Figures 4, 5**, see **Table 2**). Seed, single seed from herbarium sample. Leaf, leaf sample in silica gel.*

frequency of 1,000 and a relative burn-in of 25%; specified model of character evolution was GTR+G. Convergence was visually assessed with Tracer v. 1.6 (Rambaut et al., 2013), by combining trace files to confirm mixing and high effective sampling size (EES). Maximum clade credibility trees were drawn in FigTree v. 1.4.4 (Rambaut, 2018). Clade and

TABLE 2 | List of sequence accession data.

Coffea species and accession identifier	Source	*ITS*	*trnL–trnF*	*rpl16*	*accD–psaI*
C. affinis (1) SL	Sierra Leone, Kambui Hills	**MT250043**	**MT274308**	**MT274311**	–
C. affinis (2) SL	Sierra Leone, Kambui Hills	**MT250044**	**MT274309**	**MT274312**	**MT274306**
C. affinis (3) Guinea	Guinea-Conakry, Saliya, 2015, *Couch 757* (K)	**MT250045**	**MT274310**	**MT274313**	**MT274307**
C. sp. (1) SL cult.	Sierra Leone (cult.) *	**MN719945**	**MN715153**	**MN715208**	**MN715181**
C. sp. (2) SL cult.	Sierra Leone (cult.) *	**MN719946**	**MN715154**	**MN715209**	**MN715182**
C. anthonyi	DR Congo	DQ153620	DQ153856	DQ153738	DQ153489
C. arabica (1)	Mauritius (cult.)	DQ153609	DQ153845	DQ153727	DQ153478.1
C. arabica (2) cult.	Sierra Leone (cult.)	**MN719947**	**MN715155**	**MN715210**	**MN715183**
C. arabica × *C. racemosa* (1) Brazil cult.	Brazil (cult.)	**MN719948**	**MN715156**	**MN715211**	**MN715184**
C. arabica × *C. racemosa* (2) Brazil cult.	Brazil (cult.)	**MN719949**	**MN715157**	**MN715212**	**MN715185**
C. arabica × *C. racemosa* (backcross) Brazil cult.	Brazil (cult.)	**MN719962**	**MN715170**	**MN715225**	**MN715198**
C. bakossi	Cameroon	DQ153599	DQ153835	DQ153717	DQ153468
C. brevipes	Cameroon	DQ153591	DQ153827	DQ153709	DQ153460
C. bridsoniae	Tanzania	DQ153584	DQ153822	DQ153704	DQ153455
C. canephora (1) SL cult.	Sierra Leone (cult.)	**MN719950**	**MN715158**	**MN715213**	**MN715186**
C. canephora (2) SL cult.	Sierra Leone (cult.)	**MN719951**	**MN715159**	**MN715214**	**MN715187**
C. canephora (3) SL cult.	Sierra Leone (cult.)	**MN719952**	**MN715160**	**MN715215**	**MN715188**
C. canephora (4) SL cult.	Cameroon (cult.)	DQ153593	DQ153829	DQ153711	DQ153462
C. congensis	Cameroon	DQ153632	DQ153834	DQ153716	DQ153467
C. costatifructa	Tanzania	DQ153604	DQ153840	DQ15372	DQ153473
C. eugenioides	Tanzania	DQ153588	DQ153824	DQ153706	DQ153457
C. fadenii	Tanzania	DQ153574	DQ153813	DQ153695	DQ153446
C. heterocalyx	Cameroon	DQ153594	DQ153830	DQ153712	DQ153463
C. humilis	Ivory Coast	DQ153611	DQ153847	DQ153729	DQ153480
C. kapakata	Angola	DQ153596	DQ153832	DQ153714	DQ153465
C. kihansiensis	Tanzania	DQ153583	DQ153821	DQ153703	DQ153454
C. kimbozensis	Tanzania	DQ153575	DQ153814	DQ153696	DQ153447
C. kivuensis	Tanzania	DQ153612	DQ153848	DQ153730	DQ153481
C. liberica (1) SL cult.	Sierra Leone (cult.)	**MN719956**	**MN715164**	**MN715219**	**MN715192**
C. liberica (2) SL cult.	Sierra Leone (cult.)	**MN719957**	**MN715165**	**MN715220**	**MN715193**
C. liberica (3) SL cult.	Sierra Leone (cult.)	**MN719958**	**MN715166**	**MN715221**	**MN715194**
C. liberica (4) SL cult.	Sierra Leone (cult.)	**MN719959**	**MN715167**	**MN715222**	**MN715195**
C. liberica (5) Uganda cult.	Uganda	**MN719960**	**MN715168**	**MN715223**	**MN715196**
C. liberica (6) CAR	Central African Rep.	DQ153603	DQ153839	DQ153721	DQ153472
C. liberica (7) DRC	DR Congo	**MN719953**	**MN715161**	**MN715216**	**MN715189**
C. liberica (8) DRC	DR Congo	**MN719954**	**MN715162**	**MN715217**	**MN715190**
C. liberica (9) DRC	DR Congo	**MN719955**	**MN715163**	**MN715218**	**MN715191**
C. liberica × *C. eugenioides*$^{+}$	DR Congo	DQ153623	DQ153859	DQ153741	DQ153492
C. lulandoensis	Tanzania	DQ153580	DQ153819	DQ153701	DQ153452
C. magnistipula	Cameroon	DQ153640	DQ153876	DQ153758	DQ153509
C. mayombensis	Cameroon	DQ153592	DQ153828	DQ153710	DQ153461
C. mongensis	Tanzania	DQ153576	DQ153815	DQ153697	DQ153448
C. montekupensis	Cameroon	DQ153590	DQ153826	DQ153708	DQ153459
C. mufindiensis	Tanzania	DQ153577	DQ153816	DQ153698	DQ153449
C. pocsii	Tanzania	DQ153582	DQ153820	DQ153702	DQ153453
C. pseudozanguebariae	Tanzania	DQ153578	DQ153817	DQ153699	DQ153450
C. racemosa (1) Brazil cult.	Brazil (cult.)	**MN719961**	**MN715169**	**MN715224**	**MN715197**
C. racemosa (2) Moz	Mozambique	DQ153627	DQ153863	DQ153745	DQ153496
C. racemosa (3) Moz	Mozambique	DQ153595	DQ153831	DQ153713	DQ153464
C. racemosa (4) Moz	Mozambique	DQ153628	DQ153864	DQ153746	DQ153497
C. rhamnifolia	Somalia*	DQ153589	DQ153825	DQ153707	DQ153458
C. salvatrix	Mozambique	DQ153622	DQ153858	DQ153740	DQ153491

(Continued)

TABLE 2 | Continued

Coffea species and accession identifier	Source	*ITS*	*trnL–trnF*	*rpl16*	*accD–psaI*
C. schliebenii	Tanzania	DQ153587	DQ153823	DQ153705	DQ153456
C. sessiliflora	Tanzania	DQ153579	DQ153818	DQ153700	DQ153451
C. stenophylla (1) SL cult.	Sierra Leone (cult.)*	**MN719966**	**MN715174**	**MN715228**	**MN715201**
C. stenophylla (2) SL cult.	Sierra Leone (cult.)*	**MN719967**	**MN715175**	**–**	**MN715202**
C. stenophylla (3) IC	Ivory Coast*	DQ153597	DQ153833	DQ153715	DQ153466
C. stenophylla (4) IC	Ivory Coast	**MN719968**	**MN715176**	**MN715229**	**MN715203**
C. stenophylla (5) SL	Sierra Leone	**MN719964**	**MN715172**	**MN715227**	**MN715200**
C. stenophylla (6) SL	Sierra Leone	**MN719965**	**MN715173**	**–**	**–**
C. stenophylla (7) Brazil cult.	Brazil (cult.)	**MN719969**	**MN715177**	**MN715230**	**MN715204**
C. stenophylla × *C. liberica* (1) cult.	Sierra Leone (cult.)	**MN719972**	**MN715180**	**MN715233**	**MN715207**
C. stenophylla × *C. liberica* (2) cult.	Sierra Leone (cult.)	**MN719970**	**MN715178**	**MN715231**	**MN715205**
C. stenophylla × *C. liberica* (3) cult.	Sierra Leone (cult.)	**MN719971**	**MN715179**	**MN715232**	**MN715206**
C. togoensis	Togo*	DQ153607	DQ153843	DQ153725	DQ153476
C. zanguebariae	Mozambique	**MN719963**	**MN715171**	**MN715226**	**MN715199**

*Newly sequenced material in bold. See **Table 1** for details of accessions sequenced for this study. See Maurin et al. (2007) for details of previously sequenced material. *Accessions sequenced from herbarium material; [+]Accession listed as Coffea heterocalyx in Maurin et al. (2007).*

species alliance terminology follow Maurin et al. (2007) and Davis et al. (2011).

RESULTS

Literature Review

Coffea stenophylla (Highland Coffee of Sierra Leone, Rio-Nunez Coffee, Senegal Coffee, Sierra Leone Coffee) (Figures 1, 2A)

Knowledge of *C. stenophylla* and its commercial potential dates back to at least 1794, based on reports by Adam Afzelius (1750–1837), who worked in, and collected plants from, Sierra Leone (Hiern, 1876). *Coffea stenophylla* was described as new to science in 1834 (Don, 1834) and was characterized on the basis of having narrow leaves (hence the species epithet) and black fruits (most coffee species have red fruits). Don (1834) stated that is was a 'Native of Sierra Leone, where it is cultivated'. . . and that 'The seeds of this species are roasted and used as the common coffee, and are even considered superior to it.' De Wildeman (1904) reported *C. stenophylla* as indigenous in Guinea (in the forests that border the Southern Rivers area) and Ivory Coast. From at least the 1850s, the seeds of *C. stenophylla* were disseminated from Sierra Leone, with the accompanying vernacular names of 'Highland coffee of Sierra Leone' or 'Sierra Leone coffee' (Hooker, 1896). Commercial (cultivated) samples reached the Royal Botanic Gardens, Kew (England) in 1856 (specimens in the Economic Botany Collection, Kew). According to Chevalier (1929), *C. stenophylla* was being cultivated in quantity in Sierra Leone in the 1890s (c. 1893), and in Guinea, to the extent that is was exported as a commercial product to France. In France it apparently received an exceptionally favorable market price (Scott Elliot, 1893). Living material (seeds) of *C. stenophylla* were sent to the Royal Botanic Gardens, Kew in May 1894, and from here it was sent to India, Sri Lanka (then known as Ceylon), Trinidad (**Figure 1**), and Java (Cheney, 1925; Scott Elliot, 1893). From Guinea is was sent to Vietnam (Chevalier, 1929) and probably other countries under French colonial rule at that time. In 1904, De Wildeman (1904) provided a summary of the cultivation of *C. stenophylla* in Guinea, where it seems to have been cultivated in some quantity as Rio-Nunez coffee, after the Nunez River (a major river in Guinea). It was also cultivated in Ghana, Senegal (where it was known as Senegal coffee), and Ivory Coast, possibly through early intervention by the Portuguese (Haarer, 1962), and in Uganda (Tothill, 1940). Chevalier (1929), states that the export of *C. stenophylla* in Sierra Leone and Guinea amounted to around three to five tons (3,000 to 5,000 kg) per year, although this does not include the amount of coffee consumed in these producing countries, which may have been substantial. *Coffea stenophylla* appears to have been a prominent feature of agriculture in Sierra Leone up until at least the 1920s (Dudgeon, 1922), but it may have been in decline after that time (Chevalier, 1929), perhaps due to a fall in coffee prices (Dudgeon, 1922). Elsewhere, despite all the reports of an excellent flavor (Don, 1834; Scott Elliot, 1893; De Wildeman, 1904, 1906b; Watt, 1908; Macmillan, 1914; Dudgeon, 1922; Cheney, 1925; Chevalier, 1929; Wellman, 1961; Haarer, 1962) and a range of potential agronomic attributes (see below) it did not prevail as a coffee crop species. Chevalier (1929) reported that although it was considered by many to be an exquisite coffee ("Suivant beaucoup de dégustateurs, c'est un café exquis") it was not widespread in global cultivation, due to low yields. Likewise, despite being introduced to Uganda in 1919 and then again in 1931, small bean size and low yield prevented it from being a sustainable coffee crop plant (Tothill, 1940). Despite this, a good agronomic performance at low elevations (e.g., c. 150 m) has been reported for *C. stenophylla* (Watt, 1908; Cheney, 1925; Bramel et al., 2017), and there is a report of potential resistance to CLR (Cheney, 1925). Given that the natural, and once-cultivated, environments of these two species in Upper West Africa were at relatively low elevations (150–610 m) (Watt, 1908) and that *C. stenophylla* is reported to withstand dry conditions (Wellman, 1961; Wrigley, 1988), there may also be



FIGURE 1 | *Coffea stenophylla*, cultivated in Trinidad Botanical Garden, with Demerara sugarcanes, photograph taken around 1900. The man in the photograph is 5 ft. 8 in. (1.72 m) tall. Image: Royal Botanic Gardens, Kew.

some resilience to high temperatures and low rainfall, compared to the main crop species.

Since the 1940s, most of the literature on *C. stenophylla* has been restricted to the recycling of information from previous publications (Wellman, 1961; Haarer, 1962; Wrigley, 1988; Stoffelen, 1998; Davis et al., 2006). The exceptions to this are the reports of dedicated coffee collection missions in Upper West Africa undertaken in the 1980s. Reporting on various missions to Ivory Coast between 1984 and 1987, Le Pierrès et al. (1989) record two wild populations of *C. stenophylla* from the main forest block of the Sud-Est region of Ivory Coast (north east of Abidjan); and from Guinea, 114 examples of small scale cultivation of this species in the gardens of local houses (between Boffa and Boke, and around Boke). Berthaud (1986) demonstrated the existence of populations of *C. stenophylla* in Ivory Coast, from Ira Forest (Forêt L'Ira), three other localities (populations) in the east of the country. In addition, Berthaud (1983) recorded this species at a dry forest site in the Ouellé area in western Ivory Coast. Berthaud (1986) reported *C. stenophylla*, *C. liberica*, and *C. canephora* in Ira Forest (Forêt L'Ira); *C. stenophylla* was restricted to the upper, drier parts of the hills, whereas the other two species were found in the valley bottoms (lower, wetter areas).

Coffea affinis (Kamaya Coffee)

This species was first reported in c. 1900 from the coffee research garden of M. Boery in Guinea, having been originally collected from nearby native forests (De Wildeman, 1904). On first inspection, De Wildeman considered these plants to be similar in many characteristics to *C. stenophylla* (i.e., the presence

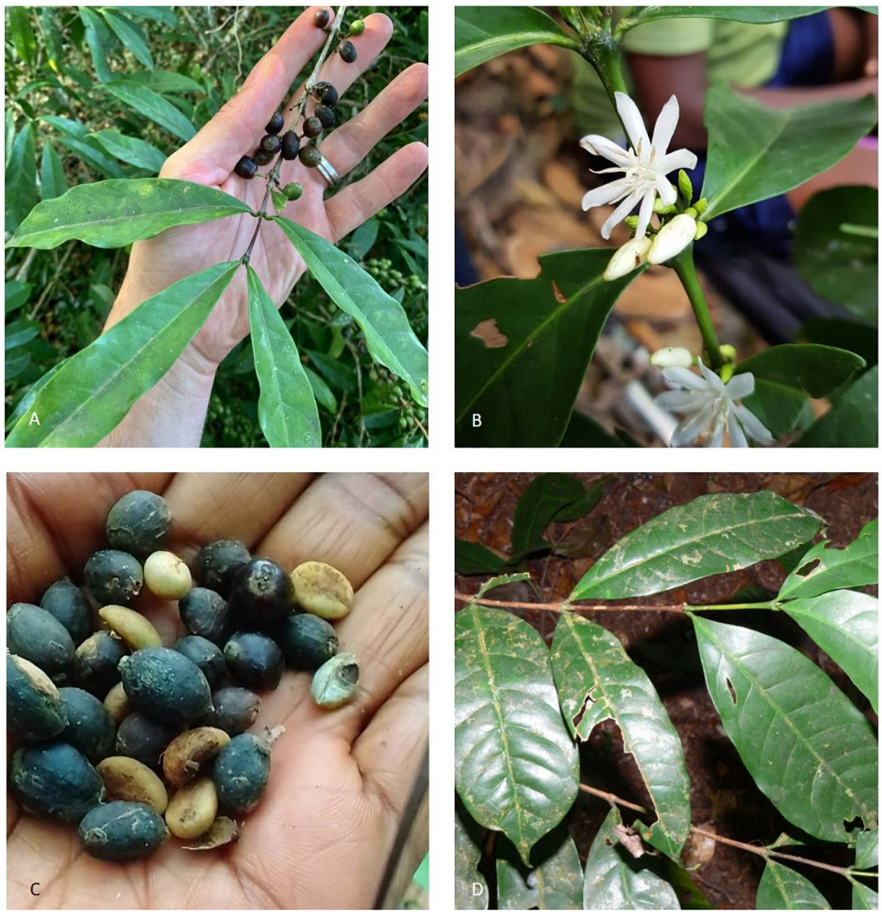


FIGURE 2 | *Coffea affinis* and *C. stenophylla*. **(A)** *C. stenophylla* in fruit, at Centre National de Recherche Agronomique (CNRA), Ivory Coast (image: Charles Denison); **(B)** *C. affinis* in flower; **(C)** *C. affinis*, fruits and seeds (partially dried); **(D)** *C. affinis*, leaves. Images (**B–D**), from Kambui Hills, Sierra Leone (images: Daniel Sarmu).

of black fruits, rather than the usual red) but different in other respects and particularly leaf size and shape (De Wildeman, 1904). When De Wildeman visited Guinea (De Wildeman, 1904) to observe these plants, which were being grown collectively as Rio-Nunez coffee, he declared that there were two species, *C. stenophylla* and another species, which he named as a new to science: *C. affinis*. According to De Wildeman (1904) *C. affinis* was akin to *C. stenophylla* in the color and shape of the fruit, and shape of the seeds, but differed in its vegetative characters (e.g., stems, leaves, stipules) and mainly by its larger leaves. De Wildeman (1904) believed that *C. affinis* was of considerable importance as new coffee crop species, due to its vigor, the quality (high value) of the coffee, and general resistance to disease (compared to Arabica coffee). A contemporaneous photograph of *C. affinis* (De Wildeman, 1906b) shows a coffee plant that differs from the narrow-leaved *C. stenophylla* by having larger, broader leaves.

Contrary to the viewpoints of De Wildeman (1904, 1906b), Chevalier (1905) suggested that *C. affinis* was native in Sierra Leone, as he had no knowledge of it growing wild in Guinea. Subsequently, Chevalier (1929) considered *C. affinis* to be a hybrid between *C. liberica* and *C. stenophylla*. The potential hybrid status of *C. affinis* was discussed at length by Portères (1937a), who argued against a hybrid origin, particularly in relation to a new coffee plant he considered indigenous to the Ivory Coast, known locally as 'Kamaya.' He drew a close association between 'Kamaya' and *C. affinis*, but owing to the uncertainty over the application of *C. affinis*, decided to name this plant *C. stenophylla* var. *camaya*. Portères (Portères, 1937a) reported that *C. stenophylla* var. *camaya* was found as single example in a coffee plantation near Abengourou (**Figure 3**), but that it originated from the wild at a nearby location close to Niabli (6° 39′ N 3° 16′ W) and that a few small plantations were established in Ivory Coast. A decade later, Chevalier (1947) suggested that *C. stenophylla* var. *camaya* and *C. affinis* were the same species, and in contrast to his earlier report was only found in its wild state in Ivory Coast. Wellman (1961) considered *C. affinis* to be indigenous to Guinea and Ivory Coast, and suggested that it was a fixed mutation of *C. stenophylla*. Other workers (Cramer, 1957; Stoffelen, 1998) referred back to the earlier opinion of Chevalier (1905), i.e., that *C. affinis* is a hybrid between *C. liberica* and *C. stenophylla* (Davis et al., 2006).

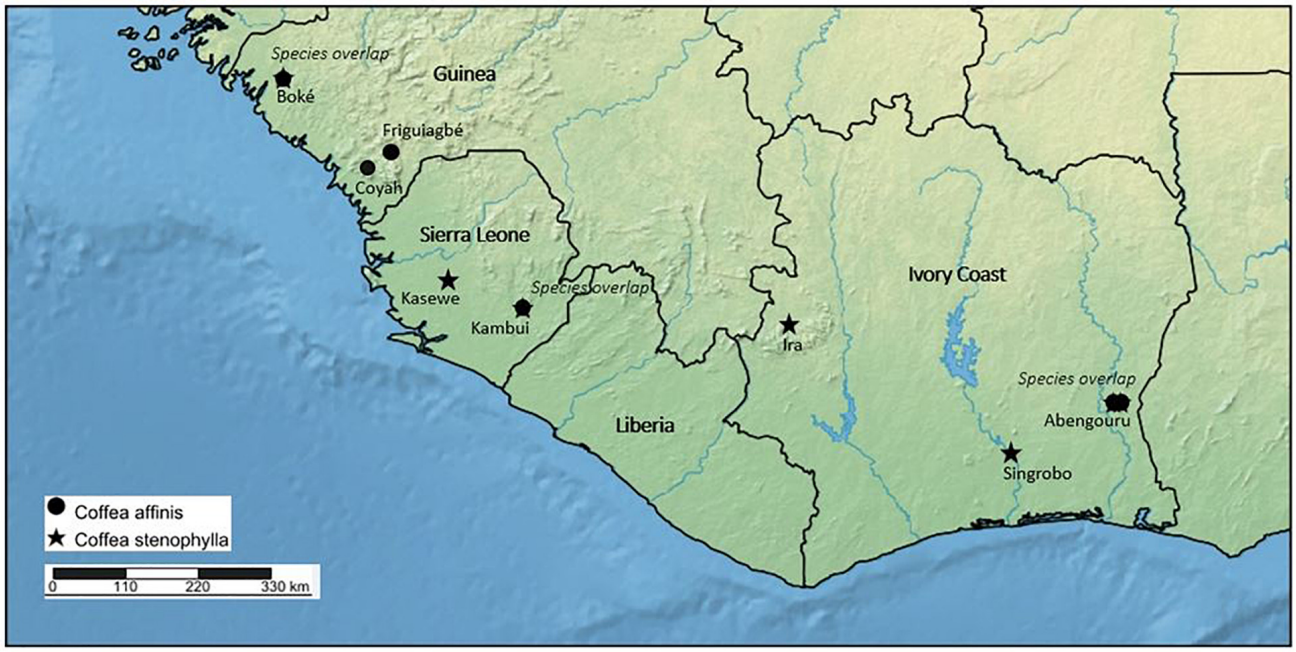


FIGURE 3 | Distribution map of *Coffea affinis* and *C. stenophylla* based on herbarium and literature survey, and fieldwork. Labels beneath species symbols indicate general collection sites [Boké, Friguiagbé, Coyah (Saliya Forest Reserve), Kasewe (Kasewe Hills Forest Reserve, near Moyamba), Kambui (Kambui Hills Forest Reserve, near Kenema, Ira [Ira Forest (Forêt L'Ira), north of Man], Abengouru and Singrobo)]. *Species overlap*, indicates where *C. affinis* and *C. stenophylla* occur in the same location.

Herbarium Collection Survey

For *C. affinis* the herbarium survey yielded 12 unique records (seven cultivated and five wild records) with a total of 28 herbarium specimens (including duplicates), and for *C. stenophylla* 50 unique records (29 cultivated, 12 wild, and 9 with no data) and 72 herbarium specimens. Herbarium records associated with these species include the hybrid *C. liberica* × *C. stenophylla*, of which there were two unique records, both from cultivated material found in research collections. Compared to many other coffee species the number of herbarium specimens is low, especially for *C. affinis*. We examined two commercial seed collections of *C. stenophylla*, one with and one without parchment (pre-milling stage, with endocarp attached) and one of clean (green, pre-roasted) coffee (endocarp removed), and one fruit collection (whole, sun-dried fruits) from the Economic Botany Collection of the Royal Botanic Gardens, Kew (K). Examination of herbarium specimens confirms many aspects of the literature survey (see above).

From the herbarium survey, *C. stenophylla* is confirmed as an indigenous (wild) species of Guinea, Sierra Leone and Ivory Coast (Davis et al., 2006), and that it was also farmed and otherwise cultivated (e.g., research stations and farms) in these countries, as per the literature (see above). Most of the herbarium specimens date from the late 1800s and early 1900s. The most recent collections for these countries are as follows: Guinea (from the wild, 1941; from (small-scale) farm cultivation 1961); Sierra Leone [wild, 1954; from (small-scale) farm cultivation 1963]; Ivory Coast (wild 1932; from farms, no data). By comparison, the literature survey reveals small scale cultivation of *C. stenophylla* in Guinea and Ivory Coast, in the mid to late 1980s (Berthaud, 1983, 1986; Le Pierrès et al., 1989). Herbarium data also show that this species was cultivated in coffee research collections, and other germplasm collections, in Africa (Ivory Coast, Guinea, Nigeria, Sao Tomé, Sierra Leone, Tanzania, Ghana, and the Democratic Republic of Congo) and in Asia (Vietnam, Java).

The herbarium survey reveals that *C. affinis* is as an indigenous (wild) species of Guinea and Ivory Coast (De Wildeman, 1904). Chevalier could not find any evidence of its wild status in Guinea, referring only to cultivated material from the gardens of Conakry and Cameyenne (Chevalier, 1905), but herbarium data provide clear evidence of collections from natural forests in Guinea (see below). In contrast to the views of Chevalier (1905), and even more recent opinion (Davis et al., 2006) we could not find any evidence of wild *C. affinis* in Sierra Leone, which is also the case for the literature survey (but see Fieldwork in Sierra Leone, below). Herbarium specimens exist that were collected from a coffee plantation [Cope *s.n.*, 7 iii 1912 (K)], labeled as *C. affinis*, which has the key leaf characteristics of this species but the flowers are absent, and the fruit color is not noted. Regarding the native status of *C. affinis* in Guinea, there are two collections in the Paris herbarium (P) collected from the environs of Boké (Boké Prefecture) that are identifiable as *C. affinis* and the labels clearly state that the plants were spontaneous (wild): Chillou *s.n.* 20 xii 1923 (three duplicates); Chillou *s.n.*, 17 xii 1923 (two duplicates); a third specimen collected from Friguiagbé (Kindia Prefecture), by the same collector (Chillou *2381*, 3 ii 1941) is also

likely to be spontaneous, although the native/cultivated status is not indicated on the specimen. The most recent collections for these countries are as follows: Guinea (from the wild, 1941; from farms, 1905); Sierra Leone (from a coffee plantation, 1912); Ivory Coast (wild 1930; from farms 1934).

Fieldwork in Sierra Leone

A collection of 20 samples (leaf samples, images, and DNA samples) were made between 2014 and 2016 resulting from our request for samples of *C. affinis* and *C. stenophylla* and of any atypical coffee morphotypes (see the section "Materials and Methods"). Of these, seven were selected for DNA analysis (**Table 1**); the remaining samples conformed to regular variants of *C. robusta* and *C. liberica*. Three samples of putative *C. stenophylla* and *C. affinis* coffee were collected from the SLARI research collection (**Table 1**). The 10 samples were either considered as potential candidates for *C. stenophylla* or *C. affinis*, or hybrids between these species. Visits to sites (in 2017) where *C. stenophylla* had been recorded in cultivation in northern Sierra Leone failed to produce any coffee sightings. In December 2018, we followed up on the farm survey, visiting five farms that had stated cultivation of *C. stenophylla*, but no plants of this species or *C. affinis* were located (only *C. canephora*). Our visit to Kasewe Hills (Southern Province) resulted in the collection of a single sterile (no flowers or fruits) immature plant, which we preliminary identified as *C. stenophylla*. We did not find any plants matching *C. stenophylla* in forest locations within the Western Peninsula National Park (near Freetown) or near Moyamba Junction, but located a small population (with mature trees up to 7 m tall) matching this species in the forested area of Kambui Hills. At both localities the plants were collected in humid evergreen (lowland) forest at c. 400 m elevation — on a ridge top in the case of Kasewe Hills and on the side of a ridge on steeply sloping ground at Kambui Hills. Further visits to Kambui Hills throughout 2019 and early 2020 yielded further *C. stenophylla*, and trees provisionally identified as *C. affinis*, in both flower and fruit (**Figures 2B–D**). DNA samples from the two locations (Kasewe Hills and Kambui Hills) were added to the DNA analyses (see below and **Table 1**).

DNA Analyses

A total of 120 sequences from four DNA regions, from 31 accessions, were generated for this study and their sequences deposited in GenBank (with NCBI accession numbers; see **Table 2**); 35 species-level reference sequences were downloaded from GenBank, from two previous studies (Maurin et al., 2007; Davis et al., 2011). The ITS alignment had a length of 804 bp, whereas the concatenated plastid dataset (*trnL–trnF*, *rpl16* and *accD–psaI*) had a total length of 3253 bp. A merged ITS/plastid matrix for a subset of species containing members of the Upper Guinea (UG) Clade, plus two outgroup species, had a length of 3989 bp. Some sequences were missing due to poor sequencing quality: *rpl16* in *C. stenophylla* (2) SL cult. and *C. stenophylla* (6) SL; and *accD–psaI* in *C. stenophylla* (2) SL cult. and *C. affinis* (1) SL. Accession information is given in **Tables 1**, **2**.

ITS Analysis (Figure 4)

The results obtained are consistent with previously obtained ITS analysis (Maurin et al., 2007; Davis et al., 2011), in terms of species relationships and their placement into geographically delimited clades. All tetraploid hybrids ($4n$ = 44) between *C. arabica* and *C. racemosa* are placed in the East African (EA) Clade (2) with *C. racemosa* (BS = 1). The diploid hybrid ($2n$ = 22) *C. liberica* × *C. eugenioides* is placed (BS = 1) with species of the East-Central Africa (EC-Afr) Clade, in an unresolved position with *C. eugenioides*. The natural tetraploid hybrid ($4n$ = 44) *C. arabica* (two accessions) is placed with species of the Lower Guinea/Congolian (LG/C) Clade in the 'Canephora Alliance' (BS = 1), sister to two accessions of *C. canephora* (BS = 1). Specimens of *C. stenophylla* collected from the wild in Sierra Leone [specimens (5) SL & (6) SL], other wild accessions of this species, and wild accessions of *C. affinis* from Sierra Leone, all fall within the Upper Guinea (UG) Clade (BS = 0.93) with *C. humilis* and *C. togoensis* [specimens (1) SL & (2) SL]. Three cultivated collections from Sierra Leone originally accessioned as *C. arabica* and *C. affinis* (×2) were placed in the Upper Guinea (UG) Clade (labeled in **Figure 4** as *C. liberica* × *C. stenophylla* SL cult. (1), (2), and (3). The historical accessions of farmed *C. affinis* from Sierra Leone (*C. affinis* in **Table 1**) were not placed with *C. stenophylla* or *C. affinis*, but in a clade with *C. liberica* and *C. montekupensis* (BS = 0.79), within the LG/C Clade [labeled as *C.* sp. (1) and *C.* sp. (2) in **Figure 4**].

Plastid Analysis (Figure 5)

The results obtained are consistent with previously obtained plastid analysis using the same markers (Maurin et al., 2007; Davis et al., 2011), in terms of the relationships between species and their placement into geographically delimited clades. The F1 tetraploid hybrids ($4n$ = 44) *C. arabica* × *C. racemosa* are placed in the East-Central Africa (EC-Afr) Clade, sister to a clade comprising *C. arabica*, *C. eugenioides*, *C. kivuensis*, and *C. anthonyi*. In contrast to the ITS analysis, the backcrossed *C. arabica* × *C. racemosa* is placed in the EA clade with wild and cultivated *C. racemosa* accessions. *Coffea liberica* × *C. eugenioides* is placed in one of the Lower Guinea/Congolian (LG/C) Clades with two species, *C. liberica* and *C. magnistipula* (BS = 0.66) in an unresolved position with three *C. liberica* accessions (BS = 0.99). The tetraploid ($4n$ = 44) hybrid species *C. arabica* is placed with species of the EC-Afr Clade, viz. *C. eugenioides*, *C. kivuensis*, and *C. anthonyi* (BS = 0.97). Three cultivated collections from Sierra Leone originally accessioned as *C. arabica* and *C. affinis* (×2) were placed within one of the two LG/C clades [labeled in **Figure 5** as *C. liberica* × *C. stenophylla* SL cult. (1), (2), and (3)] in an unresolved position in a clade with various *C. liberica* accessions (BS = 0.97), and a subclade of *C. liberica* × *C. eugenioides, C. liberica* and *C. magnistipula* (BS = 0.66).

All wild and cultivated accessions of *C. stenophylla*, and wild accessions of *C. affinis* fall within the UG Clade (BS = 1), which includes *C. togoensis* and *C. humilis*. A separate clade containing only all wild and cultivated accessions of *C. stenophylla* and all

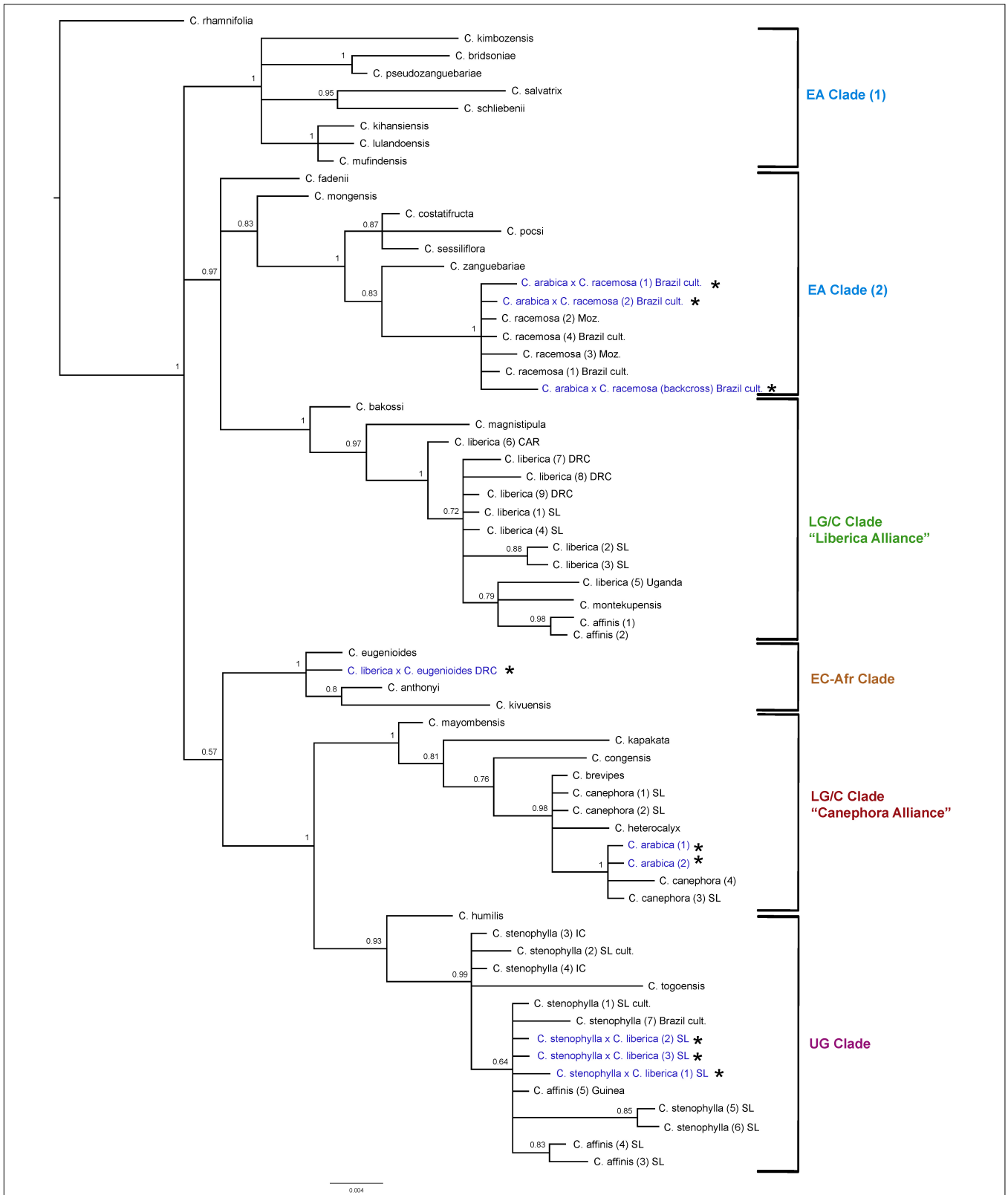


FIGURE 4 | ITS maximum clade credibility tree. Bayesian posterior probabilities are indicated above branches. See **Tables 1**, **2** for accession information. Country abbreviations: CAR, Central African Republic; DRC, Democratic Republic of Congo; IC, Ivory Coast; Moz, Mozambique; SL, Sierra Leone. Clade terminology follows Maurin et al. (2007) and Davis et al. (2011): EA, East Africa; LG/C, Lower Guinea/Congolian; EC-Afr, East-Central Africa; UG, Upper Guinea. All known and identified interspecies hybrids are marked in blue text and with a star (*).

wild species of *C. affinis*, but including *C. humilis* fall within and in an unresolved clade (BS = 0.76). Our historical accessions of farmed *C. affinis* (*C. affinis* in **Table 1**) from Sierra Leone are not placed with *C. stenophylla* or *C. affinis*, but in a clade with *C. liberica* (BS = 0.98), sister to *C. montekupensis* (BS = 1), within one of the two LG/C clades [accessions are labeled as *C.* sp. (1) and *C.* sp. (2) in **Figure 5**].

Incongruence Between ITS and Plastid Trees

There are several points of substantial incongruence between the ITS and plastid analyses (**Figures 4**, **5**), including those taxa of known (manmade) or proven (via DNA study) hybrid origin: *C. arabica* (*C. arabica* × *C. eugenioides* (Maurin et al., 2007), *C. arabica* × *C. racemosa* (Medina Filho et al., 1977a,b), and *C. liberica* × *C. eugenioides* (Maurin et al., 2007). These incongruencies are anticipated based on the knowledge, DNA and otherwise, that they are hybrids. In this study we identified three samples, originally accessioned (**Table 1**) as *C. arabica* and *C. affinis* (× 2), that were substantially incongruent in each analysis. Given the specific positions of these accessions in the analyses (see ITS and plastid results, and **Figures 4**, **5**), and their morphological features, we suggest that they represent the hybrid *C. stenophylla* × *C. liberica*. All known and identified interspecies hybrids are marked in the phylogenetic trees (**Figures 4**, **5**) in blue text and with a star (*).

Combined Analysis for the Upper Guinea (UG) Clade (Figure 6)

Combining ITS and plastid data sets for all members of the Upper Guinea (UG) clade, *C. humilis*, *C. togoensis*, *C. affinis*, and *C. stenophylla*, produced relationships congruent with previous analyses using the same data (Maurin et al., 2007; Davis et al., 2011). The four species of the UG Clade are monophyletic (BS = 1), with *C. humilis* sister to *C. togoensis, C. affinis* and *C. stenophylla* (BS = 1); *C. affinis* and *C. stenophylla* form a clade (BS = 0.88) but each species is not monophyletic. The Kambui Hills accessions of *C. affinis* [(1) SL and (2) SL] and *C. stenophylla* accessions from Kasewe Hills and Kambui Hills [(5) SL and (6) SL, respectively] are monophyletic (BS = 0.92), and are sister to two accessions of *C. stenophylla* [(3) IC and (4) IC] from Ivory Coast (BS = 0.76). The *C. affinis* accession [(3) Guinea] from Guinea falls (BS = 0.78) with *C. stenophylla* accessions [(1) SL cult. and (7) Brazil cult], which originate from unknown localities in Sierra Leone.

DISCUSSION

Historical and Present-Day Status of *C. affinis* and *C. stenophylla*

In 2018, we rediscovered *C. stenophylla* in two locations in Sierra Leone, one from where it been collected before (Kasewe Hills, in 1954) and one a new location (Kambui Hills) (**Figure 3**). In both locations, *C. stenophylla* is extremely localized, and seemingly threatened. In the Kasewe Hills, near Moyamba, we were only able to locate a single plant, in an area of high deforestation. In the Kambui Hills, near Kenema, we located a small population, the extent of which is as yet unknown, but there are ongoing threats from logging, human encroachment, and artisanal gold mining. In late 2019, we located *C. affinis* in the Kambui Hills, not far from the populations of *C. stenophylla*. This is the first record of this species from the wild in Sierra Leone. The present day status of *C. stenophylla* and *C. affinis* in Guinea and Ivory Coast is poorly known. In Guinea, *C. stenophylla* was recorded as being under limited small scale cultivation in the 1980s (Le Pierrès et al., 1989), but there were no records of wild plants at that time. From a basic survey of remaining forest cover in Guinea, using satellite imagery (Google Earth, 2019), the likelihood of finding *C. affinis* and *C. stenophylla* in many of the localities where it was previously recorded as an indigenous plant (see the section "Results"; **Figure 3**) is limited, although possible. Field survey in more remote locations in Guinea, in appropriate environments and elevations, may reveal wild populations of these species. Indeed, we report here on a recent collection [2015; *Couch 757* (K)] of a sterile (no flowers or fruit) coffee specimen from Guinea, which was identified on the basis of our DNA sequencing and morphology as *C. affinis* (see below). Despite this encouraging find, deforestation rates in Guinea are very high and ongoing (Couch et al., 2019); in 1992 it was calculated that 96% of the original forest had already been destroyed (Sayer et al., 1992). In Ivory Coast, the likelihood of finding more extensive wild populations of *C. stenophylla* and *C. affinis* are better, particularly as satellite data (Google Earth, 2019) shows the existence of native remnant vegetation in localities where this species was previously recorded, and where it is likely to be located. That said, one of the best known forest sites for this species in Ivory Coast (Berthaud, 1986), i.e., at Ira Forest (Forêt L'Ira), was largely destroyed around 2008. In summary, *C. stenophylla* and *C. affinis* are threatened throughout their indigenous ranges, and particularly in Guinea. On the IUCN Red List *C. stenophylla* is assessed as Vulnerable (VU); *C. affinis* is currently Data Deficient (DD) (International Union for Conservation of Nature (IUCN), 2020). On the basis of our literature, herbarium, and field survey, and DNA analysis, *C. stenophylla* and *C. affinis* are indigenous species of Guinea, Ivory Coast and Sierra Leone (**Figure 3**).

Historical data shows that *C. stenophylla* and *C. affinis* were farmed in some quantity in Upper West Africa, and especially *C. stenophylla* in Guinea and Sierra Leone. *Coffea stenophylla* was also widely cultivated in research stations across Africa and in various Asian countries; the presence of *C. affinis* in research stations appears to have been restricted to Upper West Africa (Guinea, Ivory Coast) during the early part of the last century. Our field surveys in Sierra Leone indicates that neither *C. stenophylla* nor *C. affinis* are under commercial cultivation (i.e., being farmed), or otherwise cultivated, in the present day. The last confirmed record of *C. stenophylla* production in Sierra Leone may have been the small plantation at Njala University grounds, which was apparently cut down after being abandoned in the 1980s (A. Lebbie pers. comm.). In Guinea, *C. stenophylla* was recorded as being under limited small scale cultivation in the 1980s (Le Pierrès et al., 1989). *Coffea stenophylla* has

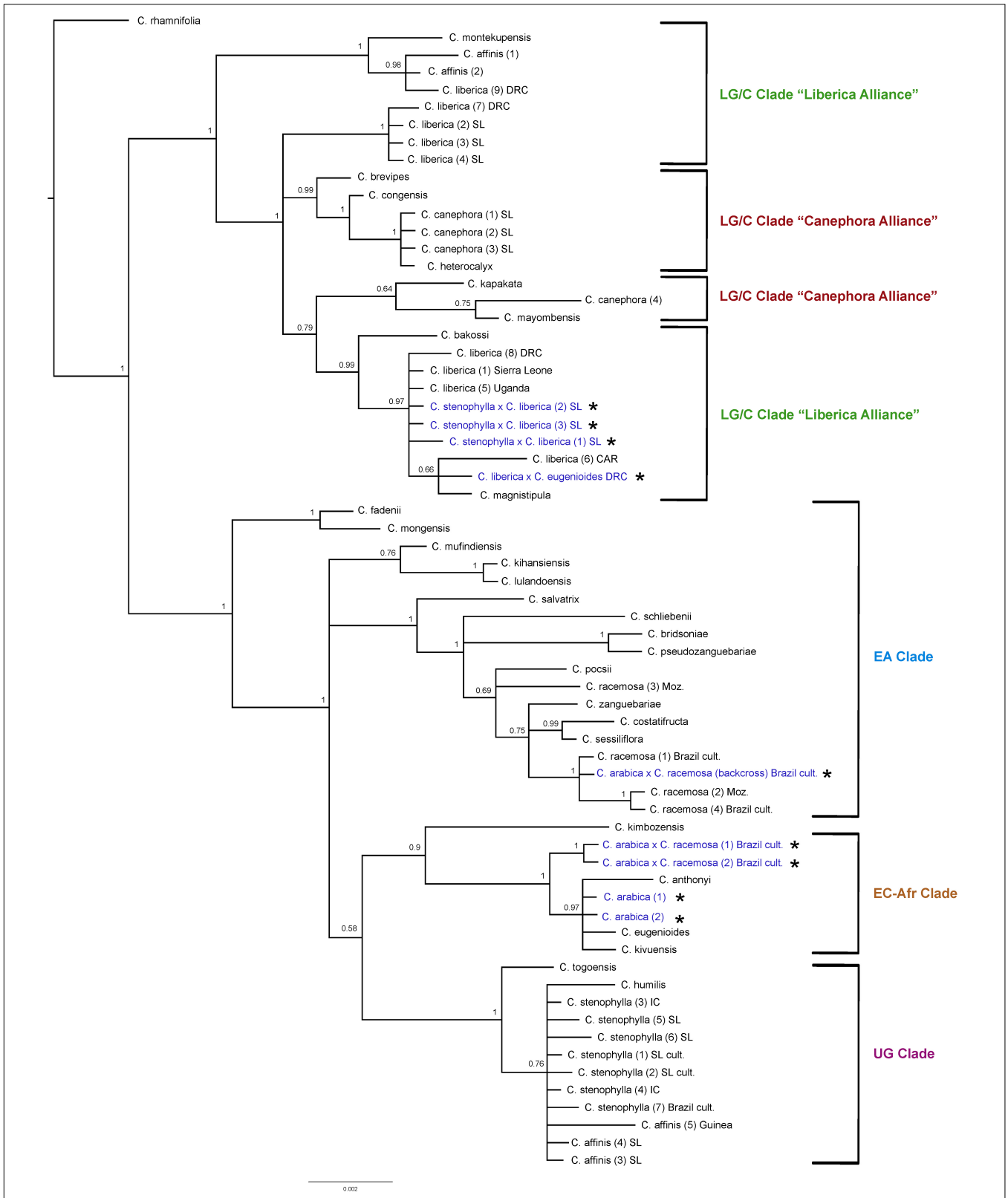


FIGURE 5 | Plastid maximum clade credibility tree. Bayesian posterior probabilities are indicated above branches. See **Tables 1**, **2** for accession information. Country abbreviations: CAR, Central African Republic; DRC, Democratic Republic of Congo; IC, Ivory Coast; Moz, Mozambique; SL, Sierra Leone. Clade terminology follows Maurin et al. (2007) and Davis et al. (2011): EA, East Africa; LG/C, Lower Guinea/Congolian; EC-Afr, East-Central Africa; UG, Upper Guinea. All known and identified interspecies hybrids are marked in blue text and with a star (*).

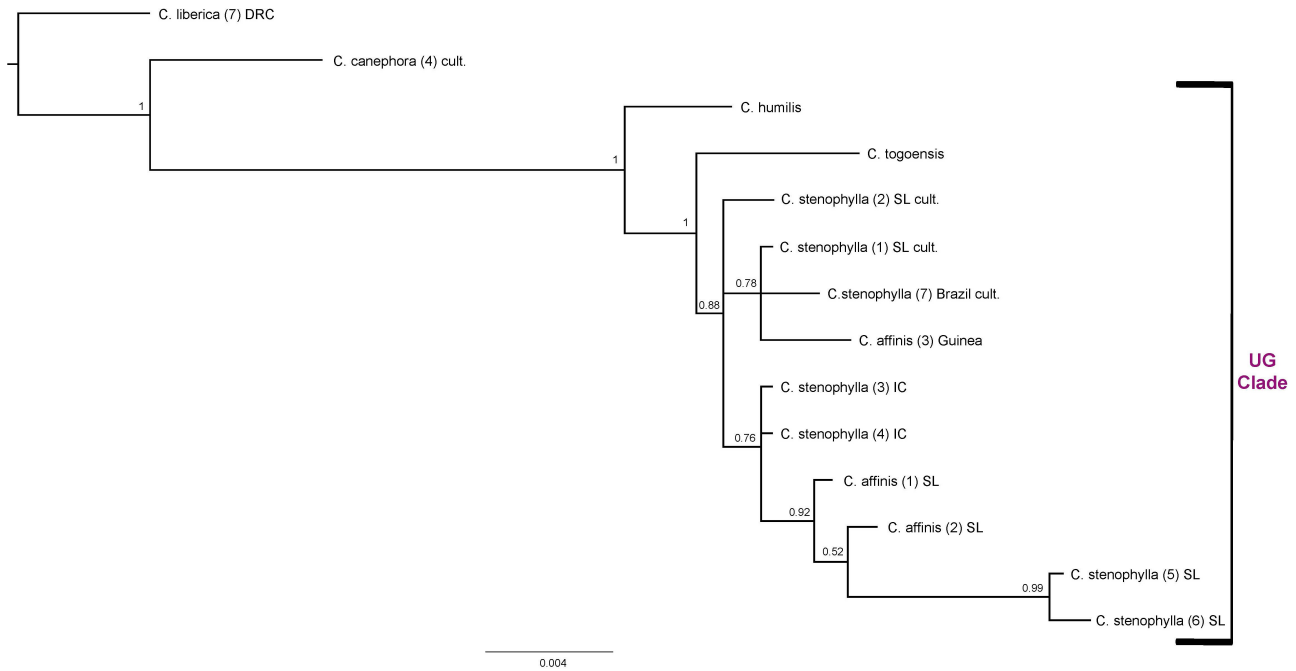


FIGURE 6 | Combined ITS and plastid maximum clade credibility tree. Bayesian posterior probabilities are indicated above branches. See **Tables 1**, **2** for accession information. Country abbreviations: IC, Ivory Coast; SL, Sierra Leone. Clade terminology follows Maurin et al. (2007) and Davis et al. (2011): UG, Upper Guinea.

been recorded recently in several coffee research collections including the Centre National de Recherche Agronomique (CNRA), in Ivory Coast (**Figure 2A**); L'Institut de recherche pour le développement (IRD), France [A. Davis pers. observ.], and Entebbe Botanical Gardens, Uganda [A. Davis pers. observ.]. It is also likely to exist in other *ex situ* collections. However, across the *ex situ* coffee germplasm network there is the problem of duplication, i.e., the same genotype(s) being represented in multiple sites (Anthony et al., 2007; Bramel et al., 2017; Davis et al., 2019). There also seems to be some confusion with the narrow-leaved variant of *C. arabica* 'Angustifolia,' which owing to its narrow leaves is sometimes mistakenly accessioned as *C. stenophylla*. *Coffea affinis* has not been recorded in coffee research collections since the beginning of the twentieth century. Genotyping by sequencing, or similar methods, are required to make a full assessment of *ex situ* collections of *C. stenophylla*, and all other coffee species.

Natural Habitat (Growing Environment) of *C. affinis* and *C. stenophylla*

Knowing the location and associated habitat of crop wild relatives is important, as it can provide an initial assessment of environmental suitability as a crop plant. Our literature survey indicates that *C. stenophylla* may have drought tolerance characteristics (Portères, 1937b; Wellman, 1961; Wrigley, 1988). Both species occur at relatively low elevations (150–700 m) (De Wildeman, 1906a; Watt, 1908; Portères, 1937b). In Ivory Coast (at Ira Forest) *C. stenophylla* occurs on the upper, drier parts of hills; in the same location, *C. canephora* and *C. liberica* was found in the valleys (i.e., the lower, wetter areas). The locations for this species in Ivory Coast are generally drier than Sierra Leone (see below), with rainfall in the region of 1,500–1,700 mm per year, a 3–4 months dry season (Portères, 1937b), and an average annual temperature of c. 25.5°C. In Guinea, De Wildeman (1906a) reported that in its natural state *C. affinis* occurs in gallery forest (forest associated with rivers) bordering waterfalls and in humid (evergreen) forest, and that it was frequently found at elevations of 400–700 m at a distance of 100 to 300 km from the sea.

In Sierra Leone our fieldwork located *C. stenophylla* at precisely 400 m at two locations (Kasewe and Kambui Hills), even when there was sufficient forest down to 200 m. We visited higher elevation locations within the Western Peninsula National Park (up to 600 m), but did not locate either of the two species. Sierra Leone is generally not a mountainous country, and most of its land is below 500 m. At Kasewe and Kambui Hills the climate is tropical monsoonal, with an annual rainfall average of 2,350–2,650 mm, an average annual temperature of c. 26°C, and a distinct 3–4 months dry season (November to March/April). It should be noted that there is a considerable difference in rainfall between the wild locations of *C. affinis* and *C. stenophylla* in Ivory Coast and Sierra Leone, although this requires careful verification.

Our herbarium survey did not provide a great deal of further information on the habitat of either of these two species. Notes on herbarium specimens for *C. stenophylla* infrequently state 'hills,' one specimen states 'very common [on] more open places on augite hills,' and another 'dans les montagnes de Sierra Leone,' which suggested an association with high ground topology.

Species Status and Systematic Affinities of *C. affinis* and *C. stenophylla*

A recent hybrid origin for *C. affinis*, as a result of crossing between *C. liberica* and *C. stenophylla* (Chevalier, 1929; Cramer, 1957; Stoffelen, 1998) is ruled out on three counts: (1) *C. affinis* is clearly fertile and productive, as evidenced from literature records (Portères, 1937a), herbarium specimens (with plentiful seed), and recent (2020) field observation (D. Sarmu pers. observ.; **Figure 2C**). Diploid interspecies hybrids of coffee are usually sterile, and while they may produce flowers, fruit set and production of viable seed are minimal unless fertility is restored via polyploidization (usually tetraploids) (Carvalho and Monaco, 1968; Charrier, 1978). (2) The fruits of *C. affinis* are always described as black; the hybrid *C. liberica* × *C. stenophylla* (see below) has purple fruits. (3) Our samples of *C. affinis* do not alternate between ITS and plastid markers, as in known hybrids [*C. arabica*, *C. arabica* × *C. racemosa*, *C. liberica* × *C. eugenioides*, and *C. liberica* × *C. stenophylla* (see below)], but instead are consistently resolved as sister to *C. stenophylla*. The idea that *C. affinis* is a fixed mutation of *C. stenophylla*, (Wellman, 1961) is also ruled out because of the variation evident in this taxon.

Our DNA analyses infer that *C. stenophylla* and *C. affinis* are closely related. Separate ITS (**Figure 4**) and plastid analysis (**Figure 5**) fail to resolve the systematic positions of the four Upper Guinea (UG) clade species (i.e., including *C. humilis* and *C. togoensis*), but a combined analysis of these data sets retrieves monophyly monophyly for *C. affinis* and *C. stenophylla* (**Figure 6**). *Coffea affinis* and *C. stenophylla* share specific characters, including: the habit of a small tree, obovate leaves with a distinct apical tip (acumen) and drying green, 2 to 4 flowers per axil, 6- to 8-merous flowers (i.e., six to eight corolla lobes and anthers per flower; and black or black–purple fruits (see **Figure 2**). *Coffea togoensis*, from Ghana and Togo, is also a small tree, and has leaves with a distinct apical tip (acumen), 6- to 8-merous flowers and black fruits, but generally has elliptic leaves, drying grayish or gray-green, 1 or 2 flowers per axil, and smaller fruits and seeds. *Coffea humilis* is unlike *C. affinis*, *C. stenophylla* and *C. togoensis,* as it is a monocaul dwarf (single-stemmed woody plant, up to 1 m high), with large (up to 22 cm long) obovate leaves, 5- to 7-merous flowers and red fruits.

Coffea stenophylla and *C. affinis* exhibit considerable morphological variation, particularly with regard to leaf size and shape. Some examples of *C. stenophylla* approach *C. affinis* in terms of leaf shape and dimensions. Further work, including morphological and more detailed molecular study, is required to determine the precise relationship between these two species. There could be grounds for subsuming *C. affinis* within *C. stenophylla,* for example as *C. stenophylla* var. *camaya* (Portères, 1937a).

Our historical accession of farmed *C. affinis* from a coffee plantation in Sierra Leone (two genotypes from a single accession [Cope *s.n.*, 7 iii 1912 (K)] is not related to either *C. affinis* or *C. stenophylla*. ITS and plastid marker data place this accession in the Lower Guinea/Congolian Clade. The specimens of this accession are a reasonable morphological match for *C. affinis*, with leaves of the same dimensions and possessing an acuminate leaf tip, but the seeds are somewhat larger and narrower than *C. affinis*; flower morphology and fruits color are unknown. The Cope *s.n.* accession is compelling, as it placed with the Liberica Alliance but does not conform to any of the known species in this alliance. The distinct, elongated leaf tip (acumen) immediately sets it apart from all variants of *C. liberica* (Stoffelen, 1998).

In both ITS (**Figure 4**) and plastid (**Figure 5**) analyses *C. liberica* is not monophyletic. Further molecular data is required for *C. liberica* and closely related taxa. As currently circumscribed, *C. liberica* is a highly polymorphic (Bridson, 1985, 1988; Stoffelen, 1998; Davis et al., 2006) encompassing a broad range of morphological variation.

Identification of *Coffea* Species Hybrids

Identification of interspecies hybrids or introgressed plants on the basis of incongruence between biparental (e.g., nuclear) and uniparental (plastid) phylogenetic trees is well established (Linder and Rieseberg, 2004). In most flowering plants plastid DNA is maternally inherited, whereas the nuclear DNA and the ITS region of ribosomal DNA is not (Chase et al., 2005). Previous analysis of ITS and plastid markers in *Coffea* have been used to identify the parents of the allotetraploid *C. arabica* (*C. eugenioides* × *C. canephora*) and the diploid hybrid *C. eugenioides* × *C. liberica* (Maurin et al., 2007). As part of this study, we undertook further tests of this method using an additional interspecies hybrids of known crossing history, viz. *C. arabica* × *C. racemosa* (Medina Filho et al., 1977a,b), which demonstrated the utility of the method, at least for recently produced hybrids (**Figures 4**, **5**). Using this method, we were able to identify the interspecies hybrid *C. liberica* × *C. stenophylla* (**Figures 4**, **5**) from cultivated material collected by us in Sierra Leone (see **Table 1**). The hybrid *C. liberica* × *C. stenophylla* has been reported before (Cramer, 1957), but never authenticated. The accession of this hybrid from Sierra Leone appears to be partially or totally sterile; in some years it produces a few fruits, but neither the viability nor fertility of seeds have been tested. The low level of fruit setting suggests that this hybrid is a diploid ($2n = 22$) rather than a polyploid; tetraploids generally have higher fertility or restored fertility (Carvalho and Monaco, 1968; Charrier, 1978). We have no means of knowing whether this hybrid was the result of a cross in cultivation or in the wild. Either origin is plausible: both species were grown together in a number of research stations in Africa, and Asia (Cramer, 1957); and *in situ* crossings have been reported. In Ivory Coast hybrid seedlings (but not mature plants) of *C. liberica* × *C. stenophylla*, were detected in the wild (Berthaud, 1986). Generally, interspecific hybridization in natural populations of *Coffea* is a rare phenomenon (Charrier, 1978; Berthaud, 1986).

Interspecies hybrids, once fertility (and thus yield) is restored via conversion to the tetraploid ($4n = 44$) state (Carvalho and Monaco, 1968; Nagai et al., 2008), are valuable for coffee crop development, as they provide the possibility of introducing useful traits. For example, CLR resistance (*C. canephora* × *C. arabica*; Clarindo et al., 2013; Avelino et al., 2015) and leaf-miner resistance (*C. arabica* × *C. racemosa*; Medina Filho et al., 1977b) for Arabica coffee. In our DNA survey of long-styled African

species [Coffee Crop Wild Relative Priority Groups I and II (Davis et al., 2019)] we confirm that hybridization is possible across all the major African coffee clades (lineages), indicating the potential to create custom interspecies hybrids across this wide spectrum of *Coffea* species diversity.

CONCLUSION

Coffea affinis and *C. stenophylla* may possess useful attributes for coffee crop plant development, including taste, disease resistance, and climate resilience. These attributes would be best accessed via breeding programs, including those involving interspecies crossing, followed by tetraploidization. Here, we confirm that (initial) hybridization is possible across all the major clades of long-styled African *Coffea* species (Maurin et al., 2007; Davis et al., 2011; Hamon et al., 2017), i.e., Coffee Crop Wild Relative (Priority) Groups I and II (Davis et al., 2019). For *C. stenophylla* we confirm via DNA sequencing that a cross can be made with *C. liberica*, supporting the work of Louarn (1992), who also demonstrated that *C. stenophylla* can be crossed with *C. canephora*, *C. congensis*, and *C. pseudozanguebariae*. Development of *C. stenophylla* and *C. affinis* via minimal domestication (e.g., the selection of trait-specific genotypes) may be possible, although this route would probably only be feasible for high-value markets, such as the upper end of the speciality coffee sector, based on the historical reports of its superior taste. Productivity (green coffee yield) appears to be lower than the major commercial species (*C. arabica*, *C. canephora* and *C. liberica*). A key caveat here, is that *C. stenophylla* has not undergone sensory or agronomic evaluation in a contemporary setting. Despite the shortfall in our understanding of these species, the available evidence, as summarized and reviewed here, is more than sufficient to warrant further research for *C. affinis* and *C. stenophylla*, and to take measures to ensure their survival in the wild (*in situ*) and in cultivation (*ex situ*). Deforestation, and other forms of land-use change, are threatening the survival of these species in the wild, in Sierra Leone, Ivory Coast and Guinea.

The decline in the use of *C. stenophylla* use as a crop plant in Upper West Africa was dramatic, from once widespread (although comparatively small scale) use in the late and early parts of the 19th and 20th centuries, to apparently nothing today. Our survey of local farming communities in Sierra Leone, reported an absence of indigenous knowledge for this species. One of the other main reasons for the decline in its use may have been the considerable agronomic and commercial success of robusta coffee (*C. canephora*), which was introduced into global cultivation around the same time as *C. stenophylla* and greatly surpassed the comparatively meagre productivity of other underutilized coffee species (Davis et al., 2019). Following on from our fieldwork in Sierra Leone (2018–2020), wild stock of *C. affinis* and *C. stenophylla* is now being propagated in quantity, for sensory and agronomic evaluation, and to safeguard its existence. Field work in Guinea and Ivory Coast are required to further ascertain the present day indigenous and cultivated (farmed) status, respectively, and to provide a conservation management plan to ensure its survival in the wild. An in-depth review of coffee research collections, including genotyping (genome banking), is required to formulate an effective *ex situ* conservation management strategy for *C. affinis* and *C. stenophylla*, and indeed many other coffee species. African coffee species provide key resources for the sustainability of the global coffee sector (Davis et al., 2019) and should receive appropriate conservation measures in the wild (*in situ*) and in cultivation (*ex situ*).

DATA AVAILABILITY STATEMENT

The accession numbers for the sequencing data presented in this article can be found in **Table 2** within the article.

AUTHOR CONTRIBUTIONS

AD devised the non-fieldwork elements of the study, undertook the literature and herbarium, participated in fieldwork, and was the lead on writing the manuscript. RG undertook the DNA sequencing and analyses, and contributed toward the writing of the manuscript. MF helped to devise the study, provided assistance with DNA sequencing, and contributed toward the writing of the manuscript. DS undertook the bulk of the fieldwork in Sierra Leone. JH devised the overall framework of the study, devised and undertook substantial fieldwork in Sierra Leone, and contributed toward the writing of the manuscript.

FUNDING

Funding for this research was provided from different sources including the EU funded Agriculture for Development Programme - Robusta Coffee Development Project contract FED/2013/322-213 administered by the Government of Sierra Leone; and a Darwin Initiative Scoping Project DARSC196 - Conservation and use of native coffee species in Sierra Leone.

ACKNOWLEDGMENTS

We acknowledge the support and collaboration of the Department for Forestry, Ministry of Agriculture, and Sierra Leone Agricultural Research Institute, in Sierra Leone. We would like to offer our sincere gratitude to Welthungerhilfe (WHH), and to staff at their offices in Freetown and Kenema (Sierra Leone) for various support during this project. At WHH we would particularly like to thank Franz Moestl, Manfried Bischofberger, and Derek Wambulwa Makokha. We would also like to thank: Arnaud Havet, at WARC, in Sierra Leone; Martin Cheek at the Royal Botanic Gardens, Kew (UK) for advice on deforestation in Upper West Africa and for showing us project material from Guinea; Charles Denison for providing images of *C. stenophylla*; and Iris Wiegant and László Csiba for their assistance with lab work at Kew.

REFERENCES

Anthony, F. (1982). *La Conservation des Ressources Génétiques chez les Caféiers Prospection, Gestion et Evaluation du Materiel en Collection en Cote d'Ivoire*. Paris: ORSTOM, 15.

Anthony, F. (1992). *Les Ressources Génétiques des Caféiers: Collecte, Gestion d'un Conservatoire et Évaluation de la Diversité Génétique. Série TDM 8 I*. Paris: ORSTOM.

Anthony, F., Dussert, S., and Dulloo, M. E. (2007). *II. Coffee Genetic Resources. Conserving Coffee Genetic Resources: Complementary Strategies for ex Situ Conservation of Coffee (Coffea Arabica L.) Genetic Resources*. Rome: CATIE.

Avelino, J., Cristancho, M., Georgiou, S., Imbach, P., Aguilar, L., Bornemann, G., et al. (2015). The coffee rust crises in Colombia and Central America (2008–2013): impacts, plausible causes and proposed solutions. *Food Secur.* 7, 303–321. doi: 10.1007/s12571-015-0446-9

Berthaud, J. (1983). *Liste du Matériel Provenant des Prospections de Côte D'ivoire. Document de Travail*. Paris: ORSTOM, 17.

Berthaud, J. (1986). *Les Ressources Génétiques Pour L'amélioration des Cafeiers Africains Diploïdes. ORSTOM, Série TDM 188*. Paris: ORSTOM.

Bramel, P., Krishnan, S., Horna, D., Lainoff, B., and Montagnon, C. (2017). *Global Conservation Strategy For Coffee Genetic Resources*. Washington, DC: Crop Trust & World Coffee Research.

Bridson, D. M. (1985). The lectotypification of *Coffea liberica* (Rubiaceae). *Kew Bull.* 40, 805–807.

Bridson, D. M. (1988). "Coffea," in *Flora of Tropical East Africa, Rubiaceae, Part 2*, eds R. M. Polhill, D. M. Bridson, and B. Verdcourt (Avereest: Balkema, Rotterdam/Brookfield), 703–723.

Carvalho, A., and Monaco, L. C. (1968). Relaciones genéticas de especies seleccionadas de *Coffea*. *Cafe* 9, 3–19.

Charrier, A. (1978). *La Structure Génétique des Caféiers Spontanés de la Région Malgache (Mascarocoffea)*. Paris: ORSTOM.

Chase, M. W., Salamin, N., Wilkinson, M., Dunwell, J. M., Kesanakurthi, R. P., Haider, N., et al. (2005). Land plants and DNA barcodes: short-term and long-term goals. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 360, 1889–1895. doi: 10.1098/rstb.2005.1720

Cheney, R. H. (1925). *A Monograph of the Economic Species of the Genus Coffea L.* New York, NY: The New York University Press.

Chevalier, A. (1905). Les caféiers sauvages de la Guinée française. Note de M.A.Chevalier, présentée par M. Ph. van Tieghem. C. R. *Hebd. Acad. Sci.* 140, 1472–1475.

Chevalier, A. (1929). Les caféiers du globe, fasc. 1: généralités sur les caféiers. *Ency. Biol.* 5, 1–196.

Chevalier, A. (1947). Les caféiers du globe, fasc. 3: systématique des caféiers et faux-caféiers. Maladies et insectes nuisibles. *Ency. Biol.* 28:212.

Clarindo, W. R., Carvalho, C. R., Eveline, T. C., and Koehler, A. D. (2013). Following the track of "Híbrido de Timor" origin by cytogenetic and flow cytometry approaches. *Genet. Resour. Crop Evol.* 60, 2253–2259. doi: 10.1007/s10722-013-9990-3

Couch, C., Cheek, M., Haba, P., Molmou, D., Williams, J., Magassouba, S., et al. (2019). *Threatened Habitats and Tropical Important Plant Areas (TIPAs) of Guinea, West Africa*. Kew: Royal Botanic Gardens.

Cramer, P. J. S. (1957). *A Review of Literature of Coffee Research in Indonesia*. Turrialba: Inter-American Institute of Agricultural Sciences.

Davis, A. P., Chadburn, H., Moat, J., O'Sullivan, R., Hargreaves, S., and Nic Lughadha, E. (2019). High extinction risk for wild coffee species and implications for coffee sector sustainability. *Sci. Adv.* 5:eaav3473. doi: 10.1126/sciadv.aav3473

Davis, A. P., Govaerts, R., Bridson, D. M., and Stoffelen, P. (2006). An annotated taxonomic conspectus of the genus *Coffea* (Rubiaceae). *Bot. J. Linn. Soc.* 152, 465–512. doi: 10.1111/j.1095-8339.2006.00584.x

Davis, A. P., and Rakotonasolo, F. (2009). A taxonomic revision of the baracoffea alliance: nine remarkable *Coffea* species from western Madagascar. *Bot. J. Linn. Soc.* 158, 355–390. doi: 10.1111/j.1095-8339.2008.00936.x

Davis, A. P., Tosh, J., Ruch, N., and Fay, M. (2011). Growing coffee: *Psilanthus* (Rubiaceae) subsumed on the basis of plastid and nuclear DNA sequences; implications for the size, morphology, distribution and evolutionary history of *Coffea*. *Bot. J. Linn. Soc.* 167, 357–377. doi: 10.1111/j.1095-8339.2011.01177.x

De Wildeman, E. (1904). Noveaux caféiers de la cote occidentale d'afrique. *Agric. Prat. Pays Chauds* 4, 113–116.

De Wildeman, E. (1906a). *Mission Émile Laurent (1903–1904), Vol. 1 (Texte)*. Brussels: F. Vandenbuggenhoudt.

De Wildeman, E. (1906b). *Mission Émile Laurent (1903–1904), Vol. 2 (Planches)*. Bruxelles: F. Vandenbuggenhoudt.

Don, G. (1834). *General System of Gardening and Botany* (Vol. 3). London: Gilbert and Rivington Printers, St. John's Square.

Doyle, J. J., and Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phyt. Bull. Bot. Soc. Amer.* 19, 11–15.

Dudgeon, G. C. (1922). *Imperial Institute Handbooks. The Agricultural and Forest Products of British West Africa*, 2nd Edn. London: John Murray.

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucl. Acid. Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340

Engelmann, F., Dulloo, M. E., Astorga, C., Dussert, S., and Anthony, F. (2007). *Conserving Coffee Genetic Resources: Complementary Strategies for ex Situ Conservation of Coffee (Coffea Arabica L.) Genetic Resources*. Rome: CATIE.

Google Earth (2019). *Google Earth Pro, Version 7.3.1*. Available online at: https://www.google.com/intl/en_uk/earth/desktop (accessed June 30, 2019).

Govaerts, R., Rhusam, M., Frodin, D., Bridson, D., Dawson, S., and Davis, A. P. (2019). *World Checklist of Rubiaceae. Facilitated by the Royal Botanic Gardens, Kew. Published on the Internet*. Available online at: http://wcsp.science.kew.org/ (accessed December 18, 2019).

Haarer, A. E. (1962). *Modern Coffee Production*. London: Leonard Hill [Books] Limited.

Hamon, P., Grover, C., Davis, A., Rakotomalala, J.-J., Raharimalala, N., Albert, V., et al. (2017). Genotyping-by-sequencing provides the first well-resolved phylogeny for coffee (*Coffea*) and insights into the evolution of caffeine content in its species. *Mol. Phylogenet. Evol.* 109, 351–361. doi: 10.1016/j.ympev.2017.02.009

Hiern, W. P. (1876). On the African species of the genus *Coffea*. *Trans. Linn. Soc. London Bot.* 1, 169–176. doi: 10.1111/j.1095-8339.1876.tb00033.x

Holmgren, P. K., Holmgren, N. H., and Barnett, L. C. (1990). *Index Herbariorum. Part 1: The Herbaria of the World, 8th edn. Regnum Vegetabile*. New York, NY: New York Botanical Garden.

Hooker, J. D. (1896). Coffea stenophylla. *Curtis Bot. Mag* 122:7475.

Huelsenbeck, J. P., and Ronquist, F. (2001). MrBAYES: bayesian inferenceof phylogenetic trees. *Bioinformatics* 17, 754–755. doi: 10.1093/bioinformatics/17.8.754

International Coffee Organization (ICO) (2019). *Trade Statistics*. Available online at: http://www.ico.org/trade_statistics.asp (accessed October 7, 2019).

International Union for Conservation of Nature (IUCN) (2020). *The IUCN Red List of Threatened Species*. Available online at: http://www.iucnredlist.org/ (accessed November 14, 2018).

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., et al. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647–1649. doi: 10.1093/bioinformatics/bts199

Le Pierrès, D., Charmetant, P., Yapo, A., Leroy, T., Couturon, E., Bontems, S., et al. (1989). *Les Caféiers Sauvages de Côte d'Ivoire et de GUINÉE: Bilan des Missions de Prospection Effectuées de 1984 à 1987. Treizième Colloque Scientifique International sur le Café*. Paris: ASIC, 420–428.

Linder, C. R., and Rieseberg, L. H. (2004). Reconstructing patterns of reticulate evolution in plants. *Am. J. Bot.* 91, 1700–1708. doi: 10.3732/ajb.91.10.1700

Louarn, J. (1992). *La fertilite des Hybrides Interspecifiques et Les Relations Genomiques Entre Cafeiers Diploides D'origine Africaine (Genre Coffea l. Sous-Genre Coffea)*. Ph.D. Thesis. Universite de Paris Sud, Centre d'Orsay, Paris.

Macmillan, H. F. (1914). *A Handbook of Tropical Gardening and Planting*, 2nd Edn. Colombo: Amen Corner: H.W. Cave & Co.

Maurin, O., Davis, A. P., Chester, M., Mvungi, E. F., Jaufeerally-Fakim, Y., and Fay, M. F. (2007). Towards a phylogeny for *Coffea* (Rubiaceae): identifying well-supported lineages based on nuclear and plastid DNA sequences. *Ann. Bot.* 100, 1565–1583. doi: 10.1093/aob/mcm257

Medina Filho, H. P., Carvalho, A., and Medina, D. M. (1977a). Germoplasma de *C. racemosa* e seu potencial no melhoramento do cafeeiro. *Bragantia* 36, XLIII–XLVI.

Medina Filho, H. P., Carvalho, A., and Monaco, L. (1977b). Melhoramento do cafeeiro. XXXVII – Observações sobre a resistência do cafeeiro ao bicho mineiro. *Bragantia* 36, 131–137. doi: 10.1590/s0006-87051977000100011

Nagai, C., Rakotomalala, J.-J., Katahira, R., Li, Y., Yamagata, K., and Ashihara, H. (2008). Production of a new low-caffeine hybrid coffee and the biochemical mechanism of low caffeine accumulation. *Euphytica* 164, 133–142. doi: 10.1007/s10681-008-9674-9

Portères, R. (1937a). Etude sur les caféiers spontanés de la section "Des Eucoffeae". Leur répartition, leur habitat, leure mise en culture et leur sélection en Cote d'Ivoire. Deuxième partie: Espèce, variétés et formes. *Ann. Agric. Afr. Occid.* 1, 219–283.

Portères, R. (1937b). Etude sur les caféiers spontanés de la section "Des Eucoffeae". Leur répartition, leur habitat, leure mise en culture et leur sélection en Cote d'Ivoire. Première partie: Répartition et habitat. *Ann. Agric. Afr. Occid.* 1, 68–91.

Posada, D. (2008). jModelTest: phylogenetic model averaging. *Molec. Biol. Evol.* 25, 1253–1256. doi: 10.1093/molbev/msn083

Rambaut, A. (2018). *FigTree, Version 1.4.4*. Available online at: http://github.com/rambaut/figtree/ (accessed February 18, 2020).

Rambaut, A., Drummond, A. J., and Suchard, M. (2013). *Tracer, Version 1.6*. Available online at: http://tree.bio.ed.ac.uk/software/ (accessed February 18, 2020).

Ronquist, F., and Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogeneticinference under mixed models. *Bioinformatics* 19, 1572–1574. doi: 10.1093/bioinformatics/btg180

Sayer, J. A., Harcourt, C. S., and Colins, N. M. (1992). *The Conservation Atlas of Tropial Forests: Africa*. Cambridge: IUCN.

Scott Elliot, G. F. (1893). Economic Plants of Sierra Leone. *Bull. Misc. Inform. (Kew)* 79, 167–169.

Stoffelen, P. (1998). *Coffea and Psilanthus in Tropical Africa: A Systematic and Palynological Study, Including a Revision of the West and Central African Species*. Leuven: Katholieke Universiteit Leuven.

Thiers, B. (2019). *Index Herbariorum: A Global Directory of Public Herbaria and Associated Staff. New York Botanical Garden's Virtual Herbarium*. Available online at: http://sweetgum.nybg.org/science/ih/ (accessed September 11, 2019).

Tothill, J. D. (1940). *Agriculture in Uganda*. London: Oxford University Press.

Vega, F. E., Rosenquist, E., and Collins, W. (2003). Global project needed to tackle coffee crisis. *Nature* 425:343. doi: 10.1038/425343a

Watt, G. (1908). *The Commerical Products of India*. London: John Murray.

Wellman, F. L. (1961). *Coffee: Botany, Cultivation and Utilization*. New York, NY: Leonard Hill [Books] Limited/Interscience Publishers, Inc.

Wrigley, G. (1988). *Coffee – Tropical Agriculture Series*. New York, NY: Longman Scientific & Technical.

Conflict of Interest: 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 © 2020 Davis, Gargiulo, Fay, Sarmu and Haggar. 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) and the copyright owner(s) 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 Plant Science

ORIGINAL RESEARCH
published: 25 May 2020
doi: 10.3389/fpls.2020.00624

A Chromosome-Scale Assembly of the Garden Orach (*Atriplex hortensis* L.) Genome Using Oxford Nanopore Sequencing

Spencer P. Hunt[1], David E. Jarvis[1], Dallas J. Larsen[1], Sergei L. Mosyakin[2], Bozena A. Kolano[3], Eric W. Jackson[4], Sara L. Martin[5], Eric N. Jellen[1] and Peter J. Maughan[1]*

[1] Department of Plant and Wildlife Sciences, Brigham Young University, Provo, UT, United States, [2] M.G. Kholodny Institute of Botany, National Academy of Sciences of Ukraine, Kyiv, Ukraine, [3] Institute of Biology, Biotechnology and Environmental Protection, Faculty of Natural Sciences, University of Silesia in Katowice, Katowice, Poland, [4] 25:2 Solutions, Rockford, MN, United States, [5] Agriculture and Agri-Food Canada, Ottawa Research and Development Centre, Ottawa, ON, Canada

OPEN ACCESS

Edited by:
Eric Von Wettberg, University of Vermont, United States

Reviewed by:
Namshin Kim, Korea Research Institute of Bioscience and Biotechnology (KRIBB), South Korea
Shanshan Chen, Zhengzhou University, China

***Correspondence:**
Eric N. Jellen
jellen@byu.edu

Specialty section:
This article was submitted to Plant Breeding, a section of the journal Frontiers in Plant Science

Received: *28 February 2020*
Accepted: *22 April 2020*
Published: *25 May 2020*

Citation:
Hunt SP, Jarvis DE, Larsen DJ, Mosyakin SL, Kolano BA, Jackson EW, Martin SL, Jellen EN and Maughan PJ (2020) A Chromosome-Scale Assembly of the Garden Orach (Atriplex hortensis L.) Genome Using Oxford Nanopore Sequencing. Front. Plant Sci. 11:624. doi: 10.3389/fpls.2020.00624

Atriplex hortensis (2*n* = 2*x* = 18, 1C genome size ∼1.1 gigabases), also known as garden orach and mountain-spinach, is a highly nutritious, broadleaf annual of the *Amaranthaceae-Chenopodiaceae* alliance (*Chenopodiaceae* sensu stricto, subfam. *Chenopodioideae*) that has spread in cultivation from its native primary domestication area in Eurasia to other temperate and subtropical regions worldwide. *Atriplex* L. is a highly complex but, as understood now, a monophyletic group of mainly halophytic and/or xerophytic plants, of which *A. hortensis* has been a vegetable of minor importance in some areas of Eurasia (from Central Asia to the Mediterranean) at least since antiquity. Nonetheless, it is a crop with tremendous nutritional potential due primarily to its exceptional leaf and seed protein quantities (approaching 30%) and quality (high levels of lysine). Although there is some literature describing the taxonomy and production of *A. hortensis*, there is a general lack of genetic and genomic data that would otherwise help elucidate the genetic variation, phylogenetic positioning, and future potential of the species. Here, we report the assembly of the first high-quality, chromosome-scale reference genome for *A. hortensis* cv. "Golden." Long-read data from Oxford Nanopore's MinION DNA sequencer was assembled with the program Canu and polished with Illumina short reads. Contigs were scaffolded to chromosome scale using chromatin-proximity maps (Hi-C) yielding a final assembly containing 1,325 scaffolds with a N50 of 98.9 Mb – with 94.7% of the assembly represented in the nine largest, chromosome-scale scaffolds. Sixty-six percent of the genome was classified as highly repetitive DNA, with the most common repetitive elements being Gypsy- (32%) and Copia-like (11%) long-terminal repeats. The annotation was completed using MAKER which identified 37,083 gene models and 2,555 tRNA genes. Completeness

of the genome, assessed using the Benchmarking Universal Single Copy Orthologs (BUSCO) metric, identified 97.5% of the conserved orthologs as complete, with only 2.2% being duplicated, reflecting the diploid nature of *A. hortensis*. A resequencing panel of 21 wild, unimproved and cultivated *A. hortensis* accessions revealed three distinct populations with little variation within subpopulations. These resources provide vital information to better understand *A. hortensis* and facilitate future study.

Keywords: *Amaranthaceae*, *Atriplex hortensis*, Hi-C, orach, orphan crop, proximity-guided assembly

INTRODUCTION

Atriplex hortensis L. ($2n = 9x = 18$), also known as garden orach or mountain-spinach, is a highly nutritious, leafy annual plant. It is a moderately xero-halophytic species that is resistant to salinity, a wide range of temperatures, and drought. Originating in Eurasia, *A. hortensis* has been a minor vegetable food source in multiple areas of the Trans-Himalayan region and has since become naturalized throughout the Americas. It exhibits incredible variation in pigmentation as a result of its variable content of betalains, as well as substantial differences in height and seed production (Tanaka et al., 2008; Simcox and Stonescu, 2014).

Atriplex hortensis has been recognized for its medicinal properties which were shown to improve digestion, increase circulation and boost the immune system (Rinchen et al., 2017). Additionally, *A. hortensis* has been used in land rehabilitation projects because of its ability to establish well, grow rapidly, reduce soil erosion and compete with native plants (McArthur et al., 1983; Simon et al., 1994; Wright et al., 2002). As a result, *A. hortensis* is important for both domestic and wild browsing animals where other forage crops are lacking. Despite its affinity for low to moderate saline areas where it has little competition from non-halophytes, *A. hortensis* can also grow where total soluble salts are low, making it well suited to a multitude of different environments (Welsh and Crompton, 1995).

As the world continues to search for new ways to feed its ever-growing population, new food sources have gained popularity that have helped provide diversity to diets while capitalizing on less desirable, underutilized or even fallow landscapes for agriculture. Given its xero-halophytic characteristics, *A. hortensis* is an intriguing candidate for contributing to world food security, especially in areas rich in saline soils. In comparison to other leafy vegetable crops, *A. hortensis* seeds and leaves are both edible and have protein contents of 26% (dry weight) in seeds, which is comparable to some legumes (Wright et al., 2002), and 35% (dry weight) in leaves, which is higher than spinach (*Spinacea oleracea* L.), a close relative of *A. hortensis* also belonging to the same subfamily *Chenopodioideae*, but to a different tribe (*Anserineae*; see Fuentes-Bazan et al., 2012). However, the seeds contain antinutritional saponins that must be removed by washing and/or seed abrasion. In this respect *A. hortensis* resembles its distant relative quinoa (*Chenopodium quinoa* Willd.); the name recently formally proposed for nomenclatural conservation (Mosyakin and Walter, 2018), which also contains saponins. Interestingly, sweet varieties of quinoa have been identified that have a nonsense mutation in the regulator of the saponin biosynthetic pathways (Jarvis et al., 2017) – suggesting similar pathways could be targeted to remove antinutritional saponins in *A. hortensis*. The seeds of *A. hortensis* have higher fat, ash, fiber and lysine contents than most cereal grains (Wright et al., 2002). Its high protein content, which includes an essential amino acid profile that meets the WHO and UN-FAO recommended adult levels, also makes *A. hortensis* very attractive as a novel protein source.

Atriplex hortensis belongs to the family *Chenopodiaceae* in the strict sense, which is now often included in the extended family *Amaranthaceae* sensu lato; this group (*Chenopodiaceae*+*Amaranthaceae*) is phylogenetically nested in the core clade of the order *Caryophyllales*, which in turn, belongs to core eudicots, the largest and most diverse clade of angiosperms (for an overview of high-level phylogeny of the group, see Hernández-Ledesma et al., 2015; APG IV, 2016, and references therein).

The merger of the traditionally recognized families *Chenopodiaceae* and *Amaranthaceae* sensu stricto into one family under the priority name *Amaranthaceae* sensu lato proposed already in the first version of the APG system (APG (Angiosperm Phylogeny Group), 1998) remained unchanged in all other APG modifications (see, APG IV, 2016, and references therein). It was widely followed by many researchers and users of botanical nomenclature, but usually not by the experts in taxonomy of *Chenopodiaceae* (s. str.), who mainly continued to accept the two families. Not discussing here the reasons of and arguments for the two concepts of familial and subfamilial delimitation in the *Amaranthaceae*/*Chenopodiaceae* alliance (which will be discussed in a separate article, now in progress), we, however, note that the merger of the two families resulted in some confusion and miscommunication in recent literature regarding the usage of family names, and especially names of infrafamilial suprageneric entities (such as subfamilies and tribes). For example, some authors use the subfamily name *Amaranthoideae* in its traditional sense for just one group in *Amaranthaceae* s. str., while others may use it to cover all formerly recognized groups in *Amaranthaceae* s. str. (including *Amaranthoideae*, *Gomphrenoideae*, etc.). To avoid any uncertainty, we conventionally use in the

Abbreviations: BLAST, basic local alignment search tool; Mbp, megabases; MYA, million years ago; ONT, Oxford nanopore technology; SNP, single nucleotide polymorphisms.

present article the following nomenclature (both formal and informal names): (1) the group uniting *Chenopodiaceae* and *Amaranthaceae* s. str. (forming together the extended *Amaranthaceae* sensu APG) is referred to under an informal designation "*Amaranthaceae*/*Chenopodiaceae* aliance;" (2) the family-rank names *Amaranthaceae* and *Chenopodiaceae* refer to the groups corresponding to the two traditionally recognized families; and (3) the sufbamily-rank name *Chenopodioideae* (in paralel with other recognized subfamilies, such as *Betoideae*, *Salsoloideae*, etc.) corresponds to just one group of *Chenopodiaceae* s. str., but not to the group covering the whole family *Chenopodiaceae* in its traditional circumscription; similarly, *Amaranthoideae* refers to the subfamily-rank subdivision of *Amaranthaceae* s. str., comparable to *Gomphrenoideae*.

Recent molecular phylogenetic and taxonomic studies have led to considerable improvements in taxonomy and in our understanding of phylogenetic relationships in the order *Caryophyllales* in general and *Atriplex* and its closer relatives in particular (Kadereit et al., 2010; Zacharias and Baldwin, 2010; Brignone et al., 2019; Morales-Briones et al., 2020). However, few molecular studies have been focused specifically on *A. hortensis* in recent years.

As it is viewed now, *Atriplex* is nested in the larger clade corresponding to the tribe *Atripliceae* (including *Chenopodieae*, which is the correct name for the group if placed in *Chenopodiaceae*, not *Amaranthaceae* s.l.) as outlined by Fuentes-Bazan et al. (2012), and/or to a smaller clade corresponding to the tribe *Atripliceae* in a narrower sense, as outlined by Kadereit et al. (2010). The *Atripliceae* in the narrow sense is sister to another clade (informally called *Chenopodieae* I; see Kadereit et al., 2010) containing *Chenopodium* s. str. (including Australian *Rhagodia* R. Br. and *Einadia* Raf.; see Fuentes-Bazan et al., 2012; Mosyakin and Iamonico, 2017) in its much restricted sense, excluding taxa formerly placed in *Chenopodium* sensu lato but now recognized in phylogenetically more distant genera *Blitum* L. (which is close to *Spinacia* L.), *Chenopodiastrum* S. Fuentes, Uotila & Borsch, *Dysphania* R. Br., *Lipandra* Moq., *Oxybasis* Kar. & Kir., and *Teloxys* Moq. (see Fuentes-Bazan et al., 2012; Hernández-Ledesma et al., 2015).

The clade of *Atripliceae* (sensu Kadereit et al., 2010) contains two main subclades (informally named as the *Archiatriplex*-clade and *Atriplex*-clade) with several smaller lineages, some of which are currently recognized as separate genera. As circumscribed now, the phylogenetically coherent and monophyletic *Atriplex* includes several groups that were earlier described and recognized as separate genera, such as *Obione* Gaertn. and some Australian and North American groups. Despite morphological distinctiveness of some of those groups, they are phylogenetically deeply rooted in *Atrilpex* and thus their recognition as separate genera is not recommended. In contrast, several genera are recognized in the *Archiatriplex*-clade, namely *Archiatriplex* G.L.Chu, *Exomis* Fenzl ex Moq., *Extriplex* E.H. Zacharias, *Grayia* Hook. & Arn., *Holmbergia* Hicken, *Manochlamys* Aellen, *Microgynoecium* Hook. f., *Proatriplex* (W.A.Weber) Stutz & G.L. Chu, and *Stutzia* E.H. Zacharias (Kadereit et al., 2010; Brignone et al., 2019). They represent relicts of earlier diversification events in the group. Also, some additional early-branching ("basal") lineages of the *Atriplex*-clade can also probably be recognized as separate genera. For example, in addition to the currently recognized genera *Halimione* Aellen and *Atriplex* s. str. (Kadereit et al., 2010), such groups as *Cremnophyton* Brullo & Pavone from Malta containing *C. lafrancoi* Brullo & Pavone (=*Atriplex lafrancoi* (Brullo & Pavone) G. Kadereit & Sukhorukov; see Kadereit et al., 2010) and the mainly Central Asian *Sukhorukovia* Vasjukov (2015) with *S. cana* (C.A. Mey.) Vasjukov (*Atriplex cana* C.A. Mey. = *Cremnophyton canum* (C.A. Mey.) G.L. Chu) may be probably assigned the generic rank after further research.

Since *A. hortensis* is the nomenclatural type of the genus, it naturally belongs to *Atriplex* subgen. *Atriplex* sect. *Atriplex* (Art. 22.1 of the *International Code for Nomenclature of algae, fungi and plants*: Turland et al., 2018). This section houses at least two other species, *A. sagittata* Borkh. (earlier often known under the synonymous name *A. nitens* Schkuhr) and *A. aucheri* Moq., which seem to be most closely related to *A. hortensis* (Sukhorukov, 2006). The clade of *A. hortensis* and its two close relatives belongs to the grade of early-branching clades of *Atriplex* s. str. containing taxa with C_3 photosynthesis (Brignone et al., 2019).

The geographic and taxonomic origins of domesticated *A. hortensis* remain elusive because at present the species is known mostly (or exclusively?) in cultivated and escaped (and locally naturalized) populations. It probably originated somewhere within the geographic ranges of its closest relatives, in Central Asia or adjacent regions, or it could be native in the Mediterranean region and/or Asia Minor (Sukhorukov, 2014).

