



OPEN ACCESS

EDITED BY

Quan Zou,
University of Electronic Science and
Technology of China, China

REVIEWED BY

Wenzheng Bao,
Xuzhou University of Technology, China
Xumin Ni,
Beijing Jiaotong University, China

*CORRESPONDENCE

Bo-Wen Zhang,
✉ zhangbw@bnu.edu.cn

[†]These authors have contributed equally to
this work

RECEIVED 15 October 2025

REVISED 25 November 2025

ACCEPTED 30 November 2025

PUBLISHED 17 December 2025

CITATION

Li M-J, Li X-X, Xu L-L and Zhang B-W (2025)
Variance of K_s distribution corrects the bias in
the divergence caused by the ancestral
population size.
Front. Genet. 16:1725551.
doi: 10.3389/fgene.2025.1725551

COPYRIGHT

© 2025 Li, Li, Xu and Zhang. 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.

Variance of K_s distribution corrects the bias in the divergence caused by the ancestral population size

Mi-Jia Li^{1†}, Xiao-Xue Li^{1†}, Lin-Lin Xu² and Bo-Wen Zhang^{1*}

¹Ministry of Education Key Laboratory for Biodiversity Science and Ecological Engineering, College of Life Sciences, Beijing Normal University, Beijing, China, ²Department of Microbiome Dynamics, Leibniz Institute for Natural Product Research and Infection Biology (Leibniz-HKI), Jena, Germany

K_s distribution, the distribution of the synonymous substitutions, has been widely used to estimate the species divergence using orthologous genes. However, conventional approaches often ignore the underlying bias that species divergence is delayed to average gene divergence by $2N_e$ generations, where N_e represents the ancestral effective population size, due to the lack of scalable methods for N_e inference. Here, we demonstrate through simulations that K_s distribution variance correlates with N_e , enabling direct estimation of ancestral population parameters from standard K_s data. Leveraging this relationship, we present *Tspecies*, a framework that corrects divergence time estimates using only substitution rates and K_s distributions, without requiring additional genomic data. Our practical application of *Tspecies* in *Liriodendron* has inferred a divergence time between North American and East Asian lineages (1.44 Ma) that align with early Pleistocene glaciation, and a large ancestral N_e ($\sim 5.29 \times 10^4$) consistent with fossil evidence. Our finding reveals the correlation between the variance of K_s distribution and N_e , and develops a computational framework to resolve the bias in K_s based dating by incorporating a readily estimated N_e .

KEYWORDS

K_s distribution, effective population size, species divergence, coalescent model, gene divergence, orthologous gene

Introduction

K_s distribution, the distribution of the synonymous substitutions of orthologs, has been widely employed to estimate the species divergence times in comparative genomics. When studying the characteristics of some specific genes in *Gossypium hirsutum* (Hao et al., 2020), *Raphanus sativus* (Hu et al., 2018), *Brassica napus* (Zhu et al., 2020), researchers have successfully employed the K_s distributions to depict the divergence times of the objects from their close relatives. The K_s value reflects the sum of independent evolutionary distances accumulated in two species following their divergence. Building on this, under the assumption of constant mutation accumulation rates in both species, the divergence time T can be calculated by dividing the $K_s/2$ value (representing the distance from either species to their common ancestor) by the substitution rate (μ). This K_s distribution-based method for estimating species divergence times has been widely adopted in comparative genomic studies and can be easily obtained with the commonly used genomic analysis toolkits such as OrthoFinder (Emms and Kelly, 2019) and KaKs_Calculator (Zhang, 2022).