Although recent studies have tested the limits of salt-tolerance of *A. hortensis* (Vickerman et al., 2002; Sai Kachout et al., 2011), there has been little to no research conducted to develop genetic tools necessary for accelerating *A. hortensis* breeding. One phenotypic characteristic in need of improvement for seed production is the panicle, which consists of two types of flowers usually mixed on the same plant. One type produces 3–5 mm diameter seed that are encased within large, papery bracteoles that are not retained well under windy conditions at maturity. The other flower type produces 1–2 mm black fruits/seeds that have no bracteoles but are instead subtended by easily removed tepals.

To better understand the underlying genetic basis of the xero-halophytic, nutritive and unique pigmentation characteristics of *A. hortensis*, and to more accurately assess phylogenetic relationships within its family and genus, we sequenced the *A. hortensis* genome. We show that ultra-long reads produced by the portable, real-time Oxford Nanopore Technology (Oxford, United Kingdom) MinION sequencing system (Lu et al., 2016) with short-read polishing and chromatin-contact mapping is an effective approach to generate a high-quality genome assembly in a moderately large and complex genome of a diploid plant species. We annotated the genome with a deeply

sequenced transcriptome from various *A. hortensis* plant tissues, and we demonstrate the quality of the chromosome-level genome assembly and annotation using Benchmarking Universal Single-Copy Orthologs (BUSCO) (Simão et al., 2015) to assess the completeness of the assembled genome. Genomic comparison to other Caryophyllales within the *Amaranthaceae-Chenopodiaceae* family identified highly syntenic and orthologous chromosomal relationships. Together, these resources provide an initial, important foundation for accelerated genetic improvement to neodomesticate this potentially valuable crop.

MATERIALS AND METHODS

Plant Material

Atriplex hortensis cv. "Golden" was obtained from Wild Garden Seed (Philomath, Oregon) and used for whole-genome sequencing and assembly. Sterilized seed were grown hydroponically in a growth chamber at BYU. An 11-h photoperiod was maintained using broad-spectrum light sources. Growing temperatures ranged from 18°C (night) to 20°C (day). Hydroponic growth solution, changed weekly, was made from MaxiBloom® Hydroponics Plant Food (General Hydroponics, Sevastopol, CA, United States) at a concentration of 1.7 g/L.

The resequencing panel consisted of 21 A. hortensis accessions: 15 from the United States Department of Agriculture collection (USDA, ARS, NALPGRU;[1]); five each from two separate commercial seed vendors (Baker Creek Heirloom Seed Company, Mansfield, Missouri and Wild Garden Seed, Philomath, Oregon); and one accession collected in the wild in Utah (BYU 1317 from Park City, Utah). Plants used in the resequencing panel were originally collected from across Europe (France, Poland, countries of the former Soviet Union, former Serbia/Montenegro and Norway) and North America (United States, and Canada). A complete list of all plant materials including passport information is provided in **Table 1**.

DNA Extraction, Library Preparation, and Oxford Nanopore Sequencing

The Golden variety of *A. hortensis* was grown hydroponically in a growth chamber at BYU as previously described. Plants were dark-treated for 72 h at which point young leaf tissue was harvested and extracted for high molecular weight (HMW) genomic DNA using the Qiagen (Germantown, MD) Genomic-tip protocol. The DNA concentration was checked using the dsDNA High Sensitivity DNA Assay on the Qubit® 2.0 Fluorimeter (Invitrogen, Merelbeke, Belgium).

Samples for DNA sequencing were prepared with and without fragmentation using Covaris g-TUBEs (Woburn, MA) and the ZYMO DNA Clean & Concentrator-5 column (Irvine, CA, United States). Samples were fragmented using both the ZYMO DNA kit and Covaris g-TUBEs following manufacturer's instructions. Samples prepared with the Covaris g-TUBEs were fragmented at several centrifugation speeds, including 3,800, 4,000, and 4,200 RPM. In total, nine libraries from the original DNA stock were prepared for sequencing using the 1D Genomic DNA by Ligation MinION library preparation kit. Libraries were sequenced on R9 flow cells on a MinION for 48 h using MinKNOW 2.0 software with the following settings: DNA, PCR-free, no multiplexing, SQK-LSK109 kit (Oxford Nanopore Technologies, Ltd., Oxford, United Kingdom). No alterations were made to voltage or time. Albacore v2.3.1, part of the MinKNOW package, was used for base calling.

[1]https://npgsweb.ars-grin.gov/

Read Cleaning, Draft Genome Assembly, and Polishing

MinIONQC (Lanfear et al., 2019) was used with default settings to summarize sequence data. NanoFilt (De Coster et al., 2018) was then used to trim and filter reads using the following options: $-q$ = 8, headcrop = 25, $-$l = 2000. Porechop v.0.2.3 (Wick et al., 2017) was used to trim adaptors from sequence data with the default options. Draft genomes were assembled using multiple assemblers, specifically Canu v.1.7.1 (Koren et al., 2017), MaSuRCA v.3.2.8 (Zimin et al., 2013), Flye v.2.3.6 (Kolmogorov et al., 2019) and wtdbg2 (Ruan and Li, 2020). Illumina reads were used to polish the Canu assembly using Nanopolish (Loman et al., 2015), and Pilon v.1.22 (Walker et al., 2014). The completeness of each of the draft genome assemblies was assessed using BUSCO v4 (Simão et al., 2015) using the flowering plant (embryophyte_odb10) orthologous gene data set. Specific commands and flags for each assembly program used are provided in **Supplementary File 1**.

Hi-C Scaffolding

Atriplex hortensis plants (cv. Golden) were dark-treated for 72 h prior to flash-freezing young leaf tissue in liquid nitrogen. Tissue samples were then shipped to Dovetail Genomics (Scotts Valley, CA, United States) for Chicago® and Hi-C proximity ligation sequencing. Dovetail Chicago® libraries are similar to Hi-C libraries but differ in that they rely on library preparation from *in vitro* rather than *in vivo* reconstituted chromatin that has been cross-linked and subsequently sheared (Moll et al., 2017). Chromosome-scale scaffolds were generated using Dovetail Genomics' HiRiSE™ assembler.

Illumina Sequencing and Transcriptome Assembly

The "Golden" variety of *A. hortensis* was grown hydroponically as previously described. Plants were either grown in a control hydroponic solution or in hydroponic solution supplemented with NaCl. For the salt treatment, NaCl was added daily at 50 mM increments to the hydroponic solution of 21-day old plants until a concentration of 350 mM NaCl was reached (7 days). Tissue for RNA extraction was harvested 24 h after 350 mM NaCl concentration was reached. Root,

TABLE 1 | Identification and passport information for plant materials used for the genome sequencing and the resequencing panel.

Accession	Source	Improvement status (name)	Collection location	Latitude/Longitude	Sequencing technology	SRR number[2]
Resequencing Panel						
BYU 1317	BYU herbarium	Wild	Park City, UT, United States	40.66796, −111.515032	Illumina[3]	SRR11123184
Red Orach	Baker Creek Heirloom Seeds	Improved	Mansfield, MO, United States	N/A[3]	Illumina	SRR11123183
PI 310383	USDA	Uncertain	Tashkent, Uzbekistan	N/A	Illumina	SRR11123172
PI 323313	USDA	Improved	Poland	N/A	Illumina	SRR11123170
PI 345962	USDA	Uncertain	Norway	N/A	Illumina	SRR11123169
PI 357340	USDA	Improved (Zolta)	Former Serbia/Montenegro	41.91667000, 22.41667000	Illumina	SRR11123168
PI 357342	USDA	Improved (Zolta Prilepska)	Former Serbia/Montenegro	41.34640000, 21.55440000	Illumina	SRR11123167
PI 357344	USDA	Improved (Lokalna Zolta)	Former Serbia/Montenegro	41.81200000, 21.99470000	Illumina	SRR11123166
PI 357346	USDA	Improved (Gradinarska)	Former Serbia/Montenegro	41.57920000, 21.57190000	Illumina	SRR11123165
PI 357347	USDA	Improved (Debarska)	Former Serbia/Montenegro	41.52500000, 20.52750000	Illumina	SRR11123164
PI 370353	USDA	Improved (Lokalna)	Former Serbia/Montenegro	41.89890000, 21.40810000	Illumina	SRR11123182
PI 370354	USDA	Improved (Mestna)	Former Serbia/Montenegro	41.94140000, 21.41280000	Illumina	SRR11123181
PI 372512	USDA	Wild	Alberta, Canada	51.4502063, −112.7061764	Illumina	SRR11123180
PI 379088	USDA	Improved (2261)	Former Serbia/Montenegro	41.84890000, 21.82030000	Illumina	SRR11123179
PI 379093	USDA	Improved (2475)	Former Serbia/Montenegro	41.38250000, 22.28750000	Illumina	SRR11123178
PI 379095	USDA	Improved, (Skopska)	Former Serbia/Montenegro	42.00000000, 21.43330000	Illumina	SRR11123177
PI 420154	USDA; 218	Wild (218)	France	N/A	Illumina	SRR11123176
Triple Purple	Wild Garden Seed Co.	Improved	Philomath, OR, United States	N/A	Illumina	SRR11123174
P1reselection	Wild Garden Seed Co.	Improved	Philomath, OR, United States	N/A	Illumina	SRR11123173
P6 reselection	Wild Garden Seed Co.	Improved	Philomath, OR, United States	N/A	Illumina	SRR11123171
Golden	Wild Garden Seed Co.	Improved	Philomath, OR, United States	N/A	Illumina	SRR11123175
Whole Genome Sequencing						
Golden	Wild Garden Seed Co.	Improved	Philomath, OR, United States	Leaf tissue	Oxford Nanopore	SRR11147376
Hi-C scaffolding						
Golden	Wild Garden Seed Co.	Improved	Philomath, OR, United States	Chicago™ Hi-C	Illumina	SRR11147368
Golden	Wild Garden Seed Co.	Improved	Philomath, OR, United States	HiRiSE™ Hi-C	Illumina	SRR11147367
Transcriptome						
Golden	Wild Garden Seed Co.	Improved	Philomath, OR, United States	Root tissue	Illumina	SRR11147369
Golden	Wild Garden Seed Co.	Improved	Philomath, OR, United States	Stem tissue	Illumina	SRR11147370
Golden	Wild Garden Seed Co.	Improved	Philomath, OR, United States	Floral tissue	Illumina	SRR11147371
Golden	Wild Garden Seed Co.	Improved	Philomath, OR, United States	Leaf tissue	Illumina	SRR11147372
Golden	Wild Garden Seed Co.	Improved	Philomath, OR, United States	Whole plantlet	Illumina	SRR11147373
Golden	Wild Garden Seed Co.	Improved	Philomath, OR, United States	Root tissue – NaCl treated[4]	Illumina	SRR11147374
Golden	Wild Garden Seed Co.	Improved	Philomath, OR, United States	Leaf tissue – NaCl treated	Illumina	SRR11147375

Accessions of A. hortensis originated throughout Europe and North America. For the whole genome sequencing, Hi-C scaffolding and transcriptome data, the source tissue and/or library type is provided in the Latitude/Longitude column. [1]Meters above sea level. [2]Sequence read archive accession number for each resequenced line. Maintained by national Center for Biotechnology Information (https://dataview.ncbi.nlm.nih.gov/objects?linked_to_id = PRJNA607334). [3]N/A indicates no data. [4]Treated with 350 mM NaCl (see section "Materials and Methods").

stem and leaf tissue was harvested from both control and treated plants. One-week old whole plantlet and inflorescence (tissue and immature seed) tissues from untreated plants were also collected.

In total, seven libraries were prepared with 180-bp inserts. Sequencing was conducted using the Illumina HiSeq platform at the Beijing Genomics Institute (Shenzhen, China). Reads were trimmed and quality controlled using the program Trimmomatic-0.35 (Bolger et al., 2014). RNA-seq data were aligned to the Hi-C assembly using HiSat v2.2.1 with the max intron length set to 50,000 bp (Kim et al., 2015). Data was then assembled into potential transcripts using StringTie (Pertea et al., 2015) with default parameters.

Repeat Analysis and Annotation

Atriplex hortensis-specific repeats were identified using RepeatModeler v.1.0.11 (Smit and Hubley, 2008–2015). RepeatMasker v.4.0.7 (Smit et al., 2013–2015) was used to classify *A. hortensis*-specific repeats using the RepBase database version 20160829. The MAKER2 v2.31.10 pipeline (Holt and Yandell, 2011) was used to annotate the *A. hortensis* genome with *ab initio* gene predictions using AUGUSTUS (Stanke et al., 2004) species-specific gene models for *A. hortensis*. Additional evidence sources for the annotations included expressed sequence tags (EST) and protein homology from the transcriptomes of *C. quinoa* (Jarvis et al., 2017) and *C. pallidicaule* Aellen (Mangelson et al., 2019) as well as the *A. hortensis* transcriptome produced from the RNA-seq data previously described. The uniprot_sprot database (downloaded 11/13/2018) was used for Basic Local Alignment Search Tool (BLASTp)-based annotation of the gene models.

Resequencing

Genomic DNA from each of the 21 *A. hortensis* accessions was extracted using the mini-salts protocol reported by Todd and Vodkin (1996). The DNA concentrations and quality were checked using the dsDNA BR Assay from Qubit® 2.0 Fluorimeter. Libraries were sent to Novogene (San Diego, CA, United States) for whole-genome Illumina HiSeq X Ten sequencing (2 × 150-bp paired-end). Reads were trimmed with Trimmomatic using default parameters (Bolger et al., 2014). Reads from each accession were then aligned to the final *A. hortensis* reference genome using Bowtie2 using the very-sensitive-local flag (Langmead and Salzberg, 2012) to produce BAM files that were further marked for PCR duplicates using the MarkDuplicates subroutine in the Picard package.[2] Single nucleotide polymorphism (SNP) genotype likelihoods and covariances were then determined from the 21 accessions using ANGSD using a *p*-value of 10E-06 for a site being variable (Korneliussen et al., 2014) to produce a genotype likelihood (beagle) file. Multivariate analysis of the covariance data was accomplished using PAST4 (Hammer et al., 2001), while population structure and admixture were then inferred using PCAngsd (Meisner and Albrechtsen, 2018) at $K = 3$ based on the DeltaK method described by Evanno et al. (2005). Bootstrapped ($n = 1000$) UPGMA phylogenies based on Euclidean similarity indices were produced using PAST4 (Hammer et al., 2001).

[2]http://broadinstitute.github.io/picard/

Cytogenetics and Genome Size Estimation

Atriplex hortensis cv. Golden seeds were germinated on petri dishes for 36 h. Root meristems were collected and immersed in ice water for 24 h. Root meristems were then treated for another 24 h in a 3:1 mixture of ethanol (95%) – glacial acetic acid. Root tips were prepared under a dissecting microscope where they were placed on slides, treated with iron-acetocarmine, warmed on an alcohol burner, and squashed. Chromosomes were examined using a Zeiss Axioplan 2 phase-contrast microscope and images were captured on an Axiocam (Carl Zeiss, Jena, Germany) CCD camera. Fluorescent *in situ* hybridization (FISH) rDNA images of mitotic chromosome preparations of *A. hortensis* cv. "Triple Purple" were taken using yellow-green fluorescing digoxygenin to highlight the NOR-35S region and red fluorescing rhodamine to highlight the 5S region using the protocol described by Maughan et al. (2006). Chromosome spreads and DNA probes for FISH were prepared using the protocol described in Kolano et al. (2012).

Genome-size estimation was conducted using a Beckman Coulter (Miami, FL, United States) Gallios flow cytometer by Agriculture and Agri-Food Canada (AAFC) as described by Yan et al. (2016). Samples were analyzed in triplicate (technical replicates) conducted over three different days. Characteristics of the florescence peaks including mean, nuclei numbers, and coefficients of variation were determined using the R package flowPloidy (Smith et al., 2018). The 2C DNA value of each sample was calculated as: (mean of sample G1peak/mean of standard G1 peak) × 2C DNA content (pg) of the radish (*Raphanus raphanistrum* subsp. *sativus*, estimated 515 Mb) standard.

RESULTS

Library Fragmentation

Since Oxford Nanopore Technologies (ONT) sequencing is still relatively new, we tested the relationship between fragmentation strategies, read length and total sequence output to discover the optimal sample preparation method. To achieve sufficient coverage, we developed nine different libraries that were each sequenced independently on different flow cells. In total, the nine libraries yielded 65.4 Gb of data from 5,525,447 reads with a read length N50 of 22,087 bp, a mean read length of 13,487 bp and a mean quality score of nine (**Table 2**). Individual DNA libraries prepared with fragmentation (Covaris g-TUBEs and ZYMO DNA concentrator-5 column kit) or without fragmentation produced dramatically varied results in terms of read lengths and total sequence yield. Not unexpectedly, the library prepared without fragmentation produced the longest read lengths (N50 = 40,434 bp) but also exhibited the lowest overall

TABLE 2 | Oxford Nanopore library preparation and sequencing statistics. Non-fragmentation and fragmentation techniques were used in sample preparation.

Library	Fragmentation	Total length (Gb)	Totals reads	N50 (bp)	Mean length (bp)	Median length (bp)	Max length (bp)	Mean quality score
1	No Fragmentation	1.26	55,551	40,434	22,617	15,607	194,834	9.1
2	Zymo	5.61	567,514	23,595	9,877	4,522	153,389	9.4
3	Zymo	2.24	133,660	33,394	16,770	9,857	199,575	9.2
4	Covaris, 4,200 RPM	13.04	1,005,270	15,878	11,760	11,017	181,817	8.3
5	Covaris, 3,800 RPM	10.08	854,994	15,277	11,788	11,104	133,274	9.1
6	Covaris, 3,800 RPM	6.94	501,526	20,681	13,760	12,431	164,726	8.9
7	Covaris, 4,200 RPM	8.86	617,385	19,276	14,343	12,932	231,794	9.1
8	Covaris, 4,200 RPM	10.72	1,221,530	12,664	8,778	8,300	149,453	9.1
9	Covaris, 4,000 RPM	6.64	568,017	17,580	11,686	11,115	128,743	8.9
Avg/Total		65.4	5,525,447	22,087	13,487	10,765	170,845	9

sequence yield (1.26 Gb). Fragmentation using the Covaris g-TUBEs at different centrifugation speeds (3,800, 4,000, and 4,200 RPM) produced variable results, but with general trends, specifically: (1) Covaris g-TUBE fragmented libraries always outproduced the non-fragmented library (average yield = 8.67 Gb), but the N50 of the read lengths of these libraries was always smaller (average N50 = 17,166 bp), and (2) lower centrifugation speeds produced longer read lengths (3,800 RPM = 17,979 bp vs 4,200 RPM = 15,939 bp), but with lower yield (3,800 RPM = 8.5 Gb vs 4,200 RPM = 10.9 Gb; **Table 2**). The two fragmentation libraries produced using the ZYMO DNA kit yielded intermediately to the non-fragmented libraries and the Covaris fragmented libraries, with an average of 3.93 Gb of sequence with a read length N50 of 28,495 bp, suggesting that the ZYMO DNA kit only minimally fragmented the DNA.

Genome Assembly

Flow cytometry indicated that the *A. hortensis* genome is approximately 1.172 Gb (**Table 3**), while karyotyping of cell nuclei showed that *A. hortensis* carries nine pairs of chromosomes ($2n = 2x = 18$). In the *A. hortensis* karyotype, chromosomes were metacentric to slightly submetacentric (**Figure 1**), and similar in length.

Multiple assemblers were tested to determine which would most optimally assemble the *A. hortensis* genome. These assemblers included Canu (Koren et al.), MaSuRCA (Zimin et al., 2013), Flye (Kolmogorov et al., 2019) and wtdbg2 (Ruan and Li, 2020). All assemblers were run with default parameters. The wtdbg2 assembler produced the largest number of contigs and the Flye assembler produced the smallest N50 (**Table 4**). Both the wtdbg2 and the Flye assemblers produced smaller (total genome size) assemblies relative to the MaSuRCA and Canu assemblers. Both the MaSuRCA and Canu assemblers produced >1 Mb contig N50s, with the Canu assembler producing the least collapsed assembly (relative to the predicted genome size of 1.2 Gb).

The MaSuRCA assembler uses a hybrid approach for assembly that initially utilizes high-quality short-reads to produce super-reads that are then scaffolded and gap-filled with the long reads to produce high-quality scaffolds that do not require error correction (Zimin et al., 2013). The Flye, Canu and wtdbg2 assemblers are based solely on the error-prone long reads and are thus considered unpolished assemblies and require polishing to correct the inherently high sequencing error rate associated with the ONT technology. We polished the Flye, Canu and wtdbg2 draft assemblies with Nanopolish, which uses the original ONT reads for consensus correction along with two rounds of Pilon, which in turn uses the high-quality Illumina short reads for correction to produce a high-quality, polished set of draft assemblies for comparison (**Table 4**). We evaluated the final assemblies using Benchmarking Universal Single Copy Orthologs (BUSCO) (Simão et al., 2015) which quantifies gene content completeness based on a large core set of highly conserved orthologous genes (COGs). After polishing, BUSCO

TABLE 3 | Flow cytometry results of *A. hortensis* (cv. "Golden") leaf tissue. A C-value of 2.4 picograms yielded a genome size estimate of 1.17 Gb.

Sample	No. technical replicates	Mean 2C value ± S.D. (pg DNA)	Gb per haploid (1C) genome[1]
1	3	2.39 ± 0.0342	1.169 ± 0.017
2	3	2.39 ± 0.0184	1.169 ± 0.009
3	3	2.41 ± 0.0061	1.178 ± 0.003
Average		2.39 ± 0.0196	1.172 ± 0.009

[1] *Haploid genome size was calculated as 1 pg = 978 Mbp per Dolezel et al. (2003).*

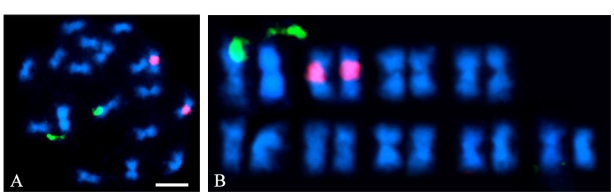


FIGURE 1 | *A. hortensis* chromosome pairs. Nine metacentric chromosome pairs are visible. **(A)** Fluorescent *in situ* hybridization (FISH) using NOR-35S (green) and 5S (red) labeled rDNA probes on mitotic chromosome preparations of *A. hortensis* cv. "Triple Purple." **(B)** Chromosomes from panel **(A)** arranged as a karyotype. Note the metacentric to submetacentric centromere positions on all nine chromosome pairs.

TABLE 4 | Assembly and Benchmarking Universal Single Copy Orthologs (BUSCO) statistics for the MaSuRCA, Flye, wtdbg2, Canu and Hi-C scaffolded Canu assemblies.

Metrics	Assembler				
	MaSuRCA[1]	Flye	wtdbg2	Canu	Canu Hi-C[1]
Number of contigs	2,850	3,009	4,659	3,183	1,325
Total size of the assembly (bp)	914,348,993	813,404,858	867,633,014	964,774,081	965,003,581
Longest (bp)	10,067,120	11,408,532	7,183,448	9,632,068	113,540,806
Shortest (bp)	2,256	178	2,807	1,072	603
Mean (bp)	320,824	270,324	186,227	303,102	728,305
Median (bp)	39,479	100,839	15,813	63,149	35,305
N50 (bp)	1,317,304	698,928	1,176,597	1,114,696	98,884,393
L50	169	282	201	223	5
G+C%	36.84	36.77	36.89	37.06	37.05
	BUSCO metrics				
%Complete COGS found [%single, %duplicate]	90.0% [85.8%, 4.2%]	90.6% [88.1%, 2.5%]	86.0% [83.5%, 2.5%]	97.3% [95.1%, 2.2%][2]	96.7% [95.0%, 1.7%]

[1] *Scaffold statistics are reported for the MaSURca and Canu Hi-C assembly. All other assembly statistics are contig metrics.*

was used to identify complete COGs within the various assemblies, which ranged from a low of 86.0% in the Flye assembly, to a high of 97.3% in the Canu assembly. The necessity of the polishing steps was reflected in the increasing BUSCO scores after successive rounds of polishing. For example, the BUSCO scores for complete COGs identified for the original, Nanopolished, Nanopolished+Pilon and Nanopolished+Pilon+Pilon Canu assemblies were 50.5, 73.8, 90.1, and 97.3%, respectively.

Both the MaSuRCA and Canu assemblers produced superior assemblies based on the total size of the contigs, contig N50, and BUSCO scores; however, the polished Canu assembly was ultimately chosen as the draft genome for Hi-C scaffolding due to concerns of repeat collapse within the MaSuRCA assembly as reflected in the smaller total size of the contigs. The polished Canu assembly resulted in 3,183 contigs, spanning 965 Mb, a contig N50 of 1.114 Mb, an L50 of 223, and a BUSCO score of 97.3% (**Table 4**).

Chromosome-Scale Scaffolding

To further improve the Canu assembly, contigs were scaffolded using chromatin-contact maps using Dovetail Chicago® and Hi-C libraries. Chicago® library contact maps are based on *in vitro* reconstituted DNA and are ideal for detecting and correcting miss-joins in *de novo* assemblies as well as short-range scaffolding (Putnam et al., 2016). A total of 163 million read pairs (70X coverage) were generated from the Chicago® library and were used to detect misalignments and scaffold the Canu assembly using the HiRiSE™ scaffolder. In total, 429 breaks and 1,421 joins were made, resulting in a net decrease in the total number of scaffolds to 2,191 and a slight decrease in N50 (817 kb) for the assembly. Whenever a join was made between contigs, an "N" gap, consisting of 100 Ns, was created. The total percent of the genome in gaps was less than 0.1%.

The Chicago®-based assembly was then further scaffolded using an *in vivo* Hi-C library created from native chromatin to produce ultra-long-range (10–10,000 kb) mate-pairs. A total of 200 million mate-pair reads, representing a physical coverage of 62×, were generated and scaffolded using the HiRiSE™ scaffolder. In total, 868 joins and no breaks were made, producing a final assembly containing 1,325 scaffolds, spanning a total sequence length of 965 Mb with an N50 and L50 of 98.9 Mb and 5 scaffolds, respectively. Nine chromosome-scale scaffolds were assembled containing 94.7% of the total sequence length. The chromosome-scale scaffolds ranged in size from 93.6 to 113.5 Mb and were numbered sequentially based on scaffold length (e.g., Ah1–Ah9). Scaffold joins produced by the Hi-C mate-pairs introduced new "N" gaps in the assembly, thereby increasing the number of gaps in the assembly to 2,295. The final number of "N" nucleotides in the final Hi-C assembly was 229,050 (<0.1%; **Table 2**).

A BUSCO analysis of the final Hi-C assembly identified 1,331 (96.8%) complete COGs from the *Embryophyta* database (n = 1375), of which only 1.7% (23) were duplicated – reflecting the diploid nature of the *Atriplex* genome and suggesting that only minor paralogous duplications have occurred. Another nine (0.7%) fragmented COGs were identified. Only 35 COGs were missing, which is indicative of a highly complete assembly.

Repeat Features

The RepeatModeler and RepeatMasker pipelines were used to annotate and mask the repeat fraction of the Hi-C assembly. Approximately 66% (639.6 Mb) of the genome was annotated as repetitive, which is slightly higher than the repetitive fraction classified for other members of the *Amaranthaceae/Chenopodiaceae* alliance with reference genomes [48% in *Amaranthus hypochondriacus* L. (Lightfoot et al., 2017), 64% in *Spinacia oleracea* (Li et al., 2019), 63% in *Beta vulgaris* L. (Flavell et al., 1974), 64.5% in *C. quinoa* (Jarvis et al., 2017)]. The most common repeat elements identified were long-terminal repeat retrotransposons (LTR-RT). The LTR-RTs are the most abundant genomic component in flowering plants (Du et al., 2010; Galindo-Gonzalez et al., 2017) and their frequency is strongly correlated with increased genome size (Michael, 2014).

TABLE 5 | Repetitive element classification for final assembly (Canu Hi-C) as reported by RepeatMasker.

Repeat class[a]	Repeat name	Count	bp masked	Masked (%)
DNA		2,937	2,206,004	0.23
	CMC-EnSpm	24,140	13,861,122	1.44
	Crypton	972	83,799	0.01
	MULE-MuDR	13,906	6,550,989	0.68
	MuLE-MuDR	2,306	1,334,054	0.14
	PIF-Harbinger	1,772	892,861	0.09
	TcMar-Mogwai	1,261	652,191	0.07
	TcMar-Stowaway	32,881	5,565,216	0.58
	hAT	548	112,059	0.01
	hAT-Ac	10,241	2,812,539	0.29
	hAT-Tag1	1,772	308,038	0.03
	hAT-Tip100	6,258	7,096,629	0.74
LINE				
	CRE-II	447	335,625	0.03
	Jockey	2,157	567,406	0.06
	L1	8,788	7,141,529	0.74
	L2	14,243	21,727,481	2.25
	Penelope	2,431	992,653	0.10
	R1	566	220,850	0.02
	RTE-BovB	6,587	2,055,784	0.21
LTR		7,950	1,885,731	0.20
	Caulimovirus	351	589,622	0.06
	Copia	62,884	107,371,058	11.13
	DIRS	4591	1,952,514	0.20
	Gypsy	167,590	307,636,295	31.9
	Pao	76	7,473	0.00
	Caulimovirus	7,950	1,885,731	0.20
RC				
	Helitron	6,262	2,307,357	0.24
SINE				
	tRNA	418	50,479	0.01
Unknown		451,250	143,307,912	14.86
Total interspersed		835,585	639,625,270	66.33
Low_complexity		31,237	1,603,682	0.17
Satellite		128	11,962	0.00
Simple repeat[b]		216,414	47,610,726	4.94
rRNA		598	309,846	0.03
Total		1,083,962	6,89,161,486	71.46

[a]*LINE, long interspersed nuclear elements; LTR, long terminal repeat; RC, Rolling circle.* [b]*The most common mono- di-, tri-, and tetra- nucleotide repeat motifs were (T)n, (TA)n, (ATT)n, (TTTA)n, respectively.*

Of the various LTR-RTs present in the *A. hortensis* genome, *Gypsy*-like (31.90%) and *Copia*-like (11.13%) elements represent greater than 40% of the genome and are in a 3:1 (*Gypsy*:*copia*) ratio, similar to the 2.9:1 ratio reported for 50 sequenced plant genomes (Ou and Jiang, 2018). An additional 5.14% (49.5 Mb) of the genome was classified as low-complexity (satellites, simple repeats, and rRNAs), while 14.86% (143 Mb) of the genome was characterized as unclassified repetitive elements (**Table 5**) – presumably representing *Atriplex*-specific repeat elements that will undoubtedly be important for understanding the evolution of the *A. hortensis* genome.

A BLAST search (Altschul et al., 1990) of the complete rRNA gene sequence found in *C. quinoa* (DQ187960.1) was conducted to identify the 35S rRNA gene (NOR) location in the *A. hortensis* genome using the *C. quinoa* sequence as query. The 35S rRNA locus was located on chromosome Ah6. Another BLAST search was conducted to identify matches for the 5S rRNA gene locus in *A. hortensis*, again using the homologous 5S rDNA repeat sequence in *C. quinoa* (DQ187967.1) as the query. The 5S rDNA sequences mapped primarily to chromosome Ah4 and to several other smaller unscaffolded contigs that did not assemble into specific chromosomes. The appearance of these smaller scaffolds in the BLAST search results was not surprising as 5S rDNA repeats are highly repetitive and of low-complexity, and thus extremely difficult to assemble and scaffold accurately. A FISH analysis of mitotic chromosome preparations for *A. hortensis* cv. Triple Purple revealed a physical location of a single NOR-35S (green) locus and of a single 5S (red) rRNA gene tandem repeat-array locus (**Figure 1**). The identification of the cytological and genomic position of the 5S rRNA and 35S rRNA gene loci gives unique identities to two of the nine chromosome pairs in the *A. hortensis* kayotype (specifically Ah4 and Ah6).

The sequence for telomeric repeats in plants is highly conserved and has been identified as TTTAGGG (Richards and Ausubel, 1988). A BLAST search of this sequence motif against the nine *A. hortensis* chromosomes identified tandemly repeated telomeric sequences on at least one end of each of the nine chromosome assemblies with a total of 13 telomere-like repetitive regions identified (**Figure 2**). Four of the nine chromosomes had telomere-to-telomere assemblies (telomeres identified on both ends of the chromosome assembly).

Genome Annotations

A *de novo A. hortensis* transcriptome, derived from 30–40 million RNA-seq reads each from stem, leaf, floral and whole plantlet tissues, consisted of 272,255 isoforms with an N_{50} of 3,325 bp and a mean length of 1,956 bp. The *A. hortensis* transcriptome, along with the EST and peptide models from *C. quinoa* and *C. pallidicuale* and the uniprot-sprot database, were provided as primary evidence for annotation in the MAKER pipeline. The RNA-seq data mapped with high efficiency to the final genome assembly, with an overall alignment rate of 92% and with 81.5% of the pair reads aligning concordantly exactly one time, with only 4.26% aligning more than once concordantly – suggestive of a high-quality genome assembly and reflective of the diploid nature of the *A. hortensis* genome. The MAKER pipeline identified a total of 39,540 gene models and 2,555 tRNA genes. The average length of genes identified by MAKER was 1,750 bp. The completeness of the annotation was assessed by BUSCO which identified 1,278 (92.9%) complete COGs from the transcript annotation (Complete: 92.9% [Single Copy: 90.4%, Duplicated: 2.5%], Fragmented: 4.7%, Missing: 2.4%). To assess the quality of the annotations, we used the mean Annotation Edit Distance (AED) which is calculated by combining annotation values corresponding to specificity and sensitivity. AED values of 0.5 and below are considered good annotations, and values of 0.30 and below are considered high quality annotations (Holt and Yandell, 2011). Over 90% of the genome models have an AED

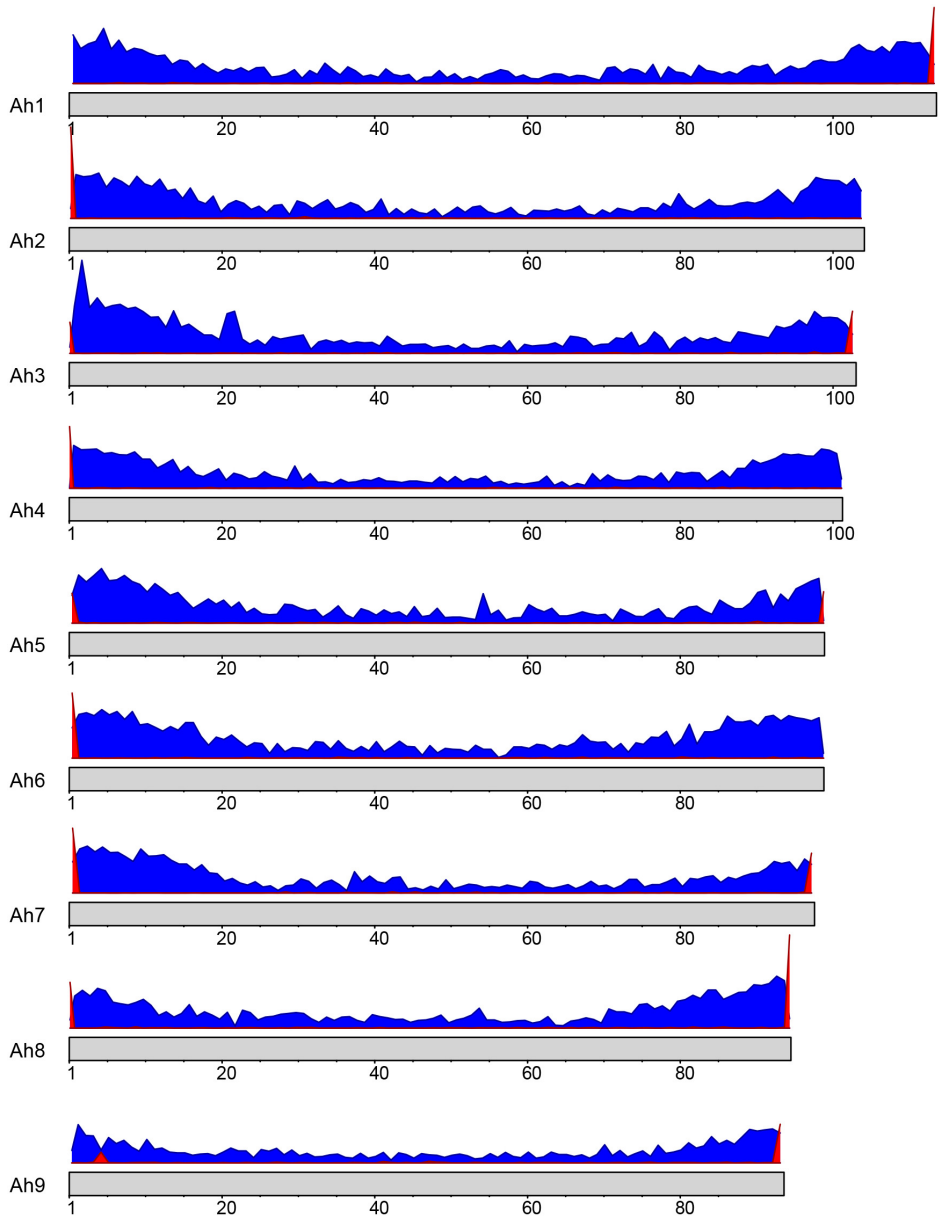


FIGURE 2 | Gene density (blue) and telomere positioning (red) on the *A. hortensis* chromosomes. A conserved telomere repeat sequence in plants (TTTAGGG; Richards and Ausubel, 1988) was used to locate telomere positions using BLAST searches. The *x*-axis represents the position on the chromosome, while the *y*-axis represents the frequency of genes or telomeric repeats in each bin.

value <0.5, with the majority (51.7%) of the models having AED values below 0.325 (**Figure 3**). An analysis of the completeness of the gene models was further assessed by comparing the matched length of the transcripts with orthologous *C. quinoa* transcripts. Orthologs were determined using BLAST analysis (*e*-value < 1e-20) with the max target set to 1. Of the 18,657 orthologs identified with *C. quinoa*, ∼80% (14,764) covered at least 70% of the *C. quinoa* orthologs. The AED score coupled with the BUSCO assessment and ortholog analysis are suggestive of a high-quality genome assembly and annotation. In addition, the observed chromosomal distribution of the annotated genes, with higher gene density near the ends of chromosomes and lower gene density in the centromeric regions (**Figure 2**), is suggestive of a high-quality genome assembly and annotation. An examination of the self-synteny map (not shown but accessible via CoGe) revealed no obvious blocks of paralogous genes.

Genomic Comparison and Features

Several species within the *Amaranthaceae/Chenopodiaceae* alliance have chromosome-scale genome assemblies, including the ancient allotetraploids *C. quinoa* (Jarvis et al., 2017) ($2n = 4x = 18$) and *Amaranthus hypochondriacus* (Lightfoot

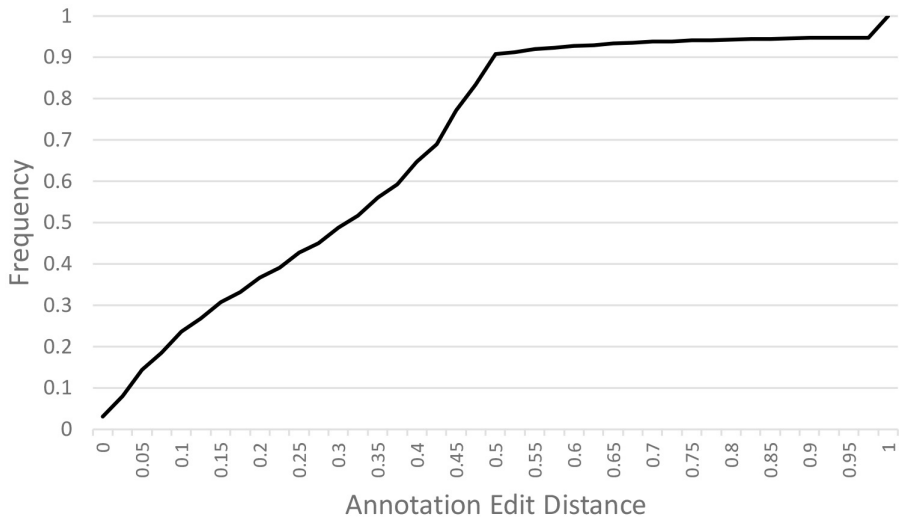


FIGURE 3 | Annotation Edit Distance (AED) calculated for MAKER predicted gene models. Annotation Edit Distance (AED) is used to measure the quality of a genome annotation. This is calculated by combining annotation values corresponding to specificity and sensitivity. AED values of 0.50 and below are good annotations and values of 0.30 and below are high-quality annotations (Holt and Yandell, 2011).

et al., 2017) ($2n = 4x = 16$), and the diploid *Beta vulgaris* (Funk et al., 2018) ($2n = 2x = 9$). Previous phylogenic research using chloroplast DNA (*rbcL* gene, *atpB-rbcL* spacer) and nuclear rDNA internal transcribed spacer (ITS), clearly demonstrated that *Atriplex* is more closely related to *Chenopodium* and *Beta,* which are both found in the same family *Chenopodiaceae* s. str. but in different subfamilies, *Chenopodioideae* (*Chenopodium* and *Atriplex*) and *Betoideae* (*Beta*), while *Amaranthus* is more distantly related to these chenopods and is found within the family *Amaranthaceae* s. str., subfamily *Amaranthoideae* (Kadereit et al., 2003, 2010; Fuentes-Bazan et al., 2012; Morales-Briones et al., 2020). Syntenic relationships between *A. hortensis* and these other genomes were explored using DAGChainer (Haas et al., 2004), which identifies syntenic blocks of collinear homologous gene pairs between genomes.

A synteny analysis of *A. hortensis* and *B. vulgaris* identified 226 shared syntenic blocks between the genomes, with 11,697 colinear gene pairs (averaging 52 gene pairs/block) spanning 469 and 616 Mb of the *B. vulgaris* and *A. hortensis* genomes, respectively. Moreover, syntenic block sizes between the species were also correlated ($R^2 = 0.36$), further reflecting a shared ancestry of two species. One-to-one orthologous relationships between *A. hortensis* and *B. vulgaris* chromosomes (**Figure 4** and **Table 6**) were clearly ascertained for six of the nine *A. hortensis* chromosomes: Ah1 = Bv5 (100% shared syntenic block sequence), Ah2 = Bv4 (99%); Ah4 = Bv8 (100%); Ah5 = Bv3 (100%); Ah7 = Bv7 (100%); Ah8 = Bv2 (100%). The remaining three *A. hortensis* chromosomes shared substantial levels of synteny with multiple *B. vulgaris* chromosomes, suggestive of intergenomic rearrangements (i.e., reciprocal translocations), with Ah3 = Bv6 (51%), Bv9 (49%); Ah6 = Bv6 (52%), Bv1 (48%); and Ah9 = Bv9 (54%), Bv1 (44%). We note that we cannot exclude that these rearrangements are possible misassembles – although our Hi-C data strongly supports the current placements.

Atriplex hortensis and *B. vulgaris* are both diploid and share a haploid chromosome number ($n = 9$), whereas *C. quinoa* is an allotetraploid member (showing amphidiploid inheritance) of the subfamily *Chenopodioideae*, having experienced an ancient allopolyploidization event (Storchova et al., 2015). Our analysis of synteny between *C. quinoa* and *A. hortensis* identified a combined total of 24,710 syntenic gene pairs, spanning 1.1 Gb and 1.3 Gb of the *C. quinoa* and *A. hortensis* genome, respectively, using a tetraploid-to-diploid (2:1) analysis. The synteny observed among the *A. hortensis* and *C. quinoa* chromosomes suggests several orthologous relationships with known homeologous *C. quinoa* chromosome pairs, including Ah1 = Cq5A (51% shared syntenic block sequence), Cq5B (49%); Ah2 = Cq4A (50%), Cq4B (48%); Ah4 = Cq8A (48%), Cq8B (51%); Ah5 = Cq3A (50%), Cq3B (50%); Ah7 = Cq7A (49%), Cq7B (43%); Ah8 = Cq2A (52%), Cq2B (43%). As with *B. vulgaris*, *A. hortensis* chromosomes Ah3, Ah6 and Ah9 have large rearrangements showing synteny to Cq1A&B, Cq6A&B and Cq9A&B.

Atriplex hortensis, *B. vulgaris* and *C. quinoa* share a base chromosome number of $x = 9$, whereas the base number in *Amaranthus* is $x = 8$, due to a chromosome loss (Am5) and a chromosome fusion (Am1; Lightfoot et al., 2017). The amaranths belong to the family *Amaranthaceae* s. str. Subfamily *Amaranthoideae* and were thus expected to be the most divergent of the three genomes compared. Indeed, while our genome comparison of *A. hortensis* with *A. hypochondriacus* clearly showed synteny (**Figure 4** and **Table 6**), the size of the 410 syntenic blocks (12,306 syntenic gene pairs) observed was the smallest of the three genomes (Bv: 2.1 Mb/block; Cq: 2.7 Mb/block; Am: 0.84 Mb/block), accompanied by the lowest block size correlation between the species (*Bv*: $R^2 = 0.36$; *Cq*: $R^2 = 0.42$; *Am*: $R^2 = 0.04$). These decreases are reflective of the more distant evolutionary relationship between *Atriplex* and *Amaranthus* within the family. We confirm the chromosome

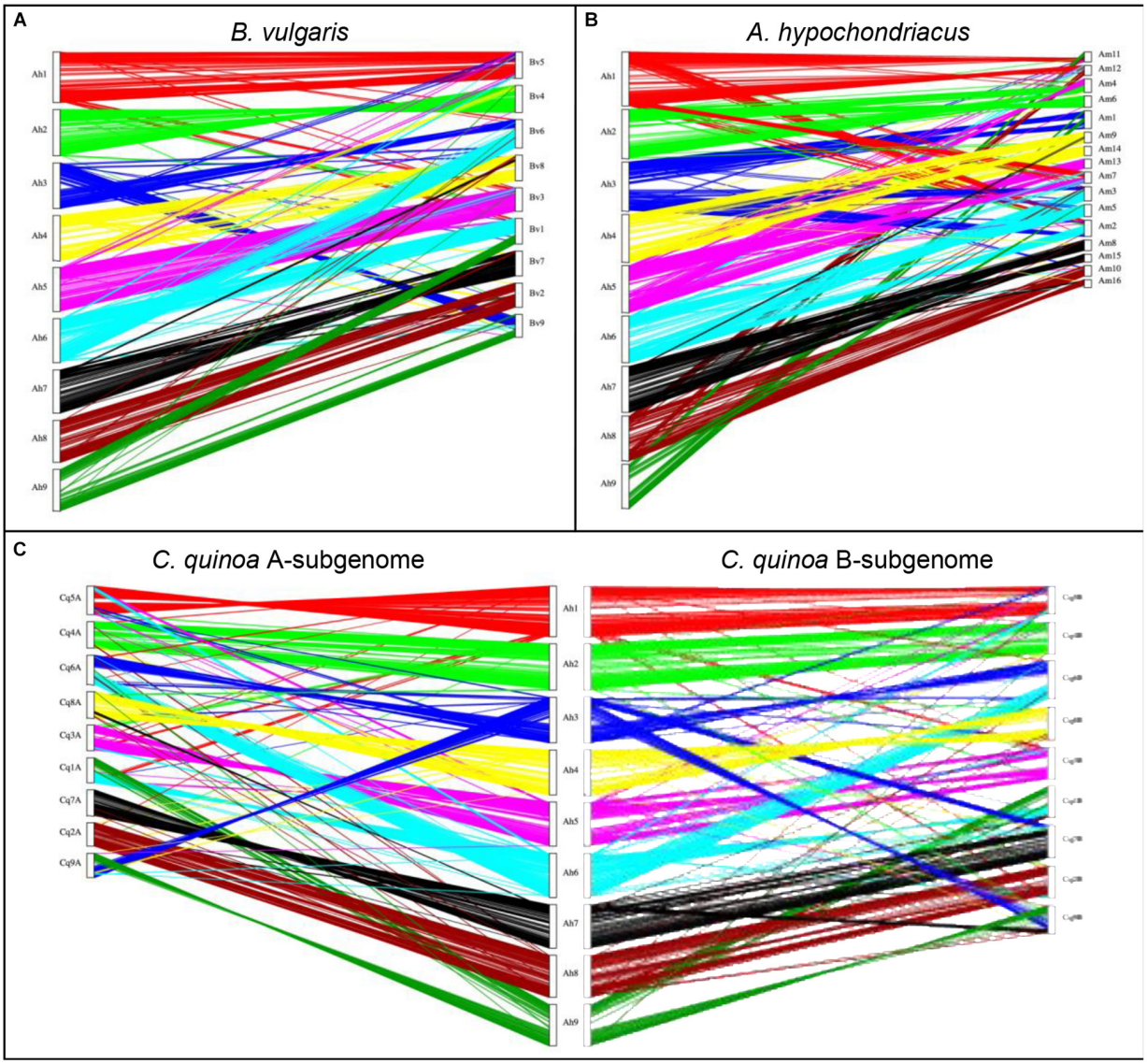


FIGURE 4 | Genomic comparison of *A. hortensis* with *B. vulgaris*, *A. hypochondriacus*, and *C. quinoa*. The dual synteny plots show syntenic regions between *A. hortensis* and *B. vulgaris* **(A)**, *A. hypochondriacus* **(B)** and *C. quinoa* **(C)** coding sequences. The dual synteny plot of the *C. quinoa* genome is divided into A- and B-subgenomes with *A. hortensis* in the center.

fusion event in *Amaranthus* as seen by the synteny plot of Ah3, where Ah3 aligns twice with Am1 (**Figure 5**; red arrow). Although many additional rearrangements are present which obscure one-to-one orthologous chromosome relationships with the known homeologous amaranth chromosomes (Lightfoot et al., 2017), several can be confirmed: Ah2 = Am4 (46%), Am6 (52%); Ah4 = Am9 (52%), Am14 (48%); Ah7 = Am8 (58%), Am15 (42%; **Table 6**).

To elucidate the timing of the evolutionary events that separate *Atriplex* from *C. quinoa*, *B. vulgaris* and *A. hypochondriacus*, we calculated the rate of synonymous substitutions per synonymous site (K_s) in duplicate gene-pairs between the species (**Figure 6**) using the CodeML (Yang, 2007) tool on the CoGe platform (genomevolution.org/coge). As expected, *C. quinoa* is most closely related to *A. hortensis*, with a clear peak present at Ks = 0.25, followed by *B. vulgaris* (Ks peak = 0.55), while *A. hypochondriacus*, as expected, is more distantly related, with a Ks peak = 0.7. The timing of the divergence events (time to last common ancestor) can be established using the Ks peak values and synonymous mutation rates, such as the core eukaryotic rate (8.1E–09) proposed by Lynch and Conery (2000) or with lineage specific rates, calibrated to the fossil record. Kadereit et al. (2003) used three paleobotanical fossils to establish a much lower synonymous substitution rate for *Chenopodioideae* (2.8–4.1E-09), which showed rate constancy among the lineages studied,

TABLE 6 | Orthologous genes were identified between *A. hortensis* and beet (A), *C. quinoa* (B) and *Am. hypochondriacus* (C) to detect orthologous chromosome relationships.

	Ah1	Ah2	Ah3	Ah4	Ah5	Ah6	Ah7	Ah8	Ah9	Total synteny spanned (bp)
(A)										
Bv5	75,985,214									75,985,214
Bv4		76,231,045		205,876						76,436,921
Bv6			32,474,644			38,024,142			1,018,214	71,517,000
Bv8		1,043,269		59,005,657						60,048,926
Bv3					76,200,679					76,200,679
Bv1						35,497,946			24,584,006	60,081,952
Bv7							68,960,315			68,960,315
Bv2								65,798,714		65,798,714
Bv9			31,029,170						29,976,099	61,005,269
(B)										
Cq5A	88,356,818		80,733							88,437,551
Cq5B	85,221,237									85,221,237
Cq4A		78,889,108						678,948		79,568,056
Cq4B		74,788,751		265,848						75,054,599
Cq6A			34,880,745			45,906,343				80,787,088
Cq6B			30,549,711			43,708,464				74,258,175
Cq8A		844,472		63,922,236	73,334					64,840,042
Cq8B		844,472		68,256,659						69,101,131
Cq3A					72,931,389	1,972,202				74,903,591
Cq3B					73,061,839					73,061,839
Cq1A						39,186,132			24,859,805	64,045,937
Cq1B						32,666,424		3,006,287	29,799,350	65,472,061
Cq7A							68,938,800			68,938,800
Cq7B			12,915,209				60,375,606	2,102,719		75,393,534
Cq2A								68,395,255		68,395,255
Cq2B		1,684,254		794,038		7,716,047		57,060,920		67,255,259
Cq9A			34,390,946						29,902,118	64,293,064
Cq9B			21,296,635				10,957,411		27,531,165	59,785,211
(C)										
Am11	34,077,858					1,414,077		5,216,151	7,553,686	48,261,772
Am12	31,526,741				604,945	329,285		9,294,151		41,755,122
Am4		41,122,991			30,756,424					71,879,415
Am6		46,411,045								46,411,045
Am1			47,400,936						31,083,078	78,484,014
Am9		609,018		45,747,049			521,887			46,877,954
Am14				41,664,113						41,664,113
Am13					44,814,461					44,814,461
Am7	6,842,070				17,248,169	1,330,917				25,421,156
Am3	1,167,442		24,822,048			25,821,038				51,810,528
Am5						25,008,790			12,043,145	37,051,935
Am2	23,542,392	1,607,269	15,871,055			16,950,212				57,970,928
Am8							55,245,256			55,245,256
Am15							40,042,725			40,042,725
Am10	1,585,265		54,032		144,414			37233979		39,017,690
Am16						283,173		16935472		17,218,645

Total syntenic bases are shown between all chromosome comparisons. Syntenic relationships are colored red and transition to white as the amount of synteny decreases.

suggesting that the *Amaranthaceae-Chenopodiaceae* have a lower nucleotide substitution rate than other angiosperms, including the *Arabidopsis* rate (1.5E-08). The CodeML workflow in CoGe identifies syntenic gene pairs between species, extracts coding sequences, and aligns protein sequences using the Needleman–Wunsch alignment algorithm, which is then back-translated to a codon alignment that is then used for Ks estimation. Using the lower substitution rates calculated by Kadereit et al. (2003), we date the last shared common ancestor between *A. hortensis* and *C. quinoa*, *B. vulgaris*, and

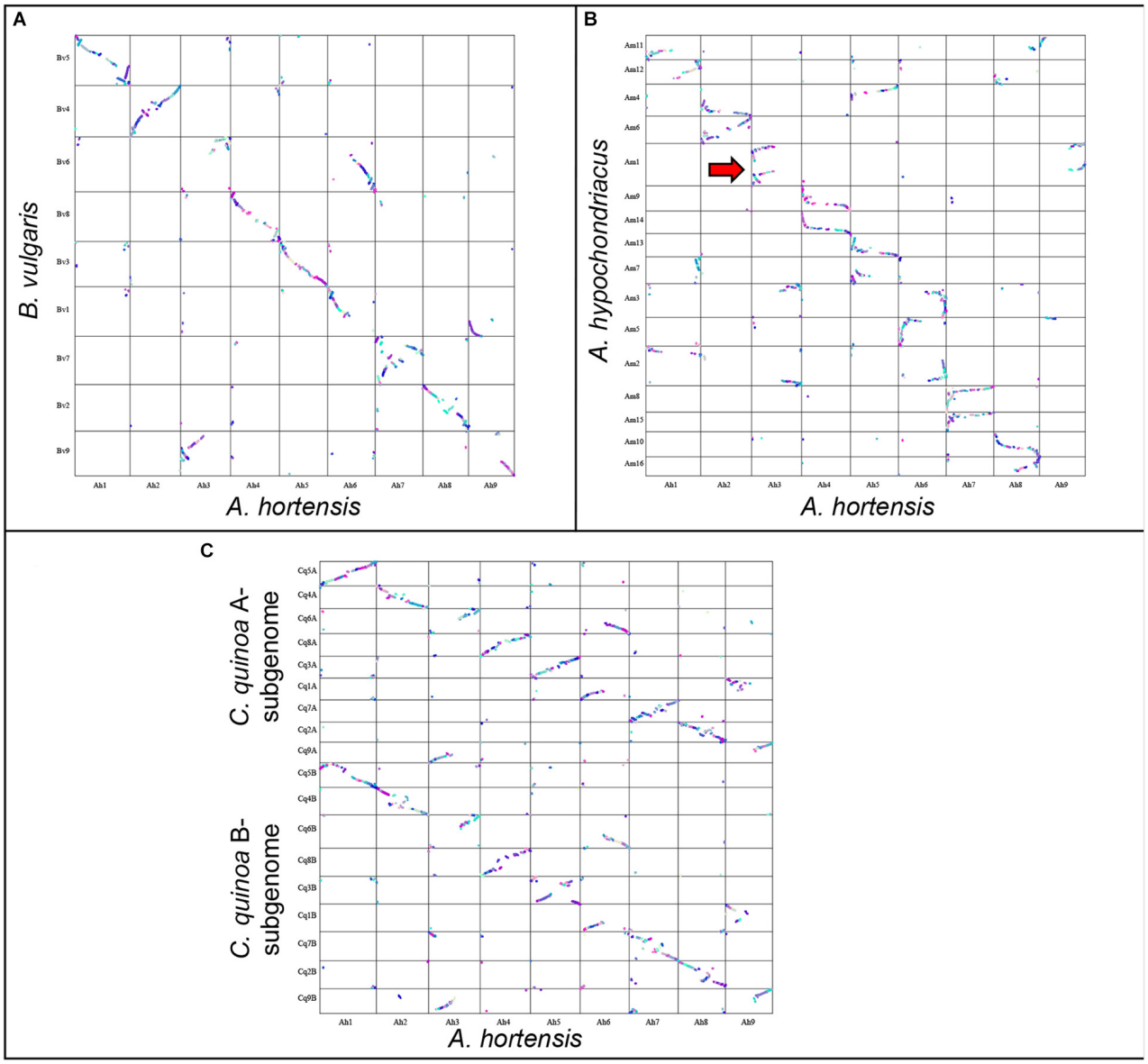


FIGURE 5 | Genomic comparison of *A. hortensis* with *B. vulgaris*, *Am. hypochondriacus* and *C. quinoa*. Synteny dotplot showing syntenic coding sequences between *A. hortensis* and *B. vulgaris* **(A)**, *Am. hypochondriacus* **(B)** and *C. quinoa* **(C)** coding sequences. Increasing color intensity is associated with increasing homology.

A. hypochondriacus to approximately 30.4 – 44.6 MYA, 67.1 – 98.2 MYA, and 85.3 –125 MYA, respectively.

Resequencing

A diversity panel consisting of 21 diverse accessions of *A. hortensis* (**Table 1**) was re-sequenced using Illumina paired-end sequencing, resulting in an average of 13X coverage (13.2 Gb) per accession. Following alignment and genotype likelihood calling with ANGSD (Korneliussen et al., 2014), a total 17,711,684 SNPs were filtered from the 846,491,542 sites analyzed using a 5% minimum minor allele frequency. A principal components analysis of the covariance data using PC1 and PC2 explained a total of 99.92% of the total variation and clearly identified three clusters of *Atriplex* accessions, which also agreed with our DeltaK analysis of the number of groups in the data set ($K = 3$; **Figure 7B**). Analysis of the 1000-bootstrap, consensus tree identified three distinct clades, with two accessions including the commercial cultivar Triple Purple and a wild accession collected in Alberta, Canada forming the first clade. The second clade consisted of four cultivated accessions of Serbia/Montenegro origin with a single wild accession at their root originating from France. The last and largest clade consisted of two subgroups with the first subgroup consisting of five cultivated lines from Serbia/Montenegro and a second subgroup consisting of four

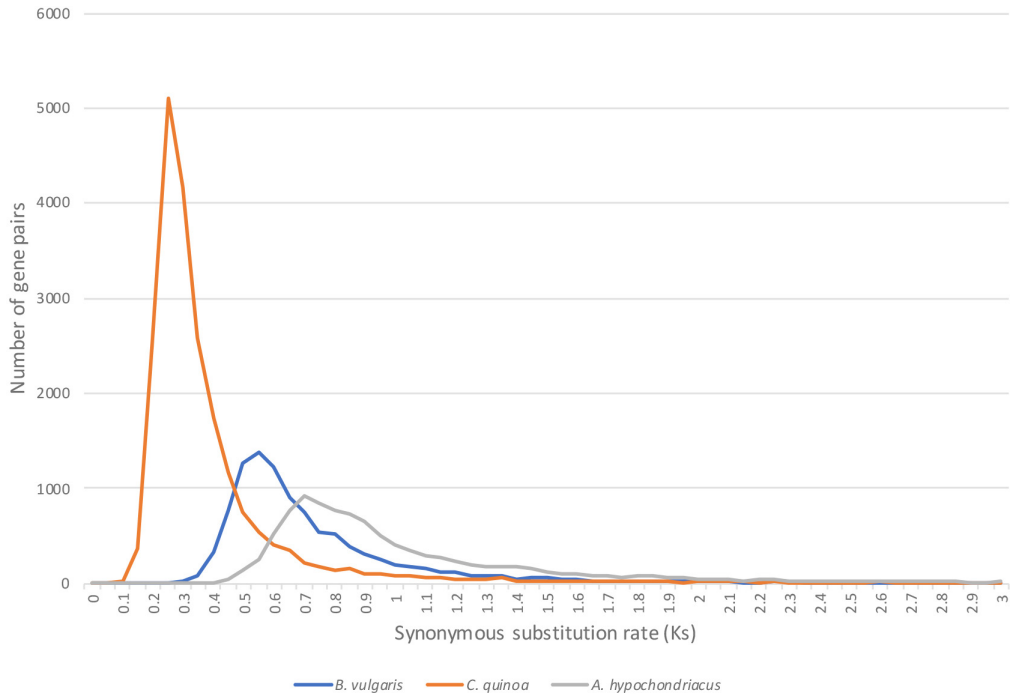


FIGURE 6 | The age of the evolutionary split between *A. hortensis* and *B. vulgaris*, *C. quinoa*, and *A. hypochondriacus* was estimated using the rate of synonymous substitutions per synonymous site (Ks) calculated from othologous gene pairs. *C. quinoa* is most closely related to *A. hortensis*, with a clear peak present at Ks = 0.25, followed by *B. vulgaris* (Ks peak = 0.55), while *A. hypochondriacus* is more distantly related, with a Ks peak = 0.7.

commercially available cultivars (obtained from Wild Garden Seed and Baker Creek Heirloom Seeds) and four accessions from disparate localities across Europe (Poland, Uzbekistan, Norway and Serbia) that were rooted by a wild accession collected in Utah, United States (**Figure 7A**). The structure plots (**Figure 7B**) indicate little to no admixture among the subpopulations, suggesting three distinct subpopulations with little to no interbreeding.

DISCUSSION

Multiple different libraries were prepared for ONT sequencing, including with and without fragmentation, to ascertain the influence of fragmentation on sequencing yield and read length – both important components of successful genome assembly. Fragmentation consistently improved throughput and yield, with the Covaris g-TUBEs producing the most effective and least variable fragmentation (i.e., based on sequencing yield and read length variation). The effect of centrifugation speed (3,800, 4,000, and 4,200 rpm) was also an important, albeit less controllable, factor. In general, higher centrifugation speeds produced higher yields, but concomitantly with decreased read lengths. Indeed, flow cell nanopores remained active for longer periods with fragmented libraries as compared to those without fragmentation. Kubota et al. (2019) demonstrated a similar correlation between DNA length and nanopore inactivity, with inactivity increasing exponentially in relation to increasing DNA molecule size. Nivala et al. (2013) suggested that one possible reason for this could be that longer molecules correlate with an increased presence of secondary and/or tertiary structures in the DNA molecules. Nanopores are restricted to the width of one DNA molecule at a time; thus, if secondary and/or tertiary structures are present in the DNA molecules, they increase the probability of clogging the nanopores, rendering them inactive. The combination of Covaris g-TUBE libraries prepared with differing centrifugation speeds resulted in a dataset with enough yield to provide ample coverage to compensate for the high error rate of ONT sequencing while still yielding long reads needed to span repetitive or otherwise problematic genomic regions (**Table 2**).

Canu (Koren et al., 2017), MaSuRCA (Zimin et al., 2013), Flye (Kolmogorov et al., 2019) and wtdbg2 (Ruan and Li, 2020) assemblers were used to assemble the ONT sequence data to ascertain which assembly program would perform best with the *A. hortensis* ONT sequence data. There were substantial differences in the overall time to finish an assembly, with wtdbg2 being the fastest of the assemblers tested. However, the MaSuRCA and Canu assemblies produced superior assemblies in terms of total contig size, N50, L50 and BUSCO statistics (**Figure 2**), with the polished Canu assembly ultimately being chosen as the draft genome for Hi-C scaffolding due to concerns of repeat collapse within the MaSuRCA assembly as reflected in its smaller total size of contigs assembled – a concern noted by Kolmogorov et al. (2019) who demonstrated the difficulty of assembling telomeric and centromeric chromosome regions

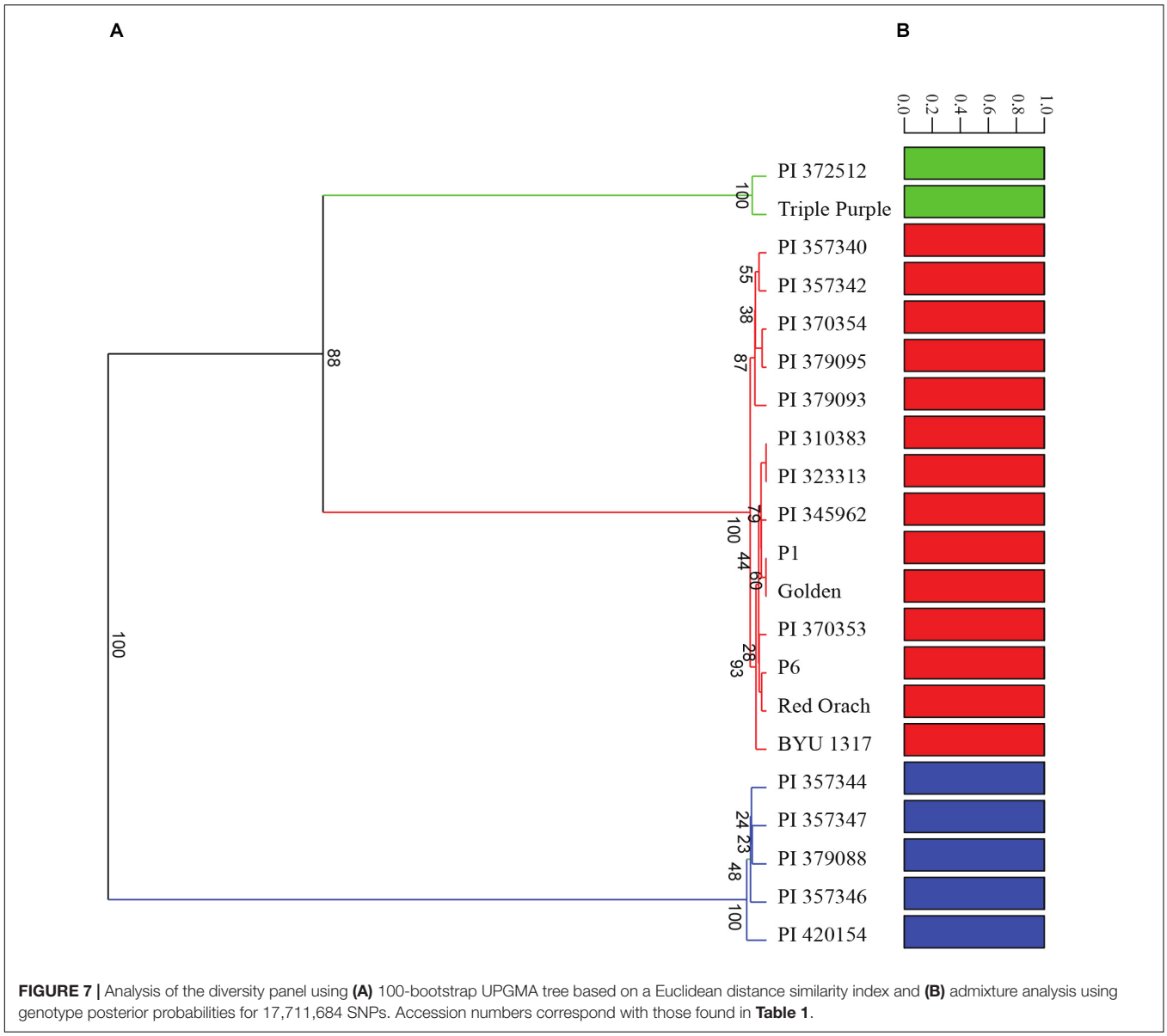


FIGURE 7 | Analysis of the diversity panel using **(A)** 100-bootstrap UPGMA tree based on a Euclidean distance similarity index and **(B)** admixture analysis using genotype posterior probabilities for 17,711,684 SNPs. Accession numbers correspond with those found in **Table 1**.

with the MaSuRCA assembler. Three rounds of polishing were conducted utilizing Nanopolish followed by two rounds of Illumina read-based Pilon correction. Nanopolish uses an index to detect misassemblies based on sequencing-generated signal levels generated from the original nanopore sequence data that correspond to likelihood ratios, while Pilon uses read alignments of high-quality Illumina reads to consensus-correct the draft genome (Walker et al., 2014; Loman et al., 2015). RACON (Vaser et al., 2017), another popular long-read consensus polisher that can use ONT and Illumina sequence for consensus correction, was also tested as a substitute for both Nanopolish and Pilon but showed no significant enhancement to the final BUSCO statistics (data not shown). We note that over-polishing an assembly can also be problematic, as seen by a decrease in BUSCO scores, and should therefore be avoided. In our assembly, a third round of polishing did not improve BUSCO scores.

Unsurprisingly, the B subgenome of *C. quinoa*, which is approximately 25% larger than the assembled A subgenome, shared more and longer syntenic blocks (209 vs 189; 2.9 vs 2.4 Mb average) with *A. hortensis*. The higher synteny with the B subgenome of *C. quinoa* may also reflect a closer ancestry of *A. hortensis* with the B subgenome – whose closest extant known *Chenopodium* species (*C. suecicum* Murr or *C. ficifolium* Sm.) are of Old-World origin, similar to that of *A. hortensis*. We note that the A-subgenome of *C. quinoa* is suspected to be of New World origin with its closest known extant species being *C. watsonii* A. Nelson, which is native to the southwestern United States (Jellen et al., 2019). It should, however, be noted that at least one diploid with the A-genome, *C. bryoniifolium* Bunge, is native to eastern Siberia (Walsh et al., 2015; Mandák et al., 2018a). Moreover, a new allohexaploid species containing the A-subgenome from *C. bryoniifolium* has been recently described from the Far East

of Russia as *C.* luteorubrum Mandák & Lomonosova (Mandák et al., 2018b). Thus, an East Asian origin of the A-genome lineage in *Chenopodium*, with its subsequent trans-Beringian migration and explosive diversification in the Americas, cannot be excluded at the present state of our knowledge; however, that scenario is less parsimonious than the New World origin of the A-genome lineage.

The genome of *A. hortensis* is highly repetitive with approximately 66.3% of the sequence containing interspersed repetitive sequence. By comparison, the genome of quinoa is 64.5% repetitive (Jarvis et al., 2017). Genomes that contain substantial repeat fractions can be difficult to assemble correctly. To overcome this challenge, Hi-C chromosome-contact maps were used for genome scaffolding which dramatically increased the continuity of the assembly, producing nine chromosome-sized scaffolds presumably representing each of the haploid chromosomes in *A. hortensis* (n = 9). Additionally, the Hi-C chromatin contact maps leverage the spatial orientation of the chromatin to identify and correct misassemblies in the overlap-layout-consensus assembly produced by Canu that potentially would have gone unnoticed. The nine chromosome pairs in *A. hortensis* are metacentric to slightly submetacentric (**Figure 1**). Due to the difficulty in assembling highly conserved and repetitive sequence regions within telomeres, the identification of 13 of the possible 18 telomeric ends is indicative of a highly complete, chromosome-scale genome assembly (**Figure 2**). The unexpected location of telomeric sequences in the subtelomeric region of one of the arms of chromosome Ah5 could reflect a potential assembly error – although careful inspection of the chromatin maps for this region do not show any indications of misassembly. Similar paracentric inversions have been seen in other species which result in telomere-specific tandem repeats being present in abnormal locations in plant chromosomes (Tek and Jiang, 2004). Nonetheless, additional research, potentially including optical mapping (e.g., BioNano genomics) and/or high-density linkage map development (neither of which have been developed for *A. hortensis*) should be targeted to this region to verify the orientation of this segment of the chromosome. Such investigations will also help verify the assemblies of chromosomes Ah3, Ah6, and Ah9, which show syntenic relationships with multiple *B. vulgaris* and *C. quinoa* chromosomes arms, thus obscuring their orthologous relationships. Such research would undoubtedly provide additional insight into the chromosomal evolution that characterizes the family *Chenopodiaceae* s. str. and the whole *Amaranthaceae*/*Chenopodiaceae* alliance – such as the homoelog loss and chromosomal fusion reported in *Amaranthus hypochondriacus* (Lightfoot et al., 2017).

It is not surprising that the North American-derived materials grouped with European accessions, as it is commonly understood that the center of origin of *A. hortensis* is the Trans-Himalayan (central Asia and Siberia) and Southeast European regions and that it was likely introduced during the third century B.C. into the Mediterranean littoral and from there to the Americas in Colonial times (Ruas et al., 2001). The species has become locally naturalized along riverbanks, roadsides, and ditches in parts of the Great Basin of North America (personal observations). There is also evidence of its use as a food in Switzerland as early as the Neolithic Age (Andrews, 1948), suggesting widespread, albeit ancient use of the species. Unfortunately, the United States National Plant Germplasm System curates only 45 *A. hortensis* accessions, of which very few are publicly available. The identification of three highly distinct clades, showing only limited admixture in our results, emphasizes the need for additional collections of wild and cultivated germplasm from throughout its native range, particularly in its presumed center of origin. Indeed, of the 45 curated accessions at the USDA, nearly two-thirds (28) are derived from a single European region corresponding to the Balkan Peninsula in Southeast Europe (Serbia-Macedonia). Subsequent phylogenetic analysis using materials from much broader geographic collections should improve our understanding of extant genetic variation and speciation processes within *A. hortensis*.

DATA AVAILABILITY STATEMENT

The raw sequences are deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive database under the BioProject ID PRJNA607334 with the following sequence read archive (SRR) accession numbers: SRR11147376 (Nanopore data), SRR11147367 – SRR11147368 (Hi-C), SRR11147369 – SRR11147375 (transcriptome) and SRR11123164 – SRR11123184 (resequencing panel; **Table 1**). Bulk data downloads, including annotations and BLAST analysis, and JBrowse viewing of the final Hi-C assembly are available at CoGe (https://genomevolution.org/coge/;Genomeid56906). The scaffold names in CoGe corresponding to specific pseudo chromosome assemblies are as follows: Scaffold_552_HRSCAF_710 = Ah1; Scaffold_579_HRSCAF_742 = Ah2; Scaffold_1312_HRSCAF_2063 = Ah3; Scaffold_481_HRSCAF_623 = Ah4; Scaffold_390_HRSCAF_510 = Ah5; Scaffold_1281_HRSCAF_1836 = Ah6; Scaffold_1313_HRSCAF_2064 = Ah7; Scaffold_1311_HRSCAF_2062 = Ah8; Scaffold_291_HRSCAF_384 = Ah9. In addition, the variant call file (VCF) for the diversity panel is available as a download from CoGe (ID# 15277).

AUTHOR CONTRIBUTIONS

PM, ENJ, and EWJ conceived and designed the study. SH and DL performed the sequencing experiments and managed the plant material. SM performed the flow cytometry experiments, and provided plant material, taxonomic information, and phylogenetic analysis. EWJ provided the assembly bioinformatics and expertise. BK preformed the fluorescent in situ hybridization experiments. PM, SH, and ENJ wrote the manuscript. All authors read and approved the final manuscript.

FUNDING

This work was supported by funding from General Mills, Inc., and internal funds of Brigham Young University.

ACKNOWLEDGMENTS

We gratefully acknowledge the assistance of the USDA, ARS, NALPGRU at Parlier, California, and Frank Morton at Wild Garden Seed Co., for seed contributions and to General Mills, Inc., for financial assistance in carrying out this project. We are also indebted to Isaac Clouse and other undergraduate students who helped take care of plants and performed DNA extractions. We are also grateful to Dr. Daniel J. Fairbanks (Utah Valley University, Orem, Utah, United States) for sharing his extensive insight of the *Atriplex* genus.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.00624/full#supplementary-material

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410.

Andrews, A. C. (1948). Orach as the spinach of the classical period. *Isis* 39, 169–172. doi: 10.1086/348950

APG IV (2016). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Bot. J. Linnean Soc.* 181, 1–20. doi: 10.1016/j.jep.2015.05.035

APG (Angiosperm Phylogeny Group) (1998). An ordinal classification for the families of flowering plants. *Ann. Missouri Bot. Garden* 85, 531–553.

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Brignone, N. F., Pozner, R. E., and Denham, S. S. (2019). Origin and evolution of *Atriplex* (*Amaranthaceae* s.l.) in the Americas: unexpected insights from South American species. *Taxon* 68, 1021–1036.

De Coster, W., D'Hert, S., Schultz, D. T., Cruts, M., and Van Broeckhoven, C. (2018). NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* 34, 2666–2669. doi: 10.1093/bioinformatics/bty149

Dolezel, J., Bartos, J., Voglmayr, H., and Greilhuber, J. (2003). Nuclear DNA content and genome size of trout and human. *Cytometry A* 51, 127–128. doi: 10.1002/cyto.a.10013

Du, J., Tian, Z., Hans, C. S., Laten, H. M., Cannon, S. B., Jackson, S. A., et al. (2010). Evolutionary conservation, diversity and specificity of LTR-retrotransposons in flowering plants: insights from genome-wide analysis and multi-specific comparison. *Plant J.* 63, 584–598. doi: 10.1111/j.1365-313X.2010.04263.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

Flavell, R. B., Bennett, M. D., Smith, J. B., and Smith, D. B. (1974). Genome size and the proportion of repeated nucleotide sequence DNA in plants. *Biochem. Genet.* 12, 257–269. doi: 10.1007/bf00485947

Fuentes-Bazan, S., Uotila, P., and Borsch, T. (2012). A novel phylogeny-based generic classification for *Chenopodium* sensu lato, and a tribal rearrangement of *Chenopodioideae* (*Chenopodiaceae*). *Willdenowia* 42, 5–24.

Funk, A., Galewski, P., and McGrath, J. M. (2018). Nucleotide-binding resistance gene signatures in sugar beet, insights from a new reference genome. *Plant J.* 95, 659–671. doi: 10.1111/tpj.13977

Galindo-Gonzalez, L., Mhiri, C., Deyholos, M. K., and Grandbastien, M. A. (2017). LTR-retrotransposons in plants: engines of evolution. *Gene* 626, 14–25. doi: 10.1016/j.gene.2017.04.051

Haas, B. J., Delcher, A. L., Wortman, J. R., and Salzberg, S. L. (2004). DAGchainer: a tool for mining segmental genome duplications and synteny. *Bioinformatics* 20, 3643–3646. doi: 10.1093/bioinformatics/bth397

Hammer, Ø, Harper, D. A. T., and Ryan, P. D. (2001). PAST: paleontological statistics software package for education and data analysis. *Palaeontol. Electronica* 4, 1–9.

Hernández-Ledesma, P., Berendsohn, W. G., Borsch, T., Mering, S. V., Akhani, H., Arias, S., et al. (2015). A taxonomic backbone for the global synthesis of species diversity in the angiosperm order *Caryophyllales*. *Willdenowia* 45, 281–383.

Holt, C., and Yandell, M. (2011). MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* 12:491. doi: 10.1186/1471-2105-12-491

Jarvis, D. E., Ho, Y. S., Lightfoot, D. J., Schmockel, S. M., Li, B., Borm, T. J. A., et al. (2017). The genome of *Chenopodium quinoa*. *Nature* 542, 307–312. doi: 10.1038/nature21370

Jellen, E. N., Jarvis, D. E., Hunt, S. P., Mangelsen, H. H., and Maughan, P. J. (2019). New seed collections of North American pitseed goosefoot (*Chenopodium berlandieri*) and efforts to identify its diploid ancestors through whole-genome sequencing. *Cienc. Investig. Agrar.* 46, 187–196.

Kadereit, G., Borsch, T., Weising, K., and Freitag, H. (2003). Phylogeny of Amaranthaceae and Chenopodiaceae and the evolution of C4 photosynthesis. *Int. J. Plant Sci.* 164, 959–986.

Kadereit, G., Mavrodiev, E. V., Zacharias, E. H., and Sukhorukov, A. P. (2010). Molecular phylogeny of Atripliceae (Chenopodioideae, Chenopodiaceae): implications for systematics, biogeography, flower and fruit evolution, and the origin of C4 photosynthesis. *Am. J. Bot.* 97, 1664–1687. doi: 10.3732/ajb.1000169

Kim, D., Landmead, B., and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12, 357–360. doi: 10.1038/nmeth.3317

Kolano, B., Tomczak, H., Molewska, R., Jellen, E. N., and Maluszynska, J. (2012). Distribution of 5S and 35S rRNA gene sites in 34 *Chenopodium* species (Amaranthaceae). *Bot. J. Linnean Soc.* 170, 220–231.

Kolmogorov, M., Yuan, J., Lin, Y., and Pevzner, P. A. (2019). Assembly of long error-prone reads using repeat graphs. *Nat. Biotechnol.* 37, 540–546. doi: 10.1038/s41587-019-0072-8

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., and Phillippy, A. M. (2017). Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Genome Res.* 27, 722–736. doi: 10.1101/gr.215087.116

Korneliussen, T. S., Albrechtsen, A., and Nielsen, R. (2014). ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics* 15:356. doi: 10.1186/s12859-014-0356-4

Kubota, T., Lloyd, K., Sakashita, N., Minato, S., Ishida, K., and Mitsui, T. (2019). Clog and release, and reverse motions of DNA in a nanopore. *Polymers (Basel)* 11:84. doi: 10.3390/polym11010084

Lanfear, R., Schalamun, M., Kainer, D., Wang, W., and Schwessinger, B. (2019). MinIONQC: fast and simple quality control for MinION sequencing data. *Bioinformatics* 35, 523–525. doi: 10.1093/bioinformatics/bty654

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923

Li, S. F., Guo, Y. J., Li, J. R., Zhang, D. X., Wang, B. X., Li, N., et al. (2019). The landscape of transposable elements and satellite DNAs in the genome of a dioecious plant spinach (*Spinacia oleracea* L.). *Mobile DNA* 10:3. doi: 10.1186/s13100-019-0147-6

Lightfoot, D. J., Jarvis, D. E., Ramaraj, T., Lee, R., Jellen, E. N., and Maughan, P. J. (2017). Single-molecule sequencing and Hi-C-based proximity-guided assembly of amaranth (*Amaranthus hypochondriacus*) chromosomes provide insights into genome evolution. *BMC Biol.* 15:74. doi: 10.1186/s12915-017-0412-4

Loman, N. J., Quick, J., and Simpson, J. T. (2015). A complete bacterial genome assembled de novo using only Nanopore sequencing data. *Nat. Methods* 12, 733–735. doi: 10.1038/nmeth.3444

Lu, H., Giordano, F., and Ning, Z. (2016). Oxford Nanopore MinION sequencing and genome assembly. *Genomics Proteomics Bioinformatics* 14, 265–279. doi: 10.1016/j.gpb.2016.05.004

Lynch, M., and Conery, J. S. (2000). The evolutionary fate and consequences of duplicate genes. *Science* 290, 1151–1155. doi: 10.1126/science.290.5494.1151

Mandák, B., Krak, K., Vít, P., Lomonosova, M. N., Belyayev, A., Habibi, F., et al. (2018a). Hybridization and polyploidization within the *Chenopodium album* aggregate analysed by means of cytological and molecular markers. *Mol. Phylogenetics Evol.* 129, 189–201. doi: 10.1016/j.ympev.2018.08.016

Mandák, B., Lomonosova, M. N., and Mosyakin, S. L. (2018b). *Chenopodium luteorubrum* (*Chenopodiaceae*/*Amaranthaceae* sensu APG): a remarkable new species from the Russian Far East. *Phytotaxa* 382, 102–112.

Mangelson, H., Jarvis, D. E., Mollinedo, P., Rollano-Penaloza, O. M., Palma-Encinas, V. D., Gomez-Pando, L. R., et al. (2019). The genome of *Chenopodium pallidicaule*: an emerging Andean super grain. *Appl. Plant Sci.* 7:e11300. doi: 10.1002/aps3.11300

Maughan, P. J., Kolano, B. A., Maluszynska, J., Coles, N. D., Bonifacio, A., Rojas, J., et al. (2006). Molecular and cytological characterization of ribosomal RNA genes in *Chenopodium quinoa* and *Chenopodium berlandieri*. *Genome* 49, 825–839. doi: 10.1139/g06-033

McArthur, E. D., Stevens, R., and Blauer, A. C. (1983). Growth performance comparisons among 18 accessions of fourwing saltbush [*Atriplex canescens*] at two sites in central Utah. *J. Range Mgmt.* 36:78.

Meisner, J., and Albrechtsen, A. (2018). Inferring population structure and admixture proportions in low-depth NGS data. *Genetics* 210, 719–731. doi: 10.1534/genetics.118.301336

Michael, T. P. (2014). Plant genome size variation: bloating and purging DNA. *Brief Funct. Genomics* 13, 308–317. doi: 10.1093/bfgp/elu005

Moll, K. M., Zhou, P., Ramaraj, T., Fajardo, D., Devitt, N. P., Sadowsky, M. J., et al. (2017). Strategies for optimizing BioNano and Dovetail explored through a second reference quality assembly for the legume model, *Medicago truncatula*. *BMC Genomics* 18:578. doi: 10.1186/s12864-017-3971-4

Morales-Briones, D. F., Kadereit, G., Tefarikis, D. T., Moore, M. J., Smith, S. A., Brockington, S. F., et al. (2020). Disentangling sources of gene tree discordance in phylotranscriptomic datasets: a case study from *Amaranthaceae* s.l. *bioRxiv* [Preprint]. doi: 10.1101/794370

Mosyakin, S. L., and Iamonico, D. (2017). Nomenclatural changes in *Chenopodium* (incl. *Rhagodia*) (*Chenopodiaceae*), with considerations on relationships of some Australian taxa and their possible Eurasian relatives. *Nuytsia* 28, 255–271.

Mosyakin, S. L., and Walter, J. (2018). (2660) Proposal to conserve the name *Chenopodium quinoa* (*Chenopodiaceae*/*Amaranthaceae* sensu APG) against *C. purpurascens* and *C. punctulatum*. *Taxon* 67, 1222–1223.

Nivala, J., Marks, D. B., and Akeson, M. (2013). Unfoldase-mediated protein translocation through an α-hemolysin nanopore. *Nat. Biotechnol.* 31, 247Ű250. doi: 10.1038/nbt.2503

Ou, S. J., and Jiang, N. (2018). LTR_retriever: a highly accurate and sensitive program for identification of long terminal repeat retrotransposons. *Plant Physiol.* 176, 1410–1422. doi: 10.1104/pp.17.01310

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T. C., Mendell, J. T., and Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* 33, 290–295. doi: 10.1038/nbt.3122

Putnam, N. H., O'Connell, B. L., Stites, J. C., Rice, B. J., Blanchette, M., Calef, R., et al. (2016). Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. *Genome Res.* 26, 342–350. doi: 10.1101/gr.193474.115

Richards, E. J., and Ausubel, F. M. (1988). Isolation of a higher eukaryotic telomere from *Arabidopsis thaliana*. *Cell* 53, 127–136. doi: 10.1016/0092-8674(88)90494-1

Rinchen, T., Narendra, S., Samar Bahadur, M., Soni, V., Phour, M., and Kumar, B. (2017). Morphological characterization of indigenous vegetable ('*Atriplex hortensis*' L.) from trans-Himalayan region of Ladakh (Jammu and Kashmir), India. *Australian J. Crop Sci.* 11, 258–263.

Ruan, J., and Li, H. (2020). Fast and accurate long-read assembly with wtdbg2. *Nat. Methods* 17, 155–158. doi: 10.1038/s41592-019-0669-3

Ruas, C. F., Ruas, P. M., Stutz, H. C., and Fairbanks, D. J. (2001). Cytogenetic studies in the genus *Atriplex* (Chenopodiaceae). *Caryologia* 54, 129–145.

Sai Kachout, S., Ben Mansoura, A., Jaffel Hamza, K., Leclerc, J. C., Rejeb, M. N., and Ouerghi, Z. (2011). Leaf–water relations and ion concentrations of the halophyte *Atriplex hortensis* in response to salinity and water stress. *Acta Physiol. Plant.* 33, 335–342.

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210–3212. doi: 10.1093/bioinformatics/btv351

Simcox, J., and Stonescu, I. (2014). *Ten promising crops for Western and Northern Europe. Analysis and Recommendations*. Utrecht: Kiemkracht, 56.

Simon, R. D., Abeliovich, A., and Belkin, S. (1994). A novel terrestrial halophilic environment: the phylloplane of *Atriplex halimus*, a salt-excreting plant. *FEMS Microbiol. Ecol.* 14, 99–109.

Smit, A. F. A., and Hubley, R. (2008–2015). *RepeatModeler Open-1.0*. Available online at: http://www.repeatmasker.org (accessed January 22, 2020).

Smit, A. F. A., Hubley, R., and Green, P. (2013–2015). *RepeatMasker Open-4.0*. Available online at: http://www.repeatmasker.org (accessed January 22, 2020).

Smith, T. W., Kron, P., and Martin, S. L. (2018). flowPloidy: An R package for genome size and ploidy assessment of flow cytometry data. Applications in plant sciences, 6(7), e01164. doi: 10.1002/aps3.1164

Stanke, M., Steinkamp, R., Waack, S., and Morgenstern, B. (2004). AUGUSTUS: a web server for gene finding in eukaryotes. *Nucleic Acids Res.* 32, W309–W312. doi: 10.1093/nar/gkh379

Storchova, H., Drabesova, J., Chab, D., Kolar, J., and Jellen, E. N. (2015). The introns in FLOWERING LOCUS T-LIKE (FTL) genes are useful markers for tracking paternity in tetraploid Chenopodium quinoa Willd. *Genet. Resour. Crop Evol.* 62, 913–925.

Sukhorukov, A. P. (2006). Zur systematik und chorologie der in russland und benachbarten staaten (in den Grenzen der ehemaligen UdSSR) vorkommenden *Atriplex*-Arten (*Chenopodiaceae*). *Ann. Naturhist. Mus. Wien Ser. B* 108, 307–420.

Sukhorukov, A. P. (2014). *The Carpology of the Family Chenopodiaceae in Relations to Problems of Phylogeny, Systematics and Diagnostics of its Representatives*. Tula: Grif i K, 400.

Tanaka, Y., Sasaki, N., and Ohmiya, A. (2008). Biosynthesis of plant pigments: anthocyanins, betalains and carotenoids. *Plant J.* 54, 733–749. doi: 10.1111/j.1365-313X.2008.03447.x

Tek, A., and Jiang, J. (2004). The centromeric regions of potato chromosomes contain megabase-sized tandem arrays of telomere-similar sequence. *Chromosoma* 113, 77–83. doi: 10.1007/s00412-004-0297-1

Todd, J. J., and Vodkin, L. O. (1996). Duplications that suppress and deletions that restore expression from a chalcone synthase multigene family. *Plant Cell* 8, 687–699. doi: 10.1105/tpc.8.4.687

Turland, N. J., Wiersema, J. H., Barrie, F. R., Greuter, W., Hawksworth, D. L., Herendeen, P. S., et al. (2018). *International Code of Nomenclature for Algae, Fungi, and Plants (Shenzhen Code) Adopted by the Nineteenth International Botanical Congress, Shenzhen, China, July 2017. [Regnum Vegetabile, vol. 159]*. Glashütten: Koeltz Botanical Books, 254.

Vaser, R., Soviæ, I., Nagarajan, N., and Šikiæ, M. (2017). Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.* 27, 737–746. doi: 10.1101/gr.214270.116

Vasjukov, V. M. (2015). New genus *Sukhorukovia* (*Chenopodiaceae*). *Botanika (Minsk)* 44, 119–123.

Vickerman, D. B., Shannon, M. C., Bañuelos, G. S., Grieve, C. M., and Trumble, J. T. (2002). Evaluation of *Atriplex* lines for selenium accumulation, salt tolerance and suitability for a key agricultural insect pest. *Environ. Pollut.* 120, 463–473. doi: 10.1016/s0269-7491(02)00116-1

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., et al. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963. doi: 10.1371/journal.pone.0112963

Walsh, B. M., Adhikary, D., Maughan, P. J., Emshwiller, E., and Jellen, E. N. (2015). *Chenopodium* polyploidy inferences from *Salt Overly Sensitive 1* (*SOS1*) data. *Am. J. Bot.* 102, 533–543. doi: 10.3732/ajb.1400344

Welsh, S. L., and Crompton, C. (1995). Names and types in perennial *Atriplex* Linnaeus (Chenopodiaceae) in North America selectively exclusive of Mexico. *Great Basin Nat.* 55, 322–334.

Wick, R. R., Judd, L. M., Gorrie, C. L., and Holt, K. E. (2017). Completing bacterial genome assemblies with multiplex MinION sequencing. *Microb. Genom.* 3:3000132. doi: 10.1099/mgen.0.000132

Wright, K. H., Pike, O. A., Fairbanks, D. J., and Huber, C. S. (2002). Composition of *Atriplex hortensis*, sweet and bitter *Chenopodium quinoa* seeds. *J. Food Sci.* 67, 1383–1385.

Yan, H., Martin, S. L., Bekele, W. A., Latta, R. G., Diederichsen, A., Peng, Y., et al. (2016). Genome size variation in the genus *Avena*. *Genome* 59, 209–220. doi: 10.1139/gen-2015-0132

Yang, Z. H. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24, 1586–1591. doi: 10.1093/molbev/msm088

Zacharias, E. H., and Baldwin, B. G. (2010). A molecular phylogeny of North American *Atripliceae* (*Chenopodiaceae*), with implications for floral and photosynthetic pathway evolution. *Syst. Bot.* 35, 839–857.

Zimin, A. V., Marçais, G., Puiu, D., Roberts, M., Salzberg, S. L., and Yorke, J. A. (2013). The MaSuRCA genome assembler. *Bioinformatics* 29, 2669–2677. doi: 10.1093/bioinformatics/btt476

Conflict of Interest: The authors declare that this study received funding from General Mills, Inc. The funder had provided financial assistance based on an initial interest in investigating development of new specialty crops. The authors declare that the funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication. EJ was initially employed by General Mills and subsequently left to form 25:2 Solutions.

The remaining 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 © 2020 Hunt, Jarvis, Larsen, Mosyakin, Kolano, Jackson, Martin, Jellen and Maughan. 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) and the copyright owner(s) 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 Plant Science

ORIGINAL RESEARCH
published: 29 May 2020
doi: 10.3389/fpls.2020.00498

Seed Structural Variability and Germination Capacity in *Passiflora edulis* Sims f. *edulis*

Nohra Rodríguez Castillo[1], Luz Marina Melgarejo[1] and Matthew Wohlgemuth Blair[2]**

[1] *Laboratory of Plant Physiology and Biochemistry, Department of Biology, Universidad Nacional de Colombia, Bogotá, Colombia,* [2] *Department of Agricultural and Environmental Sciences, Tennessee State University, Nashville, TN, United States*

Purple passion fruit, *Passiflora edulis* Sims f. *edulis*, is an important semi-perennial, fruit bearing vine originating in South America that produces a commercial tropical juice pulp for international and national consumption. Within the round purple passion fruit are a large number of membranous seed sacs each containing individual seeds. Little is known about the seed anatomy of the commercial passion fruit, differences between wild collected and commercial types, and its effect on seedling germination. Therefore, our main objective for this study was to analyze the seed anatomy variability of different germplasm as well as the effect on viability and germination of the seeds of this species. Germplasm was evaluated from three sources: (1) commercial cultivars grown in current production areas, (2) genebank accession from the national seed bank, and (3) landraces collected across different high and mid-elevation sites of the Andean region. A total of 12 morphometric descriptors related to seed anatomy were evaluated on the 56 genotypes, of which three were most informative: Angle to the vertex which is related to the shape of the seed, the thickness of the tegument and the horizontal length; separating the seed according to its source of origin. Germination was found to be positively correlated with the number (r = 0.789) and depth (r = 0.854) of seed pitting. Seeds of the commercial cultivars had more seed pits and higher germination compared to seeds of landraces or genebank accessions showing a possible effect of domestication on the crop. Interestingly, passion fruits often germinate during the rainy season as escaped or wild seedlings especially in the disturbed landscapes of coffee plantations, so some dormancy is needed but faster germination is needed for intensive cultivation. Harnessing passion fruit diversity would be useful as the semi-domesticated landraces have valuable adaptation characteristics to combine with rapid germination selected in the commercial cultivars. The variability of seed pitting with cultivars more pitted than landraces possibly resulting in faster germination may indicate that purple passion fruit is still undergoing a process of selection and domestication for this trait.

Keywords: morphometric variability, seed anatomy, germination, viability, descriptors

OPEN ACCESS

Edited by:
Petr Smýkal,
Palacký University Olomouc, Czechia

Reviewed by:
Alma Orozco Segovia,
National Autonomous University of Mexico, Mexico
Anca Macovei,
University of Pavia, Italy

****Correspondence:***
Luz Marina Melgarejo
lmmelgarejom@unal.edu.co
Matthew Wohlgemuth Blair
mblair@tnstate.edu

Specialty section:
This article was submitted to Plant Breeding, a section of the journal Frontiers in Plant Science

Received: *11 November 2019*
Accepted: *03 April 2020*
Published: *29 May 2020*

Citation:
Castillo NR, Melgarejo LM and Blair MW (2020) Seed Structural Variability and Germination Capacity in Passiflora edulis Sims f. edulis. Front. Plant Sci. 11:498. doi: 10.3389/fpls.2020.00498

INTRODUCTION

Seed structural variability is related to germination capacity and can be a step in the process of domestication that distinguishes wild accessions from semi or fully domesticated crops (Koomneef et al., 2002). Fruit crops propagated by seed may be subject to this type of selection given that they are often grown in hedgerows rather than in row crop agriculture. Purple passion fruit, *Passiflora*

edulis Sims f. *edulis*, is an important seed-propagated, semi-perennial fruit bearing vine, which bears fleshy, and spherical "pepo type" berries with purple rinds that are filled with up to 200 small black seeds (Morton, 1987). The species originated in South America and exports from there and other tropical regions make it a major source of fruit pulp (Thokchom and Mandal, 2017). While passion fruit species can be propagated clonally (Salomão et al., 2002; Gelape et al., 2019), to date there are no seedless varieties and all are seed propagated or grafted onto seedlings. Thus, seed physiology is important to study.

Within the purple passion fruit are the many membranous seed sacs containing individual seeds (Morton, 1987). These are produced within a single ovary after self-pollination assisted by carpenter bees (Garcia, 2010; de Lira et al., 2016). The amount of seeds per fruit and the seed size and shape typically vary within and between species of *Passiflora* (Pérez-Cortéz et al., 2002). When pressed for juice the pithy rind is cracked and liquid is removed from the sacs with the seed remaining as a by-product (Macedo et al., 2015). Although the seeds are edible they are poorly studied but contain interesting nutritional and biochemical properties (Devi et al., 2018). The species is of great importance for varietal maintenance and production (Morton, 1987) but has spatula type embryos that are inhibited from germination by physical and physiological dormancy (Black, 1972; Baskin and Baskin, 1998; Koomneef et al., 2002).

Little is known about the seed anatomy of purple passion fruit and its effect on seedling germination or its variability among landraces and wild collected accessions. Seed characteristics are key to the seed dormancy, seedling propagation and plant survival of passion fruits (de Oliveira et al., 2010), and plants in general (Baskin and Baskin, 1998). The sweet flesh around the seeds of passion fruits are important for seed dispersal by birds, serving as an attractant and motivator for them to pick through the thick egg-like rind around the fruit (Torres, 2018). The round fruits can fall off naturally from the vine after a long period of drying on the plant and can roll downhill to settle in a moist ditch, where they are likely to germinate as a mass of seedlings (Veiga et al., 2013). This form of whole fruit dispersal can be considered advantageous in the natural environment of *Passiflora* spp., which grow mostly in hillside cloud forests of South America (Delanoy et al., 2006).

Dormancy and germination rates are other factors that can control the number of progenies a plant can make and is variable both in space and time with large environment effects (Smýkal et al., 2014). The semi-domesticated passion fruit have been shown to have strong dormancy effects (de Oliveira et al., 2010; Torres, 2018). Slow germination for wild passion fruit seed can be overcome by stratification and temperature (Ramírez et al., 2015). The effect of seed structure on dormancy is less well studied than biochemical factors that inhibit germination in cultivated passion fruit (Araújo et al., 2007). Seed coat shape and thickness have been shown in some other plant families to affect germination (Chernoff et al., 1992; Gabr, 2014). In this study, we observed differences in the seed coat pitting abundance of the lignified tegmen of purple passion fruit genotypes which relate to viability and ultimately to germination. Seed coat characteristics of dry seed influence electrical conductivity and desiccation tolerance, which in turn affect seed preservation strategies as well as viability/germination (Baskin and Baskin, 1998; Veiga et al., 2013; Mira et al., 2015).

Dry seed characteristics, the focus of this study while important for plant establishment can be related to fruit characteristics as well (Macedo et al., 2015). The ratio of seed to pulp and juice has been investigated in purple passion fruit by comparing fresh mucilaginous and dry seed weight and is one factor that distinguishes commercial cultivars and landraces, and economic yield of the crop (Rodríguez et al., 2019, 2020). Seed/pulp ratio along with fruit size and fruit wall thickness are important parameters for crop improvement in breeding programs in East Africa and Asia (Matheri et al., 2016; Liu et al., 2017) but little work has been done on seed size, germination rate and their relation to pulp mass.

The objective of this research was to evaluate purple passion fruit genotypes from different sources for seed coat morphology and other seed characteristics, and associate these with seed viability as reflected in dormancy and time to or capacity to germinate. The germplasm sources used were commercial cultivars, genebank accessions and landraces collected from around Colombia. The hypothesis of lower dormancy and faster germination for commercial cultivars of purple passion fruit seeds was tested by comparing the average for these genotypes compared to the other two groups. The seed was not pre-treated and therefore differences may be a result of selection during commercialization or even during domestication of purple passion fruit genotypes. While seed morphology has been studied as a characteristic to distinguish different species within the Passifloraceae family, this is the first study to evaluate intra-specific variability for seed morpho-anatomical traits and adds to the data for tropical plants found in Baskin and Baskin (1998). This study showed that it was important to understand the processes of seed trait selection in cultivar development and to conserve the purple passion fruit landraces and genetics resources in the Andean region both for conservations and production of the crop. In Colombia, the purple passion fruit is the fourth most important export in the category of fresh or processed fruits, after bananas, berries, and pineapples (Fischer et al., 2009). Despite its importance, the purple passion fruit remains poorly investigated agronomically as well especially compared to the yellow passion fruit, which is more prevalent throughout tropical lowland areas of Brazil, Central America, and the Caribbean (Cerqueira et al., 2014; Marostega et al., 2017). Although purple passion fruit is now found in the highland tropics of Africa and sub-tropical regions, its original range was in mid to higher altitude elevations of the Andes (Franco et al., 2014).

MATERIALS AND METHODS

Plant Material and Sites of Collection

Germplasm of purple passion fruit (*P. edulis* f. *edulis* Sims) was collected from three different sources: 8 commercial cultivars from farmers' fields in the department of Cundinamarca and Boyacá, 14 genebank accessions were from the field grown *in situ* genebank collection kept by Agrosavia (ex. Corpoica)

at the Llanogrande Experiment station in Rionegro, Antioquia and 34 were landraces from all over Colombia collected in the year prior to the study. This gave a total of 56 genotypes. The landraces were collected from backyard gardens, roadsides or small farms where the purple passion fruits were semi-wild or isolated plants rather than the main crop. This was common to many departments in the country since purple passion fruit is widely distributed in both the coffee region and vegetable or potato/corn and bean farms of the highland region. Departments of Colombia represented by the landraces included Antioquia (2 genotypes), Boyacá (6), Cauca (2), Cundinamarca (8), Huila (7), Nariño (5), Norte de Santander (2), Putumayo (4), Quindío (2), Risaralda (2), Santander (1), and Tolima (1) as found in **Supplementary Table 1**.

Seed Viability and Germination

Seed viability was determined for the 56 genotypes by staining with 1% tetrazolium, an accepted process according to Sawma and Mohler (2002). Percentage seed germination was evaluated on a standard medium based on the method of Suárez and Melgarejo (2010) using 100 seeds from every accession grown on experimental plots described by Rodríguez et al. (2019, 2020). We were studying dried seed with mucilage and pulp removed and therefore our focus was on the pitted seed coat/integument alone.

Anatomical and Morphological Traits

Dry seed from the same source described above were evaluated for a total of 12 morphological traits: seven of these were morphometric descriptors suggested by Tangarife et al. (2009), Santos et al. (2014), Marostega et al. (2017); while five were novel characteristics of seed shape and seed pitting that were new to this study. Two traits were evaluated qualitatively; namely FOR = seed form and COL = seed color. Seed form was based on shapes being divided into elliptical or semi-elliptical, and closer to oval. Given the pointed ends to the passion fruit seed none were truly oval but rather semi-elliptical. Seed color were ranked as black or brown and had no arils on them. Seed pits were present on all accessions so were evaluated for their prevalence or area coverage of the seed coat and depth into the seed coat (integument). These pits consist in semi-circular indentations in the seed coat. After this, 10 traits were measured as quantitative variables and were given the following abbreviations: PCI = fresh weight of 100 seed; LOH = seed width; LOV = seed height; ASE = seed area; PSE = seed perimeter; NSU = number of seed pits per side; ASU = area of seed pits; PSU = perimeter of seed pits; GTE = integument width; and AVE = angle of seed tips. This last trait measured the seed shape as characterized by the seed angle between the two vertices formed by the seed tips. Each quantitative trait was measured on 100 seed in three replicates and averages reported for 56 genotypes.

Statistical Analysis

The data were analyzed for descriptive statistics comparing the three groups of purple passion fruit genotypes. Quantitative data were tested for normality and significance in analyses of variance (ANOVA) along with positive or negative correlation values with R software. The quantitative trait values for each group were then compared with mean separation using Duncan tests in SAS version 9.4 (SAS Institute, Cary, NC, United States). The full set of traits was also used for Multidimensional scaling (MDS) tests done with average values to evaluate the diversity between characteristics, genotypes and germplasm groups, based on RWizard software[1]. The MDS defined each trait as a vector whose proximity to other vectors for other traits reflected associations between traits confirming correlations, and their relative importance for the phenotyping, individually or in conjunction. Variance inflation factors (VIF) were estimated to determine which traits explained the major portion of variability and subsequently a principal component analysis (PCA) was done with the same software to separate germplasm groups (cultivars, genebank accessions, and landraces) and to identify outlier genotypes. Note that seed was dried to determine seed water content comparing fresh to dry weight of 100 seed (PCIs) and expressing on a wet-weight basis moisture content in percent (%) with the formula: $[(W2–W3)/(W2–W1)] \times 100$; considering that W1 = weight of container with lid; W2 = weight of container with lid and sample before drying; and W3 = weight of container with lid and sample after drying.

RESULTS

Seed Viability and Germination Tests

Seed viability varied significantly ($p < 0.001$) based on the source of seeds tested and group of genotypes (**Table 1**), being highest on average for commercial cultivars (91%), followed by landraces (71%), and genebank accessions (65%). Similarly, germination rates presented highly significant differences ($p < 0.001$) between germplasm groups with 80%, 63%, and 59% average germination for commercial cultivars, landraces and genebank accessions, respectively (**Table 2**). Skewing, a form of non-normality in a population, was found for the distribution of genotypes' seed viability in the genebank group but not in the landraces or cultivars. For the germination rate, skewing was much less evident. Slight negative kurtosis was found for the distribution of values for both variables.

Morphological Seed Variability

Seeds of purple passion fruit were found to vary in seed shape from elliptical to semi-elliptical or oval; and in seed color from either brown to black (**Table 3**). Determination of elliptical or oval shape was based on measurements of the vertical and horizontal cross section widths and lengths, angle between the vertices at the seed tips and other morphological characteristics. Observable differences in seed lengths, seed widths and resulting seed sizes measured as area of each genotypes' seeds measured on one side, gave an overall picture of the morphological differences in passion fruit types. These variables were associated with each other and with the qualitative differences seen in the figure part of the **Table 2**. Larger elliptical black seed were the only kind of seed in commercial cultivars and genebank entries. All seed had some degree of seed pitting.

[1]http://www.ipez.es/RWizard

TABLE 1 | Sum of Squares from the Analyses of Variance (ANOVAs) for 10 seed characteristics evaluated on 56 genotypes of purple passion fruit (*Passiflora edulis* f. *edulis* Sims) in three germplasm groups made up of 8 commercial cultivars, 14 genebank accessions and 34 landraces.

Variable	Df	ASE	ASU	LOH	LOV	NSU	PCI	PSE	PSU	Germination	Viability
Within groups	55	0,00091^{ns}	0,0013***	0,078*	0,047*	76,455***	0,163*	0,029^{ns}	0,00037*	434,36*	166,48*
Between groups	2	0,0251*	0,011***	1,989*	1,282*	1876,8***	2,481***	0,668*	0,0071*	3187,18***	3862,06***

*Abbreviations: ^{ns}not significant, *significant (0.01 < P < 0.05), **high significance (0.001 < P < 0.01), and ***highly significant (P < 0.001), Df = degrees of freedom; Variables: ASE = seed area, ASU = area of seed pitting, LOH = seed length on the horizontal axis, LOV = seed width on the vertical access PCI = fresh dry weight of 100 seeds, NSU = number of seed pits, PSE = seed perimeter, and PSU = seed pit perimeters.*

TABLE 2 | Average seed viability and germination percentage based on three germplasm groups of purple passion fruit (*Passiflora edulis* f. *edulis* Sims) made up of 8 commercial cultivars, 14 genebank accessions and 34 landraces.

	Germplasm Group	N	Mean (%)	Variance	Std. Dev.	Std. Error	CV	Skewing	Kurtosis
Seed Viability	Cultivars	8	91.04	19.27	4.39	1.79	5.2	−0.05	−1.88
	Genebank entries	14	64.56	128.25	11.32	4.28	18.48	0.37	−1.78
	Landraces	34	71.29	247.49	15.73	2.7	21.82	−0.4	−1.32
Germination Percent	Cultivars	8	80.05	62.01	7.87	3.21	10.57	−0.5	−1.18
	Genebank entries	14	59.28	88.18	9.39	3.55	16.72	−0.19	−1.54
	Landraces	34	62.93	101.65	10.08	1.73	16.05	0.15	−1.32

P < 0.01, Based on n = 100 seed per accession.

TABLE 3 | Qualitative traits measured on the seed of 56 genotypes of purple passion fruit (*Passiflora edulis* f. *edulis* Sims).

Descriptor		Source			Figure
		Comm. Cultivars	Genebank Entries	Landraces	
Seed Shape	Elliptical	8	14	32	
	Semi-elliptical	0	0	2	
Seed Color	Black	8	14	30	
	Brown	0	0	4	

$n_{(cultivars)}$ = 8; $n_{(genebank)}$ = 14; and $n_{(landraces)}$ = 34. Photograph showing the angle between the central axis of the purple passion fruit seed and the seed tips at each vertices (AVE) for semi-elliptical (a) and elliptical seed (b). Seed pitting is observable on both elliptical and oval seeds as well as the difference between color variants.

Semi-elliptical black seeds were found in two landraces, BUN009 from Boyacá, and BUN037 from Huila (for geographic coordinates see **Supplementary Table 1**). Meanwhile elliptical brown seed was found for four genotypes: namely BUN002 from Cundinamarca, BUN032 from North Santander, BUN036 from Cauca, and BUN040 from Huila. All other genotypes were black seeded with elliptical seed, the only phenotype found among commercial cultivars and genebank accessions. Seed pits were common across colors.

A diagram was drawn based on typical seed structure from our histological analysis to show the seed coat and cell layers below (left part of **Figure 1**). The first layer consisted of an external wax coating that was continuous and highly resistant to cracking and fractures. This was subtended by an exotesta and mesotesta layer of sclereid cells, then by a layer of palisade cells presumed to have originated from the integument based on previous studies (Cárdenas-Hernández et al., 2011). The palisade cells were oriented perpendicular to the sclereid cells. Seed pits consisted of an embossed pattern of indentations found in the sclereid cell layer with fewer and more narrow cells subtended directly by the exotegmen tissue and narrow layer of hyaline cells. Seed pitting was found on both sides of the seed and was evaluated for pit number, perimeter and area. All together, the three layers making up the seed coat varied in thickness based on the pattern of pitted and non-pitted areas. Every layer presented differences in the form of the cells, color and thickness, between the sections of the basal and medium parts of the seeds, with differences in width notable between accessions. Overall, the seed coat measurements were 0.45 mm thick on average below the seed pit; while in non-pitted areas the thickness of the seed coat was 0.50 mm on average. This is thicker than what has been found in *P. ligularis*, the granadilla from Cárdenas-Hernández et al. (2011), upon which our studies were based.

Correlation Values

During the seed viability testing, the seed thickness was found to differ and influence seed germination percentage and time to completely germinate, which fluctuated between 1.5 and 3.5 months. The seed coat also varied in amount, location and depth of seed pitting which in turn affected the seed coat thickness at the seed pits and between them. This was observable upon inspection by visible microscopy.

Highly significant ($P < 0.001$) positive correlation values were found for the germination percentage with the area covered in seed pitting ($r = 0.854$) as well as number of seed pits ($r = 0.789$, $P < 0.001$; **Table 4**); but not between percentage germination and seed size or length. Seed germination was also correlated significantly and positively with perimeter of seed pits ($r = 0.432$, $P = 0.047$) as well as fresh seed weight ($r = 0.547$, $P = 0.023$).

The high correlation value showed that genotypes with the most seed pitting, germinated much more quickly than those without much pitting. Microscopy showed that seed pits were the areas of thinner seed coat thickness and might capture water or allow water entry and faster seed imbibing. The seed pitted area, number of seed pits and seed pit perimeter were all positively correlated with each other as were fresh and dry seed weights and seed perimeter but not with angle between seed vertices, a measure of seed shape.

As expected, angle between vertices as a quantitative trait was related to the shape of the seed as a qualitative trait. Classification of the passion fruit genotypes into two groups found semi-elliptical seed with seed vertices at an angle ∼140° and elliptical seed with seed vertices at an angle ∼110°. Therefore, visual observation of the trait of semi versus elliptical seed was just as good as measuring angle between seed vertices.

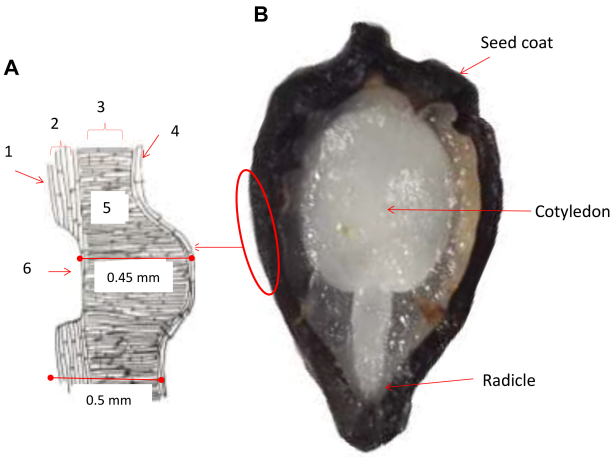


FIGURE 1 | Cross sections of a purple passion fruit (*Passiflora edulis* f. *edulis* Sims) seed coat **(A)** and the entire seed **(B)**. In the diagram of the seed coat to the left, a seed pit is shown. The seed pit is the depth of the seed coat cuticle. Approximate widths of the seed coat are 0.45 mm below the seed pit and 0.5 mm on either side of the seed pit. Numbers around the diagram indicating the following: 1 = exotesta; 2 = mesotesta; 3 = exotegmen; 4 = hyaline cells; 5 = palisade cells; and 6 = seed pit. The radicle and cotyledon of the seed embryo indicated in the seed cross section.

Principal Component Analysis Relationships

In the PCA we could distinguish the seed morphological characteristics most related to classifying the genotypes of purple passion fruit (**Figure 2**). Principal component 1 was explained by the angle of vertices and seed coat (tegument) thickness while principal component 2 was explained by the variables seed length seed width as well as fresh and dry seed weight.

These first two components and therefore these main variables, explained 85% of the variability found for seed phenotypes. In looking at the relationships between individual traits based on the proximity of their vectors we found relationships between the number of seed pits, the area covered by seed pits and percentage germination. These were the most important traits in our study in terms of variability.

Meanwhile, component traits of seed size were associated, such as seed length on vertical and horizontal axis as well as fresh and dry seed weight. Perimeter of the seeds and perimeter of the seed pits were associated with the primary group of traits as well and they were closely linked to seed viability.

Angle between vertices and seed coat thickness were outliers that were not associated with other traits showing that the genetic control for this trait is likely to be different than for the traits of number of seed pits and area covered by seed pitting or those genes controlling seed size. The vector for perimeter of seed pits was associated with those for vertical length and horizontal length as well as fresh and dry seed weight showing some association for seed size factors. These were also associated with area and perimeter of the seed. Epistatic interaction between traits are likely for the seed size traits as one group and the seed pitting traits as another group.

The most distinct genotypes in the PCA were BUN001, BUN002, BUN003, BUN004, BUN005, BUN006, and BUN007 in the left bottom quadrant; while BUN10, BUN12, BUN14, BUN41, BUN44, and BUN46 were distinct in the right bottom quadrant. Apart from these only BUN17, BUN22, and BUN35 were distinct toward the left upper quadrant.

Among the three germplasm groups, the commercial cultivars had significantly greater seed pitting, more seed pitted area, larger seed size, and associated longer seed perimeter traits than landraces or genebank accessions. Landraces, on the other hand, had the smallest values for all these characteristics, and genebank accessions were intermediate between landraces and commercial cultivars for the seed size and seed pitting traits. Meanwhile, no significant differences ($P = 0.8$) were found between groups for seed moisture content and fresh or dry seed weight (PCI and PCIs, respectively).

Variance Inflation and Multi-Dimensional Scaling

The VIF was determined for each of the quantitative morphological variables, in order to establish which ones explained the diversity among the accessions. The most significant variables including average angle of divergence, seed coat thickness, length of the seed on the horizontal axis and seed fresh weight, were then used in a MDS analysis (**Figure 3**).

The advantages of an MDS were that qualitative and quantitative traits could be analyzed together by comparing and contrasting their distributions across the population of genotypes, without any preconditions or assumptions made about their relationships based on the sampled population.

TABLE 4 | Correlation values (above diagonal) and corresponding significance (*p*-values, below diagonal) for seed traits measured on 56 genotypes of purple passion fruit (*Passiflora edulis* f. *edulis* Sims).

Traits	NSU	ASE	PSE	PSU	LOV	LOH	PCIs	PCI	ASU	AVE	GER	VIA
NSU	*****	0.093	0.310	0.302	−0.067	−0.003	0.350	0.476	0.562	−0.108	0.789	0.210
ASE	0.521	*****	0.527	−0.295	0.221	0.052	0.126	0.121	−0.091	−0.038	0.365	0.145
PSE	0.029	<0.001	*****	0.090	0.032	0.050	0.254	0.334	0.276	−0.151	0.341	0.151
PSU	0.033	0.038	0.532	*****	0.119	0.171	0.426	0.329	0.542	−0.026	0.432	0.254
LOV	0.642	0.123	0.825	0.410	*****	0.700	0.173	0.131	−0.133	−0.067	0.214	0.101
LOH	0.984	0.718	0.732	0.234	<0.001	*****	0.131	0.130	−0.116	−0.018	0.124	0.146
PCIs	0.013	0.384	0.075	0.002	0.231	0.364	*****	0.728	0.624	−0.159	0.423	0.231
PCI	<0.001	0.401	0.018	0.019	0.366	0.370	<0.001	*****	0.677	−0.198	0.547	0.114
ASU	<0.001	0.532	0.052	<0.001	0.359	0.424	<0.001	<0.001	*****	−0.293	0.854	0.168
AVE	0.457	0.792	0.296	0.858	0.643	0.900	0.271	0.169	0.039	*****	0.241	0.134
GER	<0.001	0.851	0.254	0.047	0.547	0.756	0.054	0.023	<0.001	0.125	*****	0.217
VIA	0.745	0.651	0.547	0.124	0.365	0.851	0.325	0.136	0.104	0.514	0.089	*****

$n_{(cultivars)}$ = 8; $n_{(genebank)}$ = 14; and $n_{(landraces)}$=34. Color intensity used to indicate significance values of P < 0.05, P < 0.01, and P < 0.001. Abbreviations: ASE = total seed area; ASU = area covered by seed pits; AVE = angle between vertices; GER = percentage germination; GTE = seed coat/tegument thickness; LOV = seed length on its vertical axis; LOH = seed length on its horizontal axis; NSU = number of seed pits: PCIs = dry seed weight; PCI = fresh seed weight; PER = seed perimeter; PSU = perimeter of seed pits; and VIA = seed viability percentage.

Geographical Distribution of Genotypes by Seed Size

The seed size descriptor was plotted by geographical coordinates of latitude and longitude for the collection site of each genotype on a map of Colombia (**Figure 4**). The main trend was that larger seeded genotypes with an area of 0.21 to 0.27 mm^2 were distributed only in the Central Andes Mountains while smaller seeded genotypes with an area of 0.1 to 0. 16 mm^2 were distributed throughout all other regions.

Comparing the geographical origins of the genotypes used in the study by regions shown on the map figures, we found that seeds were larger in the Central region of the Colombian Andes, probably due to the type of climate or soils in this region; and to the commercial cultivars originating in this region. Landraces from other parts of the country are more likely to be from ecologically less favorable regions outside the production zones and therefore, have unselected and smaller sized seed.

DISCUSSION

A major accomplishment of this study was to characterize seed of a large number of genotypes of purple passion fruit for characteristics important to establishment of seedlings and potentially domestication. Seed germination, dormancy, and viability plus subsequent seedling vigor are important limitations for many wild and semi-domesticated plant species (Baskin and Baskin, 1998). Among our observations about seed, we found that the purple passion fruit genotypes from our collection required at least 1.5 months to germinate but could take as long as 3.5 months or could remain dormant. Late and delayed germination was also found by Montaña et al. (2014), who observed that 50% of their seed germinated after 2 months. The rate gradually increases with longer time. Some seeds in our study only germinated after 4 months but were not used for evaluation. Previous studies of purple passion fruit also have observed long physiological dormancy periods prior to germination (Delanoy et al., 2006; Araújo et al., 2007; Barbosa et al., 2012; Veiga et al., 2013).

Another finding of our study was that the low germination values found in purple passion fruit were especially pronounced in landraces compared to commercial cultivars. Montaña et al. (2014) found similar slow germination for landraces with 50% of seeds germinating at 48 days after sowing. Current studies of passion flower seed germination are trying to determine the type of dormancy in *P. edulis* and its relatives. Some authors have suggested that passion fruit seeds are intermediate between rapid germination typical of fully domesticated annual crops and slow germination typical of many wild or semi-domesticate seeds (Silva et al., 2015). The intermediate nature of passion fruit seed is important for conservation efforts, however, more importantly new protocols for successful germination are needed (Meletti et al., 2007; Magalhães, 2010; Veiga et al., 2013; Moreno et al., 2015).

In support of a hypothesis of semi-recalcitrance, this study found that seed viability was slightly higher (cultivars 91%; landraces 71%, and genebank accessions 65%) than seed germination (cultivars 80%; landraces 63%, and genebank accessions 59%); suggesting that many seeds were alive but dormant. Barbosa et al. (2012) and Mira et al. (2015) reported percentage viability greater than 75% using conductivity tests, despite low germination rates of 40%. Germination capacity and seed viability are thought to vary depending on seed type among families of plants, or within species and genera (Baskin and Baskin, 1998). Ramírez et al. (2015) have suggested a role for mycorrhizae in germination of passion fruit seeds. Overall the purple passion fruit can be considered to have strong dormancy that is both physical and physiological in nature (Baskin and Baskin, 1998). The low germination values seen for some genebank accessions in our study (Average 59%) could be due to seed shipment and storage conditions but commercial and landrace types were fresh seed.

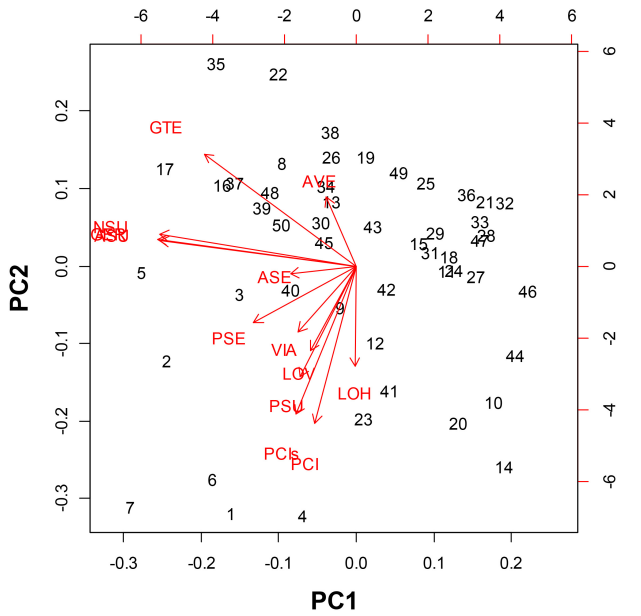


FIGURE 2 | Principal Component Analysis for seed morphology traits evaluated on 56 genotypes of purple passion fruit (*Passiflora edulis* f. *edulis* Sims.), with numbers representing cultivars, landraces and accession, and based on the following characteristics: AVE = angle between seed vertices, GTE = seed coat/gross tegument width, NSU = number of seed pits per side, ASU = area of seed pits per side, GER = percentage (%) germination, PSE = perimeter of the seed, ASE = area of the seed, VIA = percentage (%) viability, LOV = length of the seed on vertical axis, LOH = length of the seed on the horizontal axis, PSU = perimeter of the seed pits, PCI = one hundred (100) fresh seed weight; and PCIs = one hundred (100) dried seed weight.

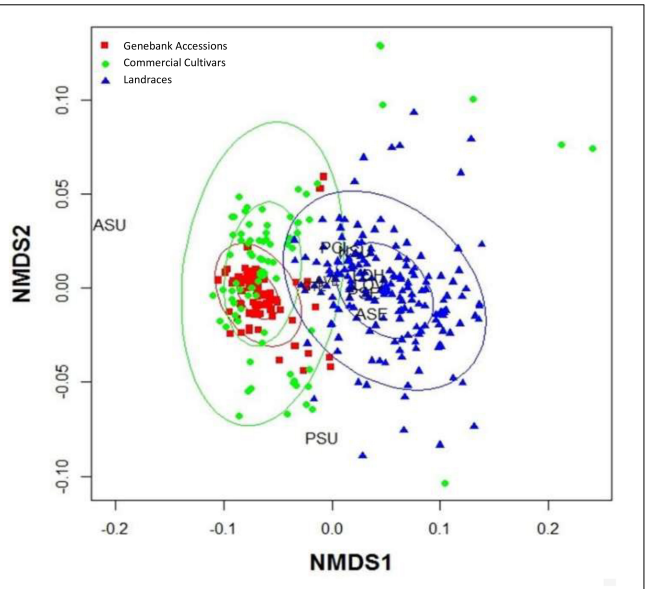


FIGURE 3 | Multidimensional Scaling (MDS) analysis for the seed traits contributing highest variance to the analysis of commercial cultivars, genebank accessions and landraces of purple passion fruit, *Passiflora edulis* Sims. f. *edulis*, namely ASU = area of seed pitting, AVE = angle between seed vertices, PCI = fresh dry weight of 100 seeds, GTE = seed coat thickness, NSU = number of seed pits, LOH = length of the seed on the horizontal axis, LOV = length of the seed on the vertical access, PSE = seed perimeter, ASE = seed area, and PSU = seed pit perimeters. Number of measurements: $n_{(cultivars)}$ = 80; $n_{(genebank)}$ = 280; and $n_{(landraces)}$ = 340.

Among the groups of genotypes used in this study, we observed that the viability and germination was higher in the commercial cultivars (91 and 80%, respectively) compared to the landrace (74% and 63%) or genebank accessions (65% and 59%). This difference could be explained by selection during the domestication process; whereby commercial cultivars have been selected for higher seed germination so as to be good genotypes for crop establishment, better than landraces. Other possibilities could relate to genetic diversity among the groups, since passion fruit flowers are known to be primarily inbreeding but up to 70% outcrossing (Bruckner et al., 1995), crosses between low and high germination genotypes are needed. Seed quality can also relate to plant physiological differences during fruit production (Carvalho and Nakagawa, 2012; Lima et al., 2017; Fischer et al., 2018). Genebank accessions were briefly stored prior to the tests so some difference with landraces could be due to seed age (Larré et al., 2007).

In other observations from this study, seed size was variable between the groups of genotypes. Seed size and weight compared to fruit weight is a determinant of juice yield (Macedo et al., 2015). Juicing success is also based on fruit size and weight or percentage of the fruit made up of the fruit wall or shell. Tradeoffs in aryl versus seed production determine plant nutritional traits of seeds since these are higher in proteins and starches requiring more nitrogen and remobilized carbon, respectively, while the pulp is higher in sugar content (Devi et al., 2018). Given the characteristics of many seed per fruit and sweet aryls surrounding each seed, evolutionary forces would tend to increase the number of seeds for the dispersal by birds but decrease the seeds for the germination from fallen and rotting fruits. Bird dispersal could also scarify seed through the digestive tract making seed coat thickness and dormancy important issues in the evolution of passion fruit species. In contrast, seedlings germinating from fallen fruit would rely on natural weathering for seed germination. The inter-relationship of all these factors on landraces, which often grow naturally in coffee plantations, would be expected to be higher than on commercial cultivars where farmers select seed harvested from fruits dried and processed in artificial settings, separate from natural processes. These evolutionary forces among others (Baskin and Baskin, 1998) could explain the variations in seed anatomy and morphology which were observed.

Passion fruit species are known to be in the process of diversification (Bruckner et al., 1995) and evolution (Cerqueira et al., 2014; Silva et al., 2017), aided by the high capacity for interspecific hybridization and karyotype plasticity (Melo et al., 2001; Melo and Guerra, 2003). Although purple passion fruit are thought to be mostly inbreeding some outcrossing (Bruckner et al., 1995; Brum et al., 2011; de Lira et al., 2016) may have result in morphological variability (Marostega et al., 2017: Ocampo and d'Eeckenbrugge, 2017). Our study confirms the importance of seed descriptors in evaluating genetic variability.

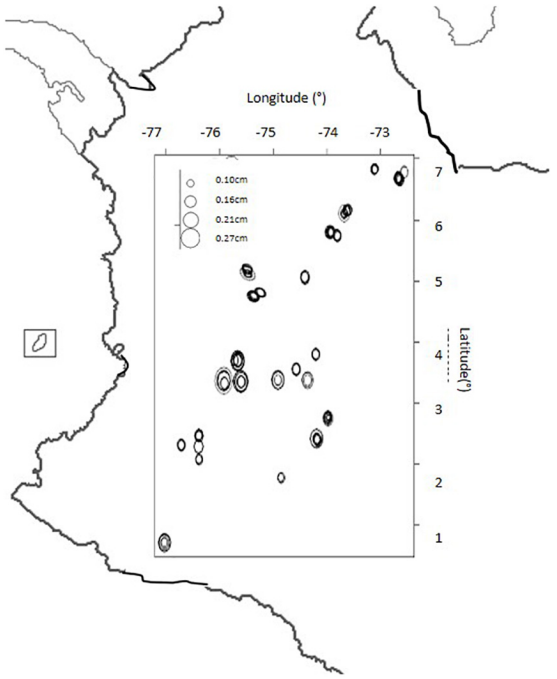


FIGURE 4 | Seed size (in cm) distribution for landrace genotypes of purple passion fruit, *Passiflora edulis* Sims. *F. edulis*, collected from Colombia based on geographical collection site with latitude and longitude numbers (in degrees) given to the top and right of the boxed area.

Seed pitting in addition to seed shape was found to be a critical seed descriptor in our study. In the case of the Passifloraceae family's evolutionary record, seed coats have varied from fovealate (notable seed pitting), to coarsely foveolate, reticulate-foveolate or transversely grooved (Martinez, 2017). Seed lengths varied greatly from very short seeds (1.5 mm) to longer seeds (14 mm); although seed shape tended to be ovoid, obovoid to elliptic. Seed pitting and seed size could have affected the seeds' relationship with the soil it is planted in and the water available for germination. Seed surface characteristics can be determining factors in water absorption capacity and resulting germination across many plant species (Koomneef et al., 2002). The greater the contact, the more water can be absorbed. Therefore, the structure of the seed coat is an important factor to measure. In the case of the purple passion fruit seed, the exposure area was amplified by level of seed pitting of the fovacous seed coats, increasing the contact surface for water and its retention and absorption. Therefore, another important result of our study was the identification of a direct relationship between germination and seed pitting irrespective of seed size, exhibiting the germination the highest correlation with number of pits (0,789) and area pits (0.854).

We observed that the germination was correlated with the three traits of number, perimeter and area of seed pitting, which indicates that if the seed has a large number of pits over a large area of the seed surface or that the seed pits are large, then the seed germination can be accelerated by several months. Although not part of our study, seed pits reduce the thickness of the seed coat in specific portions of the seed, likely favoring water uptake leading to germination. The variability in the expression of seed pitting among landraces and cultivars is probably related to selection and domestication processes for the species. The increase in fruit size may indirectly select for larger seed size and even if seed pitting remains the same in density the number of seed pits increases.

Water content could also play a role in germination capacity; however, in our study we found no significance in between group (p-value = 0.8) or within group (p-value = 0.27) variation for hydration based the weight of 100 fresh (PCI) compared to dried seeds (PCIs). Water in seed can be of four types (1) free water in between seed structure; (2) capillary water circulating in seed tissues; (3) cellular water in seed organs; and (4) molecular water bound to other metabolites, molecules or macromolecules of the seed (Craviotto et al., 2009). All of these in conjunction affect seed conservation strategies and seed quality traits of germination rate, viability and vigor. Water content of our seeds varied between 10 to 13% based on seed drying with silica gel. This value was similar to values reported for many crop species as well (soybean for example with previous reference) and in other types of passion fruits. Posada et al. (2012) found that the seed of granadilla, maracuya or yellow passion fruit and gulupa or purple passion fruit (*Passiflora ligularis*, *P. edulis* f. *flavicarpa* and *P. edulis* f. *edulis*, respectively) all had between10 al 12% moisture and that excess drying only affects germination when water content falls to 6%. In future studies we hope to see the effect of seed dessication and moisture content on seed germination and viability in genebanks but for now this was not a priority of our study but rather seed morphology and germination under normal moisture content levels of 10 to 12% were priorities.

In summary, the most important findings of our study were (1) passion fruit seed coats had a rigid structure that was associated with the long dormancy periods; (2) landraces differed from commercial cultivars and genebank accessions in germination time; and (3) there was a positively correlated relationship of germination capacity with the number of seed pits and the area of seed pitting. Seeds with a high number of seed pits over a larger area, had faster germination. Seed structure of purple passion fruit was similar to that of *P. ligularis* based on histological analysis (Cárdenas-Hernández et al., 2011). Three well differentiated layers were likely to be the internal exotegmen, the middle mesotesta, and the external exotestal. Each layer had a different tissue type with hyaline, sclereid and/or palisade cells. In closing, we can see from this study that seed morphology is relevant both today (Crochemore et al., 2003) and in the study of the fossil record of passion fruits (Martinez, 2017). Seed coat structure is critical for survival of the species and their descriptors are often used along with above ground foliar and floral characteristics to characterize genetic diversity. Viability, meanwhile, is important as it can be lost in storage especially in genebank conservation to over-drying of seed or power failure. *In vivo* conservation has its own set of advantages and disadvantages. It is therefore necessary to carry out further morphological and physiological evaluations of the seeds found for landraces and commercial varieties as a complement to the determinations of diversity in passion fruit genotypes held

by genebanks. All morphological and genotyping information can be used to find associations with productivity, yield and other morphoagronomic characteristics. These analyses are also useful in the plans for crop improvement or seed certification after varieties have been established in breeding programs.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

NC carried out research conceived by LM and MB. All authors wrote sections of the manuscript. MB edited the manuscript. NC and MB prepared figures and tables.

FUNDING

We are grateful to the Colombian funding agency, Colciencias; to the Universidad Nacional de Colombia (UNAL); and to Tennessee State University (TSU) for providing funding and support for this project. UNAL-TSU have a memorandum of understanding (2019–23) for student exchange. The grant numbers involved were Colciencias RC 0459-2013; UNAL research and extension division grant Hermes 37360; and the TSU Evans Allen fund TEN-X07.

ACKNOWLEDGMENTS

We are thankful to Dr. Gustavo Morales (Botanical Garden José Celestino Mutis- Bogotá) and Prof. Fredy Ramos (UNAL), for the collaboration in taxonomic determinations and support in field trips. Profs. Gustavo Ligarreto, Maria Isabel Chacon and Gerard Fischer (UNAL) are acknowledged for helpful advice.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.00498/full#supplementary-material

REFERENCES

Araújo, E., Daher, R., Silva, N., and Viana, A. (2007). Path analysis for physiological traits that influence seed germination of *Passiflora edulis* f. flavicarpa deg. *Crop Breed. Appl. Biotechnol.* 7, 148–154. doi: 10.12702/1984-7033.v07n02a06

Barbosa, R., Leão, E., Caprio, C., and Vieira, R. (2012). Teste de conductividade elétrica em sementes de maracujazeiro-amarelo. *Rev. Brasil. Fruticul.* 34, 646–651. doi: 10.1590/s0100-29452012000200043

Baskin, C. C., and Baskin, J. M. (1998). *Seeds: Ecology, Biogeography and Evaluation of Dormancy and Germination*. San Diego, CA: Academic Press.

Black, M. (1972). *Control Processes in Germination and Dormancy*. Oxford: Oxford University Press.

Bruckner, C. H., Casali, V., Moraes, C., Regazzi, A., and Silva, E. (1995). Self-incompatibility in passion fruit (*Passiflora edulis* Sims). *Acta Horticul.* 370, 45–57.

Brum, B., Lopes, S., Storck, L., Lúcio, A., Oliveira, P., and Milani, M. (2011). Correlações canônicas entre variáveis de semente, plântula, planta e produção de grãos em mamoneira. *Ciência Rural* 41, 404–411. doi: 10.1590/s0103-84782011000300007

Cárdenas-Hernández, J., Miranda, D. L., Magnitsky, S., and Carranza, C. (2011). Morphological and anatomical analyses of the seed coats of sweet granadilla (*Passiflora ligularis* Juss.) *seeds*. *Agron. Colombiana* 29, 377–385.

Carvalho, N., and Nakagawa, J. (2012). *Sementes: Ciência, Tecnologia e Produção*, 5.ed Edn. Jaboticabal: Funep, 590.

Cerqueira, C., Onildo, J., Santos, E., Correa, R., and Souza, A. (2014). Genetic breeding and perspectives in molecular and genetics studies. *Int. J. Mol. Sci.* 15, 14122–14152. doi: 10.3390/ijms150814122

Chernoff, M., Plitmann, U., and Kislev, M. E. (1992). Seed characters and testa texture in species of Vicieae: their taxonomic significance. *Israel J. Bot.* 41, 167–186.

Craviotto, R. M., Salinas, A., Arango, M., and Ferrari, S. (2009). Influencia de propiedades físicas sobre la conductividad eléctrica de semillas individuales de soja *Glycine max* (L.) *Merrill*. *Rev. Científica Agropecuaria* 13, 33–43.

Crochemore, M. L., Molinari, H. B., and Stenzel, M. N. (2003). Agro-morphological characterization of passion fruit (*Passiflora* spp.) *germplasm*. *Rev. Brasil. Fruticul.* 25, 5–10.

de Lira, J., Horst, C., and Souza, P. (2016). Method for overcoming the passion fruit self-incompatibility. *Rev. Brasil. Fruticul.* 38:e477.

de Oliveira, M., São José, A., Hojo, N., Magalhães, O., and Novaes, F. (2010). Superação de dormência de maracujá-do-mato (*Passiflora cincinnata* Mast.). *Rev. Brasil. Fruticul.* 32, 584–590. doi: 10.1590/s0100-29452010000500045

Delanoy, M., Van Damme, P., Scheldeman, X., and Beltran, J. (2006). Germination of *Passiflora mollissima* (Kunth) L.H.Bailey, *Passiflora tricuspis* Mast. and *Passiflora* nov sp. seeds. *Sci. Horticult.* 110, 198–203. doi: 10.1016/j.scienta.2006.07.007

Devi, S. R., Sidik, J., and Zakaria, M. (2018). Nutritive values of passion fruit (*Passiflora* Species) seeds and its role in human health. *J. Agric. Food Dev.* 4, 23–30. doi: 10.30635/2415-0142.2018.04.4

Fischer, G., Casierra, F., and Piedrahita, W. (2009). "Ecofisiología de las especies de pasifloráceas cultivadas en Colombia," in *Cultivo, poscosecha y comercialización de las pasifloras en Colombia: Maracuyá, granadilla, gulupa y curuba*, ed. G. Fischer (Bogotá: Sociedad Colombiana de Ciencias Hortícolas), 350.

Fischer, G., Melgarejo, L. M., and Cutler, J. (2018). Pre-harvest factors that influence the quality of passion fruit: a review. *Agron. Colombiana* 36, 217–226. doi: 10.15446/agron.colomb.v36n3.71751

Franco, G., Cartagena, J., and Correa, G. (2014). Análisis de crecimiento del fruto de gulupa (*Passiflora edulis* Sims), en las condiciones ecológicas del bosque húmedo montano bajo de Colombia. *Rev. Acta Divul. Cientifica* 17, 391–400. doi: 10.31910/rudca.v17.n2.2014.242

Gabr, D. G. (2014). Seed morphology and seed coat anatomy of some species of Apocynaceae and Asclepiadaceae. *Ann. Agric. Sci.* 59, 229–238. doi: 10.1016/j.aoas.2014.11.010

Garcia, A. (2010). "Pollination mechanisms in *Passiflora* species: the common and the rare flowers-ecological aspects and implications for horticulture," in *Pollination: Mechanisms, Ecology and Agricultural Advances*, eds N. D. Raskin and P. T. Vuturro (Hauppauge, NY: Nova Science Publishers).

Gelape, F., Vilela, T., Góes, T., Nunes, O., Miranda, D., and Campos, W. (2019). Advances in passion fruit (*Passiflora* spp.) propagation. *Rev. Brasil. Fruticult.* 41, 1–17.

Koomneef, M., Bentsink, L., and Hilhorst, H. (2002). Seed dormancy and germination. *Curr. Opin. Plant Biol.* 5, 33–36.

Larré, C., Zepka, A., and Moraes, D. (2007). Testes de germinação e emergência em sementes de maracujá submetidas a envelhecimento acelerado. *Rev. Brasil. Biociências* 5, 708–710.

Lima, R., Pio, A., Duarte, E., Azevedo, E., Higino, F., and Santos, C. (2017). Contribution of production and seed variables to the genetic divergence in passion fruit under different nutrient availabilities. *Pesquisa Agropecuaria Brasil.* 52, 607–614. doi: 10.1590/s0100-204x2017000800006

Liu, S., Li, A. D., Chen, C., Cai, G., Zhang, L., Guo, C., et al. (2017). *De novo* transcriptome sequencing in *Passiflora edulis* Sims to identify genes and signaling pathways involved in cold tolerance. *Forests* 8:435. doi: 10.3390/f8110435

Macedo, S., Camargos, R., Ferraz, R., Sá, A., and Machado, C. (2015). Aryl removal methods and passion fruit seed: germination and emergence. *J. Seed Sci.* 2, 125–130. doi: 10.1590/2317-1545v37n2146373

Magalhães, A. (2010). *Caracterização de frutos e sementes e germinação de Passiflora edulis Sims f. flavicarpa Degener e Passiflora cincinnata* Mast. Dissertação de Mestrado. Feira de Santana, Bahia.: Universidade Estadual de Feira de Santana.

Marostega, T., Martins, L., Baptista, P., Grillo, L., and Aparecido, M. (2017). Genetic diversity of *Passiflora* accessions based on morphophysiological seed descriptors. *Rev. Brasil. Fruticul.* 39, 365–373.

Martinez, C. (2017). Passifloraceae seeds from the late Eocene of Colombia. *Am. J. Bot.* 104, 1857–1866. doi: 10.3732/ajb.1700224

Matheri, F., Nyamai, D., Ngugi, M. P., Runo, S., Njuguna, J. K., Mwangi, M., et al. (2016). Phenotyping of selected Kenyan *Passiflora edulis* varieties and their hybrids based on quantitative morpho-agronomic traits. *J. Horticult.* 3, 181.

Meletti, L., Barbosa, W., Veiga, R., and Pio, R. (2007). Crioconservação de sementes de seis acessos de maracujazeiro. *Rev. Sci. Agraria Parana* 6, 13–20.

Melo, N. F., Cervi, A. C., and Guerra, M. (2001). Karyology and cytotaxonomy of the genus *Passiflora* L. *(Passifloraceae). Plant Systemat. Evol.* 226, 69–84. doi: 10.1007/s006060170074

Melo, N. F., and Guerra, M. (2003). Variability of the 5S and rDNA sites in *Passiflora* L. with species with distinct base chromosome numbers. *Ann. Bot.* 92, 309–316. doi: 10.1093/aob/mcg138

Mira, S., Veiga, L., Gonzalez, M., and Perez, F. (2015). Conductivity test in seeds of different passionflower species. *Pesquisa Agropecuaria Brasil.* 50, 510–513. doi: 10.1590/s0100-204x2015000600010

Montaña, L., Fischer, G., Magnitskiy, S., and Zuluaga, G. (2014). Effect of NaCl salinity on seed germination and seedling emergence of purple passion fruit (*Passiflora edulis* Sims). *Agron. Colombiana* 32, 188–195. doi: 10.15446/agron.colomb.v32n2.38287

Moreno, C., Gelape, F., Nunes, O., Lisboa, E., Pereira, A., and de Souza, F. (2015). "The genetic diversity, conservation, and use of passion fruit (*Passiflora* spp.)," in *Genetic Diversity and Erosion in Plants*, eds M. R. Ahuja and S. M. Jain (Berlin: Springer International Publishing), 215–231. doi: 10.1007/978-3-319-25954-3_5

Morton, J. F. (1987). "Passionfruit," in *Fruits of Warm Climates* ed. J. F. Morton (W. Lafayette, IN: Purdue University), 320–328.

Ocampo, J., and d'Eeckenbrugge, G. (2017). Morphological characterization in the genus *Passiflora* L.*:* an approach to understanding its complex variability. *Plant Systemat. Evol.* 303, 531–558. doi: 10.1007/s00606-017-1390-2

Pérez-Cortéz, S., Tillett, S., and Escala, M. (2002). Estudio morfológico de la semilla de 51 especies del género *Passiflora*. *Acta Bot. Venezolana* 25, 67–96.

Posada, P., Ocampo, J., and Santos, L. (2012). Estudios del comportamiento fisiológico de la semilla de granadilla (*Passiflora ligularis* Juss.) para la conservación ex situ. *Acta Agron.* 61, 39–40.

Ramírez, J. G., Muñoz, M., Osorno, L., Osorio, N., and Morales, J. (2015). Germination and growth of purple passion fruit seedlings under pre-germination treatments and mycorrhizal inoculation. *Pesquisa Agropecuaria Trop.* 45, 257–265. doi: 10.1590/1983-40632015v453 3273

Rodríguez, N. C., Demissie, D. A., Melgarejo, L. M., and Blair, M. W. (2020). Morphological and agronomic variability among cultivars, Landraces, and genebank accessions of purple passion fruit, *Passiflora edulis* f. edulis. *HortScience* (in press).

Rodríguez, N. C., Melgarejo, L. M., and Blair, M. W. (2019). Purple passion fruit, *Passiflora edulis* Sims f. edulis, variability for photosynthetic and physiological adaptation in contrasting environments. *Agronomy* 9:231. doi: 10.3390/agronomy9050231

Salomão, L. C. C., Pereira, W. E., Duarte, R. C. C., and Siqueira, D. L. D. (2002). Propagação por estaquia dos maracujazeiros doce (*Passiflora alata* Dryand.) e amarelo (*P. edulis* f. flavicarpa Deg.). *Rev. Brasil. Fruticul.* 24, 163–167. doi: 10.1590/s0100-29452002000100036

Santos, E. A., Viana, A. P., Freitas, J. C. O., Souza, M. M., Paiva, C. L., Rodrigues, D. L., et al. (2014). Phenotyping of *Passiflora edulis*, *P. setacea*, and their hybrids by a multivariate approach. *Genet Mol. Res.* 13, 9828–9845. doi: 10.4238/2014.november.27.10

Sawma, J. T., and Mohler, C. L. (2002). Evaluating seed viability by an unimbibed seed crush test in comparison with the tetrazolium test. *Weed Technol.* 16, 781–786. doi: 10.1614/0890-037x(2002)016[0781:esvbau]2.0.co;2

Silva, R., Silva, L., Gava, P., Magalhães, C., Fernandes, D., and Cruz, C. (2017). Genetic divergence between passion fruit hybrids and reciprocals based on seedling emergence and vigor. *J. Seed Sci.* 39, 417–425. doi: 10.1590/2317-1545v39n4183293

Silva, S., Camargos, R., Ferrraz, R., de Sá, A., and Machado, C. (2015). Aryl removal methods and passion fruit seed positions: germination and emergence. *J. Seed Sci.* 37, 125–130. doi: 10.1590/2317-1545v37n2146373

Smýkal, P., Vernoud, V., Blair, M. W., and Thompson, R. D. (2014). The role of the testa during development and in establishment of dormancy of the legume seed. *Front. Plant Sci.* 5:351. doi: 10.3389/fpls.2014.00351

Suárez, D., and Melgarejo, L. M. (2010). "Biología y germinación de semillas," in *Experimentos en Fisiología Vegetal*, ed. L. M. Melgarejo (Bogotá: Departamento de Biología, Facultad de Ciencias, Universidad Nacional de Colombia), 13–24.

Tangarife, M., Caetano, C., and Posada, C. (2009). Caracterización morfológica de especies del genero *Passiflora* de Colombia. *Acta Agron.* 58, 117–125.

Thokchom, R., and Mandal, G. (2017). Production preference and importance of passion fruit (*Passiflora edulis*): a review. *J. Agric Eng. Food Technol.* 4, 27–30.

Torres, M. (2018). Seed dormancy and germination of two cultivated species of Passifloraceae. *Boletín Científico Centro del Museo de Historia Natural* 22, 15–27.

Veiga, L., Mira, S., Gonzalez, M., Souza, M., Meletti, L., and Perez, F. (2013). Seed germination, desiccation tolerance and cryopreservation of *Passiflora species*. *Seed Sci. Technol.* 41, 97.

Conflict of Interest: 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 © 2020 Castillo, Melgarejo and Blair. 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) and the copyright owner(s) 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 Plant Science

ORIGINAL RESEARCH
published: 11 June 2020
doi: 10.3389/fpls.2020.00878

Weedy Rice as a Novel Gene Resource: A Genome-Wide Association Study of Anthocyanin Biosynthesis and an Evaluation of Nutritional Quality

Wenjia Wang†, Minghui Zhao†, Guangchen Zhang, Zimeng Liu, Yuchen Hua, Xingtian Jia, Jiayu Song, Dianrong Ma and Jian Sun**

Rice Research Institute, Shenyang Agricultural University, Shenyang, China

OPEN ACCESS

Edited by:
Thomas M. Davis, University of New Hampshire, United States

Reviewed by:
Dongying Gao, University of Georgia, United States
Kanako Bessho-Uehara, Carnegie Institution for Science (CIS), United States

****Correspondence:***
Dianrong Ma
madianrong@syau.edu.cn
Jian Sun
sunjian811119@syau.edu.cn

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Plant Breeding, a section of the journal Frontiers in Plant Science

Received: *31 January 2020*
Accepted: *28 May 2020*
Published: *11 June 2020*

Citation:
Wang W, Zhao M, Zhang G, Liu Z, Hua Y, Jia X, Song J, Ma D and Sun J (2020) Weedy Rice as a Novel Gene Resource: A Genome-Wide Association Study of Anthocyanin Biosynthesis and an Evaluation of Nutritional Quality. Front. Plant Sci. 11:878. doi: 10.3389/fpls.2020.00878

The pericarp color of rice grains is an important agronomic trait affected by domestication, and the color pigment, anthocyanin, is one of the key determinants of rice nutritional quality. Weedy rice, also called red rice because its pericarp is often red, may be a novel gene resource for the development of new rice. However, the genetic basis and nutritional quality of anthocyanin are poorly known. In this study, we used a genome-wide association study (GWAS) to find novel and specific QTLs of red pericarp in weedy rice. The known key gene site of red pericarp *Rc* was detected as the common genetic basis of both weedy and cultivated rice, and another 13 associated signals of pericarp color that were identified may contribute specifically to weedy rice pericarp color. We then nominated three pericarp color genes that may contribute to weedy rice divergence from cultivated rice based on selection sweep analysis. After clarifying the distribution and growth dynamics of pigment in weedy rice caryopsis, we compared its nutritional quality with cultivated rice. We found that sampled weedy rice pericarps had much greater quantities of anthocyanin, beneficial trace elements, free amino acids, and unsaturated fatty acids than the cultivated rice. In conclusion, the gene resources and novel genetic systems of rice anthocyanin biosynthesis explored in this study are of great value for the development of nutritious, high anthocyanin content rice.

Keywords: weedy rice, genome-wide association study, anthocyanin biosynthesis, nutritional quality, pericarp color

INTRODUCTION

Weedy rice (*Oryza sativa* f. *spontanea*) refers to rice plants that grow in rice fields or in surrounding fields as a weed. Weedy rice is spread throughout the rice paddy system worldwide, with temperate and subtropical areas being the hardest hit (Brespatry et al., 2001; Tai, 2002; Vaughan et al., 2017). Currently, research on the origin and evolution of weedy rice has made some progress, however,

Abbreviations: C3G, cyanidin-3-glucoside; GMO, genetically-modified organism; GWAS, genome-wide association study; KNN, k-nearest neighbor algorithm; P3G, paeoniflorin-3-glucoside; QTL, quantitative trait locus; WR, weedy rice; *WRAH*, weedy rice from Asian high latitudes; *WRSC*, weedy rice from south china.

there is no definite conclusion about the origin of weedy rice (Li L. F. et al., 2017; Qiu et al., 2017; Sun et al., 2019).

In general, the morphological characteristics of weedy rice fall between the wild rice species (*Oryza rufipogon*) and cultivated rice (*Oryza sativa* L.) (Ma et al., 2008). Weedy rice has distinctive biological characteristics such as a short growth period, black hull, strong granulation, long dormancy period, long awning, and red pericarp (Chen et al., 2004; Pipatpongpinyo et al., 2019). Because most weedy rice has a colored hull and red pericarp, it is often called red rice (Ma et al., 2008). The layers of weedy rice caryopsis, from the outside to the inside, are pericarp, seed coat, nucellus, aleurone layer, and endosperm (Sellappan et al., 2009; Juliano and Tuaño, 2019). The colored pericarp is an important feature distinguishing it from ordinary cultivated rice (Cui et al., 2016).

Brown rice refers to the caryopsis after the rice husk is removed, and the pericarp, seed coat, and nucellus are intact. Milled rice refers to rice grains with pericarps and seed coats removed. Colored rice such as red rice, gold rice, and black rice, refers to brown rice with colored pericarp (Kim et al., 2008), and the caryopsis color is mainly due to the accumulation of anthocyanins in the pericarp and seed coat (Liu et al., 2011). Colored rice contains high levels of anthocyanins and proanthocyanidins, which have strong antioxidant properties (Lei et al., 2006). Some studies have shown that rice color is related to its nutritional quality. For example, brown rice is not only rich in protein, amino acids, vitamins, vegetable fats, and trace elements such as Ca, Fe, Zn, and Se, it is also rich in biologically active substances such as flavonoids that have anti-oxidation, anti-tumor, free radical-scavenging, and hypoglycemic effects (Gu, 1992; Meng et al., 2005; Guo et al., 2011; Yue-Ting et al., 2016).

With recent developments in genetic analysis, some molecular mechanisms regulating rice pigmentation are well known. For instance the red color of pericarp is affected by the interaction of two genes, *Rc* and *Rd*, and the purple color of pericarp is controlled by the joint action of genes *Pb* and *Pp* (Furukawa et al., 2007; Rahman et al., 2013). Furukawa et al. (2007) found that the 14 bp deletion of the coding region of the *Rc* gene caused the loss of *Rc* gene function, resulting in a white pericarp. Two SNP mutations in the coding region of the *Rd* gene caused early termination of translation, resulting in the loss of *Rd* function (Furukawa et al., 2007; Sweeney et al., 2007). Studies have shown that the pericarp color of the weedy rice in southern China is mainly regulated by the *Rc* gene. The same results were found in American weedy rice (Yang et al., 2009; Gross et al., 2010). As a companion weed of cultivated rice, weedy rice infests rice fields worldwide and is believed to have multiple evolutionary origins from distinct ancestors. Thus whether other genes besides the four known ones regulate its anthocyanin content is still unknown.

The pericarp color is an important agronomic trait affected by domestication, and the color pigment, anthocyanin, is one of the important determinants of rice nutritional quality (Wang et al., 2018). Weedy rice may be a novel gene resource for the development of new rice (Sun et al., 2013). However, the genetic basis and nutritional quality of anthocyanin are poorly known in weedy rice. Therefore, we used genome-wide association analysis to explore the main sites regulating the color of weedy rice pericarp. A follow-up study was carried out on the locations and timing of anthocyanin deposition in weedy rice with different pericarp colors. Also, nutritional quality was measured and compared between cultivated and weedy rice.

MATERIALS AND METHODS

Materials

We collected 297 rice samples of six subgroups, including 46 weedy rice from Asian high latitudes (*WRAH*) and 14 middle-latitudes (*WRSC*), 69 temperate *japonica*, 12 tropical *japonica*, 145 *indica* and 11 *Aus*. The samples were divided into two sets for genome-wide association analysis (GWAS): set 1 for the red pericarp from weedy rice samples and set 2 the genetic background for the red pericarp from cultivar samples. All rice samples were maintained and cultivated in the germplasm resources field of Shenyang Agricultural University, China.

In order to further understand the production and distribution of weedy rice pericarp pigments, the typical red pericarp weedy rice WR07-14, WR07-47, WR03-32, WR07-141, WR07-142 and WR03-29 and white pericarp cultivated rice Shennong265, Akihikari, Nipponbare, Qishanzhan were selected from a GWAS panel for the observation of pigment components and their deposition process and for a nutrient quality analysis.

Methods

Genome-Wide Association Study

For GWAS, the phenotype value of pericarp color was defined as one of four levels: 0 for white, 1 for orange, 2 for red, 3 for dark red. Short Oligonucleotide Alignment Program (SOAP) software (Li et al., 2008) was used to aligning the reads of each sample with the reference genome IRGSP v1.0 (Li et al., 2008). The reads that were successfully and uniquely aligned with the reference genome at both ends were considered reliable and used to define SLAF tags. We called population SNPs based on the corresponding genome-wide SLAF tags. A total of 122,777 unimputed SNPs with a minor allele frequency > 0.05 and integrity > 0.5 for evolutionary study have been reported in our previous study (Sun et al., 2019).

In the present studies, we performed imputation to fill the missing genotype for the 122,777 unimputed SNPs by using the k-nearest neighbor algorithm (KNN) model in two sets of GWAS populations (Roberts et al., 2007; Huang et al., 2010). Consequently, the quality of this 122,777 SNPs was improved with integrity > 0.9. Genome-wide association analyses were conducted using a compressed MLM model that could effectively reduce the false positives. The equation of the compressed MLM model is $y = X\alpha + P\beta + K\mu + e$, in which y is phenotype, X is genotype, P is population structure matrix (Q matrix), and K is the kinship matrix. The P matrix was built from the top five principal components for population structure correction. The K matrix was built from the matrix of simple matching coefficients. The analyses were performed using TASSEL 5 software (Bradbury et al., 2007). The threshold of significant P-value was determined based on Bonferroni correction method. In order to increase the possibility for overlapping the selection signal, we lowered

the threshold of *P*-value by an order of magnitude from the Bonferroni correction threshold. Both thresholds of the significant *P*-value were used in the present study.

Selection Sweep Analysis and Candidate Gene Nomination

In this study, red pericarp rice (including landrace and weedy rice) were considered as the unimproved rice population, and modern cultivars with white pericarp were considered as the improved population. Nucleotide diversity (π) and Tajima's *D* were obtained for 500 kb sliding windows between unimproved and improved populations by using vcftools 1.0.3 software (Danecek et al., 2011). Then the selection sweeps were defined by π ratio ($\pi_{\text{unimproved population}}/\pi_{\text{improved population}}$) and Tajima's *D* in per 500 kb genomic slide windows and both selection signals were represented by heat maps. The cut-off of the π ratio for defining the strong selection genomic regions was set at 2.5, and the selected signal defined by negative value of Tajima's *D* was as a secondary reference. We then nominated the candidate genes according to key motifs that related to anthocyanin synthesis from the Rice Genome Annotation Project[1]. We also searched nutrient-related genes around the genomic regions of GWAS peaks.

Developmental Dynamics of Pericarp Color and the Determination of Pigment Distribution

The weedy rice that we used to observe the development of pericarp color were sampled every 2 days after flowering. The caryopsis to be used for hand-sliced sections were sampled every week after flowering. A digital SLR camera (Canon 550D) was used to record pericarp color and pigment distribution (Liu et al., 2011).

We further sampled the pericarp, seed coat, and endosperm of weedy rice 1, 2, and 3 weeks after flowering respectively in order to study the distributions of their pigments. Referring to the method of Yu et al. (2010) we dissected these parts by paraffin section. The sample was dehulled and fixed with FAA fixative, and then dehydration, transparency, waxing, embedding, slicing, dipping, dewaxing and rehydration, dyeing, and sealing of the sample were carried out in sequence (Dai et al., 2009; Yu et al., 2010). The histological structure of the weedy rice caryopsis was recorded using a stereo microscope (Zeiss Lumar. V12).

Determination of Pigment Composition and Content

We collected seeds of each sample (WR07-14, WR07-32, WR07-47 and WR07-141) at 5, 10, 15, 20, and 25 days after flowering, and we threshed, shelled, and ground the rice into flour. The total content of anthocyanins was determined by the pH differential method (Li et al., 2014). The components of the pigment in the mature weedy rice were identified through High Performance Liquid Chromatography-Mass Spectrometer (HPLC-MS). The column temperature was set to 35°C. The binary mobile phase consists of a formic acid-ammonium formate solution. The mobile phase gradient decreased from 99% to 60%. The flow rate was 0.3 mL/min (Li R. et al., 2017).

[1]http://rice.plantbiology.msu.edu/

Determination of Functional Nutrient Quality

Metal element concentrations in the pericarp of weedy rice and control samples were measured by using inorganic mass spectrometry (Gross, 2017). The amino acid concentrations were determined by chromatography using the Hitachi L8800 automatic amino acid analyzer (Hua et al., 2016). The fat was separated by soxhlet extraction for methyl esterification and then analyzed by using Gas chromatography–mass spectrometry (Agilent Technologies 7890A GC System and Agilent Technologies 240 Ion Trap). Resistant starch concentration was measured using the Shanghai Rongsheng Biotechnology Elisa kit.

Statistical Analyses

Statistical analyses for the comparison of phenotypic values ANOVA were carried out with the statistical software IBM SPSS 2.00 (IBM Crop, Armonk, NY, United States). The level of significance taken as $P < 0.05$.

RESULTS

Genome Wide Association Study of Weedy Rice Pericarp Color

The red pericarp phenotype was observed in all subgroups (**Supplementary Figure S1**). All weedy rice has colored pericarps, and some of these are extremely dark. *Aus*, *indica,* and tropical *japonica* have red pericarps and white pericarps. Most temperate *japonica* have white pericarps, and a few have red pericarps (**Figure 1**).

WRAH is a branch of temperate *japonica,* and *WRSC* is grouped with *indica* as reported in our previous research (Sun et al., 2019). In addition, the phenotypic variation of pericarp color was observed in all subgroups and did not correspond with subgroup differentiation. For these reasons, the false positives of GWAS due to population differentiation may have little impact, and we conducted GWAS across the six subgroups. In the compressed MLM of set 1 (weedy rice), we identified 13 association signals of pericarp color ($P < 1 \times 10^{-7}$), and in set 2 (cultivated rice), we identified four signals with clear peaks ($P < 1 \times 10^{-7}$) (**Figure 2**). The results showed that the genetic basis of pericarp color had both commonalities and differences between weedy rice and cultivated rice. The genetic basis of the pericarp color of weedy rice was more complex than that of the cultivated rice, which is reflected in the greater number of significant genomic association peak signals (**Figure 2**). Finally, we nominated five candidate genes according to the annotation information based on the Rice Genome Annotation Project (**Table 1**). The known key gene site of red pericarp *Rc* (Furukawa et al., 2007) was detected in both GWAS panels with significant peak signals. We also found that the association signals of red-pericarp synergistic-gene *Rd* could be detected in the weedy rice GWAS panel. The association signals of purple rice genes *Pp* and *Pb* could not be detected in either weedy rice or cultivated rice GWAS panels.

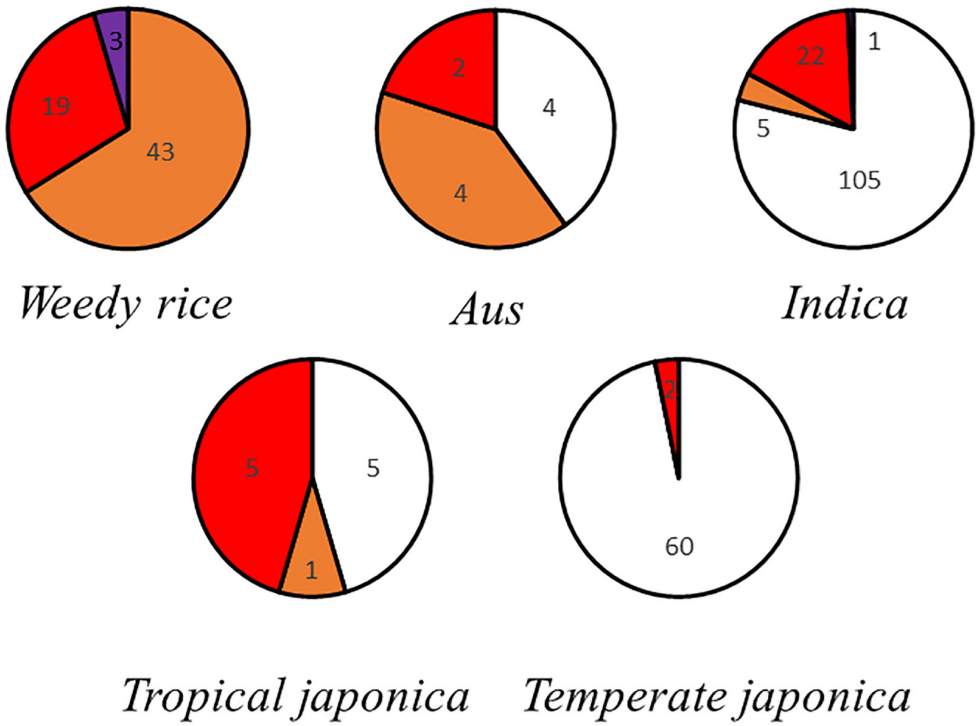


FIGURE 1 | Pericarp color grading statistics. Each circle represents a subcategory of rice, and the area occupied by each color represents the proportion of rice seeds of each pericarp color.

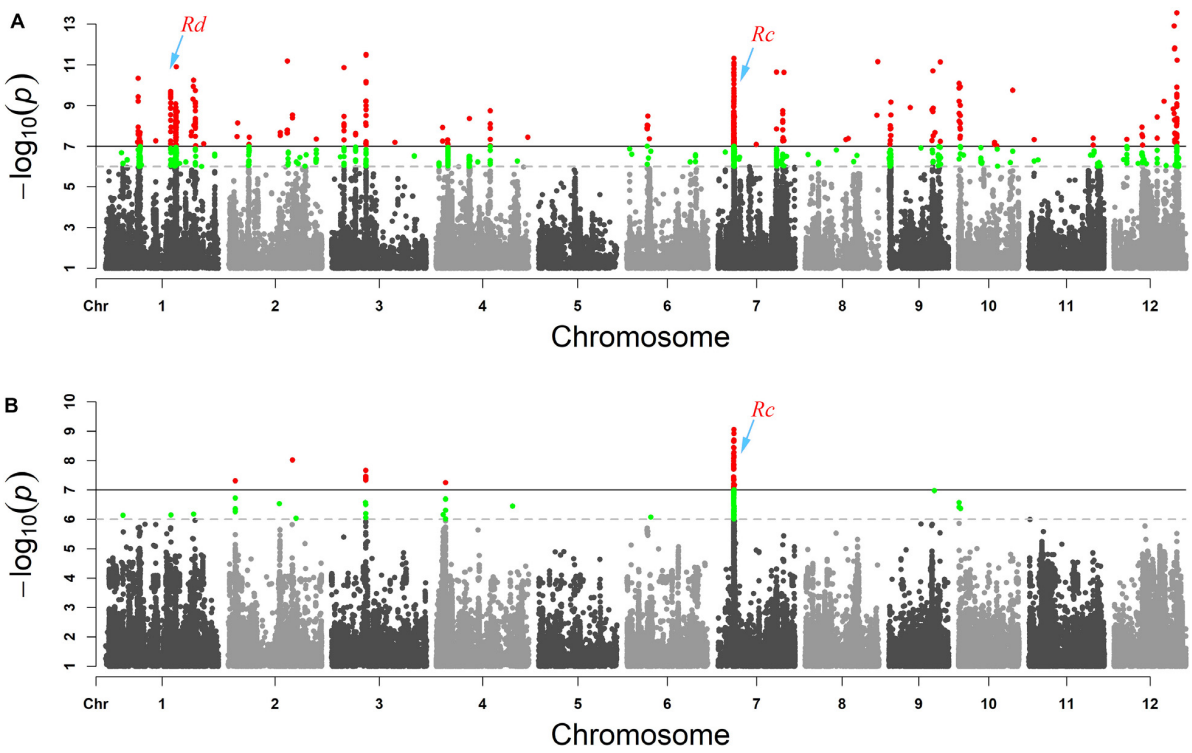


FIGURE 2 | GWAS results presented as Manhattan plot for **(A)** red pericarp weedy rice and (set 1) **(B)** red pericarp cultivated rice (set 2). The red area in the Manhattan map is the response site associated with red pericarp. The known pericarp color genes *Rd* and *Rc* are labeled.

TABLE 1 | GWAS peak selection sweep and candidate genes.

Peak *P* value position	*P* value	Genomic interval of selection sweep	π unimproved/π improved	Tajima's *D* value		Selection intensity	Pericarp color Candidate gene	Gene description
				Unimproved population	Improved population			
Chr1:12490628	4.69E−11	12316796-12821546	2.12	3.21	−0.26	Weak		
Chr1:24954656	1.3E−09	24900449-25222998	1.87	3.11	0.09	Weak	LOC_Os01g43700.1(*Rd*)	Cytochrome P450 72A1, putative, expressed
Chr1:33719449	5.7E−11	33334269-33780964	2.42	3.04	−0.78	Weak	LOC_Os01g58950.1	Cytochrome P450, putative, expressed
Chr3:4655686	1.38E−11	4045839-4658179	2.78	2.6	−1.7	Strong	LOC_Os03g08930	Basic helix-loop-helix dimerisation region bHLH domain containing protein;
Chr3:9158937	1.12E−07	9326824-9810369	3.48	2.8	−0.79	Strong		
Chr3:13155280	3.11E−12	12620491-13044531	3.83	1.88	−1.45	Strong		
Chr4:12679715	3.22E−07	12320456-13135109	2.39	2.41	−0.97	Weak		
Chr4:20685858	1.82E−09	21337500-21848954	2.4	0.9	−1.3	Weak		
Chr6:8148068	3.41E−09	7070518-7601689	3.20	3.1	−1.2	Strong		
Chr7:6197350	4.88E−12	5536887-6172772	4.09	1.44	−0.78	Strong	LOC_Os07g11020 (*Rc*)	Pericarp color; red pericarp
Chr9:656027	6.98E−10	6084684-6583948	2.62	1.72	−1.46	Strong		
Chr10:296002	8.23E−11	2605942-3053511	1.61	1.94	−0.19	Weak		
Chr12:23036595	1.23E−13	22936893-23280883	3.16	0.56	−1.76	Strong	LOC_Os12g37419	Cytochrome c oxidase polypeptide Vc, putative, expressed 6

Chr represents a chromosome.

Selection Sweep Analysis and Candidate Genes

Selection sweep analyses were conducted in the "unimproved" red pericarp population and in the "improved" cultivated white pericarp population. The selection sweep parameter of $\pi_{\text{unimproved population}}/\pi_{\text{improved population}}$ per 500 kb genomic slide windows were highlighted from yellow (0.17) to magenta (4.22) (**Figure 3A**). Another artificial selection parameter Tajima's *D* represented by a black and white heat map, supported the selection sweep to define the selection genomic regions (**Figure 3B**). In order to nominate candidate genes that may contribute to the red pericarp color of weedy rice, we considered whether the GWAS peaks were covered by the selection signal. Finally, seven GWAS peaks with the strong selection signal were detected, of which the positions of known red pericarp gene *Rc* and two new candidate genes, *LOC_Os03g08930* and *LOC_Os12g37419*, were overlapped (**Figure 3C** and **Table 1**). In addition, a Glutelin-family-protein gene *Os03g0188500* (without MSU ID) was linked with *LOC_Os03g08930*, which may together contribute to grain quality and nutrition. Although the GWAS peak related *Rd* gene was detected, its selection signal was weak. The genome-wide GWAS peaks combined with the selection sweep were shown in **Supplementary Figure S2**.

Developmental Dynamics of Weedy Rice Pericarp Color

The typical red pericarp weedy rice WR07-14, WR07-47, WR03-32, WR07-141, were selected from the GWAS panel for the observation of pigment components and the deposition process. The pericarp color of all sampled weedy rice was green in the early stage of development (0–5 days after flowering). The pigment began to deposit on both ends of the caryopsis (6–8 days after flowering), accumulating along the vascular bundle at the back of the caryopsis (9–15 days after flowering), and finally occurring throughout the entire caryopsis (16–18 days after flowering). When the seeds were mature, the pigmented pericarps were slightly faded (26–28 days after flowering) (**Figure 4A**).

We dissected the weedy rice WR07-14 caryopsis, including pericarp, seed coat, and endosperm (including aleurone layer) at different stages of development. At 1–7 days after flowering, the caryopsis was green, the seed coat was transparent, and the endosperm was white. At 8–13 days after flowering, the accumulated pigment in the pericarp was a light color, the thin seed coat was lighter in color than the pericarp, and the endosperm was still white. At 14–21 days after flowering, the pericarp presented dark red, the seed coat was colored but thinner than the pericarp, and the endosperm was white. Based on these observations, the pigment was mainly deposited in the pericarp, a small amount was deposited in the seed coat, and the pigment was absent in the endosperm (**Figure 4B**).

At 2 weeks after flowering, the pericarp and seed coat cells of the weedy rice began to shrink, and the binding between them gradually became tight. The pericarp cells began to shrink, and the pigment began to accumulate mainly in the lower part of the mesocarpal cells. A small amount of pigment was deposited in the outer pericarp. Pigments were difficult to detect in the seed

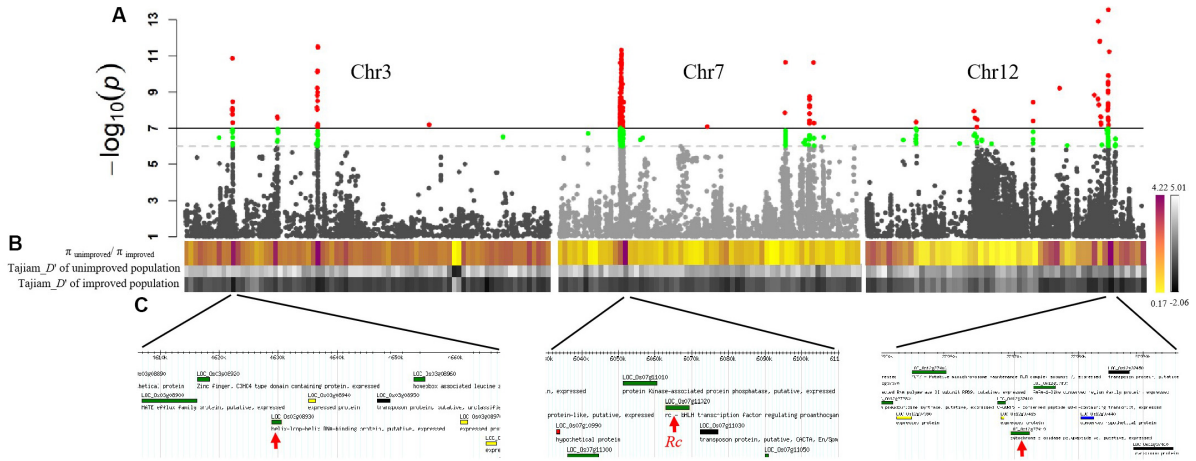


FIGURE 3 | Selection sweep analysis and candidate genes for weedy rice pericarp color. **(A)** GWAS results of weedy rice pericarp color in Chr3 Chr7 and Chr12 presented as Manhattan plot. **(B)** The upper colored line in the figure is π unimproved population/π per 500 kb genomic window improved population. Red indicates the *P* value above the threshold. *P*-values determined based on Bonferroni correction method. In order to increase the possibility of overlapping with the selection signal, we reduced *P* by an order of magnitude based on the Bonferroni correction, which is indicated by the green peak. Red regions indicates the strong selection, whereas the green regions indicates weak selection. The middle and lower grayscale lines are Tajima's *D* of the unimproved population and Tajima's *D* of the improved population, respectively. The darker the color, the greater the intensity of the selection. **(C)** For each response peak, the chromosomal locations of the candidate genes for weedy rice pericarp color are shown.

coat. The aleurone cells began to form out of the outer layer of endosperm, and they arranged themselves into a rectangular shape. At 3 weeks after flowering, the cell binding between pericarp and seed coat was more compact, and the seed coat layer was completely flattened. A large amount of pigment was deposited in the pericarp, and a small amount was deposited in the seed coat. The gap between the aleurone layer and the endosperm cells shrank, and neither the aleurone layer nor the endosperm portion was pigmented (**Figure 4C**).

Developmental Dynamics of Pigment Concentration and Composition in Weedy Rice Pericarp

To determine changes in pericarp pigmentation with weedy rice seed growth, we further measured the changes of anthocyanin concentration and composition in four typical weedy rice accessions, WR07-14, WR07-47, WR03-32, and WR07-141, since the anthocyanin is the main component of pericarp pigment. The concentration of anthocyanins rose and then fell until it stabilized, with the peak appearing sometime between the 15th to 20th day after flowering (**Figure 5A**). As expected, the anthocyanin concentration in dark red pericarp weedy rice WR07-14 was higher and accumulated more quickly than that in red pericarp weedy rice during the whole development period (**Figure 5A**). The ranking order of anthocyanin concentration and color from deep to light among rice strains was WR07-14 > WR07-47 > WR03-32 > WR07-141.

The anthocyanin components of weedy rice were detected by using the hydrochloric acid-vanillin method and high performance liquid chromatography-mass spectrometry (HPLC) (Baldi et al., 1995; Pei et al., 2017). We found that paeoniflorin (P3G) only occurred in dark red pericarp weedy rice (WR07-14 and WR07-47), based on HPLC, and proanthocyanidins only occurred in the red pericarp weedy rice (WR03-32 and WR07-141), based on the hydrochloric acid-vanillin method (**Table 2**). By comparing the peak positions of the 4 weedy rice accessions with the standards in the HPLC spectrum, we determined the anthocyanin composition of the samples (**Figure 5B**). Cyanidin (C3G) was detected in all 4 weedy rice accessions, and its concentration was higher in dark red pericarp weedy rice than in red pericarp weedy rice (**Table 2**).

Trace Element Differences Between Weedy Rice and Cultivated Rice

Cultivated red rice is considered to be a beneficial, healthful food (Rhanissa et al., 2011). However, the functional and nutritional qualities of weedy rice in red and other colors is poorly known. The concentrations of the metal elements were measured in weedy rice accessions WR07-14, WR07-47, WR03-32, WR07-141, WR07-142 and cultivated rice accessions Shennong265, Akihikari, Nipponbare, and Qishanzhan. The concentrations of various metal elements in weedy rice were significantly higher than that of control cultivated rice, sometimes 2–3 times greater (**Table 3**). The concentration of Mg was higher than that of other metal elements in both weedy rice and cultivated rice, followed by Ca > Zn and Mn > Cu > Mo > Se > Ge.

Free Amino Acid Content Differences Between Weedy Rice and Cultivated Rice

Free amino acids are important biologically active molecules which play a role in the synthesis of proteins and in providing energy for the body and brain activity (Zarei et al., 2017). A total of 17 free essential amino acids were detected in the five weedy rice and four cultivated rice varieties analyzed (**Table 4**).

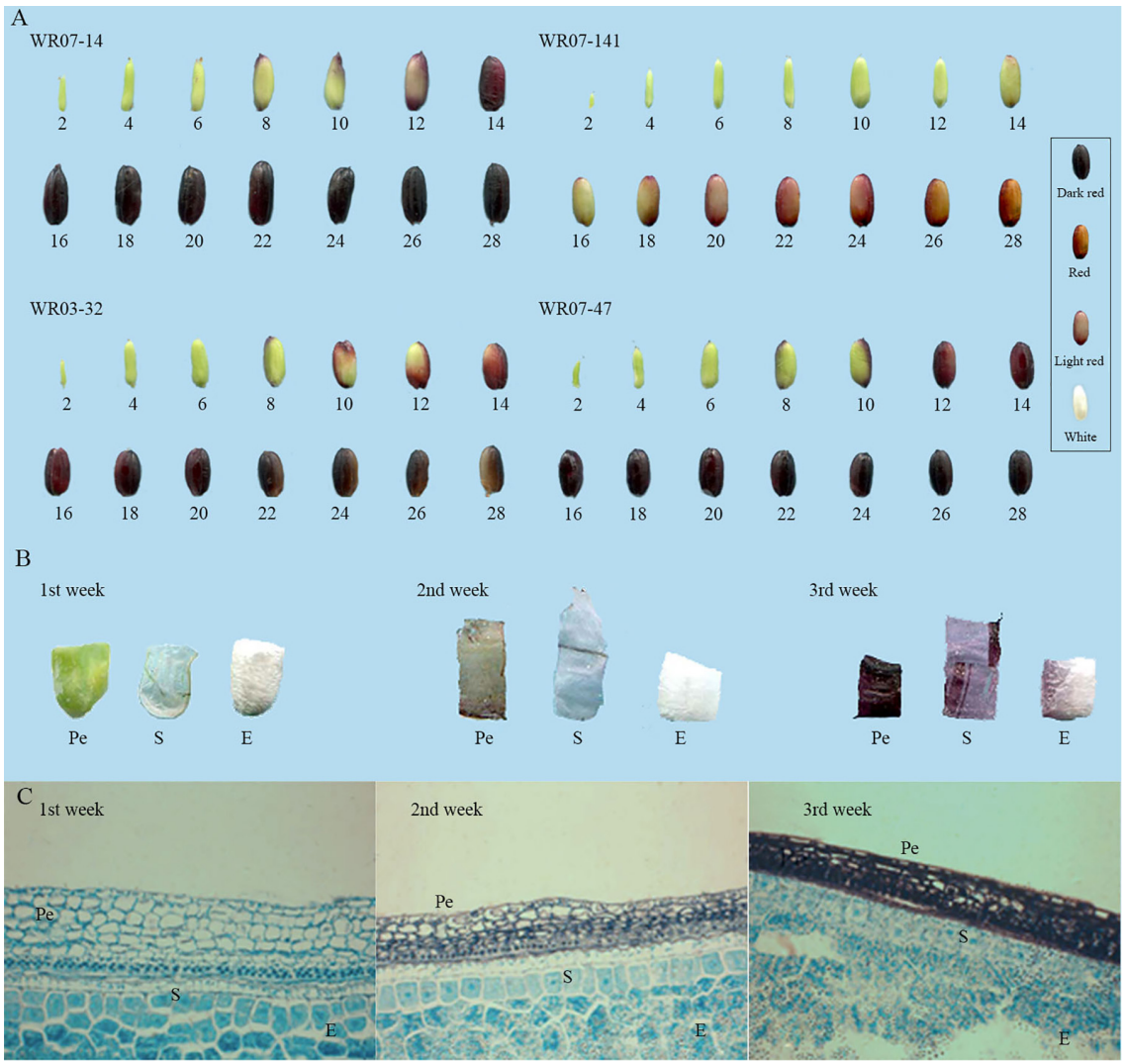


FIGURE 4 | Changes through time in the appearance of the pericarp and pigmentation sites of typical weedy rice. **(A)** Developmental changes in caryopsis appearance in different colors of weedy rice. Numbers represent the days after flowering. **(B)** Pigmentation during weeks 1–3 after flowering (Pe, pericarp; S, seed coat; E, endosperm; weedy rice WR07-14 shown). **(C)** Pigment deposition in cells during weeks 1–3 after flowering (weedy rice WR07-14 shown).