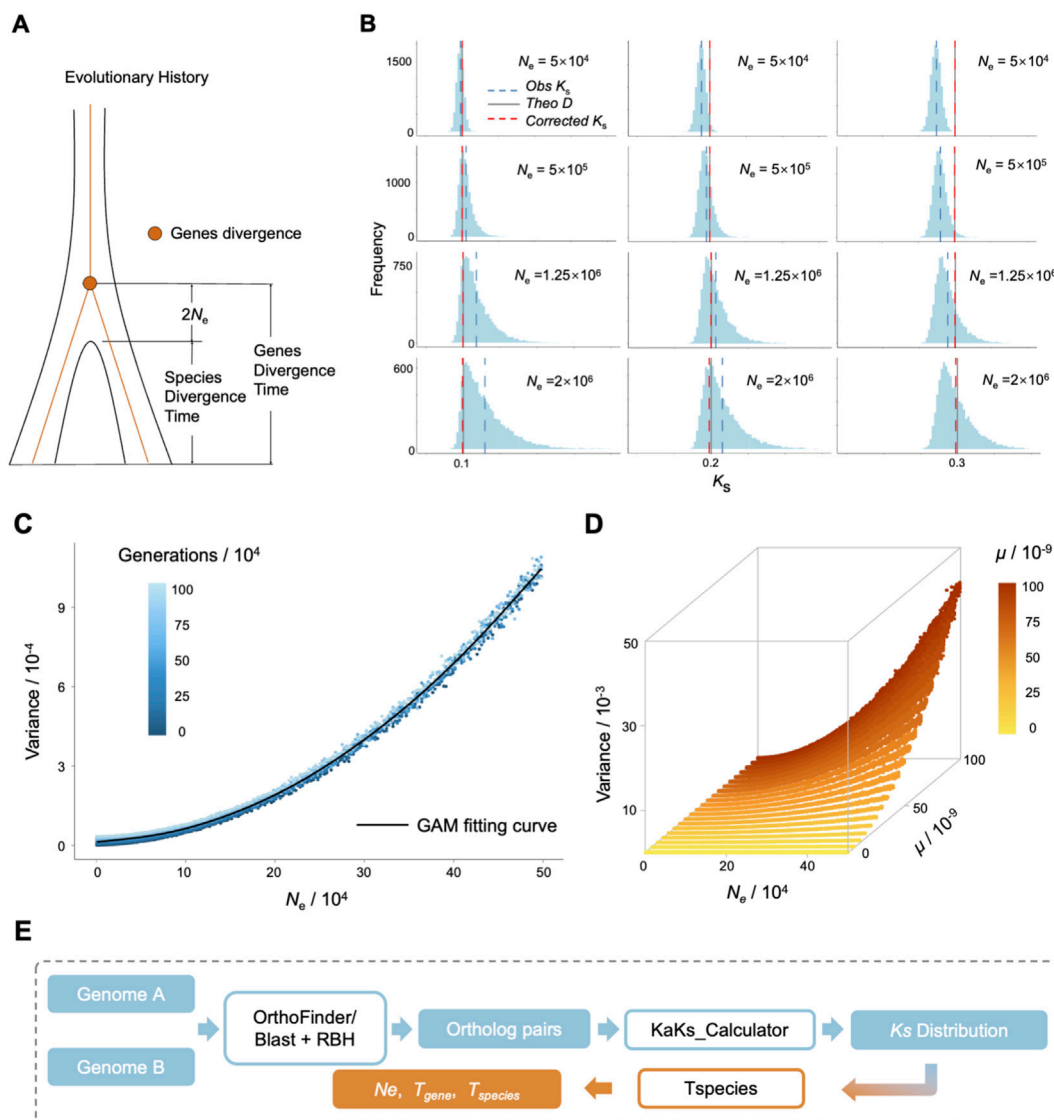


FIGURE 1

The optimization for the distribution of the synonymous substitutions (K_s). (A) Schematic diagram of species divergence time and the average divergence time of genes. (B) Observed, theoretical, and corrected K_s in the K_s distribution of different simulated scenarios. The Obs K_s refers to the K_s directly calculated by the sequential differences; The Theo D refers to the theoretical species distance, which equals to twice μ multiplied by the simulated species divergence time; The Corrected K_s refers to the estimated species divergent distance corrected with $2N_e$ and multiple substitution. (C) The Generalized Additive Model (GAM) describing the relationship between N_e and variance of K_s distribution. (D) 3D Plot of Variance of K_s distribution vs. N_e and μ . (E) Overview of the workflow utilized by *Tspecies*. Orthologous pairs were inferred through OrthoFinder or blast with reciprocal best hit (blast + RBH) before calculating K_s distribution.