Glutamate (Glu) was the most abundant free amino acid in both weedy and cultivated rice. The total amino acid concentration was higher in the weedy rice than in the cultivated rice, and most individual free amino acids (Asp, Glu, Val, and Thr) were significantly higher in concentration in the weedy rice than in the cultivated rice. However, Cys and Gly were significantly lower in the weedy rice than in the cultivated rice (**Table 4**).

Fatty Acid Content Differences Among Color Pericarp Weedy Rice

Saturated fatty acids (palmitic acid and stearic acid) and unsaturated fatty acids (oleic acid and linoleic acid) were detected in weedy rice and cultivated rice in this study. We found that the overall ratio of saturated to unsaturated fatty acids in the weedy rice was slightly lower than that in cultivated rice. The ratios of saturated fatty acid and unsaturated fatty acid content were 20.53 and 79.47% in weedy rice, and the ratios in cultivated rice is 23.03 and 76.97%. There was no significant difference in the proportion of palmitic acid (a saturated fatty acid) between weedy rice and cultivated rice. The proportion of stearic acid (a saturated fatty acid) in weedy rice was significantly lower than that of control cultivated rice (**Table 5**). These indicators imply that weedy rice fatty acid is superior to the control cultivated rice.

DISCUSSION

Our research has shown certain differences in the development of pigmentation in weedy rice compared to previous studies. Liu et al. (2011) found that the growth and endosperm development of colored rice caryopsis are basically the same as

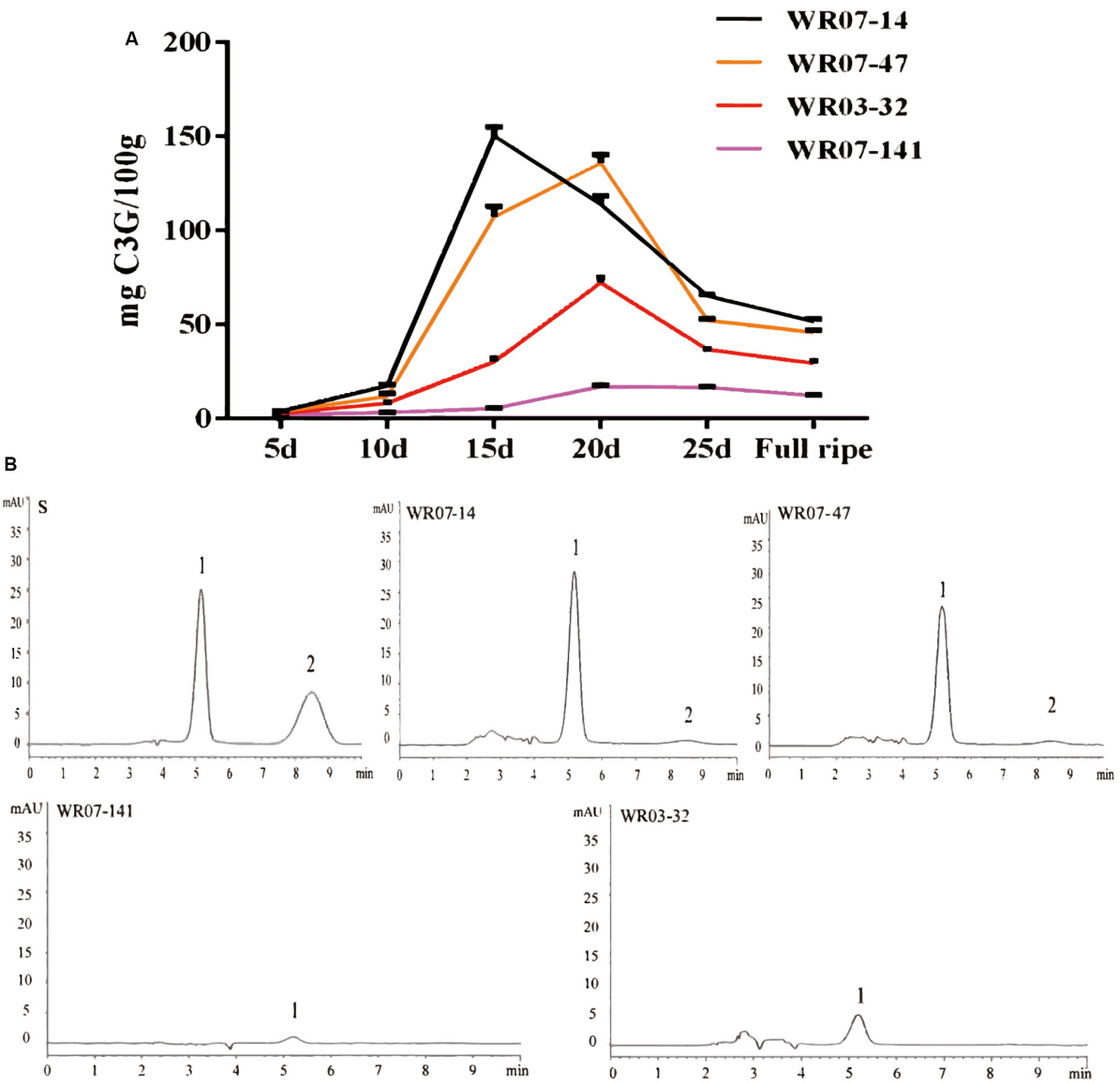


FIGURE 5 | Changes in anthocyanin content in differently colored varieties of weedy rice through different developmental stages. **(A)** The four different color polylines represent the total anthocyanin content of the 4 weedy rice accessions over 30 days. **(B)** HPLC spectra showing anthocyanin types observed in 4 weedy rice accessions varieties. Peak 1 is cyanidin-3-glucoside (C3G), and peak 2 is paeonidin-3-glucoside (P3G). Graph S represents the peak positions of the standard.

those of conventional rice varieties except for pigments such as anthocyanin. Four to five days after flowering, anthocyanin began to deposit in the pericarp of the caryopsis, and then deposited in the pericarp at both ends of the caryopsis and at the vascular bundle at the back, and finally covered the entire caryopsis 10 days after flowering. The anthocyanidin in the caryopsis mainly accumulated in the 7–20 days after flowering (Liu et al., 2011). Han observed that black rice pigment was synthesized and deposited only in the pericarp of the caryopsis, and no deposition in the seed coat and aleurone layer was observed (Han et al., 2009). The anthocyanidin was deposited on the pericarp 3 days after pollination, and the anthocyanin increased rapidly after 5–6 days of pollination. After 7 days, the anthocyanids filled the entire pericarp. The far end of the caryopsis first deposited anthocyanidose, then gradually extended to the end of the embryo and stained the entire skin (Heo et al., 2006). The pigment deposition process and location of weedy rice found in the present study were basically consistent with previous studies. Pigment deposition of weedy rice has 1–2 days delay compared to cultivated rice. In our study, accumulation of pigment in weedy rice began at both ends of the caryopsis and then accumulated in the vascular bundle on the back of the caryopsis, and the entire caryopsis was covered by pigment. The time when the weedy rice pigment began to deposit was 6–8 days after flowering, and the pigmentation time of the colored rice was slightly earlier. Most pigmentation occurred slightly in seed coats.

TABLE 2 | Anthocyanin content in different colored weedy rice varieties.

Weedy rice variety	Pericarp color	Anthocyanin (C3G)	Anthocyanin (P3G)
WR07-14	Dark red	44.92 ± 0.89^{a}	5.18 ± 0.12^{a}
WR07-141	Red	5.21 ± 0.22^{d}	N
WR03-32	Dark red	13.9 ± 0.31^{c}	N
WR07-47	Dark red	37.4 ± 0.64^{b}	5.21 ± 0.29^{a}

C3G stands for cyanidin-3-glucoside and P3G stands for peonidin-3-glucoside. The different lowercase letters in the columns of the table indicate the significant difference in the content of each component between the different weedy rice at the $p < 0.05$ level. Because the color of the control cultivated rice pericarp is white and colorless, it is 0. The unit of the anthocyanins is (mg/100 g).

It has been reported that anthocyanin accumulation was positively correlated with Superoxide Dismutase (SOD) in seeds, and anthocyanin accumulation led to darker seed coat color (Yi et al., 2019). In the present study, anthocyanin in grains increased with filling during the filling stage. However, the anthocyanin content decreased after the seed maturation. We speculate that the decrease in grain vigor after seed maturation leads to a decrease in anthocyanin content, specifically the decrease in self-oxidant activity leading to a rapid decline in the accumulation rate of anthocyanins. Further related mechanisms are worth to study in the future.

Genetic Basis of Weedy Pericarp Color and Its Significance to Domestication

During the domestication of *Oryza sativa*, pericarp color was an important trait that humans targeted to improve rice quality (Qiu et al., 2017). *Oryza rufipogon* and some of the early landraces exhibited red pericarp. However, modern rice cultivars appear white due to a lack of red pigmentation in the pericarp. *Rc* and *Rd* are two key genes related to pericarp color. The wild type *Rc* allele encodes a basic helix–loop–helix (bHLH) transcription factor, and the loss-of-function mutant *rc* allele causes the change from red to white pericarp (Sweeney et al., 2006). *Rc* gene loci were shown to have contributed to rice domestication (Sweeney et al., 2007). *Rd* is another synergistic gene of rice pericarp color that enhances the effect of *Rc* gene to promote proanthocyanidin synthesis (Furukawa et al., 2007). The purple seed coat gene *Pb* and its complementary gene *Pp* are located on chromosomes 4 and 1, respectively. When the mutant allele of *Pb* is present, the pericarp is white. When the *Pb* gene is present alone, the pericarp is brown. When the genes *Pb* and *Pp* are present at the same time, the seed coat is purple (Wang et al., 2014).

In this study, the association signals of purple rice genes *Pp* and *Pb* could not be detected in red pericarp rice (weedy or cultivated) through the GWAS panel. This implies that the two genes did not contribute to the red pericarp of weedy rice because they are two rare alleles specific to purple rice. On the other hand, we found that the weedy rice may have a novel and complex genetic system of pericarp color that involves *Rc, Rd,* and other unexplored genes. Rice pericarp color was a typical domestication-related trait of cultivated rice with a relatively simple genetic basis, i.e., the *Rc* gene experienced intense artificial selection on the pericarp color. We found some minor effect or synergistic alleles that may retain an undomesticated pericarp color gene in weedy rice. Therefore, in semi-domesticated weedy rice populations, there may exist some unexplored genetic basis of different types of agronomic traits that will further clarify the process of rice domestication. For instance, a nutrient gene (*Os03g0188500*) related to glutamic acid biosynthesis is linked to a candidate pericarp color gene (*LOC_Os03g08930*), which implies a potential selection pressure related to pigments and nutrients of weedy rice.

Significance of Weedy Rice Anthocyanin Biosynthesis

Rice anthocyanins could be a beneficial part of human diets due to their high antioxidant activities. Lei pointed out that red rice pigment was synthesized and deposited only in the pericarp of the caryopsis and absent in the seed coat and aleurone layer (Lei et al., 2006). However, a breakthrough was

TABLE 3 | Metal element content in different rice varieties.

Rice variety	Color	Fe	Ca	Mg	Mn	Cu	Zn	Mo	Se	Ge
WR07-47	Dark red	42.31	261.60	765.50	43.40	3.94	129.69	0.73	0.07	0.014
WR03-32	Dark red	23.18	164.40	762.78	28.56	2.52	59.91	0.77	0.03	0.0068
WR03-29	Red	26.71	165.90	785.04	24.46	2.23	48.78	0.58	0.02	0.0008
WR07-141	Red	38.14	177.90	986.19	29.30	3.78	73.18	1.10	0.06	0.014
WR07-142	Red	20.85	127.25	1048.90	28.91	3.28	45.77	0.46	0.05	0.0069
AKI	White	9.86	55.35	458.34	11.66	1.32	16.99	0.21	0.02	0.0034
SN265	White	9.17	83.40	435.40	14.89	1.43	22.50	0.20	0.01	0.0016
QSZ	White	8.84	77.60	534.30	10.98	1.12	19.83	0.31	0.01	0.0028
NIP	White	11.37	84.00	654.60	12.87	1.44	26.01	0.22	0.02	0.0032
WR-AVE	Red	30.24^{aA}	179.41^{aA}	869.68^{aA}	30.93^{aA}	3.15^{aA}	71.47^{aA}	0.73^{aA}	0.05^{aA}	0.0085^{aA}
CR-AVE	White	9.81^{bB}	75.09^{bB}	520.66^{bB}	12.60^{bB}	1.33^{bB}	21.33^{bB}	0.23^{bB}	0.015^{bB}	0.0028^{bB}

WR (Weedy rice), AKI (Akihikari), SN265 (Shennong265), QSZ (Qishanzhan), NIP (Nipponbare), WR-AVE (Weedy rice Average), CR-AVE (Cultivated Rice Average). AKI, SN265, QSZ, and NIP are control cultivated rice. The different lowercase letters in the columns of the table indicate the significant difference in the content of each component between weedy rice and control cultivated rice at $p < 0.05$ level. Capital letters indicate the significant difference in the content of each component between weedy rice and control cultivated rice at $p < 0.01$ level. The unit of the anthocyanins is (mg/kg).

TABLE 4 | Free amino acid content in different rice varieties.

Amino acid	WR07-47	WR03-32	WR03-29	WR07-141	WR07-142	AKI	SN265	QSZ	NIP	WR-Ave	CR-Ave
Asp	7.27	5.58	5.87	6.54	6.074	4.85	4.21	4.12	3.95	6.27^{aA}	4.28^{bB}
Thr	2.59	1.9	2.01	2.34	2.15	1.79	1.66	1.68	1.89	2.20^{aA}	1.76^{bA}
Ser	3.76	2.71	2.71	2.71	2.71	2.71	2.68	2.59	2.71	2.92^{aA}	2.67^{aA}
Glu	13.39	12.66	10.67	12.25	11.04	9.43	9.6	9.23	9.68	12.0^{aA}	9.49^{bB}
Gly	3.22	2.35	2.56	2.76	2.59	3.69	3.87	3.63	3.58	2.70^{bB}	3.69^{aA}
Ala	4.05	2.81	3.07	3.61	3.23	3.27	3.3	3.33	3.24	3.35^{aA}	3.29^{aA}
Cys	2.84	2.65	2.65	2.69	2.85	3.49	3.59	3.36	3.61	2.73^{bB}	3.51^{aA}
Val	5.88	6.3	6.89	5.59	5.63	5.21	5.13	5.22	5.19	6.06^{aA}	5.19^{bA}
Met	1.96	1.612	1.793	1.82	1.78	1.8	1.83	1.78	1.67	1.79^{aA}	1.77^{aA}
Ile	2.28	1.49	1.78	2.13	1.89	1.65	1.63	1.75	1.55	1.91^{aA}	1.65^{aA}
Leu	5.03	3.702	3.77	4.62	4.077	3.63	3.67	3.55	3.48	4.24^{aA}	3.58^{aA}
Tyr	2.25	1.34	1.58	1.97	1.74	1.43	1.48	1.33	1.49	1.78^{aA}	1.43^{aA}
Phe	5.27	4.68	5.06	5.18	5.23	4.97	5.03	4.86	4.94	5.08^{aA}	4.95^{aA}
Lys	3.1	2.41	2.55	2.89	2.85	2.5	2.55	2.45	2.41	2.76^{aA}	2.48^{aA}
His	1.54	0.97	0.97	1.33	1.2	0.99	1.02	0.93	0.93	1.20^{aA}	0.97^{aA}
Arg	6.35	4.53	4.81	5.73	5.26	4.79	4.73	4.89	4.67	5.34^{aA}	4.77^{aA}
NH3	0.23	0.19	0.2	0.18	0.21	0.19	0.2	0.22	0.18	0.20^{aA}	0.20^{aA}
Total	71.03	57.86	59.13	64.93	60.77	56.6	55.42	56	55.4	62.64^{aA}	55.87^{bA}

WR (Weedy rice), AKI (Akihikari), SN265 (Shennong265), QSZ (Qishanzhan), NIP (Nipponbare), WR-AVE (Weedy rice Average), CR-AVE (Cultivated Rice Average). AKI, SN265, QSZ, and NIP are control cultivated rice. The different lowercase letters in the columns of the table indicate the significant difference in the content of each component between weedy rice and control cultivated rice at $p < 0.05$ level. Capital letters indicate the significant difference in the content of each component between weedy rice and control cultivated rice at $p < 0.01$ level. The unit of the anthocyanins is (mg/g).

made by Zhu et al. who recently engineered a high-efficiency vector system for transgene stacking to enable anthocyanin biosynthesis in endosperm. They made a construct containing eight anthocyanin-related genes driven by endosperm-specific promoters and generated a novel biofortified germplasm, "Purple Endosperm Rice," with a high anthocyanin content (Zhu et al., 2017). However, the use of such as genetically modified organism (GMO) as a staple food still requires a long process of research and approval. Alternatively, developing natural germplasm resources with high anthocyanin contents can be achieved in the short term. In particular, marker-assisted breeding can accelerate this process once the genetic basis of desired traits is determined.

TABLE 5 | Fatty acid content in different rice varieties.

Rice variety	Saturated fatty acids		Unsaturated fatty acids		Total %
	Palmitic acid	Stearic acid	Oleic acid	Linoleic acid	
WR07-47	20.72	1.86	36.97	40.45	100.00
WR03-32	19.49	1.97	44.05	34.49	100.00
WR03-29	19.69	1.81	42.86	35.60	100.00
WR07-141	17.41	1.28	42.63	38.69	100.00
WR07-142	17.06	1.38	43.53	38.03	100.00
AKI	20.11	2.81	41.59	35.49	100.00
SN265	20.22	2.71	41.55	35.52	100.00
QSZ	20.48	2.80	41.49	35.23	100.00
NIP	20.15	2.85	41.54	35.46	100.00
WR-AVE	18.874^{aA}	1.66^{aA}	42.01^{aA}	37.45^{aA}	100.00
CR-AVE	20.24^{aA}	2.79^{bB}	41.54^{aA}	35.42^{aA}	100.00

WR (Weedy rice), AKI (Akihikari), SN265 (Shennong265), QSZ (Qishanzhan), NIP (Nipponbare), WR-AVE (Weedy rice Average), CR-AVE (Cultivated Rice Average). AKI, SN265, QSZ, and NIP are control cultivated rice. The value represents the percentage of each fatty acid component in the total. The different lowercase letters in the columns of the table indicate the significant difference in the content of each component between weedy rice and control cultivated rice at $p < 0.05$ level. Capital letters indicate the significant difference in the content of each component between weedy rice and control cultivated rice at $p < 0.01$ level.

Rice nutritional quality has attracted more attention in the traditional growing areas of Asia, where monotonous consumption of rice may lead to deficiencies of essential trace elements, amino acid, and other nutrients (Bouis, 2003). In this study, we found that sampled weedy rice had much greater quantities of anthocyanin, beneficial trace elements, free amino acids, and unsaturated fatty acids than the control cultivated rice. Unsaturated fats are considered healthier for consumption than saturated fats (Yasumatsu and Moritaka, 2008). Therefore, the gene resources and novel genetic systems of rice anthocyanin biosynthesis explored in this study are of great value for the development of high-anthocyanin content rice.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

DM, JS, and MZ conceived the project and experiment. JS, WW, and MZ performed the SLAF sequencing and population

genetic analysis. WW observed the process of pigmentation in weedy rice and cultivated rice. WW and JyS examined the composition and content of pigment in weedy rice and cultivated rice, and analyzed the nutritional quality of weedy rice and cultivated rice. WW, JS, DM, MZ, GZ, ZL, YH, and XJ provided the germplasm and performed the germplasm management. JS and WW conducted a selection strength analysis. WW, JS, MZ, and DM interpreted the data and wrote the manuscript. All authors contributed to the manuscript and approved the submitted version.

FUNDING

This work was supported by the National Key R&D Program of China (grant number 2017YFD0100501), Program of Liaoning Revitalization (XLYC1808003), and the National Natural Science Foundation of China (grant number U1708231).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.00878/full#supplementary-material

FIGURE S1 | QQ Plot diagram of the improved and unimproved red rice populations' GWAS results. **(A)** The red pericarp in the sample is the result of the cultivation of rice. The lower fitting part indicates that the model is reasonable, and the upper raised part indicates that it is affected by the color-related locus of the pericarp. **(B)** The red pericarp in the sample of weedy rice. The lower fitting part indicates that the model is reasonable, and the upper raised part indicates that it is affected by the color-related locus of the pericarp.

FIGURE S2 | Genome-wide selection sweep with GWAS peak. **(A)** Selection sweep plot of π ratio ($\pi_{\text{unimproved population}}/\pi_{\text{improved population}}$), the black color line indicates the threshold line which used to determine strong or weak selection signal. **(B)** The dark green indicates selection sweep of Tajima's *D* in unimproved population. The light green indicates selection sweep of Tajima's *D* in improved population. The red dotted line indicates the GWAS peaks.

TABLE S1 | Material information of the GWAS populations.

REFERENCES

Baldi, A., Romani, A., Mulinacci, N., Vincieri, F. F., and Casetta, B. (1995). HPLC/MS application to anthocyanins of *Vitis vinifera* L. *J. Agric. Food Chem.* 43, 2104–2109. doi: 10.1021/jf00056a027

Bouis, H. E. (2003). Micronutrient fortification of plants through plant breeding: can it improve nutrition in man at low cost? *Proc. Nutr. Soc.* 62, 403–411. doi: 10.1079/PNS2003262

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., and Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23, 2633–2635. doi: 10.1093/bioinformatics/btm308

Brespatry, C., Lorieux, M., Clément, G., Bangratz, M., and Ghesquière, A. (2001). Heredity and genetic mapping of domestication-related traits in a temperate japonica weedy rice. *Theor. Appl. Genet.* 102, 118–126. doi: 10.1007/s001220051626

Chen, L. J., Lee, D. S., Song, Z. P., Suh, H. S., and Lu, B. R. (2004). Gene flow from cultivated rice (*Oryza sativa*) to its weedy and wild relatives. *Ann. Bot.* 93, 67–73. doi: 10.1093/aob/mch006

Cui, Y., Of, S. B. S., Li, L. F., Li, Y. L., Huang, Z., Caicedo, A. L., et al. (2016). Little white lies: pericarp color provides insights into the origins and evolution of southeast asian weedy rice. *G3 Genesgenet.* 6:g3.116.035881. doi: 10.1534/g3.116.035881

Dai, W. M., Song, X. L., Chuan, W. U., Zhang, L. J., Cao, F., Liu, L. L., et al. (2009). Investigation of weedy rice (*Oryza sativa* f. spontanea) and its caryopsis microstructure in jiangsu province. *Jiangsu J. Agric. Sci.* 25:9699.

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

Furukawa, T., Maekawa, M., Oki, T., Suda, I., Iida, S., Shimada, H., et al. (2007). The *Rc* and *Rd* genes are involved in proanthocyanidin synthesis in rice pericarp. *Plant J.* 49, 91–102. doi: 10.1111/j.1365-313X.2006.02958.x

Gross, B. L., Michael, R., Shih-Chung, H., Caicedo, A. L., Yulin, J., and Olsen, K. M. (2010). Seeing red: the origin of grain pigmentation in US weedy rice. *Mol. Ecol.* 19, 3380–3393. doi: 10.1111/j.1365-294X.2010.04707.x

Gross, J. H. (2017). *Inorganic Mass Spectrometry*. Cham: Springer.

Gu, D. (1992). A study on special nutrient of purple black glutinous rice. *Sci. Agric. Sin.* 25, 36–41.

Guo, Y. M., Duan, Y. B., Shao-Ming, L. I., Huang, P., Jian, T. U., Hua-Hui, L. I., et al. (2011). Evaluation and correlation analysis on mineral concentrations and pigment content in pericarp of color rice. *J. Plant Genet. Resour.* 12, 971–974.

Han, E., Dong, B. C., Yang, X. J., Huang, C. B., Wang, X. D., and Xian-Jun, W. U. (2009). Study on flavonoids in the caryopsis of indica rice rdh. *Agric. Sci. China* 8, 249–256. doi: 10.1016/s1671-2927(09)60034-1

Heo, S. R., Jin, S. K., Chang, H. E., Park, K. U., Song, J., and Kim, E. C. (2006). Detection of enterovirus in cerebrospinal fluid by real-time nested reverse transcription polymerase chain reaction. *Korean J. Lab Med.* 26, 9–13.

Hua, B. I., Shi, X. C., Liu, L., Han, C. M., Ding, Y. X., Pei, D. N., et al. (2016). Preliminary analysis on determination of national standard for protein content by amino acid assay. *Chin. J. Biol.* 29, 188–196.

Huang, X., Wei, X., Sang, T., Zhao, Q., Feng, Q., Zhao, Y., et al. (2010). Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat. Genet.* 42, 961–967. doi: 10.1038/ng.695

Juliano, B. O., and Tuaño, A. P. P. (2019). "2 - Gross structure and composition of the rice grain," in *Rice*, 4th Edn, ed. J. Bao (Washington, D.C: AACC International Press), 31–53. doi: 10.1016/b978-0-12-811508-4.00002-2

Kim, M. K., Kim, H. A., Koh, K., Kim, H. S., Lee, Y. S., and Kim, Y. H. (2008). Identification and quantification of anthocyanin pigments in colored rice. *Nutr. Res. Pract.* 2, 46–49.

Lei, H., Zhang, H. Y., Pei-Zhou, X. U., Yun, L. I., Wang, X. D., and Xian-Jun, W. U. (2006). Genetic analysis and histological study of red seed in rice. *Acta Genet. Sin.* 33, 559–564. doi: 10.1016/s0379-4172(06)60085-1

Li, J., Hu, Z., Ma, L., Ying, L., and Dian, J. (2014). Juice yield and anthocyanins stability during the processing of blueberry. *Food Sci.* 35, 120–125. doi: 10.7506/spkx1002-6630-201402022

Li, L. F., Li, Y. L., Jia, Y., Caicedo, A. L., and Olsen, K. M. (2017). Signatures of adaptation in the weedy rice genome. *Nat. Genet.* 49, 811–814. doi: 10.1038/ng.3825

Li, R., Liu, T., Cui, S., Zhang, S., Yu, J., and Song, G. (2017). Residue behaviors and dietary risk assessment of dinotefuran and its metabolites in *Oryza sativa* by a new HPLC-MS/MS method. *Food Chem.* 235, 188–193. doi: 10.1016/j.foodchem.2017.04.181

Li, R. Q., Li, Y. R., Karsten, K., and Wang, J. (2008). SOAP: short oligonucleotide alignment program. *Bioinformatics* 24, 713–714. doi: 10.1093/bioinformatics/btn025

Liu, M., Wang, Z., and Yun-Jie, G. U. (2011). Caryopsis development and anthocyanidin accumulation of colored rice. *Chin. J. Rice Sci.* 25, 392–398. doi: 10.4162/nrp.2008.2.1.46

Ma, D., Li, M., and Sun, J. (2008). Studies on biological diversity and genetic differentiation of liaoning weedy rice. *J. Shenyang Agric. Univ.* 39, 265–269. doi: 10.1007/s10499-007-9164-4

Meng, F., Wei, Y., and Yang, X. (2005). Iron content and bioavailability in rice. *J. Trace Elem. Med. Biol.* 18, 333–338. doi: 10.1016/j.jtemb.2005.02.008

Pei, Y. Y., Xing-Xing, L. U., Cui-Fang, W. U., and Yan-Fei, X. U. (2017). Analysis of proanthocyanidins and polyphenols in peanut skins. *Food Res. Dev.* 38, 143–147.

Pipatpongpinyo, W., Korkmaz, U., Wu, H., Kena, A., Ye, H., Feng, J., et al. (2019). Assembling seed dormancy genes into a system identified their effects on

seedbank longevity in weedy rice. *Heredity* 124, 135–145. doi: 10.1038/s41437-019-0253-8

Qiu, J., Zhou, Y., Mao, L., Ye, C., Wang, W., Zhang, J., et al. (2017). Genomic variation associated with local adaptation of weedy rice during de-domestication. *Nat. Communi.* 8:15323. doi: 10.1038/ncomms15323

Rahman, M. M., Lee, K. E., Lee, E. S., Matin, M. N., Dong, S. L., Yun, J. S., et al. (2013). The genetic constitutions of complementary genes *Pp* and *Pb* determine the purple color variation in pericarps with cyanidin-3- O -glucoside depositions in black rice. *J. Plant Biol.* 56, 24–31. doi: 10.1007/s12374-012-0043-9

Rhanissa, H., William, D. J., and Trust, B. (2011). Comparative evaluation of the antioxidant potential of infant cereals produced from purple wheat and red rice grains and LC-MS analysis of their anthocyanins. *J. Agric. Food Chem.* 59:12330. doi: 10.1021/jf202662a

Roberts, A., McMillan, L., Wang, W., Parker, J., Rusyn, I., and Threadgill, D. (2007). Inferring missing genotypes in large SNP panels using fast nearest-neighbor searches over sliding windows. *Bioinformatics* 23, i401–i407. doi: 10.1093/bioinformatics/btm220

Sellappan, K., Datta, K., Parkhi, V., and Datta, S. K. (2009). Rice caryopsis structure in relation to distribution of micronutrients (iron, zinc, beta-carotene) of rice cultivars including transgenic indica rice. *Plant Sci.* 177, 557–562. doi: 10.1016/j.plantsci.2009.07.004

Sun, J., Ma, D., Tang, L., Zhao, M., Zhang, G., Wang, W., et al. (2019). Population genomic analysis and de novo assembly reveal the origin of weedy rice as an evolutionary game. *Mol. Plant* 12, 632–647. doi: 10.1016/j.molp.2019.01.019

Sun, J., Qian, Q., Ma, D.-R., Xu, Z.-J., Liu, D., Du, H.-B., et al. (2013). Introgression and selection shaping the genome and adaptive loci of weedy rice in northern China. *New Phytol.* 197, 290–299. doi: 10.1111/nph.12012

Sweeney, M. T., Thomson, M. J., Gu, C. Y., Jin, P. Y., Williamson, S. H., Bustamante, C. D., et al. (2007). Global dissemination of a single mutation conferring white pericarp in rice. *PLos Genet.* 3:e133. doi: 10.1371/journal.pgen.0030133

Sweeney, M. T., Thomson, M. J., Pfeil, B. E., and Susan, M. C. (2006). Caught red-handed: *Rc* encodes a basic helix-loop-helix protein conditioning red pericarp in rice. *Plant Cell* 18, 283–294. doi: 10.1105/tpc.105.038430

Tai, T. H. (2002). Identification of red rice, rice, and hybrid populations using microsatellite markers. *Weed Sci.* 50, 333–339. doi: 10.1614/0043-1745(2002)050[0333:iorrra]2.0.co;2

Vaughan, L. K., Ottis, B. V., Prazak-Havey, A. M., Bormans, C. A., Sneller, C., Chandler, J. M., et al. (2017). Is all red rice found in commercial rice really *Oryza sativa*? *Weed Sci.* 49, 468–476. doi: 10.1614/0043-1745(2001)049[0468:iarrfi]2.0.co;2

Wang, W., Mauleon, R., Hu, Z., Chebotarov, D., Tai, S., Wu, Z., et al. (2018). Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature* 557, 43–49. doi: 10.1038/s41586-018-0063-9

Wang, F. Q., Yang, J., Fan, F. J., Wang, J., Zhu, J. Y., Wen-Qi, L. I., et al. (2014). Delayed inheritance of purple pericarp in rice and development of functional marker for *Pb* gene. *Chin. J. Rice Sci.* 606–611.

Yang, J., Wang, J., Cao, Q., Chen, Z. D., Tang, L. H., and Zhong, W. G. (2009). Genetic analysis of red pericarp in weedy rice (*Oryza sativa* L.f.spontaneous). *Acta Bot. Boreali Occidentalia Sin.* 29, 1084–1090.

Yasumatsu, K., and Moritaka, S. (2008). Fatty acid compositions of rice lipid and their changes during storage. *J. Agric. Chem. Soc. Jap.* 28, 257–264. doi: 10.1271/bbb1961.28.257

Yi, Z., Xibo, F., Yuefeng, M., Jingqin, T., and Jianli, W. (2019). Study on the correlation seed coat colour and seed vigor of colored barley. *J. China Agric. Univ.* 24, 18–29.

Yu, X., Zhang, W., Han, Y., Jing, L., Hai, X., Zhao, M., et al. (2010). Research on the dyeing time of the caryopsis of japonica rice in paraffin section and its anatomical structure. *Crops* 80–83. doi: 10.1016/S1875-2780(09)60057-2

Yue-Ting, H. U., Wang, J. G., Liu, H. L., Sun, J., Sun, X. X., Zhao, H. W., et al. (2016). Functional nutrition quality analysis of japonica rice(*Oryza sativa* L.) varieties in the cold region. *J. Plant Genet. Resour.* 17, 840–845. doi: 10.13430/j.cnki.jpgr.2016.05.007

Zarei, I., Brown, D. G., Nealon, N. J., and Ryan, E. P. (2017). Rice bran metabolome contains amino acids, vitamins and cofactors, and phytochemicals with medicinal and nutritional properties. *Rice* 10:24. doi: 10.1186/s12284-017-0157-2

Zhu, Q., Yu, S., Zeng, D., Liu, H., Wang, H., Yang, Z., et al. (2017). Development of "purple endosperm rice" by engineering anthocyanin biosynthesis in the endosperm with a high-efficiency transgene stacking system. *Mol. Plant* 10, 918–929. doi: 10.1016/j.molp.2017.05.008

Conflict of Interest: 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 © 2020 Wang, Zhao, Zhang, Liu, Hua, Jia, Song, Ma and Sun. 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) and the copyright owner(s) 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 Plant Science

HYPOTHESIS AND THEORY
published: 11 June 2020
doi: 10.3389/fpls.2020.00789

New Food Crop Domestication in the Age of Gene Editing: Genetic, Agronomic and Cultural Change Remain Co-evolutionarily Entangled

*David L. Van Tassel[1], Omar Tesdell[2], Brandon Schlautman[1], Matthew J. Rubin[3], Lee R. DeHaan[1], Timothy E. Crews[1] and Aubrey Streit Krug[1]**

[1] The Land Institute, Salina, KS, United States, [2] Department of Geography, Birzeit University, Birzeit, Palestine, [3] Donald Danforth Plant Science Center, St. Louis, MO, United States

The classic domestication scenario for grains and fruits has been portrayed as the lucky fixation of major-effect "domestication genes." Characterization of these genes plus recent improvements in generating novel alleles (e.g., by gene editing) have created great interest in *de novo* domestication of new crops from wild species. While new gene editing technologies may accelerate some genetic aspects of domestication, we caution that *de novo* domestication should be understood as an iterative process rather than a singular event. Changes in human social preferences and relationships and ongoing agronomic innovation, along with broad genetic changes, may be foundational. Allele frequency changes at many loci controlling quantitative traits not normally included in the domestication syndrome may be required to achieve sufficient yield, quality, defense, and broad adaptation. The environments, practices and tools developed and maintained by farmers and researchers over generations contribute to crop yield and success, yet those may not be appropriate for new crops without a history of agronomy. New crops must compete with crops that benefit from long-standing participation in human cultural evolution; adoption of new crops may require accelerating the evolution of new crops' culinary and cultural significance, the emergence of markets and trade, and the formation and support of agricultural and scholarly institutions. We provide a practical framework that highlights and integrates these genetic, agronomic, and cultural drivers of change to conceptualize *de novo* domestication for communities of new crop domesticators, growers and consumers. Major gene-focused domestication may be valuable in creating allele variants that are critical to domestication but will not alone result in widespread and ongoing cultivation of new crops. Gene editing does not bypass or diminish the need for classical breeding, ethnobotanical and horticultural knowledge, local agronomy and crop protection research and extension, farmer participation, and social and cultural research and outreach. To realize the ecological and social benefits that a new era of *de novo* domestication could offer, we call on funding agencies, proposal reviewers and authors, and research communities to value and support these disciplines and approaches as essential to the success of the breakthroughs that are expected from gene editing techniques.

Keywords: domestication, coevolution, gene editing, agronomy, cultural evolution, new crops

OPEN ACCESS

Edited by:
Eric Von Wettberg, The University of Vermont, United States

Reviewed by:
Ashley DuVal, Mars, United States
Edward Marques, The University of Vermont, United States

****Correspondence:***
Aubrey Streit Krug
streitkrug@landinstitute.org

Specialty section:
This article was submitted to Plant Breeding, a section of the journal Frontiers in Plant Science

Received: *06 January 2020*
Accepted: *18 May 2020*
Published: *11 June 2020*

Citation:
Van Tassel DL, Tesdell O, Schlautman B, Rubin MJ, DeHaan LR, Crews TE and Streit Krug A (2020) New Food Crop Domestication in the Age of Gene Editing: Genetic, Agronomic and Cultural Change Remain Co-evolutionarily Entangled. Front. Plant Sci. 11:789. doi: 10.3389/fpls.2020.00789

INTRODUCTION

De novo domestication of new crops from currently wild plants could help solve a wide range of problems, including genetic and species diversification of agricultural systems (Fernie and Yan, 2019); expansion of agricultural production onto degraded sites, stressful environments, or regions highly vulnerable to climatic extremes (Zhang et al., 2018); improvement of the fit between crops and particular local ecological niches (Fernie and Yan, 2019); and intensification of the range of vital ecosystem services provided by crops (Weißhuhn et al., 2017). Many new food crops will need to produce larger, more harvestable tubers, roots, fruits and seeds. Because this kind of change is both visually obvious, including in archeological records, and something that evolved independently many times in the past, it is not surprising that these are the core traits of the domestication syndrome for food plants (Doebley et al., 2006; Dong et al., 2019; Woodhouse and Hufford, 2019) and attractive targets for genetic modification.

In much of the recent *de novo* food crop domestication literature, domestication and the fixation of alleles conferring the domestication syndrome are used as interchangeable concepts (discussed below). However, our definition is closer to that of Harlan (1992), who describes broader changes: "Domesticated plants are those brought into the *domus* [Latin for household] which may mean the dooryard, garden, field, orchard, vineyard, pasture or ranch. It may also include yards, parks, cemeteries, golf courses, roadsides, forests, and other managed areas. In ecological terms it is the change in habitat that is critical.... domestication tends in the direction of making the plant populations dependent on human interference and man-made habitats. Since the processes of domestication are evolutionary in nature, all intermediate degrees and conditions may be expected, but a fully domesticated plant is entirely dependent on human intervention for survival."

Advances in gene editing technological interventions, such as CRISPR/Cas9, could enable a new era of *de novo* domestication (Østerberg et al., 2017; Chen et al., 2019; Eshed and Lippman, 2019; Fernie and Yan, 2019; Khan et al., 2019; Wolter et al., 2019) defined as "the introduction of domestication genes into non-domesticated plants" (Fernie and Yan, 2019). Several labs have independently and successfully modified domestication-related genes in wild species as proof-of-concept (Lemmon et al., 2018; Li et al., 2018; Zsögön et al., 2018). Discovering or creating favorable alleles of relatively simply inherited "domestication genes" (DGs) (Doebley et al., 2006; Østerberg et al., 2017; Khan et al., 2019) – whether through classic selective breeding, mutagenesis or genome-editing – is helpful during *de novo* domestication.

Such research could contribute new understanding and opportunities, particularly in well-characterized genes such as the seed dormancy gene G, which appears to have been selected in parallel during the domestication of several crops in at least three plant families (Rendón-Anaya and Herrera-Estrella, 2018). Gene editing and similar technologies may be necessary to produce effective DG alleles on a realistic timeframe for the *de novo* domestication of some wild plant taxa (Eshed and Lippman, 2019; DeHaan et al., 2020) and some authors recognize that novel variation must then be introduced into diverse germplasm (Lemmon et al., 2018) and "tuned" by traditional selection on standing variation of many small-effect loci (Eshed and Lippman, 2019).

However, other recent gene editing papers may inadvertently oversimplify the rich, complex process of plant domestication. Some imply that prehistoric domestication happened when "simple choices ultimately led to the pyramiding of valuable mutations and re-combinants in key genes" (Fernie and Yan, 2019) and emphasize the importance of monogenic domestication traits (Li et al., 2018). Zsögön et al. (2018) state that the editing of six genes demonstrates "that targeted reverse genetic engineering of wild plants could rapidly create new crops."

The availability of new tools such as CRISPR, and the possibility of applying them toward *de novo* domestication of novel crops, could be misunderstood by society to mean that gene editing can produce domestication as a nearly instant and/or inexpensive event, failing to recognize the many years of foundational research that was required to characterize and identify candidate genes and develop transformation and tissue culture regeneration systems (Van Eck, 2018). The tendency to emphasize the speed and ease of domestication of plants such as *Solanum pimpinellifolium* (currant tomato) via gene editing is illustrated in the following headlines and statements: "CRISPR can speed up nature... It took thousands of years for humans to breed a pea-sized fruit into a beautiful beefsteak tomato. Now, with gene editing, scientists can change everything" (Hall, 2018); "Gene editing can potentially cram millennia of agricultural progress into the blink of an eye" (Keats, 2019); "Plant breeding at the speed of light: The power of CRISPR/Cas. . ." (Wolter et al., 2019); "CRISPR/Cas brings plant biology and breeding into the fast lane" (Schindele et al., 2020); "The CRISPR technique quickly tamed 'unruly' ground cherries" (Saey, 2018); "For the first time, researchers have created, within a single generation, a new crop from a wild plant. . .by using a modern process of genome editing" (University of Munster, 2018); "Tweaking just a few genes in wild plants can create new food crops" (Hereward and Curtis, 2018).

While new gene editing technologies may accelerate some genetic aspects of domestication, we caution that *de novo* domestication resulting in widespread and ongoing cultivation should be more accurately understood as an ongoing, iterative process rather than a singular event. New genetic techniques may allow us the opportunity to begin – not to bypass – this "lengthy and tedious" (Wolter et al., 2019) work. Whereas earlier researchers emphasized the ways that humans had discovered loss-of-function plant mutations during domestication, leaving plants dependent upon farmers for their defense and dispersal (Zohary and Hopf, 2000; Anderson, 2005), the emerging consensus is that domestication is a long-term, co-evolutionary deepening of a mutualistic relationship, involving cultural, technological, biological, and ecological factors (Zeder et al., 2006; Gepts, 2010; Meyer et al., 2012; Meyer and Purugganan, 2013; Gremillion et al., 2014; Larson et al., 2014; Allaby et al., 2017; Swanson et al., 2018; and see especially the review by Zeder, 2015).

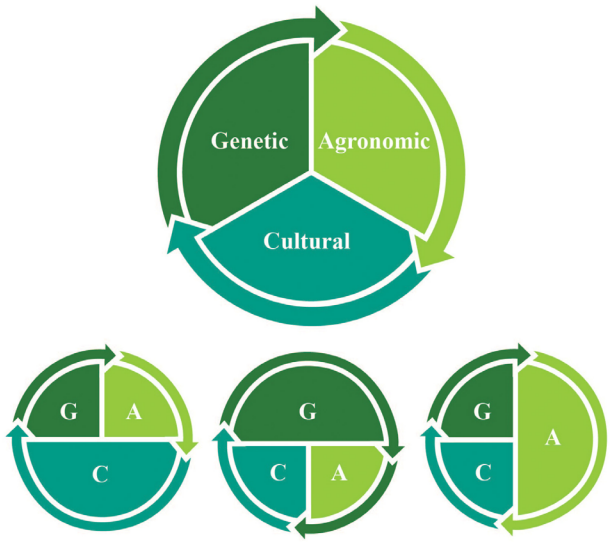


FIGURE 1 | Varying magnitudes of genetic (G), agronomic (A), and cultural (C) change and their interactions drive *de novo* domestication processes.

Historically, the genetic changes that have come to distinguish domesticated plants from their wild relatives are the outcome of an interaction of sociocultural and environmental pressures reflecting broader changes in human preference and agronomic intervention. In domestication, culturally innovative changes in human behavior that result in environmental modification (such as preparing soil and plant harvesting and processing) are "entangled" with biologically innovative changes in plant genetics (Fuller et al., 2010). Domestication is an ongoing process since people and plants live in dynamic social-ecological systems.

A more interdisciplinary approach is therefore needed to inform the review and implementation of new gene editing research on *de novo* domestication. We provide a practical framework that highlights and integrates multiple drivers of change (**Figure 1**) to conceptualize *de novo* domestication for communities of new crop domesticators, growers and consumers. Various magnitudes of genetic, agronomic, and cultural change drive cycles of selective pressures that result in the fixation of domestication genes – and broad genetic changes, innovations in management, and social support may all be required to enable domestication processes to continue at different pivotal moments and for the long term across human generations and geographies (**Table 1**). To realize the ecological and social benefits that a new era of *de novo* food crop domestication could offer, it is necessary to recognize how genetic, agronomic, and cultural changes remain co-evolutionarily entangled (**Figure 2**).

GENETIC CHANGE BEYOND DOMESTICATION GENES

Wild plants can be regarded as reservoirs of useful genes (Zhang et al., 2018) needed to create "climate-smart crops" (Li et al., 2018). However, the view that this genetic resource can be tapped by cleanly switching out a few domestication genes, hopefully "without causing an associated drag on other useful traits" (Li et al., 2018) is overly optimistic because it underestimates the additional genetic changes that will be required to improve wild plants' ability to respond to favorable environments, high plant density, increased population, and human care. Domestication is not a single genetic transformation. Other kinds of plant genetic research and genetic change are needed beyond and/or preceding modification of DGs.

Identification and Preparation of Candidates

Research may be needed to identify potential candidates for domestication (Ciotir et al., 2019), to rank species or subspecies by predicted breedability (e.g., genome size, ploidy, mating system), to identify barriers to human use (e.g., toxicity), and to prepare wild germplasm for gene-editing (e.g., genomic sequencing, mapping) (DeHaan et al., 2016). The foundation of a domestication program includes obtaining and characterizing the wild germplasm (Schlautman et al., 2020) and ascertaining that additive genetic variation for crop yield and disease resistance are available in the gene pool. A successful domestication candidate should come from a sufficiently large primary or secondary gene pool to ensure sustained breeding, as domestication by any means is only the beginning of continuous cycles of breeding for yield, quality and resistance.

Core Germplasm Development and "Tuning" of DGs

Domestication via gene-editing almost certainly implies a severe genetic bottleneck because it currently requires an efficient plant transformation/regeneration protocol (Chen et al., 2019; Fernie and Yan, 2019). Tissue culture protocols are not available for most wild plants or every genotype of crop plants; even where many amenable genotypes are available, genetic modification is laborious and may be constrained by intellectual property protection issues, thereby limiting the number of transformed lines that will be rapidly available to farmers and breeders. Mutational load could increase by drift during such a bottleneck (or later through hitchhiking with DGs) as appears to have happened during maize domestication (Wang et al., 2017). Pre-domestication inbreeding could be attempted to estimate mutation load and thus identify candidates with lower levels of preexisting load, or even to begin to purge it. Alternatively, edited genes could be immediately introgressed into pre-identified genetic diversity panels to reverse the bottleneck effect (Eshed and Lippman, 2019; Fernie and Yan, 2019), although recovering homozygosity for the edited allele while avoiding genetic bottlenecking will require sophisticated crossing strategies. Reversing bottlenecks, identifying unpredicted genetic background effects (Wolter et al., 2019), and the general need to "tune" the expression of novel major-effect alleles (Eshed and Lippman, 2019) all argue that extensive post-domestication breeding should be expected in addition to pre-domestication work.

TABLE 1 | Common evidence, processes, and participants and practices involved in the genetic, agronomic, and cultural drivers of *de novo* domestication and crop improvement.

Evidence: sources and results of domestication	Processes: iterative cycles of transmission and change in domestication	Participants and practices: major stakeholders and methods to accomplish domestication and crop improvement
Genetic		
• Extant and new cultigens • Archeological remains of extinct cultigens that are morphologically distinct from wild ancestors	• Plant lifecycle from seed to seed (months to a few years) • Selection of new scions to use in grafting (decades) • Division of tubers, rhizomes (yearly) • Formal breeding cycles, gene discovery projects (months to a few years)	• Germplasm collections at gene banks, botanical gardens • Traditional landrace development by farmers through visual or unconscious selection for reduced shattering, seed dormancy • Public and private plant breeding using sexual recombination and selection • Public or private genetic engineering (cisgenic, transgenic) • Seed exchange networks and community gene banks
Agronomic		
• Horticultural practices and skills • Management knowledge • Tools and technologies	• Training of children by families and communities (decades) • Formal educational degrees and certifications (years to decades) • Publication of major new books, articles (years to decades) • County fairs (yearly) • Farmer field days, conventions (yearly)	• University/government agronomy research into best practices for germination, inoculation, soil fertility, maintenance, harvest, storage, etc. • Journalistic and ethnobotanical interviews, publications • Development of integrated pest control strategies • Engineering of specialized machinery for harvesting or other agronomic activities • Formal and informal centers for agricultural education and extension • Practitioner innovation associations
Cultural		
• Laws and policies • Educational practices, both formal and informal • Stories, recipes, and artworks • Values, attitudes, beliefs, and norms	• Election cycles (years) • New companies or corporate leadership (years to decades) • Social movements (years to decades)	• NGOs hosting citizen science projects • National and provincial government dietary recommendations, public health campaigns • Cultural production and circulation through art, literature, cookbooks, digital and social media • Commercial and NGO advocacy, lobbying, and marketing communications • Commodity associations funding research and marketing • Regional cuisine change through immigration, cultural diffusion, urbanization, and travel • Political parties propose policy and spending, seeking votes, contributions from special interests

Breeding for Vigor and Broad Adaptation

All the world's major crops are, almost by definition, grown far beyond the environment of their origin. They have adapted to many soil types, climates and daylengths. To provide a positive return on the research investment and to attract sufficient cultural notice as to become part of human cuisine, *de novo* domesticates will arguably need to be broadly adapted. Candidate species should be evaluated in multiple environments to identify those capable of broad adaptation or those with genetic variation for yield stability in multiple years and environments. Evaluation of breeding lines in multiple environments permitted the development of broadly adapted wheat varieties. Norman Borlaug advanced only varieties "that withstood the rigors of both environments," and considered broad adaptation to have been one of the three greatest successes of wheat breeding in the late twentieth-century (Borlaug, 1983). With increasing climate instability, new crops must be able to withstand different rigors even in a single location, within or across years.

Breeding for Biotic and Abiotic Stress Tolerance

Winter hardiness (for perennials and winter annuals) and disease/insect resistance are specific adaptation traits that DG editing will not improve for undomesticated plants. Editing/engineering of resistance (R) genes could help, of course, although R genes are more likely to be species-specific "orphan" genes than DGs (Woodhouse and Hufford, 2019) and therefore researchers may rarely be able to take advantage of the gene discovery work already accomplished for existing crops. Crop pathogens reduce harvests by about 20% globally (Bebber and Gurr, 2015). Wild plants are not immune to pathogens; fungicide application increased plant biomass by 31% in a native North American grassland (Mitchell, 2003). The cultivation of wild plants could create conditions that favor the spread of some of their pests and pathogens, for example, by increasing the number, density, size and uniformity of host plants (Chen et al., 2018). Breeding for increased biotic stress resistance may therefore be necessary for new crop domestication and plant pathologists recommend that breeders include selection for quantitative and

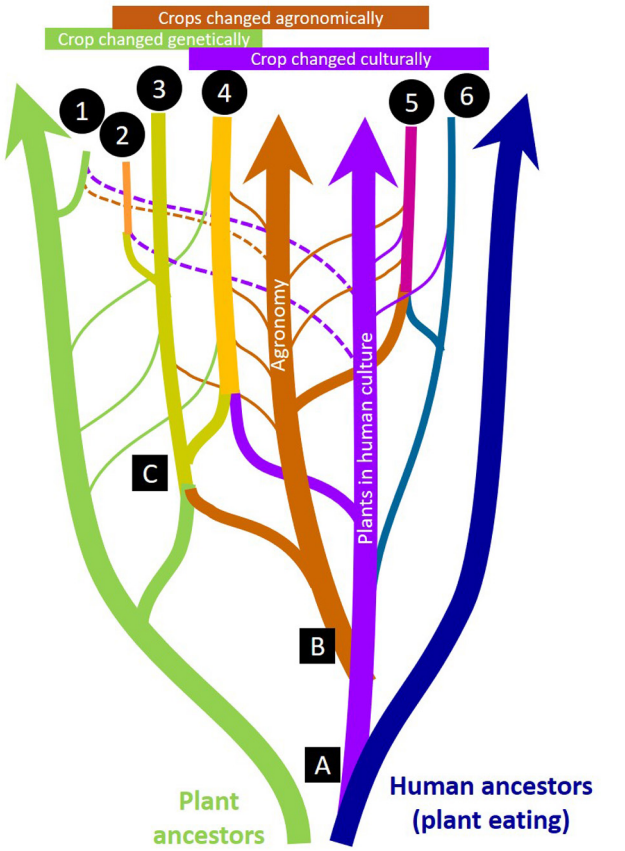


FIGURE 2 | Genetic, agronomic, and cultural changes are co-evolutionarily entangled in crop domestication. A reticulate biological/cultural "evolutionary tree" is shown in cartoon form (not to scale) to illustrate how "lineages" of plants, agronomic technology/traditions, and inherited human ethnobotanical culture "hybridize" to create different kinds of crops: (1) *De novo* incipient new crops from wild plants by breeding or gene editing (e.g., silphium), (2) New food crops developed from non-food crops by breeding or gene editing (e.g., intermediate wheatgrass), (3) Forage, fiber, and energy crops (e.g., alfalfa), (4) Major world food crops (e.g., maize, rice), (5) Cultivated crops with little genetic change (e.g., cranberry), (6) Culturally important wild-crafted food and medicinal plants (e.g., 'akkoub). Major evolutionary innovations are numbered: (1) Cultural references to plants (names, stories, recipes, etc.) appear, (2) Agronomic practices appear, (3) Genetically distinct cultigens appear. Narrow lines show "horizontal" influences such as accidental or deliberate gene flow between cultigens and wild relatives (narrow green lines) or changing cultural uses of harvested plants (narrow purple lines) or the influence of new agronomic practices and technologies (narrow brown lines). Dotted narrow lines indicate a very recent or emerging influence.

broad-spectrum resistance traits (Bebber and Gurr, 2015) in addition to R gene mediated resistance.

Breeding for Yield and Quality

Traits controlled by many genes are poor candidates for genome editing and improvement may require more conventional cycles of breeding for the foreseeable future. While some important agronomic traits are monogenic, domestication traits in some crops are polygenic (Hämälä et al., 2019). Flavor, ripening, post-harvest physiology, and nutritional profile are important traits for food crops and unlikely to be completely satisfactory in "crops" undifferentiated from wild ancestors except at a few DG loci and are likely to be genetically complex traits. For example, Zhang et al. (2015) found 28 volatile chemicals to be involved in tomato flavor/aroma and associated with 125 markers. Seed set (or seed fertility), the percentage of florets producing a viable seed, has been found to be quite low in some wild plants both in their wild habitats and under cultivation (Cornelius, 1950). However, recurrent phenotypic selection can improve this trait (Marshall and Wilkins, 2003), which would improve the yield and breedability of crops developed by gene editing. The conventional modeling of yield as a genetically complex trait with many small-effect loci continues to be supported (e.g., Martínez et al., 2016). Evidence that genetic correlations between agronomic traits constrained and slowed maize domestication and that genetic constraints increased during domestication (Yang et al., 2019) suggests that – for some species at least – it will be hard to identify many agronomic targets for genetic modification that have large, positive, and independent effects.

Value of Recognizing Broad Genetic Changes Involved in Domestication

The previous section shows that recognizing genetic change in the broader context of ongoing domestication can lead DG-based project teams to invest appropriate time, resources, and strategic planning to domestication efforts in the form of careful wild candidate selection and screening, collection, characterization and pre-adaptation of germplasm. Funding agencies should recognize the need to support multi-location, multi-trait breeding of new domesticates even after successful DG improvement.

One practical recommendation for accelerating *de novo* domestication of food crops and bypassing some stages described above is to focus on species that have been domesticated as forage, ornamental, medicinal, bioenergy or timber crops. These are species that have responded to agricultural conditions and human management and may have been further selected by breeders for adaptation to managed environments and tolerance of multiple soils and climates. Specific agronomic recommendations and laboratory protocols may be available. Last, but not least, the plant may already be culturally familiar to farmers and policy makers. Presumably, if it has become a successful forage or other kind of crop, this species has been found by humans to be pleasant to work with, non-invasive, and aesthetically compatible with rural landscapes. Forage crops do not require most of the traits conferred by DGs, yet traditional breeding has improved forage yield and usefulness to farmers.

In contrast, even locally adapted native species have been difficult to use as planted forages without some domestication (Morrison, 2016). Despite a wealth of native grass species, not a single native species is used widely enough to merit keeping statistics on the acreage of seed production in the United States (National Agricultural Statistics Service, 2019). Interestingly, the National Agricultural Statistics service primarily differentiates hay acreage as "tame," "wild," or "alfalfa" (National Agricultural Statistics Service, 2019), supporting our view that few forage

species have achieved deep cultural value (**Figure 2**). Poor seedling establishment and vigor is an example of a trait that constrains adoption of forages (Vogel, 2000). DGs leading to increased seed size could improve these traits but about half of the variation in seedling vigor is unrelated to seed size (Vogel, 2000). The value of recurrent selection for the performance of both native and introduced forages illustrates the fact that the micro-environments created by the human-crop mutualism are novel (DeHaan et al., 2007; Purugganan, 2019).

Thinopyrum intermedium (Host) Barkworth & D.R. Dewey (intermediate wheatgrass, hereafter) is being developed as a new perennial cereal grain marketed under the trade name Kernza (DeHaan and Ismail, 2017). This new crop is an example of a candidate for DG editing (DeHaan et al., 2020) that builds upon an established forage with extensive germplasm enhancement and multitrait breeding. Intermediate wheatgrass was collected from the wild in the former USSR in the 1930s (Hanson, 1959). Forage varieties were developed in the United States and Canada from 1945 onward by selection from the wild germplasm for vigor and seed fertility (Vogel and Hendrickson, 2019). Recurrent mass selection with controlled interpollination of selected genotypes was used to increase seed yield in the 1950s and other programs emphasized lodging resistance, plant health, productivity, forage quality and broad adaptation (Vogel et al., 1993). By 1988, 500 tons per year of seed was being produced for sale in Saskatchewan alone (Saskatchewan forage seed development commission, 1998). In response to the suggestion that new perennial grains could conserve soil, the Rodale Institute began evaluating wild perennial grasses as candidates in 1983 (Wagoner and Schaeffer, 1990; Wagoner and Schauer, 1990). After scoring ten agronomic traits, including seed size, seed shattering, and stem lodging, intermediate wheatgrass was identified as the top candidate and two cycles of recurrent selection for improved yield and harvestability were performed at the USDA-NRCS Big Flat Plant Materials Center (Cox et al., 2002). Forage varieties were included along with wild accessions in the starting population from which the Rodale Institute began its selections (Wagoner and Schauer, 1990). In the early 2000s, The Land Institute revived the intermediate wheatgrass domestication program and has since performed nine cycles of recurrent selection; three cycles have been completed at the University of Minnesota (Crain et al., 2020). Recent cycles of selection use genomic selection to accelerate genetic gains for domestication traits and cereal grain yield (Zhang et al., 2016).

Intermediate wheatgrass is a challenging species for genome editing. No transformation protocol is available, the species is highly heterozogous, the genome contains more than 11 million base pairs, and the genome is a complex allohexaploid (DeHaan et al., 2020). Domestication phenotypes due to knockouts in this species may require obtaining plants with all six alleles in the edited non-functional form. Such an effort involves not only completing gene edits but also possibly breeding them to homozygosity. Clearly, intermediate wheatgrass would not be a candidate for DG editing if the crop had not previously undergone more than three decades of breeding and agronomic work with the objective of obtaining a successful perennial grain crop. In essence, DG editing of this species has potential to dramatically improve its functionality as a grain crop by providing breakthrough improvements in one or two domestication traits because the species has already been the target of numerous breeding, agronomy, and utilization efforts (Tyl et al., 2020).

Genetic studies with intermediate wheatgrass thus far have revealed both the activity of known domestication genes from other grain crops in this new species, but also the potential challenges to utilizing such genes. Larson et al. (2019) identified 42 candidate genes in intermediate wheatgrass that can influence traits relevant to domestication. These genes are potential targets for selection or editing. However, even for the traits with highest heritability, such as free threshing ability, 16 significant markers associated with the trait were detected. This indicates that a large number of genomic regions were found to influence the trait, rather than a single high-impact locus. Furthermore, the full marker set only explained about 25% of variation in some environments, and marker effects depended upon environment in many cases. Therefore, even for simple domestication traits, there can be substantial interaction between genes and environment, and influence of a particular allele may depend on genetic "background," or interaction with the rest of the genome. These realities could greatly increase the complexity of genome editing approaches; perhaps particular allele forms will be necessary to override gene by environment interaction, or careful breeding will be necessary to preserve a complementary background genome that will not override the novel domestication allele introduced through editing.

AGRONOMIC CHANGE

The increased value of a domesticated plant to humans, compared with its wild ancestor, depends on its altered phenotype (e.g., large seeds, high seed yield, and reduced branching). Variation in plant phenotypes result from both differences in environment (E), genes (G) and the interaction of G × E. From the perspective of growers, the exact contribution of G or E or G × E is irrelevant; the final phenotype is relevant. In addition to G × E, agronomic management (M) changes highly relevant aspects of the environment (Hatfield and Walthall, 2015; Wang et al., 2019). There is variation in management and this variation can be transmitted across human generations, constituting a necessary component of domestication processes.

Agronomic change occurs when farmers learn to produce specific abiotic and biotic environmental changes that enhance plant productivity and harvestability (Altieri, 2004). Some of these changes are more-or-less permanent, whereas new seed must be planted each year. Perhaps this is one reason why we more often think about the genetic than the agronomic contributions to domestication. We can grow crops side by side with ancestors whereas our ancestors modified vast landscapes to increase and stabilize crop yield. They drained wetlands, terraced mountains, cleared forests, and killed off large wild herbivores and it is difficult to experimentally replicate these changes at research stations.

Genetic change in plant domestication is interdependent with the actions taken by early farmers (intentionally or unintentionally) to prepare an environment that was consistently favorable for plants they preferred (Doebley et al., 2006; Clement et al., 2015; Altman and Mesoudi, 2019; Edwards et al., 2019; Mueller, 2019). The iterative development of plants makes their final size, shape and fecundity extremely plastic compared with animals. The influence of environment upon the realized yield of all crops – regardless of their yield potential in optimal environments – is so intuitively understood by any gardener or farmer, so central to the purpose of entire agronomy departments, and so inherently complicating to plant breeding that perhaps it is easy to overlook when considering *de novo* domestication. As an example, weed control alone doubles soybean and corn yields in the corn/soy region of North America (Soltani et al., 2016, 2017) and this estimate was made for crops otherwise managed according to modern best-practices. Without high quality seed, uniform seed placement and row spacing, well-managed fertility, etc., weeds would likely have been more competitive.

Agronomic change encompasses a wide range of ecosystem alterations, from the straightforward replacement of native vegetation with crop species, to dramatic interventions in a wide range of factors that affect crop productivity, including water, temperature, nutrient, weed, insect, and pathogen management. Before the fossil fuel era, agronomic changes were primarily "ecological" in that farmers manipulated key ecosystem patterns and processes to favor productivity per unit human or animal labor (Smil, 2018). Examples of such practices included the use of crop rotations, the integration of legumes, use of periodic flooding, and as mentioned, elimination of competing vegetation with weeding (Gliessman, 2015). The discovery and rapid increase in availability of fossil fuels profoundly relaxed the energetic constraints on possible agronomic practices. People have successfully employed fossil-fuel based strategies to relax or eliminate almost every limiting factor to crop productivity, including synthetic nitrogen, biocides, plastic row covers, pumping and transportation of water for irrigation, and more (Pimentel and Pimentel, 2007). As we once again move toward cultural expectations of reduced fossil fuel dependence, we see agronomic practices shifting as well from approaches that maximize fossil fuel-dependent agronomic practices in order to maximize yields, toward agronomic solutions that maximize ecological intensification in order to optimize yields (Crews et al., 2016).

Value of Recognizing Agronomic Change Involved in Domestication

Wild species being introduced into an agricultural environment are likely to exhibit dramatic plasticity as they are released from some forms of herbivory and competition from other plant species. However, not all species or genotypes are equally plastic; some may require additional environmental modification to thrive. Since farmers are in the business of modifying the environment to benefit plants, it makes sense to screen *de novo* domestication candidates in conditions characteristic of the target agricultural system, whether that be an irrigated paddy or an orchard, a high-input cash plantation or a low-input farm. Plants without the plasticity to adapt to the target environment are poor candidates for domestication.

Learning the ideal conditions for each candidate will require either rigorous agronomy or trial-and-error experimentation by farmers. Time and funding to permit adequate characterization of crop candidate response to variations in soil texture, fertility, water availability, cold stratification and vernalization as well as other ecological factors should be included in any *de novo* domestication project.

For wild plants never previously used as crops, the research investment in agronomic drivers of domestication may be substantial and time-consuming – just as with the research investment into genetic change. Recognizing this may prompt *de novo* domestication research teams to attend to agronomic change and introduce a completely wild plant species as a crop for forage, erosion-control, pollinator habitat, or niche high-value use of seeds or fruits (medicinal, specialty vegetable or flavoring, cosmetics) prior to investing in genetic changes necessary for a major food crop.

Intermediate wheatgrass domesticators in the 1980s benefitted from planting practices, weed control, and fertilization that are recommended to farmers and researchers (Wills et al., 1998). Some of these recommendations may turn out to be counterproductive for seed producers vs. their intended audience of forage producers, but valuable insights for driving domestication via managing a new perennial cereal crop were available from forage grass seed producers (Horton et al., 1990; Saskatchewan forage seed development commission, 1998; Hybner and Jacobs, 2012; Kruger, 2015).

CULTURAL CHANGE

From an evolutionary perspective, domestication is an example of co-evolution between mutualists. Genetic changes in human partners during this co-evolutionary process are modest, though documented (Altman and Mesoudi, 2019). However, human culture can evolve by accumulating complex behaviors and transmitting them to successive generations (Altman and Mesoudi, 2019). Some of these behaviors relate to the knowledge, practices and tools used in tending crops, i.e., resulting in agronomic change as described above.

Beyond applied horticultural tools and knowledge, other cultural changes are required to enable and sustain a deepening, long-term human relationship with a new crop. Past human cultural evolution enabled initial domestication and agricultural niche construction (see review in Laland (2017). Now, cultural change describes the human community valuation of a species in social, economic, and culinary terms that may be necessary for *de novo* domestication processes to begin, continue, and succeed in bringing a new crop into widespread cultivation and use.

Human management of plants includes a variety of activities – gathering, tolerating, enhancing, and protecting *in situ* as

well as *ex situ* sowing and transplanting – which suggests multiple possible routes for domestication processes (Casas et al., 1996; Lins Neto et al., 2014). Sociocultural factors shape these human choices about management and thus, for example, which plant species are engaged in incipient landscape domestication (Betancurt et al., 2017). Diverse cultural knowledge and methods of enculturation also shape how innovation, including domestication processes and products, are pursued and maintained across generations (Larson et al., 2014) with both negative and positive social and ecological implications (Scott, 2017; Swanson et al., 2018), such as the development of cooperative behavioral norms (Zeder, 2016).

The ongoing co-evolution that defines the domestication relationship must be sustained, on the human side, by society. Some degree of cultural change in response to a potential new crop, such as at the community scale, may be required in order to influence human resource allocation (Zeder, 2015) and successfully build investments to pursue agronomic and/or genetic change in terms of breeding and management experimentation, training of farmers, and the development of food processing, storage and distribution techniques.

Vaccinium macrocarpum (cranberry, hereafter) is an example of successful, ongoing crop domestication accomplished primarily through human cultural and agronomic change with minimal plant genetic change so far. A native North American fruit crop long wild-harvested by indigenous peoples and then adopted by European settler colonists, cranberries were valued as winter sources of calories and nutrients, especially vitamin C. They were so important that laws were passed banning the collection of the wild fruits prior to a certain date each fall (Klingbeil and Rawson, 1975). Though formal documentation is lacking, we consider it likely that indigenous peoples managed wild cranberry stands for increased yield (Doolittle and Mabry, 2006).

Agronomic change in cranberry domestication was advanced when transplanting of wild vines began in 1816 (Eck, 1990) and as cultivation spread with increasing economic returns by mid-century (Peltier, 1970). Over the next 100 years, the "big four" cultivars of selected clones from wild genotypes were involved in increasing agronomic and cultural change, including: managing cranberry hydrology using dikes and canals, plant propagation through sanding, protection from biotic (insects) and abiotic pressures (cold) by flooding, and harvest using specially designed cranberry rakes, barrels, crates, and eventually machines (Cole and Gifford, 2009); and organizing into cooperatives (i.e., Ocean Spray), who in turn developed new cranberry products (juices, sauces), creating cranberry growers' associations, establishing cranberry experimental stations, planning festivals to celebrate cranberry flowering and harvest, and establishing traditions like crowning an annual cranberry queen (Peltier, 1970; Eck, 1990; Schlautman, 2016).

Nearly all the agronomic and most of the cultural change widely recognized and practiced were accomplished by the time genetic change (artificial selection) produced any sort of germplasm that was adopted by cranberry growers. The first cranberry breeding program was initiated in 1929 after growers and scientists had been unable to address the problem of cranberry false blossom disease through agronomic domestication (Eck, 1990). In 1950 the first cranberry cultivars were released after one cycle of selection from the 1929 program. A few of these varieties were eventually adopted because of their increased productivity and their perceived improved yield stability, especially cultivar "Stevens." Most cranberry breeding programs were abandoned shortly after the 1950 releases. Rutgers University started a program in the late 1980s, the University of Wisconsin started a small program in the 1990s, and a Wisconsin grower independently started a private breeding program in the 1990s (Schlautman, 2016). The next set of varieties with improved productivity were released beginning in the early 2000s (Fajardo et al., 2013), and those varieties are being adopted and planted (Gallardo et al., 2018). However, cranberry growers still plant older varieties and even continue to plant and harvest beds of cranberry clones selected from the wild in the early 1800s (Vorsa and Zalapa, 2019).

DG traits have yet to be identified in cranberry. A recent survey revealed that cranberry growers perceived fruit quality traits and, specifically, fruit firmness as needing the most attention from breeders. This is likely in direct response to premiums being paid by cranberry processors for fruits that meet standards for sweetened dried cranberry processing, while prices for cranberries used for juice concentrates (i.e., normal to low fruit quality) have decreased. Of secondary importance were traits related to disease resistance (i.e., fruit rot), abiotic plant stresses, and insects. Interestingly, yield and productivity were not assigned as important priority, nor were fruit quality traits related to shelf-life, flavor, and sweetness (Gallardo et al., 2018). The development of a "sweet" cranberry could be the DG and the genetic change the cranberry industry needs to propel the emergence of a new cultural change via the creation of a cranberry fresh fruit market and consumer demand.

As this extended consideration of cranberry shows, integration of a *de novo* domesticate into a culture's cuisine and food rituals and norms may be needed to continue the domestication process and sustain research investments in terms of agronomic and/or genetic change. While the exact size or appearance of newly domesticated food crops is difficult to predict prior to domestication, investigations into flavor, digestibility, toxicity and nutritional profile can help identify candidates or reveal additional biochemical pathways that need to be targeted for genetic change. Genetic and agronomic change remains central to work on intermediate wheatgrass, as described earlier. But a small amount of cultural change was seeded early on when the Rodale Institute began testing the nutritional and cooking properties of this new grain (Becker et al., 1991) and marketed it as "wild triga" (Wills et al., 1998). Likewise, The Land Institute and collaborators also conducted food science research (Marti et al., 2016; Banjade et al., 2019; Tyl et al., 2020). In line with previous efforts that led to the name wild triga, current marketers have expressed the need for a brief and memorable name. The Kernza trade name was initiated for this purpose, and Kernza branded grain has now been used in several restaurants, food products and beers. Customer demand for Kernza grain-containing products could, in the future, motivate corporate

TABLE 2 | Uncultivated plants commonly used in Palestinian cuisine.

Vernacular name(s)	Scientific name	Family	Food type
za'tar balat, duqah adas	*Clinopodium serpyllifolium (M. Bieb.) Kuntze*	Lamiaceae	Tisane, seasoning
za'atar	*Origanum syriacum* L.	Lamiaceae	Herb, seasoning, tisane
za'tar rumi, za'tar beid, za'itman	*Satureja thymbra* L.	Lamiaceae	Herb, seasoning, tisane
za'tar farisi	*Thymbra capitata* (L.) *Cav.*	Lamiaceae	Herb, seasoning, tisane
za'tar sabbal, sabbaleh	*Thymbra spicata* L.	Lamiaceae	Herb, seasoning, tisane
maramiyyeh	*Salvia fruticosa Mill.*	Lamiaceae	Herb, seasoning, tisane
Waraq lisan, lisseneh	*Salvia hierosolymitana Boiss.*	Lamiaceae	Cooking green
humeymsa	*Rumex cyprius Murb.*	Polygonaceae	Salad green
za'rur	*Crataegus azarolus* L.	Rosaceae	Fruit
qayqab	*Arbutus andrachne* L.	Ericaceae	Fruit
seyba'a, sneyba'a	*Lotus palaestinus Blatt.*	Fabaceae	Fresh vegetable, pulse
Burreid	*Pisum fulvum Sibth. & Sm.*	Fabaceae	Fresh vegetable, pulse
Khobs al ra'i, qurus sitti	*Medicago orbicularis* (L.) *Bartal*	Fabaceae	Fresh vegetable
kharrub	*Ceratonia siliqua* L.	Fabaceae	Sweetener
'akkoub	*Gundelia tournefortii* L.	Asteraceae	Cooking vegetable
helyoun	*Asparagus aphyllus* L.	Asparagaceae	Cooking vegetable
za'matot, qarn al ghazal	*Cyclamen persicum Mill.*	Primulaceae	Cooking vegetable

Tesdell, O. (editor). 2018 Palestinian Wild Food Plants.

and governmental investment in further genetic and agronomic components of domestication.

Value of Recognizing Cultural Change Involved in Domestication

As with agronomic change, attention to cultural change invites essential research and engagement with the humans who will grow, eat, trade, and continue to develop new crops. Without social engagement and support, appropriate tools and agronomic practices, or broadly adapted varieties, new crops will remain local and/or niche crops. Without a loyal constituency advocating for sustained research investment, they could easily be abandoned when new diseases appear or as the climate or the political and economic landscape changes. To produce the food security and agroecological benefits that new crops could provide, new crop acreage must eventually be substantial.

In order to enable ongoing social and economic support and later adoption, researchers should identify *de novo* domestication candidates that have the best chance of producing a positive return on the investments needed to domesticate and market them. Cultural and culinary uses of potential candidate plants may be documented in ethnobotany (Schlautman et al., 2018). From ethical, legal, and political standpoints as well as from a practical perspective, human relationships with candidate plants are relevant considerations in decisions about plant species' suitability for domestication. Acknowledging the reality of cultural change affirms the ongoing importance of *in situ* conservation of landraces and of wild and weedy species on farms *with farmers* for new crop domestication, as human social factors are components of these complex systems (Casas et al., 2007; Bellon et al., 2017).

Recognition of established cultural components relevant to domestication can help identify promising opportunities for new crops. In Palestine, site of some of the earliest plant and animal domestication, wild plant gathering has continued to support human communities for millennia (Ali-Shtayeh et al., 2008; Eghbarieh, 2017; Research Collaborative, 2018). Wild food plants (**Table 2**) are important for traditional Palestinian cuisine and for sustenance during climatic and political crises (Tesdell et al., 2019). Palestinian people already value these plants in social, economic, and culinary terms (Tukan et al., 1998; Marouf et al., 2015). Recognition that these relevant cultural drivers for domestication may be in place is leading local groups to begin to collect seeds of edible wild plants and to explore ways to encourage more widespread cultivation beyond recent existing cultivation and transplanting practices (Doolittle and Mabry, 2006).

Several wild food plants could be used as grain legumes or oilseed grains, in addition to their current use as vegetables (Tesdell et al., 2020). Examples include *Lotus palaestinus* Blatt., *Pisum fulvum* Sibth. & Sm., and *Gundelia tournefortii* L. ('akkoub, hereafter). 'Akkoub, a wildcrafted and increasingly cultivated plant, offers opportunities as a perennial vegetable and possible oilseed among other uses. In addition, many crops remain widely foraged in Palestine and several surrounding countries. **Table 2** offers a few prominent examples of Palestinian foraged plants of a list that could include more than 200 documented wild food plants to date. Their Arabic local names vary by region, but these wild plants persist as highly culturally important foods. Existing communities' cultural valuation of this candidate plant could help enable and support pursuit of agronomic and genetic changes involved in *de novo* domestication.

In other *de novo* domestication efforts, it may be necessary to catalyze cultural change. Public interest, motivation, and support for *de novo* domestication may not be gained through simple, one-time, or purely rational means. Instead, building and sustaining the cultural change needed to support new crops

may involve navigating a complex web of ongoing processes that include the affective, educational, ethical, legal, narrative, political, social, and other dimensions of human experiences and structures. Like scientific legitimacy more generally, agroecological legitimacy may be accomplished by meeting "credibility tests" and wisely engaging in this complex web of social life (De Wit and Iles, 2016). Honest acknowledgment and critical examination of value influences across the scientific research and development process, from inquiry to application, will likely be necessary for successful domestication of new crops – especially via gene editing (Elliott, 2017).

Participatory research and citizen science methods (Mendez et al., 2013; Ryan et al., 2018) – such as the involvement of many, geographically distributed farmers at an early stage of domestication – could help build initial communities that inform broader cultural support for the plant species. Introducing a potential food crop as a forage or specialty crop first may have the advantage of greatly increasing the number of farmers and other social groups involved and, at the same time, increasing the number of agricultural environments in which the species can be evaluated and selected.

One example of public support for new crop valuation is the Forever Green Initiative, developed at the University of Minnesota based on the principle that universities must engage with multiple stakeholders to realize their mission. When multiple stakeholders are considered, the broader societal impacts of agriculture must be considered. When agriculture threatens wildlife habitat or the provision of clean water or becomes a source of greenhouse gases, a broad base of stakeholders has the opportunity to support research directed at finding solutions to these challenges. The Forever Green Initiative has undertaken new domestication programs that aim to develop crops that will enable agricultural production to provide broad public goods. These efforts are gaining widespread interest and support (Runck et al., 2013; Anderson, 2014; Kleinhuizen, 2017; Haspel, 2019).

Documenting the ability of domestication candidates to provide regulating, cultural, and supporting ecosystem services (such as maintenance of soil and water quality, wildlife habitat, pollinator resources) could help build the case for social and economic valuation of new crops, including governmental support for continued research, farm payments, or the ability to charge consumers premium prices (Leakey and Asaah, 2013; Runck et al., 2014; Smýkal et al., 2018). As Kernza perennial grain has entered the market through the introduction of a few specialty products, abundant opportunities have emerged for storytelling around the new grain and its benefits. Messages about the ecological benefits of the grain have been placed directly on packaging such as beer cans and cereal boxes. Niche products have created a crowd funding project (Pierce, 2019) and generated numerous news reports (Bland, 2016; Garfield, 2016; Ostrander, 2017). Although this grain needs much additional research to enable profitable production at larger scale, niche specialty marketing boosts awareness and support, potentially inducing sustained funding.

In the case of *de novo* domestication that utilizes gene editing, special consideration should be given at an early stage to barriers to cultural adoption. Technological improvements in traditional crops have sometimes failed for cultural reasons. The reaction against transgenic foods leading to their non-adoption in Europe and elsewhere is an excellent example of failure to achieve cultural modification in support of a genetic approach (Kloppenburg, 2010). Social and educational researchers should investigate public perception and knowledge of the potential future domesticates via gene editing and consider strategies to foster informed public engagement in social decisions related to new crop domestication (Østerberg et al., 2017).

INTEGRATING DRIVERS OF CHANGE IN DOMESTICATION

In contrast to the idea that gene editing can nearly instantly and/or inexpensively produce domestication, the established literature and our examples of intermediate wheatgrass, ʿakkoub, and cranberry demonstrate the multiple components involved in long-term domestication. Genetic, agronomic, and cultural changes interact over time to drive domestication processes. To illustrate lessons learned about how these changes are entangled in complex ways, we turn to a final personal narrative example of a current *de novo* domestication process. This narrative is not intended to provide recommendations for other new crop candidates, as each will enter the domestication "pipeline" (DeHaan et al., 2016) at a different stage and with different liabilities.

The Land Institute (Salina, KS, United States) has been domesticating *Silphium integrifolium* Michx. (silphium, hereafter) as a perennial oilseed crop since the early 2000s (Van Tassel et al., 2017). Human cultural valuation and ecological knowledge informed the decision to begin domestication, as local silphium populations were observed to have large, good-tasting seeds and to perform well during seasonal or year-long droughts. The latter observation was confirmed anecdotally by botanists in Texas (James Manhart, personal communication) and during the droughts of the 1930s by ecologist John Weaver (Weaver et al., 1935). Seeds from several wild populations in central Kansas were collected and grown in observation plots in Salina and allowed to intermate to produce seeds. The experimental plants grew much larger than plants in nearby prairies, suggesting that silphium is well-adapted to agricultural conditions and amenable to agronomic change.

However, domestication efforts first focused on driving genetic change. The disk florets of silphium are staminate, limiting the number of seeds per head to 15–25 (equal to the number of ray florets). Therefore, the first breeding target in 2004 was to generate and identify a mutation causing bisexual disk florets or to use recurrent selection to increase the number of ray florets (Van Tassel et al., 2014). Mutagenesis was attempted but no mutants with the phenotype of interest were recovered. However, several cycles of recurrent visual selection on ligule number (ray floret number) were successful in feminizing the heads of silphium to a point where some individuals were almost male-sterile. In 2012, parents for new cycles of selection were made for partial feminization plus increased achene (hereafter "seed")

size and head diameter. Recurrent selection for non-dormant seed was attempted and reduced shattering was a breeding goal. In short, traits comprising the classic domestication syndrome were targeted, albeit mostly with the expectation of finding highly polygenic control of these traits as reported in sunflower (Burke et al., 2002). These included a greater number of larger heads and seeds, non-dormant seeds, and higher seed oil content.

In 2014 and every year thereafter, plants in research plots in Salina Kansas were infected with rust (*Puccinia silphii*) and other pathogens that have yet to be conclusively identified. *Eucosma giganteana*, a specialist moth whose larvae are head and crown parasites of several *Silphium* species, also appeared and became a serious pest with colonization of heads approaching 100% in many situations. These pests and pathogens may be more common in more humid environments and did not become common in Salina until we had created favorable microclimates (lush, dense stands) and large populations of host plants (Turner et al., 2018; Vilela et al., 2018).

The pest and pathogen pressures have made selection for yield or domestication traits difficult in Salina. In genetic terms, improved strains and breeding populations at The Land Institute have little variation for resistance to several pathogens and pests and although these populations indeed have higher seed yield than wild silphium, increased yield is accompanied by increased plant size. Standard harvesting machinery does not work as well with taller plants with stouter stalks, and increased biomass almost certainly implies greater water use. In agronomic terms, attempts to scale up plot sizes by direct seeding have had highly variable success and even transplanting in new fields has occasionally failed for unknown reasons. Together these limitations have made it hard to obtain quantities of seed needed for food or feed science research – i.e., as a strategy to drive cultural valuation.

Renewed collection of wild silphium from throughout its natural range has revealed individual plants or populations that appear to have greater resistance to a number of pests or pathogens; some also appear to be shorter and have more slender stems that may be more compatible with combine harvesters. However, these ecotypes from south or east of Kansas have much smaller seeds and smaller, less feminized seeds and heads.

Looking back, we see that the western (Kansas) ecotype attracted us (humans) with its unusually large heads and seeds. However, the initial decision to focus exclusively on genetic change for classic domestication traits, especially seed and head size, at this point distracted us from first diversifying the genetic base of the breeding populations (which would have meant reducing average seed and head size) or discovering specialized agronomic management practices as would have been necessary to first achieve broad adaptation or cultural adoption of silphium as a forage or other specialty crop.

In the United States, this kind of work has been done for many new crops, beginning with the USDA's Section of Seed and Plant Introduction in 1898 and in partnership with germplasm "banks" and agricultural universities (Hyland, 1977). However, many native plants, including *Silphium integrifolium* have not received such attention and perhaps we should have recognized the need to perform this work ourselves and spent several years collecting more diverse germplasm, increasing it, and testing it in multiple environment to identify vigorous, resistant, and stable lines.

Another alternative scenario would have been to prioritize agronomic physiology and horticultural research over genetic selection during the initial stage of domestication. Although average seed yield in Kansas was about 300 kg/ha (prior to pest outbreaks), 1600 kg/ha was achieved in one year and one location in Minnesota (Schiffner, 2018). Clearly, the species already has the genetic potential to yield well. We might have been better served spending the time to understand what environmental conditions contribute to the "yield gap" between this potential and average yield than rushing ahead with selection for increased yield potential.

Some domesticated plant phenotypes can be produced through careful management of wild germplasm. For example, surgical manipulation of plant size and branching pattern is an ancient art still actively practiced with many woody tree and vine crops. We have had some promising results trimming of silphium stalks to reduce stalk height and improve architecture, but the procedure is very sensitive to pruning height and timing. This is an excellent example of how humans develop detailed, species-specific horticultural practices, formally or informally, much of it by trial-and-error.

Furthermore, there may be some advantages in retaining developmental plasticity for plant height and branching pattern, relying on skilled farmer management to produce harvestable, seed-productive phenotypes. This is because silphium has nutritious, palatable foliage and left unpruned could be used as a forage crop. Being able to use management to switch between forage and oilseed production could allow farmers to respond to weather or market fluctuations. Deliberately alternating between allowing heads to produce seeds (oilseed) and chopping the biomass prior to seed set (forage) may interrupt the lifecycle of specialist insects.

Finding another use for silphium prior to pursuing genetic changes necessary for domestication as a perennial oilseed crop could allow us to expand the number of people interacting with this species and the number of environments in which it could be tested. As an expanding ornamental, pollinator crop, or forage crop, seed production practices would still need to be developed to make seed more available and affordable. This demand could stimulate agronomic innovation in crop management and agricultural engineering and begin the process of building traditions, networks, companies and other social structures needed for the long-term success of a new crop.

These lessons learned about the public interest and support that will be needed to sustain agronomic and genetic change in silphium have encouraged us to experiment with methods to investigate and initiate cultural change. Citizen science has been identified as a way to meet "grand challenges" in agriculture (Ryan et al., 2018). Providing people with access to potential new crops-in-process could help build interest and increase knowledge and awareness. Participants could become a publicly visible and informed network of collaborators. Learning outcomes gained through long-term citizen science projects related to new crop domestication could catalyze and fuel the cultural change needed to enable and sustain domestication.

TABLE 3 | Summary of examples of genetic, agronomic, and cultural change for plant domestication projects discussed in this article.

Example	Genetic	Agronomic	Cultural
'Akkoub (*Gundelia tournefortii*)	Very few accessions have been collected by gene banks	A few gardeners are beginning to cultivate it	Enthusiastically harvested from the wild for home use and for sale as a vegetable and medicinal; at risk of overharvesting and habitat loss; access to habitat increasingly restricted
Cranberry (*Vaccinium macrocarpum*)	Superior wild genotypes are still being clonally propagated; breeding programs are relatively new	Farmers have been constructing and irrigating bogs for about 100 years; advances in management and harvesting techniques increased yield and reduced costs	Indigenous peoples have long harvested fruit from wild bogs and have cultural knowledge about its culinary and medicinal value; colonists appropriated this knowledge and developed additional uses and markets
Intermediate wheatgrass (*Thinopyrum intermedium*)	Introduced as a forage and improved varieties released since the 1940s; recurrent selection for use as a cereal grain is ongoing	Agronomic research resulted in recommendations for forage management; grain production research is ongoing	Few people know this species but new breakfast cereal and beer products are being marketed under the Kernza name to increase awareness
Silphium (*Silphium integrifolium*)	Recurrent selection for seeds per head and other agronomic traits for the past 20 years	Not cultivated until breeding program began; a few small agronomic studies have been published	Indigenous knowledge and medicinal and cultural use, but no documented use as a food; civic science pilot project recently initiated
Currant tomato (*Solanum pimpinellifolium*)	CRISPR/Cas9 editing of 6 genes		

We began a silphium civic science pilot community in 2019. More than 40 people in 18 states joined. Like farmers who participate in on-farm research, participants in our silphium civic science community grow seedlings in backyards, gardens, farms, and public spaces and collect data over multiple seasons in a range of environments. They also share and reflect on their own personal experiences, interests, and questions about silphium's relationship with insects, soils, mammals, and humans – including its future uses by humans. Participants receive educational materials and can interact and learn with plants, with each other, and with scientists and educational researchers. While further research is needed, preliminary results indicate positive engagement.

Silphium is not currently a good candidate for gene editing because its large, intractable genome has limited genetic and genomic research in the species, because obvious monogenic DGs have not been identified in its closest crop relative, *H. annus* (Dowell et al., 2019), and because pests, diseases and soil fertility prevent silphium from achieving its existing seed yield potential in field conditions. Rather, silphium reaffirms the persistent need for science and society alike to understand the complexity of new food crop domestication. Successful *de novo* domestications will likely be the result of diverse, adaptive approaches to integrating genetic, agronomic, and cultural drivers of change over time.

CONCLUSION

In the preceding sections we have highlighted how *de novo* crop domestication, ranging from historical to contemporary, can be viewed as the varied and ongoing interaction of three factors of innovation: genetic, agronomic, and cultural. Gene editing approaches to domestication using DGs risk portraying a misunderstanding of *de novo* crop domestication as something humans do to wild plants in a simple one-time event, in a single gene or generation, using increasingly cheap technology, and drawing from just a few scientific disciplines. In contrast, *de novo* domestication literature and examples (**Table 3**) understand domestication as co-evolutionary interactions between plants and peoples that are complex, stretch across generations indefinitely, may require significant institutional and infrastructural investments, and can involve many disciplines and ways of knowing. This knowledge contextualizes gene editing approaches by remembering the additional work and integrated knowledge needed to accomplish goals that are shared across research communities: the accelerated domestication and widespread cultivation of a new generation of soil-conserving and climate-smart crops.

In the age of gene editing and at this moment of decision-making about domestication pathways, new food crop domestication might be viewed as analogous to a new public health campaign, in all its complexity. Narratives calling for new vaccines are exciting and easily communicated to private and public investors. But advances in vaccine molecules must be accompanied by sustained monitoring of virus evolution and transmission, development of improved vaccine delivery technology, support for clinics and vaccination campaigns and cultural work toward acceptance and use of vaccinations. Investment in novel recombinant or DNA vaccines could accelerate vaccine development or enable vaccination for new diseases, but this investment will not be worthwhile if vaccines are rejected on cultural or religious grounds, if vaccines are not reformulated to reflect pathogen evolution, or if vaccines can't be efficiently delivered to vulnerable populations. Similarly, recognizing the continuing importance of interacting genetic, agronomic, and cultural drivers to successful *de novo* domestication, we call on funding agencies, proposal reviewers and authors, and research communities (of new crop

domesticators, growers, and consumers) to value and support agroecology, agronomy, plant breeding, and participatory, interdisciplinary, and transdisciplinary approaches alongside genetics and genomics.

AUTHOR CONTRIBUTIONS

DV, AS, BS, OT, MR, LD, and TC: framework conceptualization, research, writing, and editing. DV: silphium and intermediate wheatgrass case studies. AS: silphium case study. BS: cranberry case study. OT: Palestinian wild foods case study. LD: intermediate wheatgrass case study.

FUNDING

This work was made possible through the charitable donations of multiple persons ("Friends of the Land") and organizations given to The Land Institute, Salina, KS, United States.