However, under the coalescent theory framework, when post-divergence gene flow is not considered, the genetic divergence time between orthologs always predates the species divergence time (Sigwart, 2009). Therefore, accurate estimation of species divergence time requires calculating the difference dT between these two temporal scales. And for dT , according to the coalescent theory, the species divergence time differs from the average gene divergence time by $2N_e$ (Figure 1A), where N_e represents the ancestral effective population size (Wakeley and Sargsyan, 2009). Current methods for inferring N_e primarily rely on population genomic data to reconstruct historical demographic dynamics (Charlesworth, 2009). However, due to limitations in sampling or sequencing, most studies lack sufficient population-level genomic data to estimate N_e (Wang et al., 2016). To

address these constraints, alternative approaches that can estimate both species divergence and N_e using single or few genomes have thus been developed, such as the Bayesian Phylogenetics & Phylogeography (BPP) program (Yang, 2015) and the F1-hybrid Pairwise Sequentially Markovian Coalescent (hPSCM) model (Cahill et al., 2016). While BPP has been widely applied and provides powerful inference under the multispecies coalescent framework, its performance can be limited in practice. For instance, it often requires high-quality genomic data (e.g., haplotype-resolved genomes) and substantial computational resources. Complex evolutionary scenarios such as polyploidy (Yan et al., 2022) or ghost introgression (Pang and Zhang, 2024) can further complicate parameter estimation.

In practice, comparative genomics studies often used this K_s distribution-based method, instead of utilizing the coalescent theory framework, for estimating species divergence times in comparison with whole genome duplication or other evolution events. Therefore, this N_e -dependent bias is often overlooked. Directly scaling N_e through K_s distributions could thereby offer a more feasible solution in such cases. Of note, the practical application of our tool on the genomes from genus *Liriodendron* indicated a divergence time of ~1.44 million years ago (Ma) between the North American and East Asian lineages, which gives a more plausible scenario of allopatric speciation driven by glacial divergence.

Methods

Simulating the divergence

To model the divergence of the two closely related species, we employed the *ms* (Hudson, 2002) and the *seq-gen* (Rambaut and Grass, 1997). Given that the samples were generated under neutral model, the rate of nucleotide substitutions rate between sequences was equivalent to the rate of synonymous substitutions. In each simulation, 10,000 pairs of homologous sequences of 1,000 bp in length were generated, and the mean and variance of the K_s distribution were calculated using R.3.6.0. To account for the range of divergence scenarios, the substitution rate (μ) was grouped μ into 23 discrete categories spanning 1×10^{-10} to 1×10^{-7} per site per generation, a range that encompasses empirically reported neutral mutation rates across eukaryotes (Wang and Obbard, 2023). The theoretical divergence time (T) was set to 10^4 to 10^6 generations, and N_e was set to 500 to 500,000. These values are covered by θ and t in *ms*. In this case, θ is defined as $4N_e\mu$, and t is the theoretical divergence time divided by four times the effective population size.

Modeling the N_e effects

We first derived the theoretical relationship between N_e and the variance of synonymous divergence under the standard coalescent model, yielding the analytical estimator:

$$N_e = \sqrt{\frac{\text{var}(K_s)}{16 \times \mu^2}}$$

Where var is the variance of the K_s distribution, and the μ is substitution rate in the unit of per site per generation. This approximation is accurate only when N_e is sufficiently large, because its validity requires the coalescent variance to dominate the Poisson variance of mutation counts. Based on the empirical evaluation from simulation mentioned above, this condition holds reliably when $N_e > 450,000$. In practice, *Tspecies* switches to the analytical coalescent estimate when the predicted value exceeds this threshold.

For smaller effective population sizes, the relationship between N_e and the variance of K_s is strongly nonlinear and lacks a convenient analytical approximation. In this regime, we

approximate the N_e - K_s variance relationship using generalized additive models (GAMs). For each mutation rate μ , we generated training data with *ms* and *seq-gen*, recording the true N_e used in the simulations and the resulting variance of K_s across 10,000 simulated loci. We then fitted a GAM of the form $N_e \sim s(\text{Var}(K_s))$, using *gam* () function in *mgcv* (Wood, 2017) with default smoothing settings. Mutation rate μ was treated as a categorical factor, and μ -specific models (and their prediction grids) are included in the *Tspecies* package. During the inference, *Tspecies* uses the GAM-based predictions whenever the estimated N_e is below 450,000, interpolating N_e from the empirical $\text{Var}(K_s)$ via the pre-fitted model for the corresponding μ .

Robustness on N_e - K_s variance relationship

To evaluate the impact of sequential length on the N_e - K_s variance relationship, we compared model predictions across $L = 500$ bp, 1,000 bp, 1,500 bp and 2,000 bp while holding other parameters constant ($\mu = 10^{-8}$ per site per generation). For each L , we generated orthologous sequence pairs using *ms* and calculated the normalized variance of K_s distributions (Supplementary Figure S1).