REFERENCES

Ali-Shtayeh, M. S., Jamous, R. M., Al-Shafie', J. H., Elgharabah, W. A., Kherfan, F. A., Qarariah, K. H., et al. (2008). Traditional knowledge of wild edible plants used in Palestine (Northern West Bank): a comparative study. *J. Ethnobiol. Ethnomed.* 4:13. doi: 10.1186/1746-4269-4-13

Allaby, R. G., Stevens, C., Lucas, L., Maeda, O., and Fuller, D. Q. (2017). Geographic mosaics and changing rates of cereal domestication. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 372:20160429. doi: 10.1098/rstb.2016.0429

Altieri, M. A. (2004). Linking ecologists and traditional farmers in the search for sustainable agriculture. *Front. Ecol. Environ.* 2, 35–42. doi: 10.2307/3868293

Altman, A., and Mesoudi, A. (2019). Understanding agriculture within the frameworks of cumulative cultural evolution, gene-culture co-evolution, and cultural niche construction. *Hum. Ecol.* 47, 483–497. doi: 10.1007/s10745-019-00090-y

Anderson, D. (2014). *U Professor Working to Save State's Pheasants and Ducks. Star Trib*. Available online at: https://www.startribune.com/u-professor-working-to-save-state-s-pheasants-and-ducks/275382751/ (accessed April 13, 2020).

Anderson, E. (2005). *Plants, Man and Life*. Mineola, NY: Dover Publications, Inc.

Banjade, J. D., Gajadeera, C., Tyl, C. E., Ismail, B. P., and Schoenfuss, T. C. (2019). Evaluation of dough conditioners and bran refinement on functional properties of intermediate wheatgrass (Thinopyrum intermedium). *J. Cereal Sci.* 86, 26–32. doi: 10.1016/j.jcs.2019.01.001

Bebber, D. P., and Gurr, S. J. (2015). Crop-destroying fungal and oomycete pathogens challenge food security. *Fungal Genet. Biol.* 74, 62–64. doi: 10.1016/j.fgb.2014.10.012

Becker, R., Wagoner, P., Hanners, G. D., and Saunders, R. M. (1991). Compositional, nutritional and functional evaluation of intermediate wheatgrass (Thinopyrum intermedium). *J. Food Process. Preserv.* 15, 63–77. doi: 10.1111/j.1745-4549.1991.tb00154.x

Bellon, M. R., Dulloo, E., Sardos, J., Thormann, I., and Burdon, J. J. (2017). In situ conservation—harnessing natural and human-derived evolutionary forces to ensure future crop adaptation. *Evol. Appl.* 10, 965–977. doi: 10.1111/eva.12521

Betancurt, R., Rovere, A. E., and Ladio, A. H. (2017). Incipient domestication processes in multicultural contexts: a case study of urban parks in San Carlos de Bariloche (Argentina). *Front. Ecol. Evol.* 5:166. doi: 10.3389/fevo.2017.00166

Bland, A. (2016). *After A Long Day Of Fighting Climate Change, This Grain Is Ready For A Beer. Natl. Public Radio, Salt*. Available online at: https://www.npr.org/sections/thesalt/2016/10/26/499336902/after-a-long-day-of-fighting-climate-change-this-grain-is-ready-for-a-beer (accessed April 19, 2020).

Borlaug, N. E. (1983). Contributions of conventional plant breeding to food production. *Science* 219, 689–693. doi: 10.1126/science.219.4585.689

Burke, J. M., Tang, S., Knapp, S. J., and Rieseberg, L. H. (2002). Genetic analysis of sunflower domestication another question that QTL mapping can answer re-lates to the identity of traits that were the primary targets. *Genetics* 161, 1257–1267.

Casas, A., Del Carmen Vázquez, M., Viveros, J. L., and Caballero, J. (1996). Plant management among the Nahua and the Mixtec in the Balsas River Basin, Mexico: an ethnobotanical approach to the study of plant domestication. *Hum. Ecol.* 24, 455–478. doi: 10.1007/BF02168862

Casas, A., Otero-Arnaiz, A., Pérez-Negrón, E., and Valiente-Banuet, A. (2007). In situ management and domestication of plants in Mesoamerica. *Ann. Bot.* 100, 1101–1115. doi: 10.1093/aob/mcm126

Chen, K., Wang, Y., Zhang, R., Zhang, H., and Gao, C. (2019). CRISPR/Cas Genome Editing and Precision Plant Breeding in Agriculture. *Annu. Rev. Plant Biol.* 70, 667–697. doi: 10.1146/annurev-arplant-050718-100049

Chen, Y. H., Ruiz-Arocho, J., and von Wettberg, E. J. (2018). Crop domestication: anthropogenic effects on insect–plant interactions in agroecosystems. *Curr. Opin. Insect Sci.* 29, 56–63. doi: 10.1016/j.cois.2018.06.004

Ciotir, C., Applequist, W., Crews, T. E., Cristea, N., DeHaan, L. R., Frawley, E., et al. (2019). Building a botanical foundation for perennial agriculture: global inventory of wild, perennial herbaceous Fabaceae species. *Plants People Planet* 1, 375–386. doi: 10.1002/ppp3.37

Clement, C. R., Denevan, W. M., Heckenberger, M. J., Junqueira, A. B., Neves, E. G., Teixeira, W. G., et al. (2015). The domestication of Amazonia before European conquest. *Proc. Biol. Sci.* 282:20150813. doi: 10.1098/rspb.2015.0813

Cole, S. A., and Gifford, L. S. (2009). *The Cranberry: Hard Work and Holiday Sauce*. Thomaston, ME: Tilbury House Publishers.

Cornelius, D. R. (1950). Seed production of native grasses under cultivation in Eastern Kansas. *Ecol. Monogr.* 20, 1–29. doi: 10.2307/1943521

Cox, T. S., Bender, M., Picone, C., Van Tassel, D. L., Holland, J. B., Brummer, E. C., et al. (2002). Breeding perennial grain crops. *Crit. Rev. Plant Sci.* 21, 59–91. doi: 10.1080/0735-260291044188

Crain, J., Bajgain, P., Anderson, J., Zhang, X., DeHaan, L., and Poland, J. (2020). Enhancing crop domestication through genomic selection, a case study of intermediate wheatgrass. *Front. Plant Sci.* 11:319. doi: 10.3389/fpls.2020.00319

Crews, T. E., Blesh, J., Culman, S. W., Hayes, R. C., Jensen, E. S., Mack, M. C., et al. (2016). Going where no grains have gone before: from early to mid-succession. *Agric. Ecosyst. Environ.* 223, 223–238. doi: 10.1016/j.agee.2016.03.012

De Wit, M. M., and Iles, A. (2016). Toward thick legitimacy: creating a web of legitimacy for agroecology. *Elem. Sci. Anth.* 4:000115. doi: 10.12952/journal.elementa.000115

DeHaan, L., Larson, S., López-Marqués, R. L., Wenkel, S., Gao, C., and Palmgren, M. (2020). Roadmap for accelerated domestication of an emerging perennial grain crop. *Trends Plant Sci.* 25, 525–537. doi: 10.1016/j.tplants.2020.02.004

DeHaan, L. R., and Ismail, B. P. (2017). Perennial cereals provide ecosystem benefits. *Cereal Foods World* 62, 278–281. doi: 10.1094/CFW-62-6-0278

DeHaan, L. R., Van Tassel, D. L., Anderson, J. A., Asselin, S. R., Barnes, R., Baute, G. J., et al. (2016). A pipeline strategy for grain crop domestication. *Crop Sci.* 56, 917–930. doi: 10.2135/cropsci2015.06.0356

DeHaan, L. R., Van Tassel, D. L., and Cox, T. S. (2007). Perennial grain crops: a synthesis of ecology and plant breeding. *Renew. Agric. Food Syst.* 20, 5–14. doi: 10.1079/RAF200496

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

Dong, Z., Alexander, M., and Chuck, G. (2019). Understanding grass domestication through maize mutants. *Trends Genet.* 35, 118–128. doi: 10.1016/j.tig.2018.10.007

Doolittle, W. E., and Mabry, J. B. (2006). "Environmental mosaics, agricultural diversity, and the evolutionary adoption of maize in the American Southwest," in *Histories of Maize: Multidisciplinary Approaches to the Prehistory, Linguistics, Biogeography, Domestication, and Evolution Of Maize*, eds J. Staller, R. Tykot, and B. Benz (Walnut Creek, CA: Left Coast Press), 704.

Dowell, J. A., Reynolds, E. C., Pliakas, T. P., Mandel, J. R., Burke, J. M., Donovan, L. A., et al. (2019). Genome-wide association mapping of floral traits in cultivated sunflower (*Helianthus annuus*). *J. Hered.* 110, 275–286. doi: 10.1093/jhered/esz013

Eck, P. (1990). *The American Cranberry*. New Brunswick, NJ: Rutgers University Press.

Edwards, J., Santos-Medellín, C., Nguyen, B., Kilmer, J., Liechty, Z., Veliz, E., et al. (2019). Soil domestication by rice cultivation results in plant-soil feedback through shifts in soil microbiota. *Genome Biol.* 20:221. doi: 10.1186/s13059-019-1825-x

Eghbarieh, R. (2017). The Struggle for Akoub and Za'atar: On Edible Plants in Palestinian Cuisine and Israeli Plant Protection Laws (Hebrew). Available online at: https://www.academia.edu/34734748/The_Struggle_for_Akoub_and_Za_atar_On_Edible_Plants_in_Palestinian_Cuisine_and_Israeli_Plant_Protection_Laws_Hebrew_ (accessed March 22, 2020).

Elliott, K. C. (2017). *A Tapestry of Values*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780190260804.001.0001

Eshed, Y., and Lippman, Z. B. (2019). Revolutions in agriculture chart a course for targeted breeding of old and new crops. *Science* 366:eaax0025. doi: 10.1126/science.aax0025

Fajardo, D., Morales, J., Zhu, H., Steffan, S., Harbut, R., Bassil, N., et al. (2013). Discrimination of American cranberry cultivars and assessment of clonal heterogeneity using microsatellite markers. *Plant Mol. Biol. Rep.* 31, 264–271. doi: 10.1007/s11105-012-0497-4

Fernie, A. R., and Yan, J. (2019). De novo domestication: an alternative route toward new crops for the future. *Mol. Plant* 12, 615–631. doi: 10.1016/j.molp.2019.03.016

Fuller, D. Q., Allaby, R. G., and Stevens, C. (2010). Domestication as innovation: the entanglement of techniques, technology and chance in the domestication of cereal crops. *World Archaeol.* 42, 13–28. doi: 10.1080/00438240903429680

Gallardo, R. K., Klingthong, P., Zhang, Q., Polashock, J., Atucha, A., Zalapa, J., et al. (2018). Breeding trait priorities of the cranberry industry in the United States and Canada. *HortScience* 53, 1467–1474. doi: 10.21273/HORTSCI13219-18

Garfield, L. (2016). *Patagonia Just Released a Beer that's Different From Any Other Brew You Can Buy*. Available online at: https://www.businessinsider.com/patagonia-new-beer-long-root-ale-2016-10 (accessed March 22, 2020).

Gepts, P. (2010). "Crop domestication as a long-term selection experiment," in *Plant Breeding Reviews Genetic Resources Chromosome Engineering & Crop Improvement*, eds R. Singh and P. Jauhar (Oxford: John Wiley & Sons, Inc), 1–44. doi: 10.1002/9780470650288.ch1

Gliessman, S. R. (2015). *Agroecology: The Ecology of Sustainable Food Systems*, 3rd Edn. Boca Raton, FL: CRC Press.

Gremillion, K. J., Barton, L., and Piperno, D. R. (2014). Particularism and the retreat from theory in the archaeology of agricultural origins. *Proc. Natl. Acad. Sci. U.S.A.* 111, 6171–6177. doi: 10.1073/pnas.1308938110

Hall, S. S. (2018). *Crispr Can Speed Up Nature — and Change How We Grow Food*. Available online at: https://www.wired.com/story/crispr-tomato-mutant-future-of-food/ (accessed March 20, 2020).

Hämälä, T., Guiltinan, M. J., Marden, J. H., Maximova, S. N., DePamphilis, C. W., and Tiffin, P. (2019). Gene expression modularity reveals footprints of polygenic adaptation in *Theobroma cacao*. *Mol. Biol. Evol.* 37, 110–123. doi: 10.1093/molbev/msz206

Hanson, A. A. (1959). *Grass Varieties in the United States (Agriculture Handbook No. 170)*. Washington, DC: Agricultural Research Service.

Harlan, J. R. (1992). "Origins and processes of domestication," in *Grass Evolution and Domestication*, eds G. P. Geoffrey and P. Chapman (Cambridge: Cambridge University Press), 175.

Haspel, T. (2019). *Perennial Grains Could be a Key Weapon Against Climate Change. But not Quite Yet. - The Washington Post. Washington Post*. Available online at: https://www.washingtonpost.com/lifestyle/food/one-day-perennial-grains-could-be-a-key-weapon-against-climate-change-but-not-until-researchers-increase-their-yield/2019/01/24/6df69be6-1f5f-11e9-8e21-59a09ff1e2a1_story.html (accessed April 13, 2020).

Hatfield, J. L., and Walthall, C. L. (2015). Meeting global food needs: realizing the potential via genetics × environment × management interactions. *Agron. J.* 107, 1215–1226. doi: 10.2134/agronj15.0076

Hereward, J., and Curtis, C. (2018). *Tweaking Just a Few Genes in Wild Plants Can Create New Food Crops – But Let ' s Get the Regulation Right*. Melbourne: The Conversation Media Trust.

Horton, H., Asay, K. H., Glover, T. F., Young, S. A., Haws, B. A., Dewey, S. A., et al. (1990). *Grass Seed Production Guide for Utah*. Available online at: https://chapter.ser.org/wp-content/blogs.dir/14/files/2013/02/GrassGuide.pdf (accessed March 22, 2020).

Hybner, R. M., and Jacobs, J. (2012). *INTERMEDIATE WHEATGRASS (Thinopyrum intermedium L.): An Introduced Conservation Grass; Plant Materials Technical Note No. MT-80*. Bozeman, MT: NRCS.

Hyland, H. L. (1977). History of U.S. Plant Introduction. *Environ. Hist. Rev.* 2, 26–32. doi: 10.2307/3984262

Keats, J. (2019). *A New Green Revolution: Scientists Are Using CRISPR to Re-domesticate Fruits and Vegetables*. Available online at: https://www.discovermagazine.com/health/a-new-green-revolution-scientists-are-using-crispr-to-re-domesticate-fruits-and-vegetables (accessed March 20, 2020).

Khan, M. Z., Zaidi, S. S. E. A., Amin, I., and Mansoor, S. (2019). A CRISPR Way for Fast-Forward Crop Domestication. *Trends Plant Sci.* 24, 293–296. doi: 10.1016/j.tplants.2019.01.011

Kleinhuizen, M. (2017). *Kernz-huh? Could the Perennial Promise of Kernza Benefit Food, Beer, and the World? The Growler*. Salina, KS: The Land Institute.

Klingbeil, G., and Rawson, J. (1975). *Wisconsin Cranberry Lore*. Madison, WI: University of Wisconsin Cooperative Extension.

Kloppenburg, J. (2010). Impeding dispossession, enabling repossession: biological open source and the recovery of seed sovereignty. *J. Agrar. Chang.* 10, 367–388. doi: 10.1111/j.1471-0366.2010.00275.x

Kruger, G. (2015). *Intermediate Wheatgrass Seed Production: A Literature Review*. Available online at: https://www.researchgate.net/profile/Gary_Kruger/publication/267196987_Intermediate_Wheatgrass_Seed_Production_A_Literature_Review/links/54da13920cf2970e4e7da7ab/Intermediate-Wheatgrass-Seed-Production-A-Literature-Review.pdf (accessed March 22, 2020).

Laland, K. N. (2017). *Darwin's Unfinished Symphony: How Culture Made the Human Mind*. Princeton, NJ: Princeton University Press.

Larson, G., Piperno, D. R., Allaby, R. G., Purugganan, M. D., Andersson, L., Arroyo-Kalin, M., et al. (2014). Current perspectives and the future of domestication studies. *Proc. Natl. Acad. Sci. U.S.A.* 111, 6139–6146. doi: 10.1073/pnas.1323964111

Larson, S., DeHaan, L., Poland, J., Zhang, X., Dorn, K., Kantarski, T., et al. (2019). Genome mapping of quantitative trait loci (QTL) controlling domestication traits of intermediate wheatgrass (*Thinopyrum intermedium*). *Theor. Appl. Genet.* 132, 2325–2351. doi: 10.1007/s00122-019-03357-6

Leakey, R. R. B., and Asaah, E. K. (2013). Underutilised species as the backbone of multifunctional agriculture -the next wave of crop domestication. *Acta Hortic.* 979, 293–310. doi: 10.17660/actahortic.2013.979.31

Lemmon, Z. H., Reem, N. T., Dalrymple, J., Soyk, S., Swartwood, K. E., Rodriguez-Leal, D., et al. (2018). Rapid improvement of domestication traits in an orphan crop by genome editing. *Nat. Plants* 4, 766–770. doi: 10.1038/s41477-018-0259-x

Li, T., Yang, X., Yu, Y., Si, X., Zhai, X., Zhang, H., et al. (2018). Domestication of wild tomato is accelerated by genome editing. *Nat. Biotechnol.* 36, 1160–1163. doi: 10.1038/nbt.4273

Lins Neto, E., Peroni, N., Casas, A., Parra, F., Aguirre, X., Guillén, S., et al. (2014). Brazilian and Mexican experiences in the study of incipient domestication. *J. Ethnobiol. Ethnomed.* 10:33. doi: 10.1186/1746-4269-10-33

Marouf, M., Batal, M., Moledor, S., and Talhouk, S. N. (2015). Exploring the practice of traditional wild plant collection in lebanon. *Food Cult. Soc.* 18, 355–378. doi: 10.1080/15528014.2015.1043103

Marshall, A. M., and Wilkins, R. W. (2003). Improved seed yield in perennial ryegrass (*Lolium perenne* L.) from two generations of phenotypic selection. *Euphytica* 133, 233–241. doi: 10.1023/A:1025593808010

Marti, A., Bock, J. E., Pagani, M. A., Ismail, B., and Seetharaman, K. (2016). Structural characterization of proteins in wheat flour doughs enriched with intermediate wheatgrass (*Thinopyrum intermedium*) flour. *Food Chem.* 194, 994–1002. doi: 10.1016/j.foodchem.2015.08.082

Martínez, A. K., Soriano, J. M., Tuberosa, R., Martinez, A. K., Koumproglou, R., Jahrmann, T., et al. (2016). Yield QTLome distribution correlates with gene density in maize Wheat Breeding for Potential Areas in Ethiopia View project Cloning of root QTL View project Yield QTLome distribution correlates with gene density in maize. *Plant Sci.* 242, 300–309. doi: 10.1016/j.plantsci.2015.09.022

Mendez, V. E., Bacon, C. M., and Cohen, R. (2013). Agroecology as a transdisciplinary, participatory and action-oriented approach. *Agroecol. Sustain. Food Syst.* 37, 3–18. doi: 10.1080/10440046.2012.736926

Meyer, R. S., DuVal, A. E., and Jensen, H. R. (2012). Patterns and processes in crop domestication: an historical review and quantitative analysis of 203 global food crops. *New Phytol.* 196, 29–48. doi: 10.1111/j.1469-8137.2012.04253.x

Meyer, R. S., and Purugganan, M. D. (2013). Evolution of crop species: genetics of domestication and diversification. *Nat. Rev. Genet.* 14, 840–852. doi: 10.1038/nrg3605

Mitchell, C. E. (2003). Trophic control of grassland production and biomass by pathogens. *Ecol. Lett.* 6, 147–155. doi: 10.1046/j.1461-0248.2003.00408.x

Vogel, K. P. (2000). "Improving warm-season forage grasses using selection, breeding, and biotechnology," in *Native Warm-Season Grasses: Research Trends and Issues CSSA Spec. Pub. No. 30*, eds K. J. Moore, and B. E. Anderson (Madison WI: ASA), 83–106. doi: 10.2135/cssaspecpub30.c6

Morrison, J. I. (2016). *Selection and Breeding to Improve Commercial Germplasm and Increase Germination Percentage of Eastern Gamagrass [Tripsacum dactyloides (L.) L.]*. Available online at: https://search.proquest.com/docview/1784199553?accountid=14544 (accessed March 22, 2020).

Mueller, N. G. (2019). Documenting the evolution of agrobiodiversity in the archaeological record: landraces of a newly described domesticate (*Polygonum erectum*) in North America. *J. Archaeol. Method Theory* 26, 313–343. doi: 10.1007/s10816-018-9375-1

National Agricultural Statistics Service (2019). *2019 Data. USDA Quick Stats*. Available online at: https://quickstats.nass.usda.gov/ (accessed March 23, 2020).

Østerberg, J. T., Xiang, W., Olsen, L. I., Edenbrandt, A. K., Vedel, S. E., Christiansen, A., et al. (2017). Accelerating the domestication of new crops: feasibility and approaches. *Trends Plant Sci.* 22, 373–384. doi: 10.1016/j.tplants.2017.01.004

Ostrander, M. (2017). *The Grain That Tastes Like Wheat, but Grows Like a Prairie Grass. The Nation*. Available online at: https://www.thenation.com/article/archive/the-grain-that-tastes-like-wheat-but-grows-like-a-prairie-grass/ (accessed March 22, 2020).

Peltier, G. (1970). *A History of the Cranberry Industry in Wisconsin*. Detroit, MI: Harlo Press.

Pierce, E. (2019). *Can cereal Save the Planet? This Sustainable New Grain May Bowl Over the Industry. Dallas Morning News*. Available online at: https://www.dallasnews.com/food/cooking/2019/05/23/can-cereal-save-the-planet-this-sustainable-new-grain-may-bowl-over-the-industry/ (accessed March 22, 2020).

Pimentel, D., and Pimentel, M. H. (2007). *Food, Energy, and Society*. Boca Raton, FL: CRC Press.

Purugganan, M. D. (2019). Evolutionary insights into the nature of plant domestication. *Curr. Biol.* 29, R705–R714. doi: 10.1016/j.cub.2019.05.053

Rendón-Anaya, M., and Herrera-Estrella, A. (2018). The advantage of parallel selection of domestication genes to accelerate crop improvement. *Genome Biol.* 19:147. doi: 10.1186/s13059-018-1537-7

Research Collaborative (2018). *al-Nabātāt al-barīyah al-ghidhā'īyah al-Filasṭīnīyah / Palestinian Wild Food Plants*, ed. O. Tesdell (Ramallah: Markaz Khalīl al-Sakākīnī al-Thaqāfī).

Runck, B., Kantar, M., Eckberg, J., Barnes, R., Betts, K., Lehman, C., et al. (2013). "Development of continuous living cover breeding programmes," in *Proceedings of the FAO Expert Workshop on Perennial Crops for Food Security*, eds C. Batello, L. Wade, S. Cox, N. Pogna, A. Bozzini, and J. Choptiany (Rome: FAO), 229–251.

Runck, B. C., Kantar, M. B., Jordan, N. R., Anderson, J. A., Wyse, D. L., Eckberg, J. O., et al. (2014). The reflective plant breeding paradigm: a robust system of Germplasm development to support strategic diversification of agroecosystems. *Crop Sci.* 54, 1939–1948. doi: 10.2135/cropsci2014.03.0195

Ryan, S. F., Adamson, N. L., Aktipis, A., Andersen, L. K., Austin, R., Barnes, L., et al. (2018). The role of citizen science in addressing grand challenges in food and agriculture research. *Proc. Biol. Sci.* 285:20181977. doi: 10.1098/rspb.2018.1977

Saey, T. H. (2018). *Gene Editing Can Speed Up Plant Domestication*. Available online at: https://www.sciencenews.org/article/gene-editing-plant-domestication (accessed March 20, 2020).

Saskatchewan forage seed development commission (1998). *Seed Production of Intermediate Wheatgrass*. Available online at: http://www.skforageseeddc.com/images/stories/pdf/SK_forage_production_guides/Intermediate.pdf

Schiffner, S. A. (2018). *Effects of Agronomic Treatments on Silphium integrifolium, a Potential Perennial Oilseed*. Master's thesis, University of Minnesota, Minneapolis, MN.

Schindele, A., Dorn, A., and Puchta, H. (2020). CRISPR/Cas brings plant biology and breeding into the fast lane. *Curr. Opin. Biotechnol.* 61, 7–14. doi: 10.1016/j.copbio.2019.08.006

Schlautman, B. (2016). *Accelerating Domestication and Genetic Gain in the American Cranberry (Vaccinium Macrocarpon Ait.): New Genetic and Genomic Resources for the Cranberry Breeders' Toolbox*. Ph.D. thesis, University of Wisconsin-Madison, Madison, WI.

Schlautman, B., Barriball, S., Ciotir, C., Herron, S., and Miller, A. (2018). Perennial grain legume domestication phase I: criteria for candidate species selection. *Sustainability* 10:730. doi: 10.3390/su10030730

Schlautman, B., Diaz-Garcia, L., and Barriball, S. (2020). Morphometric approaches to promote the use of exotic germplasm for improved food security and resilience to climate change: a kura clover example. *Plant Sci.* 290:110319. doi: 10.1016/j.plantsci.2019.110319

Scott, J. C. (2017). *Against the Grain: A Deep History of the Earliest States*. New Haven, CT: Yale University Press.

Smil, V. (2018). *Energy and Civilization: A History*. Cambridge, MA: MIT Press.

Smýkal, P., Nelson, M. N., Berger, J. D., and Von Wettberg, E. J. B. (2018). The impact of genetic changes during crop domestication. *Agronomy* 8, 1–22. doi: 10.3390/agronomy8070119

Soltani, N., Dille, J. A., Burke, I. C., Everman, W. J., VanGessel, M. J., Davis, V. M., et al. (2016). Potential corn yield losses from weeds in North America. *Weed Technol.* 30, 979–984. doi: 10.1614/wt-d-16-00046.1

Soltani, N., DIlle, J. A., Burke, I. C., Everman, W. J., Vangessel, M. J., Davis, V. M., et al. (2017). Perspectives on potential soybean yield losses from weeds in North America. *Weed Technol.* 31, 148–154. doi: 10.1017/wet.2016.2

Swanson, H. A., Lien, M. E., and Ween, G. (2018). *Domestication Gone Wild: Politics and Practices of Multispecies Relations*. Durham: Duke University Press.

Tesdell, O., Othman, Y., and Alkhoury, S. (2019). Rainfed agroecosystem resilience in the Palestinian West Bank, 1918–2017. *Agroecol. Sustain. Food Syst.* 43, 21–39. doi: 10.1080/21683565.2018.1537324

Tesdell, O., Othman, Y., Dowani, Y., Khraishi, S., Deeik, M., Muaddi, F., et al. (2020). Envisioning perennial agroecosystems in Palestine. *J. Arid Environ.* 175:104085. doi: 10.1016/j.jaridenv.2019.104085

Tukan, S. K., Takruri, H. R., and Al-Eisawi, D. M. (1998). The use of wild edible plants in the Jordanian diet. *Int. J. Food Sci. Nutr.* 49, 225–235. doi: 10.3109/09637489809086416

Turner, M. K., Ravetta, D., and Van Tassel, D. (2018). Effect of *Puccinia silphii* on yield components and leaf physiology in *Silphium integrifolium*: lessons for the domestication of a perennial oilseed crop. *Sustainability* 10:696. doi: 10.3390/su10030696

Tyl, C., DeHaan, L., Frels, K., Bajgain, P., Marks, M. D., and Anderson, J. (2020). Emerging crops with enhanced ecosystem services: progress in breeding and processing for food use. *Cereal Foods World* 65, 1–9. doi: 10.1094/CFW-65-2-0016

University of Munster (2018). *Pioneering Biologists Create a New Crop Through Genome Editing*. Available online at: https://www.sciencedaily.com/releases/2018/10/181001114217.htm (accessed March 22, 2020).

Van Eck, J. (2018). Genome editing and plant transformation of solanaceous food crops. *Curr. Opin. Biotechnol.* 49, 35–41. doi: 10.1016/j.copbio.2017.07.012

Van Tassel, D. L., Albrecht, K. A., Bever, J. D., Boe, A. A., Brandvain, Y., Crews, T. E., et al. (2017). Accelerating silphium domestication: an opportunity to develop new crop ideotypes and breeding strategies informed by multiple disciplines. *Crop Sci.* 57, 1274–1284. doi: 10.2135/cropsci2016.10.0834

Van Tassel, D. L., Asselin, S. R., Cox, S. A., Sideli, G., and Cattani, D. J. (2014). "Evaluating perennial candidates for domestication: lessons from wild sunflower relatives," in *Perennial Crops for Food Security; Proceedings of the FAO Expert Workshop*, eds C. Batello, L. Wade, S. Cox, N. Pogna, A. Bozzini, and J. Choptiany (Rome: FAO), 112–140.

Vilela, A., González-Paleo, L., Turner, K., Peterson, K., Ravetta, D., Crews, T., et al. (2018). Progress and bottlenecks in the early domestication of the perennial oilseed *Silphium integrifolium*, a sunflower substitute. *Sustainability* 10:638. doi: 10.3390/su10030638

Vogel, K. P., and Hendrickson, J. (2019). History of grass breeding for grazing lands in the northern great plains of the USA and Canada. *Rangelands* 41, 1–16. doi: 10.1016/j.rala.2018.11.006

Vogel, K. P., Reece, P. E., and Nichols, J. T. (1993). Genotype and genotype × environment interaction effects on forage yield and quality of intermediate wheatgrass in Swards. *Crop Sci.* 33, 37–41. doi: 10.2135/cropsci1993.0011183x003300010004x

Vorsa, N., and Zalapa, J. (2019). "Domestication, genetics, and genomics of the american cranberry," in *Plant Breeding Reviews*, ed. I. Goldman (Hoboken, NJ: Wiley), 279–315. doi: 10.1002/9781119616801.ch8

Wagoner, P., and Schaeffer, J. R. (1990). Perennial grain development: past efforts and potential for the future. *Crit. Rev. Plant Sci.* 9, 381–408. doi: 10.1080/07352689009382298

Wagoner, P., and Schauer, A. (1990). "Intermediate wheatgrass as a perennial grain crop," in *Advances in New Crops*, eds J. Janick and J. Simon (Portland, OR: Timber Press), 143–145.

Wang, E., Brown, H. E., Rebetzke, G. J., Zhao, Z., Zheng, B., and Chapman, S. C. (2019). Improving process-based crop models to better capture genotype × environment × management interactions. *J. Exp. Bot.* 70, 2389–2401. doi: 10.1093/jxb/erz092

Wang, L., Beissinger, T. M., Lorant, A., Ross-Ibarra, C., Ross-Ibarra, J., and Hufford, M. B. (2017). The interplay of demography and selection during maize domestication and expansion. *Genome Biol.* 18:215. doi: 10.1186/s13059-017-1346-4

Weaver, J. E., Stoddart, L. A., and Noll, W. (1935). Response of the prairie to the great drought of 1934. *Ecology* 16, 612–629. doi: 10.2307/1932592

Weißhuhn, P., Reckling, M., Stachow, U., and Wiggering, H. (2017). Supporting agricultural ecosystem services through the integration of perennial polycultures into crop rotations. *Sustainability* 9:2267. doi: 10.3390/su9122267

Wills, B. J., Douglas, G. B., Mckenzie, J., Trainor, K. D., and Foote, A. G. (1998). *Thinopyrum intermedium* (Host) Barkw. & Dewey – a review, and evaluation of intermediate and pubescent wheatgrass for dryland agriculture in New Zealand. *Proc. N. Z. Grassl. Assoc.* 60, 233–241.

Wolter, F., Schindele, P., and Puchta, H. (2019). Plant breeding at the speed of light: the power of CRISPR/Cas to generate directed genetic diversity at multiple sites. *BMC Plant Biol.* 19:210. doi: 10.1186/s12870-019-1775-1

Woodhouse, M. R., and Hufford, M. B. (2019). Parallelism and convergence in post-domestication adaptation in cereal grasses. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 374:20180245. doi: 10.1098/rstb.2018.0245

Yang, C. J., Samayoa, L. F., Bradbury, P. J., Olukolu, B. A., Xue, W., York, A. M., et al. (2019). The genetic architecture of teosinte catalyzed and constrained maize domestication. *Proc. Natl. Acad. Sci. U.S.A.* 116, 5643–5652. doi: 10.1073/pnas.1820997116

Zeder, M. A. (2015). Core questions in domestication research. *Proc. Natl. Acad. Sci. U.S.A.* 112, 3191–3198. doi: 10.1073/pnas.1501711112

Zeder, M. A. (2016). Domestication as a model system for niche construction theory. *Evol. Ecol.* 30, 325–348. doi: 10.1007/s10682-015-9801-8

Zeder, M. A., Emshwiller, E., Smith, B. D., and Bradley, D. G. (2006). Documenting domestication: the intersection of genetics and archaeology. *Trends Genet.* 22, 139–155. doi: 10.1016/j.tig.2006.01.007

Zhang, H., Li, Y., and Zhu, J. K. (2018). Developing naturally stress-resistant crops for a sustainable agriculture. *Nat. Plants* 4, 989–996. doi: 10.1038/s41477-018-0309-4

Zhang, J., Zhao, J., Xu, Y., Liang, J., Chang, P., Yan, F., et al. (2015). Genome-wide association mapping for tomato volatiles positively contributing to tomato flavor. *Front. Plant Sci.* 6:1042. doi: 10.3389/fpls.2015.01042

Zhang, X., Sallam, A., Gao, L., Kantarski, T., Poland, J., DeHaan, L. R., et al. (2016). Establishment and optimization of genomic selection to accelerate the domestication and improvement of intermediate wheatgrass. *Plant Genome* 9, 1–18. doi: 10.3835/plantgenome2015.07.0059

Zohary, D., and Hopf, M. (2000). *Domestication of Plants in the Old World: The Origin and Spread of Cultivated Plants in West Asia, Europe and the Nile Valley*, 3rd Edn. Oxford: Oxford University Press.

Zsögön, A., Èermák, T., Naves, E. R., Notini, M. M., Edel, K. H., Weinl, S., et al. (2018). De novo domestication of wild tomato using genome editing. *Nat. Biotechnol.* doi: 10.1038/nbt.4272 [Epub ahead of print].

Conflict of Interest: 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 © 2020 Van Tassel, Tesdell, Schlautman, Rubin, DeHaan, Crews and Streit Krug. 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) and the copyright owner(s) 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 Plant Science

ORIGINAL RESEARCH
published: 12 June 2020
doi: 10.3389/fpls.2020.00834

Genetic Analysis of Domestication Parallels in Annual and Perennial Sunflowers (*Helianthus* spp.): Routes to Crop Development

Sean R. Asselin[1,2], Anita L. Brûlé-Babel[1], David L. Van Tassel[3] and Douglas J. Cattani[1]*

[1] Department of Plant Science, University of Manitoba, Winnipeg, MB, Canada, [2] Agriculture and Agri-Food Canada, Swift Current Research and Development Centre, Swift Current, SK, Canada, [3] The Land Institute, Salina, KS, United States

OPEN ACCESS

Edited by:
Petr Smýkal,
Palacký University Olomouc, Czechia

Reviewed by:
Aleksandra Dimitrijevic,
Institute of Field and Vegetable Crops, Serbia
Alejandro Presotto,
Universidad Nacional del Sur, Argentina

***Correspondence:**
Sean R. Asselin
Sean.Asselin@Canada.ca;
sean.asselin@gmail.com

Specialty section:
This article was submitted to Plant Breeding, a section of the journal Frontiers in Plant Science

Received: *23 March 2020*
Accepted: *25 May 2020*
Published: *12 June 2020*

Citation:
Asselin SR, Brûlé-Babel AL, Van Tassel DL and Cattani DJ (2020) Genetic Analysis of Domestication Parallels in Annual and Perennial Sunflowers (Helianthus spp.): Routes to Crop Development. Front. Plant Sci. 11:834. doi: 10.3389/fpls.2020.00834

Parallels exist between the domestication of new species and the improvement of various crops through selection on traits which favor the sowing, harvest and retention of yield potential and the directed efforts to improve their agronomics, disease resistance and quality characteristics. Common selection pressures may result in the parallel selection of orthologs underlying these traits and homologies between crop species can be exploited by plant breeders to improve germplasm. Perennial grains and oilseeds are a class of proposed crops for improving the diversity and sustainability of agricultural systems. Maximilian sunflower (*Helianthus maximiliani* Schrad.) is a perennial crop wild relative of sunflower (*Helianthus annuus* L.) and a candidate perennial oilseed species. Understanding parallels between cultivated *H. annuus* and *H. maximiliani* may provide new tools for the development of Maximilian sunflower and other wild relatives of sunflower as crops to enhance functional diversity in cropping systems. F_2 populations of Maximilian sunflower segregating for traits associated with the domestication ideotype of cultivated sunflower including branching architecture, capitulum morphology and flowering time were developed to investigate parallels between *H. maximiliani* and *H. annuus*. Genotype-by-sequencing (GBS) was employed to genotype novel Maximilian sunflower populations and perform quantitative-trait-loci (QTL) analysis. A total of 11 QTL in five regions were identified across 21 linkage groups using 4142 GBS derived single nucleotide polymorphism markers called using the sunflower reference genome as a guide. A major QTL on linkage group 17b, associated with aspects of floral development and apical dominance, was discovered and corresponds with a known domestication QTL hotspot in *H. annuus* and candidate genes were identified. This suggests the potential to exploit orthologs for neo-domestication of *H. maximiliani* for traits such as branching architecture, timing of anthesis, and capitulum size and morphology for the development of a perennial oilseed crop from wild relatives of cultivated sunflower.

Keywords: domestication, oilseeds, perennial grains, comparative genomics, ecosystem services, genotype-by-sequencing, functional diversity, sunflower

INTRODUCTION

Annual crops comprise an estimated 60–80% of global cropland and approximately 75% of calories consumed by humans come from four annual grain crops: maize, wheat, rice, and soybean (Lobell et al., 2011). Targeting perennial species and integrating them into agroecosystems dominated by annual crops has been suggested as a method of enhancing functional diversity (i.e., the number of functionally disparate species), ecosystem function (Isbell et al., 2011), and ultimately, productivity through the introduction of new ecosystem services (Asbjornsen et al., 2014). A number of candidate perennial grain and oilseeds crops have been proposed through both the hybridization with annual crops and the identification of wild species with favorable characteristics for neo-domestication (Wagoner and Schaeffer, 1990; DeHaan et al., 2016).

Maximilian sunflower (*Helianthus maximiliani* Schrad.) is a perennial crop wild relative of cultivated sunflower (*Helianthus annuus*) and is native to much of the interior plains of North America, with a range extending from southern Canada to northern Mexico (Kawakami et al., 2011, 2014; Tetreault et al., 2016). Maximilian sunflower has been used in range seeding mixtures for high quality livestock forage, as a perennial filter strip to reduce agricultural run-off, and as a source of wildlife food and habitat (Dietz et al., 1992; USDA-NRCS, 2017). Maximilian sunflower produces an oil rich seed safe for human consumption primarily composed of linoleic and oleic acids and surveys of wild populations have shown an oil content of 31.1% (Dorrell and Whelan, 1978; Seiler, 1994; Seiler and Brothers, 1999). The natural range of Maximilian sunflower, along with its phenotypic divergence (Kawakami et al., 2011; Asselin et al., 2018), high seed production potential, and documented resistance to known common pathogens of annual sunflower (*H. annuus* L.) such as Sclerotinia rot [*Sclerotinia sclerotiorum* (Lib.) de Bary] (Taski-Ajdukovic et al., 2006; Liu et al., 2011), has attracted attention to the species for the development of a perennial oilseed. Maximilian sunflower is a candidate species for the development of dual-use perennial crop for edible oil, forage and bioenergy applications and is targeted for domestication and improvement (Cox et al., 2002; Van Tassel et al., 2014; Asselin et al., 2018).

Efforts to improve of Maximilian sunflower has, until recently, been focused on its use for conservation and rangeland applications or as a trait donor for *H. annuus*. Two open pollinated commercialized cultivars, "Aztec" and "Prairie Gold" were released by the United States Department of Agriculture Natural Resources Conservation Service (USDA-NRCS) in 1978 (Texas Agricultural Experiment Station, 1979; Dietz et al., 1992; USDA-NRCS, 2017). Aztec was developed for wildlife feed, livestock forage cover, use as a natural hedge, as filter-strips, and as an ornamental landscape plant. Prairie Gold was released for landscape reclamation and wildlife food plantings. Both cultivars were selected for vigor and stand establishment in Oklahoma and Texas (Aztec), or Kansas and further north (Prairie Gold). Agronomic research on Maximilian sunflower as a perennial grain candidate began at The Land Institute in the 1980s (Jackson, 1990). The first breeding program focused on developing Maximilian sunflower as a perennial grain was launched in 2002 at The Land Institute (Cox et al., 2002) with the intention of domesticating the species as a perennial oilseed. Selection for seed size and apical dominance has been effective and following three rounds of recurrent selection, the average seed size was increased by 240% and plants exhibiting a highly restricted branching habit were developed by 2012 (Van Tassel et al., 2014).

Crop domestication shows many parallels between species. Vavilov in the development of the *Law of Homologous Series* first recognized that heritable variation in common traits will occur in different species based on parallel selection pressures (Vavilov, 1922). The *Law of Homologous Series* states that closely related species, genera and families exhibit parallel variation in shared traits. In grains and oilseeds such as sunflower, parallel selection for a more determinate growth habit, larger inflorescence size, and increased grain size has occurred through domestication (Zohary, 2004; Purugganan and Fuller, 2009; Lin et al., 2012). A number of domestication syndrome and quality traits in annual grains are controlled by alleles with large effects and relatively simple genetic control. Domestication has led to the selection of common orthologs such as *sh1* genes which contribute to shattering tolerance in sorghum, rice and maize (Lin et al., 2012); *Q* genes in wheat (*Triticum aestivum* L.) and barley (*Hordeum vulgare* L.) which confer the brittle rachis/free-threshing trait (Simons, 2005); and variations in fatty acid desaturase genes in oilseeds such as canola (*Brassica napus* L.) and sunflower (Peng et al., 2010; Chapman and Burke, 2012). Strong apical dominance and restricted branching were key traits identified for the domestication of *H. annuus* (Burke et al., 2002; Wills and Burke, 2007). Wild type Maximilian sunflower has many similarities to wild type *H. annuus*, which exhibits profuse branching, small achenes, indeterminate flowering and lacks shattering resistance (Burke et al., 2002; Wills and Burke, 2007). Maximilian sunflower individuals with restricted branching, resembling the unbranched ideotype of cultivated sunflower, have been identified in germplasm developed at The Land Institute (Cox et al., 2010; Van Tassel et al., 2014) and offer a unique opportunity to examine parallels in the domestication process in *Helianthus*. *H. annuus* and *H. maximiliani* are both diploid species ($2n = 17$) capable of hybridization and cytogenetic analysis has shown a high proportion of normal bivalents in hybrids, suggestive of high homology between parental chromosomes (Jan, 1997; Binsfeld et al., 2001). As an extension of Vavilov's *Law of Homologous Series*, knowledge of the domestication and genomic resources available in annual sunflower could be applied as a model for the neo-domestication and improvement of Maximilian sunflower through marker-assisted-selection (MAS) for domestication orthologs.

Genotype-by-sequencing (GBS) has successfully been utilized for species in which heterologous reference genomes are available for single nucleotide polymorphism (SNP) discovery, such as for wild crop relatives of soybean (Chang et al., 2014), wheat (Edae et al., 2016), and sunflower (Baute et al., 2016; Asselin et al., 2018). The goals of this research were to examine parallels between the domestication of cultivated sunflower and the development of Maximilian sunflower as a perennial oilseed crop. This work

involved applying GBS to Maximilian sunflower using annual cultivated sunflower as a reference genome and performing quantitative-trait-loci (QTL) analysis in a novel Maximilian sunflower population segregating for apical dominance.

MATERIALS AND METHODS

Development of Maximilian Sunflower With Restricted Branching at the Land Institute

Seed from 96 wild Kansas populations were collected in the autumns of 1999 and 2000. Most of these were roadside populations in areas dominated by native rangeland. Ten plants from each population were transplanted to the field at The Land Institute in 2001 and evaluated for two growing seasons. One hundred and fifty individuals with higher than average seed production were selected and clonally replicated by rhizome division in 2003 for further assessment. Seed yield and seed mass data from the 2004–2005 growing seasons were used to select 20 genotypes for polycrossing. Polycrosses were performed by pollen bulking and subsequent manual pollination in 2006. Polycross seed from each of the 20 genotypes was sprouted in the greenhouse and 112 vigorous seedlings from each maternal family were transplanted to a new field in 2007. Although the purpose of this cycle was originally to estimate trait heritability using the 20 half-sib families, a single unbranched individual was found in 2007. Assuming that the unbranched trait was recessive, we planted 100 open-pollinated (half-sibling) seeds from this genotype in 2008, none of which exhibited restricted branching. Half-sibling plants were allowed to intermate and progeny were evaluated in the 2009 growing season for branching characteristics. Reduced branching individuals reappeared in the progeny, however, many of these plants exhibited reduced seed fertility. Plants with the least amount of branching and normal capitula morphology were polycrossed in 2009 and repeated cycles of visual selection for restricted branching and intermating occurred in 2010 and 2011 to develop seeds stock for populations with restricted branching. In 2012 crosses were performed between plants exhibiting restricted branching to generate full sibling families exhibiting limited to no branching.

Development of F_2 Mapping Populations

The mapping populations described in this study were produced through crossing highly-branched wild-type *H. maximiliani* plants from Manitoba to advanced, restricted branching populations developed by The Land Institute (TLI), Salina, Kansas. Manitoban plants were selected as a parent to introduce genetic diversity to alleviate potential inbreeding depression observed in TLI germplasm and to initialize the introduction of the restricted branching trait into an earlier flowering genetic background better suited for northern growing conditions. Maximilian sunflower exhibits clinal variation in life history traits, most notably timing of flowering which occurs in July–August in Manitoban genotypes and September–October in Kansas genotypes (Heiser et al., 1969; Kawakami et al., 2011; Cattani and Van Tassel, personal communication). Full-sib families were developed for mapping traits associated with branching, capitulum size and flowering time. Five restricted branching populations from TLI ("o/i," "I/P," "I/H," "h/L," and "h/e") were initially screened under controlled environments. Plants exhibiting a complete restriction of branches, late flowering genetic background and a single large capitulum (30–40 mm in diameter) were selected as a parents for the development of the mapping populations. The Manitoba parents consisted of wild plants collected from roadside sites across southern Manitoba. Manitoban populations, previously described under growth chamber conditions in Asselin et al. (2018), exhibit extensive branching (∼11–15 branches), an early flowering genetic background and a large number (∼12–20) of small capitula which are ∼7–21 mm in diameter. A series of crosses were produced and F_1 seed derived from a cross between an individual from the TLI population "o/i" and a wild Manitoban accession collected near Brunkild, Manitoba (49.49°N, 97.57W) were grown for sib-mating. The F_1 plants were clonally propagated from rhizome cuttings to produce materials for sib-crossing. Three F_1 individuals (herein denoted as F1A, F1B, and F1C) were intercrossed through reciprocal sib-mating to generate two F_2 populations which segregated for branching, herein denoted as crosses AB (F1A/F1B) and BC (F1B/F1C).

Phenotypic Evaluation

Plant Propagation

Seeds of the F_2 were surface sterilized using a 70% ethanol solution for 10 min, allowed to air dry, and seed coats were then manually removed to break seed dormancy. Seeds were placed in 9 cm petri dishes on filter paper moistened with distilled water containing a 0.1% solution of plant preservative mixture (PPM^{TM}, Plant Cell Technology, Washington, DC, United States). Petri plates were placed in the dark at room temperature for 48 h to germinate. Germinated seeds were transplanted once radicles had emerged. All seedlings were started in 6-well 122.9 mL volume seeding trays filled with Sunshine #4^{TM} soilless potting mix (SunGro Horticulture Ltd., Agawam, MA, United States) and transferred to 1 L pots containing a 2:2:1 ratio of soil:sand:peat by volume once they had reached the three to four true-leaf stage. Plants were grown in a growth chamber under a 23°C 16-h day/18°C 8-h night cycle. Plant positions were assigned randomly and rotated on a weekly basis to account for potential differences in airflow, humidity, and light intensity across benches within the growth chamber. Phenotypic traits measured can be broadly classified as capitulum size traits, traits relating to anthesis, traits associated with branching architecture, traits related to total plant size. Seed yield and size traits were not examined in this study directly due to the constraints of the testing location and ability to cross-pollinate all F_2 progeny examined.

Trait Evaluation

Sixteen traits were phenotyped on 163 and 177 individuals from the AB and BC crosses, respectively, for a total 340 F_2 plants (**Table 1**). As the parental and F_1 individuals underwent

TABLE 1 | Descriptions of 16 phenotypic traits measured under growth chamber conditions for 340 *H. maximiliani* F_2 plants following a 23°C 16-h day/18°C 8-h night cycle.

Capitulum traits	Description
Diameter of capitulum 1–5 (ACW)	Average diameter of the first five capitulum to reach anthesis in mm
Depth of capitulum 1–5 (ACD)	Average depth of the first five capitula to reach anthesis in mm
Depth:width ratio (DWR)	Average capitulum depth/width in mm
Size of central capitulum (CCW)	Size of central capitulum in mm
Total capitula (CC)	Total number of capitula above a minimum size of 2 mm
Branching traits	
Length of branches 1–5 (BL)	Average length of the first five apical branches above 2 cm in in length measured in cm
Highest branching node (HBN)	Measurement of the highest branching point on the main stem. Measured as the number nodes on the main stem above the highest branch with a branch length greater than 2 cm in length
Lowest branching node (LBN)	Measurement of the lowest branching point on the main stem. Measured as the number of basal nodes of the main stem below the lowest branch with a branch length greater than 2 cm
Branches (B)	Number of nodes bearing branches greater than 2 cm in length
Plant size traits	
Total nodes (TN)	Total plant nodes apparent on the main stem
Stem diameter (SD)	Stem diameter measured at the first basal node in mm at maturity
Plant height (PH)	Plant height in cm
Length of leaf 1–5 at maturity (LL)	Length of first five leaves below the central capitulum in cm
Anthesis traits	
Emergence of apical buds (RB)	Number of days to the first appearance of reproductive buds
First anthesis (FA)	Number of days to first anthesis
Average anthesis for the first 5 capitula (AA)	Average days to first anthesis for the first five capitula

photoperiod induction (resulting in skewed phenotypes), complete phenotypic analysis was conducted solely on the F_2 generation. Phenotyping of the appearance of reproductive buds and date of the first capitulum to reach anthesis through fifth capitulum to reach anthesis were conducted three times a week. As plants segregated for flowering time, the date of fifth anthesis was employed as a common physiological milestone at which all other traits were assessed. All nodes were counted starting at the cotyledonary node moving upwards along the main stem. Branches were measured as the total number of branches on the main stem greater than 2 cm in length. The highest branching node, a measurement of apical branching, was measured as the total number of unbranched nodes which were found above the highest branch on the apical portion of the stem. The lowest branching node, a measurement of basal branching, was measured as the number of unbranched nodes on the basal region of the stem between the cotyledonary node and lowest branch. Capitulum morphology was characterized to infer potential traits associated with shattering resistance. Shattering in wild *H. annuus* is in part due to the continued growth of the capitulum resulting in a convex shape (increased depth: width ratio). Domesticated *H. annuus* exhibits a low depth: width ratio, resulting in a flatter capitulum, less prone to shattering (Burke et al., 2002). As plants were not evaluated for seed traits with the depth: width ratio employed as a proxy for this trait.

Data were analyzed for normality using the SAS UNIVARIATE procedure to assure data meet assumptions of normality for downstream analysis. Principal component analysis was conducted to explore trait relationships on centered and scaled data using the R function *PRCOMP*. The number of principal components to retain was determined by examining a breakdown of eigenvectors following Cattell's rule (Cattell, 1966) and Horn's parallel analysis (Horn, 1965) was run using 1000 permutations in the R package *paran* (Dinno, 2012) and adjusted eigenvectors >1 were retained for analysis.

DNA Extraction

Leaf tissues from all plants grown under growth chamber conditions were sampled for DNA extraction. Samples from each plant were labeled with an identity number and frozen in liquid N within 1 h of collection prior to lyophilization and storage at room temperature. Genomic DNA was extracted from the parental, F_1 and a subset consisting of 190 F_2 individuals, consisting of 78 individuals from the AB cross and 112 individuals from the BC cross. A modified single-tube cetyltrimethylammonium bromide (CTAB) extraction protocol (Doyle, 1987; Li et al., 2007) was employed with 1 μL of 10 mg mL^{-1} solution of proteinase K (Promega) added to the initial CTAB incubation step. DNA quantity was determined using a dsDNA broad Range Fluorescence Assay kit and a Qubit 2.0 Fluorometer (Life Technologies) following the manufacturers' instructions for 1 μL sample sizes. DNA quality was assessed using 260/280 and 260/230 nm wavelength absorbance ratios measured using a Nanodrop 1000 spectrophotometer (Thermo Fischer Scientific). Samples that fell below a minimum DNA concentration of 50 μg/μL or absorbance ratios below 1.7 for either 260/230 or 260/280 nm were discarded and extractions were repeated to meet quality requirements.

Genotype-by-Sequencing and SNP Calling

One-hundred and ninety-five DNA samples consisting of the Manitoba and Kansas parentals, F_1 parents and 190 F_2 plants were submitted to Data2bio LLC (Ames, IA, United States) for GBS and SNPs calling. A tunable genotype-by-sequencing (tGBS) protocol was employed, which differs from conventional GBS through the use of a greater selective genome reduction procedure (Ott et al., 2017). Fewer sites are sequenced in tGBS relative to the conventional GBS, resulting in greater read depth. The advantage of this approach is that greater depth of coverage allows for more effective calling of heterozygous genotypes and reduces false homozygote calls, leading to a lower level of missing data per site.

Sequencing via tGBS was carried out using a three selective base (TGT) protocol as described in Ott et al. (2017) for genome reduction. Fragments were sequenced via eight runs on an Ion Proton™ sequencer (Thermo Fisher Scientific, Waltham, MA, United States). The conservative three selective base approach was chosen as *H. maximiliani* is an obligate outcrossing species and exhibit a high degree of heterozygosity (Asselin et al., 2018). Trimmed reads were aligned to the cultivated *H. annuus* reference genome, HA412.v.1.1.bronze.20142015,[1] using GSNAP (Wu and Nacu, 2010) with polymorphisms within the first and last 3 bp for the read being ignored (Ott et al., 2017). Polymorphisms with base calls with a PHRED score below 20 were removed from the analyses. The SNPs were called as homozygous if the most common allele was supported by a minimum of 80% aligned reads, while heterozygous SNPs were called if the two most common alleles were supported by a minimum of 30% of all aligned reads and the sum of the two most common alleles accounted for a minimum of 80% of the aligned reads. Single nucleotide polymorphisms that could be genotyped in a minimum of half of the samples, that contained an allele number of 2, a minor allele frequency >1% and a heterozygous rate of <70% were maintained for downstream analysis. Imputation was conducted separately on the AB and BC crosses using Beagle4.1 (Browning and Browning, 2016) using default parameters.

Linkage Map Construction

The AB and BC crosses were examined jointly as a single population employing markers segregating in the same fashion between the two crosses for analysis in Joinmap 5.0 (Van Ooijen, 2018). Common segregation patterns between the crosses were determined using the F_1 genotypes and observed segregation patterns in the F_2 generation to allow joint analysis. Markers were coded as three possible segregation types (<lmxll>, <nnxnp>, and <hkxhk>) corresponding to backcross-like or F_2-like segregation patterns. Homozygous allele states accounting for less than 10% of the observed total calls were considered likely genotyping errors and set to missing data prior to analysis. Markers classes in which either the parental genotype could not be accurately determined or were non-informative were removed. The remaining markers were tested for distorted segregation using a chi-square goodness-of-fit test through the test segregation function of Joinmap at $p < 0.05$ for downstream analysis. Due to the presence of a high degree of segregation distortion, markers exhibiting distortion were retained within the dataset and initially grouped based upon their assigned linkage groups in the cultivated *H. annuus* HA412.v.1.1.bronze.20142015 reference genome for ordering. Segregation distortion has been reported as having little effect on marker order (Zhang et al., 2010), but erroneously called genotypes, which may result in the appearance of highly-distorted markers, may cause pseudo-linkages between biologically unlinked sequences (Ronin et al., 2010).

Linkage map construction was conducted using the cross-pollinated (CP) analysis option in Joinmap 5.0. Identical markers were removed and reference genome positions were utilized in initial maps using the "start order" option to position markers and linkage groups were numbered based upon their reference genome analog. Markers were ordered using the regression mapping algorithm with map distances in cM being calculated using Kosambi's genetic distance. Markers which produced negative map distances were excluded from the map and maps were recalculated. In instances were insufficient linkage was present within a linkage group markers were split into separate groups (i.e., LG 3a and 3b) for analysis to produce the final map.

QTL Analysis

Restricted multiple-qtl model (rMQM) analysis was conducted using MapQTL 6.0. As a subset of the genotyped F_2 plants exhibited an abnormal "die-back" of the central capitulum, 15 plants were removed from the dataset, leaving 175 individuals for QTL analysis. An initial analysis was conducted to identify putative QTL and possible cofactors for rMQM using the interval mapping procedure at 1 cM intervals. A non-parametric one-way analysis of variance Kruskal–Wallis (K–W) rank-sum test was also conducted. The K–W test method does not require the incorporation of a genetic map, providing marker-trait associations independently of neighboring markers and was used to confirm putative co-factors and QTLs. To account for multiple testing, a stringent alpha of 0.005 was used to reduce potential spurious associations as suggested by Van Ooijen (2018). Possible rMQM cofactors were limited to no more than two per linkage group per trait and an automatic cofactor selection procedure in MapQTL 6.0 was performed using default settings. Subsequent rounds of rMQM mapping were performed for each trait until co-factors stabilized. To account for multiple testing and control of family-wise-error rate for type I errors, whole-genome LOD thresholds were calculated for each trait independently (Churchill and Doerge, 1994). Appropriate genome-wide LOD threshold values were estimated using 1000 permutations and an alpha threshold of $p < 0.05$. A genome-wide LOD threshold score of $p < 0.05$ was employed to declare significant QTL. Confidence intervals were calculated using the two-LOD drop-off method which approximates roughly a 95% confidence interval of QTL position (Lander and Botstein, 1989). Strongly overlapping/pleiotropic QTL between two or more traits were given a common name to designate the QTL region.

To infer candidate genes and provide functional annotation of the candidate SNPs generated from the analyses, *H. maximiliani* SNPs falling within the two-LOD QTL confidence interval were compared with the cultivated *H. annuus* reference genome HA412.v.1.1.bronze.20142015 assembly using Jbrowse (Skinner et al., 2009), available via the HeliaGene bioinformatics portal (Carrere et al., 2008). Significant SNPs found within a 1kb distance were considered a common entry for annotation. SNPs were compared to their positions within the *H. annuus* reference genome and a search was conducted for possible candidate genes within 100 kb of each entry. To annotate the presence of putative orthologs, a literature search of sunflower domestication studies was conducted to determine if QTL controlling similar traits

[1] www.sunflowergenome.org

have previously been reported on the same linkage groups in *H. annuus*.

RESULTS

Phenotypic Analysis

The initial TLI populations screened for restricted branching exhibited either a single capitulum and a complete lack of branching, or a complete lack of mid-plant and basal branches and a small number (2–3) short apical branches bearing capitula. It has been suggested that branching restriction in Maximilian sunflower exhibits a leaky phenotype, and this observation is consistent with previous observations from restricted branching populations (Van Tassel, personal communication). The F_1 plants were uniform in appearance, and exhibited a branched phenotype consistent with that of wild Manitoban populations. The range of phenotypes observed amongst both crosses was extensive in the F_2 generation and the phenotypic distributions are suggestive of quantitative genetic control (**Table 2**). The unbranched phenotype observed in the TLI parent was not recovered in the F_1 or F_2 generations although a number of plants exhibiting highly restricted branching were observed in the F_2 generation. Branch number ranged from one to 21 branches and the percentage of branch bearing nodes ranged from 2.38 to 48.97%. Timing of first anthesis ranged from 61 to 198 days. Plant height ranged from 90to 192 cm and diameter of the central capitulum ranged from 6.2to 25.6 mm. The total number of capitulum exhibited the highest coefficient of variation relative to other traits examined, ranging from 4 to 126 capitula per plant likely due to the presence of secondary and tertiary branches.

Principal component analysis revealed four significant components and patterns suggestive of a resource allocation trade-off relating to the degree of apical dominance (**Figure 1** and **Supplementary Table S1**). Along the first two principal components axes a negative relationship between capitulum size traits (greater apical dominance) and the total number of branches and capitula (reduced apical dominance) is apparent. The first component (PC1) which explained 30.9% of the phenotypic variance appears to be dominated by trait differences in the parental materials, with positive loadings for later anthesis and increased capitulum size and negative loadings for capitulum and branch number. The second component (PC2), explaining 21.64% of the phenotypic variance revealed patterns between timing and anthesis and capitulum size traits which contradict the first component, suggesting these traits are not strongly associated and may represent recombination between parental types. The third and fourth components (**Supplementary Figure S1**), explaining 9.84 and 8.36% of the phenotypic variance respectively were dominated by branching characteristics (PC3) and other aspects of apical dominance (highest branching node, average branch length and capitulum count) as well as stem diameter (PC4).

SNP Calling and Linkage Map Development Using the Homeologous *H. annuus* Genome as a Reference Guide

A total of 6094 polymorphic SNPs were aligned to the homeologous reference genome of *H. annuus* for QTL identification spanning all linkage groups. Following filtering 5323 SNPs were considered for joint linkage map development and QTL analysis. Due to insufficient linkage more than the expected 17 linkage groups of *H. maximiliani* were recovered, linkage groups 3, 5, 13, and 17 were split into separate groups resulting in 21 linkage groups (**Supplementary Figure S2**). To date, this is the first reported linkage map of Maximilian

TABLE 2 | Phenotypic mean, standard error of the mean (SEM) and coefficient of variation (CV) and other summary statistics of 340 F_2 *H. maximiliani* plants phenotyped for 21 traits under growth chamber conditions following a 23°C 16-h day/18°C 8-h night cycle.

Capitulum traits	Mean	SEM	Skewness	Kurtosis	Range	CV (%)
Mean diameter of capitula 1–5 (mm)	12.6	0.13	0.05	-0.41	14.18	20
Mean depth of capitula 1–5 (mm)	11.8	0.09	0.07	-0.08	12.04	15
Capitulum depth:width ratio	0.95	0.005	1.06	4.11	0.9	11
Size of central capitulum (mm)	13.89	0.18	0.35	-0.07	19.4	24
Total capitula count	39.64	1.21	1.36	1.83	122	56
Branching						
Average length of branches 1–5 (cm)	29.41	0.52	0.23	-0.04	57.8	33
Highest branching node	5.6	0.12	1	2.33	15	40
Lowest branching node	36.28	0.59	0.21	0.83	68	30
Total Branches	8.31	2.78	1.07	2.07	20	33
Plant size						
Total nodes	50.96	0.58	0.59	0.84	67	21
Stem diameter (mm)	9.53	0.07	0.26	0.51	8.5	14
Plant height (cm)	142.11	1.11	-0.07	-0.63	102	15
Length of leaf 1–5 at maturity (cm)	12.53	0.21	0.21	-0.22	22.1	31
Anthesis						
Emergence of reproductive buds (days)	67.44	0.78	-0.02	-0.03	82	21
First anthesis (days)	118.37	1.34	0.66	0.46	137	21
Average anthesis for the first 5 capitula (days)	133	1.33	0.34	-0.05	134.8	18

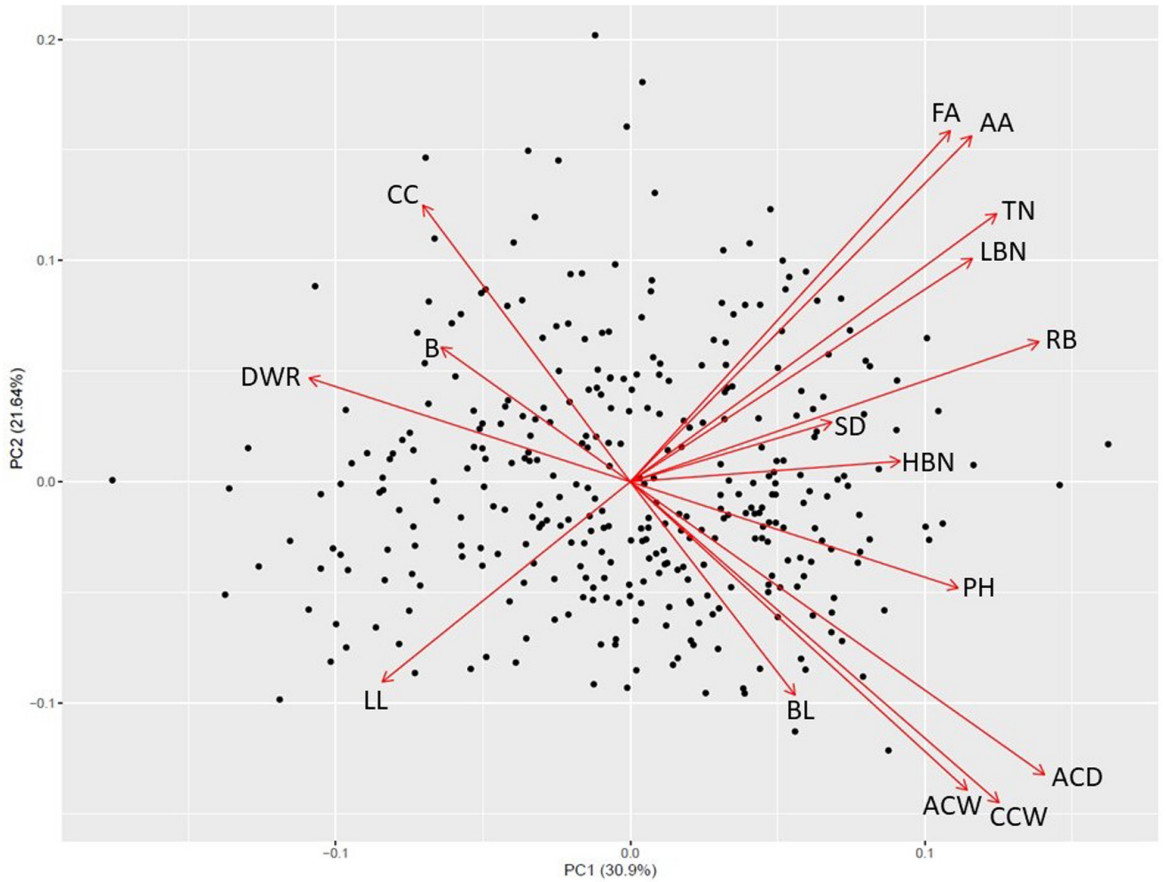


FIGURE 1 | First and second principal component axes for 340 Maximilian sunflower individuals segregating for apical dominance. FA, First anthesis; AA, average date of anthesis; TN, total nodes; LBN, lowest branching node; RB, reproductive budding; SD, stem diameter; HBN, highest branching node; PH, plant height; ACD, average capitula depth; CCW, central capitulum width; ACW, average capitula width; BL, branch length; LL, leaf length; DWR, capitulum depth:width ratio; B, branches; CC, total capitula count.

sunflower and spans a total of 3159.29 cM of the Maximilian sunflower genome, with an average marker interval of 1.31 cM. A total of 4142 markers were incorporated into the map, with 42–300 markers per linkage group (**Supplementary Table S2**).

QTL Analysis and Identification of Putative Candidate Genes Associated With Domestication

The Genome-wide QTL scans detected 11 significant QTL that were supported by both the rMQM and K-W analysis which corresponded to five genomic regions. A major co-localized QTL (herein referred to as *ultra1*) was detected on linkage group 17b explaining 12.1–21.5% of the variation in traits associated with timing of anthesis (timing of reproductive budding, first anthesis and average date of anthesis) as aspects of apical dominance (number of branches, average leaf length, average capitulum size, and average capitulum depth) (**Table 3**). Further QTL for timing of reproductive budding were found on linkage groups 6 (*rb1*) and 15 (*rb2*), and total capitulum count on linkage groups 2 (*cc1*) and 3b (*cc2*).

To identify putative genes associated with the domestication phenotype of *H. annuus* 100 kb flanking regions surrounding SNPs within the two-LOD QTL support intervals were manually examined using the sunflower genome JBrowse (INRA Sunflower Bioinformatics Resources, 2014). None of the QTL associated SNPs were found within gene sequences but several were within close proximity to annotated genes (**Supplementary Table S3**).

The SNPs associated with the QTL *ultra1* on linkage group 17b of *H. maximiliani* correspond to a gene-rich region of the *H. annuus* reference genome. A variety of putative functions were found within this region relating to hormone response and regulation and meristem development. Nearby genes in the region of *ultra1* include plant GH3 auxin-responsive promotors (IPR004993) and a ULTRAPETALA developmental regulator (IPR020533) (**Supplementary Table S3**) which is known as a key negative regulator of cell accumulation in shoot and floral meristems in *Arabidopsis thaliana* (Fletcher, 2001; Carles et al., 2004). Genes encoding Cytochrome P450 (IPR001128) and F-box domain proteins (IPR001810), both of which are candidate genes for branching in *H. annuus* (Mandel et al., 2013; Nambeesan et al., 2015), were also identified.

TABLE 3 | Significant genome-wide QTL in the tested Maximilian sunflower populations ($p < 0.05$).

QTL	Trait	Nearest Markers	Linkage group	Peak (cM)	CI[1] (cM)	%Var[2]	LOD[3]
ultra1	Timing of reproductive budding	Ha17_43720802	17b	118.57	116.76–119.31	12.1	6.22
	First anthesis	Ha17_43720799	17b	117.14	116.72–119.31	18.8	5.54
	Average date of anthesis	Ha17_43720802	17b	118.57	116.76–119.31	20.7	7.51
	Total branches	Ha17_38911771	17b	118.57	117.14–123.31	18.8	6.71
	Average leaf length	Ha17_43720799	17b	117.14	113.47–119.31	15	6.5
	Average capitulum width	Ha17_43720799	17b	117.14	111.47–119.31	17.4	6.17
	Average capitulum depth	Ha17_43720802	17b	117.57	112.47–119.31	21.5	7.34
rb1	Timing of reproductive budding	Ha6_80827033	6	95.15	92.57–99.08	11.6	7.39
rb2	Timing of reproductive budding	Ha15_4372179	15	12.5	10.50–13.17	11.7	7.66
cc1	Total capitulum count	Ha2_149340796	2	68.63	65.64–70.91	10	6.43
cc2	Total capitulum count	Ha3_54508649	3b	76.54	76.14–77.54	12.4	6.32

[1]CI, Two-LOD confidence interval as approximated using the method of Lander and Botstein, 1989, [2]%Var, Percent of variation in which the QTL explains, [3]LOD, likelihood of odds.

Interestingly, the SNPs associated with *ultra1* are approximately 200 kb downstream to a region containing a number of potentially influential genes associated with AUX/IAA proteins (IPR003311), Auxin response factors (IPR010525) and self-incompatibility S1 proteins (IPR010264) suggesting potential genetic linkage between genes associated with the domestication phenotype of *H. annuus* and self-compatibility as noted in previous studies (Burke et al., 2002; Gandhi et al., 2005; Tang et al., 2006).

The SNPs associated with *cc1* were found to be in proximity to genes which code proteins related to cell proliferation, differentiation and development (ARID/BRIGHT DNA-binding domain, IPR001606) and plant hormone production and regulation (AUX/IAA protein, IPR003311; Auxin response factor, IPR015300; Terpenoid synthase, IPR0089494 and Cytokinin riboside 5′-monophospahte phophoribohydrolase LOG, IPR005269). The annotations for *cc2* included a fibroblast growth factor receptor (IPR016248). While SNPs associated with *rb1* were not found to be within 100 kb of putative candidate genes, these SNPs flank a genomic region containing a phytochrome domain (IPR001294) which is associated with plant light response, the control of photoperiodism and flowering. The presence of a flowering-time QTL on this linkage group is also in line with previous studies which have identified a cluster of paralogous *HaFT* genes on LG 6 which are major contributors to photoperiod sensitivity in *H. annuus* (Blackman et al., 2011). Annotations for *rb2* which corresponds to LG 15 of *H. annuus* did not yield obvious functional candidates within close proximity to the associated SNPs though a QTL for flowering is known on this linkage group in wild X landrace crosses of *H. annuus* (Wills and Burke, 2007; Blackman et al., 2011).

DISCUSSION

Genetic Parallels With *H. annuus* and Putative S-Locus Linkage

The release of the sunflower reference genome has enhanced the ability to generate SNP libraries for analyzing both natural and segregating populations of cultivated sunflower as well as their wild crop relatives. The use of the sunflower reference genome greatly assisted in the development of a genetic map of Maximilian sunflower and in the identification of candidate genes.

Several promising candidate genes within 100 kb of *ultra1* were identified in this study, one of which is documented in *Arabidopsis* to influence the phenotypic traits associated with *ultra1*. *ULTRAPETALA* (*Ult1*) has been shown to regulate a number of development processes including the timing of vegetative to floral transition, maintenance of the shoot apical meristem and termination of floral meristem development in *A. thaliana* (Fletcher, 2001; Carles et al., 2004). This is in line with the associations between *ultra1* and timing of anthesis, overall capitulum size and branching in the present study, which presents a strong case for a causative candidate gene. Plants that carry a *Utl1* loss-of-function mutation show restricted branching and an increase in floral meristem size similar to the Kansas populations utilized in this study.

Another promising candidate gene in proximity to *ultra1* is a GH3 auxin-responsive promotor. The role of auxins, cytokinins, and strigolactones in the control of shoot branching and maintenance of apical dominance is well-documented (Bennett et al., 2006; Shimizu-Sato et al., 2009). Several candidate genes have been identified in sunflower domestication studies relating to the production, regulation, and transport of these hormones (Nambeesan et al., 2015).

Additional candidates underlying *ultra1* include cytochrome P450 and F-box proteins. The cytochrome P450 and F-box gene families include several well-known genes associated with the control of axillary meristem initiation and growth such as the more axillary growth (*MAX*) genes of *A. thaliana* (Stirnberg et al., 2002). One of these genes, *MAX2* encodes an F-box protein which has been shown to co-localize with a branching QTL on LG 17 of *H. annuus* (Mandel et al., 2013).

An interesting parallel between *H. maximiliani* and *H. annuus* was observed in the QTL analysis. The QTL *ultra1* on LG 17b corresponds to a region in the *H. annuus* reference genome which harbors annotations for a number of *S*-locus associated

proteins, suggesting potential linkage between *ultra1* and self-compatibility in *H. maximiliani*. Previous studies looking at wild X crop crosses in *H. annuus* have observed a preponderance for crop alleles on this linkage group and a clustering of upwards of 20 domestication and post-domestication QTL in tight linkage with the *S*-locus on LG 17 (Burke et al., 2002; Gandhi et al., 2005; Tang et al., 2006).

The role of *ULTRAPETALA* (*Ult1*) genes in sunflower has yet to be explored but several studies have identified domestication QTL in tight linkage with the *S*-locus on LG 17 of *H. annuus*. Burke et al. (2002) observed a clustering of QTL for flowering time, stem diameter, capitulum number, number of capitula/branch, plant height, number of leaves, peduncle length, achene weight, shattering and total number of selfed seeds on LG 17. Gandhi et al. (2005) subsequently mapped the *S*-locus to the same region on LG 17. Tang et al. (2006) observed several overlapping QTL for seed size and seed oil concertation in close proximity to the *S*-locus on LG 17. Wills and Burke (2007) also observed overlapping QTL for capitulum diameter, branch number, ray number, total capitula number and number of selfed seeds again in the same region on LG 17.

The presence of an *S*-locus on LG 17b in Maximilian sunflower may impact breeding efforts and is one explanation that helps explain the reduced fertility observed in the initial Kansas populations selected for apical dominance. This observation warrants further investigation as other forms of inbreeding depression reducing fertility cannot be ruled out at this stage. Tight linkage between domestication QTL such as *ultra1* and an *S*-locus could impose restrictions on domestication if linkage cannot be broken. Stringent selection for a given variant of a *ultra1* allele may result in a genetic bottleneck and reduced *S*-locus allelic diversity. Limited genetic diversity in breeding pools of Maximilian sunflower may limit seed yield through the lack of compatibility in pollen sources, reducing to ability to generate novel germplasm through directed crosses. Fine-scale genetic mapping of *ultra1* and the *S*-locus and development of further SNP markers in Maximilian sunflower would provide an efficient way to break the putative genetic linkage between these loci through the identification of recombinant progeny carrying different *S*-locus alleles.

Phenotypic Trade-Offs in the Context of Neo-Domestication

A trade-off between capitulum size and capitulum number was observed along both the first and second principal component axis in the F_2 populations and is suggestive of a putative resource allocation trade-off between capitulum size and capitulum number. In *H. annuus*, unbranched biotypes have larger capitula and seeds than their branched counterparts.

The presence of a major QTL such as *ultra1* influencing flowering time, branching and capitulum morphology would explain the first principal component of the PCA which is dominated by type differences observed in the parental materials. The presence of *ultra1* helps explain the phenotypic correlations between traits such as later flowering and increased capitulum width and depth. Interestingly, the second principal component shows the opposite relationship between these traits suggesting this trait correlation is not absolute.

While selection for a single large capitulum is a defining domestication syndrome characteristic of cultivated annual sunflower, the path perennial oilseeds will take through domestication will ultimately depends on the how standing genetic variation will contribute to yield. In sunflower, branching not only interacts with capitulum size and total capitula number but also seed weight and oil content amongst other traits (Fick et al., 1974; Dedio, 1980; Tang et al., 2006; Bachlava et al., 2010).

While there are inherent benefits to restricting branching, such as the facilitation of mechanical harvest and uniform maturity, restricting branching may also limit yield potential if selected for too stringently. Branched Maximilian sunflower plants are capable of producing a large number of capitula per $stem^{-1}$ (>100) while completely unbranched plants exhibited a single, central capitulum. The relationship between capitulum size and capitula number in Maximilian sunflower is not a 1:1 ratio. Plants with restricted branching have the tendency to exhibit larger capitula, though this increase in capitulum size does not appear to fully compensate for the loss of capitulum number and it has been suggested that this may limit seed yield (Van Tassel et al., 2014; DeHaan et al., 2016). Selection for a single central capitulum, akin to the domestication ideotype of cultivated sunflower, could decrease seed production in *H. maximiliani* if the loss of capitula through restricted branching is not compensated for by a reciprocal increase in capitulum size. Conversely, the single capitulum ideotype may be better suited for polyculture use as opposed to a monoculture, as the loss of yield associated with an unbranched stem may be compensated for by diversification within the field (Picasso et al., 2008; Picasso et al., 2011).

Helianthus annuus was initially domesticated for its edible seed, pigments and medicinal compounds (Heiser, 1951) with selection for oil content and composition occurring more recently (Burke et al., 2005; Chapman and Burke, 2012). Therefore, the defining characteristics of its domestication may not necessarily apply to the neo-domestication of Maximilian sunflower as an oilseed crop. Selection for a Maximilian sunflower ideotype as an oilseed may parallel other small-seeded *Compositae* oilseeds such as safflower (*Carthamus tinctorius*) and noug (*Guizotia abyssinica*) both of which exhibit contrasting domestication syndromes to annual sunflower (Pearl et al., 2014; Dempewolf et al., 2015), having retained a branched architecture and selection for greater seed production. Similarly, successful oilseed crops in Western Canada such as canola (*B. napus* L.) retain a branched architecture, small seed size, and are still prone to considerable shattering losses under certain conditions (Gulden et al., 2003; Cavalieri et al., 2016). Despite these characteristics, these crops are capable of sustaining economic yields supporting their use as crops. Increased branching and seed number may constrain seed size, in *H. annuus* and smaller seeds tend to bear a higher concentration of oil. Estimates of 6.7–8.5% greater oil concentration in the seed of branched individuals relative to unbranched individuals in segregating populations have been observed, presumably due to a thinner pericarp (Tang et al., 2006; Bachlava et al., 2010). Therefore,

increasing branching and seed number may prove beneficial in increasing oil yield per unit area.

Though selection for restricted branching may limit the total production of capitula, and by extension yield potential on a single plant basis, branching restriction may enhance yield on a per unit area. It has been observed that in species such as safflower lower order branches are often infertile and that in general, plants that produced many branches produced smaller capitula, fewer florets per branch, and smaller seeds in smaller quantities (Bidgoli et al., 2006). Branching restriction may enhance harvestable yield indirectly through greater synchronicity of flowering, maturity of capitula and through changes in capitulum morphology, both of which may alleviate shattering losses. In the present study, the capitulum depth:width ratio appears to decrease with branching restriction and the increase in capitulum size. Plants with a wider, flatter capitulum, more akin to those of domesticated sunflower may exhibit reduced shattering losses if these patterns are common between species. Furthermore, it may be possible to compensate for the loss of secondary capitula through increasing stem density on a per unit area, either through selection for a greater number of stems per plant or sowing higher density plantings coupled with tolerance to crowding. Ultimately, increasing the harvest index of Maximilian sunflower through conventional as well as marker-assisted breeding approaches appears to be possible through multiple pathways and further investigation into these factors under realistic field conditions is required to define the optimal ideotype for production.

In summary, we were successful in identifying QTL for important traits associated with domestication in Maximilian sunflower and identified candidate genes using the existing genomic resources of cultivated sunflower. Results presented in this study have implications for the defining an appropriate ideotype for perennial oilseed crops in the *Compositae* and the development of genomic resources in crop wild relatives. Further work is needed to compare contrasting Maximilian sunflower ideotypes and characterization is required to verify the presence of putative *S*-locus linkage and the development of markers for this and other domestication QTL. Future work will focus on the field based evaluation of yield components in proposed Maximilian sunflower ideotypes and the development of a *de novo* genetic map for Maximilian sunflower and fine mapping of genomic regions surrounding *ultra1*. Understanding both the genetic and agronomic parallels of domestication will play an important role in the neo-domestication of Maximilian sunflower and related species.

DATA AVAILABILITY STATEMENT

The SNP dataset generated in this study is available in Dryad digital repository (doi: 10.5061/dryad.tdz08kpwt). Any other datasets and germplasm used and/or analyzed during the current study will be made available through contact with the corresponding author on reasonable request.

AUTHOR CONTRIBUTIONS

SA generated the mapping populations, conducted the greenhouse and laboratory work, analyzed the data, and wrote the manuscript as a graduate student. DC and AB-B provided laboratory resources and supervision of experimental design and aided in the editing of the manuscript. DV developed the initial restricted branching populations and provided germplasm and feedback on the manuscript. All authors have reviewed and agreed on the contents of the manuscript.

FUNDING

This work was supported by the Western Grains Research Foundation graduate scholarship, by grants from Manitoba Agriculture [Advanced Rural Development Initiative (ARDI) Project No. 12-1136], and The Land Institute, Salina, KS.

ACKNOWLEDGMENTS

The authors would like to acknowledge Ardelle Slama, Robert Visser, Kathy Bay, and Ian Brown for providing greenhouse support throughout the length of the experiment and Dr. Maria Stoimenova and Dr. Curt McCartney for advice and technical support in the areas of molecular biology and genetic mapping respectively. Dr. Robert Gulden, Dr. Barbara Sharanowski, and Dr. Khalid Rashid are acknowledged for serving on SA's graduate student committee and providing feedback throughout the length of the experiment.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.00834/full#supplementary-material

REFERENCES

Asbjornsen, H., Hernandez-Santan, V., Liebman, M., Bayala, J., Chen, J., Helmers, M., et al. (2014). Targeting perennial vegetation in agricultural landscapes for enhancing ecosystem services. *Renew. Agric. Food Syst.* 29, 101–125. doi: 10.1017/S1742170512000385

Asselin, S. R., Brûlé-Babel, A. L., and Cattani, D. J. (2018). Genetic characterization of maximilian sunflower for the development of a locally adapted perennial grain oilseed. *Crop Sci.* 58, 2241–2260. doi: 10.2135/cropsci2017.10.0611

Bachlava, E., Tang, S., Pizarro, G., Schuppert, G. F., Brunick, R. K., Draeger, D., et al. (2010). Pleiotropy of the branching locus (B) masks linked and unlinked quantitative trait loci affecting seed traits in sunflower. *Theor. Appl. Genet.* 120, 829–842. doi: 10.1007/s00122-009-1212-1

Baute, G. J., Owens, G. L., Bock, D. G., and Rieseberg, L. H. (2016). Genome-wide genotyping-by-sequencing data provide a high-resolution view of wild *Helianthus* diversity, genetic structure, and interspecies gene flow. *Am. J. Bot.* 103, 2170–2177. doi: 10.3732/ajb.1600295

Bennett, T., Siberer, T., Willett, B., Book, J., Luschnig, C., and Leyser, O. (2006). The *Arabidopsis* MAX pathway controls shoot branching by regulating auxin transport. *Curr. Biol.* 16, 553–563. doi: 10.1016/j.cub.2006.01.058

Bidgoli, A. M., Akbari, G. A., Mirhadi, M. J., Zand, E., and Soufizadeh, S. (2006). Path analysis of the relationships between seed yield and some morphological and phenological traits in safflower (*Carthamus tinctorius* L.). *Euphytica* 148, 261–268. doi: 10.1007/s10681-005-9019-x

Binsfeld, P. C., Wingender, R., and Schnabl, H. (2001). Cytogenetic analysis of interspecific sunflower hybrids and molecular evaluation of their progeny. *Theor. Appl. Genet.* 102, 1280–1285. doi: 10.1007/s001220000517

Blackman, B. K., Rasmussen, D. A., Strasburg, J. L., Raduski, A. R., Burke, J. M., Knapp, S. J., et al. (2011). Contributions of flowering time genes to sunflower domestication and improvement. *Genetics* 187, 271–287. doi: 10.1534/genetics.110.121327

Browning, B. L., and Browning, S. R. (2016). Genotype imputation with millions of reference samples. *Am. J. Hum. Genet.* 98, 116–126. doi: 10.1016/j.ajhg.2015.11.020

Burke, J. M., Knapp, S. J., and Rieseberg, L. H. (2005). Genetic consequences of selection during the evolution of cultivated sunflower. *Genetics* 171, 1933–1940. doi: 10.1534/genetics.104.039057

Burke, J. M., Tang, S., Knapp, S. J., and Rieseberg, L. H. (2002). Genetic analysis of sunflower domestication. *Genetics* 161, 1257–1267.

Carles, C. C., Lertpiriyapong, K., Reville, K., and Fletcher, J. C. (2004). The ULTRAPETALA1 gene functions in early *Arabidopsis* development to restrict shoot apical meristem activity and acts through WUSCHEL to regulate floral meristem determinacy. *Genetics* 167, 1893–1903. doi: 10.1534/genetics.104.028787

Carrere, S., Gouzy, J., Langlade, N., Gamas, P., and Vincourt, P. (2008). "HELIAGENE, a bioinformatics portal for *Helianthus* sp genomics," in *Proceeding of the 17th International Sunflower Conference, Cordoba, Spain. 8–12 June 2008* (Paris: International Sunflower Association), 611–615.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behav. Res.* 1, 245–276. doi: 10.1207/s15327906mbr0102_10

Cavalieri, A., Harker, K. N., Hall, L. M., Willenborg, C. J., Haile, T. A., Shirtliffe, S. J., et al. (2016). Evaluation of the causes of on-farm harvest losses in canola in the northern great plains. *Crop Sci.* 56, 2005–2015. doi: 10.2135/cropsci2016.01.0014

Chang, S., Thurber, C. S., Brown, P. J., Hartman, G. L., Lambert, K. N., and Domier, L. L. (2014). Comparative mapping of the wild perennial *Glycine latifolia* and Soybean (*G. max*) reveals extensive chromosome rearrangements in the genus *Glycine*. *PLoS One* 9:e99427. doi: 10.1371/journal.pone.0099427

Chapman, M. A., and Burke, J. M. (2012). Evidence of selection on fatty acid biosynthetic genes during the evolution of cultivated sunflower. *Theor. Appl. Genet.* 125, 897–907. doi: 10.1007/s00122-012-1881-z

Churchill, G. A., and Doerge, R. W. (1994). Empirical threshold value for quantitative trait mapping. *Genetics* 138, 963–971. doi: 10.1534/genetics.107.080101

Cox, T. S., Bender, M., Picone, C., Van Tassel, D. L., Holland, J. B., Brummer, E. C., et al. (2002). Breeding perennial grain crops. *Crit. Rev. Plant Sci.* 21, 59–91.

Cox, T. S., Van Tassel, D. L., Cox, C. M., and Dehaan, L. R. (2010). Progress in breeding perennial grains. *Crop Pasture Sci.* 61, 513–521. doi: 10.1071/CP09201

Dedio, W. (1980). Comparison of achene characteristics and combining ability of branching and non-branching near-isogenic sunflower restorer lines. *Crop Sci.* 20, 189–190. doi: 10.2135/cropsci1980.0011183X002000020010x

DeHaan, L. R., Van Tassel, D. L., Anderson, J. A., Asselin, S. R., Barnes, R., Baute, G. J., et al. (2016). A pipeline strategy for grain crop domestication. *Crop Sci.* 56, 917–930. doi: 10.2135/cropsci2015.06.0356

Dempewolf, H., Tesfaye, M., Teshome, A., Bjorkman, A. D., Andrew, R. L., Scascitelli, M., et al. (2015). Patterns of domestication in the Ethiopian oil-seed crop noug (*Guizotia abyssinica*). *Evol. Appl.* 8, 464–475. doi: 10.1111/eva.12256

Dietz, D. R., Wasser, C. H., Dittberner, P. L., and Martin, C. O. (1992). *Defense Natural Resources Program: Maximilian Sunflower (Helianthus maximiliani), Section 7.4. 3*. Washington, D.C: US Army Corps of Engineers Wildlife Resources Management.

Dinno, A. (2012). *paran: Horn's Test of Principal Components/Factors.* R package version 1.1.

Dorrell, D. G., and Whelan, E. D. P. (1978). Chemical and Morphological Characteristics of Seeds of Some Sunflower Species 1. *Crop Sci.* 18, 969–971. doi: 10.2135/cropsci1978.0011183X001800060015x

Doyle, J. J. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* 19, 11–15.

Edae, E. A., Olivera, P. D., Jin, Y., Poland, J. A., and Rouse, M. N. (2016). Genotype-by-sequencing facilitates genetic mapping of a stem rust resistance locus in *Aegilops umbellulata*, a wild relative of cultivated wheat. *BMC Genomics* 17:1039. doi: 10.1186/s12864-016-3370-2

Fick, G. N., Zimmer, D. E., and Zimmerman, D. C. (1974). Correlation of seed oil content in sunflowers with other plant and seed characteristics. *Crop Sci.* 14, 755–757. doi: 10.2135/cropsci1974.0011183x001400050042x

Fletcher, J. C. (2001). The ULTRAPETALA gene controls shoot and floral meristem size in *Arabidopsis*. *Development* 128, 1323–1333.

Gandhi, S. D., Heesacker, A. F., Freeman, C. A., Argyris, J., Bradford, K., and Knapp, S. J. (2005). The self-incompatibility locus (S) and quantitative trait loci for self-pollination and seed dormancy in sunflower. *Theor. Appl. Genet.* 111, 619–629. doi: 10.1007/s00122-005-1934-7

Gulden, R. H., Shirtliffe, S. J., and Thomas, A. G. (2003). Harvest losses of canola (*Brassica napus*) cause large seedbank inputs. *Weed Sci.* 51, 83–86. doi: 10.1614/0043-1745(2003)051[0083:hlocbn]2.0.co;2

Heiser, C. B. (1951). The sunflower among the North American Indians. *Proc. Am. Philos. Soc.* 95, 432–448.

Heiser, C. B., Smith D. M., Clevenger, S. B., and Martin, W. C. (1969). "The North American sunflowers (*Helianthus*)," in *Memoirs of the Torrey Botanical Club*, Vol. 22 (Durham, NC: Club by the Seeman Printery), 1–218.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika* 30, 179–185. doi: 10.1007/BF02289447

INRA Sunflower Bioinformatics Resources (2014). *HA412-HO Genome Portal (Bronze version)*. Available online at: https://www.heliagene.org/HA412.v1.1.bronze.20141015/

Isbell, F., Calcagno, V., Hector, A., Connolly, J., Harpole, W. S., Reich, P. B., et al. (2011). High plant diversity is needed to maintain ecosystem services. *Nature* 477, 199–202. doi: 10.1038/nature10282

Jackson, W. (1990). "Agriculture with nature as analogy," in *Sustainable Agriculture in Temperate Zones*, eds C. A. Francis, C. Butler, and L. D. King (New York, NY: John Wiley & Sons, Inc), 381–422.

Jan, C. C. (1997). Cytology and interspecific hybridization. *Sunflow. Technol. Prod.* 35, 497–558. doi: 10.2134/agronmonogr35.c10

Kawakami, T., Darby, B. J., and Ungerer, M. C. (2014). Transcriptome resources for the perennial sunflower *Helianthus maximiliani* obtained from ecologically divergent populations. *Mol. Ecol. Resour.* 14, 812–819. doi: 10.1111/1755-0998.12227

Kawakami, T., Morgan, T. J., Nippert, J. B., Ocheltree, T. W., Keith, R., Dhakal, P., et al. (2011). Natural selection drives clinal life history patterns in the perennial sunflower species, *Helianthus maximiliani*. *Mol. Ecol.* 20, 2318–2328. doi: 10.1111/j.1365-294X.2011.05105.x

Lander, E. S., and Botstein, D. (1989). Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 136:705.

Li, J. T., Yang, J., Chen, D. C., Zhang, X. L., and Tang, Z. S. (2007). An optimized mini-preparation method to obtain high-quality genomic DNA from mature leaves of sunflower. *Genet. Mol. Res.* 6, 1064–1071.

Lin, Z., Li, X., Shannon, L. M., Yeh, C.-T., Wang, M. L., Bai, G., et al. (2012). Parallel domestication of the Shattering1 genes in cereals. *Nat. Genet.* 44, 720–724. doi: 10.1038/ng.2281

Liu, Z., Cai, X., Seiler, G. J., Gulya, T. J., Rashid, K. Y., and Jan, C. C. (2011). "Transferring sclerotinia stalk rot resistance genes from wild *Helianthus* species into cultivated sunflower," in *Proceedings of the 33rd Sunflower Research Forum*, Fargo, ND, 12–13.

Lobell, D. B., Schlenker, W., and Costa-Roberts, J. (2011). Climate trends and global crop production since 1980. *Science* 333, 616–620. doi: 10.1126/science.1204531

Mandel, J. R., Nambeesan, S., Bowers, J. E., Marek, L. F., Ebert, D., Rieseberg, L. H., et al. (2013). Association mapping and the genomic consequences of selection in sunflower. *PLoS Genet.* 9:e1003378. doi: 10.1371/journal.pgen.1003378

Nambeesan, S. U., Mandel, J. R., Bowers, J. E., Marek, L. F., Ebert, D., Corbi, J., et al. (2015). Association mapping in sunflower (*Helianthus annuus* L.) reveals

independent control of apical vs. basal branching. *BMC Plant Biol.* 15:84. doi: 10.1186/s12870-015-0458-9

Ott, A., Liu, S., Schnable, J. C., Yeh, C. T. E., Wang, K. S., and Schnable, P. S. (2017). tGBS® genotyping-by-sequencing enables reliable genotyping of heterozygous loci. *Nucleic Acids Res.* 45:e178. doi: 10.1093/nar/gkx853

Pearl, S. A., Bowers, J. E., Reyes-Chin-Wo, S., Michelmore, R. W., and Burke, J. M. (2014). Genetic analysis of safflower domestication. *BMC Plant Biol.* 14:43. doi: 10.1186/1471-2229-14-43

Peng, Q., Hu, Y., Wei, R., Zhang, Y., Guan, C., Ruan, Y., et al. (2010). Simultaneous silencing of FAD2 and FAE1 genes affects both oleic acid and erucic acid contents in *Brassica napus* seeds. *Plant Cell Rep.* 29, 317–325. doi: 10.1007/s00299-010-0823-y

Picasso, V. D., Brummer, E. C., Liebman, M., Dixon, P. M., and Wilsey, B. J. (2008). Crop species diversity affects productivity and weed suppression in perennial polycultures under two management strategies. *Crop Sci.* 48, 331–342. doi: 10.2135/cropsci2007.04.0225

Picasso, V. D., Brummer, E. C., Liebman, M., Dixon, P. M., and Wilsey, B. J. (2011). Diverse perennial crop mixtures sustain higher productivity over time based on ecological complementarity. *Renew. Agric. Food Syst.* 26, 317–327. doi: 10.1017/s1742170511000135

Purugganan, M. D., and Fuller, D. Q. (2009). The nature of selection during plant domestication. *Nature* 457, 843–848. doi: 10.1038/nature07895

Ronin, Y., Mester, D., Minkov, D., and Korol, A. (2010). Building reliable genetic maps: different mapping strategies may result in different maps. *Nat. Sci.* 02, 576–589. doi: 10.4236/ns.2010.26073

Seiler, G. J. (1994). Oil concentration and fatty acid composition of achenes of North American Helianthus (*Asteraceae*) species. *Econ. Bot.* 48, 271–279. doi: 10.1007/bf02862328

Seiler, G. J., and Brothers, M. E. (1999). Oil concentration and fatty acid composition of achenes of *Helianthus* species (*Asteraceae*) from Canada. *Econ. Bot.* 53, 273–280. doi: 10.1007/bf02866637

Shimizu-Sato, S., Tanaka, M., and Mori, H. (2009). Auxin-cytokinin interactions in the control of shoot branching. *Plant Mol. Biol.* 69, 429–435. doi: 10.1007/s11103-008-9416-3

Simons, K. J. (2005). Molecular characterization of the major wheat domestication gene Q. *Genetics* 172, 547–555. doi: 10.1534/genetics.105.044727

Skinner, M. E., Uzilov, A. V., Stein, L. D., Mungall, C. J., and Holmes, I. H. (2009). JBrowse: a next-generation genome browser. *Genome Res.* 19, 1630–1638. doi: 10.1101/gr.094607.109

Stirnberg, P., van De Sande, K., and Leyser, H. M. O. (2002). MAX1 and MAX2 control shoot lateral branching in *Arabidopsis*. *Development* 129, 1131–1141.

Tang, S. X., Leon, A., Bridges, W. C., and Knapp, S. J. (2006). Quantitative trait loci for genetically correlated seed traits are tightly linked to branching and pericarp pigment loci in sunflower. *Crop Sci.* 46, 721–734. doi: 10.2135/cropsci2005.0006-7

Taski-Ajdukovic, K., Vasic, D., and Nagl, N. (2006). Regeneration of interspecific somatic hybrids between *Helianthus annuus* L. and *Helianthus maximiliani* (Schrader) via protoplast electrofusion. *Plant Cell Rep.* 25, 698–704. doi: 10.1007/s00299-006-0134-5

Tetreault, H. M., Kawakami, T., Ungerer, M. C., and Levy, C. (2016). Low temperature tolerance in the perennial sunflower *Helianthus maximiliani*. *Am. Midl. Nat.* 175, 91–102. doi: 10.1674/amid-175-01-91-102.1

Texas Agricultural Experiment Station (1979). Aztec maximilian sunflower. *Leaflet* 1775, 1–4.

USDA-NRCS, (2017). *The PLANTS Database*. Greensboro, NC: National Plant Data Team.

Van Ooijen, J. W. (2018). JoinMap ® 5, Software for the Calculation of Genetic Linkage Maps in Experimental Populations of Diploid Species. Wageningen: Kyazma B.V.

Van Tassel, D. L., Asselin, S. R., Cox, S. A., Sideli, G., and Cattani, D. J. (2014). "Evaluating perennial candidates for domestication: lessons from wild sunflower relatives," in *Proceedings of the FAO Expert Workshop: Perennial Crops for Food Security. 28–30 Aug. 2013*, eds C. Batello, L. Wade, S. Cox, N. Pagna, A. Bozzini, and J. Choptiany (Rome: FAO), 112–140.

Vavilov, N. I. (1922). The law of homologous series in variation. *J. Genet.* 12, 47–89. doi: 10.1007/bf02983073

Wagoner, P., and Schaeffer, J. R. (1990). Perennial grain development: past efforts and potential for the future. *Crit. Rev. Plant Sci.* 9, 381–408. doi: 10.1080/07352689009382298

Wills, D. M., and Burke, J. M. (2007). Quantitative trait locus analysis of the early domestication of sunflower. *Genetics* 176, 2589–2599. doi: 10.1534/genetics.107.075333

Wu, T. D., and Nacu, S. (2010). Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* 26, 873–881. doi: 10.1093/bioinformatics/btq057

Zhang, L., Wang, S., Li, H., Deng, Q., Zheng, A., Li, S., et al. (2010). Effects of missing marker and segregation distortion on QTL mapping in F2 populations. *Theor. Appl. Genet.* 121, 1071–1082. doi: 10.1007/s00122-010-1372-z

Zohary, D. (2004). Unconscious selection and the evolution of domesticated plants. *Econ. Bot.* 58, 5–10. doi: 10.1663/0013-0001(2004)058[0005:usateo]2.0.co;2

Conflict of Interest: 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 © 2020 Asselin, Brûlé-Babel, Van Tassel and Cattani. 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) and the copyright owner(s) 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 Plant Science

PERSPECTIVE
published: 19 June 2020
doi: 10.3389/fpls.2020.00818

Common Vetch: A Drought Tolerant, High Protein Neglected Leguminous Crop With Potential as a Sustainable Food Source

*Vy Nguyen[1,2], Samuel Riley[1,2], Stuart Nagel[3], Ian Fisk[4] and Iain R. Searle[1,2]**

[1] School of Biological Sciences, School of Agriculture, Food and Wine, The University of Adelaide, Adelaide, SA, Australia, [2] Shanghai Jiao Tong University Joint International Centre for Agriculture and Health, The University of Adelaide, Adelaide, SA, Australia, [3] South Australian Research and Development Institute, Adelaide, SA, Australia, [4] Division of Food Science, Nutrition and Dietetics, School of Biosciences, University of Nottingham, Nottingham, United Kingdom

Global demand for protein is predicted to increase by 50% by 2050. To meet the increasing demand whilst ensuring sustainability, protein sources that generate low-greenhouse gas emissions are required, and protein-rich legume seeds have the potential to make a significant contribution. Legumes like common vetch (*Vicia sativa*) that grow in marginal cropping zones and are drought tolerant and resilient to changeable annual weather patterns, will be in high demand as the climate changes. In common vetch, the inability to eliminate the γ-glutamyl-β-cyano-alanine (GBCA) toxin present in the seed has hindered its utility as a human and animal food for many decades, leaving this highly resilient species an "orphan" legume. However, the availability of the vetch genome and transcriptome data together with the application of CRISPR-Cas genome editing technologies lay the foundations to eliminate the GBCA toxin constraint. In the near future, we anticipate that a zero-toxin vetch variety will become a significant contributor to global protein demand.

Keywords: legume, common vetch, *Vicia sativa*, vetch toxin, γ-glutamyl-β-cyano-alanine, plant-based protein, sustainable

OPEN ACCESS

Edited by:
Thomas M. Davis, University of New Hampshire, United States

Reviewed by:
Hamid Khazaei, University of Saskatchewan, Canada
Tomas Vymyslicky, Agricultural Research, Ltd., Czechia

***Correspondence:**
Iain R. Searle
iain.searle@adelaide.edu.au

Specialty section:
This article was submitted to Plant Breeding, a section of the journal Frontiers in Plant Science

Received: *27 February 2020*
Accepted: *21 May 2020*
Published: *19 June 2020*

Citation:
Nguyen V, Riley S, Nagel S, Fisk I and Searle IR (2020) Common Vetch: A Drought Tolerant, High Protein Neglected Leguminous Crop With Potential as a Sustainable Food Source. Front. Plant Sci. 11:818. doi: 10.3389/fpls.2020.00818

INTRODUCTION

Global demand for protein is predicted to increase by a staggering 50% by 2050 (Westhoek et al., 2011; Henchion et al., 2017). With an increasing global population and increasing demand for animal-derived protein, the sustainability of agriculture systems has been brought into question. Over the last two centuries, the expanding livestock industry has led to significant deforestation, and overgrazing of natural grassland environments such that it has caused decreased terrestrial biodiversity and increased greenhouse gas emissions and contributed to climate change and global warming (Henchion et al., 2017). In order to meet the increasing protein demand and protect our environment, more sustainable protein food sources are required. Cheap plant-based protein, such as legume seeds, represent an environmentally sustainable option that is well suited for developing countries with rapidly growing populations (Asgar et al., 2010). Moreover, to cope with increasingly unpredictable climate change and expansion of marginal cropping areas, breeding strategies for

more drought tolerant and resilient crops will be vital (Sivakumar et al., 2005; Lobell and Gourdji, 2012). A legume that could be exploited for this scenario is the common vetch (*V. sativa*). Common vetch is able to grow in marginal cropping zones whilst being resilient to variable annual weather patterns mainly through superior drought tolerance (White et al., 2005). One study demonstrated that vetch could withstand water deficit for up to 24 days and show full restoration of biotic function once regular watering had resumed (Tenopala et al., 2012). Drought and heat tolerant crops are increasingly desirable in the face of rising global temperatures and increasingly prolonged periods of drought brought on by climate change (Mba et al., 2018).

For many decades, the inability to remove the γ-glutamyl-β-cyano-alanine (GBCA) seed toxin has hindered common vetch's use in agriculture (Pfeffer and Ressler, 1967; Roy et al., 1996) leaving this resilient plant as an "orphan" legume. We envisage that the development of a zero-toxin vetch variety would facilitate its use for animal feed, specifically chickens and pigs, and human consumption (Ressler et al., 1997; Collins et al., 2002). We estimate the production costs of common vetch are approximately 50% less than competing legumes such as lentils and predict that zero-toxin varieties would rapidly surpass lentil in pig and poultry production. Zero-toxin common vetch will immediately generate new domestic markets, such as feed for the poultry industry, but it will also open new export markets for Australia and other countries thereby increasing export revenue and increasing farm profitability and indirectly increase investment for their local communities.

COMMON VETCH: A VERSATILE PASTURE CROP THAT PROVIDES MULTIPLE BENEFITS FOR THE FARM

Common vetch (*V. sativa*) which is shown in **Figure 1** belongs to the *Fabaceae* (legume) family, within the genus *Vicia*. This genus contains about 140 species including woolly-pod vetch (*V. villosa*) and faba bean (*V. faba*). Other *Fabaceae* genera also contain so-called vetches; of which two examples are *Astragalus* (containing the milkvetches) and *Lathyrus* (containing *L. ochrus*, the cyprus-vetch). Nowadays, common vetch is commonly found both in natural and agricultural settings across Europe, Asia, North America, some parts of South America, Africa, the Mediterranean, and Australia (Navrátilová et al., 2003; Ford et al., 2008).

Like other legumes, common vetch forms a symbiosis with nitrogen-fixing bacteria (*Rhizobia*) that fix atmospheric nitrogen into nitrogenous compounds available to the plant, hence reducing the need for application of expensive nitrogen fertilizer and subsequent rotation crops. Often, common vetch is used as a green manure which, when incorporated into the soil, provides valuable carbon, and nitrogen for rotation crops such as wheat and barley. Additional soil carbon often increases water-holding capacity and ability to bind nutrients including nitrate (Reeves, 1997; Bünemann et al., 2018). Furthermore, common vetch biomass can also be used for forage, fodder, pasture, silage, or hay and the seed may safely be used as a protein-rich feed component for ruminant animals (Enneking, 1995). Common vetch is well suited as a pasture species as it forms many adventitious shoots that are either buried or close to the soil surface thus giving it the ability to be resilient to heavy grazing (Rathjen, 1997).

Despite common vetch's versatile uses, the production of vetch is still limited. Data collected by the Food and Agriculture Organization (FAO) in 2017 showed that, globally, the area harvested and the production of vetch were about 0.6 million ha and 0.9 million tons, respectively[1]. Based on FAO 2017 data for legumes, vetch occupied about 0.3% of land usage and accounted for only 0.2% of the production. This is 12 times less than the area harvested and 8 times less than the production of lentils (**Figure 2**). Common vetch's limited production is mainly attributed to the anti-nutritional compounds existing in the seeds which will be further discussed in the next section.

COMMON VETCH SEEDS: IMPORTANT NUTRITIONAL ATTRIBUTES

Common vetch seed appeared in the diets of hunter-gatherers as early as 12,000–9,000 BP as evident in archaeobotanical analysis of samples from the Santa Maira cave in Alicante, Spain (Aura et al., 2005; Mikić, 2016). Today, common vetch is globally distributed and its spread is thought to have occurred through the inadvertent selection and trading of vetch seeds as a weedy contaminant with other legume seeds (Erskine et al., 1994). This has led to the suggestion that selection of common vetch amongst other palatable legumes, such as lentils, led to vegetative, and seed mimicry of vetch to lentils (Erskine et al., 1994). Although intact lentil and vetch seed can be differentiated after close examination of the seed shape and size, the split seed is harder to distinguished. Due to the similarity of split vetch and lentil seeds, there was a period of time when split vetch seeds of the variety "Blanchefleur" were oil coated and inappropriately substituted and sold as lentils. However, Tate and Enneking (1992) raised the so-called substitution issue of vetch toxicity and indicated that vetch seeds were unsuitable to be eaten by monogastric animals, including humans. Since then, vetch has been restricted to use in pastures as the seeds can be safely eaten by ruminants and some monogastric animals, such as chicken (less than 40% w/w), and pig (less than 20% w/w) in low amounts (Rathjen, 1997).

Common vetch exhibits high concentrations of crude protein (w/w) in the seed between 24 to 32% (Francis et al., 2000). This concentration is comparable to faba bean (*V. faba*) and lupin (*Lupinus angustifolius L.*) seed which have about 32% (Crépon et al., 2010; Caliskanturk et al., 2017) and 30% crude protein (Valentine and Bartsch, 1996), respectively. Common vetch seeds contain eighteen amino acids and the ratio of essential amino acids/non-essential amino acids is about 0.7 (Mao et al., 2015) which is significantly higher than the 0.38 recommended by WHO (Joint, 2007). Additionally, the principle essential amino acids are arginine and leucine at average concentrations of 2.4 and 2.1%, respectively. Glutamic and aspartic acids are the

[1] http://www.fao.org/faostat/en/#data



FIGURE 1 | Field grown *Vicia sativa* growing at the National Vetch Breeding Program in South Australia.

predominant non-essential amino acids within common vetch seed, averaging levels of 5.5 and 3.7%, respectively (Mao et al., 2015). Vetch seed also contains a much lower proportion of lipids (only 1.5–2.7%) compared with soybean seed (*Glycine max*) and this is primarily composed of unsaturated fatty acids (Mao et al., 2015).

Presently, common vetch is mainly used as animal feed for ruminants and it can be argued that higher value returns could be observed through the repurposing of common vetch as a human food crop. Obstructing the development of the higher value product is a range of anti-nutritional factors, the most significant of which are the dipeptide GBCA and the free amino acid β-cyano-L-alanine (BCA), which exist in relatively high concentrations in the seed at approximately 2.6 and 0.9%, respectively (Tate et al., 1999). Such compounds are toxic to monogastric species, such as chickens or pigs, but have no obvious effect upon ruminant species, including beef cattle (Valentine and Bartsch, 1996). The proportion (w/w) of common vetch seed within feeds that can be safely consumed without deleterious effects is 10% for piglets and 20% for adult pigs and for chickens about 40% (Harper and Arscott, 1962; Rathjen, 1997). In addition to the main toxin GBCA, other anti-nutritional compounds were also found in common vetch included vicianine, vicine, convicine, and tannins (Ritthausen and Kreusler, 1870; Rathjen, 1997).

Genetic variation and relatively high heritability of vicianine levels in common vetch accessions have allowed breeders to select and produce cultivars free of this toxin. One early developed cultivar with no vicianine was Blanchefluer (Delaere, 1996; Rathjen, 1997). Most modern common vetch cultivars are vicianine free. Vicine and convicine are well-studied in faba bean, and in humans cause the potentially fatal disease favism (Bottini, 1973). Vicine and convicine are hydrolysed by native β-glucosidases in the cotyledons of seeds to form divicine and isouramil and when consumed by humans can cause oxidation of glutathione in red blood cells. In individuals who cannot generate glutathione at normal rates due to a deficient glucose-6-phophatedehydrogenase activity this results in haemolysis and the disease favism (Arese et al., 1981). Unlike vicianine and vicine, wild accessions and breeding lines with very low GBCA levels have not been identified, and the GBCA toxin levels in current cultivars is still deemed to be too high for monogastric consumption. Chickens can only tolerate feed with less than 20% (w/w) common vetch (∼0.2% GBCA in the feed). Confounding toxic effects of other compounds with vicine and GBCA were proposed (Rathjen, 1997), however, the chemical basis of these compounds is currently unknown. Finally, anti-nutritional tannins present in vetch seed coats are often removed during the dehulling process and hence are considered less significant (Rathjen, 1997).

LIMITATION OF GBCA DETOXIFICATION USING CONVENTIONAL METHODS

Without the toxic compounds in the seeds, vetch would be highly nutritious and a valuable animal feed. Therefore, a number of methods have been investigated to remove the toxins, mainly

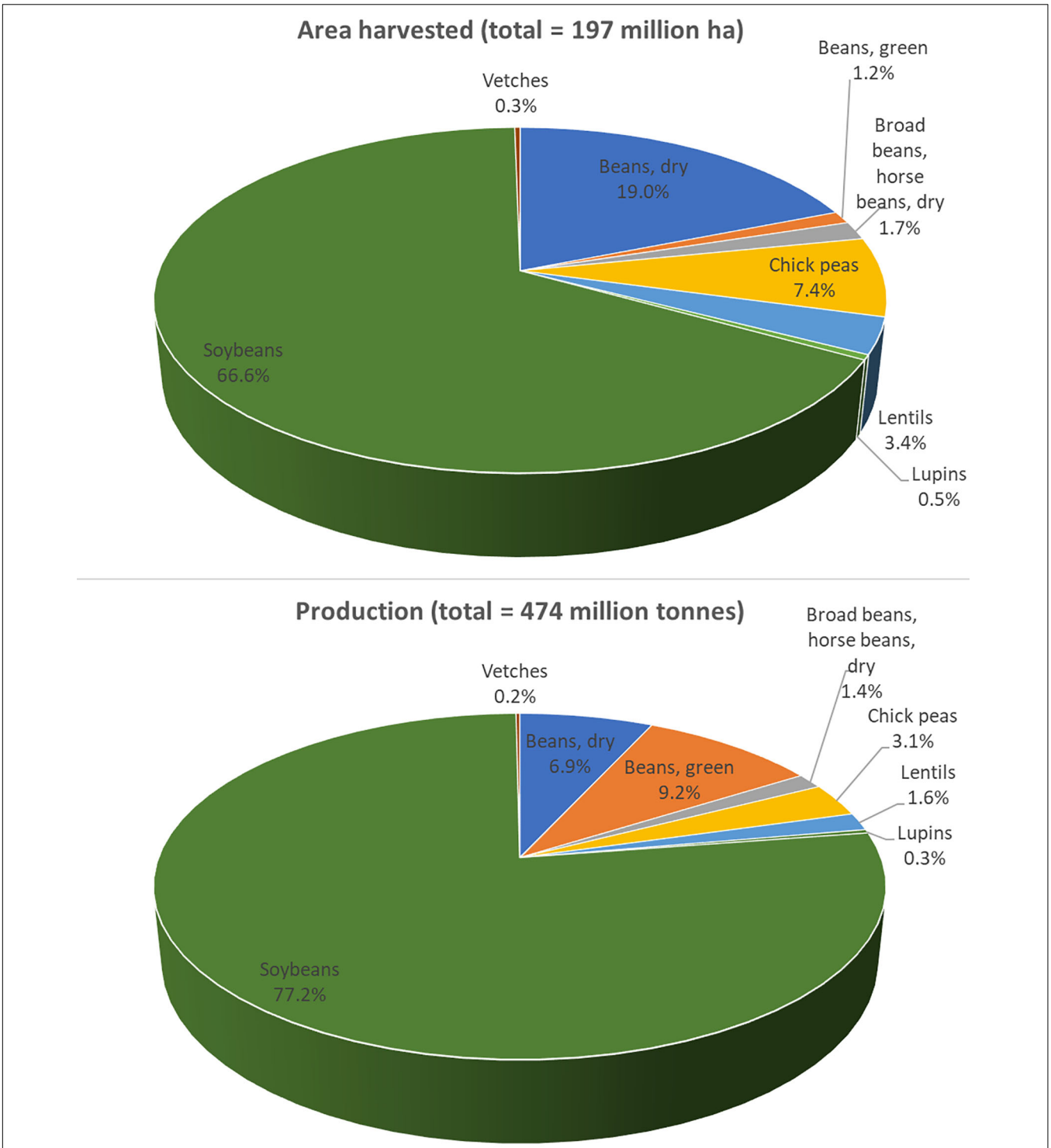


FIGURE 2 | Common vetch production when compared to other legumes globally in 2017. In general, common vetch occupied about 0.3% of the area harvested and 0.2% of the production. For vetch, total area harvested in 2017 was 560,077 ha and total production was 920,536 tons. Data was published by the Food and Agriculture Organization (FAO; http://www.fao.org/).

focusing on GBCA. Post-harvest processing efforts to lower the GBCA toxin levels within the seed have previously involved simple soaking, continuous flow through soaking, and boiling methods (Rathjen, 1997). The seed soaking method alone was insufficient to lower GBCA levels as consumption of soaked seed during feeding trials by chickens reduced egg production, and

daily food consumption and feed conversion ratios were also diminished (Farran et al., 1995). In contrast, the boiling method reduced the toxin levels in the seed such that the seed could be included in chicken feed at levels of up to 25% without negative effects on growth rates (Kaya et al., 2013). However, consumption of boiled vetch seed that had the broth periodically discarded during boiling resulted in 20% reduced growth rates in chickens when compared with conventional feed with similar protein amounts (Ressler et al., 1997). This boiling method combined with periodically discarded water had 45% decreased seed mass as water soluble vitamins, like vitamin B, water soluble proteins and carbohydrates were leached during processing (Ressler et al., 1997), and this correlated with the reduced chick growth rate. Autoclaving the seed as a processing method has also been investigated and assessed in feeding trials of laying hens. In the laying hen trial, overall growth rate was found not significantly different between animals fed with autoclaved or raw vetch seed suggesting that the non-heat labile toxins like GBCA were still bioactive (Farran et al., 1995). The lack of success of these seed processing methods in improving animal growth or health has strongly indicated the need for genetic approaches to detoxify the common vetch seeds.

UNSUCCESSFUL SEARCHES FOR ZERO-TOXIN VETCH ACCESSIONS

Using conventional breeding methods and more recently the use of molecular marker-assisted breeding for genomic selection, plant breeders have prioritized the search for common vetch varieties that have biotic and abiotic stress resistant traits as well as selecting for increasing yield and seed nutritional quality (Francis et al., 2000). However, no concerted effort has been made to select for low or zero GBCA toxin levels in Australian or overseas breeding programs. This has resulted in varieties that are only used by farmers for pasture, green or brown silage, or ruminant feed (Francis et al., 2000; Dong et al., 2019; Huang et al., 2019; Mikić et al., 2019). This is mainly due to very limited natural variation in GBCA toxin levels amongst common vetch accessions (Rathjen, 1997). Rathjen (1997) screened over 1,700 *V. sativa* accessions and failed to identify a single accession with no GBCA toxin. Later, screening of a total of 3,000 accessions identified only one line, IR28, with a low (0.3–0.4%) GBCA level but no zero-toxin line has yet been identified (Ford et al., 2008). Backcrossing IR28 to Jericho white, a spontaneous white flowered mutant of the French commercial variety Languedoc, over seven generations produced a near homozygous line named Lov 9 (Tate and Searle, unpublished). However, the GBCA levels in the Lov 9 seed from plants grown in shade houses or field conditions ranged from 0.4–1.2%, respectively (Tate and Searle, unpublished). These GBCA levels in Lov 9 seed were deemed too high for commercial release of the variety as a low toxin variety. Another strategy to develop a zero GBCA toxin common vetch variety was interspecies crosses of zero toxin species *V. villosa* and *V. pannonica* to common vetch but these resulted in embryo abortion and no viable hybrids were recovered (Searle, unpublished). Considering the limited success to date, other pathways to produce a zero-toxin common vetch variety are required.

APPLICATION OF BIOTECHNOLOGY TO PRODUCE ZERO-TOXIN VETCH

Applications of biotechnology have promised to accelerate crop improvement (Moose and Mumm, 2008). The emergence of genomics, transcriptomics, metabolomics, and proteomics data has led to the establishment of publicly available databases for most major crops. For example, the LIS – Legume Information System (Dash et al., 2015), and eFP browser (Patel et al., 2012; Hawkins et al., 2017) now contain data for legumes. By combining the information available in these databases with new bioinformatic tools, we now have the ability to dissect complex traits to determine the underlying gene architecture in a more comprehensive way. In 2018, the 1.8 Gb common vetch genome and seed transcriptome sequencing projects were initiated at the University of Adelaide, Australia, opening the opportunity to determine the genetic basis of the vetch toxin accumulation. Using this transcriptome data, we could identify the genes involved in toxin production and in the future we could investigate their functions by overexpressing or mutating candidate genes. Moreover, we envisage that application of CRISPR-Cas (clustered regularly interspaced short palindromic repeats – Cas protein) genome editing to modify agronomically important traits in crops such as wheat, barley, rice, and tomato (Liu et al., 2017) will soon be applied to more challenging species including the common vetch. Using CRISPR-Cas genome editing, the nutritional profiles of many crops have been recently demonstrated. For example, in tomato, knocking down genes in the carotenoid metabolic pathway led to a 5-fold increase in lycopene (Li et al., 2018), and in rice, generating mutations in the starch branching enzyme (Berrens et al., 2017) genes increased amylose content by up to 25% and resistant starch to 9.8% (Sun et al., 2017). One of the most significant impacts of CRISPR-Cas genome editing is the potential improvement of a key trait in a commercially released cultivar within 6 months. In contrast conventional breeding of the trait may take 5–7 years to release the new cultivar. Importantly, it only takes one generation to obtain an edited plant using genome editing.

A major challenge in non-model plant systems like common vetch, is the delivery of transgenes such as CRISPR-Cas ribonuclear complex into plant cells and subsequent plant regeneration. Unlike crops such as rice and barley where the transformation and plant regeneration systems are standardized (Sahoo et al., 2011; Harwood, 2014), efficient transformation, and plant regeneration systems are lacking for common vetch (Ford et al., 2008). Unfortunately, common vetch's GBCA toxin level have lowered the priority for research and development of these necessary biotechnological tools – for example developing a transformation system. Further investment in common vetch is required to develop transformation and plant regeneration systems to facilitate the application of genome editing for trait improvement.

FUTURE WORK AND EXPECTATIONS

The environmental benefits, the versatile growth habit, and the rich nutritional profile of common vetch make the legume an appealing crop to meet future protein food requirements for humans and animals while sustainably contributing to our agricultural system. However, the failure of conventional breeding to develop a zero-toxin common vetch variety requires a new strategy to be employed. The recent availability of new vetch genomic resources and tools for genome editing increase the likelihood of solving the vetch GBCA toxicity problem. To make this plausible, we suggest the following steps are important for vetch.

To solve the vetch GBCA toxicity, we suggest the following experiments should be prioritized:

1. Sequencing, assembling and annotating the genome.
2. Comparing gene expression by RNA-seq and undertake biochemical pathway analysis of differentially abundant genes across vetch cultivars and *Vicia* species that do not contain GBCA.
3. Identifying candidate GBCA detoxifying genes and undertake functional gene analysis.
4. Establishing an efficient transformation and plant regeneration system.
5. Applying CRISPR-Cas9 or gene expression cassettes of candidate gene(s) to manipulate gene expression levels.
6. Screening of zero-toxin vetch lines and testing in animal assays.

Although the first three objectives above can be readily achieved, the last three could take many years to be accomplished due to the absence of a demonstrably routine *in vitro* transformation and regeneration system for vetch. Therefore, expertise from the plant tissue culture field will be invaluable to achieving objective four and ultimately the final goal.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

AUTHOR CONTRIBUTIONS

VN and IS initially conceived the manuscript. All authors contributed to the writing and editing of the manuscript.

ACKNOWLEDGMENTS

We would like to thank the Hermon Slade Foundation and The University of Adelaide for the initial funding awarded to IS and VN, respectively. We also thank Jeremy Timmis for useful feedback and editing of the manuscript. This manuscript has been released as a pre-print at https://doi.org/10.1101/2020.02.11.943324 (Nguyen et al., 2020).

REFERENCES

Arese, P., Bosia, A., Naitana, A., Gaetani, S., D'aquino, M., and Gaetani, G. F. (1981). Effect of divicine and isouramil on red cell metabolism in normal and G6PD-deficient (Mediterranean variant) subjects. Possible role in the genesis of favism. *Prog. Clin. Biol. Res.* 55, 725–746.

Asgar, M. A., Fazilah, A., Huda, N., Bhat, R., and Karim, A. A. (2010). Nonmeat protein alternatives as meat extenders and meat analogs. *Comprehen. Rev. Food Sci. Food Saf.* 9, 513–529. doi: 10.1111/j.1541-4337.2010.00124.x

Aura, J. E., Carrión, Y., Estrelles, E., and Jorda, G. P. (2005). Plant economy of hunter-gatherer groups at the end of the last Ice Age: plant macroremains from the cave of Santa Maira (Alacant, Spain) ca. 12000–9000 BP. *Veg. History Archaeobot.* 14, 542–550. doi: 10.1007/s00334-005-0002-1

Berrens, R. V., Andrews, S., Spensberger, D., Santos, F., Dean, W., Gould, P., et al. (2017). An endosiRNA-based repression mechanism counteracts transposon activation during global DNA demethylation in embryonic stem cells. *Cell Stem Cell* 21, 694–703.

Bottini, E. (1973). Favism: current problems and investigations. *J. Med. Genet.* 10, 154–157. doi: 10.1136/jmg.10.2.154

Bünemann, E. K., Bongiorno, G., Bai, Z., Creamer, R. E., De Deyn, G., De Goede, R., et al. (2018). Soil quality–A critical review. *Soil Biol. Biochem.* 120, 105–125.

Caliskanturk, K. S., Gunay, D., and Sayar, S. (2017). In vitro evaluation of whole faba bean and its seed coat as a potential source of functional food components. *Food Chem.* 230, 182–188. doi: 10.1016/j.foodchem.2017.03.037

Collins, C. L., Henman, D. J., King, R. H., and Dunshea, F. R. (2002). Common vetch (*Vicia sativa* cv Morava) is an alternative protein source in pig diets. *Asia Pac. J. Clin. Nutr.* 11, S249–S249.

Crépon, K., Marget, P., Peyronnet, C., Carrouée, B., Arese, P., and Duc, G. (2010). Nutritional value of faba bean (*Vicia faba* L.) seeds for feed and food. *Field Crops Res.* 115, 329–339. doi: 10.1016/j.fcr.2009.09.016

Dash, S., Campbell, J. D., Cannon, E. K. S., Cleary, A. M., Huang, W., Kalberer, S. R., et al. (2015). Legume information system (LegumeInfo. org): a key component of a set of federated data resources for the legume family. *Nucleic Acids Res.* 44, D1181–D1188.

Delaere, I. (1996). *The Chemistry of Vicia sativa selection*. Doctor of Philosophy, The University of Adelaide, Adelaide.

Dong, R., Shen, S. H., Jahufer, M. Z. Z., Dong, D. K., Luo, D., Zhou, Q., et al. (2019). Effect of genotype and environment on agronomical characters of common vetch (*Vicia sativa* L.). *Genet. Resour. Crop Evol.* 66, 1–13.

Enneking, D. (1995). *The Toxicity of Vicia species and Their Utilization as Grain Legumes*. NedlandsW. A: University ofWestern Australia.

Erskine, W., Smartt, J., and Muehlbauer, F. J. (1994). Mimicry of lentil and the domestication of common vetch and grass pea. *Econ. Bot.* 48, 326–332. doi: 10.1007/bf02862334

Farran, M. T., Uwayjan, M. G., Miski, A. M. A., Sleiman, F. T., Adada, F. A., Ashkarian, V. M., et al. (1995). Effect of feeding raw and treated common vetch seed (*Vicia sativa*) on the performance and egg quality parameters of laying hens. *Poult. Sci.* 74, 1630–1635. doi: 10.3382/ps.0741630

Ford, R., Maddeppungeng, A. M., and Taylor, P. W. J. (2008). Vetch. *Compendium of Transgenic Crop Plants* 3, 163–170.

Francis, C. M., Enneking, D., and El Moneim, A. M. A. (2000). "When and where will vetches have an impact as grain legumes?," in *Linking Research and Marketing Opportunities for Pulses in the 21st Century*, Ed. R. Knight (Berlin: Springer), 375–384. doi: 10.1007/978-94-011-4385-1_34

Harper, J. A., and Arscott, G. H. (1962). Toxicity of common and hairy vetch seed for poults and chicks. *Poult. Sci.* 41, 1968–1974. doi: 10.3382/ps.0411968

Harwood, W. A. (2014). "A protocol for high-throughput Agrobacterium-mediated barley transformation," in *Cereal Genomics*, eds A. Furtado, and R. J. Henry (Berlin: Springer), 251–260. doi: 10.1007/978-1-62703-715-0_20

Hawkins, C., Caruana, J., Li, J., Zawora, C., Darwish, O., Wu, J., et al. (2017). An eFP browser for visualizing strawberry fruit and flower transcriptomes. *Horticult. Res.* 4:17029.

Henchion, M., Hayes, M., Mullen, A. M., Fenelon, M., and Tiwari, B. (2017). Future protein supply and demand: strategies and factors

influencing a sustainable equilibrium. *Foods* 6:53. doi: 10.3390/foods6070053

Huang, Y., Li, R., Coulter, J. A., Zhang, Z., and Nan, Z. (2019). Comparative grain chemical composition, ruminal degradation in vivo, and intestinal digestibility In Vitro of *Vicia sativa* L. Varieties Grown on the Tibetan Plateau. *Animals* 9:212. doi: 10.3390/ani9050212

Joint, W. H. O. (2007). *Protein and Amino Acid Requirements in Human Nutrition*. World health organization technical report series. Geneva: World Health Organization.

Kaya, A. A., Yörük, M., Esenbuga, N., Temelli, A., and Ekinci, Ö (2013). Retraction: The Effect of Raw and Processed Common Vetch Seed (Vicia sativa) Added to Diets of Laying Hens on Performance, Egg Quality, Blood Parameters and Liver Histopathology. *J. Poult. Sci.* 50, 228–236. doi: 10.2141/jpsa.0120013

Li, X., Wang, Y., Chen, S., Tian, H., Fu, D., Zhu, B., et al. (2018). Lycopene is enriched in tomato fruit by CRISPR/Cas9-mediated multiplex genome editing. *Front. Plant Sci.* 9:559. doi: 10.3389/fpls.2018.00559

Liu, X., Wu, S., Xu, J., Sui, C., and Wei, J. (2017). Application of CRISPR/Cas9 in plant biology. *Acta Pharm. Sin. B* 7, 292–302. doi: 10.1016/j.apsb.2017.01.002

Lobell, D. B., and Gourdji, S. M. (2012). The influence of climate change on global crop productivity. *Plant Physiol.* 160, 1686–1697.

Mao, Z., Fu, H., Nan, Z., and Wan, C. (2015). Fatty acid, amino acid, and mineral composition of four common vetch seeds on Qinghai-Tibetan plateau. *Food Chemi.* 171, 13–18. doi: 10.1016/j.foodchem.2014.08.090

Mba, W. P., Longandjo, G.-N. T., Moufouma-Okia, W., Bell, J.-P., James, R., Vondou, D. A., et al. (2018). Consequences of 1.5 C and 2 C global warming levels for temperature and precipitation changes over Central Africa. *Environ. Res. Lett.* 13:055011. doi: 10.1088/1748-9326/aab048

Mikić, A. (2016). Presence of vetches (*Vicia* spp.) in agricultural and wild floras of ancient Europe. *Genet. Resour. Crop Evol.* 63, 745–754. doi: 10.1007/s10722-016-0382-3

Mikić, A., Mihailović, V., Karagić, Đ, Milošević, B., Milić, D., Vasiljević, S., et al. (2019). Common vetch (Vicia sativa) multi-podded mutants for enhanced commercial seed production. *Proc. Appl. Bot. Genet. Breed.* 180, 78–81. doi: 10.30901/2227-8834-2019-1-78-81

Moose, S. P., and Mumm, R. H. (2008). Molecular plant breeding as the foundation for 21st century crop improvement. *Plant Physiol.* 147, 969–977. doi: 10.1104/pp.108.118232

Navrátilová, A., Neumann, P., and Macas, J. (2003). Karyotype analysis of four *Vicia* species using in situ hybridization with repetitive sequences. *Ann. Bot.* 91, 921–926. doi: 10.1093/aob/mcg099

Nguyen, V., Riley, S., Nagel, S., Fisk, I., and Searle, I. R. (2020). Common Vetch: a drought tolerant, high protein neglected leguminous crop with potential as a sustainable food source. *bioRxiv [Preprint]* doi: 10.1101/2020.02.11.943324

Patel, R. V., Nahal, H. K., Breit, R., and Provart, N. J. (2012). BAR expressolog identification: expression profile similarity ranking of homologous genes in plant species. *Plant J.* 71, 1038–1050. doi: 10.1111/j.1365-313x.2012.05055.x

Pfeffer, M., and Ressler, C. (1967). β-Cyanoalanine, an inhibitor of rat liver cystathionase. *Biochem. Pharmacol.* 16, 2299–2308. doi: 10.1016/0006-2952(67)90217-1

Rathjen, J. (1997). *The Potential of Vicia sativa L. as a Grain Legume for South Australia*. PhD thesis, University of Adelaide, Adelaide.

Reeves, D. W. (1997). The role of soil organic matter in maintaining soil quality in continuous cropping systems. *Soil Tillage Res.* 43, 131–167. doi: 10.1016/s0167-1987(97)00038-x

Ressler, C., Tatake, J. G., Kaizer, E., and Putnam, D. H. (1997). Neurotoxins in a vetch food: stability to cooking and removal of γ-glutamyl-β-cyanoalanine and β-cyanoalanine and acute toxicity from common vetch (*Vicia sativa* L.) legumes. *J. Agricult. Food Chem.* 45, 189–194. doi: 10.1021/jf9603745

Ritthausen, H., and Kreusler, U. (1870). Ueber Vorkommen von Amygdalin und eine neue dem Asparagin ähnliche Substanz in Wickensamen (*Vicia sativa*). *J. Prak. Chem.* 2, 333–338. doi: 10.1002/prac.18700020138

Roy, D. N., Sabri, M. I., Kayton, R. J., and Spencer, P. S. (1996). β-cyano-L-alanine toxicity: evidence for the involvement of an excitotoxic mechanism. *Nat. Toxins* 4, 247–253. doi: 10.1002/(sici)(1996)4:6<247::aid-nt1>3.0.co;2-m

Sahoo, K. K., Tripathi, A. K., Pareek, A., Sopory, S. K., and Singla-Pareek, S. L. (2011). An improved protocol for efficient transformation and regeneration of diverse indica rice cultivars. *Plant Methods* 7:49. doi: 10.1186/1746-4811-7-49

Sivakumar, M. V. K., Das, H. P., and Brunini, O. (2005). "Impacts of present and future climate variability and change on agriculture and forestry in the arid and semi-arid tropics," in *Increasing Climate Variability and Change*, eds J. Salinger, M. V. K. Sivakumar, and R. P. Motha (Berlin: Springer), 31–72. doi: 10.1007/1-4020-4166-7_4

Sun, Y., Jiao, G., Liu, Z., Zhang, X., Li, J., Guo, X., et al. (2017). Generation of high-amylose rice through CRISPR/Cas9-mediated targeted mutagenesis of starch branching enzymes. *Front. Plant Sci.* 8:298. doi: 10.3389/fpls.2017.00298

Tate, M. E., and Enneking, D. (1992). A mess of red pottage. *Nature* 359, 357–358. doi: 10.1038/359357a0

Tate, M. E., Rathjen, J., Delaere, I., and Enneking, D. (1999). Covert trade in toxic vetch continues. *Nature* 400:207. doi: 10.1038/22198

Tenopala, J., González, F. J., and De La Barrera, E. (2012). Physiological responses of the green manure, Vicia sativa, to drought. *Bot. Sci.* 90, 263–285.

Valentine, S. C., and Bartsch, B. D. (1996). Production and composition of milk by dairy cows fed common vetch or lupin grain as protein supplements to a silage and pasture-based diet in early lactation. *Aust. J. Exp. Agricult.* 36, 633–636.

Westhoek, H., Rood, T., Van Den Berg, M., Janse, J., Nijdam, D., Reudink, M., et al. (2011). *The protein puzzle: The Consumption and Production of Meat, Dairy and Fish in the European Union*. The Hague: Netherlands Environmental Assessment Agency.

White, P., Harries, M., Seymour, M., and Burgess, P. (2005). *Producing Pulses in the Northern Agricultural Region*. Kensington: Western Australia Department of Agriculture and Food.

Conflict of Interest: 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 © 2020 Nguyen, Riley, Nagel, Fisk and Searle. 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) and the copyright owner(s) 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 Plant Science

REVIEW
published: 02 July 2020
doi: 10.3389/fpls.2020.01002

Southern Species From the Biodiversity Hotspot of Central Chile: A Source of Color, Aroma, and Metabolites for Global Agriculture and Food Industry in a Scenario of Climate Change

Luis Letelier[1,2], *Carlos Gaete-Eastman*[1], *Patricio Peñailillo*[1], *María A. Moya-León*[1] *and Raúl Herrera*[1]*

[1] Laboratorio de Fisiología Vegetal y Genética Molecular, Instituto de Ciencias Biológicas, Universidad de Talca, Talca, Chile, [2] Núcleo Científico Multidisciplinario, Dirección de Investigación, Universidad de Talca, Talca, Chile

OPEN ACCESS

Edited by:
Eric Von Wettberg,
University of Vermont, United States

Reviewed by:
Kim E. Hummer,
United States Department of Agriculture, United States
Amparo Monfort,
Institute of Agrifood Research and Technology (IRTA), Spain
Petr Smýkal,
Palacký University, Czechia

***Correspondence:**
Raúl Herrera
raherre@utalca.cl

Specialty section:
This article was submitted to Plant Breeding, a section of the journal Frontiers in Plant Science

Received: 18 December 2019
Accepted: 19 June 2020
Published: 02 July 2020

Citation:
Letelier L, Gaete-Eastman C, Peñailillo P, Moya-León MA and Herrera R (2020) Southern Species From the Biodiversity Hotspot of Central Chile: A Source of Color, Aroma, and Metabolites for Global Agriculture and Food Industry in a Scenario of Climate Change. Front. Plant Sci. 11:1002. doi: 10.3389/fpls.2020.01002

Two interesting plants within the Chilean flora (wild and crop species) can be found with a history related to modern fruticulture: *Fragaria chiloensis* subsp. *chiloensis* (Rosaceae) and *Vasconcellea pubescens* (Caricaceae). Both species have a wide natural distribution, which goes from the Andes mountains to the sea (East-West), and from the Atacama desert to the South of Chile (North-South). The growing locations are included within the *Chilean Winter Rainfall-Valdivian Forest* hotspot. Global warming is of great concern as it increases the risk of losing wild plant species, but at the same time, gives a chance for usually longer term genetic improvement using naturally adapted material and the source for generating healthy foods. Modern agriculture intensifies the attractiveness of native undomesticated species as a way to provide compounds like antioxidants or tolerant plants for climate change scenario. *F. chiloensis* subsp. *chiloensis* as the mother of commercial strawberry (*Fragaria* × *ananassa*) is an interesting genetic source for the improvement of fruit flavor and stress tolerance. On the other hand, *V. pubescens* produces fruit with high level of antioxidants and proteolytic enzymes of interest to the food industry. The current review compiles the botanical, physiological and phytochemical description of *F. chiloensis* subsp. *chiloensis* and *V. pubescens*, highlighting their potential as functional foods and as source of compounds with several applications in the pharmaceutical, biotechnological, and food science. The impact of global warming scenario on the distribution of the species is also discussed.

Keywords: Chilean strawberry, *Fragaria chiloensis* subsp. *chiloensis*, mountain papaya, new crops, *Vasconcellea pubescens*

INTRODUCTION

Few grain species were initially used in human diet. Several other vascular plants species have been incorporated and plant breeding effort have selected inbreeds with interesting and desirable characteristics focus on productivity (Khoshbakh and Hammer, 2008; Şerban et al., 2008). In the case of fruit species, characteristic like, fruit size, color, and texture have been the main quality attributes followed by breeders during the last century, but now a great interest for new traits, such as good aroma, pathogen resistance, high level of antioxidants, healthy compounds, and better postharvest life are considered. Nowadays, fruits from native species could be incorporated in our diet due to their contribution as functional food. In particular, strawberries are rich in secondary metabolites, as well as other attributes, however the fruit is susceptible to several diseases. Hancock et al. (2010) recognizes the importance of octoploid species such as *F. chiloensis* and *F. virginiana* to enrich traits such as resistance to diseases or better performance against biotic or abiotic stresses in new commercial lines of *F.* × *ananassa*. The global warming scenario of our planet has greatest importance in recent days, as the risk to lose wild plant species is certain due to changes in the ecological conditions (Feng et al., 2010; Lippmann et al., 2019; Moran, 2020). Chile is a world center of origin for cultivated plants and also considered as a biodiversity hotspot. In this sense, two interesting taxa with a history related to modern fruticulture can be mentioned: *Fragaria chiloensis* (L.) Mill. subsp. *chiloensis* (Rosaceae), known as the Chilean strawberry, and *Vasconcellea pubescens* A. DC. (Caricaceae), known as mountain papaya. The present article describes the development, use and potential future for both species.

MOUNTAIN PAPAYA

Vasconcellea pubescens A. DC. (Caricaceae) (synonyms: *Carica pubescens* Lenné & K. Koch, *C. candamarcensis* Hook. f., and *V. cundinamarcensis* V.M. Badillo) (Badillo, 2000; Van Droogenbroeck et al., 2002; Hassler, 2019), known as highland or mountain papaya, is a diploid dicotyledonous species ($2n = 2x = 18$). It is native from subtropical Andean mountains of South America, in particular from low dry mountain forest areas in Colombia and Ecuador (2,000 to 3,000 m altitude). Ecuador, which has 15 of the 21 *Vasconcellea* species, is a center of genetic biodiversity (Van den Eynden et al., 1999; Scheldeman et al., 2011). Some genotypes of *V. pubescens* have been well adapted to central Chile before the Spanish colonization, suggesting that its introduction occurred during the Inca Empire or even earlier (Latcham, 1936). A surface of 180 ha of commercial orchards (ODEPA-CIREN, 2019), mostly located between 30° and 33° latitude south (**Figure 1**; **Supplementary Table 1**), highlights Chile as the only country in the southern hemisphere producing the fruit at commercial scale and using it as a source of processed products.

Mountain papaya is an arborescent plant reaching 1 to 8 m tall, with one central stem and palmate leaves with long petioles at the top (**Figure 2A**). Most of individual plants are dioecious, but a wide range of hermaphrodite phenotype is possible to identify in the field, so formally it is a sub-dioecious species with three sex phenotypes: female, male, and hermaphrodite (**Figure 2B**) (Badillo, 1993). *V. pubescens* presents a case of andromonoecious, where bisexuality is observed in male plants, with a high proportion of male flowers and hermaphrodite ones. Bisexual flowers are male flowers bearing viable ovules leading to the production of small fruits (Scheldeman et al., 2011). Sex in *V. pubescens* is determined by two factors, the Ypm chromosome and *pm* cytoplasmic factor (Horovitz and Jimenez, 1967), which are primitive sex chromosomes and a model for sex evolution determination (Na et al., 2014). The reproductive mode is through seeds, taking between 10 to 12 months to reach reproductive age, and maintained for 5 years at commercial scale. Then, plants progressively produce less fruit and with lower quality, nevertheless, plants older than 20 years are found in Chilean orchards. Mountain papaya tree shows a slow growth, with a continuous flower and leaf production; when lower leaves get older they fall into the ground (Sánchez, 1994).

V. pubescens is sensitive to cold, stem and leaves could be affected leading to complete plant death when temperatures fall below 2°C; even ripening of the fruit could be altered by cold. Cold episodes have been more frequent in Chile in recent years, even in coastal areas where chilling injuries on commercial orchards were absolutely absent 10 years ago. Additionally, extended drought stress promotes the continuous loss of leaves (Sánchez, 1994).

Mountain papaya tree bears fruit at the leaf axis spirally arranged along the trunk, so it is possible to observe from immature (small size, green colour) to ripe (big size, bright yellow) fruit stages in one single plant. In natural populations, birds eat ripe and overripe fruits, allowing seed dispersal and germination. Seeds have a high germination rate (60% in 30 days) without the requirement of a dormancy period (Sánchez, 1994). The fruit is oblong, truncated at the base, with five pronounced ridges; a ripe fruit is about 8–15 cm long, 5–6 cm diameter, and 200 g mean weight (García, 1975) (**Figure 3**). The fruit has a juicy yellow flesh with a strong and aromatic flavor, characterized by a high content of vitamins (A, B, and C), high level of antioxidants, and a milky latex (highly abundant in immature stages) that contains a mixture of proteolytic enzymes, commonly named as papain (Van Droogenbroeck et al., 2002; Moya-León et al., 2004). The fruit is commonly used to prepare preserves, jam, sweet candies and nectar, meanwhile its latex is widely used in the industry as meat tenderizer. Mountain papaya is frequently compared with the common and worldwide known papaya (*Carica papaya* L.), being *V. pubescens* smaller and less succulent than *C. papaya* but with a greater aroma and flavor (Van den Eynden et al., 1999; Scheldeman et al., 2003; Van den Eynden et al., 2003).

Diversity and Genetic Structure

Some phenotypic variations are observed in mountain papaya, mainly in plant height, number of branches and fruit size. Nevertheless, the most important differences are in sex

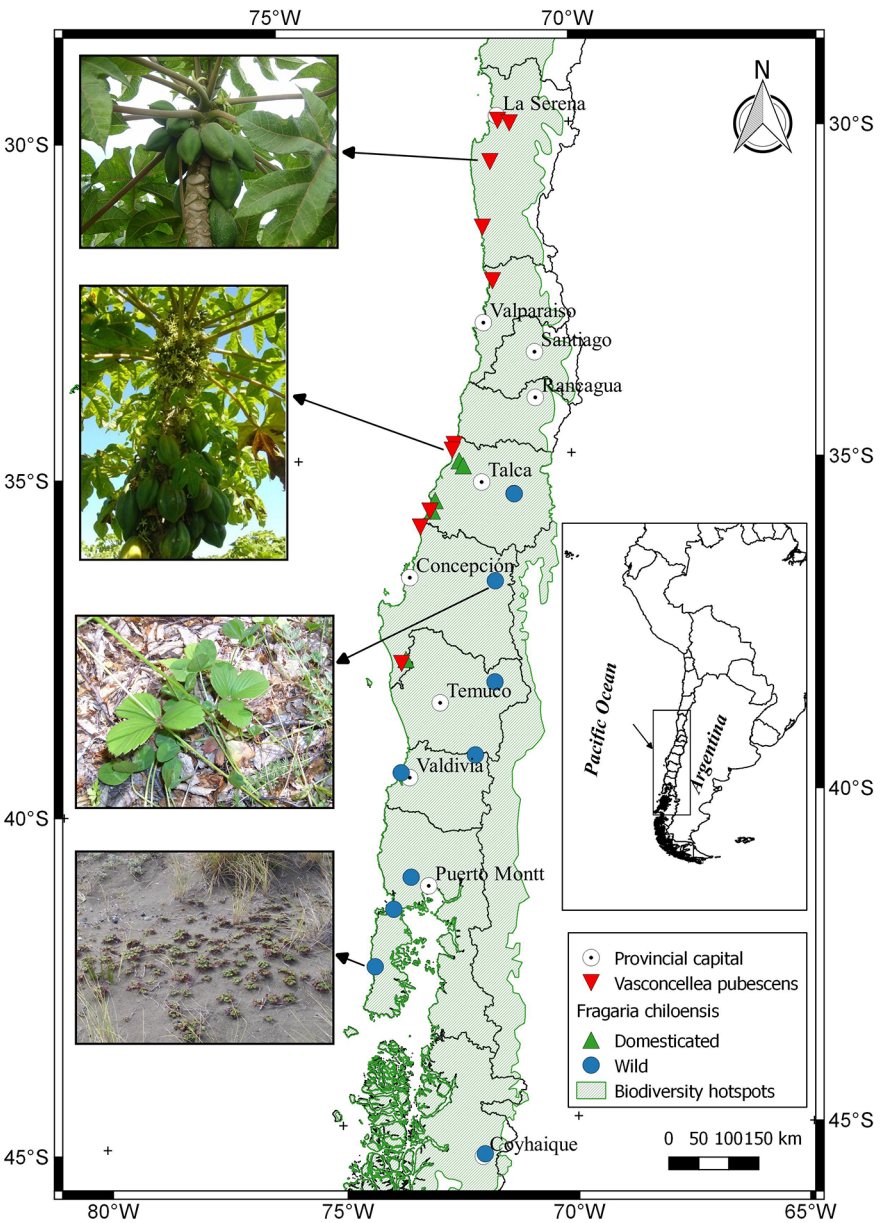


FIGURE 1 | Distribution of *Fragaria chiloensis* (wild and domesticated) and *Vasconcellea pubescens* in Chile. The map shows a partial extension (area) of the *Chilean Winter Rainfall-Valdivian Forests* hotspot. Images located on the left side of the map, from North to South, corresponded to *V. pubescens* plants grown at the localities of Limarí and Lipimávida, and for *F. chiloensis* f. *patagonica* grown at Termas de Chillán and Cucao, respectively.

determination. Although no environmental variation can be observed for pistillate and staminate plants, the proportion of male and hermaphrodite flowers in andromonoecious plants depends on climate conditions. On the other hand, the incompatibility barrier is labile in *Vasconcellea*, and therefore, this allows the possibility to increase genetic variation through the formation of spontaneous hybrid specimens between different species (Badillo, 1993; Sánchez, 1994). Interspecific hybrids have been found in natural populations (De Zerpa, 1959), however only one has commercial success: the babaco (*Vasconcellea* × *heilbornii* (V.M. Badillo) V.M. Badillo). Babaco results from the cross between *V. pubescens* and *V. stipulata* (V.M. Badillo) V.M. Badillo (Horovitz and Jimenez, 1967; Badillo, 1971; Scheldeman et al., 2011). Molecular studies using molecular markers suggest that the origin of *V.* × *heilbornii* results of a complex evolution process where *V. pubescens*, *V. stipulata*, and *V. weberbaueri* (Harms) V.M. Badillo are parental species (Scheldeman et al., 2011). Molecular analysis also confirms hybrids obtained by the cross of *V. pubescens* and *V. monoica* (Desf.) A. DC. (Kyndt et al., 2005; Kyndt et al., 2006) and interspecific crosses with *V.*

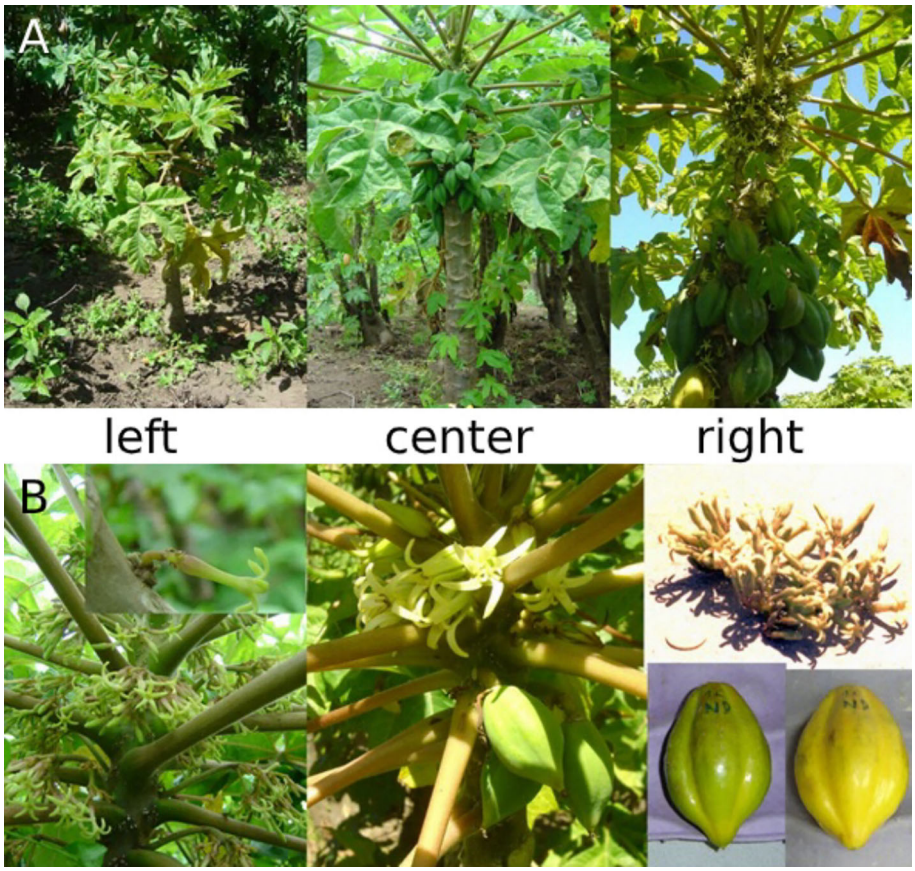


FIGURE 2 | Morphological characteristics of *V. pubescens* tree, flowers and fruit. **(A)** Mountain papaya tree at three growing stages: a two month old tree of half a meter tall producing the first flowers (left); after one year, the tree reaches one meter tall and produces fruit (center); after three years, the tree is two meters tall, it is under full fruit production (right); interestingly, fruit at different developmental stages could be observed in the same tree. **(B)** Sexual forms in *V. pubescens*. Male plants produce staminate flowers, showing their characteristic elongated form, with the stamen inside the flower (left). Female plants produce pistillate flowers, with the characteristic enlargement of the flower base, because of ovary growth (center). Andromonoecious plants produce hermaphrodite and estaminate flowers, where the first ones differentiate from pistillate flowers showing a wider enlargement of flower base (up right). The most notorious change during fruit ripening is the color change of fruit skin, varying from an intense green to bright yellow color (down right).

stipulata, *V. monoica*, *V. microcarpa* (Jacq.) A. DC. and *V. horovitziana* (V.M. Badillo) V.M. Badillo (Horovitz and Jimenez, 1967; Badillo, 1971).

A high genetic diversity along the natural distribution pattern in the centres of origin (Ecuador and northern Perú) is assumed, but scarce information exists regarding the history and genetic diversity of *V. pubescens* introduced in Chile. A low genetic variability but a high genetic polymorphism was reported using ISSR markers (Carrasco et al., 2009). Based on the results, the authors suggested that few introduction events of genetic material from the original populations were concreted during time and this material constitutes the basal genotype of Chilean material.

Phytochemistry

Aroma is one of the most important attributes of mountain papaya fruit and it is due to a complex mixture of volatile compounds produced by the fruit (Balbontín et al., 2007). Main volatile compounds produced by the fruit are esters and alcohols (linear and branched), and their production increase as ripening progresses (Balbontín et al., 2007). Most of esters identified in mountain papaya fruit are potent odour compounds, such as ethyl butanoate, ethyl acetate, ethyl hexanoate, and ethyl 2-methylbutanoate, and the most important alcohol is butanol. The dynamic of their production during ripening depends on ethylene, which agrees to consider mountain papaya as a climacteric fruit (Moya-León et al., 2004; Balbontín et al., 2007). Several other volatile compounds such as methyl cis-hex-3-enoate, isopentyl acetate, methyl 3-hydroxyhexanoate and ethyl nicotinoate have been reported in mountain papaya fruit grown in Chile, however they have not been detected in fruit grown in Colombia (Morales and Duque, 1987). On the other hand, many fruits accumulate antioxidant activity during ripening, which is a desirable feature for functional foods. Active antioxidant phenolics from mountain papaya fruit include quercetin glycosides, rutin, and manghaslin, which are not produced by common papaya (*C.*

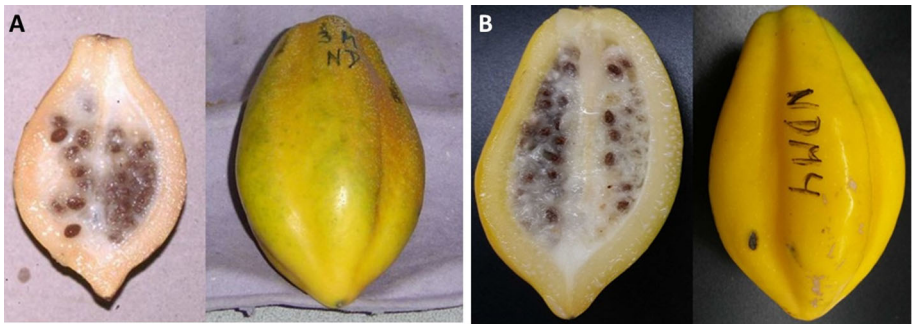


FIGURE 3 | Morphological characteristics of mountain papaya fruit at the ripe stage. Fruit shows its characteristic shape, thin pulp layer, with an inside cavity full of maternal tissue that contains many seeds (left), and the typical yellow skin color of ripe fruit with five ridges (right). The fruit was harvested at Lipimávida in seasons 2010 **(A)** and 2018 **(B)**.

papaya) (Simirgiotis et al., 2009). This suggests a metabolic divergence between both plant species.

All members of the Caricaceae family have laticifers and produce milky latex when a plant tissue is damaged. This latex helps as a defensive mechanism against predators and to heal wounded sites (Konno et al., 2004). The latex composition and proteinase quantity is different between species. The latex from *C. papaya* has been described and it consists of a mixture of hydrolase enzymes like quitinases and proteases, being the most characteristics papain, chymopapain, caricain (proteinase Ω) and glycil endopeptidase (proteinase IV) (El Moussaoui et al., 2001). There are some evidences showing that freeze-dried latex from *V. pubescens* has 5 to 8 times more proteolytic activity compare to the one obtained from *C. papaya* (Baeza et al., 1990). On the other hand, the extraction of *V. pubescens* latex and its separation into four different fractions named as CC-I to CC-IV was reported (Walraevens et al., 1993); two of them, CC-I and CC-III have been characterized and corresponded to papain and chymopapain, respectively. The mixture of latex proteinases has been used to treat gastric ulcers in rodent models (Mello et al., 2008). More recently, two different fractions were obtained from *V. pubescens* latex: a fraction showing high proteinase activity (CMS) and a fraction showing moderate to low proteinase activity (Teixeira et al., 2008). On the other hand, the fraction called P1G10 has been used for gastric ulcers and diabetic foot treatments in several wounded models (Tonaco et al., 2018). At the same time, the reduction of tumour mass in animals bearing melanoma and metastasis level was observed using this fraction (Lemos et al., 2018).

THE CHILEAN STRAWBERRY

Fragaria L. is a member of the Rosaceae family. There are about 20 species distributed in temperate Eurasia and North and South America. Staudt recognizes four subspecies for *F. chiloensis* (L.) Mill. based on morphological traits with a worldwide distribution: two of them are located in North America (*F. chiloensis* subsp. lucida (E.Vilm. ex J. Gay) Staudt, and *F. chiloensis* subsp. *pacifica* Staudt), one in Hawaii (*F. chiloensis* subsp. *sandwicensis* (Decne.) Staudt) and the last one in South-America (i.e., Ecuador, Bolivia, Perú, Argentina and Chile) (*F. chiloensis* subsp. *chiloensis* Staudt) (Staudt, 1962; Bringhurst, 1990; Staudt, 1999). Staudt also proposed two botanical forms for the subspecies *F. chiloensis* subsp. *chiloensis* based on characters such as plant size, texture of leaves, color and fruit size: *F. chiloensis* (L.) Mill. subsp. *chiloensis* f. *chiloensis* Staudt, is a landrace that produces large fruit with white/pink receptacle and white flesh; and *F. chiloensis* (L.) Mill. subsp. *chiloensis* f. *patagonica* Staudt, that produces small red fruits (Staudt, 1962; Staudt, 1999). Shortly after the conquest of the Central-South part of Chile, location where the Mapuches used to live, *F. chiloensis* subsp. *chiloensis* was taken as a war trophy by the Spaniards and the species was introduced in Perú. Historical reports indicate that Garcilaso de la Vega in 1557 observed a new strawberry species in the land and markets of Cuzco City, different in size compared to the European species (Popenoe, 1921). Thereafter, the species was introduced in Ecuador by the Spaniards; although there are not exact records, it seems to be before 1789 (Popenoe, 1921). Interestingly, it was not until 1712 that *F. chiloensis* subsp. *chiloensis* was introduced in Europe by some explorers coming from Chile (Popenoe, 1921; Darrow, 1966; Liston et al., 2014). From an agronomic perspective, mainly in Chile and in minor scale in Perú and Ecuador, *F. chiloensis* subsp. *chiloensis* was cultivated as a fruit source during the first half of 1900 century, however when new cultivars reached from Europe and USA, the species was displaced and its cultivation was not favoured (Finn et al., 2013).

In botanical terms, *F. chiloensis* subsp. *chiloensis* is a perennial herb, with strong and well developed runners; trifoliate leaves; with unique or few flowers, which can be dioic or hermaphrodite. The berry is an aggregate fruit, which develops from a flower that contains several ovaries (Marticorena, 2019). The fertilized ovaries develop into achenes that resulted embedded on surface of the receptacle. The edible receptacle or thalamus is usually considered as the most attractive part in sensory terms (**Figure 4**).

Nowadays in Chile, the two botanical forms of *F. chiloensis* subsp. *chiloensis* grow in central-south part of Chile (**Figure 1**), and can be found between the O´Higgins and the Aysén Regions



FIGURE 4 | Morphological characteristics of Chilean strawberry fruit at the ripe stage. **(A-D)** correspond to *Fragaria chiloensis* f. *chiloensis* fruit distributed from North to South, from the localities of Huelón Alto, Pelluhue, Purén and Contulmo, respectively. **(E-H)** correspond to wild *F. chiloensis* fruit distributed from North to South, from the localities of Termas de Chillán (see **Figure 1**), Curiñanco, Cuesta Gutiérrez and Cucao (see **Figure 1**), respectively.

(35°05'– 45°32' latitude South) (Lavin et al., 2000; Rodríguez et al., 2018), growing from forests (understory) to open vegetation environments (Kalkman, 2004) (**Figure 1**). These areas present contrasting conditions for plant development, with differences in light, temperature and nutrients availability (Moreira-Muñoz, 2011). The fruit phenotype of Chilean accessions is quite diverse. As a way to demonstrate this, plants collected from different geographic areas were established in one edaphoclimatic condition (Linares) (**Supplementary Table 1**). Cultivated *F. chiloensis* f. *chiloensis* species from Huelón Alto, Pelluhue, Purén, and Contulmo produces fruit of white flesh, white/pink receptacle of uniform medium size (**Figures 4A–D**). On the contrary, the fruit produced by wild plants of *F. chiloensis* f. *patagonica* collected in Termas de Chillán, Curiñanco, Cuesta Gutiérrez and Cucao produce small fruit size with red receptacle (**Figures 4E–H**). The white flesh fruit is commercially produced at small scale and its main use is for fresh consuming or to prepare beverages (Pardo and Pizarro, 2013).

F. chiloensis f. *chiloensis* (white fruit) is considered as one of the parentals for the commercial strawberry (*F.* × *ananassa* (Duchesne ex Weston) Duchesne ex Rozier) along with *F. virginiana* Mill., which have also been supported by molecular analysis (Njuguna et al., 2013; Tennessen et al., 2014; Edger et al., 2019). *F. chiloensis* f. *chiloensis* has been widely studied particularly for its distinctive exotic white-pink color and characteristic aroma (Hancock et al., 1999; Retamales et al., 2005; González et al., 2009b). Several studies have been published reporting the softening of this fruit, showing that when firmness is reduced cell wall polymers are

broken down (Figueroa et al., 2008; Moya-León et al., 2019). Several genes and enzymes involved in cell wall disassembly have been characterized (Figueroa et al., 2008; Figueroa et al., 2009; Pimentel et al., 2010; Opazo et al., 2013; Méndez-Yáñez et al., 2017; Méndez-Yáñez et al., 2020). Softening requires the fine coordination of several molecular activities, including the participation of plant hormones and transcription factors (Handford et al., 2014; Carrasco-Orellana et al., 2018; Moya-León et al., 2019). The key role played by the molecular coordinators on the expression of cell wall degrading genes impacts softening and the postharvest quality of the fruit.

The plant and fruit of *F. chiloensis* f. *chiloensis* show tolerance to pathogens and diseases (González et al., 2013). In addition, the plant species has the ability to grow under different abiotic stress conditions (low temperatures, salty soils) which gives a great potential for breeding purposes (González et al., 2009a; Aceituno-Valenzuela et al., 2018). In this sense, the species can be considered as genetic source for strawberry breeding programs.

F. chiloensis subsp. *chiloensis* is located in one of the worldwide biodiversity hotspots. A hotspot is defined as an area with high diversity of endemic species (over 1,500), where 30% or less of this area should be threatened, and there are 34 areas recognized globally (Arroyo et al., 1999; Myers et al., 2000; Mittermeier et al., 2004). The *Chilean Winter Rainfall-Valdivian Forests* hotspot covers a great part of northern and southern Chile and contains 1,957 different endemic species (Mittermeier et al., 2004). A reduction of plant communities have been observed directly by a high change in soil use from forests to forest plantations or agriculture (Arroyo et al., 1999). Several species belong to this area, which are used for human consumption (**Table 1**). Climate change models predict for the geographic area of this hotspot a decrease in the distribution range of plant species, in addition to, distribution displacements towards the south or to take refuge in the Andes Mountains (Pliscoff et al., 2012; Bambach et al., 2013).

TABLE 1 | Native plants from the Chilean winter rainfall-Valdivian forest hotspot with potential in agriculture.

Native plants	Edible organ	Folk name	Family
Araucaria araucana (Molina) K. Koch	Fruit	Pehuen	Araucariaceae
Aristotelia chilensis (Molina) Stuntz	Fruit	Maqui	Elaeocarpaceae
Berberis microphylla G. Forst.	Fruit	Calafate	Berberidaceae
Eulychnia acida Phil.	Fruit	Copao	Cactaceae
Gaultheria species	Fruit	Chauras	Ericaceae
Geoffrea decorticans (Gillies ex Hook. et Arn.) Burkart	Fruit	Chañar	Fabaceae
Gevuina avellana Molina	Fruit	Avellano	Proteaceae
Gomortega keule (Molina) Baill.	Fruit	Queule	Gomortegaceae
Gunnera tinctoria (Molina) Mirb.	Petiole	Pangue	Gunneraceae
Lardizabala biternata Ruiz et Pav.	Fruit	Cóguil	Lardizabalaceae
Peumus boldus Molina	Fruit	Boldo	Monimiaceae
Prosopis chilensis (Molina) Stuntz emend. Burkart	Fruit	Algarrobo	Fabaceae
Prumnopitys andina (Poepp. ex Endl.) de Laub.	Fruit	Lleuque	Podocarpaceae
Puya chilensis Molina	Vegetative buds	Puya	Bromeliaceae
Ribes magellanicum Poir.	Fruit	Zarzaparrilla	Grossulariaceae
Rubus geoides Sm.	Fruit	Miñe miñe	Rosaceae
Solanum tuberosum L. subsp. *tuberosum*	Tuber	Papa chilota	Solanaceae
Ugni molinae Turcz.	Fruit	Murtilla	Myrtaceae

Diversity and Genetic Structure

A center of diversity assures sustainable genetic improvement for plant species. The wide range of environmental conditions covered by the Chilean hotspot can be used as a source of genetic adaptation to different habitats, conferring tolerance to several stresses or environmental factors. Besides, the loss of genetic diversity within a species can result in the loss of useful and desirable traits and may eliminate options to use untapped resources for food production, industry and medicine (Hoisington et al., 1999; Govindaraj et al., 2015).

The genetic diversity in *Fragaria* has been studied using molecular markers such as microsatellites (SSR) or dominant markers (ISSR, AFLP and RAPD), and more recently using a high throughput system such as single nucleotide polymorphisms (SNPs) array (Bassil et al., 2015; Hilmarsson et al., 2017; Jung et al., 2017). SSR markers, for example, were used with the purpose to identify *F.* × *ananassa* cultivars (Monfort et al., 2006; Shimomura and Hirashima, 2006; Honjo et al., 2011). The IStraw90 Axiom array is a genotyping platform for *Fragaria* able to identify SNPs, indels, and also is used for mapping (Mahoney et al., 2016; Nagano et al., 2017; Sooriyapathirana et al., 2019; Oh et al., 2020). In the case of *F. chiloensis* subsp. *chiloensis* a small number of accessions were analyzed and low genetic diversity and no genetic structure was reported (Carrasco et al., 2007; Oñate et al., 2018). The vegetative propagation of the species was used as argument for the low genetic variability found along Chile (Lavin et al., 2000), however despite of that, a breeding and improvement program on the species was considered feasible. Several crosses of the Chilean strawberry and *Fragaria* spp have been performed which have been evaluated for the generation of elite genotypes (Hancock et al., 2001; Finn et al., 2013). Recently, new hybrids were successfully obtained by crossing four different *F.* × *ananassa* cultivars with *F. chiloensis* f *chiloensis*, increasing diversity and facilitating new bridge species for breeding (Luque et al., 2019). The selection from the breeding programe can be facilitated and optimized by the use of molecular techniques (Edger et al., 2019; Oh et al., 2019).

Phytochemistry

F. chiloensis subsp. *chiloensis* fruit contains an interesting composition of phenolics and high antioxidant activity. Ellagic acid and cinnamic acid glycosides are present in the white Chilean strawberry fruit (Cheel et al., 2005), and the fruit is also rich in phenolic antioxidants (Cheel et al., 2007; Simirgiotis et al., 2009). In addition, the fruit is characterized for having great aroma properties (González et al., 2009b). Differences in the volatile profiles have been described among commercial strawberry cultivars (Staudt et al., 1975; Forney et al., 2000), and most importantly between cultivated and wild species, being wild species those with the highest concentration of volatiles and better aroma properties (Ulrich et al., 2007). Commercial strawberries produce numerous volatile compounds including

esters, aldehydes, ketones, alcohols, terpenes, furanones, and sulfur compounds (Latrasse, 1991; Zabetakis and Holden, 1997), nevertheless esters are the most abundant class of compounds (25% to 90% of total volatiles) (Pérez et al., 1992), and provide the fruity notes of fresh ripe fruit (Pyysalo et al., 1979; Schreier, 1980). Some esters identified in the white Chilean strawberry fruit include ethyl acetate, methyl butanoate, 2-methyl acetate, octyl acetate, octyl butanoate, hexyl acetate, ethyl heptanoate, 2-hexenyl butanoate, benzyl acetate, and hexyl 2-methyl butanoate, which were also described in several *F.* × *ananassa* cultivars (Pyysalo et al., 1979; Zabetakis and Holden, 1997; Azodanlou et al., 2003; Berna et al., 2007; González et al., 2009b). Importantly, esters such as hexyl propanoate, ethyl 4-decenoate, 2-phenylethyl propanoate, and ethyl 2,4 decadienoate have been identified in *F. chiloensis* f. *chiloensis* and not in *F.* × *ananassa* (González et al., 2009b).

Anthocyanins are the color pigments of strawberry fruit. In red strawberries glycosylated anthocyanins derived from pelargonidin and cyanidin are normally present, being pelargonidin 3-glucoside the most abundant anthocyanin (Tulipani et al., 2008). In the white Chilean strawberry, the main anthocyanin is cyanidin 3-O-glucoside, which is mainly present in the achenes (Cheel et al., 2005; Salvatierra et al., 2010; Salvatierra et al., 2014). Importantly, anthocyanins have potential benefits and broad health-promoting effects for human considering the antioxidant activity (Chamorro et al., 2019; Hwang et al., 2019; Mazzoni et al., 2019).

The potential health benefits of strawberry fruit consumption have been analyzed, and some *in vivo* effects on the oxidative status in human and animal models are now available. Chilean white fruit aqueous extracts have a protective effect on platelet aggregation (Parra-Palma et al., 2018) and can exert protection on human epithelial gastric cells against free radical-induced damage (Avila et al., 2017). On the other hand, dietary supplementation of rats with white Chilean strawberry juice favors the normalization of oxidative and inflammatory effects in response to a liver injury induced by lipopolysaccharides (Molinett et al., 2015). These results support the idea to considerate the Chilean strawberry fruit as a functional food.

MEDITERRANEAN CROP AND CLIMATE CHANGE

In general, papaya plants grow mostly in tropical and subtropical Andes. In Chile, as shown in **Figure 1**, mountain papaya (*V. pubescens*) is cultivated in coastal areas between latitudes 30° to 36° South, which corresponds to a climate with subtropical influence characterized by moderate thermal ranges and low occurrences of frosts. Recently, Sarricolea et al. (2017) carried out a climatic classification of Chile, and proposed that the cultivated populations of *V. pubescens* could be cultivated in 4 climatic types of Köppen-Geiger (see **Figure 5** and **Supplementary Table 1**). On the other hand, regarding the cultivation of strawberry in Chile, both *F.* × *ananassa* and *F. chiloensis* f. *chiloensis* grow between latitudes 32° to 37° South, in coastal areas of Chile, where climate corresponds to a Mediterranean type, with 8 variants according to Köppen-Geiger climate classification (see **Figures 6** and **7** and **Supplementary Table 1**, Sarricolea et al., 2017).

The Intergovernmental Panel on Climate Change (IPCC) predicts that mid-latitude areas would be affected by the increment in drought, with irregular rainfall regimes and rising temperatures (Parajuli et al., 2019). These locations are the cultivation areas where main crops are worldwide concentrated, which also include our both species under study. The average temperature and precipitation within the area of cultivation for *V. pubescens* and *F. chiloensis* can be estimated using the information provided by climate models such as WorldClim. WorldClim 1 considers the interpolation of observed climatic data representative for the period of 1960–1990 (Hijmans et al., 2005), and WorldClim 2 considers climatic data between 1970-2000 (Fick and Hijmans, 2017). Although WorldClim defines 19 bioclimatic variables, only Annual Mean Temperature (BIO 1), Min Temperature of Coldest Month (BIO 6) and Annual Precipitation (BIO 12) were used in this analysis as they provide robust differences between sites. Aridity index (AI) was determined using BIO1 and BIO12, which correlates the precipitation (annual or monthly) with mean temperature (also annual or monthly), and allows the classification of study sites as desert, meadow or forest (Wang and Takahashi, 1999). High aridity index values indicate a site with more water availability. Considering that the increase in frosts could affect crops in mid-latitude areas BIO 6 variable was also considered (Parajuli et al., 2019).

In the case of *V. pubescens*´s distribution, the model predict a similar precipitation behavior for each location along the years, with values from North to South ranging from 90 mm average in Limarí to 1,515 mm average at the southern end of the distribution (Lleu-Lleu) (**Supplementary Figure 1**). Interestingly, Squeo et al. (1999) indicated that the annual rainfall for the area of La Serena (north of Limarí) was reduced from 180 mm at the beginning of 1900 to about 80 mm by the year 2000, which evidences a precipitation drop of 40% during the last century. On the other hand, a slight decrease in the average annual temperature (BIO1) and a drop in the minimum temperature for the coldest month by at least 2.4 °C can be observed (**Figure 5**, **Supplementary Figure 1**); this drop is mainly observed in southern locations. Finally, no differences in aridity index are observed between the data provided by WorldClim 1 and WorldClim 2 for mountain papaya locations (**Figure 5**, insert).

The climatic behavior for growing locations of domesticated *F. chiloensis* f. *chiloensis* is shown in **Supplementary Figure 2**. Average annual temperature (BIO1) for the locations are the same in both models; however there is a reduction of about 1.7° C in the minimum temperature of the coldest month (BIO6) in WorldClim 2 compared to WorldClim1 (**Figure 6**). On the other hand, a dramatic reduction in rainfall is observed in southern locations (Purén) (around 100 mm) in WorldClim 2 compared to WorldClim 1 (**Supplementary Figure 2**). In general terms, there is a reduction in aridity index especially in the southern locations of domesticated strawberry (**Figure 6**, insert), which is also accompanied with a reduction in minimum temperatures during winter. As suggested by

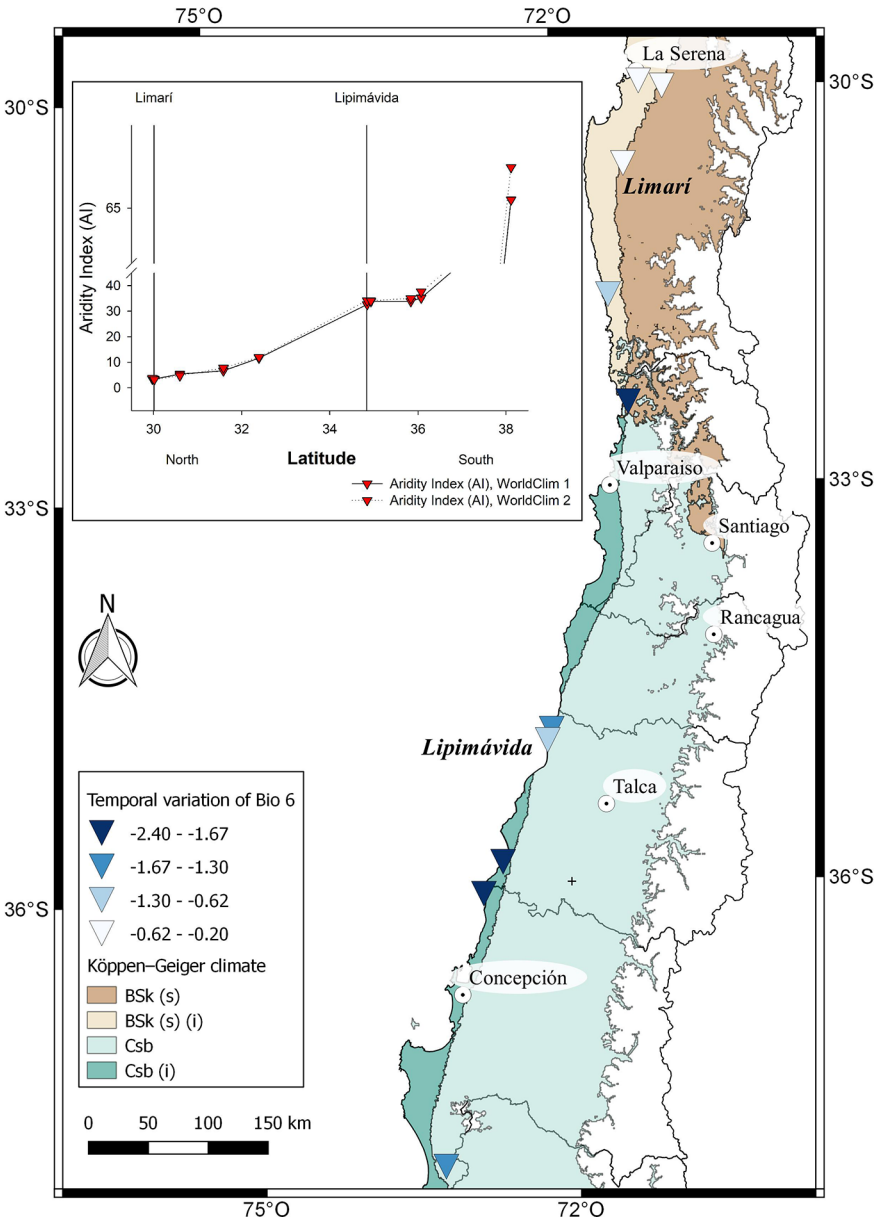


FIGURE 5 | Climatic characterization of cultivated localities of *V. pubescens* distributed from North to South. The comparison of Bioclimatic variable 6 (Min temperature of coldest month) between WorldClim 1 and 2 for each locality is shown in the map. In the insert, De Martonne Aridity Index (AI) for each locality. On the X axis, the localities of Limarí and Lipimávida are marked as reference, as previously shown in **Figure 1**.

Bambach et al. (2013) these climate changes could affect the current distribution of the species.

Finally, the analysis of climate change for the locations of *F. chiloensis* f. *patagonica* is shown in **Supplementary Figure 3**. It is more complex to interpret as there is a latitudinal effect (North – South distribution of the localities) in addition to a longitudinal component (ie, East – West; mountain - seaside). Therefore, we will focus the comparison between the localities of Termas de Chillán and Cucao (**Figures 1** and **7**). Populations distributed between 36–40° South are located in the Andes mountain (nearby Termas de Chillán), meanwhile those distributed further south (42–46° South) are located at sea level (close to Cucao) (**Supplementary Figure 3**). A significant reduction in the minimum temperature of the coldest month of 4.2°C has been reported for Termas de Chillán location in WorldClim 2 compared to WorldClim 1 (**Figure 7**); a smaller reduction in minimum temperature was reported for Cucao (2.4°C) (**Supplementary Figure 3**). On the other hand, there are no significant differences in the average annual temperature for both locations, neither in the annual precipitation in Termas de Chillán (**Supplementary Figure 3**); nevertheless an increase of about 230 mm is reported in Cucao comparing WorldClim 2 and

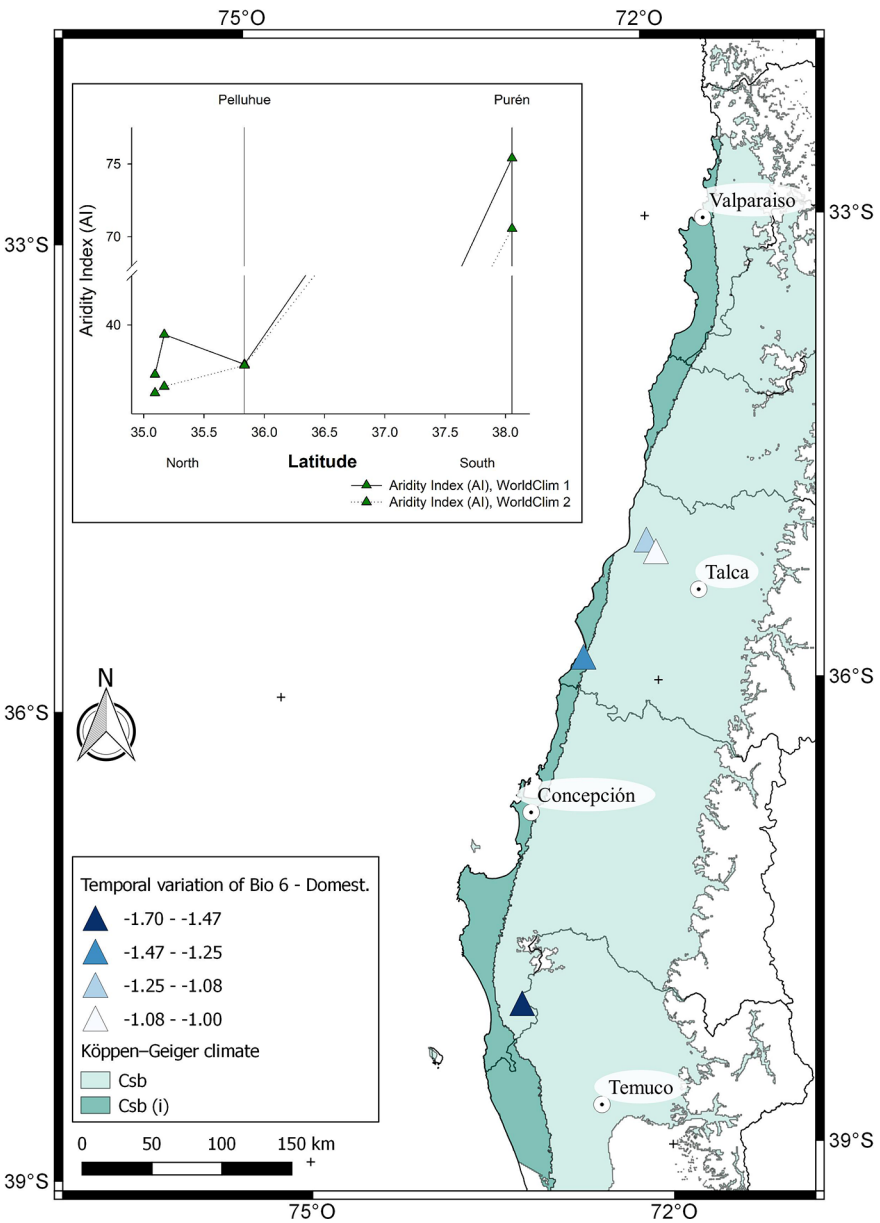


FIGURE 6 | Climatic characterization of the localities where *F. chiloensis* f. *chiloensis* is cultivated in Chile and distributed from North to South. The comparison of Bioclimatic variable 6 (Min temperature of coldest month) between WorldClim 1 and 2 for each location is shown in the map. In the insert, De Martonne Aridity Index (AI) for each locality.

WorldClim 1. In terms of aridity index, there is a significant increment of around sixteen points for the location of Cucao in WorldClim 2 compared to WorldClim 1; a smaller increment in aridity index (4 points) was also observed in Termas de Chillán (**Figure 7**, insert). The particular climate properties of the Andes Mountains (high precipitation level, lower annual average temperature) could constitute a good refuge for this wild species, restricting current distributions to these areas (Bambach et al., 2013; Alarcón and Cavieres, 2015). In the case of cultivated species, the Andes Mountains can also constitute a refuge for the species, but its displacement to southern regions could be dedicated for agriculture development (Parajuli et al., 2019).