In order to ascertain whether the predicted outcomes of the *Tspecies* undergo substantial alteration when N_e fluctuates, a simulation was conducted in *ms*: subsequent to species divergence, the number of one of the subpopulations diminished precipitously to one-half, one-fifth, and one-tenth of the ancestral population, respectively (Supplementary Figure S2). *Tspecies* was utilized to calculate the species divergence times, and the predicted outcomes were compared with the set theoretical times in simulation. Then, the relative accuracy (RA) was calculated by the following formula:

$$RA = \frac{|T - T_{true}|}{T_{true}}$$

where T is species divergence time calculated by *Tspecies*, T_{true} is theoretical species divergence time. The relative accuracy under different μ are shown in Supplementary Figure S3.

Scale generation times to real times

In *Tspecies*, the mutation rate μ is specified per site per generation. When the input mutation rate is provided per site per year, it is internally converted to a per-generation rate by multiplying by the generation time g (in the unit of years per generation). Similarly, *Tspecies* outputs the divergence time in units of generations; when a generation time g is supplied, these values are converted to years by multiplying the estimated number of generations by g .

Application in *liriodendron*

To test the model, we applied *Tspecies* to the genomes of genus *Liriodendron* with two distinct species from East Asian (*L. chinense*) and eastern North American (*L. tulipifera*). Genomes were

downloaded from the database PRJNA418360 of NCBI. The K_s distribution of reciprocal best-hit gene pairs across the genomes was calculated using the K_s analysis pipeline implemented in the wgd package (Zwaenepoel and Van de Peer, 2019). The obtained divergent distance is scaled with the reported synonymous substitution rate of 3.02×10^{-9} per site per year (Cui et al., 2006). The divergence time directly calculated by mean of the K_s distribution was compared to the time estimated under *Tspecies*. The $\delta^{18}\text{O}$ data from the Lisiecki and Raymo (2005) are converted to direct temperature estimates using the equations of Hansen (Hansen et al., 2013).

Coalescent-based bayesian analyses

The bpp v4.1.4 (Yang, 2015) was used to calculate the divergence time of East Asian (*L. chinense*) and eastern North American (*L. tulipifera*) under coalescent-based Bayesian analyses. We used A00 model with fixed species branching order to estimate the parameters. The alignments of 1,000 loci with 500 bp length were used, with MCMC chain length of 1,000,000 and the first 100,000 discarded as burn-in. Tau (for divergence time) and Theta (for effective population sizes) parameters were estimated with substitution rate of 3.02×10^{-9} per site per year and generation time of 30 years. The ESS for each parameter was confirmed to be larger than 200 in MCMC trace files to guarantee convergence.

Results

The core concept underlying this optimization is rooted in the coalescent theory. Because the K_s distribution reflects divergence among orthologous loci from two sister species, these loci coalesce exclusively within the ancestral population. Consequently, the shape of the K_s distribution should conform to the coalescent model of the ancestral lineage, with N_e representing the ancestral population size. Under this framework, the expected coalescent time for two alleles from a random locus is $2N_e$ generations earlier than the species divergence (Figure 1A). Simulation analyses confirmed this theoretical expectation, revealing a critical deviation between N_e and K_s -based divergence estimates. As illustrated in Figure 1B, direct estimation of species divergence using mean K_s values exhibited systematic biases proportional to N_e . Moreover, incorporating N_e -correction together with multi-locus substitution models (Nielsen and Slatkin, 2013) substantially reduced these deviations, bringing the corrected estimates into close agreement with the true divergence times.

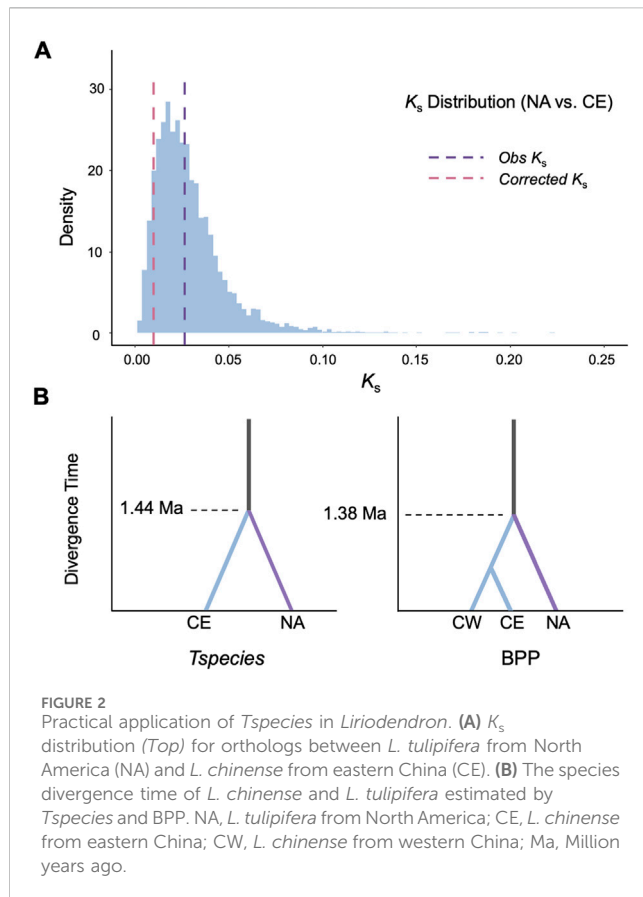
Under the coalescent model of a constant population with large N_e , the standard deviation for the coalescent time for two alleles is approximately $2N_e$. Although explicit species divergence scenarios introduce additional complexity, we hypothesized that the variance of K_s distributions could serve as a quantitative proxy for N_e . To test this hypothesis, we simulated the species divergence under varying N_e and examined the resulting relationship between N_e and K_s variance. The simulation results (Figure 1C) revealed two salient patterns: (1) a strong positive correlation between N_e and K_s variance persisted across all divergence times T ; (2) at large N_e (when $N_e > 450,000$), the relationship closely followed the analytical

expectation, whereas at smaller N_e all trajectories converged toward a single smooth curve.

To capture the observed N_e - K_s variance relationship, we fitted a generalized additive model (GAM). The GAM revealed a strongly nonlinear yet highly predictive relationship between N_e and K_s variance ($R^2 = 0.997$, Figure 1C). The smooth term for N_e showed significant complexity (effective degrees of freedom [edf] = 7.11). Model diagnostics confirmed robustness: the intercept (3.68, standard error = 0.0045) remained stable across divergence times ($t = 815.8$), while the nearly identical generalized cross-validation (GCV) score (0.0289) and scale estimate (0.0287) ruled out overfitting. Importantly, this framework retains strong predictive power for estimating N_e from empirical K_s distributions, provided that orthologous loci are sufficiently sampled across the genome.

While our model provides a computationally tractable framework to correcting divergence time estimates, it relies on simplifying assumptions, including identical sequence lengths and constant N_e across speciation events. To evaluate its robustness, we conducted sensitivity analyses under varying substitution rate (μ), sequential lengths (L) and post-divergence N_e dynamics. Simulations demonstrate that μ significantly modulates the N_e - K_s variance pattern (Figure 1D). Higher μ values intensified the positive correlation between N_e and K_s variance, underscoring the necessity of incorporating μ as a critical input parameter in our framework. The relative difference in predicted variance in different L (Supplementary Figure S1) suggested potential sequence length effects at shorter L (e.g., $L = 500$ bp), whereas longer sequences ($L \geq 1,000$ bp) produced stable estimates (mean deviation = 12.2%). Additionally, simulations considering post-divergence N_e dynamics (Supplementary Figure S2) suggested only a minor bias (~5.6%) from true divergence times, with higher μ improving accuracy (Supplementary Figure S3). Together, these results indicate that our proposed framework (Figure 1E) remains robust across a wide range of demographic and mutational scenarios.

Finally, we applied our framework to the genus *Liriodendron*, which are believed to have undergone significant population reduction throughout their evolutionary history (Chen et al., 2019). Using K_s distributions, we compared *L. tulipifera* from North America (NA) with *L. chinense* from eastern China (CE) and from western China (CW) (Figure 2A). *Tspecies* inferred the divergence time between NA and CE to be 1.44 Ma with a generation time of 30 years (or 1.21 Ma with a generation time of 20 years), which aligns well with the results obtained from the coalescent-based Bayesian estimation from BPP algorithm (1.38 Ma, Figure 2B; Supplementary Figure S4). Similarly, the estimations for divergence between NA and CW are 1.36 Ma and 1.14 Ma, with a generation time of 30 and 20 years, respectively. In contrast, direct divergence estimates between *L. tulipifera* and *L. chinense* based on mean K_s values were older, reaching 4.36 Ma (NA vs. CE), tracing back to the warmer Mid-Pliocene period. Additionally, *Tspecies* inferred the ancestral N_e of *Liriodendron* species to be approximately 5.29×10^4 , consistent with the fossil evidence indicating that this genus was once widespread across the Northern Hemisphere (Ian, 2006).



Discussion

Species divergence has been one of the most central issues in speciation studies. In recent years, with the advancements in genomic evolutionary research, speciation study has been challenged by continuously discovered hybridizations, introgressions, genome duplications, and other complex evolutionary events (Wang and Liu, 2025). An efficient and unbiased solution is hence needed in the inference of species divergence. Our results demonstrate that the variance of K_s distributions provides a robust signal of ancestral effective population size (N_e), enabling its estimation directly from standard orthologous gene datasets. By leveraging a readily quantifiable property of K_s distributions, our study offers a scalable framework for integrating coalescent theory into the estimation of species divergence.

Sensitivity analyses confirmed that the N_e – K_s variance relationship is resilient to variation in sequence length and post-divergence population dynamics. Although accurate substitution rate (μ) specification remains necessary, it is always practically required in divergence time estimation, as μ is also used as a scaling factor for time calibration. Notably, although μ is treated as a categorical variable in our GAM framework, each category corresponds to a narrow interval of μ values, and simulations show that variance within categories has only a minor influence on the N_e – K_s variance relationship (Supplementary Figure S3).

Interestingly, application to *Liriodendron* not only produced divergence estimates consistent with results from coalescent-based Bayesian approaches such as BPP, but also aligned with the climate

cooling in the Ice Ages of the early Pleistocene (Lisiecki and Raymo, 2005). Such abrupt climatic shifts are known to drive species range contractions and southward migration, ultimately driving geographic isolation between North American and Asian lineages. These aligned results underscored the population decline from the paleoclimatic and fossil evidence (Ian, 2006).

In addition to estimating species divergence, K_s distributions from paralogs are extensively used to estimate time for genome duplications (Jiao et al., 2011). Since coalescent theory also applies to the autopolyploids before their diploidization, modeling the relationship between ancestral N_e and K_s variance may similarly improve estimates in these contexts. Future work incorporating complex demographic histories, interspecies gene flows, and genome duplication events will further extend the applicability of this model across diverse evolutionary scenarios.

Data availability statement

The datasets presented in this study can be found in online repositories. Tspecies and data used in this manuscript are publicly available under an MIT license at <https://github.com/limj0987/Tspecies.git>.

Author contributions

M-JL: Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Software, Writing – original draft. X-XL: Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Software, Writing – original draft. L-LX: Resources, Validation, Writing – review and editing. B-WZ: Conceptualization, Supervision, Writing – review and editing.

Funding

The author(s) declared that financial support was received for this work and/or its publication. This project is supported by the National College Students' Innovation and Entrepreneurship Training Program and the Fundamental Research Funds for the Central Universities.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial

intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the

reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2025.1725551/full#supplementary-material>