CROP POTENTIAL

Vasconcellea species has an interesting potential as a new crop, considering potential breeding for fruit and other sub-products and its domestication in specific geographic areas (Scheldeman et al., 2011). The two most important species are the babaco

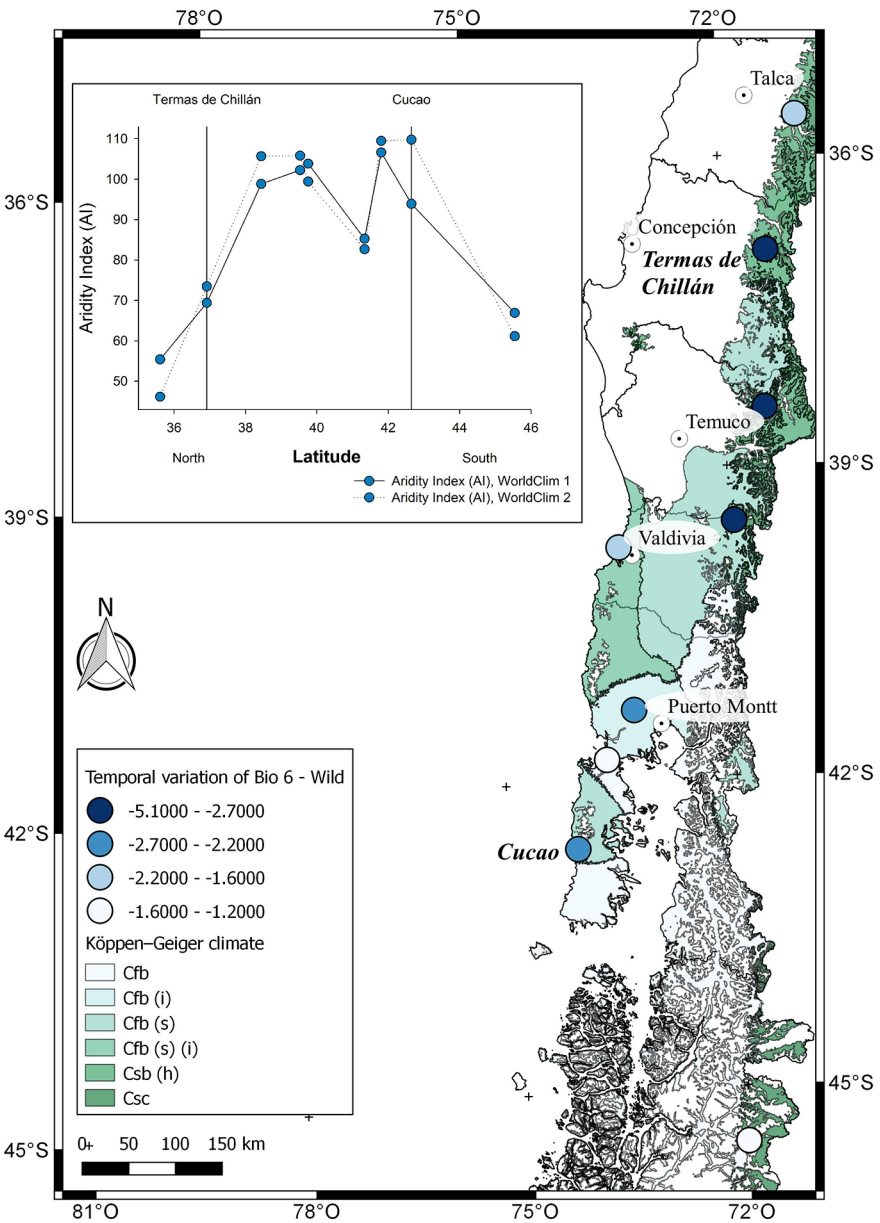


FIGURE 7 | Climatic characterization of the localities where *F. chiloensis* f. *patagonica* is distributed in Chile, from North to South. The comparison of Bioclimatic variable 6 (Min temperature of coldest month) between WorldClim 1 and 2 for each locality in which *F. chiloensis* f. *patagonica* grows are shown in the map. In the insert, De Martonne Aridity Index (AI) for each locality. On the X axis, the localities of Termas de Chillán and Cucao are marked as reference, as previously shown in **Figure 1**.

(*Vasconcellea x heilbornii*) in Ecuador and Colombia, and mountain papaya (*Vasconcellea pubescens*), in all Andean countries and particularly important at small commercial scale in Colombia (Scheldeman et al., 2011) and Chile (Moya-León et al., 2004). Interestingly, the cultivated surface of mountain papaya almost disappears after the earthquake and Tsunami of 2010 in the Maule Region, mainly because natural growing areas of the species were salinized and damaged by seawater. In the Maule Region, toxic levels of salty soil inhibit the growth of *V. pubescens* plants, because the content of salt reduces water availability to plants. In addition, the reduction in soil pH in depth (1.4 units) and surface (0.7 units) resulted as consequence of the ionic strength generated and the coarse-textured soil (Casanova et al., 2016). Remarkably, an increase in the cultivated surface of this species has been recently observed in the northern extreme of Chile (Arica and Parinacota Region, ~

18° 30' latitude South), where 13.6 ha were recorded that represents 9% of total cultivated surface for the species in Chile (ODEPA-CIREN, 2019).

F. chiloensis subsp. *chiloensis* has also potential as a new crop. Firstly, the production of phytochemicals with health benefits highlights the opportunity to use the species as a functional food. Secondly, the species has great resistance to several diseases (Rojas et al., 2013). Although Aphidborne virus (ABVs) has been detected in *F. chiloensis* subsp. *chiloensis* plants non visual symptoms of disease were determined, albeit the virus affects severely *F. vesca* plants. Similarly, *F. chiloensis* f *chiloensis* shows tolerance to *Botrytis cinerea*, a severe fungal disease which affects dramatically several fruit species (González et al., 2013). Also, it was reported that the Chilean strawberry presented higher tolerance to the fungus compared to *F.* × *ananassa*.

Ploidy and hybridization are the basis of current cultivated crops. Importantly, wild relatives have impacted positively the availability of crops, allowing the expansion of cultivation areas or increasing the tolerance to pest and diseases. In a climate change scenario, a breeding program is needed in order to obtain smart cultivars that can face the new challenges as a way to ensure sustainable production in the future (Borrell et al., 2020). Fruit homogeneity and other phenotypic characters have been the main effort of plant breeders, but new characteristic must be incorporated in modern crops, such as, tolerance to cold and high temperature environment, tolerance to drought or salinity and disease resistance (Doebley et al., 2006). May be taking into account the genetic variability of native species for those traits and using strategies like introgression or hybridization, the genetic improvement of these valuable crops can be possible.

Genomic selection has aided marker assisted selection to gain time during crop improvement (Heffner et al., 2010). Moreover, the incorporation of new tools for molecular breeding, such as genome sequencing, SNPs, or genome editing, facilitates the identification of new inbreeds; this strategy has been defined as design breeding (Heffner et al., 2011; Fernie and Yan, 2019).

Domestication of non-domesticated plants is a great opportunity to obtain cultivars well adapted to different edaphoclimatic niches. The long period of time required for selection from breeding programs can be accelerated by the use of a biotechnology approach and new genome editing techniques (Osterberg et al., 2017; Scheben et al., 2017; Wallace et al., 2018). Clearly, under the climate change scenario new type of cultivars are needed. The better adaptability to a variety of climatic conditions is a priority in current time as a way to reduce the negative impact of less food production in vulnerable environments.

Polyploidy can be a challenge for Chilean strawberry, but combining the information from genetic maps, the functional characterization of specific genes, and genome sequencing should be the starting point for applying this modern strategy (Rousseau-Gueutin et al., 2008; Davik et al., 2015; Vining et al., 2017; Edger et al., 2019; Moya-León et al., 2019). In the case of *Vasconcellea pubescens* there is no genomic or functional genomic data available until now, which can limit the breeding effort. However, wild plants are already adapted to a wide range of climatic conditions, considering their native habitat at high altitude of the Andes mountains. Chilean papaya is attractive as functional food and source of natural sub-products as papain proteases which can be the target for molecular breeding. For certain, the implementation of what is called breeding 4.0 or "*de novo* domestication" should provide new well adapted smart cultivars.

AUTHOR CONTRIBUTIONS

The present work was conceived and designed by RH, CG-E, LL, PP, and MM-L. LL and CG-E contributed collecting information and preparing figures. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

This study was funded by FONDECYT grant 1171530. The funders had no part in the design of the study or collection, analysis, and interpretation of data, neither in writing the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.01002/full#supplementary-material

REFERENCES

Şerban, P., Wilson, J. R. U., Vamosi, J. C., and Richardson, D. M. (2008). Plant diversity in the human diet: Weak phylogenetic signal indicates breadth. *BioScience* 58, 151–159. doi: 10.1641/b580209

Aceituno-Valenzuela, U., Covarrubias, M. P., Aguayo, M. F., Valenzuela-Riffo, F., Espinoza, A., Gaete-Eastman, C., et al. (2018). Identification of a type II cystatin in *Fragaria chiloensis*: a proteinase inhibitor differentially regulated during achene development and in response to biotic stress-related stimuli. *Plant Physiol. Biochem.* 129, 158–167. doi: 10.1016/j.plaphy.2018.05.021

Alarcón, D., and Cavieres, L. A. (2015). In the right place at the right time: Habitat representation in protected areas of south american *Nothofagus*-dominated plants after a dispersal constrained climate change scenario. *PloS One* 10 (3), e0119952. doi: 10.1371/journal.pone.0119952

Arroyo, M. T. K., Rozzi, R., Simonetti, J. A., Marquet, P., and Salaberry, M. (1999). "Central Chile," in *Hotspots: Earth"s Biologically Richest and Most Endangered Terrestrial Ecosystems*. Eds. R. A. Mittermeier, N. Myers, P. Robles Gil and C. Goettsch Mittermeier (México, México: Cemex, Conservation International), 161–171.

Avila, F., Theoduloz, C., López-Alarcón, C., Dorta, E., and Schmeda-Hirschmann, G. (2017). Cytoprotective mechanisms mediated by polyphenols from Chilean native berries against free radical-induced damage on AGS cells. *Oxid. Med. Cell. Longev.* 2017 (1). doi: 10.1155/2017/9808520

Azodanlou, R., Darbellay, C., Luisier, J. L., Villettaz, J. C., and Amado, R. (2003). Quality assessment of strawberries (*Fragaria* species). *J. Agric. Food Chem.* 51, 715–721. doi: 10.1021/jf0200467

Badillo, V. M. (1971). *Monografia de la familia Caricaceae* (Maracay, Venezuela: Asociación de Profesores, Univ. Cent. Venezuela).

Badillo, V. M. (1993). *Caricaceae. Segundo esquema* (Maracay, Venezuela: Rev. Fac. Agron. Univ. Cent. Venezuela, Alcance 43).

Badillo, V. M. (2000). *Carica* L. vs *Vasconcella* St. Hil. (Caricaceae): con la rehabilitación de este último. *Ernstia* 10, 74–79.

Baeza, G., Correa, D., and Salas, C. (1990). Proteolytic enzymes in *Carica candamarcensis*. *J. Sci. Food Agric.* 51, 1–9. doi: 10.1002/jsfa.2740510102

Balbontín, C., Gaete-Eastman, C., Vergara, M., Herrera, R., and Moya-León, M. A. (2007). Treatment with 1-MCP and the role of ethylene in aroma development of mountain papaya fruit. *Postharv. Biol. Technol.* 43, 67–77. doi: 10.1016/j.postharvbio.2006.08.005

Bambach, N., Meza, F. J., Gilabert, H., and Miranda, M. (2013). Impacts of climate change on the distribution of species and communities in the Chilean Mediterranean ecosystem. *Reg. Environ. Change* 13, 1245–1257. doi: 10.1007/s10113-013-0425-7

Bassil, N. V., Davis, T. M., Alperin, E., Amaya, I., Bellon, F., Brew, F., et al. (2015). Development and preliminary evaluation of a 90K Axiom SNP array in the allo-octoploid cultivated strawberry *Fragaria* × *ananassa*. *BMC Genomics* 16, 155. doi: 10.1186/s12864-015-1310-1

Berna, A. Z., Geysen, S., Li, S., Verlinden, B., Lammertyn, J., and Nicolai, B. (2007). Headspace fingerprint mass spectrometry to characterize strawberry aroma at super-atmospheric oxygen conditions. *Postharv. Biol. Technol.* 46, 230–236. doi: 10.1016/j.postharvbio.2007.05.011

Borrell, J. S., Dodsworth, S., Forest, F., Pérez-Escobar, O. A., Lee, M. A., Mattana, E., et al. (2020). The climatic challenge: Which plants will people use in the next century? *Environ. Exp. Bot.* 170, 103872. doi: 10.1016/j.envexpbot.2019.103872

Bringhurst, R. S. (1990). Cytogenetics and evolution in american *Fragaria*. *HortScience* 25, 879–881. doi: 10.21273/HORTSCI.25.8.879

Carrasco, B., Garcés, M., Rojas, P., Saud, G., Herrera, R., Retamales, J. B., et al. (2007). The Chilean strawberry [*Fragaria chiloensis* (L.) Duch.]: Genetic diversity and structure. *J. Amer. Soc Hortic. Sci.* 132, 501–506. doi: 10.21273/JASHS.132.4.501

Carrasco, B., Avila, P., Pérez-Diaz, J., Muñoz, P., Garcia, R., Lavandero, B., et al. (2009). Genetic structure of highland papayas (*Vasconcellea pubescens* (A. DC.) Badillo) cultivated along a geographic gradient in Chile as revealed by inter simple sequence repeats (ISSR). *Genet. Resour. Crop Evol.* 56, 331–337. doi: 10.1007/s10722-008-9367-1

Carrasco-Orellana, C., Stappung, J., Méndez-Yañez, A., Allan, A., Espley, R., Plunkett, B., et al. (2018). Characterization of a ripening related transcription factor FcNAC1 from *Fragaria chiloensis* fruits. *Sci. Rep.* 8, 10524. doi: 10.1038/s41598-018-28226-y

Casanova, M., Salazar, O., Oyarzun, I., Tapia, Y., and Fajardo, M. (2016). Field monitoring of 2010-tsunami impact on agricultural soils and irrigation waters: Central Chile. *Water Air Soil Pollut.* 227, 411. doi: 10.1007/s11270-016-3113-7

Chamorro, M., Reiner, G., Theoduluz, C., Ladio, A., Schmeda-Hirschmann, G., Gomez-Alonso, S., et al. (2019). Polyphenol composition and (Bio)activity of *Berberis* species and wild strawberry from the Argentinean patagonia. *Molecules* 24, 3331. doi: 10.3390/moleculaes24183331

Cheel, J., Theoduloz, C., Rodriguez, J., Saud, G., Caligari, P. D. S., and Schmeda-Hirschmann, G. (2005). E-Cinnamic acid derivatives and phenolics from chilean strawberry fruits, *Fragaria chiloensis* ssp. *chiloensis*. *J. Agric. Food Chem.* 53 (22), 8512–8518. doi: 10.1021/jf051294g

Cheel, J., Theoduloz, C., Rodriguez, J. A., Caligari, P. D. S., and Schmeda-Hirschmann, G. (2007). Free radical scavenging activity and phenolic content in achenes and thalamus from *Fragaria chiloensis* ssp. *chiloensis*, *F. vesca* and *F.* × *ananassa* cv. Chandler. *Food Chem.* 102, 36–44. doi: 10.1016/j.foodchem.2006.04.036

Darrow, G. W. (1966). *The strawberry: History, breeding and physiology* (New York, USA: Holf, Rinehart and Winston).

Davik, J., Sargent, D., Brurberg, M., Lien, S., Kent, M., and Alsheikh, M. (2015). A ddRAD based linkage map of the cultivated strawberry, *Fragaria x ananassa*. *PloS One* 10, e0137746. doi: 10.1371/journal.pone.0137746

De Zerpa, D. M. (1959). Citologia de hibridos interespecificos en *Carica*. *Agron. Trop. (Maracay)* 8, 135–144.

Doebley, J., Gaut, B., and Smith, B. (2006). The molecular genetics of crop domestication. *Cell* 127, 1309–1321. doi: 10.1016/j.cell.2006.12.006

Edger, P. P., Poorten, T. J., VanBuren, R., Hardigan, M. A., Colle, M., McKain, M. R., et al. (2019). Origin and evolution of the octoploid strawberry genome. *Nat. Genet.* 51, 541–547. doi: 10.1038/s41588-019-0356-4

El Moussaoui, A., Nijs, M., Paul, C., Wintjens, R., Vincentelli, J., Azarkan, M., et al. (2001). Revisiting the enzymes stored in the laticifers of *Carica papaya* in the context of their possible participation in the plant defence mechanism. *Cell Mol. Life Sci.* 58, 556–570. doi: 10.1007/PL00000881

Feng, S., Krueger, A. B., and Oppenheimer, M. (2010). Linkages among climate change, crop yields and Mexico-US cross-border migration. *PNAS* 107, 14257–14262. doi: 10.1073/pnas.1002632107

Fernie, A., and Yan, J. (2019). De Novo Domestication: An Alternative Route toward New Crops for the Future. *Mol. Plant* 12, 615–631. doi: 10.1016/j.molp.2019.03.016

Fick, S. E., and Hijmans, R. J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* 37, 4302–4315. doi: 10.1002/joc.5086

Figueroa, C. R., Pimentel, P., Gaete-Eastman, C., Moya, M., Herrera, R., Caligari, P. D. S., et al. (2008). Softening rate of the Chilean strawberry (*Fragaria chiloensis*) fruit reflects the expression of polygalacturonase and pectate lyase genes. *Postharv. Biol. Technol.* 49, 210–220. doi: 10.1016/j.postharvbio.2008.01.018

Figueroa, C. R., Pimentel, P., Dotto, M. C., Civello, P. M., Martínez, G. A., Herrera, R., et al. (2009). Expression of five expansin genes during softening of *Fragaria chiloensis* fruit. Effect of auxin treatment. *Postharv. Biol. Technol.* 53, 51–57. doi: 10.1016/j.postharvbio.2009.02.005

Finn, C., Retamales, J., Lobos, G., and Hancock, J. (2013). The Chilean strawberry (*Fragaria chiloensis*): over 1000 years of domestication. *HortScience* 48, 418–421. doi: 10.21273/HORTSCI.48.4.418

Forney, C. F., Kalt, W., and Jordan, M. A. (2000). The composition of strawberry aroma is influenced by cultivar, maturity and storage. *HortScience* 35, 1022–1026. doi: 10.21273/HORTSCI.35.6.1022

García, B. H. (1975). *Flora medicinal de Colombia* (Bogotá, Colombia: Imprenta Nacional).

González, G., Moya, M., Sandoval, C., and Herrera, R. (2009a). Genetic diversity in Chilean strawberry (*Fragaria chiloensis*): differential response to *Botrytis cinerea* infection. *Span. J. Agric. Res.* 7, 886–895. doi: 10.5424/sjar/2009074-1102

González, M., Gaete-Eastman, C., Valdenegro, M., Figueroa, C. R., Fuentes, L., Herrera, R., et al. (2009b). Aroma development during ripening of *F. chiloensis* fruit and participation of an alcohol acyltransferase (FcAAT1) gene. *J. Agric. Food Chem.* 57, 9123–9132. doi: 10.1021/jf901693j

González, G., Fuentes, L., Moya-León, M. A., Sandoval, C., and Herrera, R. (2013). Characterization of two PR genes from *Fragaria chiloensis* in response to *Botrytis cinerea* infection: A comparison with *Fragaria* × *ananassa*. *Physiol. Mol. Plant P.* 82, 73–80. doi: 10.1016/j.pmpp.2013.02.001

Govindaraj, M., Vetriventhan, M., and Srinivasan, M. (2015). Importance of genetic diversity assessment in crop plants and its recent advances: An overview of its analytical perspectives. *Genet. Res. Inter.* 2015, 431487. doi: 10.1155/2015/431487

Hancock, J. F., Lavin, A., and Retamales, J. B. (1999). Our sourthern strawberry heritage: *Fragaria chiloensis* of Chile. *HortScience* 34, 814–816. doi: 10.21273/HORTSCI.34.5.814

Hancock, J. F., Callow, P. W., Dale, A., Luby, J. J., Finn, C. E., Hokanson, S., et al. (2001). From the Andes to the rockies: Native strawberry collection and utilization. *Hortscience* 36, 221–225. doi: 10.21273/HORTSCI.36.2.221

Hancock, J. F., Finn, C. E., Luby, J. J., Dale, A., Callow, P. W., and Serçe, S. (2010). Reconstruction of the strawberry, *Fragaria* × *ananassa*, using genotypes of *F. virginiana* and *F. chiloensis*. *HortScience* 45, 1006–1013. doi: 10.21273/HORTSCI.45.7.1006

Handford, M., Espinoza, A., Bracamonte, M., Figueroa, A., Zapata, S., Aceituno, U., et al. (2014). Identification and characterization of key genes involved in fruit ripening of the Chilean strawberry. *New Biotechnol.* 31, S182–S182. doi: 10.1016/j.nbt.2014.05.912

Hassler, M. (2019). "World plants: Synonymic checklists of the vascular plants of the world (version Nov 2018)," in *Species 2000 & ITIS Catalogue of Life, 2019 Annual Checklist*. Eds. Y. Roskov, G. Ower, T. Orrell, D. Nicolson, N. Bailly, P. M. Kirk, T. Bourgoin, R. E. DeWalt, W. Decock, E. Nieukerken, J. van, Zarucchi and L. (Netherlands: Naturalis Biodiversity Center), Penev Digital resource at www.catalogueoflife.org/annual-checklist/2019. Species 2000: Naturalis, Leiden, the Netherlands. ISSN 2405-884X.

Heffner, E. L., Lorenz, A. J., Jannink, J. L., and Sorrells, M. E. (2010). Plant breeding with genomic selection: gain per unit time and cost. *Crop Sci.* 50, 1681–1690. doi: 10.2135/cropsci2009.11.0662

Heffner, E. L., Jannink, J., Iwata, H., Souza, E., and Sorrells, M. E. (2011). Genomic selection accuracy for grain quality traits in biparental wheat population. *Crop Sci.* 51, 2597–2606. doi: 10.2135/cropsci2011.05.0253

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., and Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* 25, 1965–1978. doi: 10.1002/joc.1276

Hilmarsson, H., Hytonen, T., Isobe, S., Goransson, M., Toivanen, T., and Hallsson, J. (2017). Population genetic analysis of a global collection of *Fragaria vesca* using microsatellite markers. *PloS One* 12, 8. doi: 10.1371/journal.pone.0183384

Hoisington, D., Khairallah, M., Reeves, T., Ribaut, J. M., Skovmand, B., Taba, S., et al. (1999). Plant genetic resources: What can they contribute toward increasd crop productivity? *Proc. Natl. Acad. Sci. U. S. A.* 96, 5937–5943. doi: 10.1073/pnas.96.11.5937

Honjo, M., Nunome, T., Kataoka, S., Yano, T., Yamazaki, H., Hamano, M., et al. (2011). Strawberry cultivar identification based on hypervariable SSR markers. *Breed. Sci.* 61, 420–425. doi: 10.1270/jsbbs.61.420

Horovitz, S., and Jimenez, H. (1967). Cruzamientos interespecíficos e intergenéricos en Caricaceas y sus implicaciones fitotécnicas. *Agron. Trop. (Maracay)* 17, 323–343.

Hwang, H., Kim, Y. J., and Shin, Y. (2019). Influence of ripening stage and cultivar on physicochemical propierties, sugar and organic acid profiles, and antioxidant compositions of strawberries. *Food Sci. Biotechnol.* 28, 1659–1667. doi: 10.1007/s10068-019-00610-y

Jung, H., Veerappan, K., Natarajan, S., Jeong, N., Hwang, I., Nagano, S., et al. (2017). A system for distinguishing octoploid strawberry cultivars using high-throuput SNP genotyping. *Trop. Plant Biol.* 10, 68–76. doi: 10.1007/s12042-017-9185-8

Kalkman, C. (2004). "Rosaceae," in *The families and genera of vascular plants: Flowering plants dicotyledons*. Ed. K. Kubitzki (Berlin, Deutschland: Springer Berlin Heidelberg), 343–386.

Khoshbakh, K., and Hammer, K. (2008). How many plant species are cultivated? *Gen. Resour. Crop Evol.* 55, 925–928. doi: 10.1007/s10722-008-9368-0

Konno, K., Hirayama, C., Nakamura, M., Tateishi, K., Tamura, Y., Hattori, M., et al. (2004). Papain protects papaya trees from herbivorous insects: role of cysteine proteases in latex. *Plant J.* 37, 370–378. doi: 10.1046/j.1365-313X.2003.01968.x

Kyndt, T., Romeijn-Peeters, E., Van Droogenbroeck, B., Romero-Motochi, J. P., Gheysen, G., and Goetghebeur, P. (2005). Species relationships in the genus *Vasconcellea* (*Caricaceae*) based on molecular and morphological evidence. *Am. J. Bot.* 92, 1033–1044. doi: 10.3732/ajb.92.6.1033

Kyndt, T., Van Droogenbroeck, B., Haegeman, A., Roldán-Ruiz, I., and Gheysen, G. (2006). Cross-species microsatellite amplification in *Vasconcellea* and related genera and their use in germplasm classification. *Genome* 49, 786–798. doi: 10.1139/g06-035

Latcham, R. (1936). *La agricultura precolombina en Chile y en los países vecinos* (Santiago, Chile: Ediciones de la Universidad de Chile).

Latrasse, A. (1991). "Fruits III," in *Volatile compounds in foods and beverages*. Ed. H. Maarse (New York, USA: Marcel Dekker, Inc.), 334–340.

Lavin, A., Del Pozo, A., and Maureira, M. (2000). Present distribution of *Fragaria chiloensis* (L.) Duch. in Chile. *Plant Genet. Resour. Newsl.* 122, 24–28.

Lemos, F. O., Dittz, D., Santos, V. G., Pires, S. F., de Andrade, H. M., Salas, C. E., et al. (2018). Cysteine proteases from *V. cundinamarcensis* (*C. candamarcensis*) inhibit melanoma metastasis and modulate expression of proteins related to proliferation, migration and differentiation. *Int. J. Mol. Sci.* 19, 2846. doi: 10.3390/ijms19102846

Lippmann, R., Babben, S., Menger, A., Delker, C., and Quint, M. (2019). Development of wild and cultivated plants under Global Warming conditions. *Curr. Biol.* 29, R1326–R1338. doi: 10.1016/j.cub.2019.10.016

Liston, A., Cronn, R., and Ashman, T. L. (2014). *Fragaria*: a genus with deep historical roots and ripe for evolutionary and ecological insights. *Am. J. Bot.* 101, 1686–1699. doi: 10.3732/ajb.1400140

Luque, A., Debes, M., Perera, M., Castagnaro, A., and Arias, M. (2019). Reproductive compatibility studies between wild and cultivated strawberries (*Fragaria* × *ananassa*) to obtain "bridge species" for breeding programmes. *Plant Breed.* 138, 229–238. doi: 10.1111/pbr.12681

Méndez-Yáñez, A., Beltrán, D., Campano-Romero, C., Molinett, S., Herrera, R., Moya-León, M. A., et al. (2017). Glycosylation is important for FcXTH1 activity as judged by its structural and biochemical characterization. *Plant Physiol. Biochem.* 119, 200–210. doi: 10.1016/j.plaphy.2017.08.030

Méndez-Yáñez, A., González, M., Carrasco-Orellana, C., Herrera, R., and Moya-León, M. A. (2020). Isolation of a rhamnogalacturonan lyase expressed during ripening of the Chilean strawberry fruit and its biochemical characterization. *Plant Physiol. Biochem.* 146, 411–419. doi: 10.1016/j.plaphy.2019.11.041

Mahoney, L., Sargent, D., Abele-Akele, F., Wood, D., Ward, J., Bassil, N., et al. (2016). A high density linkage map of the ancestral diploid strawberry, *Fragaria iinumae*, constructed with single nucleotide polymorphism markers from the IStraw90 array and genotyping by sequencing. *Plant Genome* 9, 2. doi: 10.3835/plantgenome2015.08.0071

Marticorena, A. (2019). "Rosaceae Juss," in *Flora de Chile: Berberidopsidaceae - Saxifragaceae*. Eds. R. Rodríguez and A. Marticorena (Concepción, Chile: Editorial Universidad de Concepción), 185–249.

Mazzoni, L., Giampieri, F., Suarez, J., Gasparrini, M., Mezzetti, B., Hernandez, T., et al. (2019). Isolation of strawberry anthocyanin-rich fractions and their mecganisms of action against murine breast cancer cell lines. *Food Funct.* 10, 7103–7120. doi: 10.1039/c9fo01721f

Mello, V. J., Gomes, M. T. R., Lemos, F. O., Delfino, J. L., Andrade, S. P., Lopes, M. T. P., et al. (2008). The gastric ulcer protective and healing role of cysteine proteinases from *Carica candamarcensis*. *Phytomedicine* 15, 237–244. doi: 10.1016/j.phymed.2007.06.004

Mittermeier, R. A., Robles Gil, P., Hoffmann, M., Pilgrim, J., Brooks, T., Mittermeier, C. G., et al. (2004). *Hotspots revisited: Earth"s biologically richest and most endangered terrestrial ecoregions* (Mexico City, Mexico: CEMEX).

Molinett, S., Nuñez, F., Moya-León, M. A., and Zúniga-Hernández, J. (2015). Chilean strawberry consumption protects against LPS-induced liver injury by anti-inflammatory and antioxidant capability in Sprague-Dawley rats. *Evid. Based. Compl. Alt. Med.* 2015 (1). doi: 10.1155/2015/320136

Monfort, A., Vilanova, S., Davis, T. M., and Arus, P. (2006). A new set of polymorphic simple sequence repeat (SSR) markers from a wild strawberry (*Fragaria vesca*) are transferable to other diploid *Fragaria* species and to *Fragaria* × *ananassa*. *Mol. Ecol. Notes* 6, 197–200. doi: 10.1111/j.1471-8286.2005.01191.x

Morales, A. L., and Duque, C. (1987). Aroma constituents of the fruit of the mountain papaya (*Carica pubescens*) from Colombia. *J. Agric. Food Chem.* 35, 538–540. doi: 10.1021/jf00076a024

Moran, E. V. (2020). Simulating the effects of local adaptation and life history on the ability of plants to track climate shifts. *AoB Plants* 12 (1), plaa008. doi: 10.1093/aobpla/plaa008

Moreira-Muñoz, A. (2011). *Plant geography of Chile* (London, UK and New York, USA: Springer Dordrecht Heidelberg).

Moya-León, M. A., Moya, M., and Herrera, R. (2004). Ripening of the Chilean papaya fruit (*Vasconcellea pubescens*) and ethylene dependence of some ripening events. *Postharv. Biol. Technol.* 34, 211–218. doi: 10.1016/j.postharvbio.2004.05.005

Moya-León, M. A., Mattus, E., and Herrera, R. (2019). Molecular events occurring during softening of strawberry fruit. *Front. Plant Sci.* 10, 615. doi: 10.3389/fpls.2019.00615

Myers, N., Mittermeier, R. A., Mittermeier, C. G., Da Fonseca, G. A. B., and Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature* 403, 853–858. doi: 10.1038/35002501

Na, J.-K., Wang, J., and Ming, R. (2014). Accumulation of interspersed and sex-specific repeats in the non-recombining region of papaya sex chromosomes. *BMC Genomics* 15, 335. doi: 10.1186/1471-2164-15-335

Nagano, S., Shirasawa, K., Hirakawa, H., Maeda, F., Ishikawa, M., and Isobe, S. (2017). Discrimination of candidate subgenome specific loci by linkage map construction with an S-1 population of octoploid strawberry (*Fragaria* × *ananassa*). *BMC Genomics* 18, 374. doi: 10.1186/s12864-017-3762-y

Njuguna, W., Liston, A., Cronn, R., Ashman, T.-L., and Bassil, N. (2013). Insights into phylogeny, sex function and age of *Fragaria* based on whole chloroplast genome sequencing. *Mol. Phylogenet. Evol.* 66, 17–29. doi: 10.1016/j.ympev.2012.08.026

Oñate, F. A., Hasbún, R., Mora, F., and Figueroa, C. R. (2018). Linkage disequilibrium and population structure in *Fragaria chiloensis* revealed by SSR markers transferred from commercial strawberry. *Acta Sci. Agron.* 40, e34966. doi: 10.4025/actasciagron.v40i1.34966

ODEPA-CIREN (2019). *Catastro Frutícola Nacional, superficie plantada nacional.* Santiago de Chile, Chile: Oficina de Estudios y Políticas Agrarias (ODEPA) y Centro de Información de Recursos Naturales (CIREN). Available at: https://www.odepa.gob.cl/wp-content/uploads/2019/09/Sup-plantada-nacional-web_20.08.2019.xlsx. Accessed September 2019.

Oh, Y., Zurn, J. D., Bassil, N., Edger, P. P., Knapp, S. J., Whitaker, V. M., et al. (2019). The Strawberry DNA Testing Handbook. *HortScience* 54 (12), 2267–2270. doi: 10.21273/HORTSCI14387-19

Oh, Y., Chandra, S., and Lee, S. (2020). Development of subgenome specific markers for FaRXf1 conferring resistance to bacterial angular leaf spot in allo-octoploid strawberry. *Int. J. Fruit Sci.* 20, 1–13. doi: 10.1080/15538362.2019.1709116

Opazo, M. C., Lizana, R., Pimentel, P., Herrera, R., and Moya-León, M. A. (2013). Changes in the mRNA abundance of *FcXTH1* and *FcXTH2* promoted by hormonal treatments of *Fragaria chiloensis* fruit. *Postharv. Biol. Technol.* 77, 28–34. doi: 10.1016/j.postharvbio.2012.11.007

Osterberg, J. T., Xiang, W., Olsen, L. I., Edenbrandt, A. K., Vedel, S. E., Christiansen, A., et al. (2017). Accelerating the domestication of new crops: feasibility and approaches. *Trends Plant Sci.* 22, 373–384. doi: 10.1016/j.tplants.2017.01.004

Pérez, A. G., Rios, J. J., Sanz, C., and Olías, J. M. (1992). Aroma components and free amino acids in strawberry variety Chandler during ripening. *J. Agric. Food Chem.* 40, 2232–2235. doi: 10.1021/jf00023a036

Parajuli, R., Thoma, G., and Matlock, M. D. (2019). Environmental sustainability of fruit and vegetable production supply chains in the face of climate change: A review. *Sci. Total Environ.* 650, 2863–2879. doi: 10.1016/j.scitotenv.2018.10.019

Pardo, O., and Pizarro, J. L. (2013). *Chile: Plantas alimentarias prehispánicas* (Arica, Chile: Ediciones Parina EIRL).

Parra-Palma, C., Fuentes, E., Palomo, I., Torres, C. A., Moya-León, M. A., and Ramos, P. (2018). Linking the platelet antiaggregation effect of different strawberries species with antioxidants: Metabolomic and transcript profiling of polyphenols. *Bol. Latinoam. Caribe Plantas Med. Aromat. (Blacpma)* 17, 36–52.

Pimentel, P., Salvatierra, A., Moya-León, M. A., and Herrera, R. (2010). Isolation of genes differentially expressed during development and ripening of *Fragaria chiloensis* fruit by suppression subtractive hybridization. *J. Plant Physiol.* 167, 1179–1187. doi: 10.1016/j.plph.2010.03.006

Pliscoff, P., Arroyo, M. T. K., and Lohengrin, C. (2012). Changes in the main vegetation types of Chile predicted under climate change based on a preliminary study: Models, uncertainties and adapting research to a dynamic biodiversity world. *Anales Instituto Patagonia (Chile)* 40, 81–86. doi: 10.4067/S0718-686X2012000100010

Popenoe, W. (1921). The Frutilla, or Chilean strawberry. *J. Hered.* 12, 457–466. doi: 10.1093/oxfordjournals.jhered.a102048

Pyysalo, T., Honkanen, E., and Hirvi, T. (1979). Volatiles of wild strawberries, *Fragaria vesca* L., compared to those of cultivated berries, *Fragaria ananassa* cv. Senga Sengana. *J. Agric. Food Chem.* 27, 19–22. doi: 10.1021/jf60221a042

Retamales, J. B., Caligari, P., Carrasco, B., and Saud, G. (2005). Current status of the Chilean strawberry and the research needs to convert the species into a commercial crop. *HortScience* 40, 1633–1634. doi: 10.21273/hortsci.40.6.1633

Rodríguez, R., Marticorena, C., Alarcón, D., Baeza, C., Cavieres, L., Finot, V. L., et al. (2018). Catálogo de las plantas vasculares de Chile. *Gayana Bot.* 75, 1–430. doi: 10.4067/S0717-66432018000100001

Rojas, P., Almada, R. D., Sandoval, C., Keller, K. E., Martin, R. R., and Caligari, P. D. S. (2013). Occurrence of aphidborne viruses in southernmost South American populations of *Fragaria chiloensis* ssp. *chiloensis*. *Plant Pathol.* 62, 428–435. doi: 10.1111/j.1365-3059.2012.02643.x

Rousseau-Gueutin, M., Lercerteau-Kohler, E., Barrot, L., Sargent, D., Monfort, A., Simpson, D., et al. (2008). Comparative genetic mapping between octoploid and diploid fragaria species reveals a high level of colinearity between their genomes and the essentially disomic behavoir of the cultivated octoploid strawberry. *Genetics* 179, 2045–2060. doi: 10.1534/genetics.107.083840

Sánchez, I. (1994). "Andean Fruits," in *Neglected Crops: 1492 from a different perspective, Plant Production and Protection Series N 26*. Eds. J. E. Hernándo Bermejo and J. Leon (Rome, Italy: FAO), 181–191.

Salvatierra, A., Pimentel, P., Moya-León, M. A., Caligari, P. D. S., and Herrera, R. (2010). Comparison of transcriptional profiles of flavonoid genes and anthocyanin contents during fruit development of two botanical forms of *Fragaria chiloensis* ssp. chiloensis. *Phytochemistry* 71, 1839–1847. doi: 10.1016/j.phytochem.2010.08.005

Salvatierra, A., Pimentel, P., Moya-León, M. A., and Herrera, R. (2014). Biosynthesis of flavonoids in achenes of *Fragaria chiloensis* ssp. *chiloensis*. *Bol. Latinoam. Caribe Plantas Med. Aromat. (Blacpma)* 13 (4), 406–414.

Sarricolea, P., Herrera-Ossandon, M., and Meseguer-Ruiz, Ó. (2017). Climatic regionalisation of continental Chile. *J. Maps* 13, 66–73. doi: 10.1080/17445647.2016.1259592

Scheben, A., Wolter, F., Batley, J., Puchta, H., and Edwards, D. (2017). Towards CRIPS/Cas crops – bringing together genomics and genome editing. *New Phytol.* 216, 682–698. doi: 10.1111/nph.14702

Scheldeman, X., Romero, J. P., Van Damme, V., Heyens, V., and Van Damme, P. (2003). Potential of highland papayas (*Vasconcella* spp.) in southern Ecuador. *Lyonia* 5, 73–80.

Scheldeman, X., Kyndt, T., d' Eeckenbrugge, G. C., Ming, R., Drew, R., Van Droogenbroeck, B., et al. (2011). "*Vasconcellea*," in *Wild crop relatives: Genomic and breeding resources*. Ed. C. Kole (Berlin, Heidelberg: Springer), 213–249. doi: 10.1007/978-3-642-20447-0_11

Schreier, P. (1980). Quantitative composition of volatile constituents in cultivated strawberries, *Fragaria ananassa* cv. Senga Sengana, Senga Litessa and Senga Gourmella. *J. Sci. Food Agric.* 31, 487–494. doi: 10.1002/jsfa.2740310511

Shimomura, K., and Hirashima, K. (2006). Development and characterization of simple sequence repeats (SSR) as markers to identify strawberry cultivars (*Fragaria* × *ananassa* Duch.). *J. Jpn. Soc Hortic. Sci.* 75, 399–402. doi: 10.2503/jjshs.75.399

Simirgiotis, M., Caligari, P. D. S., and Schmeda-Hirschmann, G. (2009). Identification of phenolic compounds from the fruits of the mountain papaya *Vasconcellea pubescens* A. DC. grown in Chile by liquid chromatography-UV detection-mass spectrometry. *Food Chem.* 115, 775–784. doi: 10.1016/j.foodchem.2008.12.071

Sooriyapathirana, S., Ranaweera, L., Perera, H., Weebadde, C., Finn, C., Bassil, N., et al. (2019). Using SNP/INDEL diversity patterns to identifiy a core group of genotypes from FVC11, a superior hybrid family of *Fragaria virginiana* Miller and *F. chiloensis* (L). Miller. *Genet. Resour. Crop Ev.* 66, 1691–1698. doi: 10.1007/s10722-019-00819-0

Squeo, F. A., Olivares, N., Olivares, S., Pollastri, A., Aguirre, E., Aravena, R., et al. (1999). Grupos funcionales en arbustos desérticos del norte de Chile, definidos sobre la base de las fuentes de agua utilizadas. *Gayana Bot.* 56, 1–15.

Staudt, G., Drawert, F., and Tressl, R. (1975). Gaschromatographisch-massenspektrometrische differenzierung von Erdbeerarten. II. *Fragaria nilgerrensis*. *Z. Pflanzenzuchtg.* 75, 36–42.

Staudt, G. (1962). Taxonomic studies in the genus *Fragaria*. Typification of *Fragaria* species known at the time of Linnaeus. *Can. J. Bot.* 40, 869–886. doi: 10.1139/b62-081

Staudt, G. (1999). *Systematics and geographic distribution of the American strawberry species: Taxonomic studies in the genus Fragaria (Rosaceae: Potentilleae)*, (Berkeley and Los Angeles, CA, USA: University of California Press).

Teixeira, R. D., Ribeiro, H. A. L., Gomes, M. T. R., Lopes, M. T. P., and Salas, C. E. (2008). The proteolytic activities in latex from *Carica candamarcensis*. *Plant Phys. Biochem.* 46, 956–961. doi: 10.1016/j.plaphy.2008.06.010

Tennessen, J. A., Govindarajulu, R., Ashman, T.-L., and Liston, A. (2014). Evolutionary origins and dynamics of octoploid Strawberry subgenomes revealed by dense targeted capture linkage maps. *Genome Biol. Evol.* 6, 3295–3313. doi: 10.1093/gbe/evu261

Tonaco, L. A. B., Gomes, F. L., Velazquez-Menendez, G., Lopes, M. T. P., and Salas, C. E. (2018). The proteolytic fraction from latex of *Vasconcellea cundinamarcensis* (P1G10) enhances wound healing of diabetic foot ulcers: A double-blind randomized pilot study. *Adv. Ther.* 35, 494–502. doi: 10.1007/s12325-018-0684-2

Tulipani, S., Mezzetti, B., Capocasa, F., Bompadre, S., Beekwilder, J., de Vos, C. H. R., et al. (2008). Antioxidants, phenolic compounds, and nutritional quality of different strawberry genotypes. *J. Agric. Food Chem.* 56, 696–704. doi: 10.1021/jf0719959

Ulrich, D., Komes, D., Olbricht, K., and Hoberg, E. (2007). Diversity of aroma patterns in wild and cultivated *Fragaria* accessions. *Genet. Resour. Crop Ev.* 54, 1185–1196. doi: 10.1007/s10722-006-9009-4

Van den Eynden, V., Cueva, E., and Cabrera, O. (1999). *Plantas silvestres comestibles del sur del Ecuador - Wild edible plants of southern Ecuador* (Quito, Ecuador: Ediciones Abya-Yala).

Van den Eynden, V., Cueva, E., and Cabrera, O. (2003). Wild foods from Southern Ecuador. *Econ. Bot.* 57, 576–603. doi: 10.1663/0013-0001(2003)057[0576:WFFSE]2.0.CO;2

Van Droogenbroeck, B., Breyne, P., Goetghebeur, P., Romeijn-Peeters, E., Kyndt, T., and Gheysen, G. (2002). AFLP analysis of genetic relationships among papaya and its wild relatives (Caricaceae) from Ecuador. *Theor. Appl. Genet.* 105, 289–297. doi: 10.1007/s00122-002-0983-4

Vining, K., Salinas, N., Tennessen, J., Zurn, J., Sargent, D., Hancock, J., et al. (2017). Genotyping-by-sequencing enables linkage mapping in three octoploid cultivated strawberry families. *Peer J.* 5, e3731. doi: 10.7717/peerj.3731

Wallace, J., Rodgers-Melnick, E., and Buckler, E. (2018). On the road to breeding 4.0: Unraveling the good, the bad, and the boring of crop quantitative genomics. *Ann. Rev. Genet.* 52, 421–444. doi: 10.1146/annurev-genet-120116-024846

Walraevens, V., Jaziri, M., Van Beeumen, J., Schnek, A. G., Kleinschmidt, T., and Looze, Y. (1993). Isolation and preliminary characterization of the cysteine-proteinases from the latex of *Carica candamarcensis* Hook. *Biol. Chem. H. S.* 374, 501–506. doi: 10.1515/bchm3.1993.374.7-12.501

Wang, Q., and Takahashi, H. (1999). A land surface water deficit model for an arid and semiarid region: Impact of desertification on the water deficit status in the Loess Plateau, China. *J. Climate* 12, 244–257. doi: 10.1175/1520-0442(1999)012<0244:ALSWDM>2.0.CO;2

Zabetakis, I., and Holden, M. A. (1997). Strawberry flavour: Analysis and biosynthesis. *J. Sci. Food Agric.* 74, 421–234. doi: 10.1002/(SICI)1097-0010(199708)74:4<421::AID-JSFA817>3.0.CO;2-6

Conflict of Interest: 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 © 2020 Letelier, Gaete-Eastman, Peñailillo, Moya-León and Herrera. 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) and the copyright owner(s) 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 Plant Science

ORIGINAL RESEARCH
published: 23 July 2020
doi: 10.3389/fpls.2020.01055

Utilizing Wild *Cajanus platycarpus*, a Tertiary Genepool Species for Enriching Variability in the Primary Genepool for Pigeonpea Improvement

Shivali Sharma[1]*, *Pronob J. Paul*[1], *CV Sameer Kumar*[2] and *Chetna Nimje*[3]

[1] Theme Pre-breeding, International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Hyderabad, India,
[2] Regional Agricultural Research Station, Professor Jayashanker Telangana State Agricultural University, Palem, India,
[3] Grain Quality Lab., International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Hyderabad, India

OPEN ACCESS

Edited by:
Eric Von Wettberg,
University of Vermont, United States

Reviewed by:
Paul Kiprotich Kimurto,
Egerton University, Kenya
Sivakumar Sukumaran,
International Maize and Wheat Improvement Center, Mexico
Petr Smýkal,
Palacký University, Olomouc, Czechia

***Correspondence:**
Shivali Sharma
shivali.sharma@cgiar.org

Specialty section:
This article was submitted to Plant Breeding, a section of the journal Frontiers in Plant Science

Received: 09 January 2020
Accepted: 26 June 2020
Published: 23 July 2020

Citation:
Sharma S, Paul PJ, Sameer Kumar CV and Nimje C (2020) Utilizing Wild Cajanus platycarpus, a Tertiary Genepool Species for Enriching Variability in the Primary Genepool for Pigeonpea Improvement. Front. Plant Sci. 11:1055. doi: 10.3389/fpls.2020.01055

The use of crop wild relatives in the breeding program has been well recognized to diversify the genetic base along with introgression of useful traits. *Cajanus platycarpus* (Benth.) Maesen, an annual wild relative belonging to the tertiary genepool of pigeonpea, possesses many useful traits such as early maturity, high protein content, photoperiod insensitivity, and pod borer tolerance for the genetic improvement of cultivated pigeonpea. Using this cross incompatible wild *Cajanus* species, an advanced backcross population was developed following the embryo rescue technique. In the present study, a pre-breeding population consisting of 136 introgression lines (ILs) along with five popular varieties (used as checks) was evaluated for important agronomic traits during 2016 and 2017 rainy seasons and for grain nutrient content during 2016, 2017, and 2018 rainy seasons. Large genetic variation was observed for agronomic traits such as days to 50% flowering, number of pods per plant, pod weight per plant, grain yield per plant, and grain nutrients [protein content, grain iron (Fe), zinc (Zn), calcium (Ca), and magnesium (Mg)] in the pre-breeding population. Significant genotype × environment interaction was also observed for agronomic traits as well as grain nutrients indicating the sensitivity of these traits to the environments. No significant correlations were observed between grain yield and grain nutrients except grain Zn content which was negatively correlated with grain yield. Overall, 28 promising high-yielding ILs with high grain nutrient content were identified. These ILs, in particular, ICPP # 171012, 171004, 171102, 171087, 171006, and 171050 flowered significantly earlier than the popular mega variety, ICPL 87119 (Asha) and thus hold potential in developing new short-duration cultivars. The comprehensive multi-site assessment of these high-yielding, nutrient-rich accessions would be useful in identifying region-specific promising lines for direct release as

cultivars. Moreover, these ILs are expected to replace the popular existing cultivars or for use as new and diverse sources of variations in hybridization programs for pigeonpea improvement.

Keywords: *Cajanus platycarpus*, pre-breeding, introgression lines, wild *Cajanus* species, pigeonpea, grain nutrients, short-duration, photo-insensitivity

INTRODUCTION

Pigeonpea [*Cajajus cajan* (L.) Milllspaugh] is an important often-cross pollinated grain legume crop of semi-arid tropics grown under subsistence agriculture. Globally, it is cultivated on a 7.02 m ha area with an annual production of 6.81 m t (FAOSTAT, 2017) mainly in Asia, Africa, and Latin America. India is the largest producer and consumer of pigeonpea in the world where dry, dehulled split seeds are consumed as "daal", a source of protein-rich (22–24%) food. Besides protein, pigeonpea seeds are also rich in carbohydrates, minerals, crude fibre, iron (Fe), sulphur, calcium (Ca), potassium (K), manganese (Mn) and water-soluble vitamins especially thiamine, riboflavin, and niacin (Saxena et al., 2010). In India, pigeonpea is the second most important legume after chickpea accounting for 5.39 m ha area and 4.87 m t of production (FAOSTAT, 2017). Besides mature seeds, immature green tender pods and seeds are consumed as vegetables mainly in Kenya, Tanzania, and Malawi. The crop is also grown for other uses such as fodder, medicine, rearing lac producing insects, fuelwood, and improving soil fertility through biological nitrogen fixation.

Narrow genetic base of cultivated pigeonpea and repeated use of a few elite breeding lines such as T-1 and T-90 (Kumar et al., 2004) in breeding programs are the major factors hindering its genetic improvement. Further, various biotic and abiotic stresses cause huge yield losses in pigeonpea worldwide and high levels of resistance/tolerance are not available in cultivated genepool. As a result, despite large breeding efforts in India and elsewhere, pigeonpea productivity is stagnant around 0.8-0.9 t ha^{-1}. Major biotic stresses affecting pigeonpea are pod borers (*Helicoverpa armigera* Hubner, *Maruca vitrata* Geyer), and pod fly (*Melanagromyza chalcosoma* Spencer) among insect-pests and fusarium wilt (*Fusarium udum* Butler), sterility mosaic disease (SMD), and phytophthora blight (*Phytophthora drechsleri* Tucker) among diseases. Pigeonpea crop is also sensitive to abiotic stresses such as terminal drought, water-logging, salinity, and frost/cold. The protein advisory group of the United Nations has emphasized on improvement of the nutritional quality of proteins besides improving the productivity, adaptability, and yield stability of grain legumes.

One of the key factors for a successful crop improvement program is the availability of sufficient genetic variability. Over 13,200 accessions of cultivated pigeonpea and 555 accessions of wild species belonging to genus *Cajanus* from 60 countries are conserved in ICRISAT genebank. These germplasm accessions, based on the crossability relationship with cultivated pigeonpea, are grouped into three genepools with cultivated germplasm in the primary genepool (GP 1), all cross-compatible species, *C. acutifolius* (F.Muell.) Maesen, *C. albicans* (Wight & Arn.) Maesen, *C. cajanifolius* (Haines) Maesen, *C. cinereus* (F.Muell.), *C. confertiflorus* F. Muell., *C. lanceolatus* (W.Fitzg.) Maesen, *C. latisepalus* Maesen, *C. lineatus* (Wight & Arn.) Maesen, *C. reticulatus* (Dryand.) F.Muell., *C. scarabaeoides* (L.) Thouars, *C. sericeus* (Baker) Maesen, *C. trinervius* (DC.) Maesen in the secondary genepool (GP 2), and the cross-incompatible species, *C. crassus* (King) Maesen, *C. goensis* Dalzell, *C. mollis* (Benth.) Maesen, *C. platycarpus* (Benth.) Maesen, *C. rugosus* (Wight & Arn.) Maesen, *C. heynei, C. kerstingii, C. volubilis, and other Cajaninae such as Rhynchosia* Lour., *Dunbaria W.* and *A., Eriosema (DC.)* Reichenb in the tertiary genepool (GP 3). Wild *Cajanus* species are the reservoirs of many useful genes/alleles and can be used to enrich variability in the primary genepool for developing new broad-based cultivars with increased plasticity (Sharma et al., 1993; Rao et al., 2003; Sharma and Upadhyaya, 2016; Sharma et al., 2019). Introgression of useful genes/alleles from the wild *Cajanus* species would help to break the yield plateau in pigeonpea. In chickpea, interspecific derivatives having high yield and resistance for wilt, foot rot, and root rot diseases (Singh et al., 2005) as well as for cyst nematode (Malhotra et al., 2002) were developed from crosses involving *C. reticulatum*. Similarly high-yielding, cold-tolerant lines with high biomass (ICARDA, 1995) and resistance to phytophthora root rot were developed from interspecific crosses involving *C. echinospermum* (Knights et al., 2008). Frequent utilization of wild *Cajanus* species in breeding programs is hindered due to cross incompatibility barriers and linkage drag. Among wild species, *C. platycarpus* is of particular interest to the pigeonpea breeders due to several useful traits such as extra-early flowering and maturity (Saxena et al., 1996), photoperiod insensitivity, prolific flowering and podding, high harvest index, annuality and rapid seedling growth, and resistance/tolerance to biotic and abiotic stresses such as pod borer (Sujana et al., 2008), Fusarium wilt (Saxena et al., 1990), phytophthora blight (Ariyanayagam and Spence, 1978; Pundir and Singh, 1987; Dundas, 1990), nematodes (Sharma, 1995), sterility mosaic (Lava Kumar et al., 2005) and salinity (Subbarao, 1988). Using this cross incompatible wild *Cajanus* species, *C. platycarpus*, a backcross population was developed following embryo rescue technique (Mallikarjuna et al., 2011).

Linkage drag is the most common problem associated with the utilization of wild species in breeding programs. Hence, utilization of wild species in creating new genetic variability will be successful only when introgression lines (ILs) with useful traits and acceptable agronomic performance are developed and made available to breeders for direct use in breeding programs. Therefore, the present investigation was carried out a) to study

the genetic variability in the advanced backcross population derived from *C. platycarpus* for important agronomic traits and grain nutrient, and b) to identify stable promising trait-specific ILs with minimum linkage drag for ready use in breeding programs to develop new cultivars with a broad genetic base.

MATERIALS AND METHODS

Plant Material and Field Evaluation

Using a cross incompatible wild *Cajanus* species, *C. platycarpus* accession ICPW 68, and a popular pigeonpea cultivar ICPL 85010, a backcross population was developed following embryo rescue technique (Mallikarjuna et al., 2006). ICPW 68, originated from Uttar Pradesh, India, is extra-early flowering accession having high seed protein content (Saxena et al., 1996) and pod borer resistance (Sujana et al., 2008). ICPL 85010, also known as "Sarita", is a short duration, determinate type pigeonpea variety having medium seed size (9.5 g 100-seed weight), which is cultivated in the Indian Subcontinent (Dahiya et al., 2001). Embryo rescue and tissue culture techniques were followed as described by Mallikarjuna and Moss (1995). The details of the population development have been documented by Mallikarjuna et al. (2006). The advanced backcross population consisting of 136 ILs in BC_4F_{10} generation was used in this study. The 136 ILs along with five popular varieties (used as checks) of different maturity durations [ICPL 87119 (also known as "Asha") ICP 8863 (Maruti), ICPL 20325, ICPL 85010, and ICPL 88039] were evaluated for different agronomic traits during the 2016 and 2017 rainy seasons and grain nutrients during the 2016, 2017, and 2018 rainy seasons at ICRISAT, Patancheru, Telangana, India (17°51′N, 78°27′E; 545 m). Among the checks, ICPL 87119 is a popular mega variety in the medium maturity duration group that is being widely cultivated in India over the past two decades; ICP 8863 is a medium-duration, high-yielding pigeonpea variety resistant to fusarium wilt which is popular in Karnataka, India; ICPL 20235 and ICPL 88039 are super-early and early maturing pigeonpea varieties, respectively. Accessions were planted in black soil (Vertisols) precision field in the first week of July in all years in an augmented design. Each check was placed after every 10 entries in each block and total lines were divided into three blocks. Each accession was sown in a single 4 m long row in a ridge and furrow system with a plant-to-plant spacing of 20 cm and row to row spacing of 75 cm. A standard package of practices was followed to raise a healthy crop. Manual weeding and spraying of insecticide were done to control weeds and insect-pests damage. The weather data of 2016 and 2017 crop season at ICRISAT, Patancheru, India is given in **Supplementary Figure 1**. Data were recorded on eight agronomic traits [days to first flowering, days to 50% flowering, plant height (cm), number of primary branches, number of secondary branches, number of pods per plant, pod weight per plant (g), and grain yield per plant (g)], and five grain nutrients [protein content (%), iron (Fe in mg kg^{-1}), zinc (Zn in mg kg^{-1}), calcium (Ca in g kg^{-1}), and magnesium (Mg in g Kg^{-1}) content]. Data on grain nutrients and two agronomic traits namely days to first flowering, and days to 50% flowering were recorded on a plot basis, whereas data on remaining agronomic traits (plant height, primary branches per plant, secondary branches per plant, pods per plant, pod weight per plant and grain yield per plant) were recorded on five randomly selected representative plants per plot following pigeonpea descriptors (IBPGR and ICRISAT, 1993). All the lines were harvested and threshed manually.

Estimation of Grain Nutrient Content

For estimating the nutrient contents of Fe, Zn, Mg, Ca, and protein in grains, seeds of 136 ILs along with five checks were cleaned thoroughly and special care was taken during cleaning to prevent contamination of seeds with dust and metal particles. Seeds were washed with distilled water for a few seconds and dried in hot air at 40°C for 2 h to remove the dust and metal particles. Well-cleaned random seed samples were used for estimating grain protein, Fe, Zn, Ca, and Mg contents at the Charles Renard Analytical Laboratory, ICRISAT, Patancheru, India. The four dietary minerals- Fe, Zn, Ca, and Mg contents were assessed by nitric acid and hydrogen peroxide digestion accompanied by inductively coupled plasma optical emission spectrometry (ICP-OES) (Wheal et al., 2011). The sulfuric acid-selenium digestion method was adopted for the estimation of grain protein followed by the estimation of total nitrogen (N) in a SKALAR SAN^{++} SYSTEM autoanalyzer and the measurement of protein percentage as N percent × 6.25 conversion factor (Sahrawat et al., 2002).

Statistical Analysis

Eight agronomic traits and five grain nutrients were analyzed separately for each rainy season and pooled over the two seasons for agronomic traits, and three seasons for grain nutrients using residual maximum likelihood (REML) in GenStat 15 (https://www.vsni.co.uk/) in mixed model approach considering genotypes as random effect and environment as fixed effect. The significance of environments was tested using Wald's statistic. Variance components due to genotype (σ_g^2) and genotype × environment ($\sigma_{g\times e}^2$) interaction and their standard errors (SE) were estimated. Best linear unbiased predictors (BLUPs) were obtained for agronomic traits and grain nutrients for each accession for individual environment as well as pooled over the environments. Based on BLUPs, the range, mean, variances and broad-sense heritability (H^2) were estimated. Phenotypic correlations were estimated to determine trait associations in GenStat 15. Path analysis was performed to estimate the direct effect of the traits towards grain yield using R Version 3.5.3 (R Project for Statistical Computing, http://www.r-project.org/). To avoid the multicollinearity issues, two independent traits, days to 50% flowering and pod weight per plant were excluded while performing path analysis. Using the R package *cluster* (Patterson and Thompson, 1971), the Euclidean dissimilarity matrix was constructed using agronomic traits and grain nutrients and the accessions were clustered following Ward's method. Further, accessions with high grain yield and high grain Fe, Zn, Ca, Mg, and protein content were identified. Using the Euclidian distance matrix, the most diverse accession

pairs were identified for potential use as parents in pigeonpea crossing programs.

RESULTS

Variance Components and Trait Variability

The REML analysis showed significant variations among ILs (σ_g^2) for all the eight agronomic traits in both 2016 and 2017 rainy seasons and all the five grain nutrients in 2016, 2017 and 2018 rainy seasons indicating the presence of significant variability among the ILs for these traits. Pooled analysis also showed significant genetic variance (σ_g^2) and significant G×E interactions for all agronomic traits and grain nutrients (**Table 1**).

A large variation was observed amongst the ILs for all the agronomic traits and grain nutrients in each season as well as in the pooled analysis (**Table 2**). It is evident that some of the ILs performed better than the popular variety ICPL 87119 and the recurrent parent ICPL 85010 for these traits. A flowering window of around 60 days (76–141 days range) was observed for days to 50% flowering showing substantial population variability while popular variety ICPL 87119 took ~123 days to 50% flowering. None of the ILs flowered earlier than the recurrent parent ICPL 85010 (67 days to 50% flowering). Large variation was also noted amongst the ILs for plant height (above 100 cm: ~128–272 cm) when compared with the popular variety ICPL 87119 (on an average 222 cm tall) (**Table 2**). Similarly, number of pods per plant were much higher in the ILs (up to 386 pods in 2016 and 303 pods per plant in 2017) compared to ICPL 87119 (average 171 pods per plant) and the cultivated parent ICPL 85010 (average 126 pods per plant) in both seasons. Similar pattern was observed in 2016, 2017, and in pooled analysis for other traits such as grain yield per plant (up to 83 g in ILs compared to 48 g in ICPL 87119 and 25 g in ICPL 85010), protein content (up to 23% in ILs compared to 19% in ICPL 87119 and 21% in ICPL 85010), Fe (up to 43 mg kg^{-1} in ILs compared to 32 mg kg^{-1} in ICPL 87119 and 33 mg kg^{-1} in ICPL 85010), Zn (up to 42 mg kg^{-1} in ILs compared to 29 mg kg^{-1} in ICPL 87119 and 41 mg kg^{-1} in ICPL 85010), Ca (up to 3.1 g kg^{-1} in ILs compared to ~1.0 g kg^{-1} both in ICPL 87119 and ICPL 85010), and Mg (up to 1.6 g kg^{-1} in ILs compared to ~1.3 g kg^{-1} in ICPL 87119, and 1.1 g kg^{-1} in ICPL 85010) (**Table 2**).

Associations Between Agronomic Traits and Grain Nutrients

The correlation analysis in 2016 (**Table S1**), 2017 (**Table S1**), and pooled analysis over years (**Figure 1**) showed that grain yield per plant was significantly and positively associated with days to first flowering (r = 0.38 in 2016, 0.42 in 2017 and 0.49 in pooled), days to 50% flowering (r = 0.40 in 2016, 0.39 in 2017 and 0.48 in pooled), plant height (r = 0.24 in 2016, 0.42 in 2017 and 0.37 in pooled), number of secondary branches (r = 0.50 in 2016, 0.24 in 2017 and 0.42 in pooled), pods per plant (r = 0.92 in 2016, 0.84 in 2017 and 0.90 in pooled), and pod weight per plant (r = 0.99 in 2016, 0.95 in 2017 and 0.98 in pooled) (**Table S1**, **Figure 1**). Number of primary branches had significantly positive correlation with grain yield per plant in 2017 (r = 0.45) and pooled over years (r = 0.25) but no correlation was observed in 2016 rainy season.

Similarly, among the grain nutrients, protein content in seeds showed significantly positive association with three nutrients, Fe (r = 0.23 in 2016, 0.32 in 2017, 0.34 in pooled), Zn (r = 0.44 in 2016, 0.32 in 2017, 0.40 in pooled), and Mg (r = 0.19 in 2016, 0.21 in 2017, 0.19 in pooled). Significantly positive correlation was observed between Fe and Zn content (r = 0.37 in 2016, 0.27 in 2017 and 0.37 in pooled) and between Ca and Mg content (r = 0.64 in 2016, 0.60 in 2017 and 0.67 in pooled) in year-wise and pooled analysis.

In the present study, no significant correlation was found between grain yield per plant with grain nutrients (Grain protein, Ca, Fe) in 2016 (**Table S1**), 2017 (**Table S1**), and pooled over years (**Figure 1**). Grain yield per plant showed a significantly negative correlation (r = -0.28 in 2016, -0.42 in 2017, and -0.43 in pooled) with Zn content. Further, the path analysis in 2016 (**Table S2**), 2017 (**Table S2**), and pooled analysis over the years

TABLE 1 | Variance components due to genotypes (σ_g^2), genotype × environment ($\sigma_{g\times e}^2$) interactions and their standard errors (SE) for agronomic traits and grain nutrients of *C. platycarpus* derived introgression lines evaluated during 2016, 2017, and 2018 rainy seasons at ICRISAT, Patancheru.

Traits	2016 rainy season		2017 rainy season		2018 rainy season		Pooled over seasons			
	σ_g^2	SE	σ_g^2	SE	σ_g^2	SE	σ_g^2	SE	$\sigma_{g\times e}^2$	SE
Days to first flowering	146.61**	19.460	228.28**	28.030	–	–	143.51**	20.388	41.51**	6.634
Days to 50% flowering	166.66**	21.150	210.85**	26.698	–	–	146.23**	20.620	38.42**	6.680
Plant height (cm)	353.05**	44.970	406.56**	51.870	–	–	340.57**	44.170	36.98**	8.860
Primary branches (no.)	21.67**	4.536	15.81**	3.072	–	–	8.37**	2.138	9.42**	2.649
Secondary branches (no.)	21.65**	3.290	8.728*	3.443	–	–	6.97**	1.779	7.14**	2.354
Pods per plant (no.)	5777.00**	1054.000	1909.30**	353.200	–	–	2499.10**	448.900	1282.70**	380.600
Pod weight per plant (g)	1005.20**	147.200	259.43**	47.100	–	–	322.70**	66.240	306.23**	53.780
Grain yield per plant (g)	434.12**	70.420	115.31**	20.540	–	–	150.59**	29.960	123.10**	25.090
Protein (%)	1.52**	0.281	1.53**	0.244	1.25**	0.361	0.43**	0.115	1.05**	0.150
Fe (mg kg^{-1})	19.11**	2.813	11.39**	2.669	20.77**	2.730	8.91**	1.536	8.73**	1.099
Zn (mg kg^{-1})	11.04**	1.841	11.53**	3.170	10.77**	2.051	7.84**	1.218	2.77**	1.080
Ca (g kg^{-1})	0.16**	0.024	0.18**	0.026	0.21**	0.027	0.13**	0.018	0.05**	0.008
Mg (g kg^{-1})	0.01**	0.005	0.02**	0.003	0.02**	0.004	0.01**	0.002	0.004*	0.002

*Significant at $P \leq 0.05$; **Significant at $P \leq 0.01$.

TABLE 2 | Summary statistics for various traits in pigeonpea pre-breeding population evaluated during 2016, 2017, and 2018 rainy seasons at ICRISAT, Patancheru.

Seasons		Days to first flowering	Days to 50% flowering	Plant height (cm)	Primary branches (no.)	Secondary branches (no.)	Pods per plant (no.)	Pod weight per plant (g)	Grain yield per plant (g)	Protein (%)	Fe (mg kg^{-1})	Zn (mg kg^{-1})	Ca (g kg^{-1})	Mg (g kg^{-1})
2016 rainy season	Mean	96 ± 1.05	104 ± 1.11	202.0 ± 1.62	13 ± 0.45	7 ± 0.42	206± 7.08	74.0 ± 2.82	49.8 ± 1.90	21.4 ± 0.12	32.8 ± 0.39	35.2 ± 0.30	1.2 ± 0.04	1.2± 0.01
	Range	72- 131	76 -140	127.9- 247.0	7- 23	0 - 18	70- 386	17.9 – 172.1	14.0 - 109.3	18.6 - 23.8	23.8 - 43.9	26.0- 42.0	0.7 - 2.8	1.0- 1.6
	Heritability (%)	95.1	97.0	96.7	81.4	90.6	85.2	91.7	88.6	84.9	91.7	87.8	91.4	67.5
	ICPL 85010	64	68	96.0	11	2	131	39.4	27.3	19.5	31.1	37.9	0.9	1.0
	ICPL 87119	104	112	204.2	17	16	185	66.1	52.4	18.3	30.7	27.4	0.8	1.0
2017 rainy season	Mean	111± 1.29	119 ± 1.25	240.1 ± 1.74	21 ± 0.38	5 ± 0.33	143 ± 4.08	49.1 ± 1.50	33.9 ± 1.00	21.1 ± 0.11	39.0 ± 0.33	33.5 ± 0.35	1.6± 0.04	1.4 ± 0.01
	Range	72 - 135	84-141	155.5 - 272.3	12 - 30	2 - 12	60 - 303	16.8- 99.3	10.7- 66.2	18.2 - 23.8	31.8 - 45.4	27.6 - 40.3	0.9 - 3.5	1.1 - 1.7
	Heritability (%)	98.4	97.0	96.6	83.5	63	84.8	85.4	85.9	89.0	78.2	73.7	91.5	88.0
	ICPL 85010	59	66	126.1	15	4	123	37.8	22.9	22.9	35.3	42.5	1.3	1.2
	ICPL 87119	125	133	239.7	28	13	158	56.3	43.5	19.2	32.9	31.1	1.3	1.4
2018 rainy season	Mean	–	–	–	–	–	–	–	–	20.2 ± 0.12	35.2 ± 0.4	33.9 ± 0.31	1.6 ± 0.04	1.4 ± 0.01
	Range	–	–	–	–	–	–	–	–	18.2 - 22.6	26.2 - 46.4	26.3 - 42.4	0.9 - 3.3	1.1 - 1.7
	Heritability (%)	–	–	–	–	–	–	–	–	66.2	95.4	84.0	95	81.4
	ICPL 85010	–	–	–	–	–	–	–	–	20.1	34.1	40.1	1.1	1.2
	ICPL 87119	–	–	–	–	–	–	–	–	19.3	32.3	30.0	1.5	1.4
Pooled over seasons	Mean	103 ± 1.32	111 ± 1.25	221.1 ± 1.69	17 ± 0.5	6 ± 0.54	175 ± 5.10	61.6± 1.86	41.9 ± 1.28	20.9 ± 0.78	35.7 ± 0.04	34.2 ± 0.51	1.5 ± 0.04	1.3± 0.01
	Range	72 - 135	76 - 141	138.8 - 255.9	11 - 26	1 - 14	64- 321	18.8 - 123.3	13.1 - 82.9	19.3 - 23.0	26.4 - 42.5	27.6 - 41.8	0.9 - 3.1	1.1 - 1.6
	Heritability (%)	85.7	86.4	93	54.4	54.7	71.6	63.4	65.3	49.3	72.0	80.5	85.2	78.0
	ICPL 85010	61	67	110.3	13	3	126	38.2	24.7	20.8	33.4	40.6	1.1	1.1
	ICPL 87119	115	123	222.0	23	15	171	61.4	48.2	18.9	31.8	29.3	1.2	1.3

revealed that pods per plant is the major direct contributor to grain yield (0.872 in 2016, 0.765 in 2017, and 0.813 in pooled). The other two traits which showed a major contribution towards yield were number of secondary branches per plant and plant height (**Table S2**).

Identification of Promising Introgression Lines (ILs)

To understand the potential of these ILs derived from wild species in improving cultivated pigeonpea, the average performance of these lines over two years was compared with the popular variety, ICPL 87119 and the recurrent parent, ICPL 85010. None of the ILs flowered earlier than ICPL 85010 (days to 50% flowering <70 days). However, 104 ILs flowered significantly earlier (days to 50% flowering: 81–119 days) than ICPL 87119 (123 days). The majority of the ILs (> 100 lines out of 136 lines) were found significantly better than recurrent parent ICPL 85010 for most of the agronomic traits such as pods per plant, pod weight per plant and grain yield per plant (**Table 3**). In comparison with ICPL 87119 (~48 g), 28 ILs (~20% of the backcross population) had significantly higher grain yield per plant (~52–83 g) (**Table 3**). Besides, many ILs were found significantly better than ICPL 87119 in terms of plant height (50 ILs with 139–218 cm height), number of pods per plant (53 ILs with 189–321 pods) and pod weight per plant (55 ILs with 67–123 g). A few ILs also had a higher number of primary branches compared to ICPL 87119 (**Table 3**).

For grain nutrients, all the 136 ILs were found to have higher protein content (~19–23%) than the popular variety ICPL 87119 whereas, 104 ILs (~34–42 mg kg$^{-1)}$ in grain Fe content, 102 ILs (~32–42 mg kg-^{1}) in grain Zn content, 74 ILs (1.35–3.09 g kg^{-1}) in grain Ca content, and 45 ILs (1.35–1.59 g kg^{-1}) in grain Mg content were found significantly found better than ICPL 87119 (**Table 3**). Above 50 ILs were found promising than ICPL 85010 for most of the grain nutrient contents except Zn content (**Table 3**). Top five trait-specific ILs for each agronomic trait and grain nutrients are given in **Table 3**.

A total of 28 promising high-yielding ILs, which performed better than the popular variety, ICPL 87119, were identified (**Table 4**). Most of these high-yielding ILs exhibited higher amounts of five grain-nutrient contents. Remarkably, one line ICPP 171012 was found early (50% flowering: 109 days) having a high number of pods per plant (301), pod weight per plant (108 g), grain yield per plant (73 g), along with better grain nutrient contents such as high protein (21%), grain Fe (35mg kg^{-1}), grain Zn (36.16 mg kg^{-1}), grain Ca (1.39 g kg^{-1}), and grain Mg (1.30 g kg^{-1}) content compared to the best check ICPL 87119. Similarly, ILs ICPP # 171004, 171102, 171087, 171006, and 171050 were found most promising in terms of early flowering (109–116 days), high yielding (~62–65 g) and other agronomic traits along with higher grain nutrient contents than ICPL 87119 (**Table 4**).

Cluster Analysis

Cluster analysis is performed to categorize lines into distinct groups/clusters wherein genotypes in different clusters are more diverse than within a cluster (Ward, 1963) and is useful in selecting the most diverse genotypes to be used as parents in crossing programs.

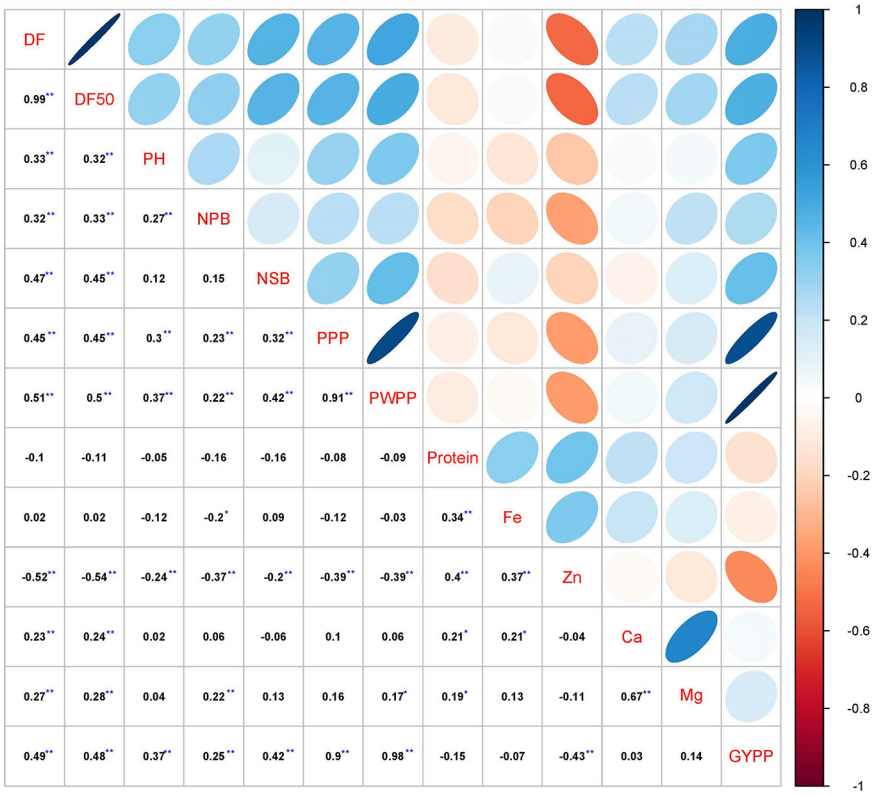


FIGURE 1 | Correlation analysis of various agronomic and grain nutrient traits in pigeonpea pre-breeding population derived from *C. platycarpus* at ICRISAT, Patancheru. Days to flowering (DF & DF50), plant height (PH), primary branches (NPB), secondary branches (NSB), pods per plant (PPP), and pod weight per plant (PWPP) were positively associated with grain yield per plant (GYPP). Zn concentration was correlated negatively with grain yield per plant. On the other hand, no significant relationship was found for grain yield per plant with grain protein content, grain Fe, Ca and Mg content. It was observed that grain Fe and Zn as well as grain Ca and Mg were positively correlated. **Significant at $P \leq 0.01$; *Significant at $P \leq 0.05$.

In this study too, a hierarchical cluster analysis based on all eight agronomic traits and five grain nutrients over two seasons was performed to group the introgression lines into different clusters. The cluster analysis following Ward's method resulted in 10 clusters (**Table S3**; **Figure 2**). Cluster 6 was the largest cluster consisting of 27 ILs followed by cluster 3 (21 ILs) and cluster 7 (20 ILs). Cluster 2 had only two genotypes, ICPL 85010 and ICPL 20325; both are early maturing cultivars (50% flowering: 68 days) (**Figure 2**). All early maturing ILs were grouped into cluster 4 (50% flowering: 89 days). Cluster 3 had the highest cluster mean for grain yield per plant (55.12 g) followed by cluster 6 (53.33 g) whereas cluster 6 had the highest number of pods per plant (225) followed by cluster 3 (213). Cluster 10 exhibited the lowest means for grain yield per plant (~21 g) and pod weight per plant (~32 g) (**Table S3**). The popular mega variety, ICPL 87119 was grouped into cluster 3 along with the highest yielding introgression line ICPP I71119. Cluster 10 was found to be the best cluster in terms of high Fe content (40 mg Kg^{-1}) but had the lowest grain yield per plant. Also, cluster 4 had highest mean for Zn content (38 mg Kg^{-1}, "Cluster 2" has not been considered as it does not hold any ILs) and cluster 8 was found to have ILs with the maximum mean value for grain Ca (2.15 g Kg^{-1}) and Mg (1.48 g Kg^{-1}) content (**Table S3**).

The Euclidian distance matrix was estimated to identify the most diverse pair of ILs among the advanced backcross population as well as to identify the most similar and diverse ILs to the popular variety ICPL 87119 and the recurrent parent ICPL 85010 (**Table S4**). ICPP 171027 was found to be the most diverse (9.09) when compared with ICPL 87119, while ICPP 171031 (2.60) was the most similar IL with ICPL 87119. Likewise, ICPP 171119 was the most diverse (12.42) and ICPP 171033 was the most similar (5.68) accession to the recurrent parent ICPL 85010. Among the 136 ILs, the most diverse pair of accessions were ICPP 171119 and ICPP 171078 with a distance of 10.76. Top 10 most diverse pairs of accession are given in **Table S4**; ICPP 171119 was found to be the most diverse IL amongst these top 10 diverse pairs of ILs.

DISCUSSION

Grain legumes are an excellent and unique source of dietary protein for human beings in many parts of the world. The dietary importance of legumes is expected to increase over the years due to an increased demand for protein and other nutrients by the

TABLE 3 | Identification of promising trait-specific introgression lines derived from pigeonpea tertiary genepool species *C. platycarpus*.

Traits	Number of ILs significantly better than ICPL 87119 (range)	ICPL 87119	Number of ILs significantly better than ICPL 85010 (range)	ICPL 85010	Top five performing ILs against mega cultivar ICPL 87119 (range)
Days to first flowering	101 (74-110)	115	0	61	ICPP #171032, 171030, 171082, 171033, 171027 (74-77)
Days to 50% flowering	104 (81-118)	123	0	67	ICPP # 171082, 171032,171027, 171030, 171024 (81-85)
Plant height (cm)	50 (138.80-217.36)	221.99	0	110.32	ICPP # 171078, 171073, 171043, 171075, 171128 (138.80-193.62)
Primary branches (no.)	1 (26)	23	93 (16-26)	13	ICPP # 171022, 171037,171104, 171031, 171098 (24-26)
Secondary branches (no.)	0	15	63 (6-14)	3	ICPP # 171029, 171094, 171069, 171021, 171129 (12-14)
Pods per plant (no.)	53 (189-321)	171	101 (143-321)	126	ICPP # 171006, 171012, 171020, 171044, 171050 (288-321)
Pod weight per plant (g)	55 (66.57-123.30)	61.38	107 (43.64-123.30)	38.21	ICPP # 171119, 171012, 171006, 171088, 171020 (103.66-123.30)
Grain yield per plant (g)	28 (51.99-82.87)	48.18	111 (29.44-82.87)	24.72	ICPP # 171119, 171012, 171020, 171088, 171029 (68.19-82.87)
Protein (%)	136 (19.33-23.01)	18.88	54 (21.10-23.01)	20.79	ICPP # 171053, 171014, 171039, 171132, 171115 (22.51-23.01)
Fe (mg kg^{-1})	104 (33.60-42.50)	31.75	80 (35.23-42.50)	33.36	ICPP # 171078, 171077, 171009,171090, 171076 (41.07-42.50)
Zn (mg kg^{-1})	102 (32.12-41.76)	29.33	0	40.56	ICPP # 171027, 171082, 171032, 171132, 171023 (39.42-41.76)
Ca (g kg^{-1})	74 (1.35-3.09)	1.20	88 (1.26-3.09)	1.11	ICPP # 171010, 171045, 171127, 171104,171077 (2.28-3.09)
Mg (g kg^{-1})	45 (1.35-1.59)	1.27	115 (1.22-1.59)	1.14	ICPP # 171077, 171014, 171022, 171051, 171127 (1.53-1.59)

TABLE 4 | Performance of 28 promising high-yielding introgression lines for important agronomic traits and grain nutrients.

Geno	DF50[#]	PPP	PWPP	GYPP	Protein	Fe	Zn	Ca	Mg
ICPP 171119	131	285*	123.30*	82.87*	20.78*	29.82	30.35[†]	0.93	1.19
ICPP 171012	109*	301*	108.45*	73.36*	20.86*	35.36*	36.16*	1.39*	1.30[†]
ICPP 171020	119[†]	294*	103.66*	70.31*	19.60*	33.88*	31.17[†]	1.20[†]	1.30[†]
ICPP 171088	127	260*	103.98*	69.96*	19.80*	33.82*	27.61	1.20[†]	1.23
ICPP 171029	121[†]	207*	97.43*	68.19*	21.32*	38.41*	32.97*	1.29[†]	1.35*
ICPP 171044	130	292*	101.16*	66.02*	21.81*	36.05*	33.44*	1.98*	1.39*
ICPP 171004	111*	248*	94.39*	64.58*	20.09*	33.89*	31.53[†]	1.46*	1.29[†]
ICPP 171102	116*	271*	90.16*	63.19*	20.91*	33.97*	31.40[†]	1.25	1.42*
ICPP 171129	117*	194*	88.80*	63.11*	21.11*	34.57*	31.52[†]	1.08	1.23
ICPP 171087	110*	196*	86.07*	62.50*	20.27*	26.35	29.73[†]	2.06*	1.41*
ICPP 171006	109*	321*	107.96*	62.20*	21.31*	38.90*	36.09*	1.60*	1.41*
ICPP 171050	116*	288*	96.41*	61.91*	20.86*	36.28*	31.51[†]	1.90*	1.31[†]
ICPP 171098	124	206*	89.29*	61.57*	19.72*	38.45*	29.49[†]	1.27[†]	1.36*
ICPP 171094	115*	223*	89.36*	60.21*	21.35*	39.50*	36.60*	1.49*	1.33[†]
ICPP 171021	112*	202*	87.82*	59.12*	20.89*	32.01[†]	32.47*	1.43*	1.31[†]
ICPP 171025	118*	218*	84.29*	58.88*	21.63*	37.59*	36.71*	2.05*	1.45*
ICPP 171001	110*	215*	81.37*	58.82*	20.05*	36.93*	36.01*	1.01	1.37*
ICPP 171085	112*	199*	89.20*	58.70*	22.16*	38.21*	36.06*	1.52*	1.47*
ICPP 171048	114*	273*	89.13*	57.53*	21.10*	37.40*	34.41*	1.77*	1.23
ICPP 171045	129	278*	89.86*	57.31*	20.60*	34.81*	31.38[†]	2.61*	1.47*
ICPP 171131	109*	231*	75.77*	56.51*	20.95*	29.98	32.03[†]	1.05	1.07
ICPP 171059	112*	202*	84.29*	55.83*	20.79*	31.06	34.03*	1.02	1.32[†]
ICPP 171097	110*	195*	81.31*	55.77*	20.08*	37.13*	35.46*	1.27[†]	1.34[†]
ICPP 171046	120[†]	247*	77.58*	53.90*	21.19*	39.00*	32.92*	1.60*	1.19
ICPP 171127	117*	216*	77.53*	53.13*	21.51*	35.26*	30.09[†]	2.39*	1.53*
ICPP 171113	128	181[†]	81.11*	53.03*	20.21*	40.05*	33.77*	1.54*	1.31[†]
ICPP 171003	114*	213*	75.03*	52.54*	19.50*	35.92*	35.95*	1.23[†]	1.31[†]
ICPP 171011	117*	204*	76.95*	51.99*	20.77*	34.25*	33.71*	1.63*	1.42*
ICPL 87119	123	171	61.38	48.18	18.88	31.75	29.33	1.20	1.27

*[#]DF50, Day to 50% flowering; PPP, Number of pods per plant; PWPP, pod weight per plant; GYPP, Grain yield per plant. *Significantly better than popular variety ICPL 87119 (Asha) at P≤ 0.05; [†]Better than ICPL 87119 on per se basis.*

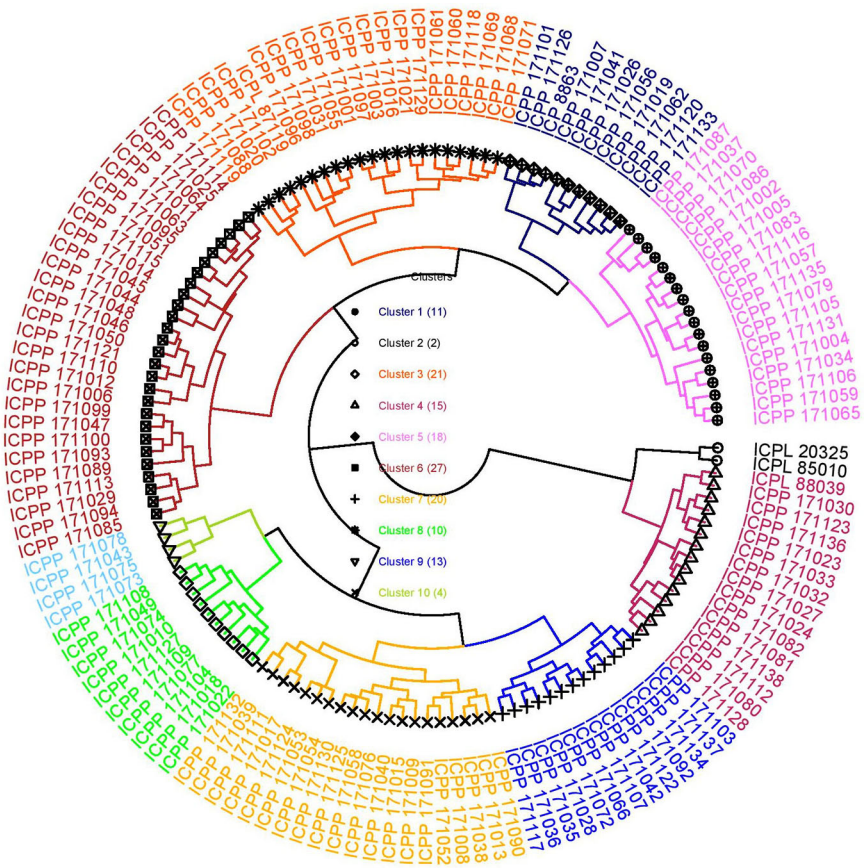


FIGURE 2 | Cluster diagram depicting different clusters formed using 136 introgression lines derived from *C. platycarpus* and five popular varieties following Ward's method based on agronomic and grain nutrient traits. The "Cluster 2" consisted of only two lines: the recurrent parent ICPL 85010 along with another popular variety ICPL 20325. The highest yielding introgression line ICPP 171119 grouped in "Cluster 3" along with mega variety ICPL 87119.

growing world population (Duranti, 2006). Moreover, the development of climate-resilient, nutrient-rich crop varieties is expected to reduce the number of malnourished people worldwide, especially in South Asia and Sub-Saharan Africa (Varshney et al., 2019). The crop varieties with improved nutrition may also be useful in addressing the UN Millennium goal of zero hunger and malnutrition, particularly in those parts of the world where plant-based protein is in high demand.

Pigeonpea is an excellent protein-rich legume crop that is mainly grown under rainfed conditions on marginal lands with minimal inputs and plays an important role in subsistence agriculture (Bohra et al., 2017; Obala et al., 2018). The modern pigeonpea cultivars have a narrow genetic base due to the frequent use of a few promising lines in the hybridization programs over the years. (Kumar et al., 2004; Bohra et al., 2010; Sharma et al., 2019). To meet the growing demand for plant-based nutrition and the need for soil health rejuvenation by implementing a proper cropping system, the pigeonpea breeding programs must embrace a few novel approaches (Anitha et al., 2019). The high-yielding nutrient-rich varieties will attract farmers not only in developing countries where the crop is being cultivated traditionally but will also find a niche in new environments in the developed countries (Atlin et al., 2017). Crop wild relatives (CWRs) are an excellent source of new alleles for different useful traits required for pigeonpea improvement (Khoury et al., 2015; Sharma and Upadhyaya, 2016; Sharma et al., 2019). But due to many hindrances such as cross incompatibility, late maturity, undesirable pod traits, poor agronomic performance, high photoperiod sensitivity, etc., breeders are disinclined to use these CWRs in crop improvement programs (Sharma et al., 2013). In this context, pre-breeding plays a vital role to enhance the use of wild relatives in breeding programs by providing the ready-to-use ILs with superior alleles for different traits introgressed from wild species.

The pre-breeding population consisting of 136 ILs used in the present study was derived from a cross-incompatible tertiary genepool species, *C. platycarpus* following embryo rescue technique (Mallikarjuna et al., 2011) with a view to introgress important traits such as early maturity, high protein content, and photoperiod insensitivity into cultivated pigeonpea. A large genetic variation was observed for important agronomic traits such as days to 50% flowering, number of pods per plant, grain yield as well as for the grain nutrients viz., protein, grain Fe, Zn, Ca, and Mg content. The genetic variation found in this

backcross population is expected to be noble with broad genetic bases as it is derived from wild species.

Based on the days to maturity (duration from planting to 75% maturity), pigeonpea cultivars/varieties are grouped into different groups such as super-early (50–60 days to flowering and/or <100 days to maturity), extra-early (60–80 days to flowering and/or 101–120 days to maturity), early (81–100 days to flowering and/or 121–140 days to maturity), mid-early (101–120 days to flowering and/or 141–160 days to maturity), medium (111–130 days to flowering and/or 161–180 days to maturity) and late (> 130days to flowering and/or >180 days to maturity) maturity duration groups (Srivastava et al., 2012). In general, major pigeonpea cultivation is dominated by varieties in the medium-maturity group (Choudhary and Nadarajan, 2011). Under changing climatic conditions, there is an emphasis on developing short-duration pigeonpea cultivars having photo- and thermo-insensitivity to fit into multiple cropping systems as well as to expand pigeonpea cultivation into new niche areas (Saxena et al., 2019).

Though short duration lines are available, there is a huge yield penalty compared to popular medium-maturing varieties such as ICPL 87119. Most of the high-yielding ILs identified in the present study flowered early and had significantly higher yield and better grain nutrient contents than all the control cultivars used in this study including popular variety, ICPL 87119. The most promising high-yielding ILs such as ICPP # 171012, 171004, 171102, 171087, 171006, and 171050 having early flowering, high yield, and better grain nutrient contents hold great potential for ready use in pigeonpea breeding programs.