References

- Cahill, J. A., Soares, A. E. R., Green, R. E., and Shapiro, B. (2016). Inferring species divergence times using pairwise sequential markovian coalescent modelling and low-coverage genomic data. *Philosophical Trans. R. Soc. B Biol. Sci.* 371, 20150138. doi:10.1098/rstb.2015.0138
- Charlesworth, B. (2009). Effective population size and patterns of molecular evolution and variation. *Nat. Rev. Genet.* 10 (3), 195–205. doi:10.1038/nrg2526
- Chen, J., Hao, Z., Guang, X., Zhao, C., Wang, P., Xue, L., et al. (2019). *Liriodendron* genome sheds light on angiosperm phylogeny and species-pair differentiation. *Nat. Plants* 5 (1), 18–25. doi:10.1038/s41477-018-0323-6
- Cui, L., Wall, P. K., Leebens-Mack, J. H., Lindsay, B. G., Soltis, D. E., Doyle, J. J., et al. (2006). Widespread genome duplications throughout the history of flowering plants. *Genome Res.* 16 (6), 738–749. doi:10.1101/gr.4825606
- Emms, D. M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 20 (1), 238. doi:10.1186/s13059-019-1832-y
- Hansen, J., Sato, M., Russell, G., and Kharecha, P. (2013). Climate sensitivity, sea level and atmospheric carbon dioxide. *Philosophical Trans. R. Soc. A Math. Phys. Eng. Sci.* 371, 20120294. doi:10.1098/rsta.2012.0294
- Hao, P., Wang, H., Ma, L., Wu, A., Chen, P., Cheng, S., et al. (2020). Genome-wide identification and characterization of multiple C2 domains and transmembrane region proteins in *Gossypium hirsutum*. *BMC Genomics* 21 (1), 445. doi:10.1186/s12864-020-06842-1
- Hu, T., Wei, Q., Wang, W., Hu, H., Mao, W., Zhu, Q., et al. (2018). Genome-wide identification and characterization of CONSTANS-Like gene family in radish (*Raphanus sativus*). *PLOS ONE* 13 (9), e0204137. doi:10.1371/journal.pone.0204137
- Hudson, R. R. (2002). Generating samples under a wright–fisher neutral model of genetic variation. *Bioinformatics* 18 (2), 337–338. doi:10.1093/bioinformatics/18.2.337
- Ian, M. R. (2006). Northern hemisphere plant disjunctions: a window on tertiary land bridges and climate change. *Ann. Bot.* 98 (3), 465–472. doi:10.1093/aob/mcl148
- Jiao, Y. N., Wickett, N. J., Ayyampalayam, S., Chanderbali, A. S., Landherr, L., Ralph, P. E., et al. (2011). Ancestral polyploidy in seed plants and angiosperms. *Nature* 473, 97–100. doi:10.1038/nature09916
- Lisiecki, L. E., and Raymo, M. E. (2005). A pliocene-pleistocene stack of 57 globally distributed benthic $\delta^{18}\text{O}$ records. *Paleoceanography* 20 (1). doi:10.1029/2004pa001071
- Nielsen, R., and Slatkin, M. (2013). *An introduction to population genetics: theory and applications*. MA: Sinauer Associates Sunderland.
- Pang, X. X., and Zhang, D. Y. (2024). Detection of ghost introgression requires exploiting topological and branch length information. *Syst. Biol.* 73 (1), 207–222. doi:10.1093/sysbio/syad077
- Rambaut, A., and Grass, N. C. (1997). Seq-gen: an application for the monte carlo simulation of DNA sequence evolution along phylogenetic trees. *Bioinformatics* 13 (3), 235–238. doi:10.1093/bioinformatics/13.3.235
- Sigwart, J. (2009). *Coalescent theory: an introduction*. Oxford University Press.
- Wakeley, J., and Sargsyan, O. (2009). Extensions of the coalescent effective population size. *Genetics* 181 (1), 341–345. doi:10.1534/genetics.108.092460
- Wang, Z., and Liu, J. (2025). Speciation studies in the genomic era. *Hered.* 47 (1), 71–100. doi:10.16288/j.ycz.24-218
- Wang, Y., and Obbard, D. J. (2023). Experimental estimates of germline mutation rate in eukaryotes: a phylogenetic meta-analysis. *Evol. Lett.* 7 (4), 216–226. doi:10.1093/evlett/grad027
- Wang, J., Santiago, E., and Caballero, A. (2016). Prediction and estimation of effective population size. *Heredity* 117 (4), 193–206. doi:10.1038/hdy.2016.43
- Wood, S. N. (2017). *Generalized additive models: an introduction with R, second edition*. Chapman and Hall/CRC.
- Yan, Z., Cao, Z., Liu, Y., Ogilvie, H. A., and Nakhleh, L. (2022). Maximum parsimony inference of phylogenetic networks in the presence of polyploid complexes. *Syst. Biol.* 71 (3), 706–720. doi:10.1093/sysbio/syab081
- Yang, Z. (2015). The BPP program for species tree estimation and species delimitation. *Curr. Zool.* 61 (5), 854–865. doi:10.1093/czoolo/61.5.854
- Zhang, Z. (2022). KaKs_Calculator 3.0: calculating selective pressure on coding and non-coding sequences. *Genomics, Proteomics and Bioinforma.* 20 (3), 536–540. doi:10.1016/j.gpb.2021.12.002
- Zhu, W., Wu, D., Jiang, L., and Ye, L. (2020). Genome-wide identification and characterization of SnRK family genes in *Brassica napus*. *BMC Plant Biol.* 20 (1), 287. doi:10.1186/s12870-020-02484-3
- Zwaenepoel, A., and Van de Peer, Y. (2019). Wgd-simple command line tools for the analysis of ancient whole-genome duplications. *Bioinformatics* 35 (12), 2153–2155. doi:10.1093/bioinformatics/bty915