Dwarf plant type is not an advantageous feature for pigeonpea as it attracts *Helicoverpa armigera* and the dwarf bushy growth habit lines have shown 40% damage due to *H. armigera* (Mallikarjuna et al., 2011). Interestingly, all the ILs in this study were tall with semi-spreading secondary and tertiary branches and indeterminate growth habit. Plant height in pigeonpea is a complex and quantitative trait (Byth et al., 1981). The ILs evaluated were found taller than their recurrent parent ICPL 85010 and were similar to ICPL 87119 which puts the ILs in an advantageous position. Also, the range of yield-contributing traits such as number of pods per plant, pod weight per plant, and grain yield per plant was very high indicating the high level of recombination in these ILs.

Grain nutrients were found to be highly influenced by genotype and genotype × environment interactions. The stability of these traits in different environments is therefore important in crop improvement programs to improve the nutritional quality of pigeonpea. Non-significant correlations were observed between grain yield and grain nutrients except grain Zn content. This indicates the possibility to develop nutrient-rich pigeonpea varieties without the trade-offs for yield. Besides, a positive association was observed between the grain nutrients, such as of protein with all four grain nutrients suggesting the simultaneous improvement of varieties with enhanced multiple nutrient contents. Grain Fe and Zn content also showed a significant correlation which is reported in several crops including wheat (Morgounov et al., 2007), sorghum (Upadhyaya et al., 2016; Phuke et al., 2017), pearl millet (Kanatti et al., 2014), proso millet (Vetriventhan and Upadhyaya, 2018), and finger millet (Upadhyaya et al., 2011).

Further, it is important to study the similarity/dissimilarity of ILs with control cultivars for use in breeding programs. The cluster analysis grouped 136 ILs with 5 control cultivars into 10 clusters wherein similar ILs were placed in the same cluster based on agronomic traits and grain nutrients. This will help the breeders to choose trait-specific and diverse ILs for use in a breeding program to introduce new useful alleles derived from wild species into their working collection and/or newly developed cultivars/varieties. In addition, the most diverse pairs of ILs have been identified based on the mean phenotypic diversity index. Involving most diverse ILs in hybridization programs would be helpful in generating new and useful recombinants. Apart from this, a few ILs such as ICPP 171119, ICPP 171098, and ICPP 171045 showing maximum diversity with the recurrent parent ICPL 85010; and ICPP 171027, ICPP 171082, and ICPP 171024 having maximum diversity with the popular variety, ICPL 87119 have been identified for use in breeding programs to develop new cultivars with a broad genetic base.

CONCLUSION

Significant variability was observed in the pre-breeding population, derived from the cross incompatible tertiary genepool species, *C. platycarpus*, for agronomic traits and grain nutrients. Moreover, it is noteworthy that many ILs performed better than the existing mega varieties not only in terms of yield but also for nutrient contents. The most promising high-yielding ILs such as ICPP # 171012, 171004, 171102, 171087, 171006, and 171050 having early flowering, high yield, and better grain nutrient contents compared to the best variety, ICPL 87119 hold great potential for ready use in pigeonpea breeding programs. A thorough multi-location evaluation of these promising trait-specific ILs will be efficacious in identifying region-specific promising lines for their possible direct release as a cultivar (s) having a diverse genetic base, especially in the short-duration group to fit into multiple cropping systems. Further, as no correlations were observed between grain yield and grain nutrients, it shows the possibility of developing nutrient-rich pigeonpea varieties without the trade-offs for yield. Positive associations of grain protein with all four grain nutrients suggests the simultaneous improvement of varieties with enhanced multiple nutrient contents. Besides this, the most diverse promising ILs can be included in the hybridization programs as the potential sources of new and diverse variations. Finally, as *C. platycarpus* is reported to possess photoperiod insensitivity, these ILs hold great potential for evaluation across locations and seasons to identify photo-insensitive lines for use in breeding programs.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/ **Supplementary Material**.

AUTHOR CONTRIBUTIONS

SS conceived the idea and evaluated the material. CS was involved in field evaluation. PP analyzed the data and assisted in preparing the first draft. CN analyzed the seed samples for grain nutrient contents in quality lab. SS prepared the final manuscript. CS, PP, and CN provided their inputs. All authors contributed to the article and approved the submitted version.

FUNDING

Funding support provided by the Global Crop Diversity Trust (GCDT) Grant Numbers GS15020 and GS18010 and CGIAR Research Program on Grain Legumes and Dryland Cereals (GLDC).

ACKNOWLEDGMENTS

This work is part of the initiative "*Adapting Agriculture to Climate Change: Collecting, Protecting and Preparing Crop Wild Relatives*" which is supported by the Government of Norway. The project is managed by the Global Crop Diversity Trust. For further information, visit the project website: http://www.cwrdiversity.org/. The partial funding support provided by the CGIAR Research Program on Grain Legumes and Dryland Cereals (GLDC) is duly acknowledged.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.01055/full#supplementary-material

REFERENCES

Anitha, S., Govindaraj, M., and Kane- Potaka, J. (2019). Balanced amino acid and higher micronutrients in millets complements legumes for improved human dietary nutrition. *Cereal Chem.* 97, 74–84. doi: 10.1002/cche.10227

Ariyanayagam, R. P., and Spence, J. A. (1978). A possible gene source for early, day-length neutral pigeon peas, *Cajanus cajan* (L.) Millsp. *Euphytica* 27, 505–509. doi: 10.1007/bf00043176

Atlin, G. N., Cairns, J. E., and Das, B. (2017). Rapid breeding and varietal replacement are critical to adaptation of cropping systems in the developing world to climate change. *Global Food Secur.* 12, 31–37. doi: 10.1016/j.gfs.2017.01.008

Bohra, A., Mallikarjuna, N., Saxena, K. B., Upadhyaya, H. D., Vales, I., and Varshney, R. K. (2010). Harnessing the potential of crop wild relatives through genomics tools for pigeonpea improvement. *J. Plant Biol.* 37, 83–98.

Bohra, A., Pareek, S., Jha, R., Saxena, R. K., Singh, I. P., Pandey, G., et al. (2017). "Modern genomic tools for pigeonpea improvement: status and prospects," in *The Pigeonpea Genome*. Eds. R. K. Varshney, R. K. Saxena and S. A. Jackson (Cham: Springer International Publishing), 41–54. doi: 10.1007/978-3-319-63797-6_5

Byth, D. E., Wallis, E. S., and Saxena, K. B. (1981). "Adaptation and breeding strategies for Pigeonpea," in *Proceedings of the International Workshop on Pigeonpeas*. vol. 1. (India: ICRISAT), 450–465.

Choudhary, A. K., and Nadarajan, N. (2011). *Breeding improved cultivars of pigeonpea in India* (Kanpur, India: Research Bulletin. Indian Pulses Research Institute).

Dahiya, S. S., Chauhan, Y. S., Srinvastava, S. K., Sekhon, H. S., Waldia, R. S., Gowda, C. L. L., et al. (2001). *Growing extra-short duration pigeonpea in rotation with wheat in the Indo-Gangetic Plain. No. 1. 2001. Report* (Hyderabad, India: ICRISAT).

Dundas, I. S. (1990). "Pigeonpea: cytology and cytogenetics-perspectives and prospects," in *The Pigeonpea*. Eds. Y. L. Nene, S. D. Hall and V. K. Sheila (Wallington, UK: CABI Publ), 117—136.

Duranti, M. (2006). Grain legume proteins and nutraceutical properties. *Fitoterapia* 77, 67–82. doi: 10.1016/j.fitote.2005.11.008

FAOSTAT (2017). *Food and Agriculture Organization of the United Nations Database.* (Rome, Italy: Food and Agriculture Organization). Available at: http://faostat.fao.org/database.

IBPGR and ICRISAT (1993). *Descriptors for pigeonpea [Cajanus cajan (L.) Millsp.]* (Rome, Italy: IBPGR and Patancheru, India: ICRISAT), 31.

ICARDA (1995). *Annual report for 1995* (ICARDA, Aleppo, Syria: Germplasm Program Legumes).

Kanatti, A., Rai, K. N., Radhika, K., Govindaraj, M., Sahrawat, K. L., and Rao, A. S. (2014). Grain iron and zinc density in pearl millet: combining ability, heterosis and association with grain yield and grain size. *SpringerPlus* 3, 763. doi: 10.1186/2193-1801-3-763

Khoury, C. K., Castañeda-Alvarez, N. P., Achicanoy, H. A., Sosa, C. C., Bernau, V., Kassa, M. T., et al. (2015). Crop wild relatives of pigeonpea [Cajanus cajan (L.) Millsp.]: Distributions, ex situ conservation status, and potential genetic resources for abiotic stress tolerance. *Biol. Cons.* 184, 259–270. doi: 10.1016/j.biocon.2015.01.032

Knights, E. J., Southwell, R. J., Schwinghamer, M. W., and Harden, S. (2008). Resistance to *Phytophthora medicaginis* Hansen and Maxwell in wild *Cicer* species and its use in breeding root rot resistant chickpea (*Cicer arietinum L.*). *J. Agric. Res.* 59383–, 387. doi: 10.1071/AR07175

Kumar, S., Gupta, S., Chandra, S., and Singh, B. B. (2004). "How wide is genetic base of pulse crops?" in Pulses in new perspective," in *Proceedings of the National Symposium on Crop Diversification and Natural Resources Management*. Eds. M. Ali, B. B. Singh, S. Kumar and V. Dhar (Kanpur: Indian Society of Pulses Research and Development, Indian Institute of Pulses Research), 211–221.

Lava Kumar, P., Latha, T. K. S., Kulkarni, N. K., Raghavendra, N., Saxena, K. B., Waliyar, F., et al. (2005). Broad-based resistance to pigeonpea sterility mosaic disease in wild relatives of pigeonpea (*Cajanus: Phaseoleae*). *Ann. Appl. Biol.* 146, 371–379. doi: 10.1111/j.1744-7348.2005.040091.x

Malhotra, R. S., Singh, K. B., Vito, M., Greco, N., and Saxena, M. C. (2002). Registration of ILC 10765 and ILC 10766 chickpea germplasm lines resistant to cyst nematode. *Crop Sci.* 42, 1756. doi: 10.2135/cropsci2002.1756

Mallikarjuna, N., and Moss, J. P. (1995). Production of hybrids between *Cajanus platycarpus* and *Cajanus cajan*. *Euphytica* 83, 43–46. doi: 10.1007/BF01677859

Mallikarjuna, N., Jadhav, D., and Reddy, P. (2006). Introgression of *Cajanus platycarpus* genome into cultivated pigeonpea, *C. cajan*. *Euphytica* 149, 161–167. doi: 10.1007/s10681-005-9063-6

Mallikarjuna, N., Senthilvel, S., Jadhav, D. R., Saxena, K. B., Sharma, H. C., Upadhyaya, H. D., et al. (2011). Progress in the utilization of *Cajanus platycarpus* (Benth.) Maesen in pigeonpea improvement. *Plant Breed.* 130, 507–514. doi: 10.1111/j.1439-0523.2011.01870.x

Morgounov, A., Gómez-Becerra, H. F., Abugalieva, A., Dzhunusova, M., Yessimbekova, M., Muminjanov, H., et al. (2007). Iron and zinc grain density in common wheat grown in Central Asia. *Euphytica* 155, 193–203. doi: 10.1007/s10681-006-9321-2

Obala, J., Saxena, R. K., Singh, V. K., Vechalapu, S., Das, R. R., Rathore, A., et al. (2018). Genetic variation and relationships of total seed protein content with some agronomic traits in pigeonpea (*Cajanus cajan* (L.) Millsp.). *Aust. J. Crop Sci.* 12, 1859–1865. doi: 10.21475/ajcs.18.12.12.p1138

Patterson, H. D., and Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika* 58, 545–554. doi: 10.2307/2334389

Phuke, R. M., Anuradha, K., Radhika, K., Jabeen, F., Anuradha, G., Ramesh, T., et al. (2017). Genetic variability, genotype × environment interaction, correlation, and GGE biplot analysis for grain iron and zinc concentration and other agronomic traits in RIL population of sorghum (*Sorghum bicolor* L. Moench). *Front. Plant Sci.* 8, 712. doi: 10.3389/fpls.2017.00712

Pundir, R. P. S., and Singh, R. B. (1987). Possibility of genetic improvement of pigeonpea (*Cajanus cajan* (L.) Millsp.) utilizing wild gene sources. *Euphytica* 36, 33–37. doi: 10.1007/bf00730644

Rao, N. K., Reddy, L. J., and Bramel, P. J. (2003). Potential of wild species for genetic enhancement of some semi-arid food crops. *Genet. Resour. Crop Ev.* 50, 707–721. doi: 10.1023/A:1025055018954

Sahrawat, K. L., Kumar, G. R., and Murthy, K. V. S. (2002). Sulfuric acid–selenium digestion for multi-element analysis in a single plant digest. *Commun. Soil Sci. Plant Anal.* 33, 3757–3765. doi: 10.1081/css-120015920

Saxena, K. B., Singh, L., Reddy, M. V., Singh, U., Lateef, S. S., Sharma, S. B., et al. (1990). Intra species variation in *Atylosia scarabaeoides* (L.) Benth., a wild relative of pigeonpea (*Cajanus cajan* (L.) Millsp.). *Euphytica* 49, 185–191. doi: 10.1007/BF00036287

Saxena, K. B., Rao, A. N., Singh, U., and Remanandan, P. (1996). Intraspecies variation in *Cajanus platycarpus* for some agronomic traits and crossability. *Int. Chickpea Pigeonpea Newslett.* 3, 49–51.

Saxena, K. B., Kumar, R. V., and Sultana, R. (2010). Quality nutrition through pigeonpea—a review. *Health* 02, 1335–1344. doi: 10.4236/health.2010.211199

Saxena, K., Choudhary, A. K., Srivastava, R. K., Bohra, A., Saxena, R. K., and Varshney, R. K. (2019). Origin of early maturing pigeonpea germplasm and its impact on adaptation and cropping systems. *Plant Breed* 138, 243–251. doi: 10.1111/pbr.12696

Sharma, S., and Upadhyaya, H. D. (2016). Interspecific hybridization to introduce useful genetic variability for pigeonpea improvement. *Indian J. Genet. Pl. Br.* 76, 496–503. doi: 10.5958/0975-6906.2016.00067.5

Sharma, S. B., Remanandan, P., and McDonald, D. (1993). Resistance to *Meloidogyne javanica* and *Rotylenchulus reniformis* in wild relatives of pigeonpea. *J. Nematol.* 25, 824–829.

Sharma, S., Upadhyaya, H. D., Varshney, R. K., and Gowda, C. L. L. (2013). Prebreeding for diversifcation of primary genepool and genetic enhancement of grain legumes. *Front. Plant Sci.* 4, 309. doi: 10.3389/fpls.2013.00309

Sharma, S., Paul, P. J., Kumar, C. V., Rao, P. J., Prasanthi, L., Muniswamy, S., et al. (2019). Evaluation and identification of promising introgression lines derived from wild *Cajanus* species for broadening the genetic base of cultivated pigeonpea (*Cajanus cajan* (L.) Millsp.). *Front. Plant Sci.* 10, 1269. doi: 10.3389/fpls.2019.01269

Sharma, S. B. (1995). Resistance to *Rotylenchulus reniformis*, *Heterodera cajani*, and *Meloidogyne javanica* in accessions of *Cajanus platycarpus*. *Plant Dis.* 79, 1033–1035. doi: 10.1094/PD-79-1033

Singh, S., Gumber, R. K., Joshi, N., and Singh, K. (2005). Introgression from wild *Cicer reticulatum to* cultivated chickpea for productivity and disease resistance. *Plant Breed.* 124477–, 480. doi: 10.1111/j.1439-0523.2005.01146.x

Srivastava, R. K., Vales, M. I., Sultana, R., Saxena, K. B., Kumar, R. V., Tanki, H. P., et al. (2012). Development of 'super-early' pigeonpeas with good yield potential from early × early crosses. *J. SAT Agric. Res.* 10, 1–6.

Subbarao, G. V. (1988). *Salinity tolerance in pigeonpea (Cajanus cajan) and its wild relatives. [Ph.D. dissertation]* (Kharagpur, India: Indian Institute of Technology).

Sujana, G., Sharma, H. C., and Manohar, R. D. (2008). Antixenosis and antibiosis components of resistance to pod borer, *Helicoverpa armigera* in wild relatives of pigeonpea. *Int. J. Trop. Insect Sci.* 28, 191–200. doi: 10.1017/S1742758408191822

Upadhyaya, H. D., Ramesh, S., Sharma, S., Singh, S. K., Varshney, S. K., Sarmad, N. D. R. K., et al. (2011). Genetic diversity for grain nutrient contents in a core collection of finger millet (*Eleusine coracana* (L.) Gaertn.) germplasm. *Field Crops Res.* 121, 42–52. doi: 10.1016/j.fcr.2010.11.017

Upadhyaya, H. D., Dwivedi, S. L., Singh, S., Sahrawat, K. L., and Singh, S. K. (2016). Genetic variation and post flowering drought effects on seed iron and zinc in ICRISAT sorghum mini core collection. *Crop Sci.* 56, 374. doi: 10.2135/cropsci2015.05.0308

Varshney, R. K., Thudi, M., Roorkiwal, M., He, W., Upadhyaya, H. D., Yang, W., et al. (2019). Resequencing of 429 chickpea accessions from 45 countries provides insights into genome diversity, domestication and agronomic traits. *Nat. Genet.* 51, 857–864. doi: 10.1038/s41588-019-0401-3

Vetriventhan, M., and Upadhyaya, H. D. (2018). Diversity and trait-specific sources for productivity and nutritional traits in the global proso millet (*Panicum miliaceum* L.) germplasm collection. *Crop J.* 6, 451–463. doi: 10.1016/j.cj.2018.04.002

Ward, J. H. (1963). Hierarchical Grouping to Optimize an Objective Function. *J. Am. Stat. Assoc.* 58, 236–244. doi: 10.1080/01621459.1963.10500845

Wheal, M. S., Fowles, T. O., and Palmer, L. T. (2011). A cost-effective acid digestion method using closed polypropylene tubes for inductively coupled plasma optical emission spectrometry (ICP-OES) analysis of plant essential elements. *Anal. Methods* 3, 2854–2863. doi: 10.1039/c1ay05430a

Conflict of Interest: 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.

The reviewer PK declared a past co-authorship with one of the authors SS to the handling editor.

Copyright © 2020 Sharma, Paul, Sameer Kumar and Nimje. 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) and the copyright owner(s) 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 Plant Science

REVIEW
published: 19 August 2020
doi: 10.3389/fpls.2020.01217

Crop Wild Relatives as Germplasm Resource for Cultivar Improvement in Mint (*Mentha* L.)

Kelly J. Vining[1*], *Kim E. Hummer*[1,2], *Nahla V. Bassil*[1,2], *B. Markus Lange*[3], *Colin K. Khoury*[4,5] *and Dan Carver*[5,6]

[1] Department of Horticulture, Oregon State University, Corvallis, OR, United States, [2] National Clonal Germplasm Repository, USDA-ARS, Corvallis, OR, United States, [3] Institute of Biological Chemistry and M.J. Murdock Metabolomics Laboratory, Washington State University, Pullman, WA, United States, [4] Decision and Policy Analysis, International Center for Tropical Agriculture (CIAT), Cali, Colombia, [5] National Laboratory for Genetic Resources Preservation, Agricultural Research Service, United States Department of Agriculture, Fort Collins, CO, United States, [6] Colorado State University, Geospatial Centroid, Fort Collins, CO, United States

Mentha is a strongly scented herb of the *Lamiaceae* (formerly *Labiatae*) and includes about 30 species and hybrid species that are distributed or introduced throughout the globe. These fragrant plants have been selected throughout millennia for use by humans as herbs, spices, and pharmaceutical needs. The distilling of essential oils from mint began in Japan and England but has become a significant industrial product for the US, China, India, and other countries. The US Department of Agriculture (USDA), Agricultural Research Service, National Clonal Germplasm Repository (NCGR) maintains a mint genebank in Corvallis, Oregon. This facility preserves and distributes about 450 clones representing 34 taxa, hybrid species, advanced breeder selections, and F_1 hybrids. Mint crop wild relatives are included in this unique resource. The majority of mint accessions and hybrids in this collection were initially donated in the 1970s by the A.M. Todd Company, located in Kalamazoo, Michigan. Other representatives of diverse mint taxa and crop wild relatives have since been obtained from collaborators in Australia, New Zealand, Europe, and Vietnam. These mints have been evaluated for cytology, oil components, verticillium wilt resistance, and key morphological characters. Pressed voucher specimens have been prepared for morphological identity verification. An initial set of microsatellite markers has been developed to determine clonal identity and assess genetic diversity. Plant breeders at private and public institutions are using molecular analysis to determine identity and diversity of the USDA mint collection. Evaluation and characterization includes essential oil content, disease resistance, male sterility, and other traits for potential breeding use. These accessions can be a source for parental genes for enhancement efforts to produce hybrids, or for breeding new cultivars for agricultural production. Propagules of *Mentha* are available for distribution to international researchers as stem cuttings, rhizome cuttings, or seed, which can be requested through the GRIN-Global database of the US National Plant Germplasm System, subject to international treaty and quarantine regulations.

Keywords: mint, peppermint, spearmint, verticillium wilt, monoterpene

OPEN ACCESS

Edited by:
Thomas M. Davis, University of New Hampshire, United States

Reviewed by:
Benoit Bertrand, Institut National de la Recherche Agronomique (INRA), France
Kateřina Smékalová, Crop Research Institute (CRI), Czechia

***Correspondence:**
Kelly J. Vining
kelly.vining@oregonstate.edu

Specialty section:
This article was submitted to Plant Breeding, a section of the journal Frontiers in Plant Science

Received: *31 October 2019*
Accepted: *27 July 2020*
Published: *19 August 2020*

Citation:
Vining KJ, Hummer KE, Bassil NV, Lange BM, Khoury CK and Carver D (2020) Crop Wild Relatives as Germplasm Resource for Cultivar Improvement in Mint (Mentha L.). Front. Plant Sci. 11:1217. doi: 10.3389/fpls.2020.01217

INTRODUCTION

The millennia of human effort involved in the domestication of economically important agricultural and horticultural crops is documented and broadly discussed among plant evolutionary biologists (Darwin, 1859; Darlington, 1963; Zohary, 1969; Zohary, 1984; Zohary, 2004). Plant domestication, the genetic modification of a wild form to create an altered plant to meet human needs, has produced many plants incapable of existing in the wild (Doebley et al., 2006). This "domestication syndrome" (Hammer, 1984) involves a combination of traits that are different from those of the wild progenitors. These domesticated plants may have larger fruit or grains, robust growth habits, loss of sexual fertility, loss of bitterness, or synchronous flowering. These plants may not compete successfully in the natural world.

As Zohary (2004) points out, seed propagated agronomic crops, many of which display domestication syndrome, have undergone stabilizing selection to protect fertility. Grain crops have rigid protection for sexual reproduction with streamlined development of flowers, fruits, and seed. Chromosomes behave normally at meiosis and deviants do not survive to reproduce. Chromosomes are balanced with little pollen or seed sterility. In contrast, clonally propagated fruit crops, which also display domestication syndrome, not only tolerate but also promote the reduction of pollen and seed fertility and lower chromosome stability. Parthenocarpy, unequal ploidy levels, aneuploids, and other innovative mass production solutions reduce seed set without reducing fruit production.

At the next level, crops maintained by clonal propagation and grown for their non-reproductive organs have the most drastic disruptions to their flowering and fruiting systems. This group, according to Zohary (2004), demonstrates bizarre chromosomal segregation and unusual ploidy levels. Mint species are an exemplar of this category.

Multiple species of mint have been used for medicinal purposes by humanity from prehistory. From savory herbs produced in monasteries to single-family needs of the kitchen garden, commercial peppermint production for menthol, and essential oil extraction from other mint species, mints represent a significant global economic commodity.

The objectives of this manuscript are to describe the domestication of mint and its uses to humanity. Trait and genotype examples and a summary of the preservation of *Mentha* genetic resources and global genebank operations will be presented. Present breeding improvements and future possibilities considering future genetic analysis will be projected.

DOMESTICATION OF MINT

Taxonomy

Mentha is a strongly scented herb genus of the *Lamiaceae* (formerly *Labiatae*) and includes 18 species, 31 subspecies or botanical varieties, and 11 recognized hybrid species (Tucker and Naczi, 2007; GRIN-Global, 2020). The mint family includes diverse additional aromatic genera, such as mountain mint (*Pycnanthemum* L.), lavender (*Lavendula* L.), sage (*Salvia* L.), rosemary (*Rosmarinus* L.), and oregano (*Origanum* L.), which are grown commercially for essential oils that are distilled from leaves and stems. Mint shoots and leaves are used for medicinal and aromatic purposes, such as dried organic extracts, distillates, condiments, and food flavorings. Likewise, *Mentha* includes many commercially valuable species, including peppermint (*Mentha* ×*piperita*), Scotch spearmint (*M.* ×*gracilis*), native spearmint (*M. spicata*), American wild mint (*M. canadensis*), and corn mint (*M. arvensis*), which are cultivated in different parts of the world for their culinary and medicinal properties. Plants in this genus are herbaceous, rhizomatous, perennial, and aromatic, with smooth, wide-spreading underground stems, which are square in cross-section. The leaves are arranged in opposite pairs, and the white, pink, or purple flowers are produced in clusters.

The name *Mentha* is derived from Classical Greek mythology, from Minthe, (Minthê), who was a beautiful Cocythian (river nymph) beloved by Hades (Pluto) god of the underworld. Minthe was metamorphosed into dust by Hades's wife, Demeter (Persephone) (Rosengarten, 1969), but Hades caused the fragrant mint plant to grow from the dust. The etymology of "mint" is from the old English minte (= mint plant), which is derived from Proto-Germanic, through Latin from the Ancient Greek, and is akin to old Norse.

At the base of Mt. Minthe, there was a temple dedicated to Hades and a grove for Demeter (Ovid Met. 10) Near Pylos, Greece, Mt. Minthe is thought to be a location of the origin of triploid sterile spearmint ($2n = 3x = 36$), which likely arose from the introgression of the conspecific endemic ancient amphidiploid *Mentha spicata* ($2n = 4x = 48$) and the diploid *M. longifolia* ($2n = 2x = 24$) (Kokkini, 1991; Lawrence, 1991).

Since the time of Linneaus, more than 3,000 specific epithets have been reported for *Mentha* (Tucker and Naczi, 2007). As a measure of mint species taxonomic diversity, the Global Biodiversity Information Facility (GBIF) network includes more than 740,000 locality data points from 454 *Mentha* taxa with occurrences throughout the world (GBIF, 2020). The plethora of names and occurrences of hybrid and naturalized mints have created confusion in the literature. Wild species of mint hybridize readily, and over evolutionary time, native hybrid-species swarms developed in conspecific regions. Subsequently, minor variances have achieved species rank.

The Plants of the World database (http://www.plantsoftheworldonline.org/) (Kew Science, 2020) describes 39 taxa including, 24 species and 15 hybrid species (**Figure 1**). Tucker and Naczi (2007) prepared a more conservative list of the number of mints, synonymizing many to recognize 18 species, 31 subspecies or botanical varieties, 11 hybrid species, and one excluded species. They chose to consider *Mentha cunninghamii* (Benth.) Benth., as *Micromeria cunninghamii* Benth. GRIN-Global (2020) taking a moderate course, includes 20 *Mentha* species.

For this manuscript, we applied the determination by Tucker and others (Tucker and Chambers, 2002; Tucker and

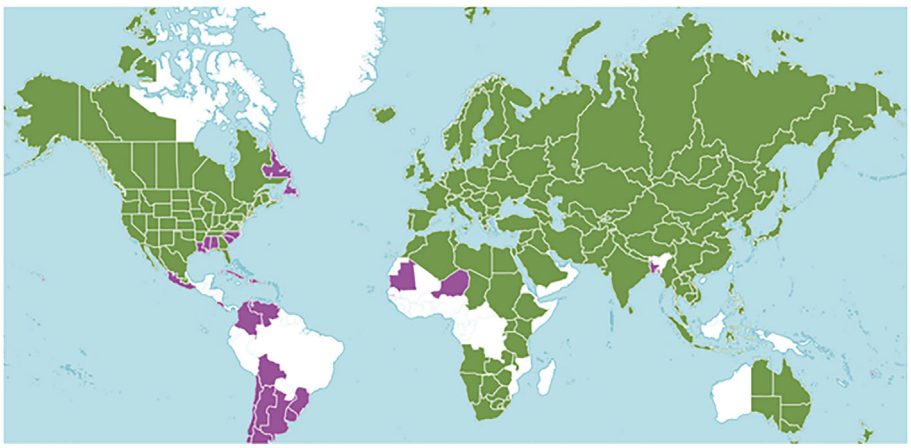


FIGURE 1 | *Mentha* distribution including 39 taxa prepared by Kew Science. The green regions represent native species and purple represents areas of introduction. http://www.plantsoftheworldonline.org/taxon/urn:lsid:ipni.org:names:30016176-2.

Naczi, 2007) for 18 recognized species, with one exception: For this review, we considered *Mentha repens* (J.D. Hook.) Briq. as *M. pulegium* L. subsp. *repens*.

Species Distribution

Mentha L. has a cosmopolitan native range (Kew Science, 2020). The *Mentha* distribution map (Kew Science) included 163 countries, provinces, and regions for native distribution and 43 regions of introduction (**Figure 1**). Considering the tendency for species hybridization within this genus and the successful expansion strategies of mint around the globe, we used a conservative estimation to develop global richness maps. We considered only the taxa accepted by Tucker and Naczi (2007). We prepared global richness maps searching GBIF data for the 17 mint species (**Figures 2** and **3**).

We separately examined world occurrences of point data within GBIF data for each of the 17 species. We filtered out the introduced, naturalized, and cultivated locality data, keeping only natural endemic occurrences. We used climatic and topographic predictors to model likely occurrence for the 17 individual species maps. The data for the separate species maps were merged to produce the global species richness maps (**Figures 2** and **3**). These

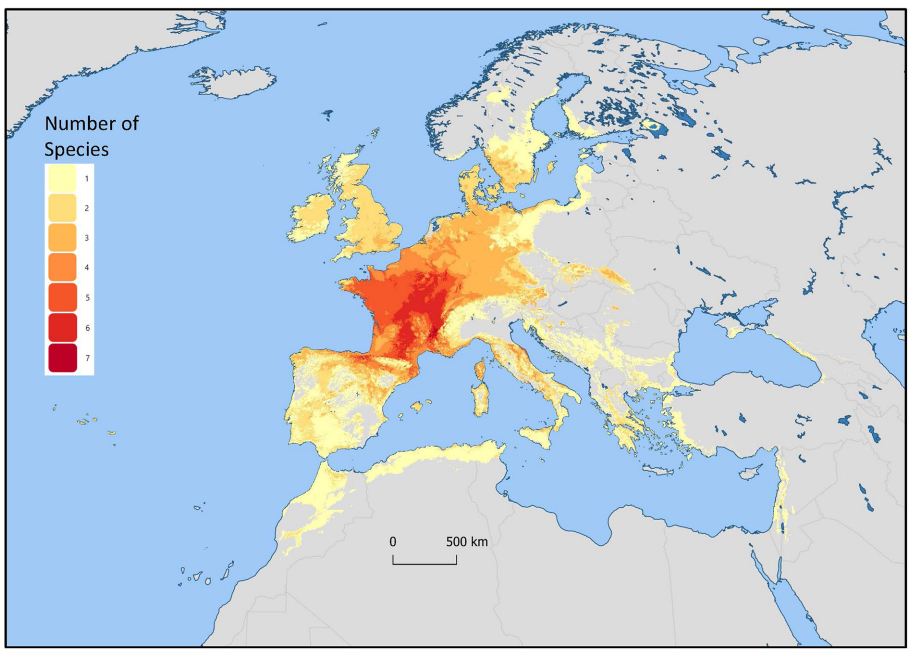


FIGURE 2 | Global richness map for 17 endemic *Mentha* L. species. The colors indicate the number of species in that region.

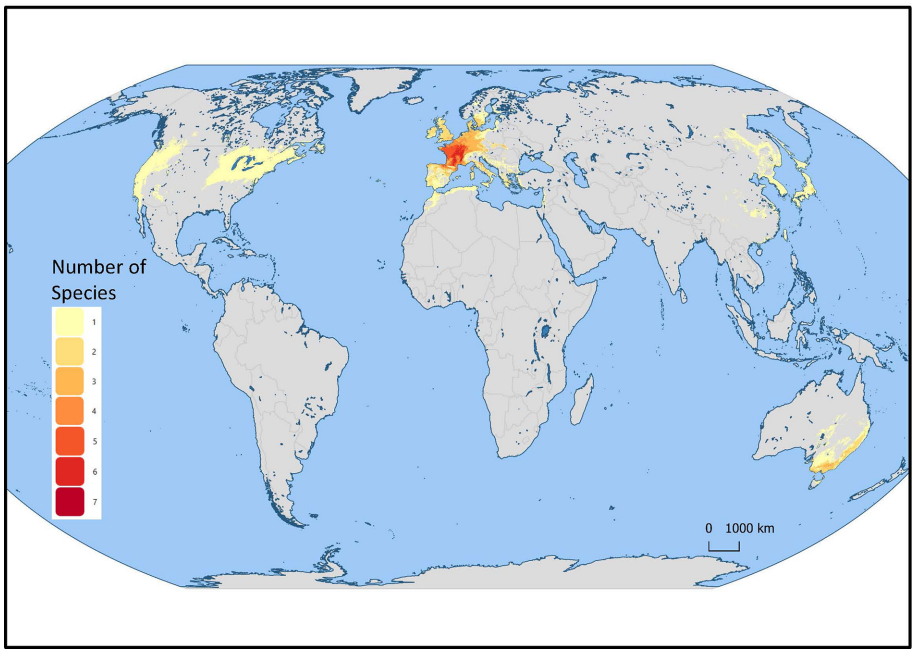


FIGURE 3 | *Mentha* L. European region of the global richness map for 17 endemic European *Mentha* species. The colors indicate the number of species in that region.

maps can be used for applied research or conservation planning and investigating the processes that have shaped these patterns.

While the highest diversity of present day species occurs in Western Europe (**Figure 2**), significant endemic mint species occur in Eastern and Western North America, Asia, Southern Australia, and Tasmania (**Figure 3**). In addition to this natural species diversity, hundreds of thousands of data points of introduced, naturalized, and cultivated mint occur throughout five continents (GBIF, 2020). The multitude of global mint occurrences (**Figure 1**) speak to the global success of this genus, starting from Centers of Diversity in Europe, Asia, North America, and Australia and spreading throughout the globe.

The ease of propagation by seeds and clonal propagules, such as rhizomes and cuttings, and the survivability of the plants during harsh and undesirable climactic conditions, allow for spread and diversification. The utilitarian application of mint for human pharmaceutical, food, and cosmetic needs encouraged domestication and cultivation throughout the world and across multiple cultures.

Prehistory

Throughout human history, medicinal and aromatic plants have been used for flavor enrichment in culinary and medicinal purpose in folk medicine (Sonmezdag et al., 2017). *Mentha* has been of great importance considering its unique aroma and nutritional value (Naghibi et al., 2010). Archeological excavations showed that the usage of lavenders, sage, and mints occurred in prehistoric times, harvested locally from the wild (Nuńez and De Castro, 1992). In western traditions, mints have been cultivated in the dry, mild, and cold districts of Asia, Europe, and North Africa since antiquity (Jamzad et al., 2003). Native peoples in America have documented traditions of uses of mints as cold remedies (Chehalis, Cowlitz), treatments for gastrointestinal ailments (Kiowa), food (Kiowa, chewing leaves), and ceremonial medicine (Navaho) (Vestal and Schultes, 1939; Elmore, 1944; Gunther, 1973).

Antiquity

Multiple mint species are referenced (Bostok, 1885) throughout Greek and Roman herbals (Aeschylus, Hippocrates, Krataeus, Dioscorides, and Galen; Romans: Cato, Ovid, and Pliny the Elder); Asian medicinal traditions, traditional Chinese medicine, and the Ayurvedic tradition of India.

The earliest published mint species images in existence can be found on gr. III fol. 129r and 132r of the Juliana Anicia Codex or Vindobonensis Codex (Janick and Hummer, 2012). The Juliana Anicia Codex is a magnificently illustrated manuscript with written information based on the *Peri Ylis Iatrikis* (De Materia Medica in Latin; Of Medical Matters of Dioscorides). This codex, one of the earliest books, was presented to the imperial Princess Juliana Anicia in Constantinople 512 CE (Collins, 2000). In the 15th century, it was purchased by the Emperor Maximilian for the Imperial Library and was moved to Vienna, where it now resides. The translation of the uses for mints as described in Byzantine Greek written in the background of the images was prepared by Lily Beck (Beck, 2005) and is presented (**Table 1**). These uses described in the Juliana Anicia Codex were reiterated, revised, and interpreted in many recensions of Dioscorides

TABLE 1 | English translation of reference to species of *Mentha* in *De Materia Medica* of Dioscorides (Beck, 2005).

Book	Translation
III, 31	[Ὑλήχων (βλήχων), *M. pulegium* L. Pennyroyal]
	The pennyroyal: it is a familiar herb, warming, thinning and promoting digestion. When drunk, it draws the menses, afterbirth, and embryos/fetuses. It brings up from the long phlegm when drunk with salt and honey and it helps people with spasms, and with sour wine mixed with water it relieves nausea and gnawing pains of the stomach. With wine, it drives down the bowel dark matter and helps those bitten by wild animals and when applied to the nostrils with vinegar, it revives those who fainted.
	Ground up dry and burned, it also strengths the gums; plastered on with barley groats, it soothes all inflammations; it is suitable to use all by itself on the gouty until the skin surfaces becomes irritated and when used with a cerate, it checks facial eruptions; it also helps patients with spleen disease when plastered on with salt. Its decoction used as a wash stops itching and is suitable in a sitz bath for uterine inflations, indurations, and twistings. Some people call it *blechon* because sheep that taste it when in bloom bleat continuously.
III, 34	[ὑδύοσμον *Mentha* sp. L. Green mint]
	The green mint, but some call it *minthe*: it is a well-known little herb having warming, astringent and drying properties; it is for this reason that its juice, when drunk with vinegar, staunches blood, destroys the round intestinal worm, rouses sexual desire, and when two or three little sprays are drunk with the juice of sour pomegranate, stops hiccups, vomiting, and cholera. Applied as a plaster with barley groats, it dissipates abscesses, placed on the forehead, it assuages headaches, and it abates distension and swelling of the breasts.
	With salt, it is a plaster for people bitten by dogs, and its juice with hydromel is suitable for earaches. Used by women as a pessary before sexual intercourse, it causes barrenness, and if rubbed on a rough tongue, it smothes it; it keeps milk from curdling when little sprays are stirred about in it, and it is through and through wholesome and spicy.
	There is also a wild green mint which has thicker leaves, all told it is larger than bergamot mint which has thicker leaves, all told it is larger than bergamont mint, rather foul smelling, and less useful for health purposes.

works and in subsequent herbals for the next several millennia (Collins, 2000; **Figures 4** and **5**).

Dioscorides mentions about 500 plants (Singer, 1927). About 130 of these were referenced in the Hippocratic collection. Thus, many of these treatments, including peppermint, were utilized in the Greek world for more than four centuries before Dioscorides writings and have survived throughout the times to medicinal applications in the present day.

Mint is mentioned in the bible: Mathew 23:23, "For you tithe mint, dill, and cumin, and have neglected the weightier matters of the law: justice and mercy and faith. It is these you ought to have practiced without neglecting the others" and Luke 11:42, "Woe to you Pharisees, because you give God a tenth of your mint, rue and all other kinds of garden herbs, but you neglect justice and the love of God. You should have practiced the latter without leaving the former undone." Parry (1925) suggests that these quotations refer to *Mentha longifolia*, long leafed mint, native to North Africa. Robert Gunther (1934) prepared a translation of John Goodyear's 1655 version of Eduosmos agrios in Dioscorides as: "But ye wilde mint, which ye Romans call *Mentastrum*, is round in the leaves and altogether greater than *Sisymbrium*, but poisonous in smell, and lesse fitting for use in health."

Mentha in the traditional Galenic medicine in the Islamic period (**Table 2**) can be traced to the writings of Dioscorides and other Greek authors.



FIGURE 4 | *M. pulegium*, pennyroyal, *Ğleḵon*, γλήχον Vienna Dioscorides, Juliana Anicia Codex, dating to 512 CE. This illustration is Book III, plate 31. The Arabic writing on the right is translated "type of thyme". The Arabic on the left is the translation of *Ğleḵon*. The faded Byzantine Greek in the background is translated by L. Beck (**Table 1**).

Egyptian

The *Ebers Papyrus* dating to 1550 BCE and purchased by Georg Ebers in Luxor in 1874, is one of the earliest surviving Egyptian medical papyri. It is kept in the library at the University of Lipzig, Germany. This 110-page scroll is about 20 m long, and the text is scribed in hieratic Egyptian, likely copied from earlier manuscripts. One treatment suggests that peppermint be mixed with flour, incense, wood of the waneb plant, a stag's horn, sycamore seeds, mason's plaster, seeds of zart, and water to create a curative paste for headaches supporting later Anglo-Saxon and Greek prescriptions for headache and migraine pain relief (Aboelsoud, 2010; Elansary and Ashmawy, 2013).

In the *Hearst Papyrus*, dated to approximately 2000 BCE, peppermint is recommended as a treatment for rhinitis, where a plaster is applied directly to the nose. In prescription 171, peppermint is mixed with wine and is used to treat what might be edema of the legs. After the mint and wine mixture is consumed, some sort of bloodletting is required, though the physician doesn't provide much in terms of specificity. Mint is still used to treat edema today, though most modern references



FIGURE 5 | Image of *M. aquatica*, labeled Eduosmos agrios, is from the Pierpont Morgan recession of Dioscorides *De material medica*. This book dates to about 1050 CE, and the plants are alphabetically arranged. Morgan Library New York.

TABLE 2 | Application of mint in Arabic traditional texts. (Iranian encyclopedia online, 2020).

ʿAli b. ʿAbbās Majusi	(date 384/994 CE	Raw mint mixed with vinegar is an effective medicine for treatment of swooning and vomiting.
Ebn Sinā	(d. 428/1037 CE	Mentioned that its sherbet* is useful for curing jaundice.
Jorjāni	(d. 531/1136 CE)	Claimed that it was necessary to eat mint after having cucumber, pumpkin, and the seeds of lettuce. Additional references of detailed accounts about mint in traditional Arabic medicine.

**Persian "sherbet", sekanjabin, a concoction of honey and vinegar flavored with mint extract.*

deal with topical creams for livestock, particularly cattle (Aboelsoud, 2010).

European Medieval and Renaissance 1400–1600

The majority of Medieval and Renaissance herbals elaborated the traditions described by ancient Greek and Roman text (Hummer and Janick, 2007). Mint species played a prominent role in many treatments. Adams et al. (2011) compared malaria treatments in Fuchs (1543) and Brunfels (1532), both of whom suggested the use of mints against this disease. Fuchs (1543) also recommended taking pennyroyal (*M. pulegium*) in vinegar against epilepsy. Essential oil extracted from pennyroyal was amongst the pro-convulsive essential oils discussed by Duke (1985), Burkhard et al. (1999), and Riddle and Estes (1992).

Nicholas Culpeper's work, the *Complete Herbal* (Culpeper, 1652) documents the use of several mint species. Unlike the Egyptian papyri, or Anglo-Saxon Leech books, Culpeper's work provides in-depth descriptions and growth habits of the plant entries. Culpeper discusses spearmint, which he refers to alternatively as *"heartmint."* The plant is labeled an *"herb of Venus,"* which directly refers to Dioscorides' *De Materia Medica*, inferring that mint possesses healing, binding, and drying qualities. Much of the entry borrows directly from this text but has some additions. Culpeper includes treatments using mint for a sore and itchy scalp, pain of the ears, venomous bites, headache, indigestion, and flatulence. He also includes prescriptions for the resolution of bad breath and soreness of the gums and palate. Horse or wild mint are then referenced. Due to menthol's innate analgesic qualities, mint is described to be used against headaches, migraines, and general pain and swelling. Mint is also recommended as a digestive aid. Mint relaxes stomach muscles, allowing food and flatulence to pass more easily.

Mint was employed topically to treat dry, itchy skin, and insect and animal bites and lessen the appearance of scars, bruises and scabs, particularly on the scalp and face. Due to its antimicrobial and antifungal properties, mint was used for millennia to treat fungal infections, such as ringworm, and to treat parasites, such as roundworm. Mint is also a diuretic and, as such, can be used to stimulate urination and possibly even treat hypertension.

INDUSTRIAL PRODUCTION OF MINT

Mint was significant to humans from early history, being collected from the wild and brought to the convents, monasteries, and kitchen gardens for direct use. Japan preceded western countries in the cultivation of mint and extraction of menthol from *M. arvensis*, begun before 0 CE (Flueckiger, 1891; Hayden, 1929). Tamba Yasudori (984 CE), in the oldest surviving Japanese medical text, stated that menthol was used in preparation of an aqueous eye medication. Japan began exporting menthol in 1873 (Gildemeister and Hoffman, 1913). In Japan, this oil was extracted from corn mint (*Mentha arvensis* L.).

Production of peppermint for menthol in large areas in the western world did not occur until 1750 in England (Parry, 1925). By 1796, about 40.5 ha (100 acres) of peppermint (*M. ×piperita*) were grown near Mitcham, England. The plant material was harvested, bundled, and sent to London for distilling. This early production of peppermint oil yielded 900–1360 kg (2,000–3,000 lb) (Parry, 1925). Over the next 20 years, the production area expanded to 10

townships surrounding Mitcham. By 1940, peppermint cultivation in the original Mitcham area ceased (Parry, 1925).

Peppermint plants were imported from England to the United States with early settlers. By 1812, peppermint was raised commercially near Ashfield, Massachusetts (Lawrence, 1991). Enterprising farmers and distillers, such as Archibald Burnett and Hiram G. Hotchkiss, brought mint westward to New York. By 1830, mint production was sufficient for nine distilleries. These stills also processed spearmint, tansy, wintergreen, spruce, and hemlock oils (Lawrence, 1991). Mint production began to move further west in the US. Both spearmint and peppermint were introduced in New York, Ohio, Michigan, and Indiana. In the western states, mint from Oregon and Washington spread into Idaho, Montana, California, Nevada, and Utah. Lawrence (1991) provides details of early mint production in the US.

In 1853, Albert M. Todd imported 'Black Mitcham' roots from England and began production in St. Joseph County, Michigan. 'White Mitcham,' the original introduction, had become known as 'American peppermint.' Todd found 'Black Mitcham' to be superior to 'White Mitcham' for production in Michigan. Todd's production and distillery in Kalamazoo became a major mint producer. By 1915, Todd began several hundred acres of Scotch spearmint (*M.* x *gracilis* Sole), because of its productivity over the native spearmint (*M. spicata*).

The main mint crops produced in the US are peppermint and spearmint, the latter including Scotch and native spearmints. 'Black Mitcham' peppermint has been the most widely grown peppermint cultivar since at least the mid-20th century (Lawrence, 1991).

The US does not report peppermint production to the Food and Agriculture Organization of the United Nations (FAOSTAT, 2018). In 2018, the countries reporting the largest peppermint production included Morocco, Argentina, Mexico, Bulgaria, and Spain. Oil distilled in China, India, and Japan is *M. arvensis* var *piperescens* and is not counted in the FAO peppermint database. In 2018, the FAO reported world mint oil production at 106,728 MT.

The annual farm gate value of peppermint production in the US has been between $100 and $150 million, while that of spearmint has ranged between $45 to $50 million during the past decade (NASS, 2020).

Recent US mint oil production was highest in 2011, at about 7 million pounds (3,175 MT) of peppermint oil and 2.2 million pounds (998 MT) of spearmint oil. Since then, US production has dropped to under 6 million pounds (2,721 MT), as production of (-)-menthol from oils of *M. arvensis* has been increasing in China and India. The demand for mint oil by Southeast Asian countries has been driving the expansion of the market in recent years (Liu and Lawrence, 2006).

PROTECTING *MENTHA* GENETIC RESOURCES

The FAO manages a list of world genebanks or botanical gardens and the crops that they preserve. This information is publically available in two global databases: Genesys PGR2 and CATIE. We queried both of these databases on 14 April 2020 and totaled entries for *Mentha* and combined results. This query determined that 65 genebanks in 42 countries as preserving *Mentha* genetic resources. The largest mint collections are held in the US, Ukraine, Germany, Japan, Portugal, and the United Kingdom (**Supplemental Table 1**).

USDA NATIONAL CLONAL GERMPLASM REPOSITORY AT CORVALLIS, OREGON

The *Mentha* collection of the USDA, NCGR in Corvallis, Oregon, includes representatives of approximately 450 accessions of 13 species and 10 hybrid species plus cultivated types (**Table 3**). This working genebank maintains the primary collections as plants in containers in greenhouse or lath house environments (Hummer et al., 2020). Most of these plants represent diverse wild species collections from 57 countries.

Most mints in this collection were originally donated to the USDA from Dr. Merrit J. Murray, breeder at A.M. Todd Company, Kalamazoo, Michigan, dating back to the 1970s. At that time, the Todd Company decided to discontinue their mint breeding program and donated the collection to Dr. Chester Ellsworth Horner, USDA researcher in Corvallis, Oregon. Dr. Al Haunold assumed responsibility upon Dr. Horner's retirement and subsequently donated the collection to the NCGR in the mid-1980s.

At present, 216 mint accessions at the NCGR are species or advanced breeder selections originally donated from M. J. Murray. Other representatives of diverse mint taxa including wild accessions have since been obtained from breeders and taxonomic collaborators in the US, Australia, New Zealand, Europe, and Vietnam.

Besides maintaining diverse species representatives, the NCGR collection includes advanced breeder lines, cultivar and germplasm releases, and what are referred to as "donor-named selections," where the NCGR has kept the names provided by donors. In addition, virologists have donated 16 pathogen-positive mints for virus identification studies and certification programs.

DISTRIBUTION FROM THE NCGR MINT GENE BANK

The NCGR was dedicated in 1981. Since that time, the NCGR has distributed > 9,700 mints as plants, cuttings, rhizomes, tissue cultures, or seed lots for 1,277 orders. About 31% of the orders were sent to US researchers (including government agencies, universities, and non-profit organizations), 33% to US individuals and private companies, and 36% were shipped to foreign requestors. Additional plant and tissue culture shipments were made for secure remote backup of the collection at the USDA National Laboratory for Genetic Resource Preservation in Ft. Collins, Colorado, the designated base germplasm collection within the US National Plant Germplasm System (NPGS).Some

TABLE 3 | *Mentha* L. species, distribution, habitat, and number of accessions in the USDA ARS National Clonal Germplasm Repository-Corvallis.

Species	NCGR-Corvallis Accessions	Range	Habitat
Mentha aquatica L.	15	Europe except extreme north	wet edges of ponds, lakes, canals
Mentha aquatica var. citrata	14	North America and Central Europe	moist fields and roadsides, cultivated
Mentha arvensis L.	3	Southern and Western Europe	moist fields and roadsides
	0	Northern and Eastern Europe	moist fields and roadsides
Mentha australis R. Br.	2	Australia - except West Australia), Tasmania	near watercourses and waterholes and damp gullies
Mentha canadensis L.	43	North America and Eastern Asia	streambankslake shores, moist fields, roadsides. Also cultivated
Mentha cervina L.	3	Western Mediterranean, Spain, Portugal, and Mediterranean France	damp land to water overwintering under water in native habitat
Mentha dahurica Fisch. ex Benth.	0	Eastern Siberia to Northern China, Hokkaido, Japan	water meadows, shores of rivers and lakes, thickets, wood margins
Mentha diemenica Spreng.	1	South Australia to Tasmania and New South Wales, Australia	wet edges of ponds, lakes, canals
Mentha gattefossei Maire	5	Grand and Moyen Atlas Mountains, Morocco	edges of dayas, damp pastures
Mentha grandifora Benth.	0	Eastern Australia from Inland Queensland to Northern New South Wales	sandy soil
M. hybrids	100	cultivated types	—
Mentha japonica (Miq.) Makino.	2	Hokkaido and Honshu, Japan	wet places in lowlands to mountains
Mentha laxiflora Benth.	0	Victoria and Southern New South Wales	forests, in gullies, regions above 508 mm rainfall
Mentha longifolia var. asiatica	2	Kazakhstan, Kyrgyzstan, Tajikistan, Turkmenistan, Uzbekistan, Russia, China, Afghanistan, India	wet places in lowlands to mountains
Mentha longifolia subsp. longifolia	18	Europe	alpine meadows stream banks above 305 m
Mentha longifolia subsp. dumortieri	0	Belgium	unknown
Mentha longifolia subsp. lavandulacea	0	Spain	unknown
Mentha longifolia subsp. erminea	0	Crete, Central and Southern Greece, Turkey	marshes
Mentha longifolia subsp. cyprica	0	Cyprus Mountains	moist ground by streams and springs
Mentha longifolia subsp. grisella	0	Hungary, Romania, Macedonia, Greece, Asia minor	unknown
Mentha longifolia subsp. diabolina	0	Eastern Europe, Asia	moist areas
Mentha longifolia subsp. mollis	0	Romania, Slovenia, Croatia, Bosnia-Herzegovina, Montenegro, Serbia, Macedonia	unknown
Mentha longifolia subsp. minutiflora	0	Hungary, Macedonia, Crete	unknown
Mentha longifolia subsp. *typhoides*	1	Aegean, Northwestern Iran, Northern Iraq, Turkey, Syria, Lebanon, Israel, Egypt	marshy fields by streams and rivers
Mentha longifolia subsp. *caucasica*	0	Caucasus	unknown
Mentha longifolia subsp. *calliantha*	0	Northwestern Iran, Eastern Anatolia	marshy ground, streamsides
Mentha longifolia subsp. *noeana*	0	Western Iran, Iraq, Southeast Anatolia	marshy ground, streamsides
Mentha longifolia subsp. *modesta*	0	Asia minor, Iran, Tibet	unknown
Mentha longifolia subsp. *royleana*	0	Asia minor, Iran, Afghanistan, Turkestan, Siberia, Tibet	river planes
Mentha longifolia subsp. *hymalaiensis*	1	Himalayas, Afghanistan	unknown
Mentha longifolia subsp. *syriaca*	0	Syria	wadis
Mentha longifolia subsp. *pellita*	0	Syria, Ethiopia	wadis
Mentha longifolia subsp. *schimperi*	0	Ethiopia, Sinai Penninsula, Yemin	wadis

(Continued)

TABLE 3 | Continued

Species	NCGR-Corvallis Accessions	Range	Habitat
Mentha longifolia subsp. *capensis*	1	South Africa, Namibia, Zambebwe, Lesotho	in water courses and moist places
Mentha longifolia subsp. *polyadenia*	1	South Africa, Lesotho	in water courses and moist places
Mentha longifolia subsp. *wissii*		Namibia, South Africa	in water courses and moist places
Mentha pulegium var. *pulegium*	9	Southern, Western, and Central Europe, north to Ireland and Central Poland and extending to western and southern Ukraine	wet places, roadsides, pond banks
Mentha pulegium var. *micrantha*	0	Southeast Russia, Western Kazakhstan	sinkholes and steppes
Mentha pulegium var. *repens*	0	Tasmania, Australia	moist soil
Mentha requienii Benth	1	Tyrrhenian Isles (Corsica, Sardinia, Montecristo, Caprera	margins of still and running water, turf, and other damp places often in shade
Mentha satureioides R.BR	0	Australia	native
Mentha spp.	6	Australia except in the extreme north	usually loamy soils
Mentha spicata subsp. spicata	96	ancient amphidiploid of M. longifolia x M. suaveolens	streambanks, lake shores, moist fields, roadsides; cultivated
Menhta spicata subsp *crispata*		cultivated	cultivated
Mentha spicata subsp. *condensata*	2	Southern Italy, Sicily, Balkan Penninsula, Agean Region	cultivated
		cultivated	cultivated
Mentha suaveolens subsp. insularis	28	Cultivated (western europe from Denmark to eastern Europe and North Africa and the Canaries	along streambaanks, lakeshores, moist field, and roadsides
Mentha suaveolens subsp. suaveolens	1	western Mediterranean	islands of the western Mediterranean above 305 m
Mentha suaveolens subsp. *timja*	0	Morocco	unknown
Mentha x dalmatica Tausch	5	Germany, Poland, Switzerland	cultivated
Mentha x dumetorum Schult.	2	Europe	cultivated
Mentha x gracilis Sole	22	Europe, United States	cultivated, naturalized
Mentha x piperita L.	48	Africa, Asia, Australasia, Europe, North and South America	cultivated, naturalized
Mentha x smithiana R. A. Graham	3	Europe	native and cultivated
Mentha x suavis Guss	1	Africa	native
Mentha x verticillata L.	3	Europe	native
Mentha x villosa Huds.	6	Europe	native and cultivated
Mentha x villosa nothovar. alopecuroides Hull (Briq.)	10	Europe	native and cultivated
Total	455		

of the most requested cultivars included *M.* × *piperita* 'Chocolate,' 'Pineapple,' 'Black Mitcham,' 'f. *lavanduliodora*,' 'Todd's Mitcham,' *M. aquatica* var. *citrata* 'Eau de Cologne,' *M.* × *gracilis* 'Scotch Spearmint,' and *M. spicata* 'Kentucky Colonel' (**Table 4**).

Many researchers noted that their purpose for requesting the germplasm was to seek diverse essential oil profiles or to perform genetic analyses. Many universities and companies requested propagules from the entire mint collection for this use. Private company results are unpublished and not publically available. Other scientists were interested in the taxonomy of specific accessions and systematic determinations (Tucker and Chambers, 2002). Others requested germplasm to seek disease resistance (Vining et al., 2005).

The policy of the US NPGS is to distribute plant genetic resources freely for crop improvement. The quarantine importation regulations on international shipment of *Mentha* are relatively minor compared to those for shipment of fruit, nut, and other more restricted horticultural crops. Fewer importation requirements facilitate the ease of transfer of germplasm for foreign requests.

GERMPLASM EVALUATION AND CHARACTERIZATION

The primary objectives of a plant genebank includes the acquisition, maintenance, distribution, evaluation, and characterization of the assigned genetic resources (Chambers and Hummer, 1992). During the past 35 years, staff at the NCGR, working in collaboration with scientists at other institutions, have examined the cytology, essential oil content, disease resistance, and the development and use of molecular markers for identity confirmation and diversity determination. Data from these studies have been added as descriptors to the GRIN-Global database.

TABLE 4 | The 50 most distributed *Mentha* accessions from 1980 through 2020 from the USDA ARS National Clonal Germplasm Repository-Corvallis.

PI number	Species	Cultivar or clonal name	Amount orders shipped
557968	*Mentha* x *piperita*	Chocolate Mint	220
557912	*Mentha suaveolens* subsp. *suaveolens*	Pineapple Mint	113
557922	*Mentha* x *gracilis*	*M* x *gracilis* 10306	104
557952	*Mentha* x *piperita*	*M*. x *piperita* f. lavanduliodora	93
557971	*Mentha x piperita*	Black Mitcham Peppermint	91
557781	*Mentha requienii*	*M. requienii* 10001	79
557939	*Mentha* x *piperita*	White Peppermint	79
277803	*Mentha canadensis*	*M. canadensis* Brazil 701	75
557953	*Mentha x piperita*	*M* x *piperita* 10332	70
557993	*Mentha aquatica* var. *citrata*	Eau de Cologne	70
557974	*Mentha* x *piperita*	*M* x *piperita*	67
557973	*Mentha* x *piperita*	Todd Mitcham Peppermint	63
617475	*Mentha japonica*	*M. japonica* 84232	63
557972	*Mentha* x *piperita*	Murray Mitcham Peppermint	61
617482	*Mentha diemenica*	*M. diemenica*	59
557771	*Mentha pulegium*	*M. pulegium* 10006	58
557935	*Mentha* x *gracilis*	Scotch Spearmint	55
557989	*Mentha aquatica* var. *citrata*	*M. aquatica* var. *citrata* 10190	55
557885	*Mentha spicata*	*M. spicata* Oregon	52
557984	*Mentha* x *piperita*	*M*. x *piperita* 10064	52
617476	*Mentha australis*	*M. australis*	52
294099	*Mentha spicata*	*M. spicata* 10085	50
557755	*Mentha longifolia*	*M. longifolia* 10028	49
557758	*Mentha longifolia*	*M. longifolia* 10063	49
557572	*Mentha aquatica*	*M. aquatica* 10176	48
557769	*Mentha longifolia* subsp. *polyadenia*	*M. longifolia* subsp. *polyadenia*	47
557770	*Mentha longifolia* subsp. *typhoides*	*M. longifolia* subsp. *typhoides* 10034	47
637833	*Mentha spicata*	Kentucky Colonel	46
557639	*Mentha gattefossei*	*M. gattefossei* 10012	45
557767	*Mentha longifolia subsp. capensis*	*M. longifolia* subsp. capensis	45
557598	*Mentha canadensis*	*M. canadensis* 10247	44
557955	*Mentha* x *piperita*	Murray Mitcham	43
557999	*Mentha suaveolens*	*M. suaveolens* 10084	43
557937	*Mentha* x *piperita*	Mitcham	42
557954	*Mentha* x *piperita*	Todd's Mitcham	42
557913	*Mentha* x *dalmatica*	*M.* x *dalmatica* 10268 Farm #13	41
557597	*Mentha canadensis*	*M. canadensis* Bachmann #2	38
557997	*Mentha aquatica* var. *citrata*	Orange Mint	38
617495	*Mentha canadensis*	*M. canadensis*	38
557634	*Mentha cervina*	*M. cervina* 10013	37
557768	*Mentha longifolia* subsp. *hymalaiensis*	*M. longifolia* subsp. *hymalaiensis*	37
557891	*Mentha suaveolens* subsp. *suaveolens*	*M. suaveolens* 10014 2n rot.	37
557584	*Mentha arvensis*	*M. arvensis* 10260	35
557985	*Mentha aquatica* var. *citrata*	*M. aquatica* var. *citrata* 10185 2n cit.	35
558001	*Mentha* x *smithiana*	*M*. x *smithiana* 10328 En. #5	35
558004	*Mentha x verticillata*	*M*. x *verticillata* 10325	35
557911	*Mentha suaveolens subsp. insularis*	*M. suaveolens* No. 125	34
557613	*Mentha canadensis*	*M. canadensis* Oregon	33
557918	*Mentha canadensis*	*M. canadensis* 10299	33
617498	*Mentha australis*	*M. australis* 029	33
		Total distributed	2810

CYTOLOGY

Harley and Brighton (1977) defined five sections within the genus *Mentha* (*Mentha* sect. *Audibertia*, sect. *Eriodontes*, sect. *Mentha*, sect. *Preslia*, and sect. *Pulegium*) including 19 species and 13 named hybrids. They also listed chromosome counts for sect. *Mentha* including most of the mint taxa recognized today. Chambers and Hummer (1994) summarized other documentation of chromosome counts and surveyed 73 *Mentha* accessions from the NCGR collection.

Many mint species have a monoploid (base) number of $x = 12$, though diverse species can have monoploid numbers of $x = 9$, $x = 10$, $x = 18$, or $x = 25$. In addition, ploidy series of several species are reported as diploid, tetraploid, hexaploid, octoploid, enneaploid, decaploid, or aneuploid.

Chambers and Hummer (1994) reported that ploidy determinations were obtained by cytological counts of dividing pollen mother cells in flower buds or meristematic cells in root tips. Unusual species counts included *M. requenii* with $2n = 2x =$

18; *M. pulegium* with diploid ($2n = 2x = 20$), triploid ($2n = 3x = 30$), and tetraploid ($2n = 4x = 40$); and *M. japonica* with $2n = 2x = 50$. This survey included chromosome counts for *M. australis* (two accessions), *M. japonica*, *M. diemenica*, and *M. cunninghamii*. For the majority of mint species with $x = 12$, diploid, triploid, tetraploid, hexaploid, octoploid, and decaploid members were observed (Chambers and Hummer, 1994). Vining et al. (2019) recently reported several unusual ploidy counts, finding diploid and triploid clones in *M. suaveolens* and octoploid and enneaploid clones of *M. aquatica* in the NCGR mint collection.

ESSENTIAL OILS

The *Lamiaceae*, and particularly *Mentha*, produce aromatic "essential" oils. These lipophilic substances, predominantly terpenes, are produced by capitate or peltate glandular trichomes that occur on the leaf and stem surfaces (Amelunxen, 1964). They can be simply extracted by crushing the leaves and stems or more completely through distillation. The volatile composition is affected by short chain terpenes that constitute the main fraction especially C10 mono- and C15 sesquiterpenes, which overwhelmingly affect the flavor and taste of these species.

Mints with unique essential oil profiles are valued by the herb, nursery, and pharmaceutical trades. More than 275 accessions of the mint collection of the NCGR were screened by gas chromatography for 37 separate essential oil profiles (Bergstein, 1983). The signature component of native spearmint oil is the C6-oxygenated p-menthane monoterpene (-)-carvone (**Figure 6**). The oil distilled from 'Black Mitcham' peppermint contains predominantly C3-oxygenated monoterpenes, with (-)-menthone and (-)-menthol being the most prominent metabolites. Essential oil analyses indicated that about half of the NCGR accessions labeled as *M. aquatica* accumulate (+)-menthofuran as the most abundant monoterpene (with occasionally high levels of (-)-limonene and/or 1,8-cineole as well), while the remainder had variable oil profiles and need to be further investigated (Vining et al., 2019). Oils from *M. suaveolens* accessions are high in either (-)-carvone, piperitenone oxide, or trans-piperitone oxide (Vining et al., 2019). The oil types of *M. longifolia* accessions are quite diverse, with different C6- or C3-oxygenated monoterpenes dominating the profile (Vining et al., 2005).

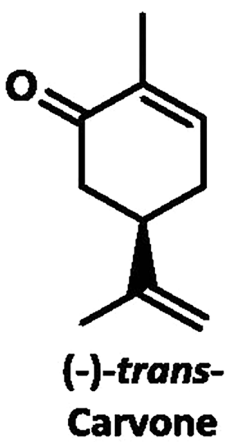


FIGURE 6 | C6-oxygenated p-menthane monoterpene (-)-carvone.

While the oil composition varies significantly depending on growing location, regional weather, harvest dates, and processing technology (Lawrence, 2006), recent assessments of NCGR accessions and cultivars were performed under controlled conditions to ensure comparability. These data sets can be searched online as descriptors through GRIN-Global (https://npgsweb.ars-grin.gov/gringlobal/method.aspx?id=496325).

DISEASE RESISTANCE

Verticillium wilt disease perpetually plagues the mint industry, and therefore breeding for durable wilt resistance is a long-standing goal. Verticillium wilt is a vascular wilt disease caused by the soil-borne fungus *Verticillium dahliae*. 'Black Mitcham' peppermint is highly wilt susceptible. However, the cultivar is also a sterile hexaploid, which hinders traditional breeding techniques. Other genetic mutation and recombination techniques have been performed over the years.

An irradiation program, working in collaboration with nuclear facilities in the 1960s, resulted in the release of two new peppermint cultivars with relatively higher wilt resistance: 'Todd's Mitcham' (Todd et al., 1977) and 'Murray Mitcham' (Murray and Todd, 1972). In the 1990s–2000s, the Mint Industry Research Council initiated a genetic engineering project with the objective of genetically transforming peppermint cultivars to confer verticillium wilt resistance. However, despite remarkable technical successes in improving oil yield and composition (but not verticillium resistance) (Lange et al., 2011; Lange, 2015), this research was discontinued due to concerns about market acceptance of genetically modified organisms.

MOLECULAR ANALYSIS

Multi-locus dominant markers such as amplified fragment length polymorphism (AFLP), random amplified polymorphic DNA (RAPD), and intersimple sequence repeat (ISSR) markers have been used with a small number of mint accessions for assessment of genetic diversity and relationships (Khanuja et al., 2000; Gobert et al., 2002; Shasany et al., 2002; Shiran et al., 2004; Vining et al., 2005; Sabboura et al., 2016; Jedrzejczyk and Rewers, 2018; Choupani et al., 2019) and for cultivar identification (Fenwick and Ward, 2001). Species-specific start codon targeted (SCoT) markers were also developed and used to assess genetic diversity in 12 accessions from four mint species (Khanm et al., 2017). Few co-dominant simple sequence repeat

(SSR) markers are available for mint (Kumar et al., 2015; Vining et al., 2019). Kumar et al. (2015) identified 54 SSRs from publicly available expressed sequence tag (EST-SSR) sequences of *M.* ×*piperita* that generated clear amplicons in 13 accessions from this species. Three SSRs with long core repeats amplified in representative samples from four species including *M. arvensis*, *M. citrata*, *M. longifolia*, and *M. spicata*, indicating transferability among species.

We recently screened these three SSRs in addition to 48 other primer pairs designed from a draft genome assembly of *M. longifolia* CMEN 585, from South Africa (Vining et al., 2017) in an eight-member testing panel (three accessions each of *M. suaveolens* and *M. aquatica* and two accessions of *M. longifolia*) (Vining et al., 2019). Screening the eight-member testing panel with these 51 SSRs identified nine new primer pairs that were polymorphic and appeared easy to score (**Figure 7**). These nine SSRs were thus used in two multiplexes to genotype 49 accessions propagated from the NCGR collection that included 24 accessions of *M. aquatica*, 23 accessions of *M. suaveolens*, and the two *M. longifolia* plants in the testing panel. The nine SSRs developed in this study separated the accessions mostly according to species and identified each accession as unique. Three sets of accessions of *M. aquatica* were closely related and were distinguished by a single allele each: CMEN 116, CMEN 117; CMEN 110, CMEN 111; and CMEN 121, CMEN 122. The consecutive local numbers of each pair of closely related accessions indicate they are in neighboring pots and may suggest plant contamination and will be investigated further.

Single nucleotide polymorphism (SNP) markers have not yet been reported in *Mentha*, and the number of SSRs remains low. More SSRs and new SNPs need to be identified and evaluated across *Mentha* species and are now possible due to the availability of a genome sequence (*M. longifolia*) (Vining et al., 2017), and the low cost of re-sequencing.

USES OF THE COLLECTION IN BREEDING EFFORTS

From the mint industry viewpoint, improved oil quality and verticillium wilt disease resistance are two priority traits for new cultivar development. Both traits are multigenic with complex mechanisms. *Mentha* germplasm evaluations have therefore focused on those traits and efforts are ongoing to understand the underlying genetics. Vining et al. (2019) also reported species accessions with undeveloped anthers, resulting in apparent male sterility. This trait could be useful for developing male-sterile cultivars. The USDA NCGR *Mentha* collection is valuable as a source of genetic diversity for mint breeding toward these goals. Several private companies have had, or currently have, mint breeding programs that have requested accessions from the *Mentha* collection. While information about these programs is strictly proprietary, patent documents show cultivar releases for *M. spicata:* (https://patents.google.com/patent/USPP8645P/en; https://patents.google.com/patent/US20120204283A1/en; https://patents.google.com/patent/US20050044600), *M. arvensis* (https://patents.justia.com/patent/PP10935), and an interspecies hybrid: (https://patents.google.com/patent/USPP12030). Besides patents, future cultivars could be released under Plant Variety Protection (PVP), or as germplasm releases through the USDA.

The origin of 'Black Mitcham' peppermint, the most widely grown American cultivar in the US and Europe, is unknown, because it resulted from natural hybridization events. However, phylogenetic studies (Tucker et al., 1980; Tucker and Naczi, 2007) suggest that genotypes of *M. longifolia* and *M. suaveolens* are the diploid ancestors of *M. spicata*, which then further hybridized with *M. aquatica* to give rise to *M.* ×*piperita*. Therefore, efforts over the past 50 years have focused on evaluation and crossing of USDA NCGR accessions of *M. longifolia*, *M. suaveolens*, and *M. aquatica*. *M. longifolia* showed the most obvious phenotypic diversity, with variation in growth habit, leaf morphology, and oil type (Vining et al., 2005). *M. suaveolens* and *M. aquatica* showed relatively lower morphological and oil type diversity (Vining et al., 2019).

In *V. dahliae* inoculation tests, both *M. longifolia* and *M. aquatica* showed a range of wilt disease resistance to susceptibility, with a few accessions showing high susceptibility, a few showing high resistance, and most displaying mild to moderate susceptibility levels (Vining et al., 2005; Vining et al., 2019). *M. suaveolens* accessions, in contrast, were highly resistant, with the exception of one moderately susceptible accession (Vining et al., 2019). The contrast in overall wilt disease resistance between *M. suaveolens* and the other two species led to speculation that verticillium wilt disease resistance was genetically linked to a spearmint oil type. However, one carvone-type *M. longifolia* accession, CMEN 584, was among the most highly wilt-susceptible accessions of that species. Segregating F_1 and F_2 populations derived from crossing CMEN 584 with wilt-resistant *M. longifolia* accession CMEN 585 also showed a range of wilt resistance to susceptibility, with wilt susceptibility not associated with any particular oil phenotype (Hummer, pers. comm.). It is possible that the higher wilt resistance in *M. suaveolens* results from homozygosity at particular loci or even a different genetic resistance from the other two species.

Transgressive segregation, a phenomenon in which progeny traits differ from those of either parent, is well documented in oil types resulting from *Mentha* crosses (Tucker, 2012). Since both monoterpene profile and verticillium wilt resistance are polygenic traits, transgressive segregation can be expected for verticillium wilt resistance as well. With the advent of the *M. longifolia* reference genome, the genetic underpinnings of these traits are now actively being studied. The complex regulatory mechanisms governing both and verticillium wilt resistance will be a challenge to breeding

In conclusion, mint has a long history of human use and influence and relatively recent (within the past century) increasing attention by mostly private breeding programs. With the advent of genome sequencing, efforts are underway to develop molecular markers for alleles of monoterpene biosynthesis genes and for verticillium wilt disease resistance

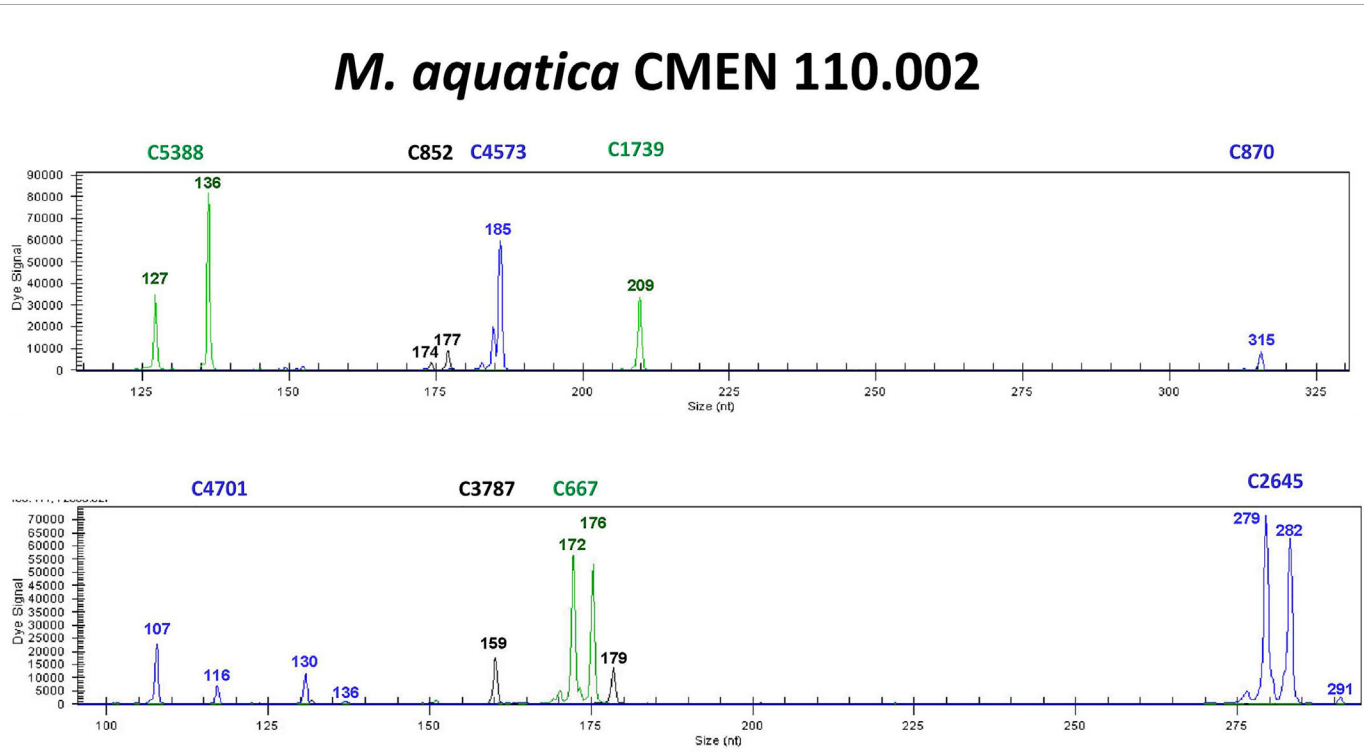


FIGURE 7 | Example of SSR genotyping with *M. aquatica*.

genes. Going forward, marker assisted selection will increase in importance to mint breeding programs.

AUTHOR CONTRIBUTIONS

All authors contributed equally to this manuscript.

FUNDING

Funding for the collection, maintenance, distribution, and evaluation of the NCGR Mentha collection was provided by USDA ARS PWA CRIS 2072-21000-049-00D.

ACKNOWLEDGMENTS

We appreciate the taxonomic assistance of Henrietta Chambers and Arthur Tucker in species determinations of the NCGR *Mentha* collections. We appreciate the technical assistance of James Oliphant, Jeanine DeNoma, Debra Hawkes, Melissa Fix, and Sunny Green.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.01217/full#supplementary-material

REFERENCES

Aboelsoud, N. H. (2010). Herbal medicine in ancient Egypt. *J. Med. Plant Res.* 4 (2), 082–086. doi: 10.5897/JMPR09.013

Adams, M., Alther, W., Kessler, M., Kluge, M., and Hamburger, M. (2011). Malaria in the renaissance: Remedies from European herbals from the 16th and 17th century. *J. Ethnopharmacol.* 133, 278–288. doi: 10.1016/j.jep.2010.10.060

Amelunxen, F. (1964). Elektronenmikroskopische Untersuchungen an den Drüsenhaaren von *Mentha piperita* L. *Planta Med.* 12, 121–139. doi: 10.1055/s-0028-1100158

Beck, L. Y. (2005). Pedanius Dioscorides of Anazarbus: De Materia Medica. (Olms-Weidmann, Hildensheim, Germany), 192–197. (ed. Trans.).

Bergstein, D. (1983). *Oil Analysis of Mentha Accessions* (Corvallis, Oregon: USDA NCGR Station publication). 6 volumes of gas chromatic analysis and tabled data.

Bostok, J. (1885). *The Natural History. Pliny the Elder*. Ed. H. T. Riley (London: Taylor and Francis, Red Lion Court, Fleet Street).

Brunfels, O. (1532). *Herbarum vivae eicones* (Mainz, Germany).

Burkhard, P. R., Burkhardt, K., Haenggeli, C. A., and Landis, T. (1999). Plant-induced seizures: reappearance of an old problem. *J. Neurol.* 246 (8), 667–670. doi: 10.1007/s004150050429

Chambers, H. L., and Hummer, K. E. (1992). Clonal repository houses valuable mint collection in Corvallis, Oregon. *Diversity* 8 (4), 31–32.

Chambers, H. L., and Hummer, K. E. (1994). Chromosome counts in the *Mentha* collection at the USDA-ARS National Clonal Germplasm Repository. *Taxon* 43 (3), 423–432. doi: 10.2307/1222717

Choupani, A., Shojaeiyan, A., and Maleki, M. (2019). Genetic relationships of Iranian endemic mint species, *Mentha mozaffariani* Jamzad and some other mint species revealed by ISSR markers. *BioTechnologia* 100 (1), 19–28. doi: 10.5114/bta.2019.83208

Collins, M. (2000). *Medieval Herbals: The Illustrative Traditions* (London, UK: The British Library), 39–105.

Culpeper, N. (1652). *The English Physitian: Or an Astrologo-Physical Discourse of the Bulgar Herbs of This Nation* (Cornhill, UK: Peter Cole Printing Press).

Darlington, C. D. (1963). *Chromosome Botany, and the Origin of Cultivated Plants. 2nd ed*. Eds. Allen, and Unwin, (London, UK: George Allen & Unwin,Ltd.).

Darwin, C. (1859). *The Origin of Species*. Ed. J. Murray (London, UK).

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

Duke, J. (1985). *CRC Handbook of Medicinal Herbs* (Boco Raton, Florida: CRC Press LLC).

Elansary, E. O., and Ashmawy, N. A. (2013). Essential oils of mint between benefits and hazards. *J. Essential Oil Bear. Plants* 16 (4), 429–438. doi: 10.1080/0972060X.2013.813279

Elmore, F. H. (1944). *Ethnobotany of the Navajo* (Sante Fe, New Mexico: School of American Research), p.73.

FAOSTAT (2018). *Food and Agriculture Organization of the United Nations, Statistical database*. Available at: http://www.fao.org/faostat/en/#data/QC (Accessed October 2, 2018).

Fenwick, A. L., and Ward, S. M. (2001). Use of random amplified polymorphic DNA markers for cultivar identification in mint. *HortScience* 36 (4), 761–764. doi: 10.21273/HORTSCI.36.4.761

Flueckiger, F. A. (1891). *Pharmakognosie des pflenzenreischs, 3rd ed*. 723.

Fuchs, L. (1543). *De historia stirpium commentarii insignes* (Basel, Switerland).

GBIF (2020). *Global Biodiversity Information Facility*. Available at: https://www.gbif.org/what-is-gbif (Accessed 17 January, 2020).

Gildemeister, E., and Hoffman, F. (1913). *The Volatile Oils* Vol. Vol 1 (Hoboken, New Jersey: J. Wiley and Sons), 191–192.

Gobert, V., Moja, S., Colson, M., and Taberlet, P. (2002). Hybridization in the section *Mentha* (Lamiaceae) inferred from AFLP markers. *Am. J. Bot.* 89 (12), 2017–2023. doi: 10.3732/ajb.89.12.2017

GRIN-Global (2020). *Germplasm Resources Information Network-Global, US National Plant Germplasm System*. Available at: https://npgsweb.ars-grin.gov/gringlobal/search.aspx (Accessed April 27, 2020).

R. T. Gunther (Ed.) (1934). *The Greek Herbal of Dioscorides (1655/1933) translated by John Goodyer* (Oxford, UK: Hafner, New York), 274–279. (ed.)Preface and.

Gunther, E. (1973). *Ethnobotany of Western Washington. The Knowledge and Use of Native Plants by Native Americans* (Seattle, WA: University of Washington Press), p.45. Revised ed.

Hammer, K. (1984). Das Domestikationssyndrom. *Kulturpflanze* 32, 11–34. doi: 10.1007/BF02098682

Harlan, J. R. (1992). *Crops and Man. 2nd ed* (Madison, Wisconsin: American Society of Agronomy and Crop Science Society of America).

Harley, R. M., and Brighton, C. A. (1977). Chromosome numbers in the genus *Mentha* L. *Bot. J. Linn. Soc.* 74, 71–96. doi: 10.1111/j.1095-8339.1977.tb01168.x

Hayden, A. H. (1929). *An Investigation of Oregon Grown Peppermint. Thesis* (Corvallis, Oregon: Oregon State Agricultural College).

Hummer, K. E., and Janick, J. (2007). *Rubus* iconography: Antiquity to the Renaissance. *Acta Hortic.* 859, 89–105. doi: 10.17660/ActaHortic.2007.759.6

Hummer, K. E., Bassil, N. V., and Vining, K. (2020). "Mentha L. and Pycnanthemum L. germplasm at the US National Clonal Germplasm Repository in Corvallis, Oregon," in *Medicinal and Aromatic Plants of North America*. Ed. A. Mathe (Springer). Available at: http://www.springer.com/series/11192

Iranian encyclopedia online (2020). Available at: http://www.iranicaonline.org/ (Accessed May 9, 2020).

Jamzad, Z., Ingrouille, M., and Simmonds, M. S. (2003). Three new species of *Nepeta* (*Lamiaceae*) from Iran. *Taxon* 52, 93–98. doi: 10.2307/3647304

Janick, J., and Hummer, K. E. (2012). The 1500th Anniversary (512–2012) of the Juliana Anicia Codex: an illustrated Dioscoridean recension. *Chronica Hortic.* 52 (3), 9–15.

Jedrzejczyk, I., and Rewers, M. (2018). Genome size and ISSR markers for *Mentha* L. (Lamiaceae) genetic diversity assessment and species identification. *Ind. Crops Prod.* 120, 171–179. doi: 10.1016/j.indcrop.2018.04.062

Kew Science (2020). Available at: http://www.plantsoftheworldonline.org/taxon/urn:lsid:ipni.org:names:30016176-2 (Accessed 29 April, 2020).
Khanm, N., Singh, S., and Dhawan, S. S. (2017). Development of species specific SCoT markers and analysis of genetic diversity among *Mentha* genotypes. *Int. J. Innovative Sci. Eng. Technol.* 4 (2), 244–264.
Khanuja, S. P. S., Shasany, A. K., Srivastava, A., and Kumar, S. (2000). Assessment of genetic relationships in *Mentha* species. *Euphytica* 111 (2), 121–125. doi: 10.1023/a:1003829512956
Kokkini, S. (1991). Essential oils as taxonomic markers in *Mentha*. *Proceedings of the Advances in Labiate Science*. Kew Gardens. London. April 2-5.
Kumar, B., Kumar, U., and Yadav, H. K. (2015). Identification of EST–SSRs and molecular diversity analysis in *Mentha piperita*. *Crop J.* 3, 335–342. doi: 10.1016/j.cj.2015.p02.002
Lange, B. M., Mahmoud, S. S., Wildung, M. R., Turner, G. W., Davis, E. M., Lange, I., et al. (2011). Improving peppermint essential oil yield and composition by metabolic engineering. *Proc. Natl. Acad. Sci. U. States America* 108, 16944–16949. doi: 10.1073/pnas.1111558108
Lange, B. M. (2015). Biosynthesis and biotechnology of *p*-menthane monoterpenes, including menthol, carvone and limonene. *Adv. Biochem. Eng. Biotechnol.* 148, 319–353. doi: 10.1007/10_2014_289
Lawrence, B. M. (1991). The spearmint and peppermint industry of North America pp 59–pp 90. *Proceedings of the 3rd International conference on aromatic and medicinal plants*. (Lyons, France).
Lawrence, B. M. (2006). *Mint – The genus Mentha*. Ed. B. M. Lawrence (Boca Raton, FL, USA: CRC Press). ed598 pages. doi: 10.1201/9780849307980
Liu, W., and Lawrence, B. M. (2006). "Chapter 5, "Production of mints and mint oil in China," in *Mint – The genus Mentha*. Ed. B. M. Lawrence (CRC Press), 165–184. Ed.
Murray, M. J., and Todd, W. A. (1972). Registration of Todd's Mitcham peppermint. *Crop Sci.* 12, 128. doi: 10.2135/cropsci1972.0011183X001200010056x
Naghibi, F., Mosaddegh, M., Mohammadi, M. M., and Ghorbani, A. (2010). Labiatae family in folk medicine in Iran: From ethnobotany to pharmacology. *Iranian J. Pharm. Res.* 4, 63–79. doi: 10.22037/IJPR.2010.619
NASS (2020). *US National Agricultural Statistical Service*. Available at: https://www.nass.usda.gov/Statistics_by_Subject/index.php?sector=CROPS (Accessed May 9, 2020).
Nuńez, D., and De Castro, C. O. (1992). "Palaeoethnobotany and archaeobotany of the Labiatae in Europe and the near East," in *Advances in Labiate Science* (London, UK: Royal Botanic Gardens, Kew), 437.
Ovid, Metamorphoses 10. 728 ff (trans. Melville) (Roman epic C1st B.C. @ to C1st A.D.).
Parry, E. J. (1925). *"Parry"s Cyclopedia of Perfumery, a Handbook"* (Philadelphia: P. Blakiston's and Sons), 551.
Riddle, J., and Estes, J. (1992). Oral contraceptives in ancient and medieval times. *Am. Sci.* 80 (3), 226–233.
Rosengarten, F. (1969). *The Book of Spices* (New York: Jove Publ., Inc.).
Sabboura, D., Yacouyb, R., and Lawand, S. (2016). Assessment of genetic relationships among mint species. *Int. J. ChemTech. Res.* 9 (3), 462–468.
Shasany, A. K., Srivastava, A., Bahl, J. R., Sharma, S., Kumar, S., and Khanuja, S. P. S. (2002). Genetic diversity assessment of *Mentha spicata* L. germplasm through RAPD analysis. *Plant Genet. Resour. Newslett.* 130, 1–5. doi: 10.1093/jhered/esi091
Shiran, B., Momeni, S., and Razmjoo, K. (2004). "Assessment of genetic diversity among Iranian mints using RAPD markers," in *Genetic Variation for Plant Breeding, Proceedings of the 17th Eucarpia General Congress*. (Tulin Austria), 121–125.
Singer, C. (1927). The herbal in antiquity and its transmission to later ages. *J. Hellenistic Stud.* 47, 1–52. doi: 10.2307/625251
Sonmezdag, A. S., Kelebek, H., and Selli, S. (2017). Identification of aroma compounds of Lamiaceae species in Turkey using the purge and trap technique. *Foods* 6 (2), 10. doi: 10.3390/foods6020010
Todd, W. A., Green, R. J., and Horner, C. E. (1977). Registration of Murray Mitcham Peppermint1 (Reg. No. 2). *Crop Sci.* 17, 188–188. doi: 10.2135/cropsci1977.0011183X001700010054x
Tucker, A. O., and Chambers, H. L. (2002). *Mentha canadensis* L. (Lamiaceae) a relict amphidiploid from the lower Tertiary. *Taxon* 51, 703–718. doi: 10.2307/3647334
Tucker, A. O., and Naczi, R. F. C. (2007). "Mentha: an overview of its classification and relationships," in *Mint: the genus Mentha, Chapter 1*. Ed. B. Lawrence (Boco Raton, FL.: CRC Press), 1–40. ed.
Tucker, A. O., Harley, R. M., and Fairbrothers, D. E. (1980). The Linnean types of *Mentha* (Lamiaceae). *Taxon.* 29, 233–255. doi: 10.2307/1220285
Tucker, A. O. (2012). Genetics and breeding of the genus *Mentha*: a model for other polyploid species with secondary constituents. *J. Med. Active Plants* 1 (1), 19–29. doi: 10.7275/R54B2Z7Q
Vestal, P. A., and Schultes, R. E. (1939). *The Economic Botany of the Kiowa Indians* (Cambridge MA: Botanical Museum of Harvard University).
Vining, K. J., Zhang, Q., Tucker, A. O., Smith, C., and Davis, T. M. (2005). *Mentha longifolia* L.) L.: A model species for mint genetic research. *HortScience* 40, 1225–1229. doi: 10.21273/HORTSCI.40.5.1225
Vining, K. J., Johnson, S. R., Ahkami, A., Lange, I., Parrish, A. N., Trapp, S. C., et al. (2017). Draft genome sequence of *Mentha longifolia* and development of resources for mint cultivar improvement. *MolecularPlant* 10, 323–339. doi: 10.1016/j.molp.2016.10.018
Vining, K. J., Pandelova, I., Hummer, K. E., Bassil, N. V., Contreras, R., Neill, K., et al. (2019). Genetic diversity survey of *Mentha aquatica* L. and *Mentha suaveolens* Ehrh., mint crop ancestors. *Genet. Resour. Crop Evol.* 66, 825–845. doi: 10.1007/s10722-019-00750-4
Yasudori, T. (984). 醫心方, Ishinpō Tokyo, Japan.
Zohary, D. (1969). "The progenitors of wheat and barley in relation to domestication and agricultural dispersal in the Old World," in *The Domestication and Exploitation of Plants and Animals*. Eds. P. J. Ucko and G. W. Dimbleby (London, UK.: Duckworth), 47–66. eds.
Zohary, D. (1984). "Modes of evolution under domestication," in *Plant Biosystematics*. Ed. W. Grant (Toronto, Canada.: Academic Press), 579–596. ed.
Zohary, D. (2004). Unconscious selection and the evolution of domestic plants. *Econom. Bot.* 58 (1), 5–10. doi: 10.1663/0013-0001(2004)058[0005:USATEO]2.0.CO;2

Conflict of Interest: 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.

The handling editor declared past co-authorship with several of the authors [NB, KV].

Copyright © 2020 Vining, Hummer, Bassil, Lange, Khoury and Carver. 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) and the copyright owner(s) 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

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership