

Plant diversification driven by genome and chromosome evolution and its reproductive and environmental correlates

Edited by

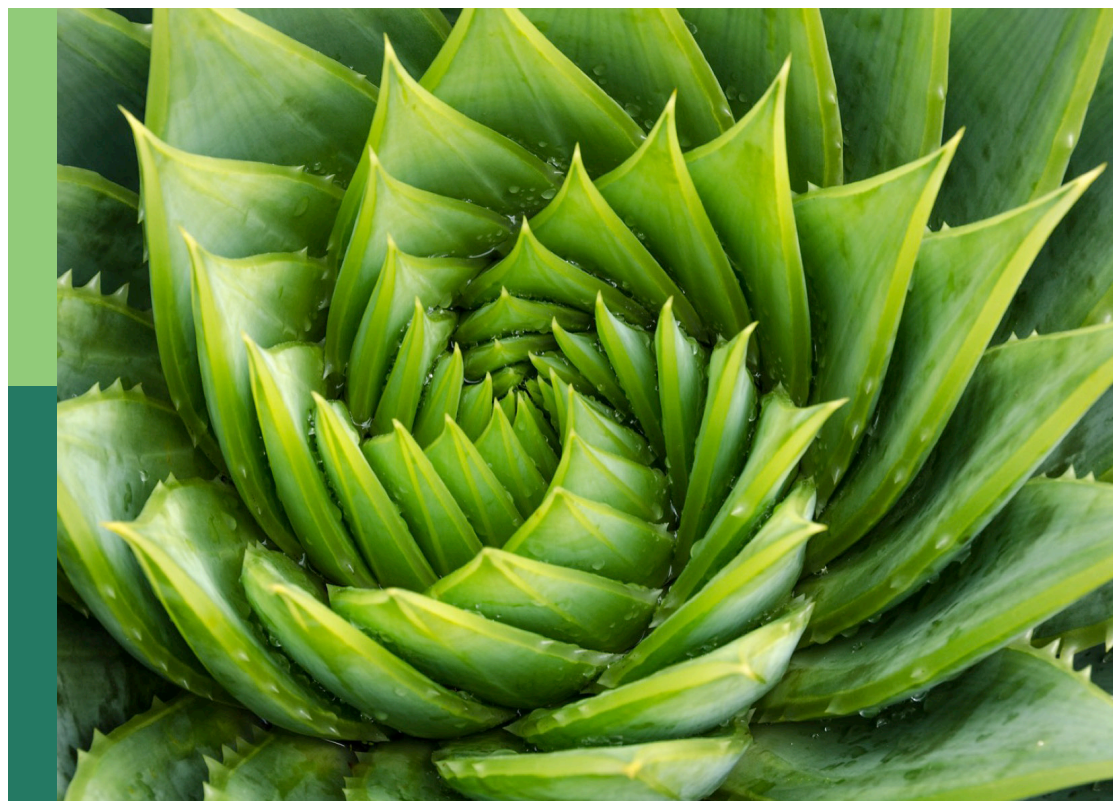
Marcial Escudero, Enrique Maguilla, Angelino Carta and Santiago Martín-Bravo

Coordinated by

Carmen Benítez Benítez

Published in

Frontiers in Plant Science



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 978-2-8325-6976-4
DOI 10.3389/978-2-8325-6976-4

Generative AI statement

Any alternative text (Alt text) provided alongside figures in the articles in this ebook 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.

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

Plant diversification driven by genome and chromosome evolution and its reproductive and environmental correlates

Topic editors

Marcial Escudero — Sevilla University, Spain

Enrique Maguilla — Universidad Pablo de Olavide, Spain

Angelino Carta — University of Pisa, Italy

Santiago Martín-Bravo — Universidad Pablo de Olavide, Spain

Topic coordinator

Carmen Benítez Benítez — Sevilla University, Spain

Citation

Escudero, M., Maguilla, E., Carta, A., Martín-Bravo, S., Benítez Benítez, C., eds. (2025). *Plant diversification driven by genome and chromosome evolution and its reproductive and environmental correlates*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-6976-4

Table of contents

- 05 **Editorial: Plant diversification driven by genome and chromosome evolution and its reproductive and environmental correlates**
Carmen Benítez-Benítez, Angelino Carta, Marcial Escudero, Enrique Maguilla and Santiago Martín-Bravo
- 08 **Genetic basis of local adaptation in the cold-tolerant mangrove *Kandelia obovata***
Chuangchao Zou, Yushuai Wang, Renchao Zhou and Tian Tang
- 25 **Striking variation in chromosome structure within *Musa acuminata* subspecies, diploid cultivars, and F1 diploid hybrids**
Denisa Beránková, Jana Čížková, Gabriela Majzlíková, Alžběta Doležalová, Hassan Mduma, Allan Brown, Rony Swennen and Eva Hřibová
- 39 **Expansions and contractions of repetitive DNA elements reveal contrasting evolutionary responses to the polyploid genome shock hypothesis in *Brachypodium* model grasses**
María Ángeles Decena, Rubén Sancho, Luis A. Inda, Ernesto Pérez-Collazos and Pilar Catalán
- 57 **The interplay between climatic niche evolution, polyploidy and reproductive traits explains plant speciation in the Mediterranean Basin: a case study in *Centaureum* (Gentianaceae)**
Ana Valdés-Flórido, Virginia Valcárcel, Enrique Maguilla, Zoila Díaz-Lifante, Cristina Andrés-Camacho, Louis Zeltner, Marina Coca-de-la-Iglesia, Nagore G. Medina, Juan Arroyo and Marcial Escudero
- 69 **Complete chloroplast genome data reveal the existence of the *Solidago canadensis* L. complex and its potential introduction pathways into China**
Yu-Tian Tao, Lu-Xi Chen, Ming Jiang, Jie Jin, Zhong-Shuai Sun, Chao-Nan Cai, Han-Yang Lin, Allison Kwok, Jun-Min Li and Mark van Kleunen
- 79 **tRNA gene content, structure, and organization in the flowering plant lineage**
Kim Carlo Monloy and Jose Planta
- 94 **Integrative taxonomy reveals cryptic diversity within the *Euphorbia nicaeensis* alliance (Euphorbiaceae) in the central Balkan Peninsula**
Angela Sharovikj Ivanova, Peter Schönschwetter, Mitko Kostadinovski, Michael H. J. Barfuss, Renata Čušterevska and Božo Frajman

- 111 **Ancient polyploidization events influence the evolution of the ginseng family (Araliaceae)**
Angélica Gallego-Narbón, Gabriel Johnson,
Mario Fernández-Mazuecos, Jun Wen and Virginia Valcárcel
- 128 **Bridging micro and macroevolution: insights from chromosomal dynamics in plants**
Carmen Benítez-Benítez, Ashwini V. Mohan,
Rogelio Sánchez-Villegas, Inés Gómez-Ramos, Ana Valdés-Flrido,
Kay Lucek, Marek Slovák, Filip Kolář, Ilia J. Leitch, Modesto Luceño,
Isabel Larridon, Angelino Carta, Joan Cuscó-Borràs, Enrique Maguilla,
Alegria Montero-Ramírez, Patrick G. Meirmans, Alison Dawn Scott,
Santiago Martín-Bravo and Marcial Escudero



OPEN ACCESS

EDITED AND REVIEWED BY
Jim Leebens-Mack,
University of Georgia, United States

*CORRESPONDENCE
Carmen Benítez-Benítez
✉ cbenitez2@us.es
Marcial Escudero
✉ aescudero2@us.es

RECEIVED 14 July 2025
ACCEPTED 06 August 2025
PUBLISHED 29 September 2025

CITATION
Benítez-Benítez C, Carta A, Escudero M,
Maguilla E and Martín-Bravo S (2025)
Editorial: Plant diversification driven by
genome and chromosome evolution and its
reproductive and environmental correlates.
Front. Plant Sci. 16:1665681.
doi: 10.3389/fpls.2025.1665681

COPYRIGHT
© 2025 Benítez-Benítez, Carta, Escudero,
Maguilla and Martín-Bravo. 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.

Editorial: Plant diversification driven by genome and chromosome evolution and its reproductive and environmental correlates

Carmen Benítez-Benítez^{1*}, Angelino Carta², Marcial Escudero^{1*},
Enrique Maguilla³ and Santiago Martín-Bravo³

¹Department of Plant Biology and Ecology, Faculty of Biology, University of Seville, Seville, Spain,

²Department of Biology, Botany Unit, University of Pisa, Pisa, Italy, ³Botany Area, Department of Molecular Biology and Biochemical Engineering, Universidad Pablo de Olavide, Seville, Spain

KEYWORDS

chromosome, microevolution, macroevolution, genome, biodiversity, diversification

Editorial on the Research Topic

Plant diversification driven by genome and chromosome evolution and its reproductive and environmental correlates

The intricate processes of plant diversification are profoundly influenced by changes at the genomic and chromosomal levels (Wang et al., 2025), which in turn impact reproductive strategies and adaptation to diverse environments (Hansen et al., 2012; Bragg et al., 2015). This Research Topic, “Plant Diversification Driven by Genome and Chromosome Evolution and Its Reproductive and Environmental Correlates,” brings together a collection of nine insightful articles that explore these multifaceted interactions across a diverse range of plant lineages. From ancient polyploidization events to contemporary local adaptations, these articles highlight the dynamic interplay between genomic architecture, chromosomal rearrangements, and evolutionary success in the plant kingdom.

Several articles in this Research Topic shed light on the fundamental role of genome and chromosome evolution in shaping plant diversity. Gallego-Narbón et al. investigated the influence of whole-genome duplication (WGD) events on the evolution of the ginseng family (Araliaceae). Their phylogenomic analyses suggest that ancient hybridization and WGDs preceded the origin and diversification of major clades within the family, underscoring the long-term impact of these events (Soltis and Soltis, 2016). Building on this theme, Benítez-Benítez et al. provided a comprehensive review bridging micro and macroevolutionary processes through chromosomal dynamics. The reviewed evidence reflects that while polyploidy and dysploidy are known drivers of speciation, other chromosomal rearrangements like insertions, deletions, inversions, and translocations are increasingly recognized for their role in local adaptation and speciation. In addition, it suggests that certain chromosomal dynamics become fixed over macroevolutionary time after the filter of speciation (Rolland et al., 2023).

Further exploring the impact of polyploidy, Sharovikj Ivanova et al. employed an integrative taxonomic approach to uncover cryptic diversity within the *Euphorbia nicaeensis* alliance (Euphorbiaceae). Their research revealed multiple polyploidization events and complex phylogenetic patterns, resulting in the description of new species and emphasizing the significant role of polyploidy driving diversification within this group. The dynamic behavior of repetitive DNA elements in polyploid genomes is addressed by Decena et al. in their study of *Brachypodium* grasses (Poaceae). They demonstrate how the expansion and contraction of repeat elements contribute to genome size variation and respond to the “polyploid genome shock hypothesis,” revealing contrasting evolutionary outcomes across different *Brachypodium* lineages. Meanwhile, Beránková et al. uncover striking variations in chromosome structure within *Musa acuminata* subspecies (Musaceae) and cultivars, underscoring the profound impact of hybridization and polyploidization on chromosomal rearrangements in cultivated bananas.

Beyond the direct impact of genome and chromosome evolution (Lucek et al., 2023; Mohan et al., 2024), this Research Topic also delves into how these changes correlate with reproductive traits and environmental adaptation. Valdés-Flrido et al. investigated the interplay between climatic niche evolution, polyploidy, and reproductive traits in the Mediterranean genus *Centaureum* (L.) Hill. Their findings suggest that polyploidization is a crucial process for plant evolution in the Mediterranean region (Escudero et al., 2018), facilitating speciation and diversification into new areas with different climates, and involving shifts in climatic niches and the evolution of novel reproductive strategies. This emphasizes the adaptive advantage conferred by genomic changes in response to environmental pressures (Hansen et al., 2012).

The genetic basis of local adaptation to environmental challenges is further investigated by Zou et al. in the cold-tolerant mangrove *Kandelia obovata* Sheue, Liu & Yong. Using whole-genome re-sequencing they identified strong population structure and selective sweeps in highly differentiated regions, with candidate genes underlying local adaptation to temperature-related variables. This study provides crucial insights into how genomic variation underlies a species' ability to adapt to specific environmental conditions (Bragg et al., 2015).

This Research Topic also showcases advanced genomic approaches that facilitate a deeper understanding of plant diversification (Soltis and Soltis, 2021). Monloy and Planta provide a comprehensive analysis of tRNA gene content, structure, and organization across the flowering plant lineage. Their comparative genomic study reveals variation in the number of nuclear tDNAs and distinct clustering patterns among different plant groups, providing a valuable foundation for future research on tRNA gene function and regulation. Furthermore, Tao et al. utilized complete chloroplast genome data to investigate the *Solidago canadensis* L. complex and its anthropogenic introduction pathways into China. Their *de novo* assembled chloroplast genomes offer important insights into phylogenetic relationships, sequence divergence, and potential

introduction routes, demonstrating the power of organellar genomics in understanding invasion biology and evolutionary history (Keller and Taylor, 2008).

In summary, the articles in this Research Topic collectively underscore the pivotal role of genome and chromosome evolution in driving plant diversification (Soltis and Soltis, 2021). From ancient polyploidization events that shaped entire plant families to the subtle chromosomal rearrangements influencing local adaptation and speciation at microevolutionary levels, the contributions highlight the dynamic and intricate mechanisms at play. Furthermore, the interplay between these genomic changes, reproductive strategies, and environmental correlates provides a more complete scenario of how plants adapt and diversify (Cowling and Pressey, 2001). The ongoing advancements in genomic technologies are clearly helping researchers to unravel these complex processes with unprecedented detail (Zuntini et al., 2024). We hope this Research Topic serves as a valuable resource for researchers interested in the genetic and chromosomal underpinnings of plant evolution and inspires further investigations into these fascinating areas.

Author contributions

CBB: Writing – original draft, Writing – review & editing. AC: Writing – original draft, Writing – review & editing. ME: Writing – original draft, Writing – review & editing. EM: Writing – original draft, Writing – review & editing. SMB: Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research and/or publication of this article. This work was supported by grants from the Spanish AEI MICINN (PID2021-122715NB-I00 DiversiChrom) to ME and CB-B. CB-B was also supported by a Postdoctoral Fellowship Program grant (JDC2022-048955-I), Innovation and Universities (MICIU/AEI/10.13039/501100011033) and by the EuropeanUnionNextGenerationEU/PRTR.

Acknowledgments

We thank all reviewers and Frontiers editors that critically scrutinized and assisted in editing the articles included in this Research Topic. We are also grateful to all the researchers who contributed their work to this collection.

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 author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Generative AI statement

The author(s) declare that no Generative AI was 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.

References

- Bragg, J. G., Supple, M. A., Andrew, R. L., and Borevitz, J. O. (2015). Genomic variation across landscapes: insights and applications. *New Phytol.* 207, 953–967. doi: 10.1111/nph.13410
- Cowling, R. M., and Pressey, R. L. (2001). Rapid plant diversification: planning for an evolutionary future. *Proc. Natl. Acad. Sci.* 98, 5452–5457. doi: 10.1073/pnas.101093498
- Escudero, M., Balao, F., Martín-Bravo, S., Valente, L., and Valcárcel, V. (2018). Is the diversification of Mediterranean Basin plant lineages coupled to karyotypic changes? *Plant Biol.* 20, 166–175. doi: 10.1111/plb.12563
- Hansen, M. M., Olivieri, I., Waller, D. M., Nielsen, E. E., and GeM Working Group (2012). Monitoring adaptive genetic responses to environmental change. *Mol. Ecol.* 21, 1311–1329. doi: 10.1111/j.1365-294X.2011.05463.x
- Keller, S. R., and Taylor, D. R. (2008). History, chance and adaptation during biological invasion: separating stochastic phenotypic evolution from response to selection. *Ecol. Lett.* 11, 852–866. doi: 10.1111/j.1461-0248.2008.01188.x
- Lucek, K., Giménez, M. D., Joron, M., Rafajlović, M., Searle, J. B., Walden, N., et al. (2023). The impact of chromosomal rearrangements in speciation: from micro-to macroevolution. *Cold Spring Harbor Perspect. Biol.* 15, a041447. doi: 10.1101/cshperspect.a041447
- Mohan, A. V., Escuer, P., Cornet, C., and Lucek, K. (2024). A three-dimensional genomics view for speciation research. *Trends Genet.* 40, 638–641. doi: 10.1016/j.tig.2024.05.009
- Rolland, J., Henao-Díaz, L. F., Doebe, M., Germain, R., Harmon, L. J., Knowles, L. L., et al. (2023). Conceptual and empirical bridges between micro-and macroevolution. *Nat. Ecol. Evol.* 7, 1181–1193. doi: 10.1038/s41559-023-02116-7
- Soltis, P. S., and Soltis, D. E. (2016). Ancient WGD events as drivers of key innovations in angiosperms. *Curr. Opin. Plant Biol.* 30, 159–165. doi: 10.1016/j.pbi.2016.03.015
- Soltis, P. S., and Soltis, D. E. (2021). Plant genomes: markers of evolutionary history and drivers of evolutionary change. *Plants People Planet* 3, 74–82. doi: 10.1002/ppp3.10159
- Wang, S., Mank, J. E., Ortiz-Barrientos, D., and Rieseberg, L. H. (2025). Genome architecture and speciation in plants and animals. *Mol. Ecol.*, e70004. doi: 10.1111/mec.70004
- Zuntini, A. R., Carruthers, T., Maurin, O., Bailey, P. C., Leempoel, K., Brewer, G. E., et al. (2024). Phylogenomics and the rise of the angiosperms. *Nature* 629, 843–850. doi: 10.1038/s41586-024-07324-0



OPEN ACCESS

EDITED BY

Enrique Maguilla,
Universidad Pablo de Olavide, Spain

REVIEWED BY

Evandro Novaes,
Universidade Federal de Lavras, Brazil
Isabel Marques,
Universidade de Lisboa, Portugal

*CORRESPONDENCE

Renchao Zhou

✉ zhrench@mail.sysu.edu.cn

Tian Tang

✉ tsstt@mail.sysu.edu.cn

RECEIVED 12 February 2024

ACCEPTED 09 April 2024

PUBLISHED 24 April 2024

CITATION

Zou C, Wang Y, Zhou R and Tang T (2024)
Genetic basis of local adaptation in the
cold-tolerant mangrove *Kandelia obovata*.
Front. Plant Sci. 15:1385210.
doi: 10.3389/fpls.2024.1385210

COPYRIGHT

© 2024 Zou, Wang, Zhou and Tang. 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.

Genetic basis of local adaptation in the cold-tolerant mangrove *Kandelia obovata*

Chuangchao Zou, Yushuai Wang, Renchao Zhou* and Tian Tang*

State Key Laboratory of Biocontrol and Guangdong Key Laboratory of Plant Resources, School of Life Sciences, Sun Yat-sen University, Guangzhou, Guangdong, China

Understanding the genetic basis of local adaptation is crucial in the context of global climate change. Mangroves, as salt-tolerant trees and shrubs in the intertidal zone of tropical and subtropical coastlines, are particularly vulnerable to climate change. *Kandelia obovata*, the most cold-tolerant mangrove species, has undergone ecological speciation from its cold-intolerant counterpart, *Kandelia candel*, with geographic separation by the South China Sea. In this study, we conducted whole-genome re-sequencing of *K. obovata* populations along China's southeast coast, to elucidate the genetic basis responsible for mangrove local adaptation to climate. Our analysis revealed a strong population structure among the three *K. obovata* populations, with complex demographic histories involving population expansion, bottleneck, and gene flow. Genome-wide scans unveiled pronounced patterns of selective sweeps in highly differentiated regions among pairwise populations, with stronger signatures observed in the northern populations compared to the southern population. Additionally, significant genotype-environment associations for temperature-related variables were identified, while no associations were detected for precipitation. A set of 39 high-confidence candidate genes underlying local adaptation of *K. obovata* were identified, which are distinct from genes under selection detected by comparison between *K. obovata* and its cold-intolerant relative *K. candel*. These results significantly contribute to our understanding of the genetic underpinnings of local adaptation in *K. obovata* and provide valuable insights into the evolutionary processes shaping the genetic diversity of mangrove populations in response to climate change.

KEYWORDS

local adaptation, population genomics, demographic history, selective sweeps, genome-environment association, mangroves

1 Introduction

Understanding the evolutionary processes and genetic basis of local adaptation has been a longstanding interest in evolutionary biology (Hoban et al., 2016; López-Goldar & Agrawal, 2021). Abiotic and biotic effects such as low and high temperature, drought, flooding, herbivore and pathogen stresses impose different selective pressures across habitats (Vanwallendael et al.,

2019). The interactions among selection, gene flow and genetic drift shape genetic variation within and between populations, leading to evolutionary differentiation at different spatial scales (Orsini et al., 2013; von Takach, Penton, et al., 2021). Local adaptation can occur when genetic differentiation allows a single population to become better adapted in a particular set of environmental conditions within its range (Angert et al., 2020). The genomic architecture of natural populations thus provides an opportunity to reveal and link molecular mechanisms underlying species' phenotypic diversity with diverse environments in which the species lives. This knowledge is important for informing management decisions in the context of rapid contemporary environmental changes (Isabel et al., 2020).

Climate adaptation is one of the most prevalent forms of local adaptation that has significant impact on the distribution of a species. Classical experimental approaches explore to which extent morphological, physiological and transcriptional responses are associated with changes in environmental factors, typically through transplants and common gardens (Carley et al., 2021; Oomen & Hutchings, 2022). Nevertheless, as most traits are polygenic, genes involved in ecological, physiological, or transcriptional changes may only be weakly related to fitness. With recent advance of sequencing technology, population genetic or genomic analyses are typically utilized to identify selective signatures linked to local adaptation, with allele frequencies expected to demonstrate ecological differentiation or correlation with specific environmental factors (von Takach, et al., 2021; Hu et al., 2022; Wang J. et al., 2022). As adaptation to climate change progresses, genetically distinct populations undergoing local adaptation may eventually become separate species if they no longer interbreed or exchange genes. Therefore, studying the genetic basis of climate adaptation can also provide insights into the evolutionary processes leading to the fixation of genetic differentiation and subsequent speciation (Levin, 2019; Dool et al., 2022).

Mangroves are a group of phylogenetically diverse woody plants that inhabit tropical and subtropical intertidal zones with great ecological and economical significance (Tomlinson, 1986; Sribianti et al., 2021). At the dynamic interface between sea and land, the structure and biodiversity of mangrove communities are sensitive to climate change (Jennerjahn et al., 2017). Mangrove forests have faced and survived several catastrophic climate change events since their origination during the Late Cretaceous-Early Tertiary period (Singh et al., 2022). Sea-level rise significantly influences the historical distribution of mangrove forests, leading to speciation in five common mangrove species via multiple cycles of mixing, isolation, and mixing (Lovelock et al., 2015; He et al., 2022). Meanwhile, low temperature stress governs the latitudinal range limits for contemporary mangrove flora (Lin, 1999) and plays a crucial role in mangrove afforestation and restoration (Ellison, 2000; Quisthoudt et al., 2012; Su et al., 2021; Chen et al., 2023). Given the ecological importance of mangroves and the potential impact of climate change on their habitats, further research into their adaptation to low temperatures is vital. The complex demographic histories of mangroves involving bottleneck, genetic drift, population expansion, and gene flow also provide a challenge for studying local adaptation.

Kandelia obovata, regarded as the most cold-tolerant mangrove species, is able to survive chilling temperatures as low as 4.2°C

(Maxwell, 1995; Sheue et al., 2003; Zhou et al., 2007). This species can be found from the Gulf of Tonkin in the northeast to southern Japan, separated from its cold-intolerant relative, *K. candel*, by the South China Sea (Sheue et al., 2003), representing a good case in point for ecological speciation. Besides the clear differences in cold tolerance between *Kandelia* species, common garden studies have identified a greater cold tolerance in the northern populations of *K. obovata* compared to the southern populations (Sheue et al., 2003; Zhao et al., 2021), indicating the occurrence of local adaptation within *K. obovata*. Whole-genome bisulfite sequencing and RNA-seq of the *K. obovata* transplants suggest that modifications of DNA methylation in MADS-box genes may contribute to the adaptation to new environments, whereas the suppressed expression of lignin biosynthesis genes appears to play a role in maladaptation (Zhao et al., 2021). Additionally, the physiological and expressional analyses have highlighted several key genes and pathways that are potentially involved in cold tolerance in *K. obovata*, such as genes involved in calcium signaling, cell wall modification, and post-translational modifications of ubiquitination pathways (He et al., 2023). However, how genetic differentiation is maintained and whether the same genes underlying local adaptation to cold stress within species can also be responsible for between-species differences in cold tolerance remain largely unknown.

In this study, we investigated the genomic architecture of *K. obovata* populations along a latitudinal gradient in China using whole-genome resequencing, based on the chromosome-anchored genome assembly of *K. obovata* (Hu et al., 2020). Our analysis aimed to infer the demographic history and detect selective signatures within *K. obovata* populations, with the underlying hypothesis that local adaptation has occurred despite gene flow, driven by strong selection pressures related to temperature. We expected the signatures of local adaptation to be more pronounced in the northern population compared to the southern population. Moreover, we identified candidate genes under selection by comparing polymorphism within *K. obovata* with divergence between *K. obovata* and *K. candel*. We expected that the genes involved in local adaptation within *K. obovata* would differ from those under selection in the lineage of *K. obovata* since its divergence from *K. candel*.

2 Materials and Methods

2.1 Plant materials and population resequencing

We sampled leaves from a total of 46 *K. obovata* individuals, collecting plants at least 20 meters apart within each of three natural populations along the southeast coast of China (Supplementary Table S1). Specifically, we collected 14 individuals from Shacheng Harbor (Fuding, Fujian Province), 17 from Yanzao Village (Shenzhen, Guangdong Province), and 15 from Bamen Bay (Wenchang, Hainan Province). The fresh leaves were air-dried with silica gel before DNA extraction. Genomic DNA was then extracted using a modified CTAB protocol (Yang et al., 2008). Subsequently, DNA libraries were obtained using Nextera Mate Pair Sample Preparation Kit (Illumina USA), followed by whole-genome

re-sequencing on an Illumina HiSeq 2500 platform (Illumina, San Diego, CA, USA) with paired-end reads of 150 bp (PE150). The average sequencing depth for each individual was approximately 40× (Supplementary Table S1).

2.2 Reads mapping, variant callings and SNP filtering

A high-quality chromosome-scaled assembly of the *K. obovata* genome, with a total length of 177 Mb, was published using PacBio, Illumina and Hi-C sequencing (Hu et al., 2020) and served as the reference genome for this study. Raw Illumina reads were initially processed by trimming and filtering using Trimmomatic v0.39 (Bolger et al., 2014), followed by mapping onto the *K. obovata* reference genome using Burrows-Wheeler-Alignment (BWA) v0.7.12-r1039 (Li & Durbin, 2009). Subsequently, sorted bam files were generated from sam files using SAMtools v1.6 (Li et al., 2009), and PCR duplicates were removed using MarkDuplicates in the Picard toolkit. Variants were called using HaplotypeCaller and genotyped using GenotypeGVCFs in Genome Analysis Tool Kit (GATK) (McKenna et al., 2010). The analysis solely focused on single nucleotide polymorphism sites (SNPs) and employed specific filtering criteria to reduce false positives: (1) SNPs with a read number less than three for each individual were removed; (2) SNPs with a minor allele frequency (MAF) less than 0.05 were discarded; (3) SNPs were further filtered using VariantFiltration with the following parameters: quality by depth (QD) < 2.0, Fisher strand (FS) > 60.0, mapping quality (MQ) < 40.0, mapping quality tank sum test (MQRankSum) > -12.5, and read pos rank sum test (ReadPosRankSum) < -8.0. Finally, SNP annotation was conducted based on the *K. obovata* genome using snpEff 4.3t (Cingolani et al., 2012). Following sequence alignment, removal of PCR duplicates, and SNP filtering, a set of high-quality SNPs sites were retained for the subsequent analyses.

2.3 Genetic diversity and population differentiation

The entire *K. obovata* genome was divided into non-overlapping 20-kb bins. Nucleotide diversity (θ_{π} ; Tajima, 1989) for each *K. obovata* population and fixation index (F_{ST} ; Weir & Cockerham, 1984) between pairwise populations were calculated within each bin using VCFtools v0.1.15 (Danecek et al., 2011). Pairwise genetic distance (D_{XY} ; Nei & Miller, 1990) between populations were calculated within each bin using pixy v1.2.7 (Korunes & Samuk, 2021).

2.4 Population structure analysis

A principal component analysis (PCA) was performed to visualize inter-individual genetic relationships using PLINK v1.90 (Purcell et al., 2007). A phylogenetic tree based on the high-quality SNPs was constructed using IQtree (Nguyen et al., 2015) with the

parameter: -m GTR+F+G4+ASC, for which the input PHYLIP file was converted using vcf2phylic.py (<https://doi.org/10.5281/zenodo.1257058>). The number of genetic clusters (K) was identified using ADMIXTURE v1.3.0 with default parameters (Alexander et al., 2009). Various values of K ranging from 2 to 5 were tested, and the best K was selected based on the minimum error rate of K value.

2.5 Demographic modelling

The demographic history of *K. obovata* was inferred based on its observed population structure through the construction of a two-dimensional joint unfolded site frequency spectrum (2D-SFS) using the easySFS tools (<https://github.com/isaacovercast/easySFS#easysfs>), with all SNPs and the projection number equal to the individual number (i.e. 14, 17 and 15) for each of the three populations, respectively. We considered various scenarios of divergence, bottleneck, expansion, and/or migration, which were represented by different models: (i) Models 1-4: Divergence, including one-step isolation in model 1 and three two-step isolations of the *K. obovata* populations in models 2-4; (ii) Model 5-7: Bottleneck, with one population experiencing the reduction of population size and then recovery; (iii) Model 8-10: Expansion, with one population experiencing exponential population changes; (iv) Model 11-14: Migration, with asymmetric gene flow and differences in occurrence between different populations; (v) Model 15: Composite model incorporating each of the estimated best scenario for divergence (Model 1), bottleneck (Model 6), expansion (Model 10), and migration (Model 14) (Supplementary Figure 1).

These demographic models were compared, and demographic parameters were inferred using a coalescent simulation-based method as implemented in fastsimcoal2.6 (Excoffier et al., 2021). The initial ranges for the parameter estimation were listed in Supplementary Table S2. The log-likelihood for a set of demographic parameters was estimated using 100,000 coalescent simulations, with 80 conditional maximum algorithm cycles in each run and global maximum likelihood estimates obtained from 100 independent runs. The maximum likelihood value of the 100 independent runs for each model was used to compare between models using the Akaike information criterion (AIC) and Akaike's weight of evidence tests. The model with the highest Akaike's weight value was considered as the optimal model. The parameter confidence intervals (CIs) for the optimal model were obtained from 100 parametric bootstrap samples, independently run 100 times in each bootstrap. When converting estimates to years, it was assumed that the mutation rate and the average generation interval time in *K. obovata* were 7.86×10^{-8} per site per generation and 20 years per generation (average time from seed germination to seed production), respectively (He et al., 2022).

2.6 Detection of positive selection

Sliding window analysis was employed to detect outlier genomic regions with strong population differentiation using pairwise

fixation index (F_{ST}) and the cross-population composite likelihood ratio (XP-CLR) test (Chen et al., 2010) with a 20-kb window size and 5-kb step size. Pairwise F_{ST} estimates were calculated using VCFtools v0.1.15. The XP-CLR test was performed using the xpcr python module (<https://github.com/hardningj/xpcr>) with default parameters. Outlier windows with both F_{ST} and normalized XP-CLR values at least 1.96 standard deviations (SD) above the mean (one-tailed p -value = 0.025) were identified as the highly differentiated regions (HDRs). Sliding window analysis of Fay and Wu's H -statistic (Fay & Wu, 2000) in HDRs was conducted for each population separately, in comparison with the whole genome, using ANGSD v0.921 (Korneliussen et al., 2014) with a 20-kb window size and 5-kb step size.

To detect positively selected genes, the McDonald-Kreitman (MK) test (McDonald & Kreitman, 1991) was conducted using whole-genome resequencing data for all *K. obovata* individuals obtained in this study and the reference genome of *K. candel* (He et al., 2022) as an outgroup. BlastN (Altschul et al., 1990) was used to identify homologous genes between the two *Kandelia* species using an e -value threshold of $< 1e-10$, resulting in the identification of 16,536 one-to-one orthologous genes. The number of non-synonymous (D_n) and synonymous (D_s) substitutions were compared to the number of non-synonymous (P_n) and synonymous (P_s) polymorphisms within *K. obovata* for coding sequences of each *Kandelia* homologous gene using *K. candel* as an outgroup. Genes with p -value lower than 0.05 in the one side Fisher's exact test with Benjamini-Hochberg multiple test (Benjamini & Hochberg, 1995) correction were classified as positively selected genes.

2.7 Genotype-environment associations

Redundancy analysis (RDA) was used to identify associations between SNP variations across diverse populations and environmental parameters (Forester et al., 2018). Climate data including the mean annual temperature, mean annual minimum temperature, and mean annual precipitation for the three sampling locations were obtained from the Central Meteorological Observatory (<http://www.nmc.cn/>). The RDA was conducted using the RDA function from the vegan package (Oksanen et al., 2007) with the input file of genotype matrix comprising all SNPs as transformed by PLINK v1.90. Genotypes at each SNP site was encoded as follows: 0 for homozygotes identical to the reference, 1 for heterozygotes, and 2 for homozygotes differing from the reference. The proportion of variance explained by the environmental variables was evaluated using the RsquareAdj function. The significance of the linear relationship between each constrained axis and the environmental variables was then assessed using the anova.cca function. In this study, two constrained axes (RDA1 and RDA2) were found to be significant. Outlier SNPs for each significant constrained axis were identified using a cutoff of 1.96 SD greater or less than mean (two-tailed p -value = 0.05) and were annotated with known genes based on the *K. obovata* genome.

2.8 *K. obovata* gene functional annotation and GO enrichment analysis

To annotate the *K. obovata* genes, BlastP (Altschul et al., 1990) was used to identify homologous genes between the *K. obovata* and *Arabidopsis thaliana* using an e -value threshold of $< 1e-10$. Gene Ontology (GO) functions of each *K. obovata* gene were annotated using eggNOG-mapper (Cantalapiedra et al., 2021). To evaluate potential over-representation of functional gene classes, we conducted GO enrichment analysis using the R package ClusterProfiler (Yu et al., 2012) by applying the annotation mentioned above. The statistical significance of over-represented GO terms within the input gene sets was assessed through Fisher's exact tests, with a significance threshold set at False Discovery Rate (FDR) < 0.2 .

3 Results

3.1 Genomic data and genetic diversity

A total of 0.82 Tb of data with an average sequencing depth of approximately 40× per individual were obtained for the 46 *K. obovata* individuals from the three populations, representing the natural distribution of *K. obovata* from north to south in China (Table 1; Figure 1A). After quality control and filtering, the clean reads were mapped to the reference genome of *K. obovata*, achieving an average mapping rate of 93.33%. The average depth of uniquely mapped reads per site was 40.23 (Supplementary Table S1). In total, we obtained 189,909 high-confidence single nucleotide polymorphism (SNP) sites, corresponding to an average density of 1.07 SNPs per kilobase in the *K. obovata* genome. Out of these SNPs, 25,988 (13.7%) were in protein-coding regions, including

TABLE 1 Sampling information and genetic measures of three *K. obovata* populations.

Terms	Populations		
	Fuding (F)	Shenzhen (S)	Wenchang (W)
Location (longitude, latitude)	120.33°E, 27.29°N	114.52°E, 22.65°N	110.83°E, 19.60°N
Population size	14	17	15
Annual average temperature (°C)	19.5	23.5	24.9
Annual average minimum temperature (°C)	5.8	12.4	15.6
Annual average precipitation (mm)	1,743.3	1,889.3	1,913.0
Number of SNPs	115,239	160,726	140,250
Nucleotide diversity (θ_π)	2.32×10^{-4}	3.27×10^{-4}	2.87×10^{-4}

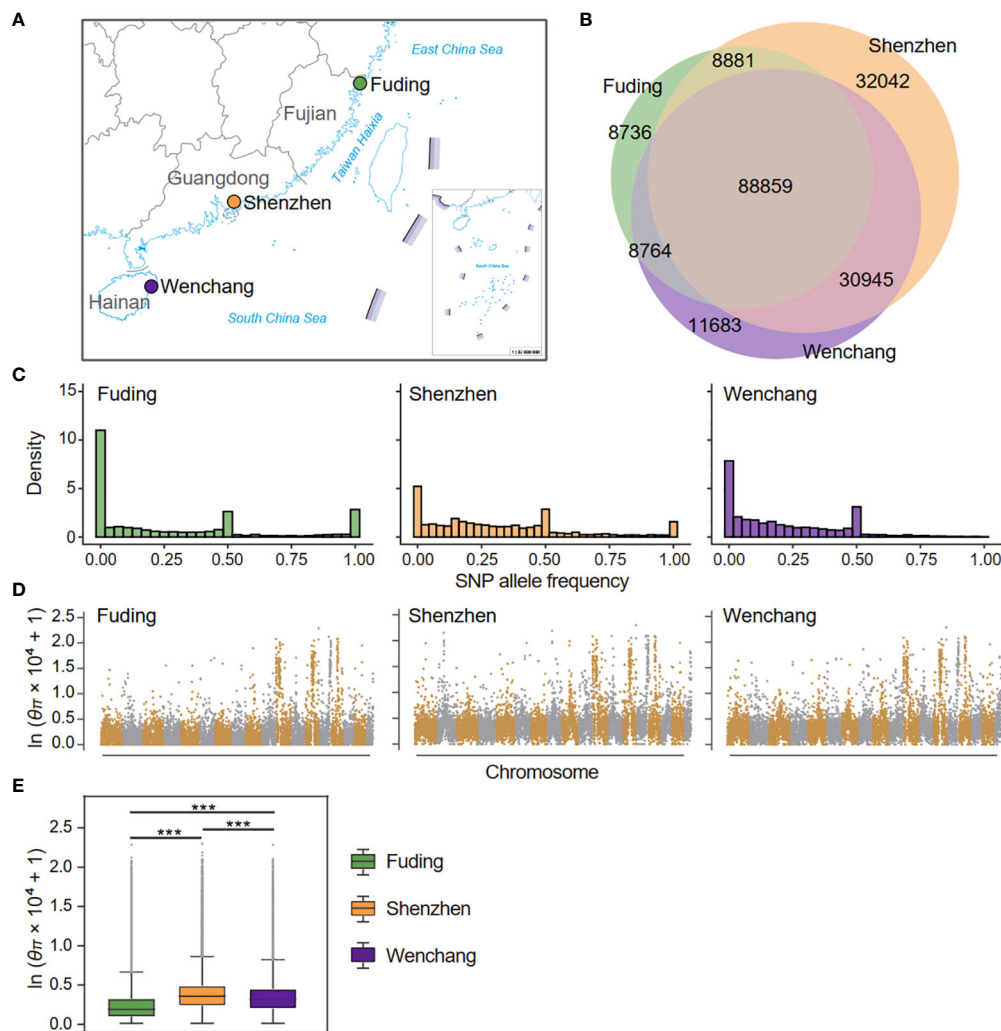


FIGURE 1

Sampling locations and genetic diversity of three populations of *Kandelia obovata* in China. (A) Map depicting the sampling locations of the Fuding ($n = 14$), Shenzhen ($n = 17$), and Wenchang ($n = 15$) populations in China. (B) Venn diagrams show the numbers of shared and private SNPs detected in the three *K. obovata* populations. (C) Site frequency spectra based on 189,909 SNPs in each population. (D) Genome-wide analysis of nucleotide diversity (θ_π) in each *K. obovata* populations. θ_π was calculated in nonoverlapping 20-kb bins and displayed in logarithmic scale across the *K. obovata* genome. (E) Boxplot displaying the distribution of θ_π in three *K. obovata* populations. Asterisks indicate the significance level of Mann-Whitney U test: *** p -value < 0.001.

exons and introns, 27,551 (14.5%) were found in putative regulatory regions, including promoters (2kb upstream of TSS) and UTRs, and the majority (56.7%) were intergenic (Supplementary Figure 2).

When comparing between populations, the number of high-confidence SNP sites was highest for the Shenzhen population (160,726), followed by the Wenchang (140,250) and Fuding (115,239) population. The number of private or population-specific SNPs were 8,736, 32,042, and 11,683 for the Fuding, Shenzhen and Wenchang populations, respectively (Figure 1B). The site frequency spectrum of individual *K. obovata* populations revealed that the Fuding populations had the highest proportion (16.7%) of fixed SNPs followed by the Shenzhen population (5.5%), while only 75 SNPs (0.6%) were fixed in the Wenchang population (Figure 1C). However, the Wenchang population had a higher percentage (9.4%) of rare SNPs (frequency < 0.05), than the Fuding (5.8%) and Shenzhen population (4.4%, Figure 1C). Meanwhile, all three populations possessed considerable medium-frequency SNPs,

indicating a high level of heterozygosity within individuals of *K. obovata* (Figure 1C).

We divided the *K. obovata* genome into 20-kb non-overlapping bins and used Tajima's π (θ_π) as a measurement of nucleotide diversity at the individual SNP level. The average nucleotide diversity per site was calculated for each bin and compared between populations of *K. obovata*. Nucleotide diversity was unevenly distributed across the *K. obovata* genome, with the most pronounced diversity observed in chromosomes 11, 13, 14 and 15 (Figure 1D). Consistent with the observed number of high-quality SNPs, the Shenzhen population exhibited the highest nucleotide diversity averaged across the genome (mean $\pm$ SD, $\theta_\pi = 3.27 \times 10^{-4} \pm 9.63 \times 10^{-4}$), followed by the Wenchang population ($\theta_\pi = 2.87 \times 10^{-4} \pm 8.64 \times 10^{-4}$), while the Fuding population showed the lowest nucleotide diversity ($\theta_\pi = 2.32 \times 10^{-4} \pm 8.87 \times 10^{-4}$) (Figure 1E). The differences in nucleotide diversity between pairwise populations were all significant (Mann-Whitney U test, all p -value < 0.05; Figure 1E).

3.2 Genetic differentiation and population structure

To assess population differentiation, the fixation index F_{ST} (Weir and Cockerham, 1984) and the average nucleotide diversity D_{XY} (Nei & Miller, 1990) were calculated for each SNP and averaged within each bin for pairwise populations of *K. obovata* (Figure 2). F_{ST} values revealed substantial population differentiation, particularly evident when comparing the Fuding population with the others: Fuding vs. Shenzhen ($F_{ST} = 0.30 \pm 0.19$) and Fuding vs. Wenchang ($F_{ST} = 0.30 \pm 0.16$), in contrast to Shenzhen vs. Wenchang ($F_{ST} = 0.24 \pm 0.13$) (Table 1 and Figure 2A). D_{XY} values were similar among pairs of populations ($D_{XY} = 0.30 \pm 0.15$ to 0.33 ± 0.11), all supporting a high level of population differentiation in *K. obovata* (Figure 2B). The principal component analysis (PCA) revealed a clear clustering pattern, indicating that individuals within a population tend to group together and are distinctly separated from individuals in other populations (Figure 2C). Similarly, the structure analysis

indicated that the pattern of ancestry was best represented by $K = 3$, which is supported by the lowest minimum K -value error rate (cross-validation error = 0.41) (Figure 2D). The three clusters that resulted from $K = 3$ clearly separated individuals according to their population origination. When $K = 2$, individuals from Fuding population formed a cohesive cluster, while individuals from Shenzhen and Wenchang clustered together. Subsequent analyses at K values of 4 and 5 revealed highly variable levels of admixture within the Shenzhen and Wenchang populations. However, there is no significant admixture within the genetic cluster in Fuding population, indicating a relatively more isolated genetic profile for this population.

3.3 Inference of demographic history

The optimal demographic model (model 15; log-likelihood = -1089415.63, AIC = 2178865.27, Δ AIC = 0), as shown in Figure 3

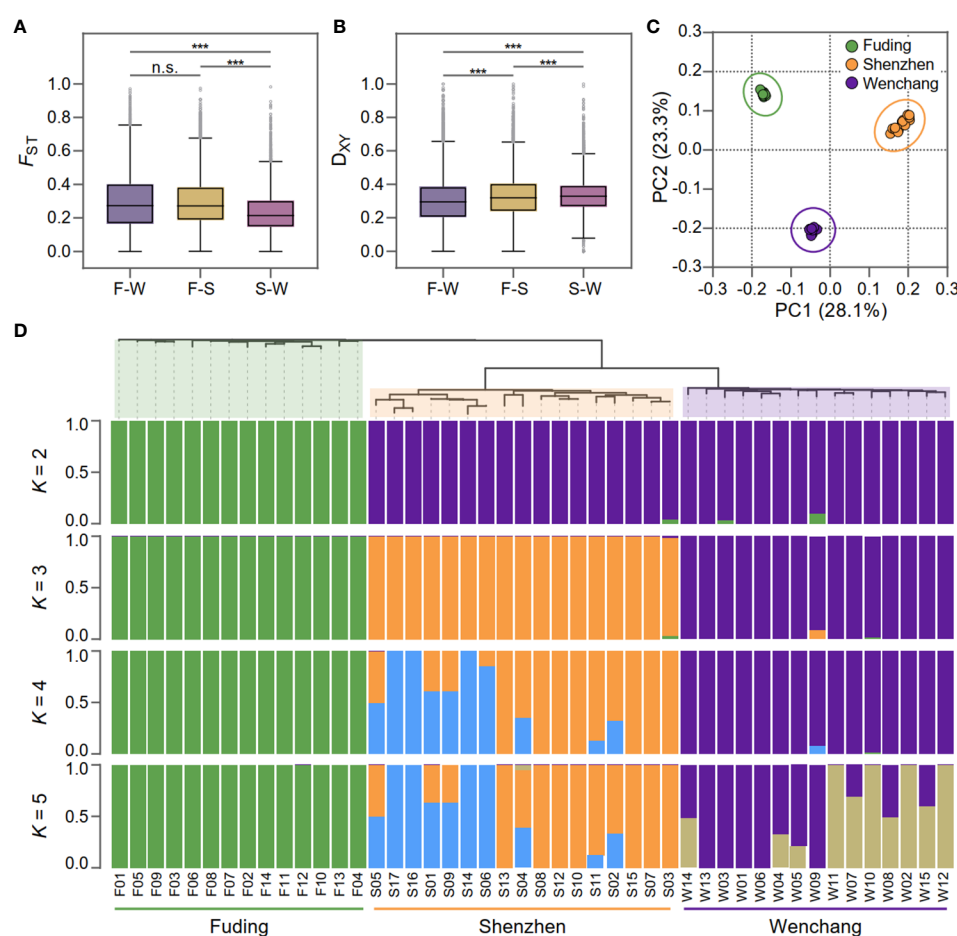


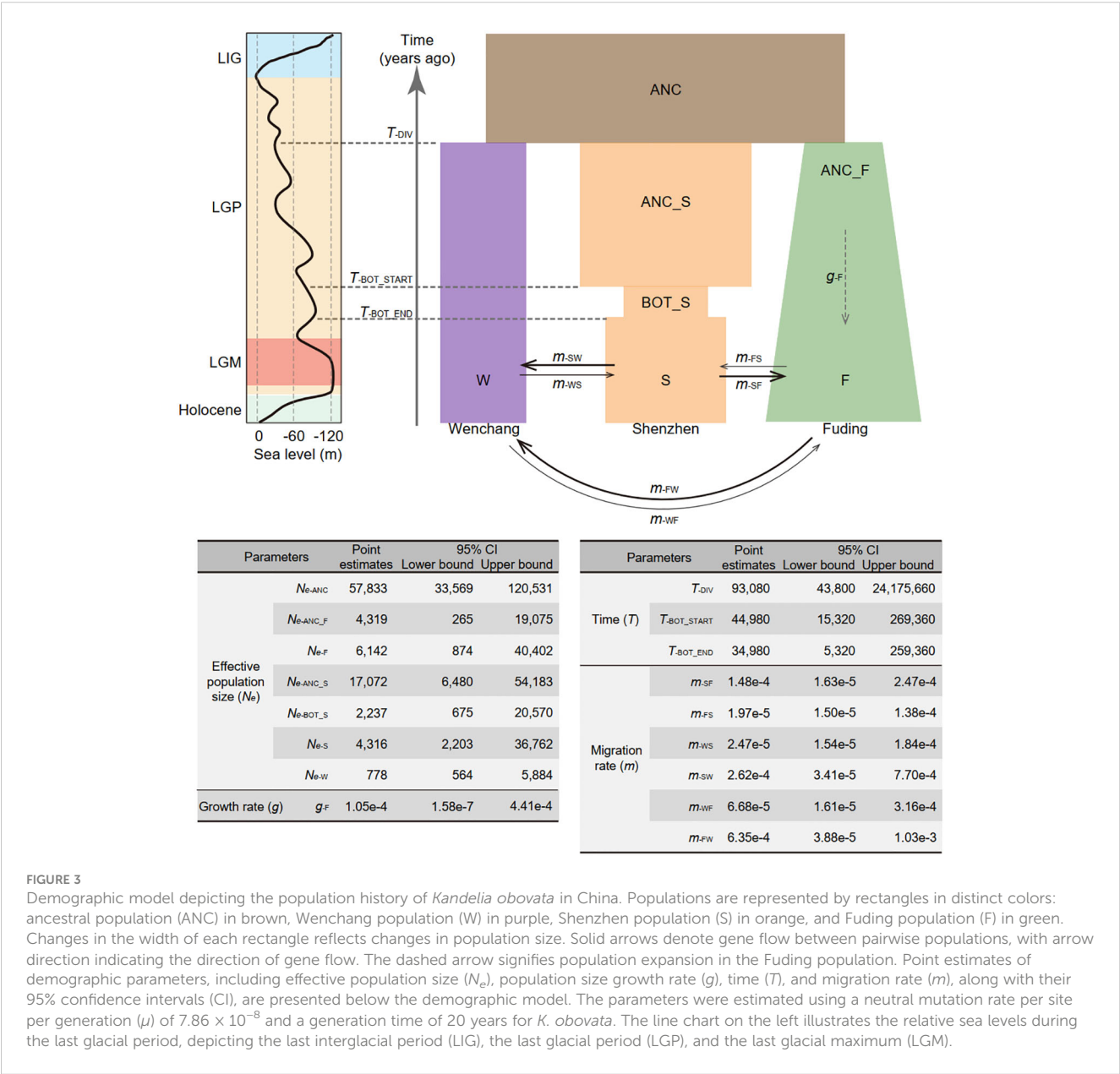
FIGURE 2

Population differentiation and genetic structure of *Kandelia obovata*. (A) Boxplot displaying the genome-wide distributions of fixed index (F_{ST}) in three pairwise populations: F-W (Fuding vs. Wenchang population), F-S (Wenchang vs. Shenzhen population), and S-W, (Wenchang vs. Shenzhen population). (B) Boxplot displaying genome-wide distributions of genetic divergence (D_{XY}) in the same three pairwise populations. Asterisks indicate the significance level of Mann-Whitney U test: ***, p -value < 0.001; "n.s.", non-significant. (C) Principal components analysis (PCA) based on 189,909 SNPs showing genetic separation among the 46 *K. obovata* samples. Principal components 1 (28.1%) and principal components 2 (23.3%) are shown. (D) Phylogenetic tree of individuals and population genetic structure. Each individual is represented by a vertical bar, which is partitioned into K ($K = 2, 3, 4$, and 5) colored segments reflecting the individual's probability of membership to each genetic cluster.

and [Supplementary Table 3](#), supports a composite history inclusive of divergence, bottleneck, expansion, and migration of *K. obovata*. With a generation time of 20 years and mutation rate of 7.86×10^{-8} per site per generation ([He et al., 2022](#)), the optimal demographic model suggests that the three *K. obovata* populations diverged from their ancestral population with an effective population size (N_e) of 57,833 (95% CI: [33,569, 120,531]) approximately 93,080 years ago (95% CI: [43,800, 24,175,600]). This coincides the late onset of the last glacial period, marked by highly unstable sea levels. Subsequently, the Fuding population underwent a gradual population expansion with a growth rate of $1.05e-4$ (95% CI: [1.58e-7, 4.41e-4]), increasing from an N_e of 4,319 (95% CI: [265, 19,075]) to 6,142 (95% CI: [874, 40,402]). In contrast, the ancestral Shenzhen population, with an N_e of 17,072 (95% CI: [265, 19,075]), experienced a ten-thousand-year bottleneck around 45,000 years ago (95% CI: [15,320, 269,360]), coinciding with the start of the last

glacial maximum. The Wenchang population maintained a consistently small value of N_e of 778 (95% CI: [564, 5,884]) over the last hundred thousand years with minimal fluctuation. Migration rates per generation between populations varied substantially, with gene flow mainly occurring from Fuding to Wenchang (m_{FW}) at a rate of $6.35e-4$ (95% CI: [3.88e-5, 1.03e-3]) and from Shenzhen to Wenchang ($m_{SW} = 2.62e-4$, 95% CI: [3.41e-5, 7.70e-4]) or Fuding ($m_{SF} = 1.48e-4$, 95% CI: [1.63e-5, 2.47e-4]), while the others were relatively small ([Figure 3](#)).

Notably, both the Shenzhen and Fuding populations have experienced changes in population size according to the best model. Despite high genetic diversity, a model that assumed a bottleneck occurred in the Shenzhen population (model 6; log-likelihood = -1091480.43, AIC = 2182978.86, $\Delta AIC = 4113.594$) was more likely than one assuming a bottleneck in the Fuding population (model 7; log-likelihood = -1091661.914, AIC =



2183341.828, $\Delta\text{AIC} = 4476.562$) (Supplementary Figure 1 and Supplementary Table 3). In contrast, although the Fuding population is marginal, a model that assumed expansion in the Fuding population (model 10; log-likelihood = -1091358.265, AIC = 2182730.53, $\Delta\text{AIC} = 3865.26$) was more likely than in the other two populations (Supplementary Figure 1 and Supplementary Table 3).

3.4 Detection of positive selection

Sliding widow analysis of population differentiation used a cutoff of at least 1.96 SD above the mean ($p < 0.025$) to detect highly differentiated outliers for both F_{ST} and XP-CLR values. For the F_{ST} values, 1452, 1457, and 1494 outliers were detected for three pairwise comparisons: Fuding and Wenchang (F-W), Fuding and Shenzhen (F-S), and Shenzhen and Wenchang (S-W), respectively (Figures 4A–C). For XP-CLR values, 125, 264, and 430 outliers were detected for the same pairwise comparisons (Figures 4A–C). The regions where both methods detected outliers were considered as the highly differentiated regions (HDRs) or the candidate regions for selection. Sliding window analysis of Fay and Wu's H -statistic (Fay & Wu, 2000), designed to detect high-frequency hitchhiking alleles associated with selective sweeps, revealed significantly lower H values in HDRs than in the whole genome for each of the three populations (Mann-Witney U test, all p -value < 0.001 ; Supplementary Figure 3), suggesting selective sweeps occurred in HDRs of each population. Negative H values were observed in HDRs of the Fuding (mean $\pm$ SD, H -statistic = -2.43 ± 3.22) and Shenzhen populations (H -statistic = -2.43 ± 2.12), but not in the Wenchang population (H -statistic = 0.13 ± 1.03 ; Supplementary Figure 3), indicating that selective signatures of HDRs are stronger in the two northern populations compared to the Wenchang population.

A total of 4, 27, and 95 HDRs were identified for the F-W, F-S, and S-W comparisons, respectively, comprising 4, 8, and 67 genes and a total of 2,868 SNPs (Figure 4D; Supplementary Table 4). Among them, 77 genes were identified in a single population pair, while one gene, *Plant U-box 45* (*PUB45*; Maker00002595), was shared by the F-W and S-W comparisons. Notably, no genes were shared by all three comparisons, resulting in a total of 78 genes in HDRs (Figure 4D). Furthermore, Gene Ontology (GO) analysis revealed that the identified genes are enriched in functional categories of supramolecular fiber organization (GO: 0097435) and those related to response to stimuli, including cellular response to extracellular stimulus (GO: 0031668), cellular response to external stimulus (GO: 0071496), response to extracellular stimulus (GO: 0009991), and cellular response to stress (GO: 0033554). Additionally, other enriched functional categories are involved in cell cycles and the regulation of developmental process (Figure 4E).

Using *K. candel* as an outgroup, we also detected genes under positive selection by the McDonald-Kreitman (MK) test (McDonald & Kreitman, 1991). Applying the G test of independence described in the original MK test, we identified 27 genes with a positive selection signature at a cutoff of an adjusted p -value < 0.05 (Supplementary Table S5). Only two genes, *AP2-like ethylene-responsive transcription factor 1* (*AIL1*; Maker00004062) participating in the ethylene-activated signaling pathway (Kim et al., 2006) and *callose synthase 5* (*CALS5*; Maker00007136)

involved in the regulation of pollen tube growth (Dong et al., 2005) were also found in the list of HDRs-related selected genes (Figure 4F). The 27 candidate genes were enriched in housekeeping functional categories including macromolecule modification (GO: 0043412), regulation of cellular process (GO: 0050794), protein modification process (GO: 0036211), and regulation of nitrogen compound metabolic process (GO: 0034641) with a significance threshold of p -value < 0.05 (Figure 4G). However, none of these overrepresentations retained significant after FDR correction, likely due to the limited size of gene list.

3.5 Genotype-environment associations

The three populations, located at different latitudes, differ in their local environments. Redundancy analysis (RDA) detected significant associations between SNPs and two environmental variables (F -statistics, both p -value < 0.001), the mean annual temperature (MAT) and the mean annual minimum temperature (MAMT), but not between SNPs and the mean annual precipitation (MAP, Supplementary Figure 4A). This suggests that temperature is the primary environmental variable shaping population differentiation. Two constrained axes (RDA1 and RDA2) explained about 29% of the total variation. Using a cutoff of 1.96 SD greater or less than mean (two-tailed p -value = 0.05), a total of 16,256 environment-associated SNPs were identified as candidate SNPs involved in local adaptation (Figure 5; Supplementary Figure 4B). The temperature-related SNPs exhibited distinct patterns of allele frequency spectrum compared to genome-wide SNPs in all three populations (Kolmogorov-Smirnov test, p -value $< 2.2 \times 10^{-16}$ for all comparisons), characterized by an elevated proportion of fixed SNPs in the Fuding and Shenzhen populations (Figure 5C). Among the 16,256 temperature-associated SNPs, 1,118 were located in the HDRs, indicating a significant enrichment of temperature-associated SNPs in HDRs compared to the whole genome (χ^2 test, p -value $< 2.2 \times 10^{-16}$; Figure 5D). These temperature-related SNPs are within 1,074 genes which were enriched in a range of functional categories involved in sexual reproduction, DNA repair, cell death, cell cycle and immune response (Supplementary Table S6).

A total of 39 genes containing temperature-associated SNPs within their exons and located within HDRs were finally identified (Figure 5E). These genes, consisting of 2, 6, and 32 for the F-W, F-S, and S-W comparisons, respectively, were considered as high-confidence candidates underlying local adaptation to temperature in *K. obovata* (Table 2). In the F-W comparison, two genes were identified, namely *Plant U-box 45* (*PUB45*; Maker00002595) and a gene (Maker00006413) containing eight temperature-associated SNPs but no homolog in *Arabidopsis* (Table 2). *Plant U-box 45*, known for its function in protein ubiquitination and cold response (He et al., 2023), was also identified as a high-confidence candidate in the S-W comparison. Notably, the high-confidence candidates in the F-S comparison with known functions included *CRY2 interacting splicing factor 1* (*CIS1*; Maker00002948) involved in the regulation of flowering (Zhao et al., 2022), *tRNA pseudouridine synthase A 5* (*TRUA5*, Maker00014866) involved in RNA modification, *Pol-like 5* (*PLL5*; Maker00017856) participating in

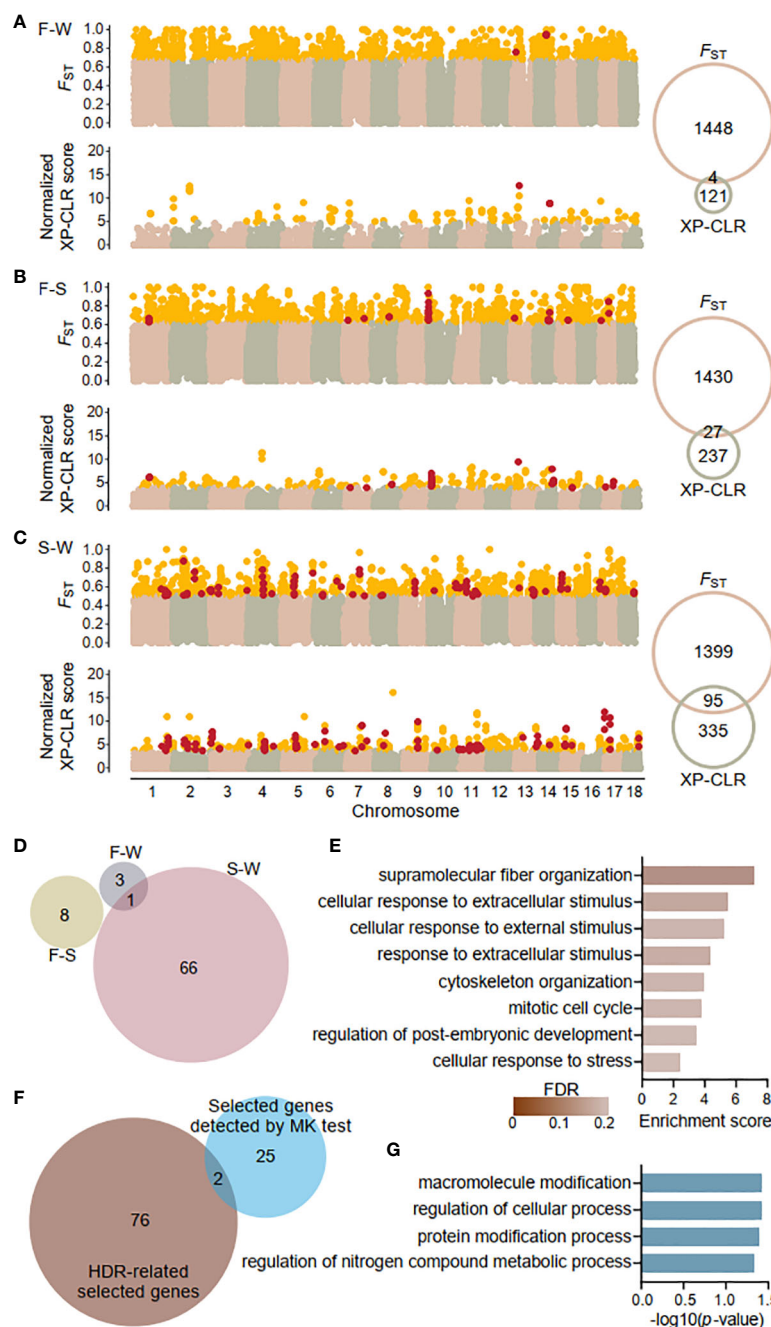


FIGURE 4

Genome wide signatures of selection and selected genes. Sliding window analysis of fixation index (F_{ST}) and cross-population composite likelihood ratio (XP-CLR) with 20-kb window size and 5-kp step size across the *K. obovata* genome for pairwise populations: (A) Fuding and Wenchang population (F-W), (B) Fuding and Shenzhen population (F-S), and (C) Shenzhen and Wenchang population (S-W). Outlier values (defined as at least 1.96 SD above the mean) are indicated in gold. Venn diagrams show the number of outliers identified by each method. Windows exhibiting both F_{ST} and XP-CLR outliers were identified as highly differentiated regions (HDRs) and indicated in red. (D) Venn diagrams showing the numbers of genes residing in HDRs that were identified in the three comparison pairs. All 78 genes identified as HDR-related selected genes are listed in [Supplementary Table S4](#). (E) Bar plot displaying the enriched Gene Ontology (GO) terms of biological process of HDR-related selected genes. (F) Venn diagrams showing the overlaps between the 78 HDR-related selected genes and the 27 selected genes identified by the McDonald-Kreitman test (MK test). Genes detected by MK test are listed in [Supplementary Table S5](#). (G) Bar plot displaying the enriched GO terms of biological process of 27 selected genes identified by the MK test.

leaf development (Song & Clark, 2005), a 2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein (Maker00003467), and a UDP-N-acetylglucosamine (UAA) transporter family gene (Maker00018193). Meanwhile, for the S-W comparison, candidate genes were found to be responsive to cold and other stresses, such as

growth-regulating factor 5 (GRF5; Maker00004032; Lantzouni et al., 2020), ATP-binding cassette G40 (ABCG40; Maker00008456; Baron et al., 2012) and Integrin-linked kinase 1 (ILK1; Maker00007966; Brauer et al., 2016), or participate in various development processes, such as *cryptochrome 1* (CRY1; Maker00010750) involved in

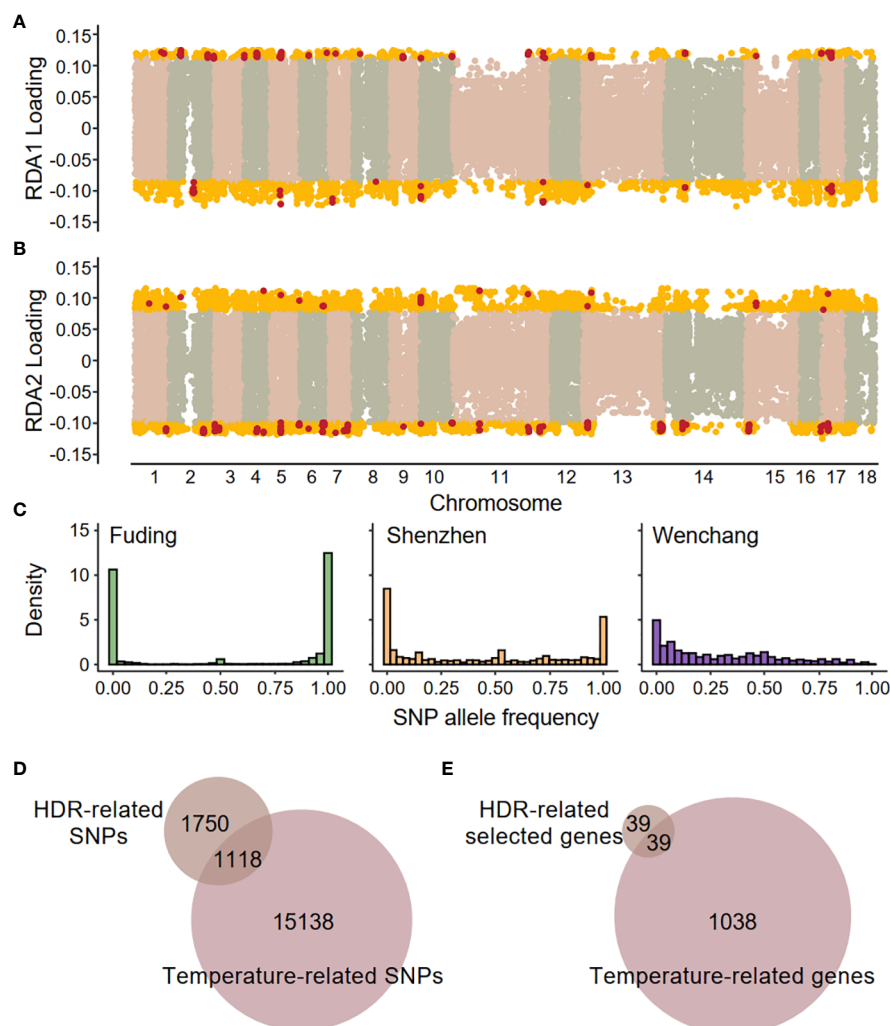


FIGURE 5

Genome-environment associations detected by redundancy analysis (RDA). SNP loadings on (A) the first RDA axis (RDA1) and (B) the second RDA axis (RDA2). The gold dots represent SNPs with significant associations along the RDA axes (defined as 1.96 standard deviations below or above the mean), and these SNPs are identified as temperature-related SNPs. The red dots represent SNPs located in highly differentiated regions (HDRs). (C) Site frequency spectra based on 16,256 temperature-related SNPs in each population. (D) Venn diagrams showing the overlaps between HDR-related SNPs and temperature-related SNPs. (E) Venn diagrams showing the overlaps between HDR-related selected genes and temperature-related genes. A total of 39 shared genes were identified as high-confidence candidate genes responsible for local temperature adaptation and are listed in Table 2.

flowering (Gao et al., 2023), *cotton Golgi-related 2* (CGR2; Maker00005232) involved in leaf morphogenesis (Kim et al., 2015) and *lonesome highway* (LHW; Maker00018105) involved in root development (Ohashi-Ito & Bergmann, 2007) (Table 2).

4 Discussion

K. obovata offers valuable opportunities for investigating the genetic basis of local adaptation in species that have undergone ecological speciation with complex demographic histories. Here, by examining how genome-wide patterns of population structure relate to genotype-environment associations, in conjunction with inferring demographic history and detecting selection signatures, we empirically contribute to the broader understanding of how

genetic variation leads to climate adaptation. Despite complex demographic history, our results demonstrate that natural selection drives local adaptation in *K. obovata* due to differential selective pressures related to temperature.

4.1 Complex demography and restricted gene flow

The complex demographic history of the *K. obovata* populations has a significant impact on their genetic diversity and population structure. The Fuding population, occurring in the northmost in China, exhibits the lowest genetic diversity and is more isolated from the other two populations. This is consistent with the expectation for marginal populations, which often show

TABLE 2 High-confidence candidate genes underlying local adaptation in *K. obovata*.

<i>K. obovata</i> gene	Populations	Number of temperature-related SNPs	Orthologs in <i>Arabidopsis</i>	Gene symbol	Description	Main GO terms of biological process
Maker00002595	F-W, S-W	2	AT1G27910	<i>PUB45</i>	Plant U-box 45	protein ubiquitination
Maker00006413	F-W	8			Carbonic anhydrase 2	carbon utilization
Maker00002948	F-S	1	AT3G52120	<i>CIS1</i>	CRY2 interacting splicing factor 1	RNA processing
Maker00003467	F-S	1	AT5G48020		2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein	
Maker00014866	F-S	1	AT5G35400	<i>TRUA5</i>	Enzyme for the pseudouridine (Ψ) to uridine (U) conversion	tRNA pseudouridine synthesis
Maker00017856	F-S	1	AT1G07630	<i>PLL5</i>	Pol-like 5, protein phosphatase 2C like gene	protein dephosphorylation
Maker00017915	F-S	1				
Maker00018193	F-S	16	AT5G59740		UDP-N-acetylglucosamine (UAA) transporter family	transmembrane transport
Maker00000528	S-W	2	AT3G02750		Protein phosphatase 2C family protein	peptidyl-threonine dephosphorylation
Maker00000994	S-W	1			Iron ascorbate-dependent oxidoreductase family	
Maker00001853	S-W	1	AT4G14300	<i>RBGD4</i>	RNA-binding glycine-rich protein D4	regulation of response to salt stress
Maker00004032	S-W	1	AT3G13960	<i>GRF5</i>	Growth-regulating factor 5	response to cold
Maker00004062	S-W	6	AT1G72570	<i>AIL1</i>	AP2-like ethylene-responsive transcription factor	ethylene-activated signaling pathway
Maker00004523	S-W	1	AT3G25905	<i>CLE27</i>	CLAVATA3/ESR-Related 27	cell-cell signaling involved in cell fate commitment
Maker00004536	S-W	1				
Maker00004832	S-W	8	AT4G02210		Myb/SANT-like DNA-binding domain protein	
Maker00005232	S-W	10	AT3G49720	<i>CGR2</i>	Cotton Golgi-related 2	leaf morphogenesis
Maker00005452	S-W	1	AT1G50060		CAP superfamily protein	biological process
Maker00005494	S-W	1	AT4G33630	<i>EX1</i>	Executer 1	response to singlet oxygen
Maker00006074	S-W	1	AT5G64380		Inositol monophosphatase family protein	fructose metabolic process
Maker00006210	S-W	1	AT1G08980	<i>AMI1</i>	Amidase 1	auxin biosynthetic process
Maker00006424	S-W	1	AT2G03480	<i>QUL2</i>	Quasimodo2 like 2	
Maker00006950	S-W	15	AT3G27350		Transcriptional regulator ATRX-like protein	
Maker00007136	S-W	3	AT2G13680	<i>CALS5</i>	Callose synthase 5	regulation of pollen tube growth
Maker00007772	S-W	1	AT1G65540	<i>LETM2</i>	Leucine zipper-EF-hand-containing transmembrane protein 2	
Maker00007966	S-W	11	AT2G43850	<i>ILK1</i>	Integrin-linked kinase 1	response to osmotic stress
Maker00008456	S-W	16	AT1G15520	<i>ABCG40</i>	ATP-binding cassette G40	response to abscisic acid

(Continued)

TABLE 2 Continued

<i>K. obovata</i> gene	Populations	Number of temperature-related SNPs	Orthologs in <i>Arabidopsis</i>	Gene symbol	Description	Main GO terms of biological process
Maker00010492	S-W	1	AT2G44970		Alpha/beta-Hydrolases superfamily protein	
Maker00010750	S-W	10	AT4G08920	<i>CRY1</i>	Cryptochrome 1	response to blue light
Maker00012336	S-W	1	AT2G35630	<i>MOR1</i>	Microtubule organization 1	cell plate formation involved in plant-type cell wall biogenesis
Maker00013529	S-W	1	AT1G04140		Transducin family protein/WD-40 repeat family protein	regulation of protein catabolic process
Maker00015311	S-W	1	AT5G42570		B-cell receptor-associated 31-like protein	intracellular protein transport
Maker00015660	S-W	1	AT1G05100	<i>MAPKKK18</i>	Mitogen-activated protein kinase kinase kinase 18	negative regulation of stomatal opening
Maker00016003	S-W	2	AT5G45360	<i>SKIP31</i>	SKP1-interacting partner 31	protein ubiquitination
Maker00016879	S-W	1	AT4G33925	<i>SSN2</i>	Suppressor of sni1 2	defense response
Maker00018105	S-W	1	AT2G27230	<i>LHW</i>	Lonesome highway	root development
Maker00018163	S-W	5	AT5G22100		RNA cyclase family protein	ribosome biogenesis
Maker00018375	S-W	1	AT4G23740		Leucine-rich repeat protein kinase family protein	
Maker00018456	S-W	1	AT4G23730		Galactose mutarotase-like superfamily protein	carbohydrate metabolic process

lower genetic diversity and increased genetic differentiation from more central populations (Eckert et al., 2008). However, the inference of the demographic model revealed that the Fuding population has experienced a mild population expansion, while the central population in Shenzhen underwent a population bottleneck around the start of the last glacial maximum (Figure 3). These divergent patterns of population size change may align with the cold-tolerant nature of *K. obovata*, with the northern population being better adapted to cold weather than the southern population (Lu et al., 2022). Additionally, these two populations exhibit distinct patterns in the site frequency spectrum compared to the Wenchang population, with the former two showing a high number of fixed SNPs while the Wenchang population has almost none (Figure 1C). These results suggest that genetic drift associated with frequent population size change in the Fuding and Shenzhen populations may have caused rapid fixation of alleles in these regions, while the southern population in Wenchang may represent the ancestor state of this species, considering that *K. obovata* is diverged from the cold intolerant *Kandelia candel* (Sheue et al., 2003). Nevertheless, it is unclear how the Wenchang population could maintain a constant population size, especially given that global cooling during the last glacial maximum is known to be a major cause of extinction for tropical woody species such as mangroves (Song et al., 2021).

It is widely acknowledged that mangroves exhibit long-distance gene flow due to the dispersal of floating propagules (Van der Stocken et al., 2022). Previous studies on *K. candel* have indicated substantial regional gene flow over considerable distances (Chiang

et al., 2001; Geng et al., 2008). In this study, the estimated gene flow from Fuding to Wenchang ($Nm_{FW} = 3.90$) is much greater than that from Wenchang to Fuding ($Nm_{WF} = 0.05$), despite the broad confidence intervals (Figure 3). This finding is consistent with the direction of current flow in the South China Sea during early winter (Shaw & Chao, 1994), coinciding with the maturation and shedding of the hypocotyl in *K. obovata* in the northern regions (Khoon et al., 2004). Thus, the asymmetric migration between the Fuding and Wenchang populations provides additional evidence for the influence of ocean currents on gene flow in mangroves. However, we observed strong population structure among the three *K. obovata* populations in China (Figure 2D), indicating local gene flow may be insufficient to counteract genetic differentiation caused by either genetic drift or natural selection. Notably, the best-fitting demographic model (Figure 3) reveals substantial gene flow occurring only from the Fuding and Shenzhen populations to the Wenchang population ($Nm_{FW} = 3.90$ and $Nm_{SW} = 1.13$), while all other Nm estimates, ranging from 0.02 to 0.64, are relatively small. Contrastingly, a recent study on grey mangrove (*Avicennia marina*) populations across the Red Sea, the Arabian Sea and the Persian/Arabian Gulf revealed a moderate population genetic structure correlating with geographic distance, which supports clades both among and within the seas surrounding the Arabian Peninsula (Friis et al., 2024).

What might have caused the limited gene flow among geographically proximate *K. obovata* populations? One possible explanation is the impact of human activities, which have led to increased fragmentation. Human activities such as land

development, urbanization, and habitat destruction can create barriers to gene flow, resulting in population isolation and reduced genetic exchange between populations (Guo et al., 2016). Moreover, the small effective population size for all populations in this study may have also played a role in limiting gene flow. With a small effective population size, genetic drift has a more pronounced effect, leading to increased genetic differentiation between populations (Song et al., 2006). This effect may be particularly significant for the Shenzhen and Fuding populations, which have experienced population bottleneck and expansion, respectively.

4.2 Local adaptation and genotype-environment associations

Local adaptation involves significant changes in allele frequency. Therefore, conducting a genome scan to identify outlier values of the allelic differentiation is the preferred method for detecting loci associated with local adaptation. It has been recognized that high neutral differentiation among populations can make it more difficult to detect high outlier loci (Pérez-Figueroa et al., 2010). This is evident in the case for *K. obovata*, where the average F_{ST} values are approximately 0.3 across the genome (Figure 2B). Utilizing the 1.96 SD cutoff of the empirical distribution, we discovered that the number of F_{ST} outliers was largely comparable for all three pairwise population comparisons, whereas the number of XP-CLR outliers varied substantially among comparisons (Figure 5A-C). Given that F_{ST} is the quotient of two variances, the large expected variability in F_{ST} among neutral loci, influenced by complex demography, population structure, and migration, may impact the power to discern high outlier F_{ST} values, potentially resulting in a high rate of false positive or false negative (Le Corre & Kremer, 2012). In contrast, the XP-CLR method, which is independent of window size and robust to uncertainty regarding demography history, identifies regions in the genome where the change in allele frequency at the locus occurred too quickly to be due to random drift (Chen et al., 2010). The different results between the two methods lies in the fact that the detection of F_{ST} outliers depends on neutral loci, while the XP-CLR method is sensitive to recent selective sweeps. By combining both methods, we observed an enrichment of environment-associated SNPs in highly differentiated regions (HDRs) that showed both outlier F_{ST} and outlier XP-CLR values, indicating the effectiveness of our analyses in identifying loci associated with local adaptation in the presence of high neutral differentiation among populations.

The three *K. obovata* populations, located at varying latitudes, are subject to differing intensities of selective pressure. Genotype-environment associations in *K. obovata* were detected exclusively for temperature-related variables, indicating that temperature is the primary factor driving differential selective pressures leading to local adaptation. This supports the notion that winter temperature shapes mangrove distributions and assemblage composition in China (Wu et al., 2018). Local temperature has been consistently identified as significant drivers of varying selective pressures in many organisms, such as grey mangrove (Friis et al., 2024), lichen-

forming fungi (Valim et al., 2023), and mountain pine (Méndez-Cea et al., 2023). In this study, the Fuding and Shenzhen populations showed a higher proportion of fixation in temperature-associated SNPs (Figure 5C) compared with genome-wide SNPs (Figure 1C), implying that natural selection has facilitated the rapid evolution of genes related to local adaptation. However, the evidence for local adaptation in the Wenchang population is weak, as reflected by the allele frequency spectra (Figure 5C) and results of Fay and Wu's H (Supplementary Figure 3), likely due to the minimal selective pressure for temperature adaptation in this tropical region.

Gene flow may have also impacted local adaptation in the Wenchang population, which receives migrants from the Fuding and Shenzhen populations ($Nm_{FW} = 3.90$ and $Nm_{SW} = 1.13$; Figure 3). Contrary to expectations based on local adaptation gradients, many more HDRs were found between the Shenzhen and Wenchang populations than between the Fuding and Wenchang populations (Figure 4C). This inconsistency could stem from more substantial gene flow from Fuding to Wenchang compared to from Shenzhen to Wenchang (Figure 3). Both theoretical and empirical studies have shown that gene flow can either promote or counteract local adaptation depending on the extent of standing variation and the strength of natural selection (Tigano & Friesen, 2016). The scant number of selected candidate genes identified between the Wenchang and Fuding populations suggests that the weak selection in the Wenchang population is insufficient to overcome the homogenizing effect of gene flow. Nevertheless, we still identified a candidate gene under selection, *PUB45*, in both the comparison between the Fuding and Wenchang populations and the comparison between the Shenzhen and Wenchang populations (Figure 4D). This gene, which belongs to ubiquitin ligases enzymes, was reported to play a role in the cold acclimation of *K. obovata* seedlings in a previous study using transcriptome analysis (He et al., 2023). Collectively, our findings highlight the complexities involved in identifying loci responsible for local adaptation, especially in the presence of complex demographic histories, population structure, and gene flow.

4.3 Similarity and discrepancy between genes involved in local adaptation and those responsible for inter-species divergence

The question of whether the same genes responsible for local adaptation within a species also contribute to differences between species is a fundamental topic in evolutionary biology. *K. obovata* and its cold-intolerant relative *K. candel* are separated by the South China Sea, exhibiting genetic discontinuity and differential adaptation (Sheue et al., 2003). Among the 78 candidate genes within HDRs in *K. obovata*, only two genes, *AIL1* and *CALS5* overlap with those identified being under selection by the McDonald-Kreitman (MK) test when comparing *K. obovata* to *K. candel* (Figure 4F). This overlap may be due to both intra- and interspecific variations being influenced by similar selective pressures in cold environments. *AIL1* has been identified as an

ethylene response factor and a transcription factor responsive to cold in grape (Ma et al., 2022). *CALS5* plays a crucial role in preserving the proper formation of callose walls during pollen development and responding to biotic stress (Dong et al., 2005). Evidence supporting the contribution of the same genetic variation to adaptive traits within and between species has been found in various organisms, such as antifreeze proteins in fish, indicating that the genetic mechanisms involved may be complex due to the polygenic nature of adaptive traits and the influence of gene-environment interactions (Berthelot et al., 2019).

The lack of overlap for the majority of candidate genes under selection, as identified by between-population comparisons with F_{ST} and XP-CLR and by inter-species comparison with MK test, can be explained by several factors. Firstly, intra-specific selection within *K. obovata* may target different genetic variations than inter-specific selection between *K. obovata* and *K. candel*. Although temperature adaptation is considered the primary force driving ecological speciation in *Kandelia*, the two species also differ in shoot, leaf, floral, fruit and hypocotyl characters, in addition to physiological differentiation (Sheue et al., 2003). The 27 selected genes detected by MK test are enriched in functional categories related to several protein modifications (Figure 4G). This is consistent with the notion that inter-specific selection often reflects more fundamental differences in niche occupation and overall lifestyle that have accrued over longer periods of evolutionary time (Tarjuelo et al., 2017). Second, the time scale of selection differs within and between species. Genes that were important in the initial divergence of the cold-tolerant *K. obovata* and cold-intolerant *K. candel* might not be the same genes that are currently under selection within *K. obovata*, which could be adapting to more recent or localized environmental changes. In line with this idea, four out of the top five enriched functional categories of the HDR candidate genes are associated with response to stimulus (Figure 4E). The discrepancies in genes identified as under selection between different methodologies may be a common issue (Balick et al., 2022; Wang X. et al., 2022; Benjelloun et al., 2023).

One limitation of this study is that only three *K. obovata* populations in China were surveyed. Considering the strong population structure among the surveyed populations (Figure 2D), we might capture both adaptive and neutral changes when identifying candidates under selection as high differential outliers between populations. Nevertheless, the enrichment of temperature-associated SNPs within HDRs indicates that most candidates are likely adaptive. On the other hand, the small number of populations surveyed may lead to an underestimation of genes responsible for local adaptation, as different environmental pressures might act on different gene interaction networks via gene-environment interactions (Tiffin & Ross-Ibarra, 2014). However, as temperature is the major factor shaping the evolutionary trajectories of *Kandelia* species, our results should capture the majority genetic variation critical for climate adaptation within and between species. A wider geographic sampling across the

complete range of *K. obovata* and *K. candel* will be necessary to conduct further research on candidate functional loci associated with their ecological divergence.

5 Conclusion

Our investigation into the local adaptation of *Kandelia obovata* populations across China provides critical insights into the genetic dynamics and adaptive evolution of mangroves under global climate change. Through whole-genome resequencing analysis, we uncovered a strong population structure with complex demographic events such as expansion, bottlenecks, and gene flow, highlighting the intricate historical context within which local adaptation has occurred. Notably, genetic differentiation is high among the geographically close *K. obovata* populations, likely due to limited gene flow as a result of human activities. Genome-wide scans of population differentiation pinpointed regions under selective sweeps, with more intense signals in northern populations. Our findings emphasize the importance of temperature in driving genetic adaptations, as opposed to precipitation, which showed no discernible genotype-environment associations. The southmost Wenchang population exhibited minimal selective sweep signatures, reflecting the low selective pressure in this tropic region, potentially confounded by gene flow from other populations. We identified 39 candidate genes with high confidence responsible for local adaptation, enriched in stimulus response functions and largely different from those genes involved in the speciation of *K. obovata* from *K. candel*, which are associated with basic cellular functions. These findings set the stage for further research to explore the molecular basis of local adaptation and resilience of mangroves to environmental stress. Such knowledge is vital for conservation strategies and predictive modeling of species responses in an era of rapid climate change.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Author contributions

CZ: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation. YW: Writing – review & editing, Software, Methodology, Investigation, Formal analysis. RZ: Writing – review & editing, Supervision, Resources, Funding acquisition, Conceptualization. TT: Writing – review & editing, Writing – original draft, Validation, Supervision, Project administration, Funding acquisition, Conceptualization.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was funded by the National Science Foundation of China (32270234, 31970245, and 32170217), Natural Science Foundation of Guangdong Province- Outstanding Youth Team Project (2023B1515040002), and Guangdong Basic and Applied Basic Research Foundation (2022A1515010571 and 2021A1515010808).

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.

References

- Alexander, D. H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19, 1655–1664. doi: 10.1101/gr.094052.109
- 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. doi: 10.1016/S0022-2836(05)80360-2
- Angert, A. L., Bontrager, M. G., and Aringren, J. (2020). What Do We Really Know about Adaptation at Range Edges? *Annu. Rev. Ecol. Evol. Syst.* 51, 341–361. doi: 10.1146/annurev-ecolsys-012120-091002
- Balick, D. J., Jordan, D. M., Sunyaev, S., and Do, R. (2022). Overcoming constraints on the detection of recessive selection in human genes from population frequency data. *Am. J. Hum. Genet.* 109, 33–49. doi: 10.1016/j.ajhg.2021.12.001
- Baron, K. N., Schroeder, D. F., and Stasolla, C. (2012). Transcriptional response of abscisic acid (ABA) metabolism and transport to cold and heat stress applied at the reproductive stage of development in *Arabidopsis thaliana*. *Plant Sci.* 188, 48–59. doi: 10.1016/j.plantsci.2012.03.001
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. B.* 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x
- Benjelloun, B., Leempoel, K., Boyer, F., Stucki, S., Streeter, I., Orozco-terWengel, P., et al. (2023). Multiple genomic solutions for local adaptation in two closely related species (sheep and goats) facing the same climatic constraints. *Mol. Ecol.* 00, e17257. doi: 10.1111/mec.17257
- Berthelot, C., Clarke, J., Desvignes, T., William Detrich, H., Flicek, P., Peck, L. S., et al. (2019). Adaptation of proteins to the cold in Antarctic fish: A role for methionine? *Genome Bio. Evol.* 11, 220–231. doi: 10.1093/gbe/evy262
- 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
- Brauer, E. K., Ahsan, N., Dale, R., Kato, N., Coluccio, A. E., Piñeros, M. A., et al. (2016). The Raf-like kinase ILK1 and the high affinity K⁺ transporter HAK5 are required for innate immunity and abiotic stress response. *Plant Physiol.* 171, 1470–1484. doi: 10.1104/pp.16.00035
- Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P., and Huerta-Cepas, J. (2021). eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Mol. Biol. Evol.* 38, 5825–5829. doi: 10.1093/molbev/msab293
- Carley, L. N., Mojica, J. P., Wang, B., Chen, C. Y., Lin, Y. P., Prasad, K. V. S. K., et al. (2021). Ecological factors influence balancing selection on leaf chemical profiles of a wildflower. *Nat. Ecol. Evol.* 5, 1135–1144. doi: 10.1038/s41559-021-01486-0
- Chen, G., Gu, X., Capinha, C., Lee, S. Y., Cui, B., Yang, F., et al. (2023). Large-scale changes in macrobenthic biodiversity driven by mangrove afforestation. *J. Appl. Ecol.* 60, 2066–2078. doi: 10.1111/1365-2664.14462
- Chen, H., Patterson, N., and Reich, D. (2010). Population differentiation as a test for selective sweeps. *Genome Res.* 20, 393–402. doi: 10.1101/gr.100545.109
- Chiang, T. Y., Chiang, Y. C., Chen, Y. J., Chou, C. H., Havanond, S., Hong, T. N., et al. (2001). Phylogeography of *Kandelia candel* in East Asiatic mangroves based on nucleotide variation of chloroplast and mitochondrial DNAs. *Mol. Ecol.* 10, 2697–2710. doi: 10.1046/j.0962-1083.2001.01399.x
- Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., et al. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* 6, 80–92. doi: 10.4161/fly.19695
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., et al. (2011). The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158. doi: 10.1093/bioinformatics/btr330
- Dong, X., Hong, Z., Sivaramakrishnan, M., Mahfouz, M., and Verma, D. P. S. (2005). Callose synthase (CalS5) is required for exine formation during microgametogenesis and for pollen viability in *Arabidopsis*. *Plant J.* 42, 315–328. doi: 10.1111/j.1365-313X.2005.02379.x
- Dool, S. E., Picker, M. D., and Eberhard, M. J. B. (2022). Limited dispersal and local adaptation promote allopatric speciation in a biodiversity hotspot. *Mol. Ecol.* 31, 279–295. doi: 10.1111/mec.16219
- Eckert, C. G., Samis, K. E., and Loughheed, S. C. (2008). Genetic variation across species' geographical ranges: The central-marginal hypothesis and beyond. *Mol. Ecol.* 17, 1170–1188. doi: 10.1111/j.1365-294X.2007.03659.x
- Ellison, A. M. (2000). Mangrove restoration: Do we know enough? *Restor. Ecol.* 8, 219–229. doi: 10.1046/j.1526-100x.2000.80033.x
- Excoffier, L., Marchi, N., Marques, D. A., Matthey-Doret, R., Gouy, A., and Sousa, V. C. (2021). Fastsimcoal2: Demographic inference under complex evolutionary scenarios. *Bioinformatics* 37, 4882–4885. doi: 10.1093/bioinformatics/btab468
- Fay, J. C., and Wu, C. I. (2000). Hitchhiking under positive Darwinian selection. *Genetics* 155, 1405–1413. doi: 10.1093/genetics/155.3.1405
- Forester, B. R., Lasky, J. R., Wagner, H. H., and Urban, D. L. (2018). Comparing methods for detecting multilocus adaptation with multivariate genotype–environment associations. *Mol. Ecol.* 27, 2215–2233. doi: 10.1111/mec.14584
- Friis, G., Smith, E. G., Lovelock, C. E., Ortega, A., Marshall, A., Duarte, C. M., et al. (2024). Rapid diversification of grey mangroves (*Avicennia marina*) driven by geographic isolation and extreme environmental conditions in the Arabian Peninsula. *Mol. Ecol.* 33 (4), e17260. doi: 10.1111/mec.17260
- Gao, J., Zhang, R., Zheng, L., Song, L., Ji, M., Li, S., et al. (2023). Blue light receptor CRY1 regulates HSFA1d nuclear localization to promote plant thermotolerance. *Cell Rep.* 42 (9), 113117. doi: 10.1016/j.celrep.2023.113117
- Geng, Q., Lian, C., Goto, S., Tao, J., Kimura, M., Islam, M. S., et al. (2008). Mating system, pollen and propagule dispersal, and spatial genetic structure in a high-density population of the mangrove tree *Kandelia candel*. *Mol. Ecol.* 17, 4724–4739. doi: 10.1111/j.1365-294X.2008.03948.x
- Guo, Z., Huang, Y., Chen, Y., Duke, N. C., Zhong, C., and Shi, S. (2016). Genetic discontinuities in a dominant mangrove *Rhizophora apiculata* (Rhizophoraceae) in the Indo-Malesian region. *J. Biogeogr.* 43, 1856–1868. doi: 10.1111/jbi.12770
- He, S., Wang, X., Du, Z., Liang, P., Zhong, Y., Wang, L., et al. (2023). Physiological and transcriptomic responses to cold waves of the most cold-tolerant mangrove, *Kandelia obovata*. *Front. Plant Sci.* 14. doi: 10.3389/fpls.2023.1069055
- He, Z., Feng, X., Chen, Q., Li, L., Li, S., Han, K., et al. (2022). Evolution of coastal forests based on a full set of mangrove genomes. *Nat. Ecol. Evol.* 6, 738–749. doi: 10.1038/s41559-022-01744-9

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/fpls.2024.1385210/full#supplementary-material>

- Hoban, S., Kelley, J. L., Lotterhos, K. E., Antolin, M. F., Bradburd, G., Lowry, D. B., et al. (2016). Finding the genomic basis of local adaptation: Pitfalls, practical solutions, and future directions. *Am. Nat.* 188, 379–397. doi: 10.1086/688018
- Hu, Y., Feng, C., Yang, L., Edger, P. P., and Kang, M. (2022). Genomic population structure and local adaptation of the wild strawberry *Fragaria nilgerrensis*. *Hortic. Res.* 9, uh059. doi: 10.1093/hr/uh059
- Hu, M. J., Sun, W. H., Tsai, W. C., Xiang, S., Lai, X. K., Chen, D. Q., et al. (2020). Chromosome-scale assembly of the *Kandelia obovata* genome. *Hortic. Res.* 7, 75. doi: 10.1038/s41438-020-0300-x
- Isabel, N., Holliday, J. A., and Aitken, S. N. (2020). Forest genomics: Advancing climate adaptation, forest health, productivity, and conservation. *Evol. Appl.* 13, 3–10. doi: 10.1111/eva.12902
- Jennerjahn, T. C., Gilman, E., Krauss, K. W., Lacerda, L. D., Nordhaus, I., and Wolanski, E. (2017). “Mangrove ecosystems under climate change,” in *Mangrove Ecosystems: A Global Biogeographic Perspective: Structure, Function, and Services*, 211–244. Eds. V. H. Rivera-Monroy, S. Y. Lee, E. Kristensen and R. R. Twilley (Cham: Springer International Publishing). doi: 10.1007/978-3-319-62206-4_7
- Khoon, G. W., Aksornkoae, S., Tri, N. H., Vongwattana, K., Fan, H., Santos, N., et al. (2004). *Mangroves in the South China Sea* (UNEP/GEF/SCS Technical Publication No. 1).
- Kim, S. J., Held, M. A., Zemelis, S., Wilkerson, C., and Brandizzi, F. (2015). CGR2 and CGR3 have critical overlapping roles in pectin methylesterification and plant growth in *Arabidopsis thaliana*. *Plant J.* 82, 208–220. doi: 10.1111/tjp.12802
- Kim, S., Soltis, P. S., Wall, K., and Soltis, D. E. (2006). Phylogeny and domain evolution in the APETALA2-like gene family. *Mol. Biol. Evol.* 23, 107–120. doi: 10.1093/molbev/msj014
- Korneliusen, T. S., Albrechtsen, A., and Nielsen, R. (2014). ANGSD: analysis of next generation sequencing data. *BMC Bioinf.* 15, 1–13. doi: 10.1186/s12859-014-0356-4
- Korunes, K. L., and Samuk, K. (2021). pixy: Unbiased estimation of nucleotide diversity and divergence in the presence of missing data. *Mol. Ecol. Resour.* 21, 1359–1368. doi: 10.1111/1755-0998.13326
- Lantzouni, O., Alkofer, A., Falter-Braun, P., and Schwechheimer, C. (2020). Growth-regulating factors interact with dellas and regulate growth in cold stress. *Plant Cell* 32, 1018–1034. doi: 10.1105/tpc.19.00784
- Le Corre, V., and Kremer, A. (2012). The genetic differentiation at quantitative trait loci under local adaptation. *Mol. Ecol.* 21, 1548–1566. doi: 10.1111/j.1365-294X.2012.05479.x
- Levin, D. A. (2019). Plant speciation in the age of climate change. *Ann. Bot.* 124, 769–775. doi: 10.1093/aob/mcz108
- 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
- Lin, P. (1999). *Mangrove ecosystem in China* (Beijing, China: Science Press).
- López-Goldar, X., and Agrawal, A. A. (2021). Ecological interactions, environmental gradients, and gene flow in local adaptation. *Trends Plant Sci.* 26, 796–809. doi: 10.1016/j.tplants.2021.03.006
- Lovelock, C. E., Cahoon, D. R., Friess, D. A., Guntenspergen, G. R., Krauss, K. W., Reef, R., et al. (2015). The vulnerability of Indo-Pacific mangrove forests to sea-level rise. *Nature* 526, 559–563. doi: 10.1038/nature15538
- Lu, W. X., Zhang, B. H., Zhang, Y. Y., and Yang, S. C. (2022). Differentiation of cold tolerance in an artificial population of a mangrove species, *Kandelia obovata*, is associated with geographic origins. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.695746
- Ma, X., Zhao, F., Su, K., Lin, H., and Guo, Y. (2022). Discovery of cold-resistance genes in *Vitis amurensis* using bud-based quantitative trait locus mapping and RNA-seq. *BMC Genomics* 23, 551. doi: 10.1186/s12864-022-08788-y
- Maxwell, G. S. (1995). Ecogeographic variation in *Kandelia candel* from Brunei, Hong Kong and Thailand. *Hydrobiologia* 295, 59–65. doi: 10.1007/BF00029111
- McDonald, J. H., and Kreitman, M. (1991). Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* 351, 652–654. doi: 10.1038/351652a0
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernysky, 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
- Méndez-Cea, B., García-García, I., Gazol, A., Camarero, J. J., de Andrés, E. G., Colangelo, M., et al. (2023). Weak genetic differentiation but strong climate-induced selective pressure toward the rear edge of mountain pine in north-eastern Spain. *Sci. Total Environ.* 858, 159778. doi: 10.1016/j.scitotenv.2022.159778
- Nei, M., and Miller, J. C. (1990). A simple method for estimating average number of nucleotide substitutions within and between populations from restriction data. *Genetics* 125, 873–879. doi: 10.1093/genetics/125.4.873
- Nguyen, L. T., Schmidt, H. A., Von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300
- Ohashi-Ito, K., and Bergmann, D. C. (2007). Regulation of the *Arabidopsis* root vascular initial population by Lonesome highway. *Development* 134 (16), 2959–2968. doi: 10.1242/dev.006296
- Oksanen, J., Legendre, P., O'Hara, B., Stevens, M. H. H., Oksanen, M. J., and Suggests, M. (2017). *Community Ecology Package*. Available at: <https://CRAN.R-project.org/package=vegan>.
- Oomen, R. A., and Hutchings, J. A. (2022). Genomic reaction norms inform predictions of plastic and adaptive responses to climate change. *J. Anim. Ecol.* 91, 1073–1087. doi: 10.1111/1365-2656.13707
- Orsini, L., Vanoverbeke, J., Swillen, I., Mergeay, J., and De Meester, L. (2013). Drivers of population genetic differentiation in the wild: Isolation by dispersal limitation, isolation by adaptation and isolation by colonization. *Mol. Ecol.* 22, 5983–5999. doi: 10.1111/mec.12561
- Pérez-Figueroa, A., García-Pereira, M. J., Saura, M., Rolán-Alvarez, E., and Caballero, A. (2010). Comparing three different methods to detect selective loci using dominant markers. *J. Evol. Biol.* 23, 2267–2276. doi: 10.1111/j.1420-9101.2010.02093.x
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., et al. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795
- Quisthoudt, K., Schmitz, N., Randin, C. F., Dahdouh-Guebas, F., Robert, E. M. R., and Koedam, N. (2012). Temperature variation among mangrove latitudinal range limits worldwide. *Trees-Struct. Funct.* 26, 1919–1931. doi: 10.1007/s00468-012-0760-1
- Shaw, P. T., and Chao, S. Y. (1994). Surface circulation in the South China Sea. *Deep-Sea Res. Pt. I: Oceanographic Research Papers* 41 (11–12), 1663–1683. doi: 10.1016/0967-0637(94)90067-1
- Sheue, C. R., Liu, H. Y., and Yong, J. W. H. (2003). *Kandelia obovata* (Rhizophoraceae), a new mangrove species from Eastern Asia. *Taxon* 52, 287–294. doi: 10.2307/3647398
- Singh, S., Rangarajan, S., and Thattai, D. (2022). Mangroves: Sea level rise and elevation - A review. *AIP Conference Proceedings*. 2515 (1), 020015. doi: 10.1063/5.0102952
- Song, S. K., and Clark, S. E. (2005). POL and related phosphatases are dosage-sensitive regulators of meristem and organ development in *Arabidopsis*. *Dev. Biol.* 285, 272–284. doi: 10.1016/j.ydbio.2005.06.020
- Song, S., Dey, D. K., and Holsinger, K. E. (2006). Differentiation among populations with migration, mutation, and drift: Implications for genetic inference. *Evolution* 60, 1–12. doi: 10.1554/05-315.1
- Song, H., Kemp, D. B., Tian, L., Chu, D., Song, H., and Dai, X. (2021). Thresholds of temperature change for mass extinctions. *Nat. Commun.* 12, 4694. doi: 10.1038/s41467-021-25019-2
- Sribianti, I., Muthmainnah, Hikmah, and Kiswandi. (2021). Economic valuation of mangrove ecosystem environmental services based on green economy. *IOP Conf. Series: Earth Environ. Sci.* 886 (1), 012116. doi: 10.1088/1755-1315/886/1/012116
- Su, J., Friess, D. A., and Gasparatos, A. (2021). A meta-analysis of the ecological and economic outcomes of mangrove restoration. *Nat. Commun.* 12, 5050. doi: 10.1038/s41467-021-25349-1
- Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123, 585–595. doi: 10.1093/genetics/123.3.585
- Tarjuelo, R., Morales, M. B., Arroyo, B., Mañosa, S., Bota, G., Casas, F., et al. (2017). Intraspecific and interspecific competition induces density-dependent habitat niche shifts in an endangered steppe bird. *Ecol. Evol.* 7, 9720–9730. doi: 10.1002/ece3.3444
- Tiffin, P., and Ross-Ibarra, J. (2014). Advances and limits of using population genetics to understand local adaptation. *Trends Ecol. Evol.* 29, 673–680. doi: 10.1016/j.tree.2014.10.004
- Tigano, A., and Friesen, V. L. (2016). Genomics of local adaptation with gene flow. *Mol. Ecol.* 25, 2144–2164. doi: 10.1111/mec.13606
- Tomlinson, P. B. (1986). *The botany of mangroves*. (Cambridge University Press). doi: 10.2307/4109956
- Valim, H. F., Grande, F. D., Wong, E. L. Y., and Schmitt, I. (2023). Circadian clock and temperature-associated genes contribute to overall genomic differentiation along elevation in lichenized fungi. *Mol. Ecol.* 33 (4), e17252. doi: 10.1111/mec.17252
- Van der Stocken, T., Vanschoenwinkel, B., Carroll, D., Cavanaugh, K. C., and Koedam, N. (2022). Mangrove dispersal disrupted by projected changes in global seawater density. *Nat. Clim. Change* 12, 685–691. doi: 10.1038/s41558-022-01391-9
- Vanwallendael, A., Soltani, A., Emery, N. C., Peixoto, M. M., Olsen, J., and Lowry, D. B. (2019). A molecular view of plant local adaptation: incorporating stress-response networks. *Annu. Rev. Plant Biol.* 70, 559–583. doi: 10.1146/annurev-arplant-050718-100114
- von Takach, B., Ahrens, C. W., Lindenmayer, D. B., and Banks, S. C. (2021a). Scale-dependent signatures of local adaptation in a foundation tree species. *Mol. Ecol.* 30, 2248–2261. doi: 10.1111/mec.15894
- von Takach, B., Penton, C. E., Murphy, B. P., Radford, I. J., Davies, H. F., Hill, B. M., et al. (2021b). Population genomics and conservation management of a declining tropical rodent. *Heredity* 126, 763–775. doi: 10.1038/s41437-021-00418-9
- Wang, J., Hu, Z., Liao, X., Wang, Z., Li, W., Zhang, P., et al. (2022). Whole-genome resequencing reveals signature of local adaptation and divergence in wild soybean. *Evol. Appl.* 15, 1820–1833. doi: 10.1111/eva.13480
- Wang, X., Hu, Y., He, W., Yu, K., Zhang, C., Li, Y., et al. (2022). Whole-genome resequencing of the wheat A subgenome progenitor *Triticum urartu* provides insights into its demographic history and geographic adaptation. *Plant Commun.* 3, 100345. doi: 10.1016/j.xplc.2022.100345

- Weir, B. S., and Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution* 38 (6), 1358–1370. doi: 10.1111/j.1558-5646.1984.tb05657.x
- Wu, Y., Ricklefs, R. E., Huang, Z., Zan, Q., and Yu, S. (2018). Winter temperature structures mangrove species distributions and assemblage composition in China. *Global Ecol. Biogeogr.* 27, 1492–1506. doi: 10.1111/geb.12826
- Yang, G., Zhou, R., Tang, T., and Shi, S. (2008). Simple and efficient isolation of high-quality total RNA from *Hibiscus tiliaceus*, a mangrove associate and its relatives. *Prep. Biochem. Biotech.* 38, 257–264. doi: 10.1080/10826060802164991
- Yu, G., Wang, L. G., Han, Y., and He, Q. Y. (2012). ClusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS: A J. Integr. Biol.* 16, 284–287. doi: 10.1089/omi.2011.0118
- Zhao, Y., Zhong, Y., Ye, C., Liang, P., Pan, X., Zhang, Y. Y., et al. (2021). Multi-omics analyses on *Kandelia obovata* reveal its response to transplanting and genetic differentiation among populations. *BMC Plant Biol.* 21, 1–14. doi: 10.1186/s12870-021-03123-1
- Zhao, Z., Dent, C., Liang, H., Lv, J., Shang, G., Liu, Y., et al. (2022). CRY2 interacts with CIS1 to regulate thermosensory flowering via FLM alternative splicing. *Nat. Commun.* 13, 7045. doi: 10.1038/s41467-022-34886-2
- Zhou, R., Zeng, K., Wu, W., Chen, X., Yang, Z., Shi, S., et al. (2007). Population genetics of speciation in nonmodel organisms: I. Ancestral polymorphism in mangroves. *Mol. Biol. Evol.* 24, 2746–2754. doi: 10.1093/molbev/msm209



OPEN ACCESS

EDITED BY

Enrique Maguilla,
Universidad Pablo de Olavide, Spain

REVIEWED BY

Mathieu Rouard,
Alliance Bioversity International and CIAT,
France
Blondy Beatriz Canto-Canche,
Centro de Investigación Científica de
Yucatán, Mexico

*CORRESPONDENCE

Eva Hřibová
✉ hribova@ueb.cas.cz

RECEIVED 16 February 2024

ACCEPTED 03 June 2024

PUBLISHED 04 July 2024

CITATION

Beránková D, Čížková J, Majzlíková G,
Doležalová A, Mduma H, Brown A,
Swennen R and Hřibová E (2024) Striking
variation in chromosome structure within
Musa acuminata subspecies, diploid
cultivars, and F1 diploid hybrids.
Front. Plant Sci. 15:1387055.
doi: 10.3389/fpls.2024.1387055

COPYRIGHT

© 2024 Beránková, Čížková, Majzlíková,
Doležalová, Mduma, Brown, Swennen and
Hřibová. 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.

Striking variation in chromosome structure within *Musa acuminata* subspecies, diploid cultivars, and F1 diploid hybrids

Denisa Beránková¹, Jana Čížková¹, Gabriela Majzlíková¹,
Alžběta Doležalová¹, Hassan Mduma², Allan Brown²,
Rony Swennen^{3,4} and Eva Hřibová^{1*}

¹Institute of Experimental Botany of the Czech Academy of Sciences, Centre of Plant Structural and Functional Genomics, Olomouc, Czechia, ²International Institute of Tropical Agriculture, Banana Breeding, Arusha, Tanzania, ³International Institute of Tropical Agriculture, Kampala, Uganda, ⁴Division of Crop Biotechnics, Laboratory of Tropical Crop Improvement, Katholieke Universiteit Leuven, Leuven, Belgium

The majority of cultivated bananas originated from inter- and intra(sub)specific crosses between two wild diploid species, *Musa acuminata* and *Musa balbisiana*. Hybridization and polyploidization events during the evolution of bananas led to the formation of clonally propagated cultivars characterized by a high level of genome heterozygosity and reduced fertility. The combination of low fertility in edible clones and differences in the chromosome structure among *M. acuminata* subspecies greatly hampers the breeding of improved banana cultivars. Using comparative oligo-painting, we investigated large chromosomal rearrangements in a set of wild *M. acuminata* subspecies and cultivars that originated from natural and human-made crosses. Additionally, we analyzed the chromosome structure of F1 progeny that resulted from crosses between Mchare bananas and the wild *M. acuminata* 'Calcutta 4' genotype. Analysis of chromosome structure within *M. acuminata* revealed the presence of a large number of chromosomal rearrangements showing a correlation with banana speciation. Chromosome painting of F1 hybrids was complemented by Illumina resequencing to identify the contribution of parental subgenomes to the diploid hybrid clones. The balanced presence of both parental genomes was revealed in all F1 hybrids, with the exception of one clone, which contained only Mchare-specific SNPs and thus most probably originated from an unreduced diploid gamete of Mchare.

KEYWORDS

oligo painting FISH, chromosome translocation, comparative cytogenetics, *Musa acuminata*, F1 hybrids

Introduction

Edible banana clones are an important trade commodity in tropical and subtropical countries and a staple food crop in Eastern Africa. They have originated by natural intraspecific and interspecific hybridization, and polyploidization in some cases. Most edible bananas originated from crosses between two wild *Musa* species, *M. acuminata* (donor of A genome) and *M. balbisiana* (donor of B genome) (Simmonds and Shepherd, 1955), with the possible contribution of other *Musa* species (Martin et al., 2020a; Sardos et al., 2022). Both species have small genomes and contain the same number of chromosomes ($2n = 22$). Diversity studies showed higher variability in genome sizes as well as higher genetic variability within *M. acuminata*, which contains several subspecies, compared to *M. balbisiana* (Janssens et al., 2016; Sardos et al., 2016a, 2020; Christelová et al., 2017). Domestication of bananas began in the Holocene, around 7,000 years BP, in Southeast Asia, when human migration brought banana species and subspecies from different regions into close proximity and therefore enabled natural crosses (Perrier et al., 2011).

Hybridization and polyploidization events during banana evolution led to the formation of parthenocarpic, clonally propagated cultivars, which are characterized by a high level of heterozygosity. Seven different translocation groups among *M. acuminata* were described by Shepherd (1999), who studied chromosome pairing during meiosis. As was recently shown, some of these large chromosomal translocations are characteristic of individual wild diploid species or even subspecies of *M. acuminata* (Baurens et al., 2019; Martin et al., 2017; Šimoníková et al., 2019; Dupouy et al., 2019; Martin et al., 2020b, 2020; Liu et al., 2023). Chromosome structural heterozygosity is common and results in aberrant chromosome pairing during meiosis and reduced or even zero production of fertile gametes (Dodds and Simmonds, 1948; Fauré et al., 1993; Shepherd, 1999; Goigoux et al., 2013). The reduced fertility of edible banana cultivars, together with differences in chromosome structure among *M. acuminata* subspecies greatly hampers the breeding of improved banana cultivars (Batte et al., 2019; Baurens et al., 2019) with lower susceptibility to diseases and pests. Diseases in commercial plantations are currently being controlled by the frequent application of fungicides; however, this treatment has a negative effect on the environment and health of banana workers (De Bellaire et al., 2010; Cordoba and Jansen, 2014).

Nowadays, the most-grown groups of bananas are triploid cultivars, including dessert bananas such as Cavendish and Gros Michel (AAA genome), cooking East African Highland bananas (AAA genome), and plantains (AAB genome). The Cavendish subgroup of bananas itself is the world's most exported fruit, with an annual global export quantity of about 20 million tons (FAO, 2023). Several recent studies focused on the characterization of parentage relationships of selected diploid and triploid edible banana clones, including the triploid sweet banana cultivar Cavendish, showing the contribution from Mchare bananas (Raboin et al., 2005; Perrier et al., 2009; Martin et al., 2023b). The Mchare cultivar subgroup forms a phenotypically distinct group of diploid (AA genome) cooking bananas, which are nowadays found only in East Africa and some East African Islands. The study of

Martin et al. (2023b) showed that the Mchare bananas contributed the unreduced $2x$ gamete to the origin of Cavendish bananas and the closely related Gros Michel cultivar subgroup (Martin et al., 2023b). The identification of Mchare bananas as the direct parents of successful cultivars highlights the importance of their use in breeding programs and warrants a more detailed analysis of their genome structure in the landraces and their hybrid offspring.

Improved triploid banana cultivars can be obtained through the development of tetraploids ($4x$), followed by the production of secondary triploid hybrids (Bakry and Horry, 1992; Tomekpe et al., 2004; Ortiz, 2013; Ortiz and Swennen, 2014; Nyine et al., 2017). Improved diploid varieties are produced by crosses of the existing cultivars with improved diploids ($2x \times 2x$ crosses). Traditional breeding requires the production of $4x$ or improved $2x$ genotypes that produce seeds, followed by re-establishing seed-sterile end products. Unfortunately, this process is hampered by the almost complete sterility of edible cultivars and flower incompatibility when stigma development does not correlate with bract lifting (Amah et al., 2021) and fast loss of pollen viability (Bayo et al., 2024). In combination with limited information on genome structural heterozygosity, the breeding process leads to very low seed sets. For instance, four seeds per Matooke (genome AAA) and only 1.6 seeds per Mchare (genome AA) can be obtained on average from one bunch (Brown et al., 2017; Batte et al., 2019).

Large genome structural changes can be identified by new sequencing technologies that span large sequence regions, e.g., mate-pairs Illumina sequences or long-read sequences produced by PacBio or Oxford Nanopore Technologies (Baurens et al., 2019; Dumschott et al., 2020; Pucker et al., 2022). Molecular cytogenetic techniques represent another option for the identification of large genome rearrangements. Localization of whole chromosome painting probes onto mitotic chromosomes *in situ* provides a powerful tool to identify large chromosomal translocations and to perform comparative analysis of chromosome structure in plants (e.g., Braz et al., 2018; Hou et al., 2018; Yu et al., 2022). Oligo-painting fluorescence *in situ* hybridization (FISH) is based on the *in silico* identification of large sets of unique oligomers specific to genome/chromosomal regions of interest (reviewed in Jiang, 2019) that can serve as probes. In bananas, chromosome-arm-specific sets of 45-nt-long sequences have already been identified and used as probes for *in situ* localization (Šimoníková et al., 2019). It has been shown that banana-specific painting probes designed based on the reference genome sequence of *M. acuminata* ssp. *malaccensis* 'DH Pahang', can be used to study genome rearrangements in other banana species and subspecies, including those that play an important role in the evolution of most edible cultivars (Šimoníková et al., 2020). Their use permitted the identification of general chromosome structures in different *M. acuminata* (A genome) subspecies, *M. balbisiana* (B genome) and *M. schizocarpa* (S genome), and most importantly, in edible banana clones (Šimoníková et al., 2020). These results showed the presence of specific chromosome structures in different species and subspecies. Furthermore, the presence of translocation events detected only in one chromosome set in some wild diploid species (e.g., *M. acuminata* ssp. *siamea* 'Pa Rayong', and *M. acuminata* ssp. *burmannica* 'Tavoy') was observed and indicated their hybrid origin (Šimoníková et al., 2020).

Our present study is focused on the characterization of large chromosomal translocations in a set of wild *M. acuminata* subspecies and edible banana cultivars that originated from natural intrasubspecific crosses. One of the main aims was to provide insight into the chromosomal evolution of diploid banana species and their cultivars and to identify variability and mode of the genome rearrangements. In addition to natural cultivars, we analyzed the chromosome structure of F1 progeny that resulted from crosses between Mchare bananas (the ancestor of Cavendish and Gros Michel) and the wild *M. acuminata* ssp. *burmannicoides* ‘Calcutta 4’ genotype. As Mchare bananas were found to be of hybrid origin, we wanted to analyze if some of the chromosome structures were preferentially transmitted to the progeny. Comparative chromosome painting revealed large variations in the genome structure within *M. acuminata* cultivars and the presence of translocation events, which were not observed in wild species analyzed so far. Our findings support previous assumptions about a more complex mode of *Musa acuminata* evolution and also show that the origin of edible banana clones was most probably accompanied by repeated introgressions and backcrosses (De Langhe et al., 2010; Martin et al., 2023a).

Materials and methods

Plant material and diversity tree construction

Most *Musa* genotypes were obtained from the International *Musa* Transit Centre (ITC, Bioversity International, Leuven, Belgium) as *in vitro* plantlets, transferred to soil, and kept in a greenhouse. Root tips of Mchare clones and their F1 hybrids were collected and fixed from the plants stored at a field collection of the International Institute of Tropical Agriculture (IITA), NM-AIST, Arusha, Tanzania. The accessions used in the current study are listed in Table 1.

The genetic diversity was analyzed using a standardized SSR genotyping platform (Christelová et al., 2011). To achieve higher resolution, the studied set of wild *M. acuminata* accessions and diploid clones was supplemented with selected accessions from our previous studies (Šimoníková et al., 2019, 2020).

Using a set of M13-tailed fluorescent-labeled primers, 19 highly polymorphic SSR loci were amplified, and the allele sizes were measured on the ABI 3730xl DNA analyzer (Applied Biosystems, Foster City, CA, USA), followed by the data analysis using GeneMarker v1.75 (Softgenetics, State College, PA, USA) (Christelová et al., 2011). Dendrograms of selected *M. acuminata* species and cultivated clones were constructed using the Neighbor-Net inference method in the SplitsTree4 program (Huson and Bryant, 2006). Neighbor-Net constructs phylogenetic networks to visualize distance data to show evolutionary relationships and conflict in the data (Bryant and Huson, 2023) by use of the split decomposition method (Bandelt and Dress, 1992a, b).

Oligo-painting FISH

To characterize chromosome structure within the *Musa* accessions, chromosome-arm-specific painting probes developed by Šimoníková et al. (2019) were used. Individual chromosome arms synthesized as immortal libraries by Arbor Biosciences (Ann Arbor, Michigan, USA) were labeled directly by CY5 fluorochrome or by digoxigenin or biotin, according to Han et al. (2015), with minor modifications: the oligomer libraries were amplified using debubbling PCR according to Immortal Labelling Protocol v2.2. (Daicel Arbor Biosciences; <https://arborbiosci.com/>) instead of the emulsion PCR.

Please note that in the reference genome assembly of *M. acuminata* ‘DH Pahang’ (D’Hont et al., 2012), which was originally used to develop banana-specific chromosome-arm painting probes, pseudomolecules 1, 6, and 7 are oriented inversely to the traditional way karyotypes are presented, where the short arms are on the top.

TABLE 1 List of *Musa* accessions analyzed in this work.

Species	Subspecies/ subgroup	Accession name	ITC code ^a	Genbank DOI	Genomic constitution	Chromosome number (2n)
<i>M. acuminata</i>	<i>banksii</i>	‘Banksii’	0341	10.18730/9JSM\$	AA	22
	<i>banksii</i>	‘Higa’	0428	10.18730/9JXPG	AA	22
	<i>banksii</i>	–	0896	10.18730/9KSY*	AA	22
	<i>malaccensis</i>	–	1886	10.18730/P5GEA	AA	22
	<i>malaccensis</i>	–	1887	10.18730/SAK1W	AA	22
	<i>zebrina</i>	‘Zebrina’	1139	10.18730/9M6P~	AA	22
	<i>siamea</i>	‘Khao (Phrae)’	0660	10.18730/9KCD6	AA	22
Cultivars	Unknown	‘Tuu Gia’	0610	10.18730/9K95D	AA	22
	–	‘Himone’	0886	10.18730/9KS68	AA	22
	–	‘Maleb’	0809	10.18730/9KKRK	AA	22
	–	‘Marakudu’	1210	10.18730/9MB6X	AA	22

(Continued)

TABLE 1 Continued

Species	Subspecies/ subgroup	Accession name	ITC code ^a	Genbank DOI	Genomic constitution	Chromosome number (2n)
	–	‘Vudu Beo’	1211	10.18730/9MB8Z	AA	22
	Sucrier	‘Mai’a hapai’	1172	10.18730/9M8EF	AA	22
	Rose	‘Rose’	0712	10.18730/9KER7	AA	22
	Mchare	‘Kahuti’	–	–	AA	22
	Mchare	‘Mchare mlelembo’	–	–	AA	22
	Mchare	‘Mchare laini’	–	–	AA	22
F1 hybrids						
Mchare x ‘Calcutta 4’		‘NM275–4’ (Mchare laini × Calcutta 4)	–	–	AA	22
		‘NM258–3’ (Mchare laini × Calcutta 4)	–	–	AA	22
		‘NM209–3’ (Mchare laini × Calcutta 4)	–	–	AA	22
		‘NM237–8’ (Ijihu Inkudu × Calcutta 4)	–	–	AA	22
		‘T.2269–1’ (Huti white × Calcutta 4)	–	–	AA	22
		‘T.2274–6’ (Huti white × Calcutta 4)	–	–	AA	22
		‘T.2274–9’ (Huti white × Calcutta4)	–	–	AA	22
		‘T.2619–15’ (Mchare mlelembo × Calcutta 4)	–	–	AA	22

^aCode assigned by the International Transit Centre (ITC, Leuven, Belgium).

The orientation of chromosomes in the present study corresponds with the traditional way karyotypes are presented, as was depicted in our previous study (Šimoníková et al., 2019). The “L” stands for the long arms of chromosomes, and the “S” stands for the short arm of chromosomes in the entire manuscript.

Mitotic metaphase chromosome spreads were prepared according to Šimoníková et al. (2020) from root meristems using the dropping method of protoplast suspension described by Doležel et al. (1998). Fluorescence *in situ* hybridization and image analysis were performed as mentioned previously (Šimoníková et al., 2019). A hybridization mixture containing 50% (v/v) formamide, 10% (w/v) dextran sulfate in 2 × SSC, and 10 ng/μL of labeled probes was added onto a slide and denatured for 90 s at 80°C, followed by overnight hybridization performed in a humid chamber at 37°C. The sites of digoxigenin- and biotin-labeled probes were detected using anti-digoxigenin-FITC (Roche Applied Science, Penzberg, Germany) and streptavidin-Cy3 (ThermoFisher Scientific/Invitrogen, Carlsbad, CA, USA), respectively. The stringent washes, detection of probe signals, and final chromosome counterstaining with DAPI and mounting of the preparations in Vectashield Antifade Mounting Medium (Vector Laboratories, Burlingame, CA, USA) were performed according to Beránková and Hříbová (2023).

Microscopic and image analysis

The slides were examined with an Axio Imager Z.2 Zeiss microscope (Zeiss, Oberkochen, Germany) equipped with a Cool Cube 1 camera (Metasystems, Altlussheim, Germany) and appropriate optical filters and a PC with ISIS software 5.4.7 (Metasystems, Altlussheim, Germany). The final image adjustment and creation of idiograms were done in Adobe Photoshop CS5. Different probe combinations hybridizing on a minimum of 10 preparations with mitotic metaphase chromosome spreads were used for the final karyotype reconstruction of each genotype.

Illumina sequencing and data analysis

Genomic DNA was isolated with the NucleoSpin PlantII kit (Macherey-Nagel, Düren, Germany) according to the manufacturer’s recommendations and further sheared by Bioruptor Plus (Diagenode, Liège, Belgium) to achieve an insert size of about 500 bp. Libraries for sequencing were prepared from 2 μg of fragmented DNA using the TruSeq® DNA PCR-free kit (Illumina) and sequenced on a NovaSeq 6000 (Illumina, San

Diego, CA, USA), producing 2×150 -bp paired-end reads to achieve a minimal sequence depth of $25\times$. Raw data were trimmed for low-quality bases and adapter sequences and to the same length using fastp v.0.20.1 (Chen et al., 2018).

Analysis of the proportion of individual parental subgenomes in the F1 hybrid clones was done using the vcfHunter pipeline (<https://github.com/SouthGreenPlatform/vcfHunter>), according to Baurens et al. (2019). Briefly, trimmed reads were aligned to reference the genome sequence of *M. acuminata* ssp. *malaccensis* ‘DH Pahang’ v4 (Belser et al., 2021) by BWA-MEM v0.7.15 (Li, 2013), followed by removing redundant reads using MarkDuplicate from Picard Tools v2.7.0, and locally realigned around indels using the IndelRealigner tool of the GATK v3.3 package (McKenna et al., 2010). Bases with a mapping quality of ≥ 10 were counted using the process_reseq_1.0.py python script (<https://github.com/SouthGreenPlatform/vcfHunter>). Variant calling and SNP filtering steps were performed according to Baurens et al. (2019) using the VcfPreFilter.1.0 python script (alleles supported by at least three reads and with a frequency 0.25 were kept as variant) and the vcfFilter.1.0.py python script (< 6 -fold coverage for the minor allele were converted to missing data) (<https://github.com/SouthGreenPlatform/vcfHunter>). Finally, the proportion of parental genomes in the F1 hybrid clones along the individual chromosomes of the reference genome sequence was called using biallelic SNPs (SNPs specific to Mchare cultivars and *M. acuminata* spp. *burmannicoides* ‘Calcutta 4’) in CDS genome regions using vcf2allPropAndCov.py and vcf2allPropAndCovByChr.py python

scripts (<https://github.com/SouthGreenPlatform/vcfHunter>), according to Baurens et al. (2019).

Results

To enlarge the knowledge of general chromosome structure in *Musa acuminata* and to shed light on the evolution of *M. acuminata*, we provided cytogenetic analysis in 25 diploid *M. acuminata* accessions, including wild species and natural cultivars. We also analyzed chromosome structure in F1 hybrids obtained from ‘Mchare’ $\times$ *M. acuminata* ssp. *burmannicoides* ‘Calcutta 4’ crosses, to shed light on the transfer of Mchare chromosomes differing in their structures in comparison to ‘Calcutta 4’—chromosomes 1, 3 and 8—and corresponding reshuffled chromosome structures: reciprocal translocations between long arms of chromosomes 1 and 4 (1L/4L, 4L/1L) and Robertsonian translocation between chromosome 3 and 8 (3S/8L and 8S/3L).

Evolutionary relationships within *M. acuminata*

To assess evolutionary relationships among selected wild species and cultivars of *M. acuminata*, the SSR genotyping data were used to create a phylogenetic network (Figure 1). Neighbor-

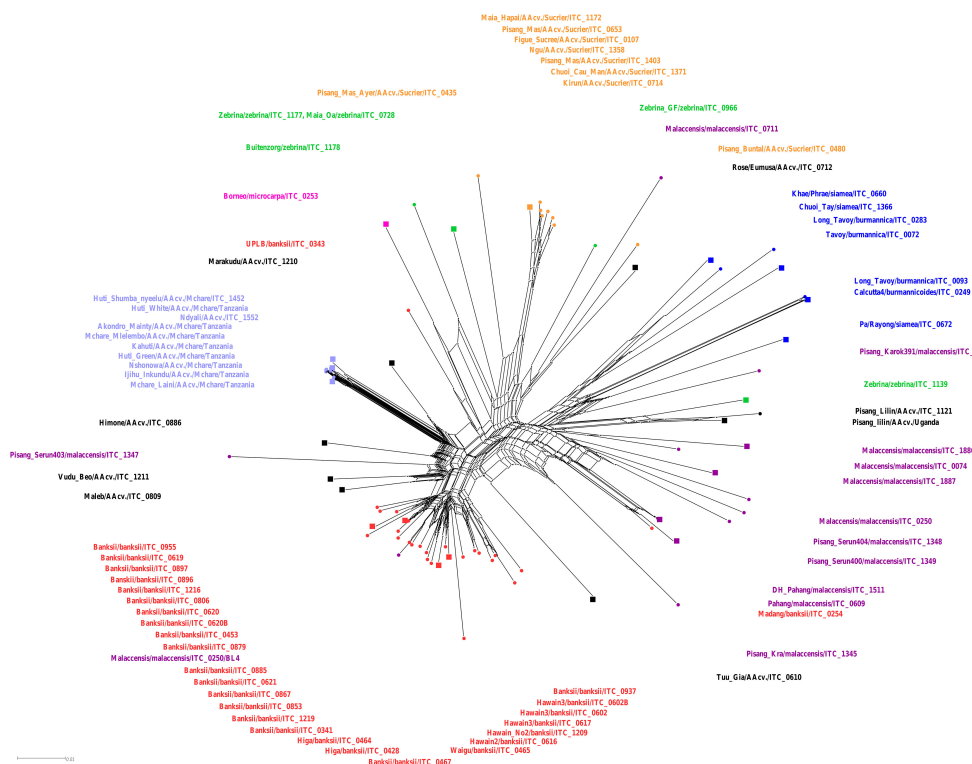


FIGURE 1

Neighbor-Net analysis of *Musa* accessions performed by SplitsTree. Accessions representing different *M. acuminata* subspecies and groups of cultivars are depicted in colors: *banksii* ssp. in red; *malaccensis* ssp. in violet; *burmannica/burmannicoides/siamea* in blue; *zebrina* in green; *microcarpa* in pink; Mchare genotypes in light blue; Sucrier genotypes in orange; and other analyzed banana AA cultivars in black. The accessions that were used for chromosome painting are depicted as squared nodes in the Neighbor-Net tree.

Net of wild *acuminata* subspecies and selected natural diploid hybrids resulted in split-separated populations of ssp. *banksii* from ssp. *burmannica/siamea* group, ssp. *malaccensis* group, and Mchare and Sucrier (Pisang Mas) groups of cultivars. Other analyzed cultivars were clustered in close proximity with the closely related *acuminata* subspecies (Figure 1). *M. acuminata* ssp. *microcarpa*, which was represented only by one accession ('Borneo'), is closely related to *M. acuminata* ssp. *zebrina* ('Maia Oa', 'Zebrina', and 'Buitenzorg' accessions). Some of the accessions did not cluster together with their presumed relatives. For instance, 'Madang' cv. (ITC0254) and 'Zebrina' ITC1139 clustered together with *malaccensis* accessions; 'Zebrina GF' (ITC0966) and 'Malaccensis' ITC0711 were included within the Sucrier group. The 'Malaccensis' ITC0250/BL4 was most probably a mislabeled sample because it clustered within the *banksii* clade. 'Pisang Serun' (*malaccensis*; ITC1347), together with 'Himone', 'Vudu Beo', and 'Maleb' cultivars, shared splits with the Mchare group of accessions, signifying their close relationships (Figure 1). Integration of clones representing F1 progeny obtained after Mchare × 'Calcutta 4' (ssp. *burmannicoides*) crosses did not change the position and composition of the other group of *acuminata* subspecies (Supplementary Figure S1). Most F1 progeny occupied one distinct clade alongside Mchare cultivars. The presence of 'Calcutta 4' accession within the F1 hybrids suggests successful hybridization. Two representatives of F1 progeny (NM237-8 and NM237-1) clustered in close proximity to 'Borneo', and 'NM209-14' F1 hybrid clustered with 'Rose', which can indicate that they are not successful hybrids with *M. acuminata* 'Calcutta 4', and could be mislabeled (Supplementary Figure S1).

Karyotype structure of *M. acuminata*

To perform comparative karyotyping and reveal chromosome structural changes within *M. acuminata* accessions, oligo-painting FISH was used (Figures 2A–M; Supplementary Figure S2). The painting probes were originally designed for the reference genome sequence of *M. acuminata* ssp. *malaccensis* 'DH Pahang' (D'Hont et al., 2012). Therefore, all chromosome structural rearrangements mentioned in the study are described based on comparison to the standard chromosome set of the 'DH Pahang' reference genome sequence, according to Šimoníková et al. (2019).

The chromosome painting of two *malaccensis* accessions (ITC1886 and ITC1887) revealed the presence of reciprocal translocation of short segments of the long arms of chromosomes 4 and 1, which was present in the homozygous state (Figure 2A). The same type of translocation was also revealed in a heterozygous state in the edible cultivars 'Vudu Beo', cv. 'Rose' and 'Mai'a hapai' (Supplementary Figures S2A–C). Cultivar 'Rose' contained additional chromosomal structural changes, which were observed in one chromosome set. Translocation of the short arm of chromosome 7 to the long arm of chromosome 1, which resulted in the formation of a small telocentric chromosome consisting of only 7L (Supplementary Figure S2B). Cultivar 'Mai'a hapai', which is a representative of Sucrier subgroup, also contained additional large rearrangements between chromosomes 1L and 7, again in the

heterozygous state. This rearrangement led to the formation of a recombined chromosome containing the short segment of 7L and short arm of 7S translocated to 1L and a telocentric chromosome containing short segment of chromosome 1L and 7S and the long arm of chromosome 7 (Figure 2B; Supplementary Figure S2C). In addition, cultivar 'Mai'a hapai' contained Roberstonian translocation between chromosomes 3 and 8 in the homozygous state (Figures 2C–E). Interestingly, one genotype is described as ssp. *zebrina* (Zebrina, ITC1139) had the same chromosome structure as the cultivar 'Mai'a hapai' (Supplementary Figure S2C).

The karyotype analysis of *M. acuminata* ssp. *siamea* 'Khae (Phrae)' showed the presence of translocations between chromosome arms 1L and 9S and 2L and 8L (Figures 2F, G), which were described previously in closely related accessions of *burmannica/burmannicoides/siamea* (Šimoníková et al., 2019). The genome of 'Khae (Phrae)' contained additional translocations, which included chromosomes 7 and 8. A recombined chromosome containing a short arm of chromosome 7, a short segment of 7L, and a short arm of chromosome 8 was found in the homozygous state (Figures 2I, J; Supplementary Figure S2D). The reciprocal recombined chromosome, containing a short arm of chromosome 7, a short segment of 7L, and the long arm of chromosome 8 was found in the heterozygous state, as was the recombined chromosome, consisting of a segment of chromosome 10L, a long arm of chromosome 7, a short segment of 7S, and a long arm of chromosome 8 (Figure 2H; Supplementary Figure S2D).

No chromosome rearrangement was detected in the two representatives of the *banksii* subspecies (ITC0341 and ITC0896) as compared to the 'DH Pahang' reference banana genome (Supplementary Figure S2E). On the other hand, a small segment of the short arm of chromosome 9 was inserted into the long arm of chromosome 5 near the centromeric region in another *banksii* representative, 'Higa' (Supplementary Figure S2F). No chromosome rearrangements were observed in the edible cultivar 'Marakudu' (Supplementary Figure S2E). Edible banana cultivars 'Himone' and 'Maleb' share the same chromosome structures, containing one reciprocal translocation between the short arm of chromosome 3 and the long arm of chromosome 8, in a heterozygous state (Supplementary Figure S2G), which was previously identified in *zebrina* subspecies (Šimoníková et al., 2020). Cultivar 'Tuu Gia' comprised reciprocal translocation between chromosomes 1L and 9S and translocation between 2L and 8L, which were also present in *burmannica/burmannicoides/siamea* accessions (Supplementary Figure S2H; Šimoníková et al., 2020). Another chromosome rearrangement involved reciprocal translocation between chromosomes 7 and 8, which led to the formation of chromosome structures consisting of the long arm of chromosome 7, a small segment of chromosome 8S, and the long arm of chromosome 8, and to a recombined chromosome consisting of the short arm of chromosome 7 and a large segment of the short arm of chromosome 8. All of these translocations were observed in a heterozygous state (Supplementary Figure S2H).

The three Mchare banana cultivars analyzed in our present study (Mchare mlelembo, Mchare Laini, and Kahuti) had the same genome structure as previously analyzed cultivars belonging to the Mchare group (Šimoníková et al., 2020). All Mchare genotypes

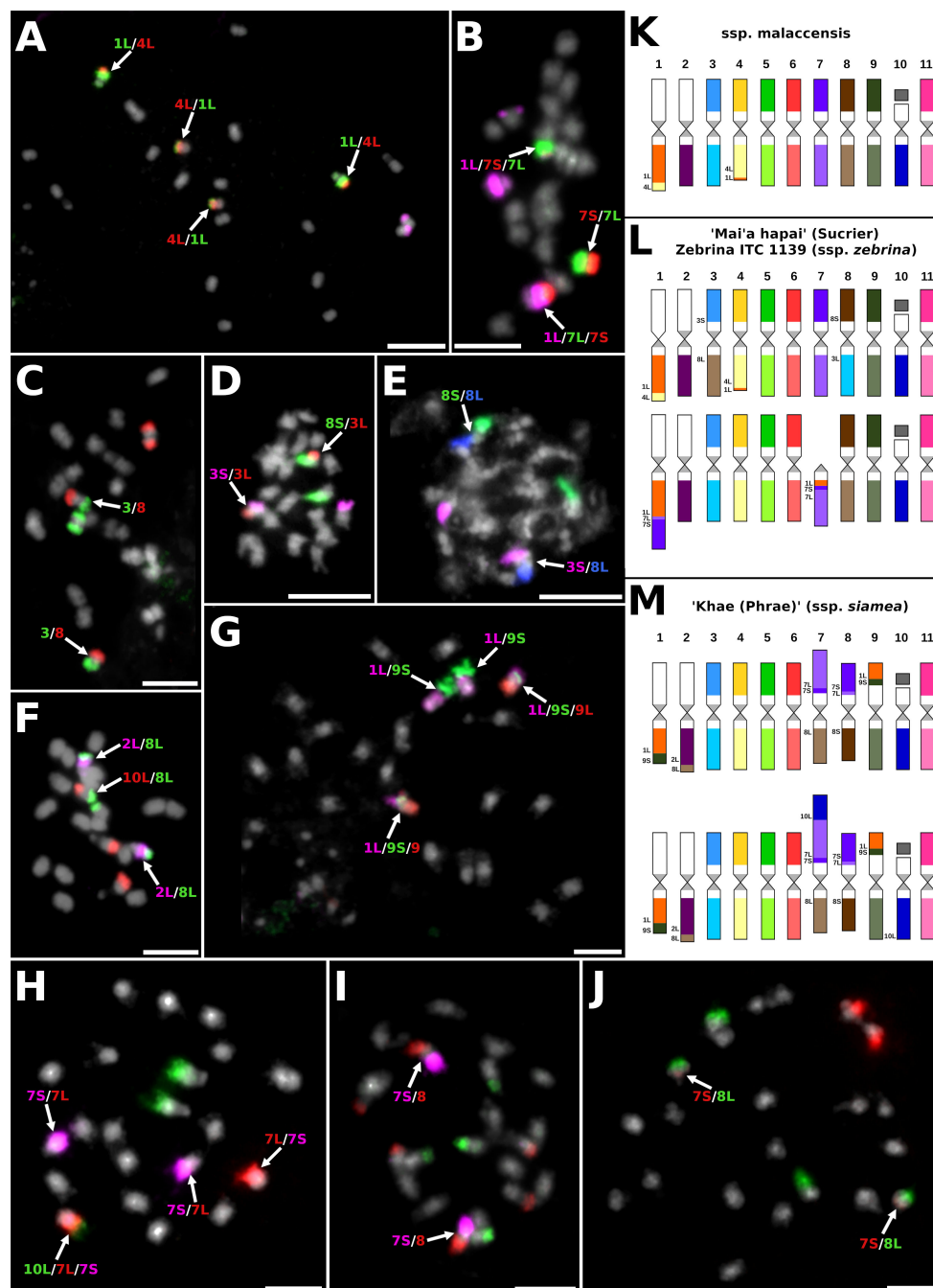


FIGURE 2

Examples of chromosome translocations identified by oligo-painting FISH on mitotic metaphase plates of *M. acuminata* subspecies and their natural hybrids: (A) ITC1886 *M. acuminata* ssp. *malaccensis* ($2n = 2x = 22$), probes for long arm of chromosome 1, entire chromosome 4, and short arm of chromosome 9 labeled in green, red, and purple, respectively. (B) ITC1172 'Mai'a hapai' (AA, subgr. Sucrier, $2n = 2x = 22$), probes for long arm of chromosome 1, short arm of chromosome 4, and long arm of chromosome 4 were labeled in purple, red, and green, respectively. (C) ITC1172 'Mai'a hapai', probes for chromosomes 3 and 8 were labeled in green and red, respectively. (D) ITC1172 'Mai'a hapai', probes for short arm of chromosome 3, long arm of chromosome 3, and short arm of chromosome 8 were labeled in purple, red, and green, respectively. (E) ITC1172 'Mai'a hapai', probes for short arm of chromosome 3, short arm of chromosome 8, and long arm of chromosome 8 were labeled in purple, green, and blue pseudocolor. (F) ITC0660 'Khao (Phrae)' (ssp. *siamea*, $2n = 2x = 22$), probes for long arm of chromosome 2, long arm of chromosome 8, and long arm of chromosome 10 were labeled in purple, green, and red, respectively. (G) ITC0660 'Khao (Phrae)', probes for long arm of chromosome 1, short arm of chromosome 9, and long arm of chromosome 9 were labeled in purple, green, and red, respectively. (H) ITC0660 'Khao (Phrae)', probes for long arm of chromosome 10, short arm of chromosome 7, and long arm of chromosome 7 were labeled in green, purple, and red, respectively. (I) ITC0660 'Khao (Phrae)', probes for short arm of chromosome 7, short arm of chromosome 8, and long arm of chromosome 9 were labeled in purple, red, and green, respectively. (J) ITC0660 'Khao (Phrae)', probes for short arm of chromosome 7 and long arm of chromosome 8 were labeled in red and green, respectively. Chromosomes were counterstained with DAPI (light grey pseudocolor). Arrows point to chromosomes with translocations. Bars = 5 μ m. Idiograms of four *Musa acuminata* representatives ($2n = 2x = 22$): (K) *M. acuminata* ssp. *malaccensis* (ITC1886 and ITC1887); (L) *M. acuminata* ssp. *siamea* 'Khao (Phrae)' (ITC0660); and (M) clone 'Mai'a hapai' (Sucrier; ITC1172). Chromosomes are oriented with their short arms on the top and long arms on the bottom in all idiograms, and translocated parts of the chromosomes contain extra labels.

contained two reciprocal translocations involving chromosomes 4 and 1 and chromosomes 3 and 8. Both reciprocal translocations were found in the heterozygous state (Supplementary Figure S2I). All chromosome changes (translocations) observed in our present study are listed in Supplementary Table S1.

Karyotype structure of Mchare × ‘Calcutta 4’ progeny

The chromosome painting was also used to study the genome structure of eight F1 hybrids, which were obtained by crosses between different Mchare genotypes (female parent) and the wild species *M. acuminata* ssp. *burmannicoides* ‘Calcutta 4’ (male parent) (Figures 3A–N; Supplementary Figure S2J). As it was revealed previously (Šimoníková et al., 2020), the genome of *M. acuminata* ssp. *burmannicoides* ‘Calcutta 4’ contains a specific chromosome structure for two pairs of chromosomes that differ from the chromosome structure found in Mchare genotypes. Based on this information, the F1 progeny of crosses between Mchare and ‘Calcutta 4’ should contain one chromosome set inherited from the Mchare cultivar and one chromosome set inherited from ‘Calcutta 4’. Thus, the hypothetical genome composition of such F1 hybrid clones can be represented by 16 different combinations of chromosomes varying in their structure between the parental genotypes (Supplementary Figure S3).

The genome of seven F1 hybrid clones contained one set of translocation chromosomes specific to ‘Calcutta 4’, and any of these seven F1 hybrid clones contained both sets of translocation chromosomes specific to Mchare (Figures 3A–D, G–N; Supplementary Figure S2J). The presence of any translocated chromosome specific to ‘Calcutta 4’ was not observed after chromosome painting in one F1 hybrid clone ‘NM237–8’ (Figures 3E, F). Oligo-painting FISH of this F1 hybrid clone, which had arisen from a cross between ‘Ijihu Inkudu’ (Mchare type) and ‘Calcutta 4’, resulted in the same genome structure as Mchare clones (Figure 3M; Supplementary Figure S2I). One F1 hybrid clone (‘NM275–4’) inherited 1L/4L and 4L/1L translocation chromosomes from the Mchare genome (Figures 3A, K), and five of the remaining F1 hybrids (‘NM258–3’, ‘NM209–3’, ‘T.2274–6’, ‘T.2274–9’, and ‘T.2274–15’) inherited 3S/8L and 8S/3L recombined chromosomes from the Mchare genome (Figures 3H, I, L, N). A short segment of chromosome 9L was inserted into the long arm of chromosome 6 near the centromeric region in ‘NM275–4’ and ‘NM258–3’ F1 hybrid clones (Figures 3C, K, L). None of the recombined chromosomes specific to Mchare was transmitted to the F1 hybrid clone ‘T.2269–1’ (Supplementary Figure S2J).

Genome constitution of Mchare × ‘Calcutta 4’ F1 hybrids

To further complete information on the genome composition of analyzed F1 hybrid clones, we performed Illumina resequencing of the F1 hybrids and their parental genomes. The proportion of parental genomes in the F1 progeny was identified by the VcfHunter program

pipeline (Baurens et al., 2019). SNPs specific to parental genomes were depicted based on the alignment of Illumina reads to reference the genome sequence of *M. acuminata* ssp. *malaccensis* ‘DH Pahang’. The coverage ratio of parental-specific SNPs along 11 chromosomes of the reference genome sequence (Supplementary Table S2) showed that most F1 hybrid clones contained whole haploid sets of chromosomes representing the individual parents (Figure 4A; Supplementary Figure S4). The only exception was revealed for the clone ‘NM237–8’, which contained only SNPs specific to Mchare (Figure 4B) and any chromosome region specific to *M. acuminata* ‘Calcutta 4’ was not revealed. This observation corresponds with the results of chromosome painting, which also did not detect any chromosome carrying ‘Calcutta 4’-specific chromosome structure (Figures 2C, J).

Discussion

Complex phylogenetic relationships within *Musa* and edible banana cultivars shown by Neighbor-Net analysis correspond to previous studies (e.g., Perrier et al., 2011; Janssens et al., 2016; Christelová et al., 2017; Sardos et al., 2022; Martin et al., 2023a). Some of the analyzed accessions, mostly those with hybrid origins, which were further confirmed by oligo-painting, showed unexpected/conflicting positions in the Neighbor-Net graph. The high level of admixture and mosaic genome structure of cultivated banana clones originating from hybridizations between subspecies of *M. acuminata* was proposed recently by Martin et al. (2020a; 2023a, b). Reticulate evolution, caused by polyploidy and hybridization (i.e., allopolyploidy), can promote rapid diversification of numerous plant lineages (reviewed in Stull et al., 2023).

Ancient polyploidization and hybridization events are known to be the major drivers of plant diversification and speciation. These processes lead to the multiplication of chromosome sets in genomes or “genome upsizing” (Ibarra-Laclette et al., 2013; Wendel, 2000), followed by the postpolyploid diploidization processes, which are thought to be associated with extensive loss of DNA or “genome downsizing”, and structural chromosomal changes (Schubert and Vu, 2016; Mandáková et al., 2017; Wang et al., 2021; Farhat et al., 2023). Until now, this phenomenon has been studied in detail mainly in the *Poaceae* and *Brassicaceae* families (e.g., Salse, 2016; Mandáková et al., 2017; Mandáková and Lysak, 2018). Recent developments in oligo-painting FISH and long sequencing technologies that enable the production of chromosome-scale genome sequences even in nonmodel species permit the study of this phenomenon in other plant species.

In the current study, we analyzed and compared the chromosome structure in *M. acuminata* and their natural and artificial hybrids by chromosome-arm-specific oligo-painting FISH, which facilitates revealing large chromosome translocations (Šimoníková et al., 2019, 2020). Genetic linkage between different chromosome-specific markers suggested the presence of five different reciprocal chromosome translocations (between chromosomes: 1 and 4; 2 and 8; 1 and 9; 1 and 7; and 3 and 8) in *M. acuminata* subspecies and their hybrids (Martin et al., 2017; Dupouy et al., 2019; Martin et al., 2020b). These and additional

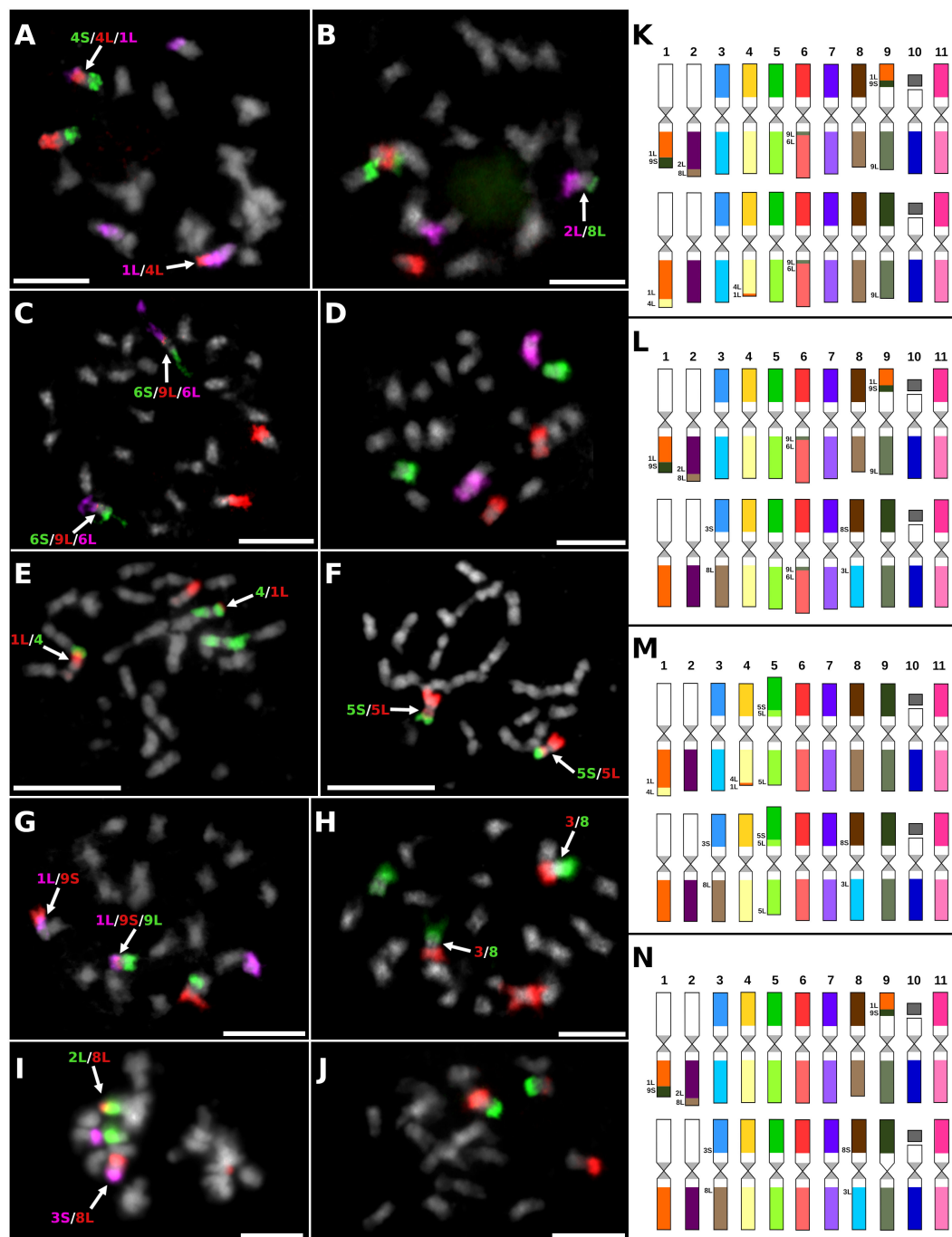


FIGURE 3

Examples of chromosome translocations identified by oligo-painting FISH on mitotic metaphase plates of seven F1 hybrids obtained in improvement programs of Mchare: (A) 'NM275-4' ($2n = 2x = 22$), probes for long arm of chromosome 1, short arm of chromosome 4, and long arm of chromosome 4 were labeled in purple, green, and red, respectively. (B) 'NM258-3' ($2n = 2x = 22$), probes for long arm of chromosome 2, short arm of chromosome 8, and long arm of chromosome 8 were labeled in purple, red, and green, respectively. (C) 'NM253-3' ($2n = 2x = 22$), probes for short arm of chromosome 6, long arm of chromosome 6, and long arm of chromosome 8 were labeled in green, purple, and red, respectively. (D) 'NM209-3' ($2n = 2x = 22$), probes for chromosomes 5, 6, and 7 were labeled by red, purple, and green, respectively. (E) 'NM237-8' ($2n = 2x = 22$), probes for long arm of chromosome 1 and entire chromosome 4 were labeled in red and green, respectively. (F) 'NM237-8' ($2n = 2x = 22$), probes for short arm of chromosome 5 and long arm of chromosome 5 were labeled in red and green, respectively. (G) 'T.2274-6' ($2n = 2x = 22$), probes for long arm of chromosome 1, short arm of chromosome 9, and long arm of chromosome 9 were labeled in purple, red, and green, respectively. (H) 'T.2274-6' ($2n = 2x = 22$), probes for chromosomes 3 and 8 were labeled in red and green, respectively. (I) 'T.2774-9' ($2n = 2x = 22$), probes for long arm of chromosome 2, short arm of chromosome 3, and long arm of chromosome 8 were labeled in green, purple, and red, respectively. (J) 'T.2619-15' ($2n = 2x = 22$), probes for short arm of chromosome 9 and long arm of chromosome 9 were labeled in green and red, respectively. Chromosomes were counterstained with DAPI (light grey pseudocolor). Arrows point to translocated chromosomes. Bars = 5 μm. Idiograms of F1 hybrid clones that originated from crosses between Mchare cultivars and *M. acuminata* ssp. *burmannicoides* 'Calcutta 4' ($2n = 2x = 22$): (K) clone 'NM275-4'; (L) clone 'NM258-3'; (M) clone 'NM237-8'; and (N) clones 'T.2274-6', 'T.2619-9', 'T.2774-15', and 'NM209-3'. Chromosomes are oriented with their short arms on the top and long arms on the bottom in all ideograms, and translocated parts of the chromosomes contain extra labels.

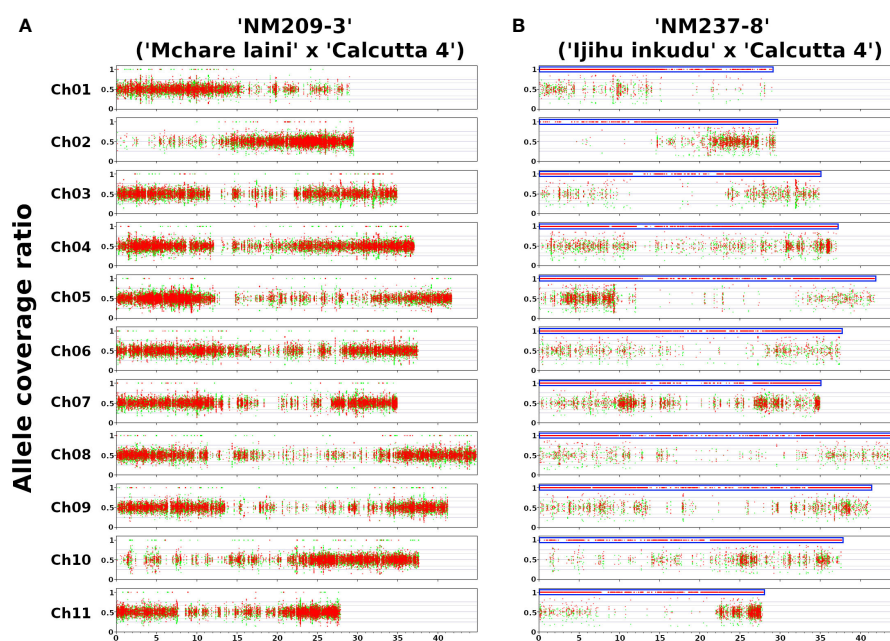


FIGURE 4

Genome structure of F1 hybrid clones gained after crosses of Mchare banana cultivars (female parent) and *M. acuminata* ssp. *burmannicoides* 'Calcutta 4' (male parent). The y-axis represents the coverage ratio of alleles specific to Mchare genotypes (red dots) and to the *M. acuminata* ssp. *burmannicoides* 'Calcutta 4' (green dots) along 11 chromosomes of *M. acuminata* ssp. *malaccensis* 'DH Pahang' reference genome sequence (x-axis). An allele coverage ratio (ACR) of 0.5 represents the equal contribution of the two parental genomes, and an ACR of 1.0 depicts the exclusive representation of one parental genome. (A) The expected ratio of 50%:50% of 'Mchare' versus 'Calcutta 4'-specific SNPs (ACR = 0.5) was revealed in most F1 hybrid clones, including 'NM209-3'. (B) A ratio of 100%:0% of 'Mchare' versus 'Calcutta 4'-specific SNPs (ACR = 1.0) was revealed only in one F1 hybrid clone 'NM237-8'. The blue boxes indicate chromosome regions with exclusive representation of 'Mchare'-specific SNPs in the F1 hybrid clone 'NM237-8'. The detection of the 'Calcutta 4'-specific SNPs along the chromosome arms of reference genome sequence *M. acuminata* 'DH Pahang' was most probably caused by the fact that the reference genome sequence used for reads alignment was created from the more distinct subspecies *malaccensis*. A small proportion of alleles specific to *burmannica*/*burmannicoides* subspecies were present also in the Chicame cultivar of Mchare bananas from Comoros, as shown in Martin et al. (2020a).

translocations were detected by chromosome-arm-specific oligo-painting FISH (Šimoníková et al., 2019, 2020). Eight additional chromosome rearrangements in *M. acuminata* were revealed in our present work (Supplementary Table S1).

Even though the chromosome painting of the additional three *banksii* genotypes did not reveal any presence of large chromosomal translocations, we detected a short segment of chromosome 9S inserted into the peri-centromeric region of chromosome 5 in the 'Higa' accession (Supplementary Figure S2F). Unfortunately, based on the oligo-painting method, we are not able to distinguish whether this short segment represents duplication or a translocation event.

Despite the very close evolutionary relationships of *burmannica*, *burmannicoides*, and *siamea* subspecies (e.g., Perrier et al., 2009; Christelová et al., 2017; Dupouy et al., 2019), their representatives contained a large number of variable chromosome structures in their genomes. All species contained already described large translocations between chromosomes 1 and 9, and 2 and 8, which seem to be specific for this evolutionary group (Šimoníková et al., 2020; Martin et al., 2020b). On the other hand, using chromosome painting, additional translocations in the genomes of *burmannica*–*burmannicoides*–*siamea* representatives were revealed. Šimoníková et al. (2020) described the presence of translocations between chromosomes 3 and 4 in 'Pa Rayong' (ssp.

siamea), and reciprocal translocations between chromosomes 7 and 8 and 3 and 8 in the genome of 'Tavoy' (ssp. *burmannica*), all presented in the homozygous state. The presence of reciprocal translocation between chromosomes 7 and 8 in the homozygous state was suggested based on genetic linkage patterns in 'Khæ Phrae' and confirmed cytogenetically by BAC-FISH in 'Long Tavoy' in the work of Martin et al. (2020b). The translocation breakpoints were located between 21.8 and 26.3 Mb of 'Pahang' reference chromosome 7 and between 22.6 and 32.1 Mb of 'Pahang' reference chromosome 8 (Martin et al., 2020b). In comparison, chromosome oligo-painting of 'Khæ Phrae' performed in our current work revealed the presence of specific chromosome structures containing chromosomes 7, 8, and 10 (Figure 2M; Supplementary Figure 2D). These reciprocally recombined chromosomes, which contained a short arm of chromosome 7, a short segment of 7L, the long arm of chromosome 8, a segment of chromosome 10L, the long arm of chromosome 7, a short segment of 7S, and the long arm of chromosome 8, were found in the heterozygous state (Figures 2H–J; Supplementary Figure 2D). Unfortunately, the chromosome oligo-painting method did not enable to localize translocation breakpoints, so we were not able to find out if the short segments of chromosomes 7L or 7S in the chromosome structures mentioned above are consequences of translocation or duplication events. Additional work utilizing

long-read sequencing technologies and further whole genome assembly have to be used to unambiguously answer these questions.

In our current study, the *burmannica-burmannicoides-siamea* phylogenetic clade contained also one accession (Pisang Karok391) originally described as representative of ssp. *malaccensis*, which thus seems to be mislabeled. On the other hand, the *malaccensis* group of subspecies with the inclusion of 'Pisang lilin' and accession described as Zebrina (*M. acuminata* ssp. *zebrina*) represents a sister cluster to *burmannica-burmannicoides-siamea*, indicating their common evolutionary history. As it was mentioned earlier, the polyploidization and hybridization events, which were accompanied by large chromosomal rearrangements (Schubert and Vu, 2016; Mandáková et al., 2017; Wang et al., 2021; Farhat et al., 2023), are known to be the major drivers of plant diversification and speciation. We can therefore speculate that the presence of specific translocations events in the *burmannica-burmannicoides-siamea* group of subspecies might have had a direct effect on their own diversification and separation from the *malaccensis* subspecies (Janssens et al., 2016). The close relationship between *malaccensis* accessions and the Pisang lilin cultivar is supported by the presence of reciprocal translocation between chromosomes 1 and 4 in their genomes.

Phylogenetic analysis showed that the only representative of ssp. *microcarpa* 'Borneo' is closely related to geographically close accession 'Buitenzorg' (described as *zebrina* ssp.) and one accession described as 'Pisang Mas Ayer' (Sucrier), a sister group of the other two accessions of the *zebrina* subspecies (Figure 1; Supplementary Figure 1). Unfortunately, 'Buitenzorg' and 'Pisang Mas Ayer' were not available during our study for the oligo-painting FISH to analyze their chromosome structure. We can thus only speculate that these two accessions can be of complex/hybrid origin, explaining their unexpected/conflicting position in the Neighbor-Net graph. This speculation can also be supported by previous results of chromosome painting that did not reveal the presence of any specific translocation event in the genome of the 'Borneo' accession, which could indicate a close relationship to the *zebrina* subspecies or Sucrier cultivars that were analyzed so far (Šimoníková et al., 2020). High level of molecular heterogeneity in ssp. *microcarpa* was mentioned in the study of Perrier et al. (2009). Previous studies showed that 'Borneo' accession shared alleles with *banksii* and always clustered together with *banksii* accessions (Perrier et al., 2009; Martin et al., 2023a). Perrier et al. (2009) showed that other representatives of *microcarpa* were closer to *zebrina*. Unfortunately, in all the studies mentioned, only a small number (at the most two representatives) of *microcarpa* accessions were analyzed; thus, we cannot speculate about the speciation processes that led to the origin and diversification of this subspecies. Nevertheless, the high level of molecular heterogeneity mentioned by Perrier et al. (2009) can indicate that hybridization events could play a major role in the speciation of the *microcarpa* subspecies. To shed more light on the evolution and genome structure of ssp. *microcarpa*, a larger number of *microcarpa* accessions from different geographic areas have to be collected and analyzed.

The *zebrina* genotypes did not form a distinct phylogenetic clade, and most of them clustered in close proximity to Sucrier cultivars. The genomes of ssp. *zebrina* and Sucrier cultivars contain reciprocal Robertsonian translocation between chromosomes 3 and 8 in the homozygous and heterozygous state, respectively. The chromosome

painting showed that the Sucrier cultivar also contained additional translocation events, all in the heterozygous state. Reciprocal translocation between chromosomes 1 and 4, specific to the *malaccensis* group, and translocation between chromosomes 1 and 7 (Figures 2B, L; Supplementary Figure S2C) indicate the involvement of a geographically close subspecies *zebrina* and *malaccensis* in the origin of Sucrier cultivars. Studies by Martin et al. (2020a, 2023a) indicated that at least four different progenitors were involved in the origin of Sucrier cultivars—*malaccensis*, *zebrina*, *banksii*, and an unknown M₂ progenitor. The same results of chromosome painting were obtained for Zebrina ITC1139. Even though the genomes of Zebrina ITC1139 and Sucrier cultivar 'Ma'i hapai' shared the chromosome painting structures, an unexpected phylogenetic position of 'Zebrina' ITC1139 was found. 'Zebrina' ITC1139 clustered together with 'Pisang lilin'. This result can indicate the hybrid origin of the 'Zebrina' ITC1139, whose genome can arise by several rounds of hybridization, similar to Sucrier, but contains a large proportion of *malaccensis*-like genome regions.

The genomic constitution of edible banana cultivars showed a high level of admixture (Martin et al., 2020a; Martin et al., 2023a). The Robertsonian translocation between chromosomes 3 and 8 was also observed in the genomes of two additional banana hybrid clones, 'Himone' and 'Maleb'. Neighbor-Net inference showed that these clones, together with the other two cultivars, 'Vudu Beo' and 'Marakudu', are closely related to Mchare. Similarly to other studies (Christelová et al., 2017; Perrier et al., 2019; Martin et al., 2020a), Mchare representatives formed a distinct evolutionary clade, which indicates their exceptional position in banana evolution. Previous phylogenetic studies were not able to unambiguously identify the mode of evolution and origin of this important group of edible banana clones, which represents a unique genetic source within *Musa*. It was shown that two subspecies of *M. acuminata*, ssp. *zebrina* and *banksii* shared some alleles with the Mchare clones (Perrier et al., 2009; Hippolyte et al., 2012; Perrier et al., 2019). The involvement of *zebrina* in the origin of Mchare bananas is supported by oligo-painting FISH, which confirmed the presence of a Robertsonian translocation between chromosomes 3 and 8 in the genomes of all Mchare clones analyzed in the present study and previously by Šimoníková et al. (2020). On the other hand, no Mchare banana cultivar was found in the region, where *banksii* and *zebrina* subspecies occur and could have contributed to the origin of these cultivars (Perrier et al., 2009; Hippolyte et al., 2012).

Chromosome painting was used to analyze the genome structure of several F1 Mchare hybrids resulting from different crosses between Mchare and *M. acuminata* ssp. *burmannicoides* 'Calcutta 4'. Regarding the fact that oligo-painting FISH can only reveal those chromosomes consisting of specific structures (the presence of translocations), it can only provide partial information on the whole genome composition of the F1 progenies. As expected for intersubspecific crosses, all progenies, except NM237–8, contained one chromosome set from both parents. This was confirmed by Illumina resequencing, which identified the contribution of parental subgenomes to the hybrid clones. The identification of parental-specific SNPs and their distribution along the chromosomes of reference genome sequence *M. acuminata* 'DH Pahang' revealed a balanced presence of both parental genomes in

all F1 hybrids. The only exception was the clone NM237–8, which contained only Mchare-specific SNPs, suggesting that the clone might have originated from an unreduced Mchare gamete. Genome composition analysis using Illumina resequencing did not reveal the presence of any aneuploid genome regions in F1 hybrids.

This analysis was done with the reference genome sequence of *M. acuminata* ssp. *malaccensis* (version 4), which is distinct from Mchare and ‘Calcutta 4’. Therefore, we cannot exclude the presence of genome regions that do not contain a proportional representation of parental subgenomes. Future high-quality chromosome-scale assembly of Mchare and/or other parental genomes used in the crosses will provide detailed information on the genome composition of the hybrid progenies.

Data availability statement

Illumina sequencing data are available on NCBI Sequence Read Archive (SRA) under the BioProject ID: 1033816 (SRA experiments: SRX22339926 - SRX22339938). The SNP files which were used for in silico painting of F1 hybrid clones are available in the Dryad repository (<https://doi.org/doi:10.5061/dryad.44j0zpcnq>).

Author contributions

DB: Investigation, Methodology, Project administration, Validation, Writing – original draft, Writing – review & editing. JČ: Investigation, Validation, Visualization, Writing – review & editing. GM: Formal analysis, Investigation, Software, Writing – review & editing. AD: Investigation, Writing – review & editing. HM: Resources, Writing – review & editing. AB: Resources, Writing – review & editing. RS: Resources, Writing – review & editing. EH: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Visualization, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was funded by the Ministry of Education, Youth, and Sports of the Czech Republic (Program INTER-EXCELLENCE, INTER-TRANSFER, grant award LTT19009) and a grant agreement between the Bill & Melinda Gates Foundation and the International Institute of Tropical Agriculture (Investment ID INV-002979).

Acknowledgments

We thank Ir. Ines van den Houwe for providing the plant material and Radomíra Tušková for excellent technical assistance. The computing was supported by the project “e-Infrastruktura CZ” (e-INFRA CZ LM2018140), supported by the Ministry of Education, Youth, and Sports of the Czech Republic, and by the

ELIXIR-CZ project (LM2018131), part of the international ELIXIR infrastructure.

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 author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

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/fpls.2024.1387055/full#supplementary-material>

SUPPLEMENTARY FIGURE 1

Neighbor-Net analysis of *Musa* accessions and F1 hybrids between Mchare and *M. acuminata* ssp. *burmannicoides* ‘Calcutta 4’ performed by SplitsTree. Accessions representing different *M. acuminata* subspecies and group of cultivars are depicted in colors: *banksii* ssp. in red; *malaccensis* ssp. in violet; *burmannica*/*burmannicoides*/*siamea* in blue; *zebrina* in green; *microcarpa* in pink; Mchare genotypes in light blue; Sucrier genotypes in orange; other analyzed banana AA cultivars in black; and F1 progeny of Mchare and *M. acuminata* ssp. *burmannicoides* ‘Calcutta 4’ in brown. The accessions which were used for chromosome painting are depicted as squared nodes in the neighbor net tree.

SUPPLEMENTARY FIGURE 2

Idiograms and short translocations (duplications) observed by oligo painting FISH in genomes of analyzed *M. acuminata* species and its edible banana clones. (A) Idiogram of cv. ‘Vudu Beo’ ITC1211; (B) Idiogram of cv. ‘Rose’ ITC0712 and oligo painting FISH with the probes for long arms of chromosomes 1 and 4, and oligo painting FISH with the probes for long arm of chromosome 1, short arm of chromosome 7 and long arm of chromosome 7; (C) Idiogram of *M. acuminata* ssp. *zebrina* ITC1139 and cv. ‘Mai’a hapai’ ITC1172, and oligo painting FISH with the probes for long arm of chromosome 1, and probes for short and long arms of chromosome 7; (D) Idiogram of *M. acuminata* ssp. *siamea* ‘Khae’ (Phrae) ITC0660, and oligo painting FISH with the probes for short arm of chromosome 7, long arm of chromosome 7 and long arm of chromosome 10L, and oligo painting FISH with the probes for short arm of chromosome 7 and probes for entire chromosomes 8 and 9; (E) Idiogram of *M. acuminata* ssp. *banksii* ITC0341 and ITC 0896, and cv. ‘Marakudu’ ITC1210; (F) Idiogram of *M. acuminata* ssp. *banksii* ‘Higa’ ITC0428 and oligo painting FISH with the probes for short arm of chromosome 9, short arm of chromosome 5 and long arm of chromosome 5; (G) Idiogram of cv. ‘Himone’ ITC0886 and cv. ‘Maleb’ ITC0809; (H) Idiogram of cv. ‘Tuu Gia’ ITC0610 and oligo painting FISH with the probes for long arm of chromosome 1 and entire chromosome, oligo painting FISH with the probes for long arm of chromosome 2 and entire

chromosome 8, and oligo painting FISH with the probes long arms of chromosomes 7 and 8; (I) Idiogram of cv. 'Mchare mlelembo', cv. 'Mchare laini' and cv. 'Kahuti'; and idiogram of F1 hybrid clone gained after crosses of Mchare banana cultivars (female parent) and *M. acuminata* ssp. *burmannicoides* 'Calcutta 4' (male parent): (J) 'T2269-1' ($2n=2x=22$).

SUPPLEMENTARY FIGURE 3

Hypothetical karyotypes of F1 hybrid clones, which could be obtained after crosses between Mchare cultivars (female parent) and *M. acuminata* ssp. *burmannicoides* 'Calcutta 4' (male parent). The chromosomes with translocations specific to Mchare genome are marked with a pink asterisk,

and chromosomes with translocations specific to 'Calcutta 4' are marked with a green asterisk. Blue rectangles indicate karyotypes of F1 hybrid clones detected in our study.

SUPPLEMENTARY FIGURE 4

Genome structure of F1 hybrid clones gained after crosses of Mchare banana cultivars (female parent) and *M. acuminata* ssp. *burmannicoides* 'Calcutta 4' (male parent). Coverage ratio of alleles specific to Mchare genotypes (red dots) and to the *M. acuminata* ssp. *burmannicoides* 'Calcutta 4' (green dots) along 11 chromosomes of *M. acuminata* ssp. *malaccensis* 'DH Pahang' reference genome sequence.

References

- Amah, D., Turner, D. W., Gibbs, J., Gil, G., and Swennen, R. (2021). Overcoming the fertility crisis in bananas (*Musa* spp.). In: *Achieving sustainable cultivation of bananas*. Chapter 1, 1–50. Kema, G. H. J., and Drenth, A. (eds.), *Achieving sustainable cultivation of bananas Volume 2: Germplasm and genetic improvement*. (Cambridge, UK: Burleigh Dodds Science Publishing). doi: 10.19103/AS.2020.0070.13
- Bakry, F., and Horry, J. P. (1992). Tetraploid hybrids from interpollid 3x/2x crosses in cooking bananas. *Fruits* 47, 641–655.
- Bandelt, H. J., and Dress, A. W. M. (1992a). A canonical decomposition theory for metrics on a finite set, *Adv. Math.* 92, 47–105. doi: 10.1016/0001-8708(92)90061-O
- Bandelt, H. J., and Dress, A. W. M. (1992b). Split decomposition: A new and useful approach to phylogenetic analysis of a distance data. *Mol. Phylogenet. Evol.* 1, 242–252. doi: 10.1016/1055-7903(92)90021-8
- Batte, M., Swennen, R., Uwimana, B., Akech, V., Brown, A., Tumuhimise, R., et al. (2019). Crossbreeding East African Highland Bananas: Lessons learnt relevant to the botany of the crop after 21 years of genetic enhancement. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00081
- Baurens, F. C., Martin, G., Hervouet, C., Salmon, F., Yohomé, D., Ricci, S., et al. (2019). Recombination and large structural variations shape interspecific edible banana genomes. *Mol. Biol. Evol.* 36, 97–111. doi: 10.1093/molbev/msy199
- Bayo, S. J., Massawe, V., Ndakidemi, P. A., Venkataramana, P., Mlaki, A., Mduma, H., et al. (2024). Pollen amount and viability in Mchare and selected wild (AA) banana (*Musa acuminata*) genotypes: Prospects for Breeding. *HortSci* 59, 632–638. doi: 10.12123/HORTSCI17608-23
- Belser, C., Baurens, F. C., Noel, B., Martin, G., Cruaud, C., Istace, B., et al. (2021). Telomere-to-telomere gapless chromosomes of banana using nanopore sequencing. *Commun. Biol.* 4, 1047. doi: 10.1038/s42003-021-02559-3
- Beránková, D., and Hřibová, E. (2023). Bulk oligo-FISH for chromosome painting and chromosome barcoding. *Methods Mol. Biol.* 2672, 445–463. doi: 10.1007/978-1-0716-3226-0_27
- Braz, G. T., He, L., Zhao, H., Zhang, T., Semrau, K., Rouillard, J. M., et al. (2018). Comparative oligo-FISH mapping: An efficient and powerful methodology to reveal karyotypic and chromosomal evolution. *Genetics* 208, 513–523. doi: 10.1534/genetics.117.300344
- Brown, A., Tumuhimise, R., Amah, D., Uwimana, B., Nyine, M., Mduma, H., et al. (2017). "Bananas and Plantains (*Musa* spp.)." in *Genetic Improvement of Tropical Crops* (Springer, Cham).
- Bryant, D., and Huson, D. H. (2023). NeighborNet: improved algorithms and implementation. *Front. Bioinf.* 3. doi: 10.3389/fbinf.2023.1178600
- Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884–i890. doi: 10.1093/bioinformatics/bty560
- Christelová, P., De Langhe, E., Hřibová, E., Čížková, J., Sardos, J., Hušáková, M., et al. (2017). Molecular and cytological characterization of the global *Musa* germplasm collection provides insight into the treasure of banana diversity. *Biodivers. Conserv.* 26, 801–824. doi: 10.1007/s10531-016-1273-9
- Christelová, P., Valárik, M., Hřibová, E., Van den Houwe, I., Channeliere, S., Roux, N., et al. (2011). A platform for efficient genotyping in *Musa* using microsatellite markers. *AOB Plants* 2011, plr024. doi: 10.1093/aobpla/plr024
- Cordoba, D., and Jansen, K. (2014). Same disease – different research strategies: Bananas and Black Sigatoga in Brazil and Colombia. *Singap. J. Trop. Geogr.* 35, 345–361. doi: 10.1111/sjtg.12072
- De Bellaire, L. L., Fouré, E., Abadie, C., and Carlier, J. (2010). Black leaf streak disease is challenging the banana industry. *Fruits* 65, 327–342. doi: 10.1051/fruits/2010034
- De Langhe, E., Hřibová, E., Carpentier, S., Doležel, J., and Swennen, R. (2010). Did backcrossing contribute to the origin of hybrid edible bananas? *Ann. Bot.* 106, 849–857. doi: 10.1093/aob/mcq187
- D'Hont, A., Denoeud, F., Aury, J. M., Baurens, F. C., Carreel, F., Garsmeur, O., et al. (2012). The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. *Nature* 488, 213–217. doi: 10.1038/nature11241
- Dodds, K. S., and Simmonds, N. W. (1948). Sterility and parthenocarpy in diploid hybrids of *Musa*. *Heredity* 2, 101–117. doi: 10.1038/hdy.1948.6
- Doležel, J., Doleželová, M., Roux, N., and Van den Houwe, I. (1998). A novel method to prepare slides for high resolution chromosome studies in *Musa* spp. *Infomusa* 7, 3–4.
- Dumschott, K., Schmidt, M. H.-W., Chawla, H. S., Snowdon, R., and Usadel, B. (2020). Oxford Nanopore sequencing: new opportunities for plant genomics? *J. Exp. Bot.* 71, 5313–5322. doi: 10.1093/jxb/eraa263
- Dupouy, M., Baurens, F. C., Derouault, P., Hervouet, C., Cardi, C., Cruaud, C., et al. (2019). Two large reciprocal translocations characterized in the disease resistance-rich burmannica genetic group of *Musa acuminata*. *Ann. Bot.* 2019, 124:31–124329. doi: 10.1093/aob/mcz078
- FAO (2023). *Banana market review – Preliminary results 2023* (Rome: FAO).
- Farhat, P., Mandáková, T., Divišek, J., Kudoh, H., German, D. A., and Lysak, M. A. (2023). The evolution of the hypotetraploid *Catolobus pendulus* genome - the poorly known sister species of *Capsella*. *Front. Plant. Sci.* 14, 1165140. doi: 10.3389/fpls.2023.1165140
- Fauré, S., Bakry, F., and Gonzalez, L. D. (1993). "Cytogenetic studies of diploid bananas," in *Breeding banana and plantain for resistance to diseases and pests*. Montpellier: CIRAD-FLHOR. Ed. J. Ganry, 77–92. International symposium on genetic improvement of bananas for resistance to diseases and pests, Montpellier, France.
- Goigoux, S., Salmon, F., and Bakry, F. (2013). Evaluation of pollen fertility of diploid and doubled-diploid clones of mlali and their potential use for banana breeding. *ISHS Acta Hort.* 986, 195–204. doi: 10.17660/ActaHortic.2013.986.20
- Han, Y., Zhang, T., Thammapichai, P., Weng, Y., and Jinag, J. (2015). Chromosome-specific painting in *Cucumis* species using bulked oligonucleotides. *Genetics* 200, 771–779. doi: 10.1534/genetics.115.177642
- Hippolyte, I., Jenny, C., Gardes, L., Bakry, F., Rivallan, R., Pomies, V., et al. (2012). Foundation characteristics of edible *Musa* triploids revealed from allelic distribution of SSR markers. *Ann. Bot.* 109, 937–951. doi: 10.1093/aob/mcs010
- Hou, L., Xu, M., Zhang, T., Xu, Z., Wang, W., Zhang, J., et al. (2018). Chromosome painting and its applications in cultivated and wild rice. *BMC Plant Biol.* 18, 110. doi: 10.1186/s12870-018-1325-2
- Huson, D. H., and Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* 23, 254–267. doi: 10.1093/molbev/msj030
- Ibarra-Laclette, E., Lyons, E., Hernández-Guzmán, G., Pérez-Torres, C. A., Carretero-Paulet, L., Chang, T., et al. (2013). Architecture and evolution of a minute plant genome. *Nature* 498, 94–98. doi: 10.1038/nature12132
- Janssens, S. B., Vandeloof, F., De Langhe, E., Verstraete, B., Smets, E., Van den Houwe, I., et al. (2016). Evolutionary dynamics and biogeography of Musaceae reveal a correlation between the diversification of the banana family and the geological and climatic history of Southeast Asia. *New Phytol.* 210, 1453–1465. doi: 10.1111/nph.13856
- Jiang, J. (2019). Fluorescence *in situ* hybridization in plants: recent developments and future applications. *Chromosome Res.* 27, 153–165. doi: 10.1007/s10577-019-09607-z
- Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv: Genomics*. doi: 10.48550/arXiv.1303.3997
- Liu, X., Arshad, R., Wang, X., Li, W. M., Zhou, Y., Ge, X.-J., et al. (2023). The phased telomere-to-telomere reference genome of *Musa acuminata*, a main contributor to banana cultivars. *Sci. Data* 10, 631. doi: 10.1038/s41597-023-02546-9
- Mandáková, T., and Lysak, M. A. (2018). Post-polyploid diploidization and diversification through dysploid changes. *Curr. Opin. Plant Biol.* 42, 55–65. doi: 10.1016/j.pbi.2018.03.001
- Mandáková, T., Pouch, M., Harmanová, K., Zhan, S. H., Mayrose, I., and Lysak, M. A. (2017). Multispeed genome diploidization and diversification after an ancient allopolyploidization. *Mol. Ecol.* 22, 6445–6462. doi: 10.1111/mec.14379
- Martin, G., Baurens, F. C., Hervouet, C., Salmon, F., Delos, J. M., Labadie, K., et al. (2020b). Chromosome reciprocal translocations have accompanied subspecies evolution in bananas. *Plant J.* 104, 1698–1711. doi: 10.1111/tjpi.15031

- Martin, G., Baurens, F. C., Labadie, K., Hervouet, C., Salmon, F., Marius, F., et al. (2023b). Shared pedigree relationships and transmission of unreduced gametes in cultivated banana. *Ann. Bot.* 131, 1149–1161. doi: 10.1093/aob/mcad065
- Martin, G., Cardi, C., Sarah, G., Ricci, S., Jenny, C., Fondi, E., et al. (2020a). Genome ancestry mosaics reveal multiple and cryptic contributors to cultivated banana. *Plant J.* 102, 1008–1025. doi: 10.1111/tpj.14683
- Martin, G., Carreel, F., Coriton, O., Hervouet, C., Cardi, C., Derouault, P., et al. (2017). Evolution of the banana genome (*Musa acuminata*) is impacted by large chromosomal translocations. *Mol. Biol. Evol.* 34, 2140–2152. doi: 10.1093/molbev/msx164
- Martin, G., Cottin, A., Baurens, F. C., Labadie, K., Hervouet, C., Salmon, F., et al. (2023a). Interspecific introgression patterns reveal the origins of worldwide cultivated bananas in New Guinea. *Plant J.* 113, 802–818. doi: 10.1111/tpj.16086
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernysky, 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
- Nyine, M., Uwimana, B., Swennen, R., Batte, M., Brown, A., Christelová, P., et al. (2017). Trait variation and genetic diversity in banana genomic selection training population. *PLoS One* 12, e0178734. doi: 10.1371/journal.pone.0178734
- Ortiz, R. (2013). Conventional banana and plantain breeding. *Acta Hort.* 986, 77–194. doi: 10.17660/ActaHortic.2013.986.19
- Ortiz, R., and Swennen, R. (2014). From crossbreeding to biotechnology-facilitated improvement of banana and plantain. *Biotechnol. Adv.* 32, 158–169. doi: 10.1016/j.biotechadv.2013.09.010
- Perrier, X., Bakry, F., Carreel, F., Jenny, F., Horry, J. P., Lebot, V., et al. (2009). Combining biological approaches to shed light on the evolution of edible bananas. *Ethnobot. Res. Appl.* 7, 199–216. doi: 10.17348/era.7.0.199-216
- Perrier, X., De Langhe, E., Donohue, M., Lentfer, C. J., Vrydaghs, L., Bakry, F., et al. (2011). Multidisciplinary perspectives on banana (*Musa* spp.) domestication. *Proc. Natl. Acad. Sci. U.S.A.* 108, 11311–11318. doi: 10.1073/pnas.1102001108
- Perrier, X., Jenny, C., Bakry, F., Karamura, D., Kitavi, M., Dubois, C., et al. (2019). East African diploid and triploid bananas: a genetic complex transported from South-East Asia. *Ann. Bot.* 123, 19–36. doi: 10.1093/aob/mcy156
- Pucker, B., Irisarri, I., de Vries, J., and Xu, B. (2022). Plant genome sequence assembly in the era of long reads: Progress, challenges and future directions. *Quant. Plant Biol.* 3, e5. doi: 10.1017/qpb.2021.18
- Raboin, L. M., Careel, F., Noyer, J. L., Baurens, F. C., Horry, J. P., Du Montcel, H. T., et al. (2005). Diploid ancestors of triploid export banana cultivars: Molecular identification of $2n$ restitution gamete donor and n gamete donors. *Mol. Breed.* 16, 333–341. doi: 10.1007/s11032-005-2452-7
- Salse, J. (2016). Deciphering the evolutionary interplay between subgenomes following polyploidy: A paleogenomics approach in grasses. *Am. J. Bot.* 103, 1167–1174. doi: 10.3732/ajb.1500459
- Sardos, J., Breton, C., Perrier, X., Van den Houwe, I., Carpentier, S., Paofa, J., et al. (2022). Hybridization, missing wild ancestors and the domestication of cultivated diploid bananas. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.969220
- Sardos, J., Perrier, X., Doležel, J., Hříbová, E., Christelová, P., Van den Houwe, I., et al. (2016a). DArT whole genome profiling provides insights on evolution and taxonomy of edible banana (*Musa* spp.). *Ann. Bot.* 118, 1269–1278. doi: 10.1093/aob/mcw170
- Schubert, I., and Vu, G. T. H. (2016). Genome stability and evolution: attempting a holistic view. *Trends Plant Sci.* 21, 749–757. doi: 10.1016/j.tplants.2016.06.003
- Shepherd, K. (1999). *Cytogenetics of the genus Musa* (Montpellier: International Network for the Improvement of Bananas and Plantain).
- Simmonds, N. W., and Shepherd, K. (1955). The taxonomy and origin of the cultivated bananas. *Bot. J. Linn. Soc.* 55, 302–312. doi: 10.1111/j.1095-8339.1955.tb00015.x
- Šimoníková, D., Němečková, A., Čížková, J., Brown, A., Swennen, R., Doležel, J., et al. (2020). Chromosome painting in cultivated bananas and their wild relatives (*Musa* spp.) reveals differences in chromosome structure. *Int. J. Mol. Sci.* 21, 7915. doi: 10.3390/ijms21217915
- Šimoníková, D., Němečková, A., Karafiátová, M., Uwimana, B., Swennen, R., Doležel, J., et al. (2019). Chromosome painting facilitates anchoring reference genome sequence to chromosomes *in situ* and integrated karyotyping in banana (*Musa* spp.). *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.01503
- Stull, G. W., Pham, K. K., Soltis, P. S., and Soltis, D. E. (2023). Deep reticulation: the long legacy of hybridization in vascular plant evolution. *Plant J.* 114, 743–766. doi: 10.1111/tpj.16142
- Tomekpe, K., Jenny, C., and Escalant, J. V. (2004). A review of conventional improvement strategies for *Musa*. *Infomusa* 13, 2–6.
- Wang, X., Morton, J. A., Pellicer, J., Leitch, I. J., and Leitch, A. R. (2021). Genome downsizing after polyploidy: mechanisms, rates and selection pressures. *Plant J.* 107, 1003–1015. doi: 10.1111/tpj.15363
- Wendel, J. F. (2000). Genome evolution in polyploids. *Plant Mol. Biol.* 42, 225–249. doi: 10.1023/A:1006392424384
- Yu, F., Zhao, X., Chai, J., Ding, X., Li, X., Huang, Y., et al. (2022). Chromosome-specific painting unveils chromosomal fusions and distinct allopolyploid species in the *Saccharum* complex. *New Phytol.* 233, 1953–1965. doi: 10.1111/nph.17905



OPEN ACCESS

EDITED BY

Marcial Escudero,
Sevilla University, Spain

REVIEWED BY

Hikmet Budak,
Arizona Western College, United States
Magdalena Vaio,
Universidad de la República, Uruguay

*CORRESPONDENCE

Pilar Catalán

✉ pcatlán@unizar.es

Rubén Sancho

✉ rubenben85@hotmail.com

[†]These authors have contributed
equally to this work and share
first authorship

RECEIVED 17 April 2024

ACCEPTED 19 June 2024

PUBLISHED 10 July 2024

CITATION

Decena MÁ, Sancho R, Inda LA,
Pérez-Collazos E and Catalán P (2024)
Expansions and contractions of repetitive
DNA elements reveal contrasting evolutionary
responses to the polyploid genome shock
hypothesis in *Brachypodium* model grasses.
Front. Plant Sci. 15:1419255.
doi: 10.3389/fpls.2024.1419255

COPYRIGHT

© 2024 Decena, Sancho, Inda, Pérez-Collazos
and Catalán. 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.

Expansions and contractions of repetitive DNA elements reveal contrasting evolutionary responses to the polyploid genome shock hypothesis in *Brachypodium* model grasses

María Ángeles Decena^{1,2†}, Rubén Sancho^{1,2*†}, Luis A. Inda^{1,3},
Ernesto Pérez-Collazos^{1,2} and Pilar Catalán^{1,2*}

¹Escuela Politécnica Superior de Huesca, Universidad de Zaragoza, Huesca, Spain, ²Grupo de Bioquímica, Biofísica y Biología Computacional (Instituto de Biocomputación y Física de Sistemas Complejos (BIFI) Universidad de Zaragoza), Unidad Asociada al Consejo Superior de Investigaciones Científicas (CSIC), Zaragoza, Spain, ³Centro de Investigaciones Tecnológicas y Agroalimentarias de Aragón (CITA), Zaragoza, Spain

Brachypodium grass species have been selected as model plants for functional genomics of grass crops, and to elucidate the origins of allopolyploidy and perennality in monocots, due to their small genome sizes and feasibility of cultivation. However, genome sizes differ greatly between diploid or polyploid *Brachypodium* lineages. We have used genome skimming sequencing data to uncover the composition, abundance, and phylogenetic value of repetitive elements in 44 representatives of the major *Brachypodium* lineages and cytotypes. We also aimed to test the possible mechanisms and consequences of the “polyploid genome shock hypothesis” (PGSH) under three different evolutionary scenarios of variation in repeats and genome sizes of *Brachypodium* allopolyploids. Our data indicated that the proportion of the genome covered by the repeatome in the *Brachypodium* species showed a 3.3-fold difference between the highest content of *B. mexicanum*-4x (67.97%) and the lowest of *B. stacei*-2x (20.77%), and that changes in the sizes of their genomes were a consequence of gains or losses in their repeat elements. LTR-Retard and Tekay retrotransposons were the most frequent repeat elements in the *Brachypodium* genomes, while Ogre retrotransposons were found exclusively in *B. mexicanum*. The repeatome phylogenetic network showed a high topological congruence with plastome and nuclear rDNA and transcriptome trees, differentiating the ancestral outcore lineages from the recently evolved core-perennial lineages. The 5S rDNA graph topologies had a strong match with the ploidy levels and nature of the subgenomes of the *Brachypodium* polyploids. The core-perennial *B. sylvaticum* presents a large repeatome and characteristics of a potential post-polyploid diploidized origin. Our study evidenced that expansions and contractions in the repeatome were responsible for the three contrasting responses to the PGSH. The exacerbated genome expansion of the ancestral allotetraploid *B. mexicanum* was a consequence of chromosome-wide proliferation of TEs and not of WGD, the additive repeatome pattern of young allotetraploid *B. hybridum* of stabilized post-WGD genome evolution, and

the genome contraction of recent core-perennials polyploids (*B. pinnatum*, *B. phoenicoides*) of repeat losses through recombination of these highly hybridizing lineages. Our analyses have contributed to unraveling the evolution of the repeatome and the genome size variation in model *Brachypodium* grasses.

KEYWORDS

Brachypodium, evolution, genome size diversification, polyploid genome shock, repeatome, transposable elements, 5S rDNA loci

Introduction

The “polyploid genome shock hypothesis” (PGSH), first proposed by McClintock (1984), postulates rapid genome restructuring following hybridization and whole genome duplication (WGD). PGSH is hypothesized to be a response to the sudden combination of different genomes into a single nucleus, and genetic and epigenetic regulatory adjustments necessary to keep pace between them (Bird et al., 2018; Edger et al., 2019), along with the possible disruption of genome integrity induced by WGD (Scarlett et al., 2023). However, it is less clear how this genomic re-patterning may occur and affect, in the short and long term, the new polyploid genome. Research in some wild and synthetic allopolyploid plants indicates that genomic reshuffling is common in the first generations after WGD, while the allopolyploid genome tends to stabilize and diploidize over time (De Storme and Mason, 2014; Wendel et al., 2018; Mason and Wendel, 2020; Deb et al., 2023). However, other plant allopolyploids have evidenced either long-term genomic instability, perpetuated over generations (Chalhoub et al., 2014; Mason and Wendel, 2020), or a complete absence of genomic restructuring, with the immediate creation of the amphidiploid (Scarlett et al., 2023; Deb et al., 2023; Mu et al., 2023a). One of the components of the genome more severely impacted by the potential restructuring is the repetitive DNA fraction, or repeatome (mobile elements (retrotransposons and transposons), and tandem ribosomal DNA and satellite repeats), which is largely present in the nuclear genome of most plants (Macas et al., 2015; Hidalgo et al., 2017; Pellicer et al., 2018). Three contrasting evolutionary scenarios have been proposed to explain the potential consequences of PGSH on polyploid repeatome turnovers. In some angiosperms, a rapid increase of repeats has been detected in the genomes after rounds of polyploidizations (Chen et al., 2020). In others, the polyploid genomes show repeatome sizes equivalent to those of their diploid progenitor species (McCann et al., 2018). And in other groups, high-level polyploids exhibit a considerable reduction of their repeatome with respect to that of their diploid and low-level polyploid relatives (Chen, 2007; Parisod et al., 2010; Moreno-Aguilar et al., 2022). The ability of centromeric retrotransposon families to proliferate has been interpreted as the potential mechanism for the increased repeatome of the first group of plants, while stabilized post-WGD genome evolution would explain the additive patterns of the second group,

and the trend of other repeat families to recombine and lose repeats may have caused the repeatome shrinkage in the third group (Michael, 2014; Chen et al., 2020; Scarlett et al., 2023). Although the proliferation or removal of the repetitive elements from the genomes could have resulted from recombination or double-strand break repair processes (Hawkins et al., 2009; Vu et al., 2017), the driving forces that balance the expansions and contractions of the repeatome are little known (Fedoroff, 2012; Drouin et al., 2021). Analysis of the three alternative evolutionary scenarios of the PGSH has also been hampered by the lack of a suitable specific group to test all their cases.

The importance and impact that the dynamics of repetitive elements has had on genome size (GS) variation and its evolution across the angiosperms has been corroborated in several studies (Dodsworth et al., 2015; Hidalgo et al., 2017; Pellicer et al., 2018). In plants with available reference genomes, the dynamics of transposable elements (TEs) insertions have also been linked to the expression of some core or dispensable genes, which are differentially regulated (Gordon et al., 2017), and to epigenetic effects (Chen, 2007; Fedoroff, 2012; Negi et al., 2016). However, analysis of the repetitive families in most angiosperms that lack assembled and annotated genomes has been performed using genome skimming data and repeatome graphical topology methods (Weiss-Schneeweiss et al., 2015; Garcia et al., 2020; Vitales et al., 2020a; Moreno-Aguilar et al., 2022). The quantification and annotation of repeats in plant genomes relies on the fact that similarity-based clustering of low-coverage genome sequencing reads, which confidentially represents between 0.01 and 0.50 times the coverage of total haploid genome, is proportional to the genomic abundance and length of the corresponding repeat types and could be used to quantify them (Macas et al., 2015; Pellicer et al., 2018; Novák et al., 2020). Furthermore, comparative analysis based on repeat sequence similarities has confirmed the phylogenetic signal of the repeatome in several angiosperm groups (Vitales et al., 2020a, b; Herklotz et al., 2021) and its utility to infer ancestral and recent polyploidization and diploidization events (Moreno-Aguilar et al., 2022). In addition, 5S rDNA graph-based clustering approaches have corroborated the identity of the ancestral progenitor genomes of several polyploid plants (Garcia et al., 2020) and have also uncovered the paleopolyploid nature of current diploidized plant species (Vozárová et al., 2021; Moreno-Aguilar et al., 2022).

The cool seasonal genus *Brachypodium*, consisting of approximately 23 taxa (Catalan et al., 2016; Catalán et al., 2023), has been selected as a model functional system for cereal and biofuel crops and to investigate the evolution of polyploidy in grasses. Annotated reference genomes and considerable genomic resources have been produced for its three annual species (*B. distachyon*, *B. stacei*, *B. hybridum*) (Scholthof et al., 2018; Hasterok et al., 2022; Mu et al., 2023a, b; Chen et al., 2024) and for the slender perennial *B. sylvaticum* (Lei et al., 2024). Comparative genomic studies of the annual species evidenced that the allotetraploid *B. hybridum* of recurrent origin did not undergo significant genomic restructuring, showing equivalently inherited parental transposon contents in its two subgenomes (Gordon et al., 2020; Scarlett et al., 2023; Mu et al., 2023a). However, cytogenetic analysis of the less investigated perennial taxa detected considerable differences between the large genome sizes of ancestral polyploids (*B. mexicanum*-4x, 3.7 pg (2C); *B. boissieri*-6x, 3.1 pg) and the relatively small sizes of recently evolved polyploids (e. g., *B. rupestre*-4x, *B. phoenicoides*-4x, 1.4 pg) (Sancho et al., 2022). *Brachypodium* shows a remarkable descending dysploidy trend from ancestral x=10 karyotypes to intermediately-to-recently evolved x=9, x=8 and x=5 karyotypes (Lusinska et al., 2019; Sancho et al., 2022). Phylogenetic subgenome detection algorithms applied to transcriptome data and karyotype barcoding analysis further identified seven diploid subgenomes in the studied *Brachypodium* polyploids, three of them present in extant diploid progenitor species and four orphan (only detected in polyploid species) (Sancho et al., 2022). Except for the thoroughly investigated annual species *B. distachyon*, where transposon landscape analysis revealed high transposable activity of LTR-*Copia* Angela elements and a large contribution to genome size of highly methylated LTR-*Gypsy* Retand elements (Stritt et al., 2020), and the identification of centromeric CRBd retrotransposons elements in six *Brachypodium* species (Li et al., 2018) and the characterization of centromeric species-specific satellite DNA families in the three annual species (Chen et al., 2024), no other study has comprehensively explored the composition and dynamics of repetitive elements across a large representation of *Brachypodium* taxa.

We were particularly interested in using *Brachypodium* as a test case study for the three alternative evolutionary scenarios of PGSH. These *Brachypodium* samples constitute exemplary case studies to investigate the putative role of repeat elements' dynamics in the evolution of these genomes and to test the potential mechanisms and consequences of the "polyploid genome shock hypothesis" in three different evolutionary scenarios of proliferation, maintenance, and reduction of repeats and genome sizes of allopolyploids that occur within this genus. We also attempted to comparatively analyze the repeatome variations in diploid *Brachypodium* species that show substantial differences in genome sizes (Sancho et al., 2022). The repeatome analysis was also used to assess the potential phylogenetic value of repeat elements in the monotypic Brachypodieae tribe. The objectives of our study were: (i) to characterize and quantify the repetitive elements of 44 representative samples of the main *Brachypodium* species and cytotypic lineages identifying both shared and private repeats; (ii) to analyze the expected correlation between genome size and repeat abundance; (iii) to identify repeat types that could have contributed to the expansions or contractions of the genomes; (iv) to evaluate

the phylogenetic signal of repeats using phylogenetic reconstructions; and (v) to assess the three alternative responses to the "polyploid genome shock hypothesis" of *Brachypodium* polyploids and the putative paleo-polyploid origin of some large genome diploids using mobile and tandem repeat data analysis.

Methods

Sampling, ploidy levels and genome skimming sequencing

Genomic sequences from 44 accessions of 11 *Brachypodium* species, representing its main out-core (*B. stacei*, *B. mexicanum*, *B. boissieri*, *B. retusum pro partim* (ancestral subgenome), *B. distachyon*, *B. hybridum*), and core-perennial (*B. arbuscula*, *B. sylvaticum*, *B. retusum pro partim* (recent subgenome), *B. pinnatum*, *B. rupestre*, *B. phoenicoides*) lineages (Sancho et al., 2022) were incorporated to the study (Table 1; Supplementary Table S1; Supplementary Figure S1). 41 perennial samples were sequenced *de novo* using genome skimming, while genome data from three annual samples was downloaded from NCBI (*B. distachyon* Bd21-3 (SRR4236817); *B. stacei* ABR114 (SRR3944701); *B. hybridum* ABR113 (SRR3945056; SRR3945058)). We obtained a large cytotype representation of *Brachypodium* through the study of 15 different cytotypes found within these taxa, including diploids, tetraploids and hexaploids (Table 1; Supplementary Table S1). The samples were classified into species and group-lineages based on previous taxonomic and phylogenetic studies (Schippmann, 1991; Catalan et al., 2016; Díaz-Pérez et al., 2018; Sancho et al., 2022). The 44 selected samples represent all currently recognized evolutionary lineages within *Brachypodium*. They include all the main annual and perennial diploid and polyploid lineages of *Brachypodium*, formed at different evolutionary times and spanning the last 12 Ma (Catalan et al., 2016; Díaz-Pérez et al., 2018; Sancho et al., 2022).

The cytogenetic and karyotypic knowledge of *Brachypodium* has been recently expanded in the recent study of Sancho et al. (2022). Chromosome counting (2n values) analysis was performed on DAPI-stained meristematic root cells following Jenkins and Hasterok (2007). Genome sizes (GS, 2C values) were estimated from fresh leaf tissue using propidium iodide staining of cell nuclei and flow cytometry measurements (Sysmex Ploidy Analyser) following Doležal et al. (2007). Ploidy levels of the *Brachypodium* samples under study were inferred from chromosome and GS estimates obtained in this study and from the repeat data for two samples of *B. mexicanum* (Bmex348H, Bmex504) showing similar repeatome content than their conspecific reference genome sample (Bmex347) (see Table 1; Supplementary Table S1). Total DNA for 41 (perennial *Brachypodium* samples) out of the 44 samples studied was extracted from fresh and silica gel-dried leaf tissues of plants growing in the common garden of the University of Zaragoza – High Polytechnic School of Huesca and from herbarium specimens (*B. mexicanum* 348H, Herbarium B) (Supplementary Table S1). DNA isolation was accomplished using a modified CTAB protocol (Doyle and Doyle, 1987) using ~20mg of tissue. DNA concentration (100–200ng/ul) and absorbance (260/230 nm of 1.8 to 2.1 and 260/280 nm of 1.8 to 2.0)

TABLE 1 Samples included in the repeatome analysis of *Brachypodium*. Species, sample's code, chromosome number (2n), genome size (2C, pg), inferred ploidy level (nx), monoploid genome size (1Cx, pg) and locality of origin.

Species	Code	Chromosome number (2n)	Genome Size (2C/pg)	Ploidy (nx)	Monoploid genome (1Cx/pg)	Locality
<i>B.distachyon</i>	Bdis_Bd21-3	10	0.631 ± 0	2	0.316	Iraq: Salakudin
<i>B.stacei</i>	Bsta_ABR114	20	0.564 ± 0	2	0.282	Spain: Balearic isles
<i>B. hybridum</i>	BhybABR113	30	1.265 ± 0	4	0.633	Portugal: Lisbon
<i>B.arbuscula</i>	Barb502	18	0.713 ± 0.004	2	0.357	Spain: Canary Isles
<i>B.boissieri</i>	Bboi3	48	3.236 ± 0.072	6	0.539	Spain: Granada
	Bboi10	48	3.152 ± 0.04	6	0.525	Spain: Granada
	Bboi15	48	3.149 ± 0.032	6	0.525	Spain: Granada
<i>B.mexicanum</i>	Bmex347	40	3.774	4	0.944	Mexico: Hidalgo
	Bmex348H	40*	~3.774*	4*	~0.944*	Mexico: Puebla
	Bmex504	40*	~3.774*	4*	~0.944*	Ecuador: Loja
<i>B.phoenicoides</i>	Bpho6-1R	28	1.443 ± 0.019	4	0.361	Spain: Huesca
	Bpho422	28	1.469 ± 0.012	4	0.367	Slovakia: Ružomberok
	Bpho452	38	2.176 ± 0.017	6	0.363	Morocco: Rif Mountains
	Bpho552	38	2.204 ± 0.039	6	0.367	Spain: Cadiz
	Bpho553	38	2.183 ± 0.013	6	0.364	Spain: Malaga
	Bpho554-1	38	2.155 ± 0.02	6	0.359	Spain: Granada
<i>B.pinnatum</i>	Bpin505	18	0.822 ± 0.009	2	0.411	Norway: Oslo
	Bpin34	28	1.449 ± 0.018	4	0.362	Great Britain: North Wiltshire
	Bpin514	28	1.537 ± 0.012	4	0.384	Turkey: Samsun
	Bpin520	28	1.499 ± 0.014	4	0.375	Netherlands: Scherpenzeel
<i>B.retusum</i>	Bret400	32	1.704 ± 0.024	4	0.426	Spain: Huesca
	Bret407	32	1.715 ± 0.017	4	0.429	Spain: Huesca
	Bret453-4	32*	1.840 ± 0.097	4	0.460	Morocco: Rif Mountains
	Bret454	32*	1.862 ± 0.196	4	0.466	Morocco: Tazza-Bou Idir
	Bret504	32	1.669 ± 0.026	4	0.417	France: Hérault
	Bret555	32	1.715 ± 0.017	4	0.429	Spain: Granada
	Bret403	42	2.373 ± 0.958	6	0.396	Spain: Huesca
	Bret408	42	2.431 ± 0.033	6	0.405	Spain: Navarra
	Bret551	42*	2.109 ± 0.025	6	0.352	Spain: Málaga
	Bret557	42	2.464 ± 0.026	6	0.411	Spain: Cadiz
	Bret561	42	2.362 ± 0.046	6	0.394	Spain: Zaragoza
<i>B.rupestre</i>	Brup7	28	1.562 ± 0.016	4	0.391	Russia: Moscow
	Brup439-1	28	1.55 ± 0.022	4	0.388	Spain: Huesca
	Brup441	28	1.483 ± 0.008	4	0.371	Spain: Leon

(Continued)

TABLE 1 Continued

Species	Code	Chromosome number (2n)	Genome Size (2C/pg)	Ploidy (nx)	Monoploid genome (1Cx/pg)	Locality
	Brup442	28	1.56 ± 0.03	4	0.39	Spain: Huesca
	Brup443	28	1.498 ± 0.012	4	0.375	Spain: Guipuzcoa
	Brup444	28	1.492 ± 0.021	4	0.373	Spain: Lugo
	Brup182	38	2.258 ± 0.026	6	0.376	Croatia: Istria
	Brup600	38	2.216 ± 0.013	6	0.369	France: Nans les Pins
	Brup605	38	2.265 ± 0.013	6	0.378	France: Pourrieres
<i>B.sylvaticum</i>	Bsyl54-1	18*	0.888 ± 0.008	2	0.444	Morocco: Rif Mountains
	Bsyl466-6	18*	0.928 ± 0.013	2	0.464	Spain: Huesca
	Bsyl477-1	18*	0.932 ± 0.017	2	0.466	Spain: Lerida
	Bsyl501-6	18*	0.947 ± 0.01	2	0.474	France: Alpes Maritimes

Ploidy levels inferred from chromosome counting and genome sizes (this study and previous records). Asterisks indicate the inferred 2n and 2C values and ploidy level of two *B. mexicanum* samples in accordance with those of their reference taxonomic cytotype based on similar repeat contents. See [Supplementary Table S1](#) for additional information on taxon authorship, detailed localities and vouchers, and cytogenetic and genomic data sources.

were estimated using Qubit® 3.0 (Life Technologies, Grand Island, NY) and Biodrop µLITE (Harvard Biosciences), respectively. Genome skimming sequencing was performed from a PCR-free multiplexed pool of KAPA libraries through the Illumina technology in paired-end mode (2 x 101 bp) at the Spanish Centro Nacional de Análisis Genómicos (CNAG, Barcelona). Illumina paired-end reads were checked using FastQC_v0.11.9 (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). The 41 *Brachypodium* genomic samples used in downstream analysis contained between 10.09 – 30.22 million reads (average 16.7 million reads) with insert sizes ranging between 123 – 329 bp ([Supplementary Table S2](#)). In addition, sequence data from the three annual *Brachypodium* samples were retrieved from NCBI. The downloaded sequences were filtered using Trimmomatic-0.39 ([Bolger et al., 2014](#)) with the following parameters: SLIDINGWINDOW:15:28 (window of bases: quality threshold) and CROP, HEADCROP and MINLEN according to the per-base sequence content and the length of 101 bp, based on the read length of other samples.

Repeat clustering and annotation, and 5S rDNA graph-clustering analysis

The repeatome of the *Brachypodium* samples under study was analyzed using RepeatExplorer2 (RE2), a computational pipeline that uses similarity graph-based clustering of filtered PE reads for the identification of the composition and proportion of repetitive elements ([Novák et al., 2010, 2013, 2020](#)). Two clustering analyses were performed. First, each sample was analyzed independently using RE2 through the Galaxy platform (<https://repeateexplorer-elixir.cerit-sc.cz>; Galaxy Version 2.3.8.1) following the protocol of [Novák et al. \(2020\)](#). The clustering analysis of individual samples

was fed with variable amount of PE reads per sample to achieve the recommended genome coverage (0.01–0.5×) of each taxon ([Supplementary Table S2](#)). Clustering was conducted employing the default RE2 settings (90% similarity, minimum overlap = 55; cluster size threshold = 0.01 %) discarding organellar clusters. Automated RE2 cluster annotation was used to quantify the clusters and calculate the proportions of repetitive elements in each sample. Sequence data from the three annual *Brachypodium* samples, enriched with organelle sequences, were pre-processed to remove them using the DUK software ([Li et al., 2011](#)) with a k-mer size of 24, a cut-off threshold for 2 matched reads, and the respective plastomes of *B. distachyon* Bd21-3 (LT558596.1), *B. stacei* ABR114 (NC_036837.1; [Sancho et al., 2018](#)) and *B. hybridum* ABR113 (NC_036836.1; S-plastotype), as references. Pre-processing steps were applied to format the sequences according to the requirements for subsequent analyses. Thus, split_pairs.v0.5 (https://github.com/ead-csic-compbio/split_pairs; [Contreras-Moreira et al., 2016](#)) was used to obtain the interleaved paired input format, and seqtk.v.1.3-r117 (<https://github.com/lh3/seqtk>) to convert fastq to fasta format. Sequences headers were formatted following the specifications required by the downstream analysis. This independent RE2 analysis resulted in the automatic repeat annotation and quantification of the studied repeatomes.

Secondly, the comparative analysis of all the *Brachypodium* samples under study was carried out using the RE2 program installed on our local server (command `repex_tarean/seqclust`) using the following parameters: `/repex_tarean/seqclust -p -l Brachy_clustering.log -c 0 -P 2 -v Brachy_clustering Brachy_RE.fasta -C -tax VIRIDIPLANTAE3.0 -opt ILLUMINA`. This comparative clustering analysis was performed employing the same RE2 settings used for the individual analyses. Organelle clusters and/or clusters with missing data were also removed. The resulting clusters were used for subsequent phylogenetic analysis.

Clustering graph analysis of the 5S data was performed with the Tandem Repeat Analyzer (TAREAN) algorithm implemented in RE2 (Novák et al., 2010; Garcia et al., 2020), available in the Galaxy environment, using the same input indicated above for the individualized RE2 analysis (Table 1; Supplementary Tables S1, S2). The shapes of the 5S rDNA clusters were characterized using a connected component index parameter (C) and their k-mer score was calculated as the sum of the frequencies of all k-mers used for consensus sequence reconstruction (Garcia et al., 2020). The graph topologies of the 5S rDNA cluster were visually inspected and classified into three groups of graph (type 1, simple circular-shaped graph; type 2, complex graph with two loops; type 3, complex graph with three loops); in the complex graphs, interconnected loops represent IGS spacers (Garcia et al., 2020). The 5S graphs were inspected to detect potential variation of 5S rDNA loci (5S ribotypic families) and to identify presumable hybrids and allopolyploids.

Correlations of repeat amounts and genome size variation in *Brachypodium*

To analyze the potential contribution of the different repeat types and the repeatome to the variation in monoploid genome size (1Cx) observed between and within the *Brachypodium* lineages and samples studied, we performed a test search using the data from the individual analysis and the linear regression model analyses (Pearson correlation coefficient) with the *ggscatter* function from the *ggpubr* package (Kassambara, 2023) in R v.4.0.5 (R Core Team, 2021). Estimation of the monoploid genome size (Cx) from the holoploid genome size (C) is not straightforward in *Brachypodium* allopolyploids, since most of them (all except autopolyploid *B. boissieri*) show dysploid subgenomes (Sancho et al., 2022, and unpub. data). However, we have assumed that the arithmetic mean of the number of genomes/subgenomes is the best approach to estimate it (e. g., allotetraploids (4x): $Cx = C/2$; auto- and allohexaploids (6x): $Cx = C/3$). The respective contributions of repeats to pairwise differences in genome sizes were estimated following Macas et al. (2015) and Moreno-Aguilar et al. (2022) using absolute amounts (Mbp) of repeats calculated for individual species (Supplementary Tables S3, S4). We also tested whether there were significant differences in repeat amount for different repeat families obtained from the individual analysis through Kruskal–Wallis rank tests using the *PMCMRplus* package (Pohlert, 2023) in R.

Landscape genomic diversity analysis of repeat types in *Brachypodium*

To investigate the levels of conservatism or diversity of the repeat types that contributed most to genome size variation in *Brachypodium* (44 studied samples) we performed a genome landscape search for the global variability of these individual repeat types across the *Brachypodium* genomes following Macas et al. (2015) and Moreno-Aguilar et al. (2022). We pooled the pairwise similarity values of reads, retrieved from the RE2 outputs (hitsort files) of the global comparative analysis (all samples together), for each sample

and repeat type in a separate dataset and evaluated their similarities with respect to similarities of reads from the same repeat in all other samples. We then calculated the ratios of intraspecific versus interspecific similarity matches (Hs/Ho hit ratios), considering that conservative sequence repeats will produce similarity hits with approximately the same frequency for Hs and Ho, while diversified sequence repeats will generate similarity hits with different frequencies (Macas et al., 2015). We also calculated similarity hit ratios for the 5S tandem-repeat rDNA to compare its gene-conserved vs IGS-variable Hs/Ho ratios with those obtained from the other repeat elements analyzed (Moreno-Aguilar et al., 2022).

Repeatome phylogenomic network of *Brachypodium*

We performed phylogenomic analyses with the repeat data obtained from the comparative clustering of the *Brachypodium* repetitive elements. The repeatome super-network was inferred following the steps described by Vitales et al. (2020a). The most abundant repeats (top 342 clusters), defined as possessing more than 0.01% of the total input reads in the dataset, were employed as the starting data set for phylogenetic analyses. Organelle clusters (plastid, mitochondrial) were removed prior to the phylogenetic inference. For each cluster, the initial data set consisted of the matrices of the observed/expected number of edges between species, which is a measure of the pairwise similarity between the species' reads. These matrices were extracted from the RE results folder using a custom Perl script (Moreno-Aguilar et al., 2022). Incomplete matrices lacking pairwise similarity or with zero values were excluded. These similarity matrices were transformed into distance matrices by calculating the inverse of the values. The NJ function from the *ape* v.5.4-1 package (Paradis and Schliep, 2019) in R was used to build the neighbor-joining trees for each of the 55 surviving clusters. The super-network was constructed using the default parameters in SplitsTree4 v.4.17.0 (Huson and Bryant, 2006). Potential phylogenetic information from the repeatome data set was assessed by topological comparisons of the repeatome network with *Brachypodium* phylogenomic trees retrieved from transcriptome (Sancho et al., 2022) and plastome and nuclear 35S and 5S rDNA data (Díaz-Pérez et al., 2018, and unpub. data).

Results

Characterization and quantification of the *Brachypodium* repeatome

The annotated repeats recovered by RE2 in the individual analysis showed remarkable differences in repeat types and contents among the 44 *Brachypodium* samples studied (Supplementary Tables S3, S4; Figure 1). The proportion of the monoploid genome occupied with repeats ranged from 67.97% (*B. mexicanum*-4x) to 20.77% (*B. stacei*-2x), with a genus-wide average of 28.65% (Supplementary Tables S3, S4; Figure 1). The amount and 1Cx-percent coverage of repetitive elements varied considerably

within both *Brachypodium* polyploids and diploids. Among the polyploids, the highest percentages corresponded to the ancestral outcore perennial tetraploid samples of *B. mexicanum* (56.6–67.9%; mean 60.81%), which exceeded those of all the remaining samples, followed by ancestral outcore perennial hexaploid samples of *B. boissieri* (29.35–32.2%; mean 30.92%), and the lowest to the tetraploid and hexaploid samples of the intermediately evolved *B. retusum* and the recently evolved core perennial *B. pinnatum*, *B. phoenicoides* and *B. rupestre* (22.5–27.9%), and the tetraploid annual sample of *B. hybridum* (22.05%). Within the intermediately-to-recently-evolved groups, some species showed higher, non-overlapping ranges of repeatome percentages in diploids and low polyploids than in high polyploids [e.g., *B. pinnatum* diploids (26.7%) vs tetraploids (22.5–24.9%; mean 23.9%), *B. phoenicoides* tetraploids (25.9–27.9%; mean 26.9%) vs hexaploids (22.7–24.3%; mean 23.2%)], while others showed overlapping ranges [e.g., *B. retusum* tetraploids (25.9–27.9; mean 27.2%) vs hexaploids (24.9–27.2%; mean 26.1%), *B. rupestre* tetraploids (23.4–25.4%; mean 24.6%) vs hexaploids (24.2–24.9%; mean 24.6%)] (Supplementary Tables S3, S4; Figure 1). Among the diploids, the highest percentages corresponded to the recent core-perennial *B. sylvaticum* samples (32.9–36.1%, mean 34.2%), which notably exceeded those of other core-perennial diploids (*B. pinnatum* 26.7%, *B. arbuscula* 22.54%) and of ancestral outcore diploid annuals (*B. distachyon* 22.75%, *B. stacei* 20.77%). The allotetraploid annual sample of *B. hybridum* showed a 1Cx-percentage coverage of repeatome equivalent to the mean between those of its diploid annual progenitor species *B. stacei* and *B. distachyon* (Supplementary Tables S3, S4; Figure 1).

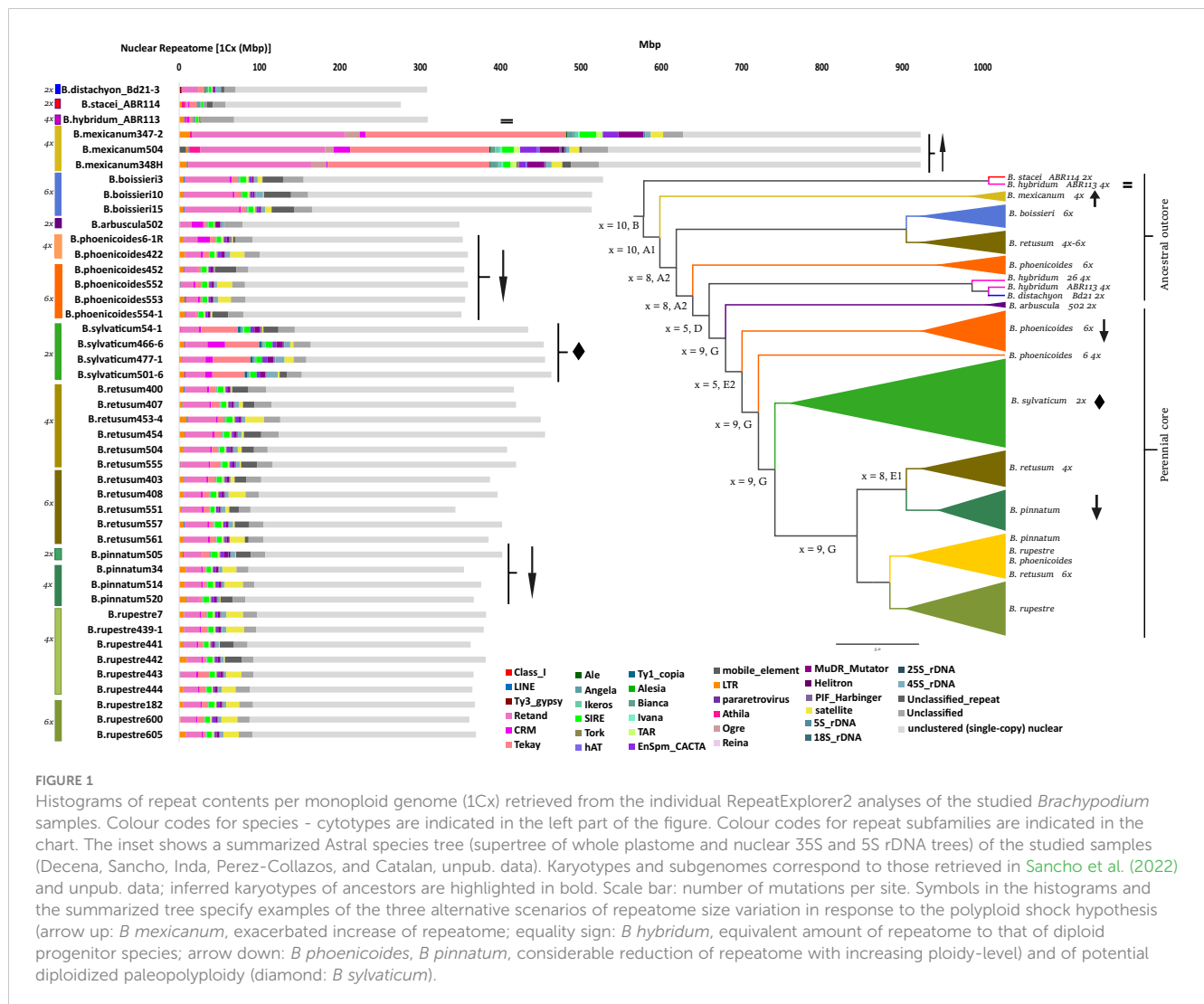
LTR-Gypsy retrotransposons represented the major repeat fractions in all *Brachypodium* genomes studied, followed by satellite repeats, LTR-Copia retrotransposons, and Class II TIR-transposons (Supplementary Tables S3, S4; Figure 1). The LTR-Gypsy Retand (mean 6.75%) and Tekay (3.91%) elements were the most represented repeats in all genomes. Of all elements, Retand repeats covered the highest percentages of genomes in almost all species and similar values in most core-perennial clade samples (16.6–20.7% *B. mexicanum*, 10.5–13.2% *B. boissieri*, 5.2–8.2% *B. retusum*, 4–6.2% core perennials, 6.8% *B. distachyon*, 1.8–0.06% *B. stacei*, *B. hybridum*; non-significant differences in Kruskal-Wallis tests, Supplementary Tables S3, S4). Tekay repeats showed considerable percent differences among species, being highly abundant in the *B. mexicanum* (18.8–27%) and *B. sylvaticum* (8.7–10.6%) genomes (more frequent than the Retand elements), and less abundant in annuals (1–3%) and the remaining core perennial genomes (0.6–3.7%) (significant differences in Kruskal-Wallis tests). LTR-Copia SIRE elements (1–2.2%; mean 1.65%) were relatively evenly distributed across *Brachypodium* genomes and showed similar coverage percentages to Tekay elements in most core perennial species (0.6–3.7%). Other types of repeats showed in general proportions <1% in most genomes with some exceptions in those of the *B. mexicanum* samples. Class II-TIR Mutator (0.05–3.4%; mean 0.98%) and CACTA transposons (0.2–2.3%; mean 0.86%) were also evenly distributed across the *Brachypodium* genomes, although the former were more abundant in *B. mexicanum* (Mutator: 2.8%; CACTA: 1.8%) and *B. sylvaticum* (Mutator: 1.64%; CACTA: 1.09%) than in the genomes of other

species. Similarly, LTR-Copia Ikeros (0.22%) and Angela elements (0.2%) were more frequent in *B. mexicanum* (~0.5%, ~0.7%) and *B. sylvaticum* (~0.38%, ~0.4%), and Angela also in the annual species (0.6–1.5%), than in the remaining *Brachypodium* genomes (<0.3%). LTR-Copia TAR elements (0.45%) were present in all genomes except *B. sylvaticum*, while Tork elements (0.18%) were absent in *B. sylvaticum*, most *B. mexicanum* and a few core perennial genomes. LTR-Gypsy Ogre elements were found exclusively in the *B. mexicanum* genomes (1.14–1.93%). LTR-Copia Bianca (0.17%), Ivana (0.05%), Ale (0.02%) and Alesia (<0.01%), LTR-Gypsy Athila (0.13%) and Reina (<0.01%), and Class II Harbinger (0.26%), Helitron (0.02%) and hAT (0.01%) elements were only residually present in a few genomes. Nonspecific tandem satellite repeats (2.23%) were generally well, moderately, or poorly represented in most *Brachypodium* genomes, although their frequencies were unevenly distributed among different groups (Supplementary Tables S3, S4). The variation in the satellite fraction between intraspecific samples showing the same cytotype could be due to different factors, such as multiple origins, but also their different dynamics (explosion vs deletion) in separate evolutionary lines, or even incomplete coverage of genomes/subgenomes by the genome scan data. The *B. mexicanum* obese genomes had the highest percentages of genome coverage for most repeat families (Supplementary Tables S3, S4; Figure 1). Kruskal-Wallis rank tests performed for each of the *Brachypodium* repeat elements found significant differences for the Tekay, Angela, Bianca, TAR, Tork, Helitron, and satellite repeats when examined across all samples (Supplementary Table S4).

Global variability and genomic landscape of the *Brachypodium* repeatome

Regression model analysis of repeat content and differences in monoploid genome size between *Brachypodium* samples showed a strong correlation when data from all major repeats were combined ($R^2 = 0.98$, $p < 2.2E-16$), which represents a 49.4% difference in genome size between species (Table 2; Figure 2). Most repetitive elements (22) presented high correlations. Among them, Retand had the highest correlation values ($R^2 = 0.96$, $p = 5.41E-30$), followed by Ikeros ($R^2 = 0.91$, $p = 9.07E-24$), Tekay ($R^2 = 0.87$, $p = 7.45E-20$), Mutator ($R^2 = 0.84$, $p = 2.08E-18$), Ogre ($R^2 = 0.83$, $p = 1.79E-17$), and others (Table 2). The repeat family that accounted for the highest contribution to pairwise differences in genome sizes was Retand (21.7%), followed by Tekay (6.69%), while contributions from the other repeats were <3% (Table 2; Supplementary Figure S2).

Global variability analysis of individual repeat types in *Brachypodium* genomes showed different histogram profiles of Hs/Ho hit ratios (Supplementary Figure S3). Using the histogram of 5S rDNA sequences as a reference, where a narrow main peak near zero on the log(Hs/Ho) x-axis indicated that the ratios of intraspecific Hs to interspecific Ho hit frequencies were close to one, reflecting hence the high sequence conservation of the 5S genes, while a wide right tail of log(Hs/Ho) values ranging from 0.1 to 3, indicated the high divergence of the 5S rDNA IGS sequences



(Moreno-Aguilar et al., 2022), the histograms of the ten analyzed repeats showed contrasting patterns. Although most histograms had overall Gaussian distributions for $\log(H_s/H_o)$ hit values, most of them presented main peak values >0.5 and a distribution skewed towards positive values 1-3 (Retand, Mutator, Ale, Ivana, TAR, SIRE, satellite), while the others had main peak values close to zero (Tekay, Ikeros, Ogre, CACTA) but also with tails skewed towards positive x-axis values (Supplementary Figure S3). These results suggested greater overall conservatism of Tekay sequences and greater diversification of Retand sequences in the *Brachypodium* genome landscape with respect to these two major repeats types, and similar dynamics for the other minor repeats (Table 2).

The *Brachypodium* repeatome phylogenetic network and 5S rDNA graph-clusters

Comparative analysis of RE2 repeats recovered different types and numbers of shared or species-specific repetitive elements in each *Brachypodium* lineage (Supplementary Tables S4, S5; Figure 1). RE2

annotated different numbers of top clusters in the studied taxa (342 *Brachypodium* clusters, 322 nuclear and 20 organellar; total number of reads in top clusters 2,914,070 (48.1% of total clustered reads) (Supplementary Table S5), representing presumably orthologous repeat families from different samples that were clustered due to their high repeat sequence similarity (Macas et al., 2015). We reduced the number of top clusters used to build the NJ trees to 55 clusters after discarding organellar clusters and clusters with NA or zero read values for some samples (Supplementary Table S5). The phylogenomic network constructed from the distance-based NJ trees revealed the clear divergences of the ancestral *Brachypodium* outcore lineages and the less resolved relationships of the recent core perennial lineages (Figure 3). Among the former group, a *B. mexicanum* cluster was highly isolated from the others, although it was more closely related to the also ancestral *B. stacei* lineage. The allotetraploid *B. hybridum* lineage nested between its two diploid progenitor species *B. stacei* and *B. distachyon* lineages, while the outcore *B. boissieri* cluster was placed close to the *B. distachyon* lineage (Figure 3). Within the intricate core-perennial group, a slightly older cluster included most *B. retusum* 4x and 6x samples, the diploid *B. arbuscula* lineage separated from the rest, all diploid *B. sylvaticum* samples grouped into an isolated cluster,

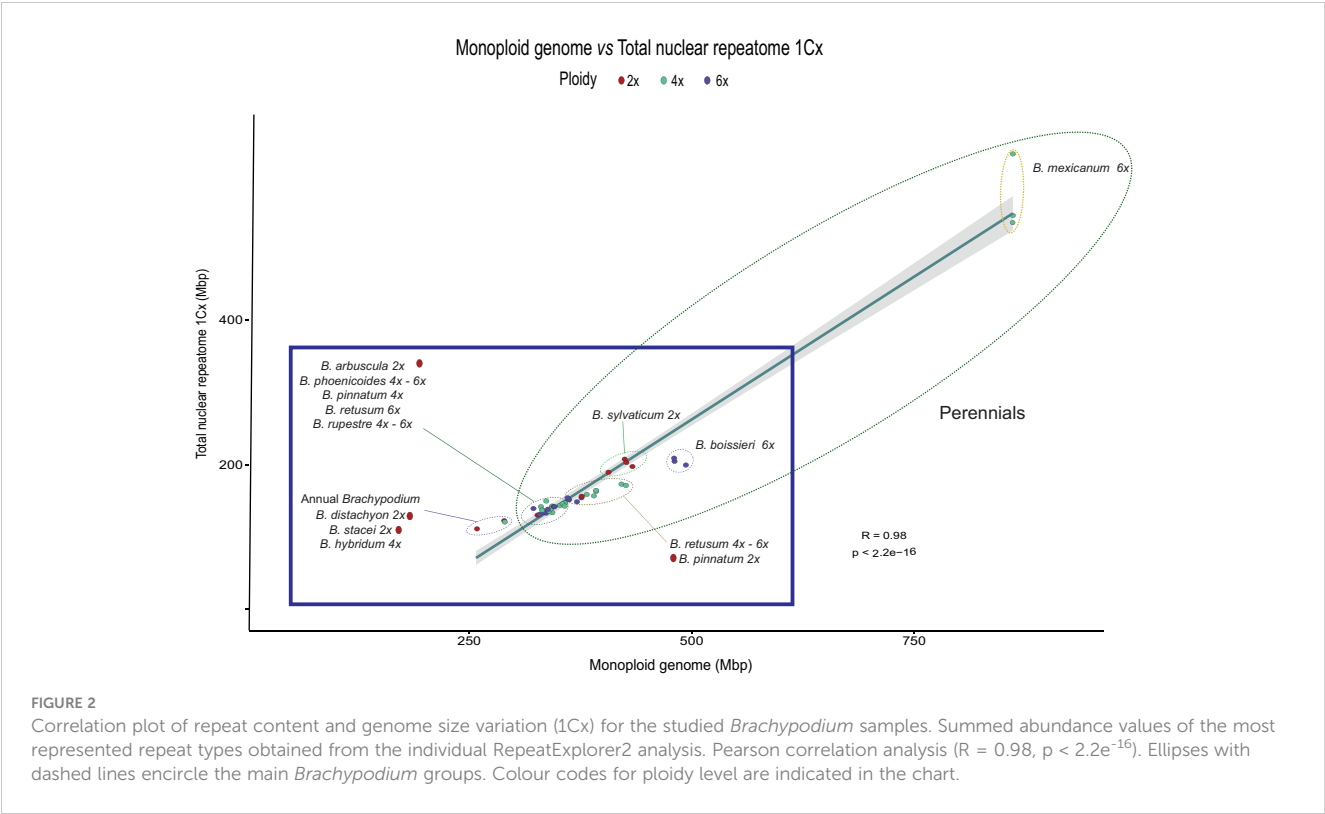
TABLE 2 Pearson linear correlation of repeat abundance with genome size variation (1Cx) in *Brachypodium* and contribution of individual repeats to the genome size differences between species.

Repeat	Correlation to genome size		Abundance in analysed genomes [Mb/1Cx]		Average contribution to pairwise differences in genome sizes [%]
	R ²	P-value	min.	max	
Retand	0.955	5.41E-30	0.19	191.01	21.7
Tekay	0.865	7.45E-20	1.97	249.42	6.69
SIRE	0.698	1.75E-12	2.92	20.48	1.62
MuDR_Mutator	0.842	2.08E-18	0.15	31.28	1.45
EnSpm_CACTA	0.765	8.92E-15	0.62	21.22	1.24
rDNA	0.169	5.62E-03	0.19	13.89	1.12
TAR	0.744	5.40E-14	0.00	8.03	0.551
CRM	0.129	1.66E-02	0.00	21.87	0.516
Ikeros	0.912	9.07E-24	0.34	6.09	0.45
Angela	0.344	2.90E-05	0.00	7.01	0.426
LTR	0.106	3.13E-02	0.13	12.92	0.388
PIF_Harbinger	0.13	1.61E-02	0.00	3.23	0.3
Bianca	0.578	2.18E-09	0.00	5.44	0.265
Ivana	0.784	1.51E-15	0.00	2.58	0.112
Ogre	0.825	1.79E-17	0.00	17.81	0
Ale	0.804	1.92E-16	0.00	1.85	0
hAT	0.517	3.83E-08	0.00	3.05	0
Athila	0.298	1.25E-04	0.00	13.38	0
mobile_element	0.298	1.26E-04	0.00	8.58	0
Helitron	0.269	3.10E-04	0.00	3.23	0
Reina	0.213	1.62E-03	0.00	0.18	0
satellite	0.0133	4.56E-01	0.03	22.99	1.32
Tork	0.0354	2.21E-01	0.00	3.14	0.13
LINE	0.0503	1.43E-01	0.00	0.87	0.0724
pararetrovirus	0.0635	9.90E-02	0.00	1.11	0.0109
Alesia	0.016	4.13E-01	0.00	0.28	0
Ty3_gypsy	0.016	4.13E-01	0.00	1.39	0
Ty1_copia	0.00277	7.35E-01	0.00	3.27	0
Class_I	0.000885	8.48E-01	0.00	0.46	0
Unclassified	0.326	5.12E-05	14.19	41.75	2.86
Unclassified repeat conflicting	0.004	6.83E-01	0.00	33.65	0.618
All	0.961	3.97E-31	57.76	627.19	49.4

Only the most represented *Brachypodium* repeat types are shown. Significant values are highlighted in bold.

and a more recent cluster included the representative samples of the *B. pinnatum* complex taxa (*B. pinnatum* 2x and 4x, *B. rupestre* 4x and 6x, *B. phoenicoides* 4x and 6x) (Figure 3).
The analysis of the 5S rDNA clusters of the 44 *Brachypodium* samples studied showed different types of simple and complex

graphs (Table 3; Figure 4) that corresponded to short (5S-S) and long (5S-L) 5S sequences. Comparative chromosome barcoding (CCB) analysis using FISH probes has shown that all *Brachypodium* polyploids studied to date, except the autohexaploid *B. boissieri* (subgenomes 8A2 + 8A2 + 8A2), are allopolyploid or



autoallopolyploids (Sancho et al., 2022, and unpub. data). These allopolyploids vary from those that have totally different subgenomic karyotypes [*B. hybridum* (4x): 10S + 5D; *B. retusum* (4x): 8A2 + 8E1; *B. phoenicoides* (4x) and *B. rupestre* (4x): 9G + 5E2; *B. retusum* (6x): 8A2 + 8E1 + 5E2] to those with similar segmental karyotypes [*B. mexicanum* (4x): 10A1.1 + 10A1.2], and those with some duplicated subgenomic karyotypes [*B. phoenicoides* (6x) and *B. rupestre* (6x): 9G + 5E2 + 5E2] (Lusinska et al., 2019; Sancho et al.

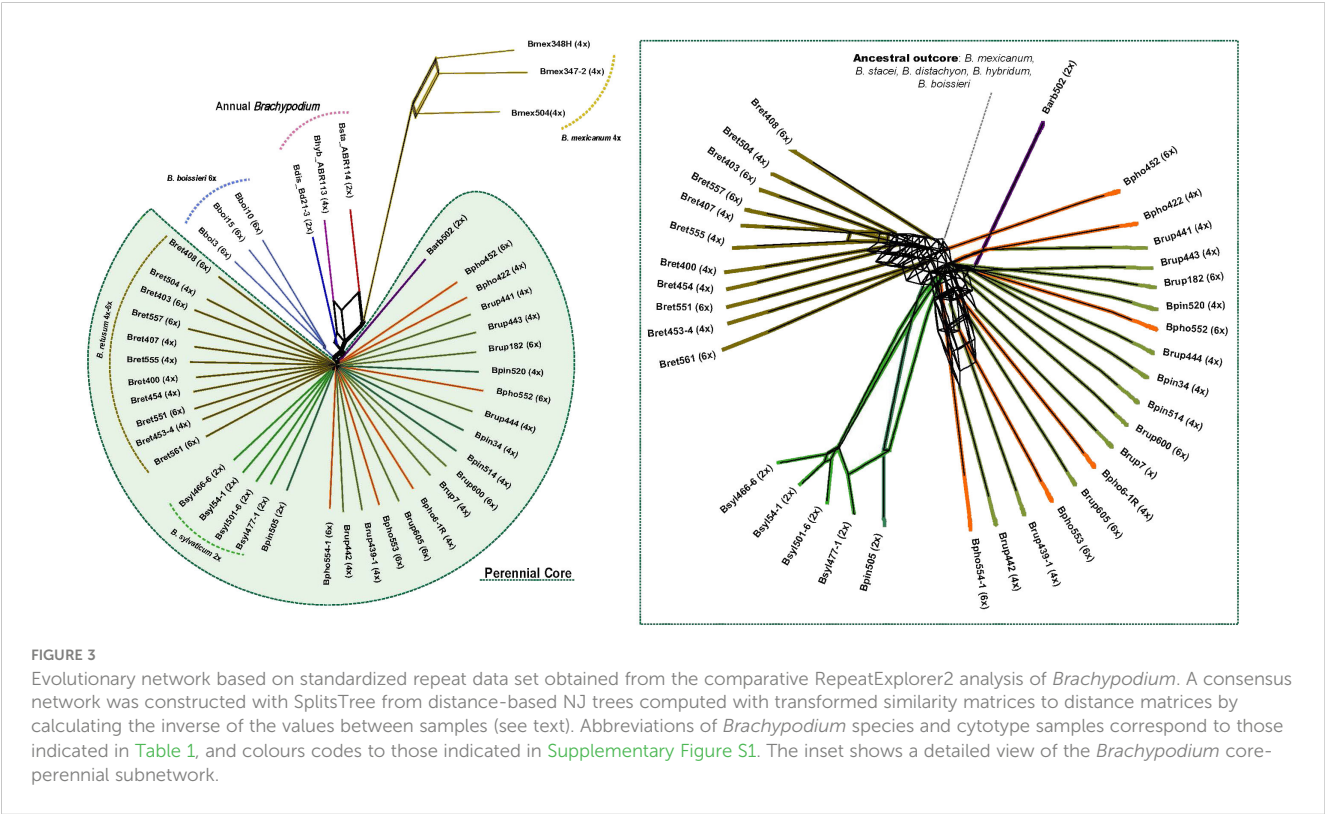


TABLE 3 Genomic pair-end read features of 5S rDNA loci and cluster graph parameters of the studied *Brachypodium* cytotypes.

Taxon-Cytotype	Code	No.Reads in cluster	Genome proportion (%)	Consensus repeat length (bp)	k-mer coverage	Connected component index C	Graph shape (type)
<i>B.stacei</i> -2x	Bsta_ABR114	970	0.071	270	0.974	0.997	1
<i>B.distachyon</i> -2x	Bdis_Bd21-3	1450	0.095	370	0.844	0.995	1
<i>B.hybridum</i> -4x	Bhyb_ABR113	315	0.01	272	0.818	0.8	2
<i>B.arbuscula</i> -2x	Barb502	4887	0.28	400	0.801	0.983	1
<i>B.boissieri</i> -6x	Bboi3	1153	0.1	271	0.82	0.989	1
	Bboi10	2072	0.16	271	0.823	0.964	2
	Bboi15	1253	0.12	271	0.815	0.974	1
<i>B.mexicanum</i> -4x	Bmex347	5483	0.18	273	0.758	0.992	1
	Bmex348H	6625	0.17	260	0.719	0.979	1
	Bmex504	9701	0.26	339	0.767	0.981	1
<i>B.phoenicoides</i> -4x	Bpho6-1R	9781	0.44	375	0.825	0.998	2
	Bpho422	2766	0.16	427	0.614	0.996	3
<i>B.phoenicoides</i> -6x	Bpho452	2048	0.1	430	0.712	0.976	2
	Bpho552	3282	0.17	376	0.767	0.967	2
	Bpho553	3404	0.15	378	0.585	0.986	2
	Bpho554-1	2356	0.11	376	0.717	0.999	2
<i>B.pinnatum</i> -2x	Bpin505	1580	0.079	373	0.926	0.98	1
<i>B.pinnatum</i> -4x	Bpin34	1028	0.059	373	0.919	0.989	2
	Bpin514	1425	0.1	373	0.817	0.965	2
	Bpin520	1427	0.072	373	0.929	0.999	2
<i>B.retusum</i> -4x	Bret400	1486	0.081	273	0.616	0.99	2
	Bret407	2298	0.12	276	0.669	0.932	3
	Bret453-4	1081	0.069	408	0.762	0.981	3
	Bret454	2938	0.17	373	0.577	0.938	3
	Bret504	2248	0.13	373	0.647	0.944	3
	Bret555	2196	0.13	373	0.614	0.957	3
<i>B.retusum</i> -6x	Bret403	2737	0.12	375	0.705	0.946	3
	Bret408	969	0.059	373	0.818	0.971	2
	Bret551	1684	0.094	274	0.542	0.943	3
	Bret557	2692	0.14	430	0.816	0.974	3
	Bret561	1776	0.1	395	0.563	0.97	2
<i>B.rupestre</i> -4x	Brup7	1908	0.11	373	0.81	0.971	2
	Brup439-1	1812	0.12	373	0.731	0.973	2
	Brup441	2391	0.13	373	0.86	0.967	1
	Brup442	2481	0.15	388	0.811	0.969	1
	Brup443	1788	0.097	373	0.877	0.977	1
	Brup444	2646	0.15	397	0.912	0.98	1
<i>B.rupestre</i> -6x	Brup182	2283	0.12	365	0.814	0.977	2

(Continued)

TABLE 3 Continued

Taxon-Cytotype	Code	No.Reads in cluster	Genome proportion (%)	Consensus repeat length (bp)	k-mer coverage	Connected component index C	Graph shape (type)
<i>B.sylvaticum</i> -2x	Brup600	1314	0.094	376	0.828	0.978	2
	Brup605	1067	0.07	373	0.668	0.993	2
	Bsyl54-1	4704	0.22	391	0.858	0.997	1
	Bsyl466-6	1672	0.074	392	0.862	0.995	1
	Bsyl477-1	3092	0.14	393	0.815	0.961	1
	Bsyl501-6	2089	0.091	393	0.752	0.956	1

Graph shape types (type 1, simple circular-shaped graph with one loop; types 2 and 3, complex graph with two and three loops, respectively, where the interconnected loops represent IGS spacers).

(2022); and unpub. data). Notably, in most cases the retrieved graphs matched the expected types for their respective ploidy levels (Table 1; Supplementary Table S1), the nature of the polyploidy (auto- vs segmental- vs allo-polyploidy), and the number and identity of the subgenomes (Sancho et al., 2022; unpub. data). Therefore, the graph topologies of the diploid taxa corresponded to a simple circular type-1 graph that probably represents a single 5S gene family and locus (outcore *B. stacei* and *B. distachyon* and core

perennial *B. arbuscula*, *B. pinnatum* and *B. sylvaticum* 2x samples). In contrast, most allotetraploid samples from *B. hybridum*, *B. pinnatum*, *B. phoenicoides* and *B. retusum* had complex type-2 graphs showing two IGS loops interconnected by a junction section (coding region of the 5S gene), suggesting they may have two 5S ribotypes. The separation of the two IGS loops was less clear in the graphs of the *B. mexicanum*-4x and *B. pinnatum*-4x samples, while the tetraploid *B. rupestre* samples showed both type-2 (Brup439-1,

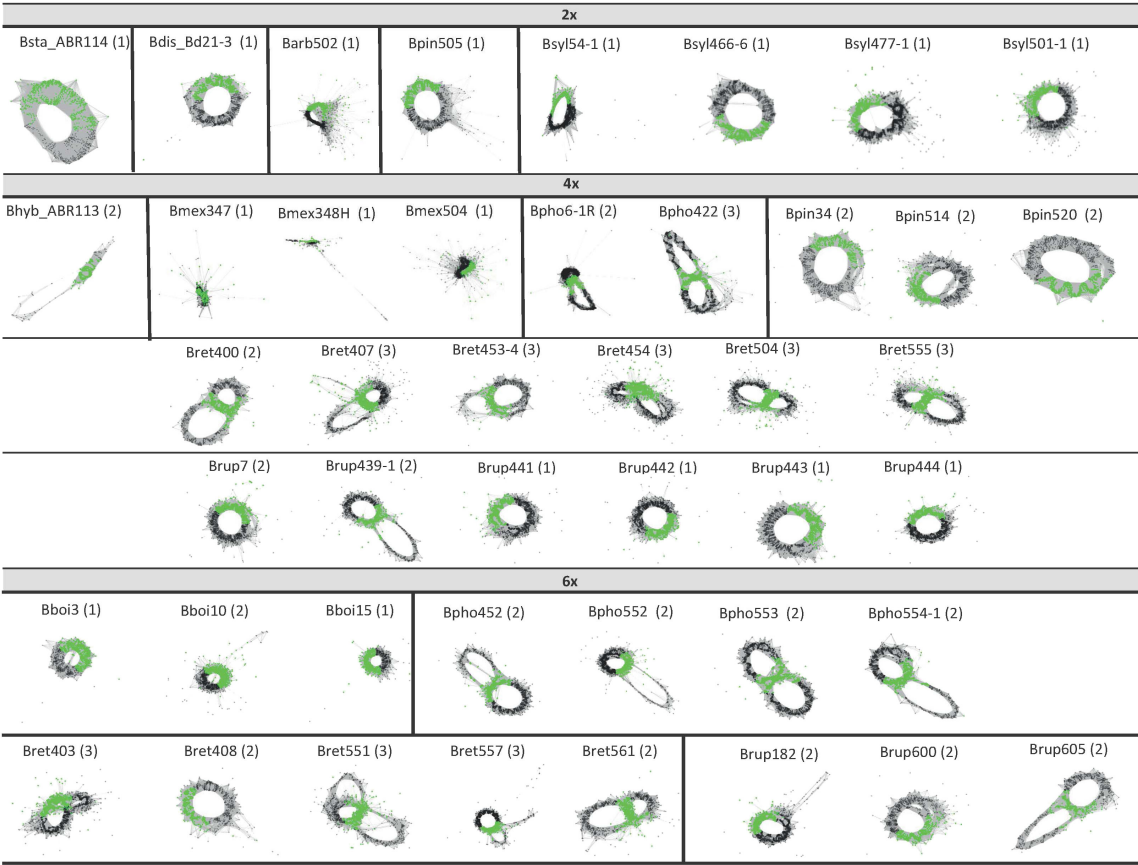


FIGURE 4
5S clustering graph plots of *Brachypodium* samples generated by the individual RepeatExplorer2 analysis sorted by ploidy level. Diploids (2x) show graph type 1, while some tetraploids and hexaploids show graph types 2 and 3, respectively (see also Table 3).

Brup7) as type-1 graphs (remaining samples) (Figure 4). Among the hexaploid taxonomic cytotypes, some *B. retusum*-6x samples (Bret403, Bret551, Bret557) presented complex type-3 graphs with three interconnected IGS loops, indicating that they might have three 5S loci, and the other samples (Bret408, Bret561) type-2 graphs, while all the *B. phoenicoides*-6x and *B. rupestre*-6x samples presented type-2 graphs and all *B. boissieri*-6x samples type-1 graphs (Figure 4). Only a few allotetraploid samples (Bpho422, Bret407) showed evidence of more IGS loops than expected (type-3 graphs) (Figure 4).

Discussion

Delineation of the *Brachypodium* repeatome and its impact on the striking genome size diversification of its lineages

Our comprehensive analysis of the *Brachypodium* repeatome has revealed the composition and frequency of the main repetitive DNA elements across the genome landscape of all its lineages (Supplementary Table S4; Figure 1). Our data confirm the decisive contribution of the repeatome to the genome size diversification of the studied *Brachypodium* genomes. The repeatome represents a major or considerable percentage of the holoploid genome of the surveyed samples. One of the most noticeable results was the enormous differences in genome sizes, and their correlated repeatome amounts, detected between species and lineages (Supplementary Tables S3, S4; Figures 1, 2). For a genus selected as a monocot model system due to the small genome size of its flagship species *B. distachyon* (IBI. The International *Brachypodium* Initiative, 2010; Gordon et al., 2017), differences between the smallest genome sizes found within its annual species (*B. stacei*, holoploid genome 551 Mbp, monoploid genome 275 Mbp), which has the lowest repeatome content (20%), and the largest genome sizes of the slender *B. mexicanum*-4x perennial samples (holoploid genome 3690 Mbp, monoploid genome 922 Mbp), presenting the highest repeatome contents (67.9%), are 6.7-fold and 3.3-fold, respectively (Table 1; Supplementary Tables S1, S3, S4). Although most of the genomes of the *Brachypodium* species analyzed are small (annual species, monoploid genome 275–309 Mbp) or relatively small (most core-perennial species; ≤ 352 Mbp) and their respective repeatome percentages are also consistently low (20.7–22.7% annuals; $\leq 27.9\%$ core perennials), *B. mexicanum*-4x plus the ancestral *B. boissieri*-6x (508 Mbp; 31%) and the recent core-perennial *B. sylvaticum*-2x (450 Mbp; 34%) lineages depart from this trend, and the intermediately evolved *B. retusum*-4x-6x (410 Mbp; 26.7%) also differs slightly from it (Supplementary Tables S3, S4; Figures 1, 2).

Surprisingly, the main differences in genome sizes and repeatome amounts have been found between the most ancestral $x=10$ karyotype lineages, the smallest genomes of *B. stacei* (S karyotype) and the largest genomes of *B. mexicanum* (P and U karyotypes) (Figure 1, Sancho et al., 2022). Although the genome (and repeatome) contractions observed in *B. stacei* and in the also ephemeral lineages of *B. distachyon* (intermediately evolved $x=5$ D

karyotype; Sancho et al., 2022) and *B. hybridum* (S+D karyotypes) (Figure 1) is a general feature detected in other annual angiosperms (Suda et al., 2015; Pellicer et al., 2018; Hloušková et al., 2019), the gross genomes of the weakly-rhizomatose perennial *B. mexicanum* and the strongly-rhizomatose perennial *B. boissieri* (ancestral $x=8$ A2 karyotype; Sancho et al., 2022) (Figure 1) points toward to a *Brachypodium* common ancestor with an expanded genome that preceded the diversification of its oldest outcore lineages. A similar evolutionary scenario has been hypothesized for the obese-genome ancestor of the *Hesperis* subclade (~1600 Mbp), within the otherwise small-genome Brassicaceae clade, which includes the model dicot *Arabidopsis thaliana* with one of the smallest genome sizes of angiosperms (157 Mbp; Hloušková et al., 2019). Our repeatome data, together with the extremely high collinearity of CCB syntenic blocks detected between the *B. stacei* and *B. mexicanum* chromosomes (Sancho et al., 2022) and high similarity of CCB karyotypes of the P and U subgenomes of *B. mexicanum* (A1.1 and A1.2 in Sancho et al., 2022), suggest that the 3.3-fold differences in the size of their monoploid genomes (for the same number of chromosomes) were caused by expansions of LTR-Gypsy retrotransposons in *B. mexicanum* chromosomes (probably coupled with some potential losses in the *B. stacei* chromosomes) and not by WGD (Supplementary Tables S3, S4; Figures 1, 2). The inflated genome of the mesopolyploid *B. mexicanum* (10.4–8.6 Ma; Sancho et al., 2022) likely resulted from the proliferation of Tekay (22–27%) and Retand (17–21%) repeat families, and the enrichment in other less abundant elements (Mutator, 2–3%; Angela, 0.7%, Ogre, 1–2%). Interestingly, Ogre retrotransposons, frequent in the genomes of dicotyledonous legumes (Macas et al., 2015) and also common in the genomes of Brassicaceae (Hloušková et al., 2019), were only residually present in some genomes of Loliinae grasses (Moreno-Aguilar et al., 2022), and have been found exclusively in *B. mexicanum* within our low-pass genomic survey of the genus (Supplementary Tables S3, S1; Figure 1). The relatively large genome of the also ancestral *B. boissieri* (5.4–3.7 Ma, Sancho et al., 2022) was probably the result of the burst of Retand retrotransposons (10–13%), which were also predominant but less enriched in the genomes of the intermediately evolved strong-rhizomatous perennial *B. retusum* (5–8%). The overall decrease in the amounts of Retand ($\leq 6\%$) and other repetitive elements in the genomes of the core-perennial diploids and their derived neopolyploid lineages (Table 2; Supplementary Table S3; Figure 1) was likely a consequence of post-WGD diploidizations and genome downsizings due to the removal of the excess of repeats (Michael, 2014; Hloušková et al., 2019).

The large genome reductions observed in annual *Brachypodium* species of ancestral origin (Supplementary Tables S3, S4; Figure 1) could also be related to the transition in life form. Evidence suggests that annuality has evolved convergently from perenniality in different lineages of flowering plants, and that it could have been facilitated by evolutionary precursors (correlated developmental, physiological, and genomic traits) in the temperate pooid grasses, which also include *Brachypodium* (Hjertaas et al., 2023). It has also been demonstrated that plants with small genomes can grow in more diverse habitats and tend to be annuals, while those with large genomes are restricted to

narrow ecological niches and are perennials (Suda et al., 2015; Pellicer et al., 2018). Although the annual *Brachypodium* species share similar mesic and arid habitats and ranges as other Mediterranean perennial relatives (Catalan et al., 2016), they show shorter generation times and therefore greater dispersal ability and long-distance colonization of new niches and continents than perennials (Scholthof et al., 2018), likely facilitated by their extremely reduced genomes (Supplementary Tables S3, S4; Figure 1). The large dysploid reduction from the $x=10$ ancestral S karyotype of *B. stacei* to the $x=5$ D intermediate karyotype of *B. distachyon* resulted from four nested chromosome fusions; however, the high collinearity of the two genomes was corroborated by their almost similar genomic sizes (Gordon et al., 2020). Our repeatome analysis further support the analogous genomic coverage of repeatome in these diploids (*B. stacei* 20.7%; *B. distachyon* 22.7%), with the differences caused by a possible recent proliferation of Retand elements in the youngest lineage (*B. stacei* 1.2%; *B. distachyon* 6.8%) while the Tekay and Angela elements were slightly higher in the oldest lineage (*B. stacei* 3.3% and 1.5% vs *B. distachyon* 2.6% and 0.9%) (Supplementary Tables S3, S4; Figure 1). Our data reinforce the findings of Stritt et al. (2020) on the main contribution of Retand elements to the variation of genome sizes among *B. distachyon* accessions, extending it to the entire genus level. These authors also postulated a high dynamic activity and source of intraspecific polymorphisms of very young Angela elements in the *B. distachyon* genome landscape; however, our Hs/Ho ratios indicated a greater diversification of Retand and, to a lesser extent, Tekay sequences in *Brachypodium* genomes (Supplementary Figure S3), likely due to the low contribution of the Angela repeats to the genome landscape at the genus level (Supplementary Table S3; Table 3). Our analysis has also confirmed the low repeatome content of the annual allotetraploid *B. hybridum* (karyotype $x=10S + 5D$), which showed a balanced percentage (22%) between those of its diploid progenitor species (Supplementary Tables S3, S4).

The striking large repeatome coverage of the recently evolved *B. sylvaticum* core-perennial diploid genome (34.2%) relative to other diploid (24.7%) and polyploid (<26%) core-perennial genomes sharing the same recently evolved karyotype $x=9$ (Sancho et al., 2022) does not correlate with parallel differences in 1Cx genome sizes, which have similar values for *B. sylvaticum*-2x (456 Mbp) as for *B. pinnatum*-2x (401 Mbp) and other core-perennial 2x-4x-6x cytotypes (349-382 Mbp) (Table 1, Supplementary Tables S1, S3, S4). This unexpected result could be a consequence of a relatively recent polyploidization and subsequent diploidization of the western lineage of *B. sylvaticum* from the Late Pliocene - Early Pleistocene (2.78 – 2.17 Ma; Figure 1; Catalán et al., 2023). *B. sylvaticum* samples showed a proliferation of Tekay retrotransposons (10%) compared to the other core perennial lineages (1-2%), and also higher proportions of Mutator, Ikeros and Angela elements (Supplementary Tables S3, S4). This finding, together with other cytogenetic characteristics, such as a greater number of 25 rDNA loci than expected for a diploid (4-6; Wolny and Hasterok, 2009; Breda et al., 2012; Garcia et al., 2017) would suggest a probable post-polyploid diploidized origin. High repeatome coverages were also found in diploidized paleo-polyploids of Loliinae grasses (Moreno-Aguilar et al., 2022).

Alternative evolutionary responses to the polyploid genome shock hypothesis by different *Brachypodium* allopolyploids

Our study has shown that expansions and contractions in the repeatome are responsible for the three contrasting responses to the PGSH in different allopolyploid *Brachypodium* lineages and that each response was caused by different biological, cytological and temporal scenarios.

Therefore, the exacerbated genome expansion of the old allotetraploid *B. mexicanum* was not a consequence of WGD *per se* but more likely of proliferations of Tekay and Retand retrotransposons and other repetitive DNA elements in the genomes of its progenitor species (Supplementary Tables S3, S4; Figure 1). TE annotations in the reference genome of this species (Bmex347; Sancho et al., unpub. data) indicated that amplifications of the LTR and other transposable elements were not only limited to (peri)centromeric regions but also to telomeres and chromosome arms. Since TEs could intersperse with coding regions (Hloušková et al., 2019), this distribution would support amplifications of the entire *B. mexicanum* chromosomes through its $x=10$ karyotype. It is still surprising why the ancestral genome of *B. mexicanum* has the propensity to tolerate or benefit from such repeatome bloating and subsequent genome expansions, which were probably inherited from the common ancestor but not eliminated over time. A plausible explanation could be that the increase in the length of the chromosome arms was compensated by an increase in centromere size and copy number of centromeric tandem repeats that would guarantee correct segregation of obese chromosomes during cell division as it is observed in other grasses (Zhang and Dawe, 2012).

In contrast, the additive repeatome pattern exhibited by the annual allotetraploid *B. hybridum* relative to those of its genome-reduced diploid progenitor species (Supplementary Tables S3, S4; Figure 1) is a likely response to post-WGD stabilized genome evolution. The three detected independent recurrent origins of this young neopolyploid, spanning the last 1.4 Mya to 20 Kya, ended in the same phenotypic allotetraploid that showed no evidence of homeologous recombination, subgenomic dominance or pronounced TE activations (Gordon et al., 2020; Scarlett et al., 2023; Mu et al., 2023a). It was probably caused by the high evolutionary and structural divergence of the *B. stacei* and *B. distachyon* progenitor genomes (karyotypes $x=10S$ and $x=5D$) which probably favored the non-recombinant integrity of the resulting subgenomes in the hybrid and the immediate creation of the amphidiploids (Mu et al., 2023a).

The observed reductions in repeatome and genome size with increasing level of ploidy in the recently evolved core-perennial *B. pinnatum* and *B. phoenicoides* lineages (Supplementary Tables S3, S4; Figure 1) are likely the result of loss of repeats through recombination that resulted in the repeatome contraction (Michael, 2014). Although the polyploid cytotypes of *B. phoenicoides* share a recent subgenome with a *B. pinnatum*-type diploid karyotype ($x=9$), a second intermediately subgenome with a reduced karyotype ($x=5E2$) is present once in the allotetraploids ($x=9G + 5E2$) and twice in the allohexaploids ($x=9G + 5E2 + 5E2$).

(Sancho et al., 2022; unpub. data), thus favoring more frequent recombination between identical or very similar subgenomes and therefore more potential repeatome losses in high polyploids. Parallel to the case of the highly hybridogenic high-polyploid Loliinae lineages, which experienced large genomic rearrangements causing massive repeatome and genome contractions (Moreno-Aguilar et al., 2022), the *B. pinnatum* and *B. phoenicoides* lineages of the core-perennial clade also showed high rates of interspecific hybridization (Khan and Stace, 1999), thus favoring reductions in their repetitive elements and genomes.

Repeatome-based phylogenomics and concordance between 5S rDNA graphs and *Brachypodium* (sub)genomes

As in previous angiosperms studies (Dodsworth et al., 2015; McCann et al., 2018, 2020; Viales et al., 2020a; Herklotz et al., 2021; Moreno-Aguilar et al., 2022), shared repeat clusters retrieved from the *Brachypodium* RE2 comparative analysis have demonstrated to contain phylogenetic information for its main lineages (Figures 1, 4; Supplementary Tables S4, S5). The topology of the repeatome network constructed from independent distance-based NJ trees (Figure 3) is highly congruent with those based on plastome and 35S and 5S rDNA gene trees (Figure 1; Díaz-Pérez et al., 2018; and unpub. data) and the *Brachypodium* transcriptome-based subgenomic tree (Sancho et al., 2022). The unrooted network showed the great divergences of the ancestral outcore lineages and the recent separations of the core-perennial lineages (Figure 3). The network recovered the high isolation of the *B. mexicanum* lineage from the other lineages; this large divergence resulted from the higher amounts of repeats for the common elements of some repetitive families within the representatives of the genus (Supplementary Table S5). The *B. mexicanum* group included two closely related samples from Mexico and a less related sample from Ecuador (Figure 3). Although all the *B. mexicanum* samples studied show a similar repeat composition (Figure 1), the divergence of the South American Andean sample from the North American Mexican samples coincides with that observed in previous phylogenetic analyses (Díaz-Pérez et al., 2018), indicating the plausible existence of two geographically separated lineages. The closeness of the *B. stacei* lineage to the *B. mexicanum* cluster supports the shared ancestry of the two $x=10$ karyotypes, and the intermediate location of *B. hybridum* between its two progenitor lineages reinforces its additive pattern. Interestingly, the ancestral *B. boissieri* cluster was resolved as closer to the *B. distachyon* lineage than to the more ancestral *B. mexicanum* and *B. stacei* lineages, matching the relationships recovered in the plastome tree but diverging from that of the nuclear transcriptome trees (Figure 3; Sancho et al., 2022). Therefore, the repeatome data also support the additional contribution of a more recently evolved ancestor to the nuclear genome of this putative ancestral autohexaploid species (karyotype $x=8A2$). Within the recently evolved core-perennial group, the earlier divergence of the *B. retusum* cluster from the rest supports its intermediate evolutionary position between the outcore *B. boissieri* and core-

perennial lineages. However, its greater proximity to the core-perennial group suggests a mixed pattern of repeats from ancestral ($x = 8A2$) and more recent progenitors ($x = 8E1 + 5E2$) or a convergent evolution towards recent core-perennial progenitor repeats (Figure 3), in parallel with plastome- and transcriptome-based findings (Sancho et al., 2022). The respective divergences of the *B. arbuscula* and *B. sylvaticum* lineages from the rest, and the closeness of the taxa and cytotypes of the *B. pinnatum* complex (*B. pinnatum*, *B. rupestre*, *B. phoenicoides*) were probably the result of their specific repeat compositions, particularly those of the highly differentiated *B. sylvaticum* group, and coincided with those recovered from plastomes and rDNA genes (Figures 1, 3; Díaz-Pérez et al., 2018; Sancho et al., 2022; and unpub. data).

The 5S rDNA graph topologies (Table 3; Figure 4) showed a great match with the number and nature of the genomes and subgenomes of the *Brachypodium* samples studied (Figure 1; Sancho et al., 2022; unpub. data), corroborating their value to uncover ploidy levels, ancient 5S families, and known and orphan subgenomes in angiosperms (García et al., 2020; Vozárová et al., 2021; Moreno-Aguilar et al., 2022). The single type-1 graphs of diploids corresponded to their respective extant monoplod genomes [*B. stacei*, $x = 10S$; *B. distachyon*, $x = 5D$; *B. arbuscula*, *B. pinnatum*, *B. sylvaticum*, $x = 9G$], type-2 graphs represented the two different subgenomes of allotetraploids [*B. hybridum*, $x = 10S + 5D$ (both extant); *B. phoenicoides*, $x = 9G$ (extant) + $5E2$ (orphan); *B. retusum*, $x = 8A2 + 8E1$ (both orphan)], and type-3 graphs to the three different subgenomes of some allohexaploids [*B. retusum*, $x = 8A2 + 8E1 + 5E2$ (all orphan)] (Figure 1; unpub. data). Interestingly, the poorly resolved type-2 graphs of *B. mexicanum*-4x may indicate that they correspond to close 5S ribotypic families, which coincides with the two similar P and U $x=10$ orphan subgenomes of this putative ancestral segmental allotetraploid. The type-2 graphs of the *B. phoenicoides*-6x and *B. rupestre*-6x samples would correspond to their two different ribotypes and subgenomes $x = 9G$ (extant) and $x = 5E2$ (orphan) plus a duplicated $x = 5E2$ copy (orphan) in these auto-/allohexaploids, and the type-1 graph of *B. boissieri*-6x samples to the triplicated $x = 8A2$ subgenomes of this ancestral autohexaploid (Figures 1, 4; Sancho et al., 2022; unpub. data). The maintenance of 5S rDNA loci in high allopolyploid *Brachypodium* species is consistent with their conserved patterns in other angiosperm allopolyploids (García et al., 2017). The few cases of fewer 5S graph loops than expected, according to ploidy level and distinct subgenomes (*B. rupestre*-6x; type-1 graphs), could be due to convergent evolution or, more likely, to failure of the low-pass genome sequencing in the detection of the different 5S IGS sequences. In contrast, the few cases of more 5S graph loops than expected (allotetraploids Bpho422 and Bret407; type-3 graphs) could be a consequence of intragenomic IGS heterogeneity of any of the 5S loci (García et al., 2020).

Conclusions

A genus-wide analysis of the repetitive elements in the genomes of model *Brachypodium* grasses has uncovered three alternative evolutionary scenarios to the PGSH (expansion, stasis, and

contractions of repeatome). None of them are related to WGD but instead reflect parental legacies (*B. mexicanum*, *B. hybridum*) or contraction through recombination in highly hybridogenous polyploids (*B. pinnatum*, *B. phoenicoides*). The model perennial species *B. sylvaticum* may be a diploidized ancestral polyploid. The 5S rDNA graphs describe the types and copies of genomes present in *Brachypodium* species and cytotypes.

Data availability statement

The datasets presented in this study can be found in the online repository https://github.com/Bioflora/Brachypodium_Repeatome. The accession number(s) can be found in the article/Supplementary Material.

Author contributions

MD: Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing – review & editing. RS: Data curation, Formal analysis, Investigation, Methodology, Software, Supervision, Validation, Writing – original draft, Writing – review & editing. LI: Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing – review & editing. EP-C: Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing – review & editing. PC: Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study was funded by the Spanish Ministry of Science and Innovation PID2022-140074NB-I00, TED2021-131073B-I00 and PDC2022-133712-I00, and by the Spanish Aragon Government LMP82-21 and Bioflora A01-23R research grants. MD was supported by a Spanish Ministry of Science predoctoral FPI fellowship. RS was supported by a Bioflora postdoctoral contract.

Acknowledgments

We thank Joaquin Ascaso, Mario Mairal and Manuel Pimentel for assistance during field trip sampling, Ruben Gorgojo and Beatriz Montes for their help with chromosome counting and flow cytometry analyses, the INIA, JGI and USDA germplasm banks and the Herbarium B for providing seeds or vouchers for some of

the *Brachypodium* samples studied, and two reviewers for their valuable comments to an early version of the manuscript. Genome skimming data from the studied *Brachypodium* samples were generated at the Centro Nacional de Análisis Genómicos (CNAG) and MacroGen laboratories (Spain). The repeatome and evolutionary analyses were performed in the Bioflora laboratory of the Escuela Politécnica Superior de Huesca (EPSHU, Universidad de Zaragoza, Spain).

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.

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/fpls.2024.1419255/full#supplementary-material>

SUPPLEMENTARY FIGURE 1

Geographical distribution of the studied 44 *Brachypodium* samples. (see Table 1, Supplementary Table S1). Colour codes for taxa and symbol codes for ploidy level (diploid: circle, tetraploid: triangle, hexaploid: square) are indicated in the corresponding charts. (A) *B. mexicanum*. (B) *B. arbuscula*, *B. boissieri*, *B. distachyon*, *B. hybridum*, *B. rupestre*, *B. stacei*. (C) *B. phoenicoides*, *B. pinnatum*, *B. retusum*, *B. sylvaticum*.

SUPPLEMENTARY FIGURE 2

Correlation plots of repeat content and genome size variation (1Cx) for the studied *Brachypodium* samples. Individual plots for the most represented repeat types found across the *Brachypodium* species and cytotype samples (see Table 2, Supplementary Table S3, Figure 2). Colour codes for species – cytotypes are indicated in the chart.

SUPPLEMENTARY FIGURE 3

Global variability of the main repeat types and their sequence conservation across the *Brachypodium* genomes. Histograms showing distributions of read similarity (Hs/Ho) hit ratios [frequencies of read similarity hits to reads from the same species (Hs) or to reads from all other species (Ho) (log scale, x-axis) and the number of reads (y-axis)]. Hs/Ho ratios close to one (0 on the logarithmic scale) indicate sequence conservatism while larger values indicate sequence diversification.

References

- Bird, K. A., VanBuren, R., Puzey, J. R., and Edger, P. P. (2018). The causes and consequences of subgenome dominance in hybrids and recent polyploids. *New Phytol.* 220, 87–93. doi: 10.1111/nph.15256
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics* 30 (15), 2114–2120. doi: 10.1093/bioinformatics/btu170
- Breda, E., Wolny, E., and Hasterok, R. (2012). Intraspecific polymorphism of ribosomal DNA loci number and morphology in *Brachypodium pinnatum* and *Brachypodium sylvaticum*. *Cell. Mol. Biol. Lett.* 17, 526–541. doi: 10.2478/s11658-012-0025-4
- Catalán, P., Ángeles Decena, M., Sancho, R., Viruel, J., Pérez-Collazos, E., Inda, L. A., et al. (2023). Phylogenetics of the Palearctic model grass *Brachypodium sylvaticum* uncovers two divergent oriental and occidental micro-taxa lineages. *Bot. Pacifica* 12, 21–28. doi: 10.17581/bp.2023.12119
- Catalán, P., López-Álvarez, D., Díaz-Pérez, A., Sancho, R., and López-Herránz, M. L. (2016). “Phylogeny and Evolution of the Genus *Brachypodium*,” in *Genetics and genomics of Brachypodium*. Ed. J. Vogel (Springer, New York), 9–38. doi: 10.1007/7997_2015_17
- Chalhoub, B., Denoeud, F., Liu, S., Parkin, I. A. P., Tang, H., Wang, X., et al. (2014). Early allopolyploid evolution in the post-neolithic Brassica napus oilseed genome. *Science* 345, 950–953. doi: 10.1126/science.1253435
- Chen, Z. J. (2007). Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. *Annu. Rev. Plant Biol.* 58, 377–406. doi: 10.1146/annurev.arplant.58.032806.103835
- Chen, Z. J., Sreedasyam, A., Ando, A., Song, Q., De Santiago, L. M., Hulse-Kemp, A. M., et al. (2020). Genomic diversifications of five Gossypium allopolyploid species and their impact on cotton improvement. *Nat. Genet.* 52, 525–533. doi: 10.1038/s41588-020-0614-5
- Chen, C., Wu, S., Sun, Y., Zhou, J., Chen, Y., Zhang, J., et al. (2024). Three near-complete genome assemblies reveal substantial centromere dynamics from diploid to tetraploid in *Brachypodium* genus. *Genome Biol.* 25, 1–28. doi: 10.1186/s13059-024-03206-w
- Contreras-Moreira, B., Cantalapedra, C. P., García-Pereira, M. J., and Sancho, R. (2016). *Split_pairs.pl v.0.5*. Available at: https://github.com/cead-csic-compbio/split_pairs.
- Deb, S. K., Edger, P. P., Pires, J. C., and McKain, M. R. (2023). Patterns, mechanisms, and consequences of homoeologous exchange in allopolyploid angiosperms: a genomic and epigenomic perspective. *New Phytol.* 238, 2284–2304. doi: 10.1111/nph.18927
- De Storme, N., and Mason, A. (2014). Plant speciation through chromosome instability and ploidy change: Cellular mechanisms, molecular factors and evolutionary relevance. *Curr. Plant Biol.* 1, 10–33. doi: 10.1016/j.cpb.2014.09.002
- Díaz-Pérez, A., López-Álvarez, D., Sancho, R., and Catalán, P. (2018). Reconstructing the origins and the biogeography of species' genomes in the highly reticulate allopolyploid-rich model grass genus *Brachypodium* using minimum evolution, coalescence and maximum likelihood approaches. *Mol. Phylogenet. Evol.* 127, 256–271. doi: 10.1016/j.ympev.2018.06.003
- Dodsworth, S., Chase, M. W., Kelly, L. J., Leitch, I. J., Macas, J., Novak, P., et al. (2015). Genomic repeat abundances contain phylogenetic signal. *Syst. Biol.* 64, 112–126. doi: 10.1093/sysbio/syu080
- Doležel, J., Greilhuber, J., and Suda, J. (2007). Estimation of nuclear DNA content in plants using flow cytometry. *Nat. Protoc.* 2, 2233–2244. doi: 10.1038/nprot.2007.310
- Doyle, J., and Doyle, J. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* 19, 11–15.
- Drouin, M., Hénault, M., Hallin, J., and Landry, C. R. (2021). Testing the genomic shock hypothesis using transposable element expression in yeast hybrids. *Front. Fungal Biol.* 2. doi: 10.3389/ffunb.2021.729264
- 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
- Fedoroff, N. V. (2012). Transposable elements, epigenetics, and genome evolution. *Science* 338, 758–767. doi: 10.1126/science.338.6108.758
- García, S., Kovařík, A., Leitch, A. R., and Garnatje, T. (2017). Cytogenetic features of rDNA genes across land plants: analysis of the Plant rDNA database. *Plant J.* 89, 1020–1030. doi: 10.1111/tpj.13442
- García, S., Wendel, J. F., Borowska-Zuchowska, N., Ainouche, M., Kuderova, A., and Kovarik, A. (2020). The utility of graph clustering of 5S ribosomal DNA homoeologs in plant allopolyploids, homoploid hybrids, and cryptic introgressants. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.00041
- Gordon, S. P., Contreras-Moreira, B., Levy, J. J., Djamei, A., Czedik-Eysenberg, A., Tartaglio, V. S., et al. (2020). Gradual polyploid genome evolution revealed by pan-genomic analysis of *Brachypodium hybridum* and its diploid progenitors. *Nat. Commun.* 11, 1–16. doi: 10.1038/s41467-020-17302-5
- Gordon, S. P., Contreras-Moreira, B., Woods, D. P., Des Marais, D. L., Burgess, D., Shu, S., et al. (2017). Extensive gene content variation in the *Brachypodium distachyon* pan-genome correlates with population structure. *Nat. Commun.* 8, 2184. doi: 10.1038/s41467-017-02292-8
- Hasterok, R., Catalan, P., Hazen, S. P., Roulin, A. C., Vogel, J. P., Wang, K., et al. (2022). *Brachypodium*: 20 years as a grass biology model system; the way forward? *Trends Plant Sci.* 27, 1002–1016. doi: 10.1016/j.tplants.2022.04.008
- Hawkins, J. S., Proulx, S. R., Rapp, R. A., and Wendel, J. F. (2009). Rapid DNA loss as a counterbalance to genome expansion through retrotransposon proliferation in plants. *Proc. Natl. Acad. Sci. U. S. A.* 106, 17811–17816. doi: 10.1073/pnas.0904339106
- Herklotz, V., Kovařík, A., Wissemann, V., Lunerová, J., Vozárová, R., Buschmann, S., et al. (2021). Power and weakness of repetition – evaluating the phylogenetic signal from repeatomes in the family rosaceae with two case studies from genera prone to polyploidy and hybridization (Rosa and fragaria). *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.738119
- Hidalgo, O., Pellicer, J., Christenhusz, M., Schneider, H., Leitch, A. R., and Leitch, I. J. (2017). Is there an upper limit to genome size? *Trends Plant Sci.* 22, 567–573. doi: 10.1016/j.tplants.2017.04.005
- Hjertaas, A. C., Preston, J. C., Kainulainen, K., Humphreys, A. M., and Fjellheim, S. (2023). Convergent evolution of the annual life history syndrome from perennial ancestors. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1048656
- Hloušková, P., Mandáková, T., Pouch, M., Trávníček, P., and Lysak, M. A. (2019). The large genome size variation in the Hesperis clade was shaped by the prevalent proliferation of DNA repeats and rarer genome downsizing. *Ann. Bot.* 124, 103–120. doi: 10.1093/aob/mcz036
- Huson, D. H., and Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* 23 (2), 254–267. doi: 10.1093/molbev/msj030
- IBI. The International *Brachypodium* Initiative (2010). Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature* 463, 763–768. doi: 10.1038/nature08747
- Jenkins, G., and Hasterok, R. (2007). BAC ‘landing’ on chromosomes of *Brachypodium distachyon* for comparative genome alignment. *Nat. Protoc.* 2, 88–98. doi: 10.1038/nprot.2006.490
- Kassambara, A. (2023). ggpubr: ‘ggplot2’ Based Publication Ready Plots. *R package version 0.6.0*. Available online at: <https://rpkgs.datanovia.com/ggpubr/>.
- Khan, M. A., and Stace, C. A. (1999). Breeding relationships in the genus *Brachypodium* (Poaceae: Pooideae). *Nordic J. Bot.* 19, 257–269. doi: 10.1111/j.1756-1051.1999.tb01108.x
- Lei, L., Gordon, S. P., Liu, L., Sade, N., Lovell, J. T., Del Mar Rubio Wilhelmi, M., et al. (2024). The reference genome and abiotic stress responses of the model perennial grass *Brachypodium sylvaticum*. *G3 Genes Genomes Genet.* 14, 1–15. doi: 10.1093/g3journal/jkad245
- Li, M., Copeland, A., and Han, J. (2011). *DUK - A Fast and Efficient Kmer Based Sequence Matching Tool*. (United States: Conference: DOE JGI User Meeting).
- Li, Y., Zuo, S., Zhang, Z., Li, Z., Han, J., Chu, Z., et al. (2018). Centromeric DNA characterization in the model grass *Brachypodium distachyon* provides insights on the evolution of the genus. *Plant J.* 93, 1088–1101. doi: 10.1111/tpj.13832
- Lusinska, J., Betekhtin, A., Lopez-Alvarez, D., Catalan, P., Jenkins, G., Wolny, E., et al. (2019). Comparatively barcoded chromosomes of *Brachypodium* perennials tell the story of their karyotype structure and evolution. *Int. J. Mol. Sci.* 20, 1–19. doi: 10.3390/ijms20225557
- Macas, J., Novak, P., Pellicer, J., Cizkova, J., Koblikova, A., Neumann, P., et al. (2015). In depth characterization of repetitive DNA in 23 plant genomes reveals sources of genome size variation in the legume tribe fabaeae. *PLoS One* 10, 1–23. doi: 10.1371/journal.pone.0143424
- Mason, A. S., and Wendel, J. F. (2020). Homoeologous exchanges, segmental allopolyploidy, and polyploid genome evolution. *Front. Genet.* 11. doi: 10.3389/fgene.2020.01014
- McCann, J., Jang, T. S., Macas, J., Schneeweiss, G. M., Matzke, N. J., Novák, P., et al. (2018). Dating the species network: Allopolyploidy and repetitive DNA evolution in American daisies (Melampodium sect. Melampodium, Asteraceae). *Syst. Biol.* 67, 1010–1024. doi: 10.1093/sysbio/syy024
- McCann, J., Macas, J., Novák, P., Stuessy, T. F., Villaseñor, J. L., and Weiss-Schneeweiss, H. (2020). Differential genome size and repetitive DNA evolution in diploid species of melampodium sect. Melampodium (Asteraceae). *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.00362
- McClintock, B. (1984). The significance of responses of the genome to challenge. *Science* 226, 792–801. doi: 10.1126/science.15739260
- Michael, T. P. (2014). Plant genome size variation: Bloating and purging DNA. *Briefings Funct. Genomics Proteomics* 13, 308–317. doi: 10.1093/bfpg/elu005
- Moreno-Aguilar, M. F., Inda, L. A., Sánchez-Rodríguez, A., Arnillas, I., and Catalán, P. (2022). Evolutionary dynamics of the repeatome explains contrasting differences in genome sizes and hybrid and polyploid origins of grass liliinae lineages. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.901733
- Mu, W., Li, K., Yang, Y., Breiman, A., Lou, S., Yang, J., et al. (2023b). Scattered differentiation of unlinked loci across the genome underlines ecological divergence of

the selfing grass *Brachypodium stacei*. *Proc. Natl. Acad. Sci.* 120, e2304848120. doi: 10.1073/pnas.2304848120

Mu, W., Li, K., Yang, Y., Breiman, A., Yang, J., Wu, Y., et al. (2023a). Subgenomic stability of progenitor genomes during repeated allotetraploid origins of the same grass *brachypodium hybridum*. *Mol. Biol. Evol.* 40, 1–20. doi: 10.1093/molbev/msad259

Negi, P., Rai, A. N., and Suprasanna, P. (2016). Moving through the stressed genome: Emerging regulatory roles for transposons in plant stress response. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.01448

Novák, P., Neumann, P., and Macas, J. (2010). Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. *BMC Bioinf.* 11, 1–12. doi: 10.1186/1471-2105-11-378

Novák, P., Neumann, P., and Macas, J. (2020). Global analysis of repetitive DNA from unassembled sequence reads using RepeatExplorer2. *Nat. Protoc.* 15, 3745–3776. doi: 10.1038/s41596-020-0400-y

Novák, P., Neumann, P., Pech, J., Steinhaisl, J., and Macas, J. (2013). RepeatExplorer: A Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. *Bioinformatics* 29, 792–793. doi: 10.1093/bioinformatics/btt054

Paradis, E., and Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35, 526–528. doi: 10.1093/bioinformatics/bty633

Parisod, C., Holderegger, R., and Brochmann, C. (2010). Evolutionary consequences of autopolyploidy. *New Phytol.* 186, 5–17. doi: 10.1111/j.1469-8137.2009.03142.x

Pellicer, J., Hidalgo, O., Dodsworth, S., and Leitch, I. J. (2018). Genome size diversity and its impact on the evolution of land plants. *Genes (Basel)* 9, 88. doi: 10.3390/genes9020088

Pohlert, T. (2023). *PMCMRplus: Calculate Pairwise Multiple Comparisons of Mean Rank Sums Extended*. doi: 10.32614/CRAN.package.PMCMRplus

R Core Team. (2021). *R: A language and environment for statistical computing* (Vienna, Austria: R Foundation for Statistical Computing). Available at: <https://www.R-project.org/>.

Sancho, R., Cantalapiedra, C. P., López-Álvarez, D., Gordon, S. P., Vogel, J. P., Catalan, P., et al. (2018). Comparative plastome genomics and phylogenomics of *Brachypodium*: flowering time signatures, introgression and recombination in recently diverged ecotypes. *New Phytol.* 218, 1631–1644. doi: 10.1111/nph.14926

Sancho, R., Inda, L. A., Díaz-Pérez, A., Des Marais, D. L., Gordon, S., Vogel, J. P., et al. (2022). Tracking the ancestry of known and ‘ghost’ homeologous subgenomes in model grass *Brachypodium* polyploids. *Plant J.* 109, 1535–1558. doi: 10.1111/tpj.15650

Scarlett, V. T., Lovell, J. T., Shao, M., Phillips, J., Shu, S., Goodstein, D. M., et al. (2023). Multiple origins, one evolutionary trajectory : gradual evolution characterizes

distinct lineages of allotetraploid *Brachypodium*. *Genetics* 223, iyac146. doi: 10.1093/genetics/iyac146

Schippmann, U. (1991). Revision der europäischen Arten der Gattung *Brachypodium* Palisot de Beauvois (Poaceae). *Boissiera* 45, 1–250. (Ed.) Conservatoire et jardin botaniques de Genève.

Scholthof, K. B. G., Irigoyen, S., Catalan, P., and Mandadi, K. K. (2018). *Brachypodium*: A monocot grass model genus for plant biology. *Plant Cell* 30, 1673–1694. doi: 10.1105/tpc.18.00083

Stritt, C., Wyler, M., Gimmi, E. L., Pippel, M., and Roulin, A. C. (2020). Diversity, dynamics and effects of long terminal repeat retrotransposons in the model grass *Brachypodium distachyon*. *New Phytol.* 227, 1736–1748. doi: 10.1111/nph.16308

Suda, J., Meyerson, L. A., Leitch, I. J., and Pyšek, P. (2015). The hidden side of plant invasions: The role of genome size. *New Phytol.* 205, 994–1007. doi: 10.1111/nph.13107

Vitales, D., Álvarez, I., Garcia, S., Hidalgo, O., Feliner, G. N., Pellicer, J., et al. (2020b). Genome size variation at constant chromosome number is not correlated with repetitive DNA dynamism in *Anacyclus* (Asteraceae). *Ann. Bot.* 125, 611–623. doi: 10.1093/aob/mcz183

Vitales, D., Garcia, S., and Dodsworth, S. (2020a). Reconstructing phylogenetic relationships based on repeat sequence similarities. *Mol. Phylogenet. Evol.* 147, 106766. doi: 10.1016/j.ympev.2020.106766

Vozárová, R., Herklotz, V., Kovařík, A., Tynkevich, Y. O., Volkov, R. A., Ritz, C. M., et al. (2021). Ancient origin of two 5S rDNA families dominating in the genus *rosa* and their behavior in the canina-type meiosis. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.643548

Vu, G. T. H., Cao, H. X., Reiss, B., and Schubert, I. (2017). Deletion-bias in DNA double-strand break repair differentially contributes to plant genome shrinkage. *New Phytol.* 214, 1712–1721. doi: 10.1111/nph.14490

Weiss-Schneeweiss, H., Leitch, A. R., Mccann, J., Jang, T. S., and Macas, J. (2015). “Employing next generation sequencing to explore the repeat landscape of the plant genome,” in *Next Generation Sequencing in Plant Systematics Regnum Vegetabile*, eds. E. Hörandl and M. Appelhans, (Königstein, Germany: Koeltz Scientific Books) pp. 155–179. doi: 10.14630/000006

Wendel, J. F., Lisch, D., Hu, G., and Mason, A. S. (2018). The long and short of doubling down: polyploidy, epigenetics, and the temporal dynamics of genome fractionation. *Curr. Opin. Genet. Dev.* 49, 1–7. doi: 10.1016/j.gde.2018.01.004

Wolny, E., and Hasterok, R. (2009). Comparative cytogenetic analysis of the genomes of the model grass *Brachypodium distachyon* and its close relatives. *Ann. Bot.* 104, 873–881. doi: 10.1093/aob/mcp179

Zhang, H., and Dawe, R. K. (2012). Total centromere size and genome size are strongly correlated in ten grass species. *Chromosom. Res.* 20, 403–412. doi: 10.1007/s10577-012-9284-1



OPEN ACCESS

EDITED BY

Gerald Matthias Schneeweiss,
University of Vienna, Austria

REVIEWED BY

Umesh Chandra Lavania,
University of Lucknow, India
Judita Zozomova-Lihova,
Institute of Botany (SAS), Slovakia

*CORRESPONDENCE

Ana Valdés-Flrido
✉ valdesf.ana@gmail.com

RECEIVED 28 May 2024

ACCEPTED 19 July 2024

PUBLISHED 09 August 2024

CITATION

Valdés-Flrido A, Valcárcel V, Maguilla E,
Díaz-Lifante Z, Andrés-Camacho C, Zeltner L,
Coca-de-la-Iglesia M, Medina NG, Arroyo J
and Escudero M (2024) The interplay
between climatic niche evolution, polyploidy
and reproductive traits explains plant
speciation in the Mediterranean Basin: a case
study in *Centaurium* (Gentianaceae).
Front. Plant Sci. 15:1439985.
doi: 10.3389/fpls.2024.1439985

COPYRIGHT

© 2024 Valdés-Flrido, Valcárcel, Maguilla,
Díaz-Lifante, Andrés-Camacho, Zeltner,
Coca-de-la-Iglesia, Medina, Arroyo and
Escudero. 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.

The interplay between climatic niche evolution, polyploidy and reproductive traits explains plant speciation in the Mediterranean Basin: a case study in *Centaurium* (Gentianaceae)

Ana Valdés-Flrido^{1*}, Virginia Valcárcel^{2,3}, Enrique Maguilla⁴,
Zoila Díaz-Lifante¹, Cristina Andrés-Camacho¹, Louis Zeltner⁵,
Marina Coca-de-la-Iglesia², Nagore G. Medina^{2,3},
Juan Arroyo¹ and Marcial Escudero¹

¹Department of Plant Biology and Ecology, Faculty of Biology, University of Seville, Seville, Spain,

²Departamento de Biología, Universidad Autónoma de Madrid, Madrid, Spain, ³Centro de
Investigación en Biodiversidad y Cambio Global (CIBC-UAM), Universidad Autónoma de Madrid,
Madrid, Spain, ⁴Department of Molecular Biology and Biochemical Engineering, Pablo de Olavide
University, Seville, Spain, ⁵Laboratoire de Botanique Evolutive, Université de Neuchâtel,
Neuchâtel, Switzerland

Speciation and diversification patterns in angiosperms are frequently shaped by niche evolution. *Centaurium* Hill is a Mediterranean genus with ca. 25 species, of which 60% are polyploids (tetra- and hexaploids), distributed mainly in the Mediterranean Basin and in areas with temperate and arid climates of Asia, Europe, North-Central Africa and North America. The evolutionary history of this genus has been studied using morphological, biogeographical and molecular approaches, but its climatic niche characterization and its relation with genome evolution (chromosome number and ploidy level) has not been addressed yet. Thus, this study aims to identify the role of the evolution of climatic niche, ploidy level, life cycle and floral traits in the diversification of *Centaurium*. Climatic niche characterization involved estimating present climate preferences using quantitative data and reconstructing ancestral niches to evaluate climatic niche shifts. The evolution of climatic niche towards selective optima determined by ploidy level (three ploidy levels) and different binary traits (polyploidy, floral size, floral display, herkogamy and life cycle) was addressed under the Ornstein-Uhlenbeck model. Chromosome number evolution was inferred using the ChromoSSE model, testing if changes are clado- or anagenetic. Chromosome number evolution and its link with cladogenesis, life cycle and floral traits was modeled on the phylogeny. The reconstruction of the ancestral niches shows that *Centaurium* originated in a mild climate and diversified to both humid and cold as well as to dry and warmer climates. Niche conservatism was estimated in the climatic niche of the ancestors, while the climatic niche of the current taxa experienced transitions from their ancestors' niche. Besides, the evolution of climatic niche towards multiple selective optima determined by the studied traits was supported, life cycle optima receiving the highest support. The reconstruction of chromosome number transitions shows that the rate of

speciation process resulting from chromosomal changes (chromosomal cladogenesis) is similar to that of non-chromosomal cladogenesis. Additionally, dependent evolution of floral size, floral display and herkogamy with chromosome number variation was supported. In conclusion, polyploidization is a crucial process in the Mediterranean region that assisted speciation and diversification into new areas with different climates, entailing niche shifts and evolution of reproductive strategies.

KEYWORDS

centauries, climatic niche, chromosome evolution, Mediterranean climate, phylogeny, polyploidy, speciation

Introduction

The crucial influence of climate in plant speciation has long been recognized (Hua and Wiens, 2013). Two of the mechanisms driving speciation are niche divergence and niche conservatism. On the one hand, niche divergence refers to the evolutionary process for which a species undergoes changes in its requirements (e.g. climatic requirements), adapting to new environmental conditions that differ from the original niche (Schluter, 2000) and leading ultimately to reproductive isolation (Schluter, 2001, 2009; Rundle and Nosil, 2005). On the other hand, under the niche conservatism hypothesis (Wiens, 2004) the species maintain similar tolerances over time. Niche conservatism has been suggested to foster speciation by supporting isolation of populations that may then undergo allopatric speciation (Wiens, 2004; Qiao et al., 2024).

Another mechanism that fosters speciation in angiosperms are chromosome rearrangements, and especially those mechanisms implying chromosome number variations (Stebbins, 1950; Grant, 1981; Soltis et al., 2009, 2015). A significant correlation between species richness and diversification rates with chromosome number variation has been inferred for the whole clade of the angiosperms (Carta and Escudero, 2023). These variations comprise the multiplication of the entire set of chromosomes (i.e., polyploidy) or changes in chromosome number resulting from fusion and fission (i.e., dysploidy) (Mayrose et al., 2010; Schubert and Lysak, 2011). The evolutionary significance of the different types of chromosome changes in terms of temporal persistence across evolution has long been debated for angiosperms. For example, there is evidence of several rounds of polyploidization across the angiosperm tree of life (Leitch and Bennett, 2004; Van de Peer et al., 2009; Wood et al., 2009; Jiao et al., 2011; Mayrose et al., 2011; Van de Peer, 2011; Escudero and Wendel, 2020), which suggests long-term persistence of polyploid lineages. Similarly, gains and losses of single chromosomes (dysploidy) are also widespread among angiosperms (Brochmann et al., 2004) and seem to persist even longer over time than polyploidy (Escudero et al., 2014). Interestingly, the majority of hyper-diverse genera in angiosperms originated after recurrent events of polyploidy (Landis et al., 2018). Whole genome duplications (WGD) provide genetic support for ecological differentiation and adaptation

(Levin, 2002) and, as such, they may boost speciation and trigger diversification increases. Different cytotypes may have different tolerances to temperature, drought, salinity, and herbivory, as well as different pollinators and, subsequently, differences in floral structure (Levin, 2004). In line with this, some studies have evaluated whether whole genome duplication events are linked to climate niche shifts with some of them supporting niche divergence between polyploids and diploids (Theodoridis et al., 2013; Thompson et al., 2014) while others do not (Godsoe et al., 2013; Castro et al., 2019). Therefore, investigating in which instances WGD is associated with niche shifts can help clarify the role of polyploidy in evolutionary processes.

The evolution of flowering plants has also been modulated by the association with pollinators. The change from outbreeding to selfing is one of the most frequent evolutionary transitions in angiosperms (Stebbins, 1950, 1957, 1974; Barrett, 2002), and is usually accompanied by a set of morphological and functional changes of flowers, jointly termed selfing syndrome (Darwin, 1876; Ornduff, 1969; Richards, 1986). These morphological changes include the reduction in the distance between anthers and stigmas within the same flower (reduction of the degree of herkogamy) (Husband and Barrett, 1992; Barrett, 1993), the reduction of floral size (Goodwillie et al., 2010), or the reduction in floral display (Sicard and Lenhard, 2011), among others, with the latter two being related to the reduction of attraction of pollinators (Goodwillie et al., 2010). Besides, the evolution of selfing is promoted by an annual or short-lived life cycle, particularly in colonizer species due to their frequent need to endure a scarcity or dearth of pollinators (Barrett, 2010; Pannell, 2015).

Here, we focus on the genus *Centaureum* Hill (Gentianaceae), which comprises ca. 25 species (Mansion and Struwe, 2004; Diaz-Lifante, 2012; POWO, 2023). This genus is distributed around temperate and arid climates of Asia, Europe, North-Central Africa, and North America (Figure 1), although its center of diversification is the Mediterranean Basin (Mansion and Struwe, 2004; Maguilla et al., 2021). The taxonomy and morphology (particularly floral traits) of this genus has been widely studied (Zeltner, 1970; Mansion and Struwe, 2004; Diaz-Lifante, 2012; Jiménez-Lobato et al., 2019), and also some phylogenetic studies have been performed to date (Mansion and Struwe, 2004; Mansion et al., 2005; Jiménez-Lobato et al., 2019;

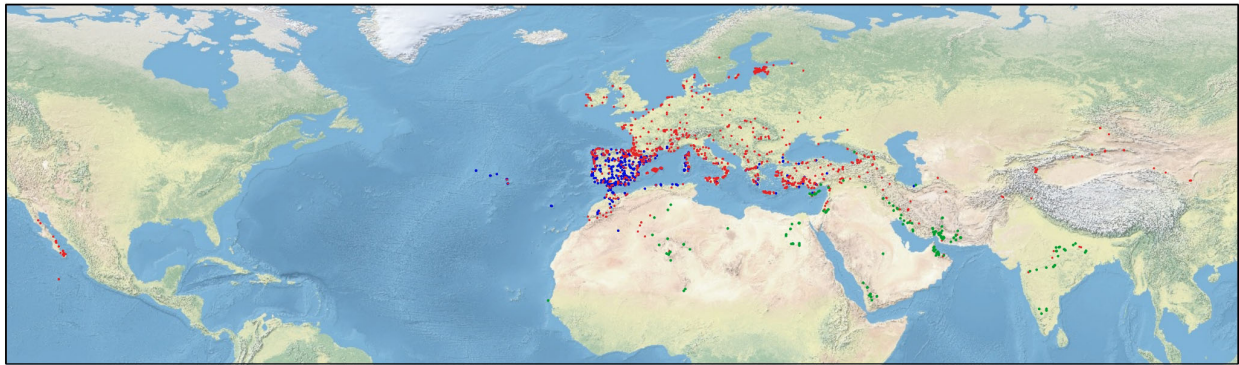


FIGURE 1
Distribution map of *Centaurium*. Blue circles represent occurrences of diploid taxa, red circles represent tetraploid taxa and green ones, hexaploid taxa.

Maguilla et al., 2021; Valdés-Flórido et al., 2024). These previous studies confirmed generalized patterns of hybridization within the genus (Mansion and Struwe, 2004; Mansion et al., 2005; Guggisberg et al., 2006; Valdés-Flórido et al., 2024), leading to complex morphological variability and thus, different taxonomic treatments (Mansion and Struwe, 2004; Mansion et al., 2005; Díaz-Lifante, 2012). Additionally, cytogenetic studies show that *Centaurium* is a polyploid complex in which different ploidy levels are found: diploid ($2x$), tetraploid ($4x$) and hexaploid ($6x$) (Zeltner, 1970). Recent studies have been accomplished to study the evolution of chromosome numbers in *Centaurium* (Maguilla et al., 2021; Escudero et al., 2023) using ChromEvol (Glick and Mayrose, 2014). The haploid chromosome number $x = 10$ was estimated as the most probable base chromosome number for the genus, although a secondary base chromosome number of $x = 9$ was also inferred (Escudero et al., 2023). Maguilla et al. (2021) and Escudero et al. (2023) showed that the main chromosome number transitions in *Centaurium* have been both descending dysploidy and polyploidy, with polyploidization estimated to mainly occur towards the tips of the phylogeny. Allopolyploidy (polyploidy involving hybridization) involved lineages of the two different base numbers (9 and 10), as previously reported by Zeltner (1970). Interestingly, ploidy levels follow a geographical pattern in *Centaurium*, where diploid species are distributed mainly in the Mediterranean Basin, tetraploids are found generally at higher latitudes in northern Europe and eastern Asia, and the hexaploids can be found at southern latitudes in India, the southwest of the Mediterranean Basin and the Arabian Peninsula (Mansion et al., 2005). In line with this, Maguilla et al. (2021) showed a strong relation of polyploidy with geographical movements. These results confirm that polyploidy occurred in species and subspecies that are distributed outside the area of origin (i.e., the Mediterranean Basin), suggesting that polyploidy may have facilitated range expansions by increasing establishment success in the newly colonized areas. Thus, polyploidization events do not seem to facilitate dispersal events *per se*, but rather the establishment in the new areas or under different niches (Maguilla et al., 2021).

The evolution of reproductive traits and its association with polyploidy has also been studied in *Centaurium* (Jiménez-Lobato et al., 2019). Specifically, traits associated with selfing syndrome

were studied: herkogamy, floral size, floral display, and polyploidy. No pattern of correlated evolution was found in the genus: *Centaurium* evolved its selfing ability early in its history, while polyploidy and flower traits evolved subsequently, probably independently (Jiménez-Lobato et al., 2019).

Despite the great climatic variations of the area of distribution of *Centaurium* (from temperate and humid areas to warmer and dry ones), the relation of the large variability in morphological traits, life cycle (annual/biennial vs. perennial) and ploidy levels with the climatic niche has not been studied yet. In this context, we aim to (i) estimate the species' climatic niche as inferred from occurrences and reconstruct its evolution, (ii) model chromosome evolution and cladogenesis, (iii) test the possible link between niche evolution and selective optima related to polyploidy and reproductive traits, and (iv) test the dependent or independent evolution of chromosome number with floral traits and life cycle. To do so, we used a comprehensive geographical and climate database, which included chromosome counts, data of morphological traits and life cycle of the species, as well as a phylogenetic reconstruction based on RADseq markers (Valdés-Flórido et al., 2024). We hypothesize that changes in ploidy level are related to the climatic niche in *Centaurium*. Thus, we expect to find polyploids to have niches with more extreme climates, and diploids with niches of milder conditions. The null hypothesis expected from the models of climatic niche evolution regarding morphological traits and life cycle is that the evolution occurs toward a single optimum. Besides, we hypothesize that chromosome evolution played a role in cladogenetic events in *Centaurium* and expect to find independent evolution of floral traits and chromosome number changes, as suggested by previous studies (Jiménez-Lobato et al., 2019).

Materials and methods

Occurrence records

This study included 28 *Centaurium* taxa (see Supplementary Table S1), following the taxonomic treatment proposed by

Mansion et al. (2005) and then reviewed by Díaz-Lifante (2012). The occurrence records were obtained from different herbaria: University of Seville (SEV), University of Oviedo (FCO), Royal Botanic Gardens of Madrid (MA), University of Santiago de Compostela (SANT), University of Valencia (VAL), University of Málaga (MGC) and University of Neuchâtel (NEU), the flora of the Iberian Peninsula project (*Flora Iberica*, Díaz-Lifante, 2012), and from our field sampling. To fill potential gaps in taxa distribution, GBIF data was also downloaded (<https://www.gbif.org/es/>). The original database contained 3,413 records.

To ensure the taxonomic and geographic quality of the database, a filtering process similar to that described by Coca-de-la-Iglesia et al. (2023) was conducted. This procedure includes a first step of data cleaning: removing records that lack coordinates, the ones that have invalid coordinates (i.e., latitude or longitude = 0), those in which coordinates have low precision (i.e., the coordinates that had fewer than 2 decimals), as well as duplicates. Given that the original databases included records from GBIF, which may increase taxonomic uncertainty due to species being incorrectly identified, the Coca-de-la-Iglesia et al.'s procedure includes filtering steps that consist of cross-referencing records with political countries describing the natural distribution of taxa according to Díaz-Lifante (2012). Subsequently, maps were built and visually inspected to double-check the natural distribution of each taxon and evaluate spatial gaps per taxon (Supplementary Figure S1). Country codes were assigned using the Level-3 code of the World Geographical Scheme for Recording Plant Distributions (WGSRPD) from the Biodiversity Information Standards (TDWG) (<http://www.tdwg.org/standards/109>). Records falling outside the natural distribution of the species were removed. The third and final cleaning step included recalculating the off-land coordinates so they fell in the nearest terrestrial climatic cell of the template (Coca-de-la-Iglesia et al., 2023). The records located more than 5 km from the coastal limit were removed. As a result, we compiled a clean database with 1,690 records.

Climatic niche characterization

We extracted 19 bioclimatic variables from WorldClim (Hijmans et al., 2005) with a resolution of 30 seconds. To avoid collinearity effects, pairwise Pearson correlation was calculated among the bioclimatic variables. The correlation values were transformed into a dendrogram to visualize the results (Supplementary Figure S2). We then removed correlated variables and selected those below 0.5 using R (R Core Team, 2023) that had sound biological interpretation for our species. The subsequent climatic analyses were performed using four variables representing temperature (bio1 -annual mean temperature-, bio2 -mean diurnal range-, bio7 -temperature annual range-, bio8 -mean temperature of wettest quarter-) and three variables representing precipitation (bio12 -annual precipitation-, bio15 -precipitation seasonality-, and bio17 -precipitation of driest quarter-). A Principal Component Analysis (PCA) was performed using the function *dudi.pca* from the R package *ade4* (Dray and Dufour, 2007), with the least correlated climatic variables for the studied *Centaureum* taxa.

Climatic niche reconstruction analyses

Climatic niche reconstruction analyses were performed using the phylogeny of the genus of Valdés-Florido et al. (2024) as the input tree. The original phylogeny was reconstructed using a maximum likelihood approach based on a SNP matrix consisting of 8,497 SNPs out of 215,894 total recovered loci derived from RADseq markers, comprising the same 28 taxa (Valdés-Florido et al., 2024). We pruned the original phylogeny to keep only one tip per taxon for the climatic niche reconstruction using the R package *ape* (Paradis and Schliep, 2019). Most taxa (species and subspecies) were monophyletic, but two taxa were polyphyletic (*C. littorale* ssp. *uliginosum* and *C. somedanum*) and one was paraphyletic (*C. scilloides*). As there was no spatial or morphological pattern to deal with polyphyletic taxa, we randomized the removal of tips for these three non-monophyletic taxa.

As climate input data we used the first axis of the PCA (PC1), as it was the principal component that better explains the data variability. To include the current climatic heterogeneity of each taxon in the ancestral reconstruction, we used the probabilistic approach developed in Coca-de-la-Iglesia et al. (2024). This probabilistic approach consists of generating 1,000 climatic matrices from the original species PC1 values as proportional to the climatic density function of each species. Then, this procedure uses the *fastAnc* function from *phytools* 1.5-1 R package (Revell, 2012) with 1,000 iterations, performing a fast estimation of the ancestral states following a Maximum Likelihood approach, under a Brownian motion (BM) model. Finally, for each of the ancestral nodes this procedure summarizes the climatic density function.

Evolution of climatic niche regarding polyploidy, life cycle and morphological traits

We used the Ornstein-Uhlenbeck (OU) model (Hansen, 1997; Blomberg et al., 2020) to test the evolution of climatic niche towards selective optima determined by ploidy levels (three optima: 2x, 4x and 6x) and different binary traits (two optima): polyploidy (diploids vs. polyploids), floral size (small to medium vs. large), floral display (low vs. high), herkogamy (low herkogamy vs. high herkogamy) and life cycle (annual/biennial vs. perennial). To conduct these analyses, we used the phylogenetic tree previously employed for climatic niche reconstruction. The binary trait characterization used (polyploidy, floral size, floral display, herkogamy and life cycle data) was that from Jiménez-Lobato et al. (2019) (Supplementary Table S1). The continuous variables were the climatic ones used in the climatic niche reconstruction. The null hypothesis is that in the evolutionary history of the genus there is only a single selective optimum for the studied traits. Conversely, the alternative hypothesis suggests the presence of multiple optima. To test the evolution of climatic optima determined by the three ploidy levels (2x, 4x and 6x), we followed two steps. First, we designed a transition matrix allowing only unidirectional shifts from 2x to 4x and from 4x to 6x, to estimate the

model of ploidy evolution and map the ancestral states in the phylogeny. The model was fitted using the *fitDiscrete* function from the R-package *geiger* (Harmon et al., 2008; Pennell and Harmon, 2013) and the stochastic mapping was performed with the *make.simmap* function from the R-package *phytools* (Revell, 2012) under a “all rates different” (ARD) model for reproductive traits and life cycle. Second, we estimated the climatic optima based on the three ploidy levels using the Ornstein-Uhlenbeck model. A similar approach was used for the binary trait diploid vs. polyploid. The evolution of the climatic niche toward selective optima for the morphological traits, life cycle and three-state ploidy levels was tested using the *OUI* function from the R-package *OUI* (Beaulieu et al., 2012; Beaulieu and O’Meara, 2020), under the “OUM” model. Besides, the null hypothesis was tested under the “OU1” model of the same R function. We used the Akaike information criterion (AIC) with the R package *diversitree* (FitzJohn, 2012) to select the best-fitting model.

Clado- and anagenetic chromosome evolution in *Centaurium*

The ChromoSSE model (Freyman and Höhna, 2018), implemented in the RevBayes platform (Höhna et al., 2016), was used to infer the chromosome number evolution along the phylogeny of the genus and to test if these changes were either clado- or anagenetic. The used input tree was the pruned tree used for the climatic reconstruction, and the chromosome counts were recovered from Maguilla et al. (2021). The ChromoSSE model consisted of cladogenetic parameters (no change, fission, fusion and polyploid events), a global rate of extinction and anagenetic parameters (fission, fusion and polyploidization events). Here, we set the demi-polyploidy parameter to 0, as no demi-polyploidization event was inferred in previous studies (Maguilla et al., 2021; Escudero et al., 2023). We then visualized the results using the RevGadgets R package (Tribble et al., 2022).

Dependent vs. independent model of trait evolution and chromosome number

The dependent evolution between chromosome number evolution and life cycle (annual/biennial vs. perennial) and reproductive traits (floral size, floral display and herkogamy), respectively, was studied with the R package *ChromePlus* (Blackmon et al., 2019). This approach infers a different model of chromosome evolution for each state of the binary trait under study. The used phylogeny was the one used for the climatic niche reconstruction, the chromosome counts were recovered from Maguilla et al. (2021), and the floral traits and life cycle characterization from Jiménez-Lobato et al. (2019). We used the R package *diversitree* (FitzJohn, 2012) to compare the AIC of a model of dependent evolution of each of the binary traits and chromosome evolution against a model of independent evolution. The estimated parameters included transitions (tran01 and tran10) and chromosome parameters: chromosome gain (gain0

and gain1), chromosome loss (loss0 and loss1) and polyploidy (polyploidy0 and polyploidy1).

Results

The PCA analysis with the selected bioclimatic variables captured 73% of the total variance in the first two Principal Components (PC1: 52% and PC2: 21%; Figure 2). Positive values on PC1 represent areas characterized by cold and humid conditions with abundant precipitation throughout the year, lacking any temperature or precipitation seasonality. However, negative values on PC1 represent warmer and drier areas that can either have precipitation seasonality (when coupled with positive values in the PC2) or high temperature contrast (annual and diurnal), when coupled with negative values in the PC2.

Climatic niche diversification in *Centaurium*

The ancestors of *Centaurium* displayed preferences for mild dry and warm climates (Figure 3). These ancestral climatic niches are characterized by milder conditions (i.e., values between -1 and 1 in the PC1), whereas the current climatic niche of the taxa encompasses more extreme conditions (i.e., values up to -5 and 4.5 in the PC1). The two ancestors with the most extreme niches are that of node 51 that display a warmer niche (with values around -2.5 in the PC1), and that of node 46 that displays the coldest niche of ancestors (with values around 2.5 in the PC1).

The most recent common ancestor of *Centaurium* originated in warm areas (node 29, Figure 3). The subsequent nodes of the phylogeny follow different trends. On the one hand, the clade from node 30 onwards (Figure 3) evolved towards both mild and colder climates, except for the current niche of *C. malzacianum*. On the other hand, the clade from node 49 onwards evolved towards warmer and drier climates. Climatic niche evolution described a phylogenetic pattern, with shifts to current cold niches occurring along descendants of the cold early ancestor (node 30), whereas shifts to current warm niches occurred in the clade of the warm early ancestor (node 49). The cold early ancestor (node 30, Figure 3) evolved towards a clade with 19 species with cold and humid niches (node 31, Figure 3), as well as one species, *C. malzacianum*, that evolved towards a warm and dry climate. From the clade of node 31 (Figure 3), the genus splits into two clades (nodes 32 and 41, Figure 3). From the clade of node 32 (Figure 3), *Centaurium* mostly evolved towards a niche with warm and dry conditions, as it only evolved towards a cold and humid climate niche for *C. erythraea* subsp. *erythraea* and *C. erythraea* var. *subcapitatum*. In contrast, from the clade of node 41 (Figure 3), it evolved towards both cold and humid as well as warmer and drier niches. Lineages within the clade of node 42 (Figure 3) evolved towards a humid and cold climate niche: *C. scilloides*, *C. portense*, *C. somedanum*, *C. chloodes*, as well as the two subspecies of *C. littorale* (*C. littorale* subsp. *littorale* and *C. littorale* subsp. *uliginosum*) are contained in this clade. The evolution towards a warm and dry climatic niche mostly

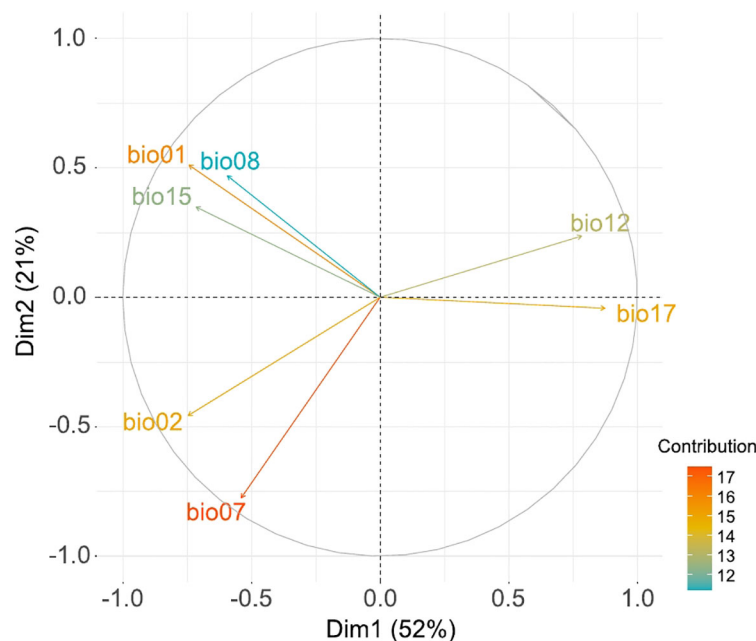


FIGURE 2

Principal Component Analysis performed with the selected bioclimatic variables: bio01 (annual mean temperature), bio02 (mean diurnal range), bio07 (temperature annual range), bio08 (mean temperature of wettest quarter), bio12 (annual precipitation), bio15 (precipitation seasonality) and bio17 (precipitation of driest quarter). Arrows represent the contribution degree of each bioclimatic variable.

occurred for lineages within the clade of node 49 (Figure 3), as well as for *C. malzacianum*, as mentioned before.

Evolution of climatic niche regarding polyploidy, life cycle and morphological traits

The evolution of climatic niche towards a single selective optimum (i.e., the null hypothesis) was not supported (Table 1). The best supported

model was that with multiple selective optima for life cycle (Table 1), with the climatic niche optimum of annual/biennial being related to warmer and drier climates and the climatic optimum of perennial related to colder and more humid climates. Nevertheless, ploidy level (with two and three states) and reproductive traits (floral size, floral display and herkogamy) were also significantly supported when compared to the null hypothesis of a single optimum. The climatic niche optimum for diploids was associated with mild climates, whereas tetra- and hexaploids were related to harsh climates, with tetraploids inhabiting colder and more humid climates than hexaploids. *Centaureum* species with small flowers and high

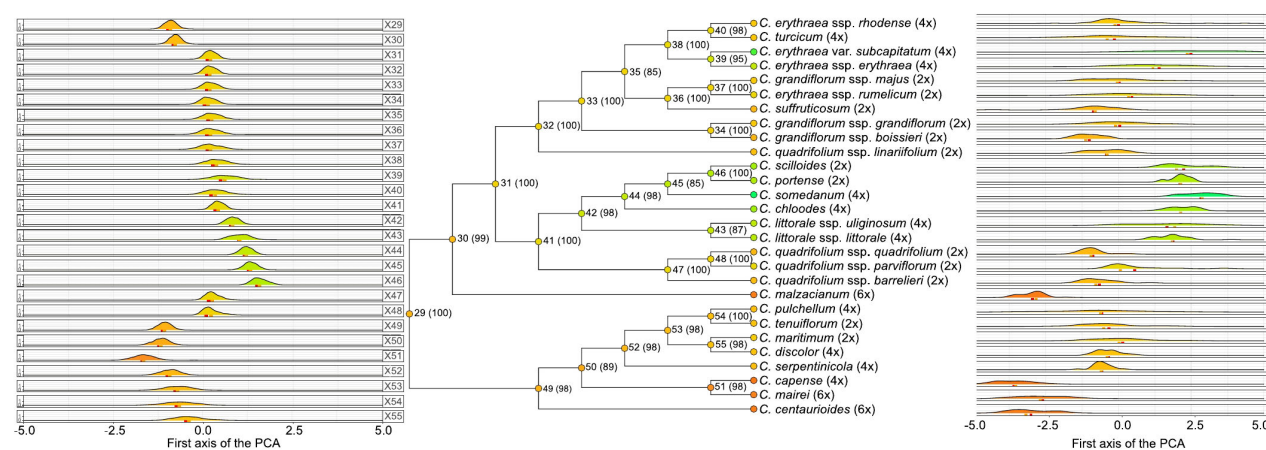


FIGURE 3

Climatic characterization of *Centaureum*. Left-side density plots represent climatic characterization of the nodes while the right-side density plots represent actual climatic characterization of the studied species. Colors in the nodes and in the tips correspond to the mean value of the PC1. Numbers at the nodes indicate node numbers, with bootstrap support in parentheses. The ploidy level of each taxon is shown at the tips.

floral display had climatic niche optima in warmer and drier conditions, while species with larger flowers and low floral display were associated with colder and more humid conditions. The climatic niche optimum for high herkogamy was associated with milder conditions, while the optimum for low herkogamy was linked to more extreme conditions, including higher humidity and lower temperatures (Table 1). The ancestral state reconstruction for each studied trait under stochastic mapping is shown in Supplementary Figure S3.

Clado- and anagenetic chromosome number evolution in *Centaurium*

Chromosome evolution analyses with ChromoSSE model showed transitions in chromosome number along the *Centaurium* phylogeny (Figure 4). The most likely ancestral chromosome number for the genus was $x = 7$. The posterior distribution of the cladogenetic parameters show no_change to be the most important event, followed by polyploidy cladogenetic events (Figure 5A). The posterior distribution of the anagenetic parameters show polyploidy to be the most important event, followed by descending dysploidy events (Figure 5B). Consequently, cladogenetic changes are mostly polyploidization events, while the anagenetic ones are dysploidy and polyploidization ones. Polyploid speciation was inferred in both ancestral and recent nodes that lead to terminal branches, while dysploidy events mostly occur in the terminal branches. Polyploid speciation (from $n = 10$ to $n = 20$) was inferred in the cladogenesis of *C. erythraea* subsp. *rhodense*, *C. turcicum*, *C. erythraea* var. *subcapitatum* and *C. erythraea* subsp. *erythraea*, in *C. somendanum*, *C. chloodes*, the subclade of *C. littorale* subsp. *uliginosum* and *C. littorale* subsp. *littorale*, *C. tenuiflorum*, *C. discolor* and *C. serpentinicola*. Besides, anagenetic chromosome transitions with dysploidy (mostly descending) were inferred for *C. malzacianum*, *C. pulchellum*, *C. capense*, *C. mairei* and *C. centaurioides*.

Dependent vs. independent model of trait evolution and chromosome number

The results of the dependent and independent models of each of the floral traits and life cycle traits and chromosome evolution are

shown in Table 2 and in Supplementary Table S2. The dependent evolution of floral size, floral display and herkogamy and chromosome number evolution is better supported than the model of independent evolution of these traits. In the models studying herkogamy and floral display, the polyploidization rates in the state “2” (high herkogamy and high floral display) were significantly higher than those in state “1” (low herkogamy and low floral display). This pattern also occurred in the model studying floral size, although the difference between rates of polyploidization was not that significant. On the contrary, the model of independent evolution of life cycle and chromosome number variation was better supported, according to the AIC.

Discussion

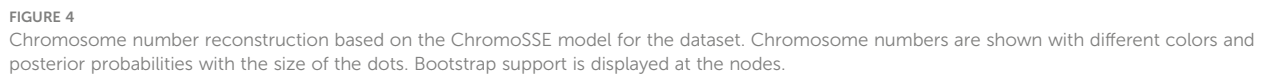
Climatic niche evolution: transitions from warm and mild climates of diploids and tetraploids to harsh climates of hexaploids

Our results show the first reconstruction of the climatic niche of *Centaurium*. The most recent common ancestor of *Centaurium* originated in a mild warm and dry climate (Figure 3). According to Jiménez-Lobato et al. (2019) and Maguilla et al. (2021), *Centaurium* originated in the Miocene (ca. 10 Mya) in the Mediterranean Basin. Our climatic reconstruction shows that *Centaurium* evolved from an ancestral warm and dry climate niche, which can be associated with the Mediterranean Basin (Jiménez-Lobato et al., 2019; Maguilla et al., 2021), to both colder and warmer climate niches (Figure 3). Besides, the colder and humid climate niches showed a remarkable variation, as some *Centaurium* species thrive in coastal areas (i.e., the tetraploid *C. littorale* subsp. *littorale*, the tetraploid *C. chloodes* or the diploid *C. scilloides*), and others in inland regions (i.e., the tetraploids *C. portense* and *C. somedanum*; see Supplementary Figure S1) (Díaz-Lifante, 2012). These two regions hold remarkable climatic differences: coastal areas display relatively stable and moderate temperature fluctuations throughout the year, as well as a constant high level of humidity. In contrast, inland regions often experience more significant temperature fluctuations. However, despite the large climatic differences between these areas, *Centaurium* species tend to occupy the same niches.

TABLE 1 Results of the analyses of climatic niche evolution towards selective optima, regarding polyploidy, floral traits, and life cycle.

	<i>H0</i>	<i>ploidy</i>	<i>ploidybin</i>	<i>FS</i>	<i>FD</i>	<i>Hk</i>	<i>LC</i>
<i>AIC</i>	396.67	385.97	383.78	382.30	383.18	382.44	380.52
<i>theta</i>	28.11	diploid = 8.69E+01 tetraploid = -3.21E+09 hexaploid = -2.52E+09	diploid = 5.21E+09 polyploid = -9.89E+01	small/medium = -4.48E+02 large = 1.22E+10	low = 8.93E+01 high = -1.01E+10	low = 7.97E+09 high = -1.79E+02	annual/biannual = -5.45E+01 perennial = 2.20E+09
<i>sigma2</i>		1.19E+05	1.21E+05	1.12E+05	1.15E+05	1.12E+05	1.05E+05
<i>alpha</i>		1.58E-07	7.84E-08	6.60E-08	8.25E-08	1.73E-07	4.46E-07

AIC corresponds to the Akaike’s information criterion, *theta* to the optima of each model, corresponding to values of the PC1, *sigma squared* corresponds to the variance of the models, and *alpha* shows the tendency to return towards the optimum. *H0* is the null hypothesis, *ploidy* corresponds to the model tested with the three ploidy levels in the genus, *ploidybin* represents the model testing ploidy as binary traits (diploid vs. polyploid), *FS* represents floral size, *FD* is floral display, *Hk* shows herkogamy and *LC*, life cycle. The best fitting model for each studied trait is in bold font, considering the Akaike’s information criterion (AIC).



warmer or colder ones. This maintenance of the ancestors' mild climate invokes niche conservatism, at least in the ancestor's climatic niche. Niche conservatism is the tendency of a species to retain its climatic requirements across time (Peterson et al., 1999), and has been proposed to facilitate allopatric speciation (Wiens, 2004). Considering the sympatric origin of the species due to hybridization events (Zeltner,

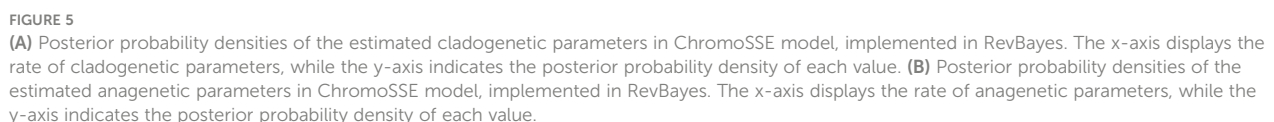


TABLE 2 Models of dependent and independent chromosome number changes correlated with floral size (FS), floral display (FD), herkogamy (Hk) and life cycle (LC).

Model	AIC
dependent FS	1.40E+02
independent FS	1.45E+02
dependent FD	1.30E+02
independent FD	1.38E+02
dependent Hk	1.37E+02
independent Hk	1.42E+02
dependent LC	1.30E+02
independent LC	1.24E+02

The best-fitted model is in bold font, considering the Akaike’s information criterion (AIC).

1970; Mansion et al., 2005; Jiménez-Lobato et al., 2019; Maguilla et al., 2021; Valdés-Flórido et al., 2024) with no geographical barrier to get isolated, and the climatic niche conservatism reported here, *Centaureum* could represent an example of climatic niche conservatism playing an important role for sympatric speciation. However, we should consider that our climate data are geographically very coarse, whereas the differentiation necessary for speciation may happen at a much finer resolution. The importance of niche conservatism in sympatric speciation was suggested by Wiens (2004). Besides, climatic niche breadth changes along the evolutionary history of the genus. In the reconstructed ancestral climatic niche, climatic niches from the species of node 30 onwards are broader than those of the node 49 onwards (Figure 3). Climatic niches reconstructed in the nodes are narrower than the niches reconstructed at the tips (Figure 3). The species that have retained the ancestral niche have mildly expanded it (e.g., the tetraploid *C. erythraea* subsp. *rhodense* and the diploid *C. quadrifolium* subsp. *quadrifolium*), whereas others have undergone a complete niche shift (e.g., the hexaploids *C. malzacianum* and *C. mairei*). Notably, only the two subspecies of *C. erythraea* have marginally broadened their niche, yet persisting within the intermediate zones as part of their niche. Conversely, the hexaploids (i.e., *C. malzacianum*, *C. mairei*, *C. centaureoides*) and one tetraploid (*C. capense*) have undergone a radical shift to niches characterized by warmer, drier, and more disparate conditions, distinct from those of any other species. Additionally, the niche of taxa as *C. scilloides* has evolved towards niches characterized by wetter, colder, and less disparate environments. As mentioned, some *Centaureum* species have originated by hybridization. For example, the hexaploids and *C. capense* have been suggested to be allopolyploids (Valdés-Flórido et al., 2024), so that there has not been any geological barrier involved in the speciation process. Thus, the pattern of niche broadening could be beneficial to reduce niche competitiveness, and then enhance the likelihood of the lineage to survive (Carscadden et al., 2020).

Interplay between polyploid evolution, morphological traits, and climatic niche

Polyploidy, both allo- and autopolyploidy, is a key event in the evolution of *Centaureum* (Zeltner, 1970; Mansion et al., 2005; Jiménez-

Lobato et al., 2019; Maguilla et al., 2021; Valdés-Flórido et al., 2024). In fact, the climatic niche of *Centaureum* evolves towards selective optima dictated by ploidy levels. In this genus, polyploidization has been proposed to have occurred due to hybridization events (allopolyploidy) or as result of autopolyploidy (Zeltner, 1970; Mansion et al., 2005; Jiménez-Lobato et al., 2019; Maguilla et al., 2021; Valdés-Flórido et al., 2024). Thus, polyploidy in *Centaureum* is significantly linked to hybridization as well as to climatic niche evolution, the latter evolving towards ploidy level optima (Table 1). Climatic niches associated with mild conditions are related to diploids, while harsh ones to tetra- and hexaploids (Table 1), with tetraploids inhabiting colder and more humid areas than hexaploids. Regarding life cycle, our results supported that climatic niche evolves towards life cycle optima. The climatic niche optimum of annual/biennial is related to warmer and drier climates and the climatic optimum of perennial is related to colder and more humid climates (Table 1). The related evolution of life cycle considering climate has also been confirmed in *Centaureum* (Jiménez-Lobato et al., 2019). Four transitions from perennial life cycle to annual/biennial were inferred: one between 10.25 Mya and 7.29 Mya, two in the Pleistocene (one around 0.7 Mya and the other around 1.72 Mya), and the last transition with uncertain timing (Jiménez-Lobato et al., 2019). Some of these transitions to annual/biennial life forms are associated with climatic changes in the Mediterranean, as the arid period occurred around the Messinian Salinity Crisis (MSC) around 5.96 - 5.33 Mya or the seasonal drought of the Mediterranean climate (3.4 - 2.8 Mya).

In flowering plants, selection on floral traits and plant mating strategies leading to a system in which seed production is ensured (i.e., reproductive assurance), is key to guarantee the maintenance of the lineage (Lloyd and Schoen, 1992). The climatic niche evolution towards different herkogamy optima was also supported, as well as the models of the other studied traits (i.e., floral display and floral size). The climatic niche optimum for high herkogamy was associated with milder conditions and the climatic optimum of low herkogamy was related to colder and more humid conditions (Table 1). Species with small/medium flowers and with high floral display are related to warmer and drier conditions, and those with larger flowers and low floral display, to colder and humid ones (Table 1). These floral traits are related to selfing syndrome, modifying the attraction to pollinators, or the spatial distance between the anthers and stigma, which facilitate selfing (Lloyd and Schoen, 1992; Goodwillie et al., 2010). For example, in *C. erythraea* it has been reported that changes in floral size and floral display as well as in the levels of herkogamy are related to the quantity of pollinators in the environment. In pollinator-poor environments there are fewer and smaller flowers of this species with lower levels of herkogamy compared to pollinator-rich environments, conferring reproductive assurance via selfing (Brys and Jacquemyn, 2011). Thus, the evolution of floral traits is dependent on climatic niche as well as on the pollinator’s richness, which ultimately depends on climate (Rech et al., 2016).

Chromosome evolution, reproductive traits, and life cycle

The results of the ChromoSSE analysis suggest a most probable base chromosome number of $x = 7$ (Figure 4). However, this result

is questionable, as the ChromoSSE model does not include the possibility of base-number reconstruction parameter in the analyses, contrary to other approaches such as ChromEvol (Glick and Mayrose, 2014). Despite $x = 7$ being the most probable base chromosome number, chromosome numbers changed throughout the evolutionary history of the genus, ultimately stabilizing at $x = 10$ (Figure 4). This base chromosome number of $x = 10$ was also suggested by previous studies (Zeltner, 1970; Maguilla et al., 2021; Escudero et al., 2023). Polyploidy has been identified as the most important anagenetic event, as well as the second most important cladogenetic event, with that of no_change being the most important cladogenetic event. This estimation is congruent with all the previous studies of the genus that suggest the crucial role of polyploidy in *Centaureum* diversification (Zeltner, 1970; Mansion and Struwe, 2004; Mansion et al., 2005; Guggisberg et al., 2006; Escudero et al., 2023; Valdés-Florido et al., 2024). Whereas most of the transition events from diploid to tetraploid are linked to transitions from drier and warmer to colder and wetter climatic niches and from southern to northern ranges (considering the distribution of the genus, Figure 1), transitions leading to the hexaploids *C. malzacianum*, *C. centaurioides* and *C. mairei*, and the tetraploid *C. capense*, coincide with the transition from temperate to warmer and drier climatic niche in the southern limit of the genus range (Maguilla et al., 2021).

The global distribution pattern of polyploids reveals a latitudinal trend, so that polyploid frequency increases with increasing latitude, i.e., areas of cold climates (Stebbins, 1971; Rice et al., 2019). Several studies (Levin, 2019; Lavania, 2020; Vimala et al., 2021) support autopolyploids to have originated due to temperature stress, so that they are distributed in colder areas. However, our results suggest a different pattern with the allo-hexaploids and the allo-tetraploid *C. capense* (Valdés-Florido et al., 2024), distributed in warmer and southern areas (Supplementary Figure S1). This trend of polyploids inhabiting warmer areas is also observed in autopolyploids as *Solidago canadensis* L. (Cheng et al., 2021). In their experimental study, Cheng et al. (2021) demonstrated that angiosperms can also expand their distribution towards warmer regions through heat tolerance evolution given by polyploidization. Our results suggest that this pattern could occur not only in autopolyploids, but also in allopolyploid species.

Chromosome number variations (i.e., polyploidy) have been suggested to promote morphological changes (Stebbins, 1950). This association is confirmed in *Centaureum*, as the dependent evolution of chromosome number and floral size, floral display and herkogamy has been shown by our analyses (Table 2; Supplementary Table S2). In fact, some studies in *Centaureum* revealed morphological differences between diploids and tetra- and hexaploids (Valdés-Florido et al., 2024). However, Jiménez-Lobato et al. (2019) tested the dependent evolution of these traits and polyploidy using Pagel's model, finding independent evolution for those traits. This is incongruent with our results of the dependent evolution of morphological traits and chromosome number evolution using a more complex model. In this model, the morphological traits are modeled by a Markov model for binary traits whereas chromosome evolution is modeled using a specific

model for chromosome evolution (Table 2; Supplementary Table S2). Additionally, the higher rates of polyploidization estimated in *Centaureum* taxa with high herkogamy and high floral display (Table 2; Supplementary Table S2) are explained by polyploidization events in terminal short branches of the phylogeny (this model of chromosome evolution did not include chromosome evolution at cladogenesis and most of the polyploid events are inferred in the terminal branches).

Perenniality has traditionally been related to polyploidy (Stebbins, 1950; Thompson and Lumaret, 1992; Otto and Whitton, 2000; Rice et al., 2019). Polyploids typically exhibit slower growth compared to diploids due to the extended duration required for DNA replication and cell division resulting from higher chromosome numbers (Stebbins, 1950; Levin, 2002). Consequently, they are expected to be perennial rather than annual. However, our analyses suggest that life cycle has not been related to the evolution of chromosome number in *Centaureum*.

Conclusions

Our study on the climatic niche evolution and chromosome number dynamics within the genus *Centaureum* provides valuable insights into the interplay between climatic adaptation and genomic and morphological changes. We confirm that the genus originated in a climate with mild conditions, based on the reconstruction of the climatic niche of the ancestors, which could be the Mediterranean Basin (Jiménez-Lobato et al., 2019; Maguilla et al., 2021). We also show the association between climatic niche and polyploidization events. Additionally, the presence of polyploids also in warmer regions indicates that some *Centaureum* species (tetra- and hexaploid species) do not always follow traditional latitudinal trends (Stebbins, 1971; Rice et al., 2019). The dependent evolution between floral traits and chromosome number evolution has been confirmed, as well as the independent evolution between life cycle and chromosome number evolution. This study offers valuable insights into the strategies that angiosperms (here Mediterranean lineages) employ to grow in diverse life cycles and climates as well as the importance of considering both climatic and genomic factors to understand species diversification.

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 author/s.

Author contributions

AV-F: Writing – original draft, Writing – review & editing. VV: Writing – review & editing. EM: Writing – review & editing. ZD-L: Writing – review & editing. CA-C: Writing – review & editing. LZ: Writing – review & editing. MC-I: Writing – review

& editing. NM: Writing – review & editing. JA: Writing – review & editing. ME: Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The present work was funded by the Spanish Government and FEDER funds (European Commission) through granted projects PID2021-122715NB-I00 to ME, JA and AV-F, and PGC2018-099608-B-I00 to ME, JA, EM, AV-F, ZD-L and CA-C and through a FPI fellowship from MICINN to AV-F (PRE2019-087452).

Acknowledgments

We thank two anonymous reviewers and the associate editor for suggestions that improved earlier versions of the manuscript and all the herbaria that provided plant material: University of Seville (SEV), Royal Botanic Gardens of Madrid (MA), University of Santiago de Compostela (SANT), University of Valencia (VAL), University of Málaga (MGC), University of Neuchâtel (NEU) and University of Oviedo (FCO).

References

- Barrett, S. C. H. (1993). "The evolution and function of heterostyly," in *Oxford Surveys in Evolutionary Biology*. (Oxford, UK: Oxford University Press) vol. 9, 283–326.
- Barrett, S. C. H. (2002). Evolution of sex: the evolution of plant sexual diversity. *Nat. Rev. Genet.* 3, 274–284. doi: 10.1038/nrg776
- Barrett, S. C. H. (2010). Understanding plant reproductive diversity. *Philos. Trans. R. Soc. B: Biol. Sci.* 365, 99–109. doi: 10.1098/rstb.2009.0199
- Beaulieu, J. M., Jhwueng, D. C., Boettiger, C., and O'Meara, B. C. (2012). Modeling stabilizing selection: Expanding the Ornstein-Uhlenbeck model of adaptive evolution. *Evol.* 66, 2369–2383. doi: 10.1111/evo.2012.66.issue-8
- Beaulieu, J. M., and O'Meara, B. C. (2020). *OUwie: Analysis of evolutionary rates in an OU framework. R package version 2.4*. Available at: <https://github.com/thej02214/OUwie>.
- Blackmon, H., Justison, J., Mayrose, I., and Goldberg, E. E. (2019). Meiotic drive shapes rates of karyotype evolution in mammals. *Evol.* 73, 511–523. doi: 10.1111/evo.13682
- Blomberg, S. P., Rathnayake, S. I., and Moreau, C. M. (2020). Beyond Brownian motion and the Ornstein-Uhlenbeck process: Stochastic diffusion models for the evolution of quantitative characters. *Am. Nat.* 195, 145–165. doi: 10.1086/706339
- Brochmann, C., Brysting, A. K., Alsos, I. G., Borgen, L., Grundt, H. H., and Scheen, A. (2004). Polyploidy in arctic plants. *Biol. J. Linn. Soc.* 82, 521–536. doi: 10.1111/j.1095-8312.2004.00337.x
- Brys, R., and Jacquemyn, H. (2011). Variation in the functioning of autonomous self-pollination, pollinator services and floral traits in three *Centaurea* species. *Ann. Bot.* 107, 917–925. doi: 10.1093/aob/mcr032
- Carscadden, K. A., Emery, N. A., Arnillas, C. A., Cadotte, M. W., Afkhami, M. E., Gravel, D., et al. (2020). Niche breadth: Causes and consequences for ecology, evolution, and conservation. *Q. Rev. Biol.* 95, 179–214. doi: 10.1086/710388
- Carta, A., and Escudero, M. (2023). Karyotypic diversity: a neglected trait to explain angiosperm diversification? *Evol.* 77, 1158–1164. doi: 10.1093/evolut/qpaa014
- Castro, M., Loureiro, J., Serrano, M., Tavares, D., Husband, B. C., Siopa, C., et al. (2019). Mosaic distribution of cytotypes in a mixed- ploidy plant species, *Jasione montana*: Nested environ- mental niches but low geographical overlap. *Bot. J. Linn.* 190, 51–66. doi: 10.1093/botlinnean/boz007
- Cheng, J., Li, J., Zhang, Z., Lu, H., Chen, G., Yao, B., et al. (2021). Autopolyploidy-driven range expansion of a temperate-originated plant to pan-tropic under global change. *Ecol. Monogr.* 91, e01445. doi: 10.1002/ecm.1445
- Coca-de-la-Iglesia, M., Valcárcel, V., and G. Medina, N. (2023). Biodiversity online databases: An applied R protocol to get and curate spatial and climatic data. *Bio Protoc.* 13, e4847. doi: 10.21769/BioProtoc.4847
- Coca-de-la-Iglesia, M., Valcárcel, V., Gallego-Narbón, A., Wen, J., and Medina, N. G. (2024). Inferring ancestral climatic niches in a world without clear-cut climatic boundaries. *Authorea*. doi: 10.22541/au.171180716.62099650/v1
- Darwin, C. R. (1876). *The effects of cross and self-fertilisation in the vegetable kingdom* (London, UK: John Murray).
- Díaz-Lifante, Z. (2012). "Centaurea," in *Flora Iberica 11*. Eds. C. Romero and A. Quintanar (CSIC, Madrid: Real Jardín Botánico), 49–81.
- Dray, S., and Dufour, A. (2007). The *ade4* package: implementing the duality diagram for ecologists. *J. Stat. Softw.* 22, 1–20. doi: 10.18637/jss.v022.i04
- Escudero, M., Maguilla, E., Márquez-Corro, J. I., Martín-Bravo, S., Mayrose, I., Shafir, A., et al. (2023). "Using ChromEvol to Determine the Mode of Chromosomal Evolution," in *Plant Cytogenetics and Cytogenomics*, vol. 2672. Eds. T. Heitkam and S. García (Humana, New York, NY). doi: 10.1007/978-1-0716-3226-0_32
- Escudero, M., Martín-Bravo, S., Mayrose, I., Fernández-Mazuecos, M., Fiz-Palacios, O., Hipp, A., et al. (2014). Karyotypic changes through dysploidy persist longer over evolutionary time than polyploid changes. *PLoS One* 9, e85266. doi: 10.1371/journal.pone.0085266
- Escudero, M., and Wendel, J. F. (2020). The grand sweep of chromosomal evolution in angiosperms. *New Phytol.* 228, 805–808. doi: 10.1111/nph.16802
- FitzJohn, R. G. (2012). Diversitree: comparative phylogenetic analyses of diversification in R. *MEE* 3, 1084–1092. doi: 10.1111/j.2041-210X.2012.00234.x
- Freyman, W. A., and Höhna, S. (2018). Cladogenetic and anagenetic models of chromosome number evolution: a Bayesian model averaging approach. *Syst. Biol.* 67, 195–215. doi: 10.1093/sysbio/syx065
- Glick, L., and Mayrose, I. (2014). ChromEvol: assessing the pattern of chromosome number evolution and the inference of polyploidy along a phylogeny. *MBE* 31, 1914–1922. doi: 10.1093/molbev/msu122
- Godsoe, W., Larson, M. A., Glennon, K. L., and Segraves, K. A. (2013). Polyploidization in *Heuchera cylindrica* (Saxifragaceae) did not result in a shift in climatic requirements. *Am. J. Bot.* 100, 496–508. doi: 10.3732/ajb.1200275
- Goodwillie, C., Sargent, R. D., Eckert, C. G., Elle, E., Geber, M. A., Johnston, M. O., et al. (2010). Correlated evolution of mating system and floral display traits in flowering plants and its implications for the distribution of mating system variation. *New Phytol.* 185, 311–321. doi: 10.1111/j.1469-8137.2009.03043.x

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 author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

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/fpls.2024.1439985/full#supplementary-material>

- Grant, V. (1981). *Plant speciation* (New York: Columbia University Press).
- Guggisberg, A., Bretagnolle, F., and Mansion, G. (2006). Allopolyploid origin of the Mediterranean endemic, *Centaureum bianoris* (Gentianaceae), inferred by molecular markers. *Syst. Bot.* 31, 368–379. doi: 10.1600/03636440677585937
- Hansen, T. F. (1997). Stabilizing selection and the comparative analysis of adaptation. *Evol* 51, 1341–1351. doi: 10.2307/2411186
- Harmon, L. J., Weir, J. T., Brock, C. D., Glor, R. E., and Challenger, W. (2008). GEIGER: Investigating evolutionary radiations. *Bioinform* 24, 129–131. doi: 10.1093/bioinformatics/btm538
- Hijmans, R. J., Cameron, S., Parra, J., Jones, P. G., and Jarvis, A. (2005). *WorldClim, version 1.3* (Berkeley: University of California). Available at: <http://worldclim.org>.
- Höhna, S., Landis, M. J., Heath, T. A., Boussau, B., Lartillot, N., Moore, B. R., et al. (2016). RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. *Syst. Biol.* 65, 726–736. doi: 10.1093/sysbio/syw021
- Hua, X., and Wiens, J. J. (2013). How does climate influence speciation? *Am. Nat.* 182, 1–12. doi: 10.1086/670690
- Husband, B. C., and Barrett, S. C. H. (1992). Genetic drift and the maintenance of the style length polymorphism in tristylous populations of *Eichhornia paniculata* (Pontederiaceae). *Hered* 69, 440–449. doi: 10.1038/hdy.1992.148
- Jiao, Y., Wickett, N. J., Ayyampalayam, S., Chanderbali, A. S., Lena Landherr, L., Ralph, P. E., et al. (2011). Ancestral polyploidy in seed plants and angiosperms. *Nature* 473, 97–100. doi: 10.1038/nature09916
- Jiménez-Lobato, V., Escudero, M., Díaz Lifante, Z., Andrés Camacho, C., De Castro, A., Mansion, G., et al. (2019). Evolution of reproductive traits and selfing syndrome in the sub-endemic Mediterranean genus *Centaureum* Hill (Gentianaceae). *Bot. J. Linn. Soc.* 191, 216–235. doi: 10.1093/botlinnean/boz036
- Landis, J. B., Soltis, D. E., Li, Z., Marx, H. E., Barker, M. S., Tank, D. C., et al. (2018). Impact of whole-genome duplication events on diversification rates in angiosperms. *Am. J. Bot.* 105, 348–363. doi: 10.1002/ajb2.1060
- Lavania, U. C. (2020). Plant speciation and polyploidy: in habitat divergence and environmental perspective. *Nucleus* 63, 1–5. doi: 10.1007/s13237-020-00311-6
- Leitch, I. J., and Bennett, M. D. (2004). Genome downsizing in polyploid plants. *Biol. J. Linn. Soc.* 82, 651–663. doi: 10.1111/j.1095-8312.2004.00349.x
- Levin, D. A. (2002). *The role of chromosomal change in plant evolution* (New York: Oxford University Press). doi: 10.1093/oso/9780195138597.001.0001
- Levin, D. A. (2004). The ecological transition in speciation. *New Phytol.* 161, 91–96. doi: 10.1046/j.1469-8137.2003.00921.x
- Levin, D. A. (2019). Plant speciation in the age of climate change. *Ann. Bot.* 124, 769–775. doi: 10.1093/aob/mcz108
- Lloyd, D. G., and Schoen, D. J. (1992). Self- and cross-fertilization in plants. I. Functional dimensions. *IJPS* 153, 358–369. doi: 10.1086/297040
- Maguilla, E., Escudero, M., Jiménez-Lobato, V., Díaz-Lifante, Z., Andrés-Camacho, C., and Arroyo, J. (2021). Polyploidy expands the range of *Centaureum* (Gentianaceae). *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.650551
- Mansion, G., and Struwe, L. (2004). Generic delimitation and phylogenetic relationships within the subtribe Chironiinae (Chironieae: Gentianaceae), with special reference to *Centaureum*: evidence from nrDNA and cpDNA sequences. *Molec. Phylogen. Evol.* 32, 951–977. doi: 10.1016/j.ympev.2004.03.016
- Mansion, G., Zeltner, L., and Bretagnolle, F. (2005). Phylogenetic patterns and polyploid evolution within the Mediterranean genus *Centaureum* (Gentianaceae - Chironieae). *Taxon* 54, 931–950. doi: 10.2307/25065479
- Mayrose, I., Barker, M. S., and Otto, S. P. (2010). Probabilistic models of chromosome number evolution and the inference of polyploidy. *Syst. Biol.* 59, 132–144. doi: 10.1093/sysbio/syp083
- Mayrose, I., Zhan, S. H., Rothfels, C. J., Magnuson-Ford, K., Barker, M. S., Rieseberg, L. H., et al. (2011). Recently formed polyploid plants diversify at lower rates. *Science* 333, 1257. doi: 10.1126/science.1207205
- Ornduff, R. (1969). Reproductive biology in relation to systematics. *Taxon* 18, 121–133.
- Otto, S. P., and Whitton, J. (2000). Polyploid incidence and evolution. *Annu. Rev. Genet.* 34, 401–437. doi: 10.1146/annurev.genet.34.1.401
- Pannell, J. R. (2015). Evolution of the mating system in colonizing plants. *Mol. Ecol.* 24, 2018–2037. doi: 10.1111/mec.13087
- Paradis, E., and Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinform* 35, 526–528. doi: 10.1093/bioinformatics/bty633
- Pennell, M. W., and Harmon, L. J. (2013). An integrative view of phylogenetic comparative methods: Connections to population genetics, community ecology, and paleobiology. *Ann. N. Y. Acad. Sci.* 1289, 90–105. doi: 10.1111/nyas.12157
- Peterson, A. T., Soberón, J., and Sánchez-Cordero, V. (1999). Conservatism of ecological niches in evolutionary time. *Science* 285, 1265–1267. doi: 10.1126/science.285.5431.1265
- POWO, Plants of the World Online (2023). Royal botanic gardens, kew. Available at: <http://www.plantsoftheworldonline.org>.
- Qiao, H., Peterson, A. T., Myers, C. E., Yang, Q., and Saupe, E. E. (2024). Ecological niche conservatism spurs diversification in response to climate change. *Nat. Ecol. Evol.* 8, 729–738. doi: 10.1038/s41559-024-02344-5
- R Core Team (2023). *R: A Language and Environment for Statistical computing* (Vienna, Austria: R Foundation for Statistical Computing). Available at: <https://www.R-project.org/>.
- Rech, A. R., Dalsgaard, B., Sandel, B., Sonne, J., Svenning, J.-C., Holmes, N., et al. (2016). The macroecology of animal versus wind pollination: Ecological factors are more important than historical climate stability. *Plant Ecol. Diversity* 9, 253–262. doi: 10.1080/17550874.2016.1207722
- Revell, L. J. (2012). phytools: An R package for phylogenetic comparative biology (and other things). *MEE* 3, 217–223. doi: 10.1111/j.2041-210X.2011.00169.x
- Rice, A., Šmarda, P., Novosolov, M., Drori, M., Glick, L., Sabbath, N., et al. (2019). The global biogeography of polyploid plants. *Nat. Ecol. Evol.* 3, 265–273. doi: 10.1038/s41559-018-0787-9
- Richards, A. J. (1986). Plant breeding systems. *J. Trop. Ecol.* 3, 279–280. doi: 10.1017/S0266467400002170
- Rundle, H. D., and Nosil, P. (2005). Ecological speciation. *Ecol. Lett.* 8, 336–352. doi: 10.1111/j.1461-0248.2004.00715.x
- Schluter, D. (2000). *The ecology of adaptive radiation* (Oxford: Oxford University Press). doi: 10.1093/oso/9780198505235.001.0001
- Schluter, D. (2001). Ecology and the origin of species. *TREE* 16, 372–380. doi: 10.1016/S0169-5347(01)02198-X
- Schluter, D. (2009). Evidence for ecological speciation and its alternative. *Science* 323, 737–741. doi: 10.1126/science.1160006
- Schubert, I., and Lysak, M. A. (2011). Interpretation of karyotype evolution should consider chromosome structural constraints. *Trends Genet.* 27, 207–216. doi: 10.1016/j.tig.2011.03.004
- Sicard, A., and Lenhard, M. (2011). The selfing syndrome: a model for studying the genetic and evolutionary basis of morphological adaptation in plants. *Ann. Bot.* 107, 1433–1443. doi: 10.1093/aob/mcr023
- Soltis, D. E., Albert, V. A., Leebens-Mack, J., Bell, C. D., Paterson, A. H., Zheng, C., et al. (2009). Polyploidy and angiosperm diversification. *Am. J. Bot.* 96, 336–348. doi: 10.3732/ajb.0800079
- Soltis, P. S., Marchant, D. B., Van de Peer, Y., and Soltis, D. E. (2015). Polyploidy and genome evolution in plants. *Curr. Opin. Genet. Dev.* 35, 119–125. doi: 10.1016/j.gde.2015.11.003
- Stebbins, G. L. (1950). *Variation and evolution in plants* (New York: Columbia University Press). doi: 10.7312/steb94536
- Stebbins, G. L. (1957). Self fertilization and population variability in the higher plants. *Am. Nat.* 91, 337–354.
- Stebbins, G. L. (1971). *Chromosomal evolution in higher plants* (London: Edward Arnold).
- Stebbins, G. L. (1974). *Flowering plants: evolution above the species level* (Cambridge, MA, USA: Belknap Press).
- Theodoridis, S., Randin, C., Broennimann, O., Patsiou, T., and Conti, E. (2013). Divergent and narrower climatic niches characterize polyploid species of European primroses in *Primula* sect. *Aleuritia*. *J. Biogeogr* 40, 1278–1289. doi: 10.1111/jbi.12085
- Thompson, K. A., Husband, B. C., and Maherali, H. (2014). Climatic niche differences between diploid and tetraploid cytotypes of *Chamerion angustifolium* (Onagraceae). *Am. J. Bot.* 101, 1868–1875. doi: 10.3732/ajb.1400184
- Thompson, J. D., and Lumaret, R. (1992). The evolutionary dynamics of polyploid plants: origins, establishment and persistence. *TREE* 7, 302–307. doi: 10.1016/0169-5347(92)90228-4
- Tribble, C. M., Freyman, W. A., Landis, M. J., Lim, J. Y., Barido-Sottani, J., and Kopperud, B. T. (2022). RevGadgets: An R package for visualizing Bayesian phylogenetic analyses from RevBayes. *MEE* 13, 314–323. doi: 10.1111/2041-210X.13750
- Valdés-Flórido, A., González-Toral, C., Maguilla, E., Cires, E., Díaz-Lifante, Z., Andrés-Camacho, C., et al. (2024). Polyploidy and hybridization in the Mediterranean: unravelling the evolutionary history of *Centaureum* (Gentianaceae). *Ann. Bot.* 134, 247–262. doi: 10.1093/aob/mcae066
- Van de Peer, Y. (2011). A mystery unveiled. *Genome Biol.* 12, 113. doi: 10.1186/gb-2011-12-5-113
- Van de Peer, Y., Maere, S., and Meyer, A. (2009). The evolutionary significance of ancient genome duplications. *Nat. Rev. Genet.* 10, 725–732. doi: 10.1038/nrg2600
- Vimala, Y., Lavania, S., and Lavania, U. C. (2021). Chromosome change and karyotype differentiation—implications in speciation and plant systematics. *Nucleus* 64, 33–54. doi: 10.1007/s13237-020-00343-y
- Wiens, J. J. (2004). Speciation and ecology revisited: phylogenetic niche conservatism and the origin of species. *Evol* 58, 193–197. doi: 10.1111/j.0014-3820.2004.tb01586.x
- Wood, T. E., Takebayashi, N., Barker, M. S., Mayrose, I., Greenspoon, P. B., and Rieseberg, L. H. (2009). The frequency of polyploid speciation in vascular plants. *Proc. Nat. Acad. Sci. U.S.A.* 106, 13875–13879. doi: 10.1073/pnas.0811575106
- Zeltner, L. (1970). Recherches de biosystème sur les genres *Blackstonia* Huds. Et *Centaureum* Hill (Gentianaceae). *Bull. Soc. Neuchâteloise Sci. Nat.* 93, 1–164.



OPEN ACCESS

EDITED BY

Marcial Escudero,
Sevilla University, Spain

REVIEWED BY

Kyong-Sook Chung,
Jungwon University, Republic of Korea
Jose Planta,
University of the Philippines Diliman,
Philippines

*CORRESPONDENCE

Jun-Min Li
✉ lijm@tzc.edu.cn

RECEIVED 19 September 2024

ACCEPTED 02 December 2024

PUBLISHED 20 December 2024

CITATION

Tao Y-T, Chen L-X, Jiang M, Jin J, Sun Z-S,
Cai C-N, Lin H-Y, Kwok A, Li J-M and
van Kleunen M (2024) Complete chloroplast
genome data reveal the existence of the
Solidago canadensis L. complex and its
potential introduction pathways into China.
Front. Plant Sci. 15:1498543.
doi: 10.3389/fpls.2024.1498543

COPYRIGHT

© 2024 Tao, Chen, Jiang, Jin, Sun, Cai, Lin,
Kwok, Li and van Kleunen. This is an open-
access article distributed under the terms of
the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/)
(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.

Complete chloroplast genome data reveal the existence of the *Solidago canadensis* L. complex and its potential introduction pathways into China

Yu-Tian Tao^{1,2}, Lu-Xi Chen¹, Ming Jiang¹, Jie Jin²,
Zhong-Shuai Sun¹, Chao-Nan Cai^{1,3}, Han-Yang Lin^{1,3},
Allison Kwok⁴, Jun-Min Li^{1,3*} and Mark van Kleunen^{1,3,5}

¹Zhejiang Provincial Key Laboratory of Plant Evolutionary Ecology and Conservation, Taizhou University, Taizhou, China, ²School of Electronics and Information Engineering, Taizhou University, Taizhou, China, ³School of Advances Study, Taizhou University, Taizhou, China, ⁴Environmental and Life Sciences Graduate Program, Trent University, Peterborough, ON, Canada, ⁵Department of Biology, University of Konstanz, Konstanz, Germany

Solidago canadensis, native to North America, is an invasive species in many areas of the world, where it causes serious damage to natural ecosystems and economic losses. However, a dearth of genetic resources and molecular markers has hampered our understanding of its invasion history. Here, we *de novo* assembled 40 complete chloroplast genomes of *Solidago* species, including 21 *S. canadensis* individuals, 15 *S. altissima* individuals, and four *S. decurrens* individuals, the sizes of which ranged from 152,412 bp to 153,170 bp. The phylogenetic trees based on the complete chloroplast genome sequences and nuclear genome-wide SNP data showed that *S. canadensis* and *S. altissima* cluster together and form a monophyletic pair, as sister to *S. decurrens*, indicating the existence of the *S. canadensis* L. complex in China. Three potential introduction pathways were identified. The chloroplast-genome structure and gene contents are conservative in the genomes of the *S. canadensis* L. complex and *S. decurrens*. The analysis of sequence divergence indicated five variable regions, and 10 chloroplast protein-coding genes that underwent positive selection were identified. Our findings shed new light on the invasion history of *S. canadensis* and the data sets generated in this study will facilitate future research on its chloroplast genome evolution.

KEYWORDS

Solidago, chloroplast genome, phylogenetic analysis, introduction pathways, molecular marker

1 Introduction

The genus *Solidago* L. (Asteraceae) has currently ca. 140 recognized species (Semple and Cook, 2006; Beck and Semple, 2015). The majority of *Solidago* species are native to North America, while a few species are native to South America or Eurasia (Beaudry and Chabot, 1957; McNeil, 1976). However, at least 50 of the North American *Solidago* species have been introduced as ornamental plants to Europe in the 17th and 18th century (Loudon, 1850), and most likely also elsewhere. Some of those species have become naturalized in Europe, as well as in Asia and other parts of the world (Weber, 2017; Van Kleunen et al., 2019). One of them, *Solidago canadensis* L., is now one of the most aggressive invasive plant species in Europe and in China (Xu et al., 2012; Zhao et al., 2015). However, due to differences in morphology and ploidy, there is uncertainty about whether the plants in China and Europe belong to the same species, and thus whether they both belong to *S. canadensis*.

The taxonomy of *S. canadensis* is complicated, and due to its wide variability in leaf sizes, pubescence, rhizome lengths, flowerhead sizes, and ploidy levels, multiple varieties are recognized (Croat, 1972; Melville and Morton, 1982; Beck and Semple, 2015). Some of the taxa that some taxonomists consider to be varieties of *S. canadensis* are considered to be separate species by others. Consequently, uncertainty exists around the true identity of what is now usually called *Solidago canadensis* in the invaded ranges. For example, a plant-morphological study by Weber (1997) concluded that the taxon in Europe is highly variable (Weber, 1997), and that, particularly due to its characteristic nodding shoot tips, it has to be closely related to *S. canadensis* var. *scabra* (Muhl.) Torr. & Gray, which is considered a synonym of *S. altissima* L. (Fernald, 1950; Croat, 1972). Consequently, many older publications refer to this invasive species in Europe as *S. altissima* (Weber and Schmid, 1998). However, *S. altissima* is hexaploid, whereas the taxon in Europe is diploid. Since the early 2000s, most studies use the name *S. canadensis* (Van Kleunen and Schmid, 2003). However, Verloove et al. (2017) reported the first evidence for an established hexaploid population of *S. altissima* in Europe (Verloove et al., 2017). In Asia, a highly invasive hexaploid *Solidago* has been identified in Japan as *S. altissima* (<https://www.nies.go.jp/biodiversity/invasive/DB/detail/80600e.html>), but *S. canadensis* has also been identified as naturalized in Japan. In China, however, an highly invasive hexaploid *Solidago* has been identified as *S. canadensis*, but *S. altissima* has also been recorded as naturalized in China (Ma and Qiang, 2007). So, both in Europe and Asia, *S. altissima* and *S. canadensis* have been identified, but doubts about the exact taxonomic status remain.

Based on botanical identification literature, *S. altissima* should distinguish itself from *S. canadensis* by having fewer leaf serrations, more disk florets, longer ray florets, a higher involucre height, larger pollen, and a nodding shoot tip, and by being hexaploid instead of diploid (Croat, 1972; Melville and Morton, 1982; Weber, 1997). Based on an extensive comparison of morphological characteristics, Semple et al. (2015) concluded that *S. altissima* and *S. canadensis* should be treated as separate species

(Semple et al., 2015). However, some plants, like *S. canadensis* in Europe, have characteristics of both species, and therefore molecular marker data might give more precise information on its identity (Weber, 1997; Semple et al., 2023). Recently, Tian et al. (2023) used nuclear ribosomal ITS and chloroplast trnH-psbA intergenic spacer sequences on specimens from North America, Europe and China, and found that specimens morphologically identified to be *S. canadensis* and *S. altissima* were >99% identical, and should be considered as one species complex (Tian et al., 2023). However, other molecular markers might provide more conclusive evidence.

Here, we used next-generation sequencing to obtain complete chloroplast genome sequences of 40 *Solidago* individuals from the native range in the US, and the non-native ranges in Asia (China and Japan) and Europe. An advantage of the chloroplast genome, compared to the nuclear genome, is that it is smaller and has more copy numbers (Ashworth, 2017). Furthermore, the chloroplast genome is usually maternally inherited and does not undergo gene recombination (Wang et al., 2023). Therefore, complete chloroplast genomes have frequently been used to study population genetics and species evolution, and to determine phylogenetic relationships among plant species (Zhang and Chen, 2022; Jiang et al., 2023). The aims of our study were to (i) obtain complete chloroplast genomes of specimens morphologically identified to be *S. canadensis* and *S. altissima*, to evaluate the usefulness of the chloroplast genome in species identification and phylogenetics, and (ii) to obtain insight into whether the invasive *Solidago* in Europe and Asia are most likely *S. canadensis* or *S. altissima*. Furthermore, we searched for highly variable genome regions that could be used to develop barcode markers, and we tested for genes with positive selection to identify genes that may play a role in evolutionary adaptation.

2 Materials and methods

2.1 Plant materials and genome data sources

For *S. canadensis* and *S. altissima*, seeds and leaf samples of a total of 36 individuals were collected from the native range (11 from the United States and 10 from Canada) and the introduced range (10 from China, 4 from Europe, and 1 from Japan). In detail, seeds of reference native *S. canadensis* L. and *S. altissima* L. (Asteraceae) were obtained from The Germplasm Resources Information Network of the United States Department of Agriculture (GRIN/USDA) with accession numbers W6 52837 and W6 57335, respectively. The seeds were sown and leaf material of one offspring plant from each of the *S. canadensis* and *S. altissima* accessions were used for chloroplast-genomic analysis. Additionally, leaf samples of *S. altissima* herbarium specimens were obtained: three samples from the Wuhan Institute of Botany (HIB; accession numbers HIB0188587, HIB0188588, and HIB0188589) and one sample from the Chinese Field Herbarium (CFH; accession number CSH0115776). We also obtained leaf

material of one *S. canadensis* specimen from the Specimen Museum of Zhejiang University (accession number HZU60102855; [Supplementary Table S1](#)). We also used 29 leaf samples of *S. canadensis* collected in the field, including nine samples from the United States, six samples from China, and four samples from Europe. In addition, we used 10 samples of *S. altissima* from Peterborough, Canada. The healthy leaf samples were washed and then dried in silica gel for preservation until DNA extraction. In total, we had 21 *S. canadensis* samples and 15 *S. altissima* samples. For the samples collected in the field, *S. canadensis* and *S. altissima* were mainly distinguished by the length and arrangement of epidermal hairs on stems and leaves, and the shape and margins of leaves. One individual each of *S. canadensis*, *S. altissima* and *S. decurrens* are shown in [Supplementary Figure S1](#).

As a control to show that the chloroplast genomic analysis can distinguish *Solidago* species, we also used samples of *S. decurrens*, which is native in China. We had one *S. decurrens* herbarium specimen from the Specimen Museum of Zhejiang University (accession number HZU60120963), and three leaf samples of *S. decurrens* collected in the field in China ([Supplementary Tables S1, S2](#)). In addition, we also obtained complete chloroplast genome sequence data of *S. decurrens* (NC_053705.1) from the National Center for Biotechnology Information (NCBI, <https://www.ncbi.nlm.nih.gov/>). As additional outgroups, we also obtained from NCBI complete chloroplast genome sequence data of six other Asteraceae species, including *Aster flaccidus* (MN122101.1) ([Zhou et al., 2019](#)), *Conyza bonariensis* (NC_035884.1) ([Hereward et al., 2017](#)), *Erigeron canadensis* (MT806101.1), *Heteroplexis incana* (NC_048508.1), *Lagenophora cuchumatana* (NC_034819.1) ([Vargas et al., 2017](#)), and *Symphotrichum subulatum* (NC_050667.1) ([Hu, 2020](#)).

2.2 Total DNA extraction

Approximately 1 g of each of the dried leaf samples was used for genomic DNA extraction with the modified CTAB method ([Doyle and Doyle, 1987](#)). A library with insertion sizes of 300–500 bp was constructed for paired-end sequencing using the Illumina Novaseq 6000 platform (Illumina, San Diego, CA, United States).

2.3 Chloroplast genome assembly and annotation

Chloroplast genomes were *de novo* assembled using NOVOPlasty version 4.2 with default parameters ([Dierckxsens et al., 2017](#)). The GeSeq tool was used to annotate the assembled chloroplast genome sequences ([Tillich et al., 2017](#)). Results of software based annotation were verified by manual inspection using Geneious Prime 2021.1.1 (Biomatters Ltd., Auckland, New Zealand). Chloroplast gene circular maps were drawn with the OrganellarGenomeDRAW tool (OGDRAW) ([Greiner et al., 2019](#)).

2.4 Nuclear genome SNP data calling

For the nuclear phylogenetic tree construction, genome-wide SNP data was obtained according to the following steps. First, the raw sequencing data were cleaned using Trimmomatic (version 0.39) ([Bolger et al., 2014](#)) to remove adapters and low-quality sequences (the reads were scanned with a 4-base wide sliding window, cut when the average quality per base dropped below 30, and dropped when they were less than 120 bases long). Then the cleaned reads were mapped to the nuclear genome of *S. canadensis* using the software Burrows–Wheeler Alignment (BWA, version 0.7.17) with default parameters ([Li and Durbin, 2009](#)). The variant calling was performed using the Genome Analysis Toolkit (GATK) Best Practices Pipeline (version 4.2.0.0) ([Van der Auwera et al., 2013](#)), and variants were filtered with the VariantFiltration tool from GATK (-Window 4, -filter “QD < 2.0 || MQ < 40.0 || DP < 10 || FS > 60.0”). Finally, the filtered SNPs of the *Solidago* species were merged using the software BCFtools (version 1.11) with default parameters ([Danecek et al., 2021](#)).

2.5 Phylogenetic tree construction

In the phylogenetic tree construction, the six non-*Solidago* species that belong to the same Asteraceae Tribe (*A. flaccidus*, *C. bonariensis*, *E. canadensis*, *H. incana*, *L. cuchumatana*, and *S. subulatum*), served as outgroups for the chloroplast genome tree. In the nuclear genome tree, *S. subulatum* was served as outgroup. The whole-chloroplast-genome sequences were aligned using MAFFT v7.471 ([Kato and Standley, 2013](#)) and manually adjusted. To determine the phylogenetic relationships, a maximum-likelihood tree of the chloroplast genome was constructed based on the manually adjusted aligned chloroplast genomes. The nuclear genome tree was built based on the genome-wide SNP data. The best-fit substitution model was selected by the software MrModeltest v2.4, which selects general time-reversible models with invariant sites and the gamma-rate heterogeneity (GTR+G+I) as the best-fitting substitution model. The maximum likelihood tree was constructed using IQ-TREE version 1.6.12 ([Nguyen et al., 2015](#)) with 1000 bootstrap replicates.

For principal component analysis (PCA), eigenvalues for consensus genotyping from SNP data of chloroplast genomes in *Solidago* species were calculated in plink v1.9 ([Purcell et al., 2007](#)). Then PCA plots were created by plotting the first two eigenvalues by ggplot2 package ([Wickham, 2016](#)) in R.

2.6 Comparative genome analysis

Chloroplast genome comparisons were carried out and visualized with the online software mVISTA ([Frazer et al., 2004](#)) based on annotation information of the sample *Solidago canadensis* 21 ([Supplementary Table S2](#)) in Shuffle LAGAN mode. DNAsnp

v5.10 software (Librado and Rozas, 2009) was used to perform nucleotide diversity (π) and sequence-polymorphism analysis, and the settings were a step size of 200 bp and a window length of 600 bp.

Positive selection analysis was performed by the software EasyCodeML (Gao et al., 2019) across the chloroplast genomes, using a site-specific model with five site models (M0, M1a & M2a, M7 & M8) to calculate the synonymous (dS) and non-synonymous (dN) substitution rates, and their ratio ($\omega = dN/dS$). EasyCodeML uses CodeML of the PAML software (Gao et al., 2019). The site-specific model allowed the ω ratio to vary among sites while maintaining a fixed ω ratio in all the branches (Jiang et al., 2023). Specifically, M1a (nearly neutral) vs M2a (positive selection), and M7 (β) vs M8 (β & ω) were applied to find sequences that had undergone positive selection (Yang and Nielsen, 2002). Likelihood ratio tests (LRT) were used for selection-strength evaluation with the comparison of M1a vs M2a and M7 vs M8. A p value of the chi-square statistic smaller than 0.05 was considered to be significant. The site models of M2a and M8 were implemented using Bayes Empirical Bayes (BEB) inference (Yang et al., 2005) in order to

estimate the posterior probabilities and positive selection pressures on the selected genes.

3 Results

3.1 Chloroplast-genome features of *S. canadensis*, *S. altissima*, and *S. decurrens*

There was a large overlap in the lengths of the complete circular chloroplast genomes of *S. canadensis* (152,412 to 153,170 bp; $n = 21$), *S. altissima* (152,957 to 153,126 bp; $n = 15$) and *S. decurrens* (152,728 to 152,863 bp; $n = 5$). Each chloroplast genome had a typical quadripartite structure, including a large single-copy region (LSC, 84,214 bp to 85,061 bp, 84,809 bp to 85,026 bp, and 84,599 bp to 84,736 bp, respectively for *S. canadensis*, *S. altissima*, and *S. decurrens*) and a small single-copy region (SSC, 18,066 bp to 18,143 bp, 18,066 bp to 18,165 bp, and 18,089 bp to 18,095 bp, respectively, for *S. canadensis*, *S. altissima*, and *S. decurrens*), separated by two inverted repeats (IRA and IRB, 25,015 bp to 25,134 bp, 25,013 bp to

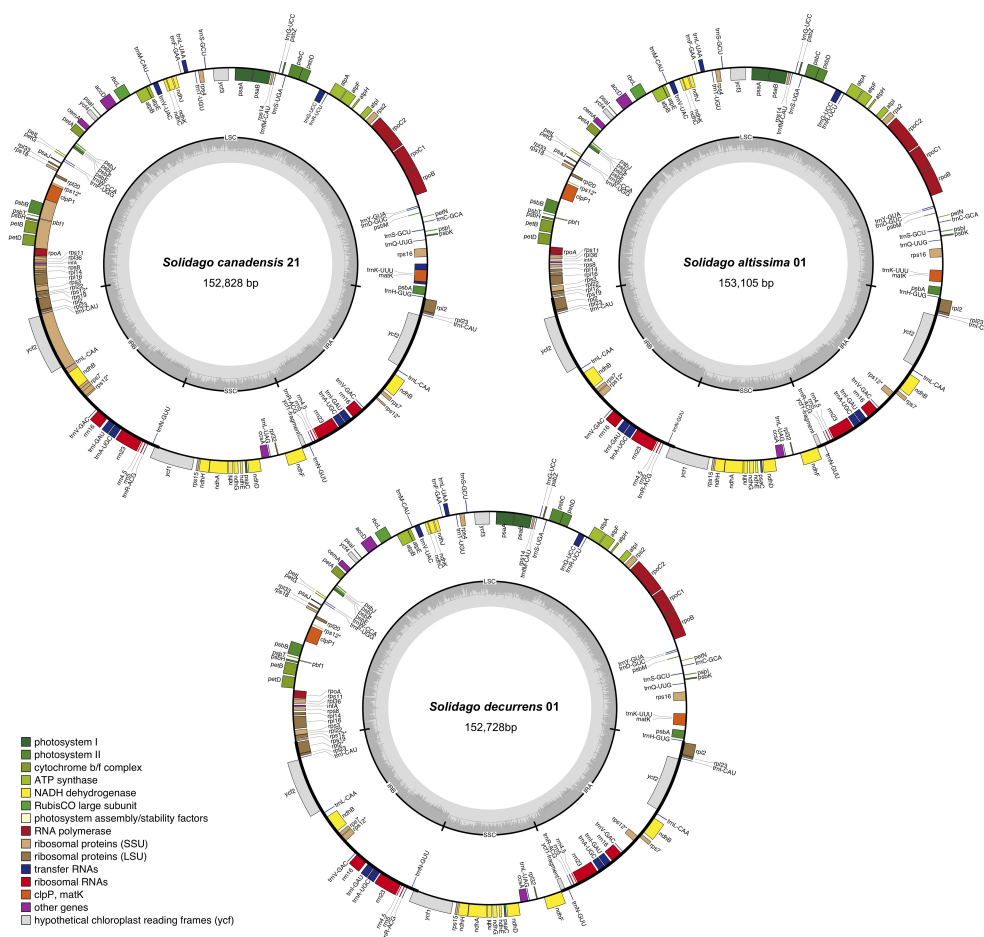


FIGURE 1

Gene map of the chloroplast genomes of *Solidago canadensis*, *S. altissima*, and *S. decurrens*. Genes that are located inside the circle were transcribed in the clockwise direction, and those located outside the circle were transcribed in the counter-clockwise direction. Genes belonging to different functional groups were marked in different colors. The dark-gray columns in the inner circle were related to the GC content, and the lighter-gray columns to the AT content.

25024 bp, and 25,019 bp to 25,020 bp, respectively for *S. canadensis*, *S. altissima*, and *S. decurrens*) (Supplementary Tables S1, S2 and Figure 1).

The chloroplast-genome features of the sampled *S. canadensis*, *S. altissima* and *S. decurrens* plants were largely identical. All the chloroplast genomes of *S. canadensis*, *S. altissima*, and *S. decurrens* showed similar GC contents ranging from 37.16% to 37.29%. The gene numbers and gene orders of the chloroplast genomes were identical. They contained a total of 128 genes, including 85 protein-coding genes, 35 tRNA genes, and 8 rRNA genes (Supplementary Table S2). Of the 128 genes, six protein-coding genes (*rpl2*, *rpl23*, *ycf2*, *ndhB*, *rps7*, and *ycf1*), seven tRNA genes (*trnI*-CAU, *trnL*-CAA, *trnV*-GAC, *trnI*-GAU, *trnA*-UGC, *trnR*-ACG, and *trnN*-GUU), and 4 rRNA genes (*rrn16*, *rrn23*, *rrn4.5*, and *rrn5*) were duplicated in the IR regions. The majority of encoded chloroplast genes are involved in photosynthesis-related metabolic processes (Table 1). There were 15 genes harboring a single intron (*trnK*-UUU, *rps16*, *rpoC1*, *atpF*, *trnG*-UCC, *trnL*-UAA, *trnV*-UAC, *petB*, *petD*, *rpl16*, *rpl2*, *ndhB*, *trnI*-GAU, *trnA*-UGC, and *ndhA*), and tree genes containing two introns (*ycf3*, *rps12*, and *clpP1*) (Supplementary Table S3).

3.2 Phylogenetic analyses

The topology of the phylogenetic trees based on the complete chloroplast genome sequences and nuclear genome-wide SNP data both had strong support (100% bootstrap value) and showed that all *S. canadensis* and *S. altissima* samples formed a monophyletic group, distinct from *S. decurrens* (Figure 2). Principal component analysis (PCA) based on the chloroplast genome sequences also significantly distinguished *S. decurrens* from *S. canadensis* and *S. altissima*, while *S. canadensis* and *S. altissima* were grouped together

(Figure 3). These results suggest that *S. canadensis* and *S. altissima* are one and the same species or belong to one species complex.

Generally, the phylogenetic trees reconstructed from the chloroplast genome data and nuclear genome data showed a largely congruent invasion history of native and invasive *S. canadensis* L. complex (Figure 2). In the monophyletic group, individuals of the *S. canadensis* L. complex in China showed a close relationship with individuals from North America. In addition, one Japanese individual showed close affinities to the Chinese individuals. The European individuals were grouped together and also showed a close relationship with North American individuals. Taken together, the results of the phylogenetic analysis indicate multiple origins and complex invasive routes of the *S. canadensis* L. complex into China.

3.3 Comparative analyses of chloroplast genomes

To investigate the divergence of chloroplast genomes among the accessions of *S. altissima*, *S. canadensis*, and *S. decurrens*, multiple alignments of chloroplast genomes were compared by mVISTA, with the annotated accession *S. canadensis* 21 as reference (Supplementary Figure S2). The results revealed a high similarity of chloroplast genome sequences, but the LSC (from *trnH*-GUG to *rps19*) and SSC (from *ycf1* to *ndhF*) regions showed relatively less conservation compared to the IRA (from *ycf1* to *rpl2*) and IRB (from *rps19* to *ycf1*) regions (Supplementary Figure S2). Moreover, higher divergence was also observed in non-coding regions when compared to coding regions.

Nucleotide diversity (Pi) values were calculated within 600 bp windows to discover hotspots of sequence divergence. The Pi values

TABLE 1 List of annotated genes in the chloroplast genome of *S. canadensis* L. complex and *S. decurrens*.

Group of genes	Gene names
Photosystem I	<i>psaA</i> , <i>psaB</i> , <i>psaC</i> , <i>psaI</i> , <i>psaJ</i>
Photosystem II	<i>psbA</i> , <i>psbB</i> , <i>psbC</i> , <i>psbD</i> , <i>psbE</i> , <i>psbF</i> , <i>psbH</i> , <i>psbI</i> , <i>psbJ</i> , <i>psbK</i> , <i>psbL</i> , <i>psbM</i> , <i>psbT</i> , <i>psbZ</i> , <i>psb1</i>
Cytochrome b/f complex	<i>petA</i> , <i>petB</i> , <i>petD</i> , <i>petG</i> , <i>petL</i> , <i>petN</i>
ATP synthase	<i>atpA</i> , <i>atpB</i> , <i>atpE</i> , <i>atpF</i> , <i>atpH</i> , <i>atpI</i>
NADP dehydrogenase	<i>ndhA</i> , <i>ndhB</i> , <i>ndhC</i> , <i>ndhD</i> , <i>ndhE</i> , <i>ndhF</i> , <i>ndhG</i> , <i>ndhH</i> , <i>ndhI</i> , <i>ndhJ</i> , <i>ndhK</i>
RubisCO large subunit	<i>rbcl</i>
RNA polymerase	<i>rpoA</i> , <i>rpoB</i> , <i>rpoC1</i> , <i>rpoC2</i>
Ribosomal proteins (SSU)	<i>rps2</i> , <i>rps3</i> , <i>rps4</i> , <i>rps7</i> , <i>rps8</i> , <i>rps11</i> , <i>rps12</i> , <i>rps14</i> , <i>rps15</i> , <i>rps16</i> , <i>rps18</i> , <i>rps19</i>
Ribosomal proteins (LSU)	<i>rpl2</i> , <i>rpl14</i> , <i>rpl16</i> , <i>rpl20</i> , <i>rpl22</i> , <i>rpl23</i> , <i>rpl32</i> , <i>rpl33</i> , <i>rpl36</i>
Hypothetical chloroplast reading frames (ycf)	<i>ycf1</i> , <i>ycf2</i> , <i>ycf3</i> , <i>ycf4</i>
Other genes	<i>accD</i> , <i>ccsA</i> , <i>cemA</i> , <i>clpP1</i> , <i>infA</i> , <i>matK</i>
Ribosomal RNAs	<i>rrn16</i> , <i>rrn23</i> , <i>rrn4.5</i> , <i>rrn5</i>
Transfer RNAs	<i>trnH</i> -GUG, <i>trnK</i> -UUU, <i>trnQ</i> -UUG, <i>trnS</i> -GCU, <i>trnC</i> -GCA, <i>trnD</i> -GUC, <i>trnY</i> -GUA, <i>trnR</i> -UCU, <i>trnG</i> -UCC, <i>trnS</i> -UGA, <i>trnG</i> -UCC, <i>trnM</i> -CAU, <i>trnS</i> -GCU, <i>trnT</i> -UGU, <i>trnL</i> -UAA, <i>trnF</i> -GAA, <i>trnV</i> -UAC, <i>trnM</i> -CAU, <i>trnW</i> -CCA, <i>trnP</i> -UGG, <i>trnI</i> -CAU, <i>trnL</i> -CAA, <i>trnV</i> -GAC, <i>trnI</i> -GAU, <i>trnA</i> -UGC, <i>trnR</i> -ACG, <i>trnN</i> -GUU, <i>trnL</i> -UAG, <i>trnN</i> -GUU, <i>trnR</i> -ACG, <i>trnA</i> -UGC, <i>trnI</i> -GAU, <i>trnV</i> -GAC, <i>trnL</i> -CAA, <i>trnI</i> -CAU

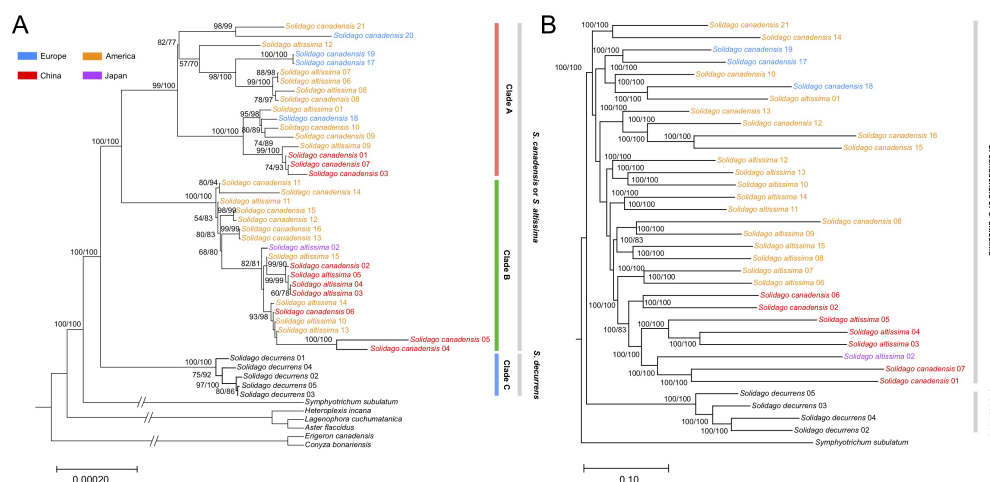


FIGURE 2

Maximum likelihood tree of collected individuals of *Solidago canadensis*, *S. altissima*, and *S. decurrens*. (A) Maximum likelihood tree constructed based on chloroplast genomes. Six published chloroplast genomes in Tribe. Astereae Cass were used as outgroups, including *Symphytotrichum subulatum*, *Aster flaccidus*, *Lagenophora cuchumatana*, *Heteroplexis incana*, *Conyza bonariensis*, and *Erigeron canadensis*. (B) The maximum likelihood tree constructed based on nuclear genome-wide SNP data. *Symphytotrichum subulatum* was used as the outgroup. Due to missing nuclear genome data, the accessions *Solidago canadensis* 03, 04, 05, 09, 11, and 20 were not included in this analysis. A bootstrap value above 70% is labeled beside each node, numbers in parentheses are SH-aLRT support (%)/ultrafast bootstrap support (%). Individuals of the *Solidago canadensis* L. complex sampled in North America are colored orange, those from Europe blue, the one from Japan purple, and those from China red.

of the whole chloroplast genomes varied from 0 to 0.00594, and there was a higher diversity in the LSC and SSC regions than in the IR regions (Figure 4). A total of five highly variable regions with Pi values higher than 0.002 were identified, including *trnG-UCC* - *psbD*, *trnT-UGU* - *trnL-UAA* and *petB* - *petD* in the LSC region, and *ycf1* - *ndhA* and *trnL-UAG-ndhF* in the SSC region (Figure 4). The nucleotide diversity of individuals of the *S. canadensis* L. complex between China and Europe was higher than between

individuals from China and North America (Supplementary Figure S3), which may be due to a limited sample size for Europe.

The non-synonymous (dN) and synonymous (dS) substitution rates of all the protein coding genes were analyzed across chloroplast genomes in the *S. canadensis* L. complex and *S. decurrens*. Most of the genes showed evidence of purifying selection. A total of 10 protein coding genes with significant posterior probabilities suggest that some sites of these genes were

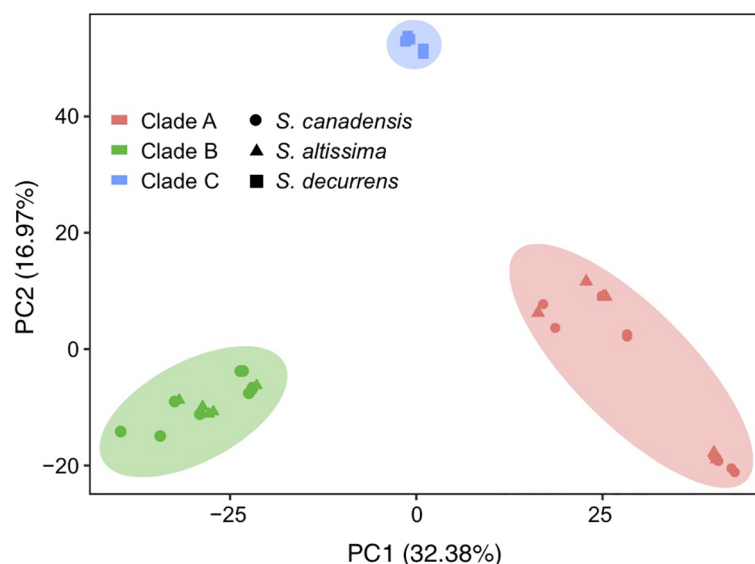


FIGURE 3

Principal component analysis of all chloroplast genomes based on SNPs. The chloroplast genomes of *S. canadensis* are represented as circles, those of *S. altissima* as triangles, and those of *S. decurrens* as squares. The chloroplast genomes located in clade A of the phylogenetic tree in Figure 2 are colored red, those in clade B are colored green, and those in clade C are colored blue.

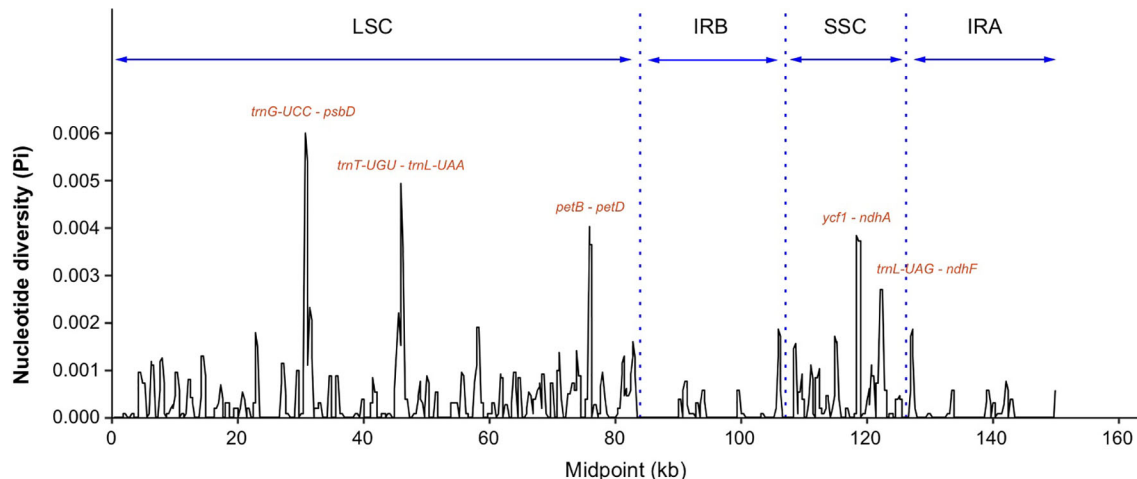


FIGURE 4

Sequence divergence analysis of chloroplast genomes of the *Solidago canadensis* L. complex and *S. decurrens*. Sliding window analysis of nucleotide diversity (Π) had a window length of 600 bp and a step size of 200 bp. Each highly polymorphic region labeled by gene name was annotated on the graph. LSC, large single-copy region; SSC, small single-copy region; IRA and IRB were short for two inverted repeats. *psbD*, Photosystem II reaction center protein D; *petB*, Photosynthetic electron transfer; *petD*, Photosynthetic electron transfer D; *ycf1*, Hypothetical chloroplast reading frame; *ndhA*, NADP dehydrogenase subunit A; *ndhF*, NADP dehydrogenase subunit F; *trnG-UCC*, Transfer RNA glycine with UCC codon; *trnT-UGU*, Transfer RNA threonine with UGU codon; *trnL-UAA*, Transfer RNA leucine with UAA codon; *trnL-UAG*, Transfer RNA leucine with UAG codon.

under positive selection (Supplementary Table S4). The gene *ycf1* showed the highest number ($n = 8$) of amino acid sites with evidence for positive selection (Supplementary Table S4). Interestingly, in the gene *rps19*, the 46th amino acid was alanine in the *S. altissima* and *S. canadensis* accessions, while it was serine in the *S. decurrens* accessions (Supplementary Figure S4).

4 Discussion

In this study, we found that the genome size (152,412–153,170 bp), GC content (37.16–37.29%), the quadripartite structure and gene composition (Supplementary Table S2) of the 40 *Solidago* accessions were consistent with those of other Tribe. Astereae Cass chloroplast genomes, and showed a highly conserved structure with minor differences among the *Solidago* species. The chloroplast genome features of the sampled *S. canadensis* L. complex and *S. decurrens* plants were largely identical (Table 1 and Supplementary Tables S2, S3). Our results confirm the results of Tian et al. (2023), who used both nuclear molecular markers and morphological traits, that *S. decurrens* is different from the other accessions (Tian et al., 2023). More importantly, our whole chloroplast genome sequences revealed that there were no differences between the putative *S. canadensis* and *S. altissima* accessions, and between their native and non-native origins (Figures 2 and 3).

Solidago canadensis has spread worldwide owing to its high competitive ability, which allows it to dominate plant communities (Likhanov et al., 2021). *Solidago altissima* has also been reported as invasive in different parts of the world (Weber and Schmid, 1998). The taxonomic status of *S. canadensis* and *S. altissima* is still controversial, as it is not clear yet whether they are two separate species or belong to a single species or species complex (Croat, 1972;

Weber, 2000; Semple et al., 2015). The main morphological traits that have been used to distinguish *S. canadensis* and *S. altissima* are the height and width of inflorescences (Croat, 1972), the numbers of disk and ray florets (Weber, 1997), pollen size (Melville and Morton, 1982) and the length and arrangement of the epidermal hairs of the stems and leaves (Szymura and Szymura, 2016). However, some studies showed that *S. canadensis* and *S. altissima* do not differ in functional traits, biomass production and allocation (Szymura and Szymura, 2013, 2015; Szymura et al., 2015). In addition, Tian et al. (2023) based on both molecular methods and morphological characteristics concluded that *S. canadensis* and *S. altissima* should be considered as a single species complex (Tian et al., 2023). As pointed out by others, next generation sequencing may help to reconstruct the phylogenetic relationships of *Solidago* species (Peirson et al., 2013; Shota et al., 2018; Semple et al., 2023). Indeed sequence data of chloroplast genomes and nuclear genomes have served as the technology of choice for phylogenetic investigations (Peirson et al., 2013; Guo et al., 2021). Similar to what earlier studies implied (Tian et al., 2023), our analyses of whole-chloroplast genome sequences and genome-wide SNP data (Figures 2 and 3) showed *S. canadensis* and *S. altissima* formed one monophyletic group as sister to *S. decurrens*. This supports the conclusion that *S. canadensis* and *S. altissima* belong to one species complex (i.e. the *S. canadensis* L. complex).

Accessions from China appeared in two different clades and one of these clades also included the accession from Japan and Europe (Figure 2). This is in line with the idea that there have been multiple introduction events of *S. canadensis* from its native North America range into China. The morphological analysis of Tian et al. (2023) based on 11 phenotypic traits also indicated that both the Chinese and European individuals in their study showed close affinity with North America individuals of *S. canadensis* L. complex. The

Chinese accessions may have been introduced directly from North America or they may have used Japan or Europe as stepping stones (Dong et al., 2006; Yang et al., 2011). In any case, the results indicate that the *S. canadensis* L. complex came to China via multiple routes, as has also been suggested by other studies (Huang and Guo, 2004; Lin et al., 2023). For Europe, historical records indicate that *S. canadensis* was introduced as an ornamental plant in the 17th century (Loudon, 1850). This was several centuries before the first plants were introduced to China (in the 1930s) (Xu et al., 2012; Zhao et al., 2015). Therefore, it is likely that the invasive accessions of *S. canadensis* in Europe came directly from North America and were not introduced via China or Japan. So, despite some limitations regarding the number of studied individuals, our results still provide new insights into the complicated invasion history of the *S. canadensis* L. complex in Europe and Asia.

DNA barcode markers, which are usually located in highly variable genome regions, are widely used for species identification. However, as some of the classical DNA barcodes are insufficient to distinguish between different *Solidago* species, it is crucial to find more highly variable genome regions that could be developed as potential barcode markers for identification of *Solidago* species. We found that most variable regions were located in the LSC and SSC regions, and that non-coding regions were more variable than coding regions (Supplementary Figure S2). This is a common phenomenon in the chloroplast genomes of most angiosperms (Li D.-M. et al., 2021). In this study, five highly variable chloroplast-genome regions that are located in the intergenic region of *S. canadensis* L. complex and *S. decurrens* were identified based on mVISTA and nucleotide diversity. These regions include *trnG-UCC* - *psbD*, *trnT-UGG* - *trnL-UAA*, *petB* - *petD*, *ycf1* - *ndhA*, and *trnL-UAG* - *ndhF* (Figure 4), and could serve as potential DNA barcodes for identification and phylogenetic analysis of *Solidago* species. The most divergent regions of *ycf1* - *ndhA*, *trnL-UAG* - *ndhF*, *trnT-UGU* - *trnL-UAA*, and *trnG-UCC* - *psbD* as shown in Figure 4 were consistent with results of previous studies (Li C. et al., 2021; Xia et al., 2022), indicating that these regions indeed evolve rapidly in *Solidago*.

Recently, *S. canadensis* has been added to the updated list of 33 key invasive alien species in China (http://www.moa.gov.cn/govpublic/KJJYS/202211/t20221109_6415160.htm; accessed on 1 March 2023). This indicates that the management and effective control of this species remains of pressing concern in China. Our study adds further evidence (see also Lin et al., 2023; Dong et al., 2016) that there may have been multiple introduction events of *S. canadensis* into China, and possibly via multiple pathways. To prevent invasions by more *S. canadensis* accessions, either from North America or other invaded regions, border quarantine measures should be strengthened. This does not only apply to *S. canadensis* but also to other alien species that may adversely affect local biodiversity. In addition, chemical, mechanical, and biological control approaches should be integrated to minimize the harm of *S. canadensis* to the local ecological environment (Wu and Qiang, 2005; Ma and Qiang, 2007; Jiang et al., 2008). Furthermore, with the application of next generation sequencing, the use of species-specific markers will be a very useful fast and cheap molecular tool for early detection of new alien species or accessions of alien

species that are already invasive (Ardura, 2019). The molecular markers detected by our study may help to establish a rapid and effective early warning monitoring system to promptly detect *S. canadensis* so that it can be eradicated before it becomes invasive.

Positive selection is considered to play an important role in adaptive evolution, while negative (purifying) selection is a ubiquitous evolutionary force responsible for conservation of the genome across long evolutionary timescales (Cvijovic et al., 2018; Moseley et al., 2018). To detect signs of adaptive evolution among species, the ratio of substitution rates at synonymous and non-synonymous sites is frequently used (Kryazhimskiy and Plotkin, 2008; Williams et al., 2020). For our *Solidago* accessions, we identified 10 genes with sites that showed signals of positive selection. Among them, *rpoB*, *rbcl*, *accD*, *psbB*, *rps3*, *rps19*, *ndhH*, *ndhD*, and *ccsA* genes have also been found in other plant species (Li C. et al., 2021; Wen et al., 2021; Tao et al., 2023). Moreover, these genes could be used for identification and phylogenetic research of *Solidago* species. The gene *ycf1* had numerous such sites, indicating that this gene may play a key role in adaptive evolution of *Solidago* (Supplementary Figure S4 and Supplementary Table S4). Furthermore, the genes *rps3* and *rps19*, encoding ribosomal subunit proteins that are considered to be essential for chloroplast biogenesis and function, showed signs of positive selection. This suggests that the studied *Solidago* species may increase evolutionary adaptability by regulating encoding ribosomal subunit proteins in chloroplast genomes. In addition, the likelihood ratio test (LRT) results showed that the p-values of *rpoB*, *rps19*, and *ndhD* genes were below 0.05, especially for the *rpoB* (encoding the beta-subunit of RNA polymerase) and *ndhD* gene (encoding NADP dehydrogenase), corroborating that these sites in the *S. canadensis* L. complex species have been under positive selection. As an important modulator of photosynthetic electron transport, recent study has revealed that positive selection of the *ndhD* gene was fairly common in all the main lineages of land plants (Wen et al., 2021). Moreover, the positive selection site of the *ndhD* gene in five *S. canadensis* L. complex individuals from China (*Solidago canadensis* 04 and 06) or three from North America (*Solidago altissima* 10, 13, and 14) contain same amino acid, indicating the Chinese individuals of the *S. canadensis* 04 and 06 may be directly introduced from North America. For the *rpoB* gene, *Solidago canadensis* 17 and 19 collected from Europe showed same amino acid in the 569th positive selection site, inferring the direct introduction route from North America to Europe. In general, positive selection would possibly contribute to various invasive environments for this introduced species. Therefore, positive selection of these chloroplast genes may promote environmental adaptation of *Solidago*, but unravelling details of the adaptation mechanisms will need further in-depth research.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Author contributions

Y-TT: Writing – original draft, Formal analysis, Writing – review & editing. L-XC: Data curation, Writing – review & editing. MJ: Data curation, Writing – review & editing. JJ: Methodology, Writing – review & editing. Z-SS: Data curation, Writing – review & editing. C-NC: Writing – review & editing. H-YL: Data curation, Writing – review & editing. AK: Data curation, Writing – review & editing. J-ML: Conceptualization, Writing – review & editing. MK: Conceptualization, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the National Key R&D Program of China (2016YFC1201100), Ten Thousand Talent Program of Zhejiang Province (No. 2019R52043) and the Taizhou Science and Technology Project (No. 21ywb76).

Acknowledgments

We thank the Taizhou Bigdata AI Research Center for providing computing resources.

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.

References

- Ardura, A. (2019). Species-specific markers for early detection of marine invertebrate invaders through eDNA methods: Gaps and priorities in GenBank as database example. *J. Nat. Conserv.* 47, 51–57. doi: 10.1016/j.jnc.2018.11.005
- Ashworth, V. (2017). Revisiting phylogenetic relationships in Phoradendreae (Viscaceae): utility of the *trnL-F* region of chloroplast DNA and presence of a homoplasious inversion in the intergenic spacer1. *Botany* 95, 247–258. doi: 10.1139/cjb-2016-0241
- Beaudry, J. R., and Chabot, D. L. (1957). Studies on *Solidago* L. I. *S. altissima* L. and *S. canadensis* L. *Contr. Inst. Bot. Univ. Montréal* 70, 65–72.
- Beck, J. B., and Semple, J. C. (2015). Next-generation sampling: pairing genomics with herbarium specimens provides species-level signal in *Solidago* (Asteraceae). *Appl. Plant Sci.* 3, 1500014. doi: 10.3732/app.1500014
- 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
- Croat, T. (1972). *Solidago canadensis* complex of the great plains. *Brittonia* 24, 317–326. doi: 10.2307/2805669
- Cvijovic, I., Good, B. H., and Desai, M. M. (2018). The effect of strong purifying selection on genetic diversity. *Genetics* 209, 1235–1278. doi: 10.1534/genetics.118.301058
- Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., et al. (2021). Twelve years of SAMtools and BCFtools. *Gigascience* 10, giab008. doi: 10.1093/gigascience/giab008
- Dierckxsens, N., Mardulyn, P., and Smits, G. (2017). NOVOPlasty: *de novo* assembly of organelle genomes from whole genome data. *Nucleic Acids Res.* 45, e18. doi: 10.1093/nar/gkw955
- Dong, M., Lu, B.-R., Zhang, H.-B., Chen, J.-K., and Li, B. (2006). Role of sexual reproduction in the spread of an invasive clonal plant *Solidago canadensis* revealed using intersimple sequence repeat markers. *Plant Spec. Biol.* 21, 13–18. doi: 10.1111/j.1442-1984.2006.00146.x
- Dong, W., Xu, C., Li, D., Jin, X., Li, R., Liu, Q., et al. (2016). Comparative analysis of the complete chloroplast genome sequences in psammophytic *Haloxylon* species (Amaranthaceae). *PeerJ* 4, e2699. doi: 10.7717/peerj.2699
- Doyle, J. J., and Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* 19, 11–15.
- Fernald, M. L. (1950). *Gray's manual of botany. 8th edition* (New York: American Book Co).
- Frazer, K. A., Pachter, L., Poliakov, A., Rubin, E. M., and Dubchak, I. (2004). VISTA: computational tools for comparative genomics. *Nucleic Acids Res.* 32, W273–W279. doi: 10.1093/nar/gkh458
- Gao, F., Chen, C., Arab, D. A., Du, Z., He, Y., and Ho, S. Y. W. (2019). EasyCodeML: a visual tool for analysis of selection using CodeML. *Ecol. Evol.* 9, 3891–3898. doi: 10.1002/ece3.2019.9.issue-7
- Greiner, S., Lehwark, P., and Bock, R. (2019). OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res.* 47, W59–W64. doi: 10.1093/nar/gkz238
- Guo, Y.-Y., Yang, J.-X., Bai, M.-Z., Zhang, G.-Q., and Liu, Z.-J. (2021). The chloroplast genome evolution of Venus slipper (*Paphiopedilum*): IR expansion, SSC contraction, and highly rearranged SSC regions. *BMC Plant Biol.* 21, 248. doi: 10.1186/s12870-021-03053-y
- Hereward, J. P., Werth, J. A., Thornby, D. F., Keenan, M., Chauhan, B. S., and Walter, G. H. (2017). Complete chloroplast genome of glyphosate resistant *Conyza bonariensis* (L.) Cronquist from Australia. *Mitochondrial DNA Part B* 2, 444–445. doi: 10.1080/23802359.2017.1357441

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/fpls.2024.1498543/full#supplementary-material>

SUPPLEMENTARY FIGURE 1

Solidago canadensis in the field (A), an herbarium specimen of *S. altissima* (B), and a potted plant of *S. decurrens* (C).

SUPPLEMENTARY FIGURE 2

Sequence alignment of chloroplast genomes with accession *Solidago canadensis* 21 as a reference. The y-scale indicates sequence identity from 50% to 100%. Gray arrows indicate the positions and directions of each gene. Red indicates non-coding sequences (CNS); Purple indicates the exons of protein-coding genes (exon); lime green indicates tRNAs and rRNAs.

SUPPLEMENTARY FIGURE 3

Nucleotide diversity analysis between Chinese and European/North American individuals of the *S. canadensis* L. complex.

SUPPLEMENTARY FIGURE 4

Partial alignment of three of the 10 positively selected genes with posterior probabilities higher than 0.95. (A–C) Partial aligned amino acid sequences of the *rpoB*, *rps19*, and *ndhD* genes, respectively. The red blocks represent amino acids with a high posterior probability.

- Hu, X. (2020). Characterization of the complete chloroplast genome of *Aster subulatus* Michx. *Mitochondrial DNA Part B* 5, 597–598. doi: 10.1080/23802359.2019.1710599
- Huang, H., and Guo, S. (2004). Review on ecological studies on three invasive species of European genus *Solidago*. *Guangxi Sci.* 11, 69–74.
- Jiang, D., Cai, X., Gong, M., Xia, M., Xing, H., Dong, S., et al. (2023). Complete chloroplast genomes provide insights into evolution and phylogeny of *Zingiber* (Zingiberaceae). *BMC Genomics* 24, 30. doi: 10.1186/s12864-023-09115-9
- Jiang, H., Fang, F., and Suo, S. (2008). Influences of parasitism by *Cuscuta japonica* plants on eco-physiological characteristics of *Solidago canadensis*. *Acta Ecol. Sin.* 1, 399–406.
- Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010
- Kryazhimskiy, S., and Plotkin, J. (2008). The population genetics of dN/dS. *PLoS Genet.* 4, e1000304. doi: 10.1371/journal.pgen.1000304
- Li, C., Cai, C., Tao, Y., Sun, Z., Jiang, M., Chen, L., et al. (2021). Variation and evolution of the whole chloroplast genomes of *Fragaria* spp. (Rosaceae). *Front. Plant Sci.* 12, 754209. doi: 10.3389/fpls.2021.754209
- Li, D.-M., Li, J., Wang, D.-R., Xu, Y.-C., and Zhu, G.-F. (2021). Molecular evolution of chloroplast genomes in subfamily Zingiberoideae (Zingiberaceae). *BMC Plant Biol.* 21, 558. doi: 10.1186/s12870-021-03315-9
- 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
- Librado, P., and Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25, 1451–1452. doi: 10.1093/bioinformatics/btp187
- Likhanov, A., Oliynyk, M., Pashkevych, N., Churilov, A., and Kozyr, M. (2021). The role of flavonoids in invasion strategy of *Solidago canadensis* L. *Plants* 10, 1748. doi: 10.3390/plants10081748
- Lin, H., Chen, L., and Li, J. (2023). Multiple introductions and distinct genetic groups of Canada goldenrod (*Solidago canadensis*) in China revealed by genomic single-nucleotide polymorphisms. *Plants* 12, 1734. doi: 10.3390/plants12091734
- Loudon, J. C. (1850). *An Encyclopaedia of gardening*. 2nd edition. Longman publishing.
- Ma, L., and Qiang, S. (2007). Identification and classification of cut-flower Huangyinghua using DNA marker techniques. *Acat Phytotaxonomica Sin.* 45, 497–504. doi: 10.1360/aps06135
- McNeil, J. (1976). *Flora Europaea*. Eds. L. Solidago, T. G. Tutin, V. H. Heywood, N. A. Burges, D. H. Valentine, S. M. Walters and D. A. Webb (Cambridge: Cambridge University Press).
- Melville, M. R., and Morton, J. K. (1982). A biosystematic study of the *Solidago canadensis* (Compositae) complex. I. The Ontario populations. *Can. J. Bot.* 60, 976–997. doi: 10.1139/b82-123
- Moseley, R. C., Mewalal, R., Motta, F., Tuskan, G. A., Haase, S., and Yang, X. (2018). Conservation and diversification of circadian rhythmicity between a model crassulacean acid metabolism plant *Kalanchoë fedtschenkoi* and a model C3 photosynthesis plant *Arabidopsis thaliana*. *Front. Plant Sci.* 9, 1757. doi: 10.3389/fpls.2018.01757
- Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300
- Peirson, J. A., Dick, C. W., and Reznicek, A. A. (2013). Phylogeography and polyploid evolution of North American goldenrods (*Solidago* subsect. *Humiles*, Asteraceae). *J. Biogeogr.* 40, 1887–1898. doi: 10.1111/jbi.2013.40.issue-10
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795
- Semple, J. C., and Cook, R. E. (2006). “*Solidago linnaeus*,” (Oxford Univ. Press, New York), 107–166.
- Semple, L. C., McMinn-Sauder, H., Stover, M., Lemmon, A. R., Lemmon, E. M., and Beck, J. B. (2023). Goldenrod herbariomics: Hybrid-sequence capture reveals the phylogeny of *Solidago*. *Am. J. Bot.* 110, e16164. doi: 10.1002/ajb2.v110.7
- Semple, J. C., Rahman, H., Bzovsky, S., Sorour, M. K., Komobis, K., Lopez Laphitz, R., et al. (2015). A multivariate morphometric study of the *Solidago altissima* complex and *S. canadensis* (Asteraceae: Astereae). *Phytoneuron* 2015, 1–31.
- Shota, S., Takuma, K., Ryuta, K., Masayuki, M., Takako, N., Naoko, I., et al. (2018). Phylogeographic analysis of the East Asian goldenrod (*Solidago virgaurea* complex, Asteraceae) reveals hidden ecological diversification with recurrent formation of ecotypes. *Ann. Bot.* 121, 489–500. doi: 10.1093/aob/mcx182
- Szymura, M., and Szymura, T. H. (2013). Soil preferences and morphological diversity of goldenrods (*Solidago* L.) from south-western Poland. *Acta Soc. Bot. Pol.* 82, 107–115. doi: 10.5586/asbp.2013.005
- Szymura, M., and Szymura, T. H. (2015). Growth, phenology, and biomass allocation of alien *Solidago* species in central Europe. *Plant Spec. Biol.* 30, 245–256. doi: 10.1111/psbi.2015.30.issue-4
- Szymura, M., and Szymura, T. H. (2016). Historical contingency and spatial processes rather than ecological niche differentiation explain the distribution of invasive goldenrods (*Solidago* and *Euthamia*). *Plant Ecol.* 217, 565–582. doi: 10.1007/s11258-016-0601-1
- Szymura, M., Szymura, T. H., and Kreitschitz, A. (2015). Morphological and cytological diversity of goldenrods (*Solidago* L. and *Euthamia* Nutt.) from south-western Poland. *Biodiv. Res. Conserv.* 38, 41–49. doi: 10.1515/biorc-2015-0010
- Tao, L., Duan, H., Tao, K., Luo, Y., Li, Q., and Li, L. (2023). Complete chloroplast genome structural characterization of two *Phalaenopsis* (Orchidaceae) species and comparative analysis with their alliance. *BMC Genomics* 24, 359. doi: 10.1186/s12864-023-09448-5
- Tian, Z., Cheng, J., Xu, J., Feng, D., Zhong, J., Yuan, X., et al. (2023). Cytogeography of naturalized *Solidago canadensis* populations in Europe. *Plants* 12, 1113. doi: 10.3390/plants12051113
- Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, E. S., Fischer, A., Bock, R., et al. (2017). GeSeq – versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* 45, W6–W11. doi: 10.1093/nar/gkx391
- Van der Auwera, G., Carneiro, M. O., Hartl, C., Poplin, R., Angel, G. D., Levy-Moonshine, A., et al. (2013). From fastq data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. *Curr. Protoc. Bioinf.* 43, 11.10.1–11.10.33. doi: 10.1002/0471250953.2013.43.issue-1
- Van Kleunen, M., Pyšek, P., Dawson, W., Essl, F., Kreft, H., Pergl, J., et al. (2019). The global naturalized alien flora (Glo NAF) database. *Ecology* 100, e02542. doi: 10.1002/ecy.2542
- Van Kleunen, M., and Schmid, B. (2003). No evidence for an evolutionary increased competitive ability in an invasive plant. *Ecology* 84, 2816–2823. doi: 10.1890/02-0494
- Vargas, O. M., Ortiz, E. M., and Simpson, B. B. (2017). Conflicting phylogenomic signals reveal a pattern of reticulate evolution in a recent high-Andean diversification (Asteraceae: Astereae: *Diplostephium*). *New Phytol.* 214, 1736–1750. doi: 10.1111/nph.2017.214.issue-4
- Verloove, F., Zonneveld, B. J. M., and Semple, J. C. (2017). First evidence for the presence of invasive *Solidago altissima* (Asteraceae) in Europe. *Willdenowia* 47, 69–75. doi: 10.3372/wi.47.47107
- Wang, X., Bai, S., Zhang, Z., Zheng, F., Song, L., Wen, L., et al. (2023). Comparative analysis of chloroplast genomes of 29 tomato germplasms: genome structures, phylogenetic relationships, and adaptive evolution. *Front. Plant Sci.* 14, 1179009. doi: 10.3389/fpls.2023.1179009
- Weber, E. (1997). Morphological variation of the introduced perennial *Solidago canadensis* L. sensu lato (Asteraceae) in Europe. *Bot. J. Linn. Soc.* 123, 197–210. doi: 10.1111/j.1095-8339.1997.tb01413.x
- Weber, E. (2000). Biological flora of central Europe: *Solidago altissima* L. *Flora* 195, 123–134. doi: 10.1016/S0367-2530(17)30960-X
- Weber, E. (2017). *Invasive plant species of the world: a reference guide to environmental weeds*. 2nd ed. (Wallingford: CABI publishing).
- Weber, E., and Schmid, B. (1998). Latitudinal population differentiation in two species of *Solidago* (Asteraceae) introduced into Europe. *Am. J. Bot.* 85, 1110. doi: 10.2307/2446344
- Wen, F., Wu, X., Li, T., Jia, M., Liu, X., and Liao, L. (2021). The complete chloroplast genome of *Stauntonia chinensis* and compared analysis revealed adaptive evolution of subfamily Lardizabaloidae species in China. *BMC Genomics* 22, 161. doi: 10.1186/s12864-021-07484-7
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis* (Switzerland Springer: Springer-Verlag New York).
- Williams, M., Zapata, L., Werner, B., Barnes, C., Sottoriva, A., and Graham, T. (2020). Measuring the distribution of fitness effects in somatic evolution by combining clonal dynamics with dN/dS ratios. *Elife* 9, e48714. doi: 10.7554/eLife.48714
- Wu, H., and Qiang, S. (2005). Bio-ecological characteristics of *Solidago canadensis* L. and its control. *Weed Sci.* 1, 52–56.
- Xia, X., Peng, J., Yang, L., Zhao, X., Duan, A., and Wang, D. (2022). Comparative analysis of the complete chloroplast genomes of eight *Ficus* species and insights into the phylogenetic relationships of *Ficus*. *Life* 12, 848. doi: 10.3390/life12060848
- Xu, H., Qiang, S., Genovesi, P., Ding, H., Wu, J., Meng, L., et al. (2012). An inventory of invasive alien species in China. *NeoBiota* 15, 1–26. doi: 10.3897/neobiota.15.3575
- Yang, Z., and Nielsen, R. (2002). Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol. Biol. Evol.* 19, 908–917. doi: 10.1093/oxfordjournals.molbev.a004148
- Yang, Z., Wong, W. S. W., and Nielsen, R. (2005). Bayes empirical bayes inference of amino acid sites under positive selection. *Mol. Biol. Evol.* 22, 1107–1118. doi: 10.1093/molbev/msi097
- Yang, R., Zan, S., Tang, J., and Chen, X. (2011). Invasion mechanisms of *Solidago canadensis* L.: a review. *Acta Ecol. Sin.* 31, 1184–1195.
- Zhang, M., and Chen, N. (2022). Comparative analysis of *Thalassionema* chloroplast genomes revealed hidden biodiversity. *BMC Genomics* 23, 327. doi: 10.1186/s12864-022-08532-6
- Zhao, S. Y., Sun, S. G., Dai, C., Gituru, R. W., Chen, J. M., and Wang, Q. F. (2015). Genetic variation and structure in native and invasive *Solidago canadensis* populations. *Weed Res.* 55, 163–172. doi: 10.1111/wre.2015.55.issue-2
- Zhou, H., She, R., Xu, Y., Zhao, P., and Zhang, S. (2019). Characterization of the complete chloroplast genome of sunflower family species *Aster Flaccidus* (Compositae). *Mitochondrial DNA Part B* 4, 4096–4097. doi: 10.1080/23802359.2019.1692710



OPEN ACCESS

EDITED BY

Marcial Escudero,
Sevilla University, Spain

REVIEWED BY

Tzvetanka D. Dinkova,
National Autonomous
University of Mexico, Mexico
Yiliang Ding,
John Innes Centre, United Kingdom
Marcus Lechner,
Philipps-University Marburg, Germany

*CORRESPONDENCE

Jose Planta

✉ jgplanta@up.edu.ph

RECEIVED 26 August 2024

ACCEPTED 02 December 2024

PUBLISHED 23 December 2024

CITATION

Monloy KC and Planta J (2024) tRNA gene
content, structure, and organization in the
flowering plant lineage.
Front. Plant Sci. 15:1486612.
doi: 10.3389/fpls.2024.1486612

COPYRIGHT

© 2024 Monloy and Planta. This is an open-
access article distributed under the terms of
the [Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). 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.

tRNA gene content, structure, and organization in the flowering plant lineage

Kim Carlo Monloy and Jose Planta*

National Institute of Molecular Biology and Biotechnology, College of Science, University of the
Philippines Diliman, Quezon City, Philippines

Transfer RNAs (tRNAs) are noncoding RNAs involved in protein biosynthesis and have noncanonical roles in cellular metabolism, such as RNA silencing and the generation of transposable elements. Extensive tRNA gene duplications, modifications to mature tRNAs, and complex secondary and tertiary structures impede tRNA sequencing. As such, a comparative genomic analysis of complete tRNA sets is an alternative to understanding the evolutionary processes that gave rise to the extant tRNA sets. Although the tRNA gene (tDNA) structure and distribution in prokaryotes and eukaryotes, specifically in vertebrates, yeasts, and flies, are well understood, there is little information regarding plants. A detailed and comprehensive analysis and annotation of tDNAs from the genomes of 44 eudicots, 20 monocots, and five other non-eudicot and non-monocot species belonging to the Ceratophyllaceae and the ANA (Amborellales, Nymphaeales, and Austrobaileyales) clade will provide a global picture of plant tDNA structure and organization. Plant genomes exhibit varying numbers of nuclear tDNAs, with only the monocots showing a strong correlation between nuclear tDNA numbers and genome sizes. In contrast, organellar tDNA numbers varied little among the different lineages. A high degree of tDNA duplication in eudicots was detected, whereby most eudicot nuclear genomes (91%) and only a modest percentage of monocot (65%) and ANA nuclear genomes (25%) contained at least one tDNA cluster. Clusters of tRNA^{Tyr}–tRNA^{Ser} and tRNA^{Ile} genes were found in eudicot and monocot genomes, respectively, while both eudicot and monocot genomes showed clusters of tRNA^{Pro} genes. All plant genomes had intron-containing tRNA^{Met} and tRNA^{Tyr} genes with modest sequence conservation and a strictly conserved tRNA^{Ala-AGC} species. Regulatory elements found upstream (TATA-box and CAA motifs) and downstream (poly(T) signals) of the tDNAs were present in only a fraction of the detected tDNAs. A and B boxes within the tDNA coding region show varying consensus sequences depending on the tRNA isotype and lineage. The chloroplast genomes, but not the mitogenomes, possess relatively conserved tRNA gene organization. These findings reveal differences and patterns acquired by plant genomes throughout evolution and can serve as a foundation for further studies on plant tRNA gene function and regulation.

KEYWORDS

tRNA genes, tDNA, tRNA gene content, tRNA gene organization, tRNA gene structure

1 Introduction

Transfer RNAs (tRNAs) are short, noncoding molecules acting as intermediaries between the genetic information in nucleic acids and protein sequences. Although the mechanistic roles of tRNAs in ribosomal protein biosynthesis are well understood, they have noncanonical functions in several aspects of cellular metabolism. Plant tRNAs have been implicated in tetrapyrrole and cytokinin biosynthesis (Chery and Drouard, 2022), plant cell growth and immunity (Soprano et al., 2018), and regulation of auxin response in *Arabidopsis* (Leitner et al., 2015). Increased attention has also been given to tRNA-derived fragments (tRFs), a class of small RNAs produced from the enzymatic cleavage of tRNAs. Initially thought as mere tRNA degradation byproducts, tRFs have been linked to gene regulation, ribosome biogenesis, plant–pathogen interactions, and stress response in plants (Park and Kim, 2018; Alves and Nogueira, 2021; Wang et al., 2023; Panstruga and Spanu, 2024). tRFs have been reported to be involved in the RNA silencing pathway and are the major source of the transposable element SINEs (short interspersed nuclear elements; Bermudez-Santana et al., 2010; Phizicky and Hopper, 2010; Soprano et al., 2018). All tRNA genes are postulated to be derived from an ancestral “proto-tRNA” (Eigen et al., 1989), and during evolution, a tRNA repertoire was generated from gene duplication and numerous mutational events. These processes gave rise to the core and dispensable sets of tRNA genes.

Despite the growing knowledge and interest in plant tRNA biology, studies on how tRNAs are structured and organized on a genome-wide scale in plants still number too few. A survey of the content, distribution, and clustering of tRNA genes and pseudogenes in many eukaryotes, including nine genomes from the green lineage, has been reported (Bermudez-Santana et al., 2010). More recent studies have also reported the evolution of tRNA gene content in the three domains of life, involving 13 plant genomes (Santos and Del-Bem, 2023), as well as the tRNA anticodon frequency of 128 plant genomes (Mohanta et al., 2020). Databases of tRNA gene sets from hundreds of plant nuclear and organellar genomes, covering diverse families of plants, have also been developed (e.g., Cognat et al., 2022; Mokhtar and Allali, 2022), whose curators were able to provide a general survey of the tRNA gene populations of 51 and 256 plant species, respectively. However, these mostly only provided insights on the tRNA gene content of these plants, and separate studies fully utilizing the information from these databases are yet to be found. To date, the first comprehensive study that focused on tRNA gene content, structure, and distribution in plants covered both the nuclear and organellar genomes of only five angiosperms—consisting of three eudicots and two monocots—and one green alga (Michaud et al., 2011). However, given the species diversity within the flowering plants, a more comprehensive and systematic comparative study is needed to provide a global landscape of plant tRNA structure and organization. The increased availability of plant genomes will provide common patterns and taxon-specific particularities of plant tRNAs.

Compared to other eukaryotic genomes, plant genomes possess a smaller variation in the number of tRNA genes and a varying abundance of tRNA gene clusters (Bermudez-Santana et al., 2010).

The following tRNA gene organization has been reported among flowering plant genomes: a predominantly A-/T-rich region spanning 50 nucleotides upstream of the tRNA gene, an upstream CAA motif and a downstream poly(T) termination signal found in most tRNAs, and intron-containing tRNA^{Met} and tRNA^{Tyr} genes (Michaud et al., 2011). Except for *Arabidopsis*, a similar chromosomal distribution of tRNA genes in terms of the numbers of tRNA genes per megabase of the chromosome was also reported within angiosperms, which hinted at the possibility of excessive tRNA gene duplications in some plant genomes (Michaud et al., 2011). Although a significant correlation between genome size and number of tRNA genes have been reported among 74 eukaryotic genomes (Bermudez-Santana et al., 2010), five plant genomes (Michaud et al., 2011), and eight monocot genomes (Planta et al., 2022), a more recent regression analysis involving a higher coverage of plant genomes (128 genomes) instead reported a weak correlation (Mohanta et al., 2020).

In the case of organellar genomes, previous studies also reported the lack of certain tRNA isoacceptors in some plant plastomes and mitogenomes (Michaud et al., 2011; Mohanta et al., 2020). Although possessing significantly fewer tRNA genes than the nuclear genome, the organellar genomes from mitochondria and chloroplasts can also encode their tRNAs. The chloroplast genome is assumed to encode all the tRNA species required for protein synthesis, and unlike the mitochondria, chloroplasts do not import cytosolic tRNAs (Maréchal-Drouard et al., 1993). A relaxed wobble rule might also explain the small number of organellar tRNAs that can read all codons of the universal genetic code (Crick, 1966; Percudani, 2001).

Several different sequencing-based approaches have been developed to quantify highly modified tRNAs. However, modifications on tRNAs can impair cDNA synthesis by premature reverse transcriptase (RT) stops (Pinkard et al., 2020; Padhiar et al., 2024). These methods [e.g., ARM-seq (Cozen et al., 2015), DM-TGIRT-seq (Zheng et al., 2015), YAMAT-seq (Shigematsu et al., 2017), Nano-tRNAseq (Lucas et al., 2024); see Padhiar et al. (2024) for a comprehensive review] incorporate pre-treatment of RNA before library construction and the use of modified adapters; pre-treatment of RNA produces less complex secondary structures and fewer modifications that can lead to premature RT stops (Padhiar et al., 2024). Plant tRNA expression and post-transcriptional modifications have been characterized in *Arabidopsis thaliana* by modifying RNA-seq to involve a demethylating enzyme and using a tRNA-specific adapter (Shigematsu et al., 2017; Warren et al., 2021). While these are promising advancements in direct tRNA sequencing, at its current state, computationally predicting tRNA genes from whole-genome sequencing data is still the preferred method in most tRNA gene studies (Chan et al., 2021).

This study compared and analyzed the tRNA gene content, structure, and organization of 69 nuclear plant genomes—including available chloroplast and mitochondrial genome counterparts (Supplementary Table 1). Included in our analyses are 44 genomes from the eudicot lineage, 20 from the monocot lineage, four from the ANA clade (Amborellales, Nymphaeales, and Austrobaileyales), and one from Ceratophyllaceae, the sister clade to eudicots. The eudicot and monocot genomes were chosen to cover as much family in the flowering plant lineage; the chosen

plant genomes span 32 families—two from the ANA clade (Amborellaceae and Nymphaeaceae), Ceratophyllaceae, nine from monocots, and 20 from dicots (Supplementary Figure 1). Having these lineages within the scope of this study should provide a better and more inclusive analysis of tRNA genes in plants. Using the widely adopted tool tRNAscan-SE (Chan et al., 2021), tRNA genes from these genomes were computationally predicted and then filtered for a “high-confidence” set of tRNA genes discarding pseudogenes. To characterize these “high-confidence” tRNA genes, we also screened the tDNAs for regulatory sequences commonly associated with the RNA polymerase III-transcribed plant tRNA genes: the upstream TATA-box and CAA motifs (Choisne et al., 1998; Yukawa et al., 2000; Dieci et al., 2006; Michaud et al., 2011), the intragenic A and B boxes (Choisne et al., 1998; Dieci et al., 2006), and the downstream poly(T) stretches (Yukawa et al., 2000; Braglia et al., 2005; Arimbasseri and Maraia, 2015).

Comparative genomics analyses revealed that the number of nuclear tRNA genes varied mainly among the plant genomes studied, even among genomes of the same lineage. In contrast, the number of organellar tRNA genes had slight variation and was consistent regardless of plant lineage. Moreover, gene duplications in tRNA gene clusters appeared more prevalent in eudicots. All nuclear genomes were found to have a strictly conserved tRNA^{Ala-AGC} species and intron-containing tRNA^{Met} and tRNA^{Tyr} genes that exhibited modest sequence conservation. Regulatory sequences found in the nuclear tRNA genes include the upstream TATA-box and CAA motifs (found upstream of 22%–32% and 78%–82% of tRNA genes detected, respectively), the intragenic A and B boxes (found in all tRNA genes detected) with general lineage- and isotype-specific motifs, and the downstream poly(T) termination signals (found downstream of 67%–72% of tRNA genes detected). Overall, this study revealed differences and patterns acquired by plant genomes throughout evolution and can serve as a foundation for further studies on plant tRNA gene function and regulation.

2 Materials and methods

2.1 Phylogenetic tree construction

Nuclear and organellar genomes from 69 flowering plant species encompassing the ANA, Ceratophyllaceae, eudicot, and monocot lineages used in this study are listed in Supplementary Table 1 and were obtained either from Phytozome (Goodstein et al., 2012) or the NCBI database (Sayers et al., 2021). Our analyses focused on the basal angiosperms—the Amborellaceae and Nymphaeaceae families—20 eudicot families, Ceratophyllaceae, and nine monocot families (see Supplementary Figure 1; https://www.plabipd.de/pubplant_cladogram1.html). The nuclear genomes in our dataset also have at least an available organellar genome (chloroplast, mitochondrial, or both). To enhance our tRNA gene clustering analysis, we incorporated genomes with chromosome-scale assemblies from the ANA, eudicot, and monocot lineages.

A phylogenetic tree was constructed from concatenated *matK* and *rbcL* sequences of each genome (Supplementary Table 2)

obtained from the NCBI database (Sayers et al., 2021). Alignment and trimming were performed with MAFFT ver. 7.453 (default parameters; Katoh and Toh, 2008) and trimAI ver. 3-2021.11 (with “-strictplus” option; Capella-Gutiérrez et al., 2009), respectively, and the tree was generated using the IQ-TREE web server (Trifinopoulos et al., 2016). Default parameters were used for the IQ-TREE run. The constructed tree was viewed and edited using TreeGraph ver. 2.15.0-887 (Stöver and Müller, 2010) and FigTree ver. 1.4.4 (Rambaut, 2024).

2.2 tRNA gene detection in plant genomes and alignment of tRNA genes and introns

For nuclear genomes, tRNAscan-SE ver. 2.0.9 (with “-Hy” option) was used for the detection of tRNA genes, or tDNAs, and the primary results were parsed with the post-filtering tool EukHighConfidenceFilter (with “-r” option) of the tRNAscan-SE package listing the high-confidence sets of tDNAs most likely to be involved in ribosomal translation (Chan et al., 2021). To ensure only nuclear tDNAs are detected, we checked each nuclear genome FASTA file and manually removed chloroplast and mitochondrial sequences that were found. The number of high-confidence, intron-containing, and unique tDNA sequences were tabulated for each tRNA isoacceptor of each genome. The “-O” and “-Hy” options were used to detect tRNA genes from chloroplast and mitochondrial genomes. To visualize the overall tRNA gene content in our dataset, heatmaps were generated using the superheat R package (Barter and Yu, 2017). Linear regression analyses were also performed using the built-in *lm* function in R (R Core Team, 2021; ver. 4.4.2), which was based on the works of Chambers (1992) and Wilkinson and Rogers (1973). We considered p-values lower than 0.05 to be statistically significant.

All the nuclear genomes used for tRNA gene detection were found to have at least one intron-containing tRNA^{Met} and tRNA^{Tyr} gene. Intronic sequences of these tRNA isoacceptors (extracted using an in-house Perl script) were separately aligned for each of the eudicot, monocot, and ANA lineages to identify conserved nucleotide bases as well as similarities and differences between the consensus intronic sequences of each lineage. Alignment was performed using Multalin ver. 5.4.1 (Corpet, 1988) with the following parameters: “symbol comparison table—DNA-5-0,” “gap penalty at extremities—both,” and “one iteration only—no.” Alignments were then manually modified, if necessary, using AliView ver. 1.21 (Larsson, 2014). Sequence logo plots for the ANA, eudicot, and monocot tRNA^{Met} and tRNA^{Tyr} intronic sequences were then separately generated using WebLogo 3 (Crooks et al., 2004).

2.3 Analysis of tRNA gene regulatory elements and conservation of tRNA species

Sequences 50 and 300 bases immediately upstream and 50 bases immediately downstream of each tDNA sequence were extracted

from each genome with the toolkit TBTools (Chen et al., 2020). PlantCARE (Lescot et al., 2002), a database for cis-acting plant regulatory elements database, was utilized to search for TATA-box motifs in tDNA upstream sequences. Other regulatory elements, such as the upstream CAA triplet and the downstream poly(T) signals, were searched through command-line text manipulation. On the other hand, intragenic regulatory elements (A and B boxes) were manually extracted from the alignment of tRNA genes for each isoacceptor and lineage. Sequence logo plots showing upstream A/T content and intragenic A/B box motifs were generated using WebLogo 3 (Crooks et al., 2004).

Command-line BLASTn was used with default settings to compare the high-confidence tRNA gene set of *Amborella trichopoda* with the high-confidence tRNA gene sets of the rest of the nuclear genomes following the procedure of Tang et al. (2009). From this search, one tRNA^{Ala-AGC} species from *A. trichopoda* was found to be identical in the other 68 nuclear genomes, and the secondary sequence of this tDNA was visualized using the RNAfold web server (Institute for Theoretical Chemistry RNAfold web server). This discovery prompted us to investigate the secondary structure conservation of all nuclear tRNA^{Ala-AGC} sequences further using structural alignment and single covariation analysis. Consensus tRNA^{Ala-AGC} secondary structures for each lineage were generated using RNAalifold (Bernhart et al., 2008).

Following Tourasse and Darfeuille's (2020) procedure, structural alignment was performed with MAFFT ver. 7.511 (Katoh and Toh, 2008) in the X-INS-i mode. These structural alignments were then analyzed by single covariation analysis through the web-based version of R-chie (Lai et al., 2012). Before single covariation analysis, a reference secondary structure was generated for tRNA^{Ala-AGC} by uploading the tRNA^{Ala-AGC} sequence into the Mfold web server (Zuker, 2003). For eudicots, monocots, and ANA, the reference secondary structures are from *A. thaliana*, *O. sativa*, and *N. colorata*, respectively. With these reference secondary structures, a single covariation analysis was performed in R-chie by mapping the structures onto the alignments (Tourasse and Darfeuille, 2020). Results were visualized with arc diagrams (with colors representing the various covariation scores) superimposed on the corresponding multiple sequence alignments allowing for the simultaneous comparison of secondary structures and sequences (Lai et al., 2012).

2.4 Analysis of tRNA gene clustering

We considered tDNAs to be clustered if at least three tDNAs are within 1 kb of each other (a density of ≥ 3 tDNAs/kb). The “merge” function of BEDTools was used to obtain a list of clustered tDNAs (Quinlan and Hall, 2010). The BED files for each nuclear genome were created from their respective GFF3 files, which were generated by converting each EukHighConfidenceFilter output file to GFF3 format using an in-house Perl script. Long tDNA clusters with more than 10 repeated tRNA gene units were visualized using the ChromoMap R package (Anand and Rodriguez Lopez, 2022).

2.5 Inferring tRNA gene duplication and loss events

To infer and gain insights into what duplication or loss events may have transpired in certain tRNA isoacceptors throughout the evolution of flowering plants, Notung ver. 2.9.1.5 (Chen et al., 2000; Zmasek and Eddy, 2001; Durand et al., 2006; Vernot et al., 2007; Stolzer et al., 2012; Darby et al., 2017) was used. This inference was made in Notung by reconciling the manually prepared gene and species trees.

A separate gene tree was created for tRNA^{Pro}, tRNA^{Ile}, and tRNA^{Ala-AGC}. All tDNA sequences of the specific isoacceptor were aligned using the Clustal Omega server to create a gene tree (Madeira et al., 2022). After converting the generated ClustalW files into the MEGA format, a maximum likelihood tree was generated using the MEGA11 software (Tamura et al., 2021) with the following parameters: “test of phylogeny—bootstrap method,” “no. of bootstrap replications—100,” “model/method—Jukes–Cantor model,” “rates among sites—uniform rates,” “gaps/missing data treatment—partial deletion,” “site coverage cutoff (%)—95,” “ML heuristic method—Nearest-Neighbor-Interchange (NNI),” “initial tree for ML—make initial tree automatically,” and “branch swap filter—very strong.” These parameters were based on the protocol of Mohanta and Bae (2017). The species tree, on the other hand, was based on the phylogenetic tree made by Janssens et al. (2020). Plant genomes in this study that were missing in the said tree were manually added, the placements of which were based on the cladogram found in the Published Plant Genomes website (https://www.plabipd.de/plant_genomes_pa.ep; Usadel Lab Published plant genomes). These trees were labeled and rerooted via the phylogenetic tree viewer FigTree ver. 1.4.4 (Rambaut, 2024).

2.6 Analysis of organellar tRNA genes

To visualize the tRNA gene organization in chloroplast and mitochondrial genomes, gene maps were created using the online tool MG2C ver. 2.1 (Chao et al., 2021). The BED file outputs of tRNAscan-SE were used to determine the tRNA gene locations in the respective organellar genome.

3 Results

Plants with sequenced chloroplast, mitochondrial, or nuclear genomes (Supplementary Table 1) were used for the comparative analysis of tRNA gene content, structure, and organization. *Aquilegia coerulea* and *Acorus americanus* were included in the analysis as these are members of the basal-most eudicot clade and the sister lineage to all other monocots (Filiault et al., 2018; Givnish et al., 2018), respectively. *Amborella trichopoda*, *Nymphaea colorata*, *Nymphaea thermarum*, and *Euryale ferox* under the ANA (Amborellales, Nymphaeales, and Austrobaileyales) clade are sisters to all other angiosperms. *Ceratophyllum demersum* belongs to the species-poor lineage of Ceratophyllales and is sister

to eudicots (Yang et al., 2020). Given the phylogenetic positions of these species (Supplementary Figure 1), including these sequences will facilitate better comparative analysis of the tRNA gene arrangement and structure in flowering plants.

3.1 Nuclear tDNA content

There is a wide variation in the number of tRNA genes, or tDNAs, among the plant genomes studied, even within the same lineage (Figure 1). Among these lineages, monocots have the largest range in tDNAs (152–1,491 tDNAs; Figure 1A). Compared to the more ancestral ANA clade, several eudicot and monocot genomes have evolved to have a greater number of tDNAs, with some even exceeding 1,400 tDNAs, as in the eudicot *Sinapis alba* ($n = 1,407$) and the monocots *Thinopyrum intermedium* ($n = 1,491$) and *Triticum aestivum* ($n = 1,472$). On the other hand, *E. ferox* had the highest tDNA count of 583 among the ANA species studied

(Figures 1A, B). *Spirodela polyrhiza* had the smallest number of tDNAs at 152 between the eudicots and monocots. Regarding the number of tDNAs, no general pattern was observed within the eudicots and monocots suggesting that lineage does not influence the number of tDNAs. Genome sizes are also not correlated with the number of tDNAs (Figure 1C), as there is a low correlation between genome size and tRNA gene count in our angiosperm dataset ($R^2 = 0.41$, p -value < 0.0001). Grouping the plants into their respective lineage showed that eudicots have the least correlation ($R^2 = 0.29$, p -value = 0.0002), while the monocots showed a relatively high correlation ($R^2 = 0.79$, p -value < 0.0001). At least for the monocot lineage, one can expect an increased number of tDNAs with a larger genome size. On the other hand, since the linear regression for ANA has a very high p -value (0.7677; likely due to having only four data points), we cannot make conclusions regarding the correlation between genome size and tRNA gene count in the ANA lineage.

No distinct patterns can also be observed between lineages regarding tRNA isoacceptor content (Figure 2 and Supplementary

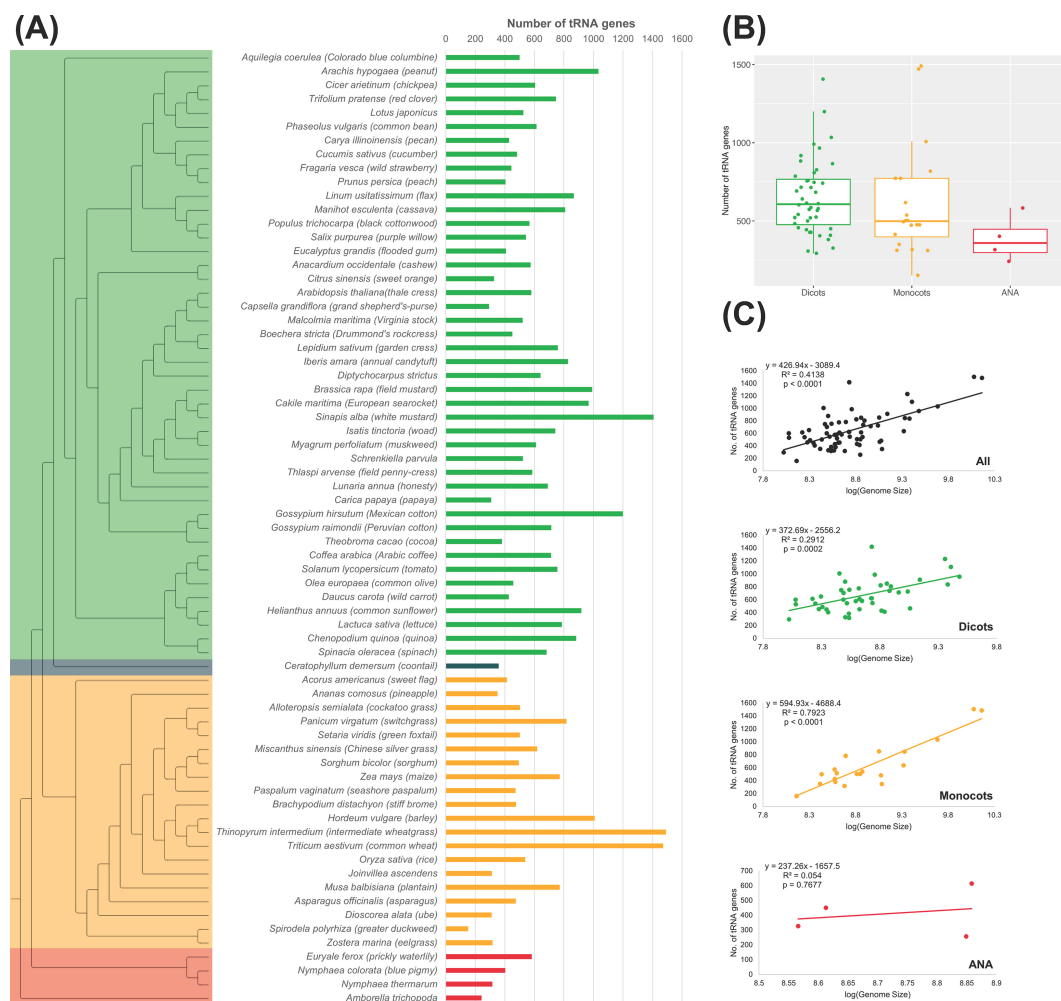


FIGURE 1

tRNA gene counts in plant nuclear genomes. **(A)** The phylogenetic tree on the left illustrates the evolutionary relationship among the 69 plant genomes examined. In the tree, eudicots are represented in green, monocots in orange, ANA (Amborellales, Nymphaeales, and Austrobaileyales) in red, and *Ceratophyllum* in blue. Adjacent to the tree, a bar graph shows the number of high-confidence tRNA genes found in each species. **(B)** Distribution of tRNA gene counts across different lineages. **(C)** Correlation between genome size and the number of tRNA genes is presented for all genomes as well as for each lineage.

File 1). The most abundant tRNA isotypes include tRNA^{Ala}, tRNA^{Pro}, tRNA^{Ser}, tRNA^{Arg}, and tRNA^{Leu}. All genomes, however, lacked tRNA^{Pro}-GGG and tRNA^{Leu}-GAG tDNAs, while tRNA^{Gly}-ACC, tRNA^{Arg}-GCG, and tRNA^{Phe}-AAA tDNAs were each found in only one genome (*A. americanus*, *Gossypium raimondii*, and *Arachis hypogaea*, respectively; Figure 2A). Out of the six tRNA isoacceptors for tRNA^{Arg}, *T. aestivum* only contained tRNA^{Arg}-TCT (Figure 2A). At the same time, *Helianthus annuus* and *S. alba* completely lacked a nuclear tRNA^{Gly} and tRNA^{Asp}, respectively (Figure 2B).

On average, less than half of all tRNA genes of each lineage are unique (Figure 3). Specifically, 35%, 39%, and 47% of the total tDNAs are unique in the eudicot, monocot, and ANA genomes, respectively. The more ancestral ANA clade had higher percentages of unique tDNA sequences in general, with *A. trichopoda* having the highest at 67%. The more recent lineages, eudicots and monocots, showed a general decrease in tRNA gene uniqueness suggesting a higher prevalence of tRNA gene duplications in these lineages.

All the plant genomes analyzed in this study have intron-containing tRNA^{eMet} and tRNA^{Tyr} (Supplementary Figures 2–3).

The mean length of these introns is similar for all lineages (Table 1), though there are extreme outliers. Five monocot tRNA^{eMet} introns had lengths ranging from 59 to 86 bp, three of which are in the *T. intermedium* genome (Supplementary Table 3). On the other hand, two long tRNA^{Tyr} introns were found in the *Miscanthus sinensis* genome (172 and 64 bp in size, respectively), while two identical 85-bp tRNA^{Tyr} introns were each found in the *G. hirsutum* and *G. raimondii* genomes (Supplementary Table 3). Aligning all tRNA^{eMet} and tRNA^{Tyr} introns reveals a modest conservation in the former and a relatively lesser conservation in the latter. For tRNA^{eMet}, a GCT motif at the start of the intron and a GAGT motif near the end appear to be conserved in angiosperms (Supplementary Figure 2). For tRNA^{Tyr}, a CAG motif around the middle of the intron appears to be the only relatively conserved residue (Supplementary Figure 3). Although rare, introns were also found in non-Met and non-Tyr tDNAs (Table 2). While most tRNA isotypes had at least one intron-containing tDNA, no intron-containing tRNA^{Ala}, tRNA^{Asp}, and tRNA^{His} were found in any plant nuclear genomes studied.

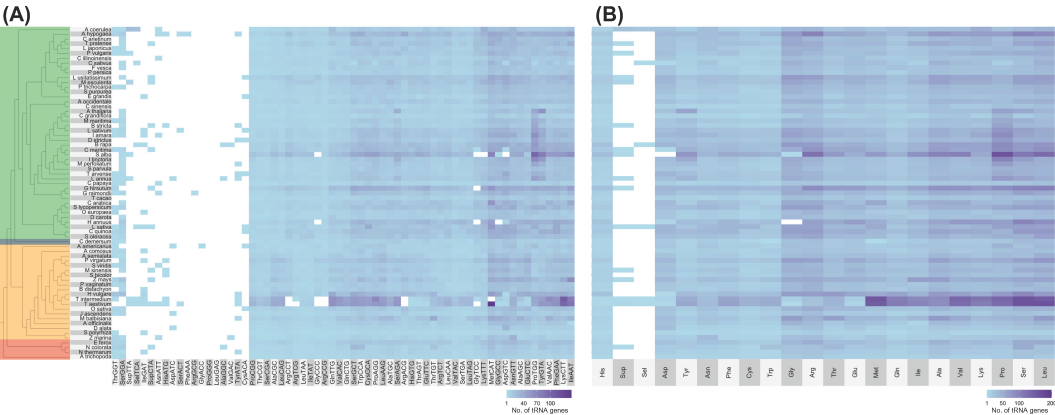


FIGURE 2 Number of tRNA isoacceptor genes found in plant nuclear genomes. Alongside the heatmap, which displays the number of tRNA genes categorized by (A) isoacceptors and (B) isotypes, is the exact phylogenetic tree shown in Figure 1A. The color coding indicates different groups: green represents eudicots, orange denotes monocots, red signifies ANA (Amborellales, Nymphaeales, and Austrobaileyales), and blue corresponds to *Ceratophyllum*. In the heatmap, white shows that no tRNA gene was found. Refer to Supplementary File 1 for the tRNA gene counts of all plant genomes examined.

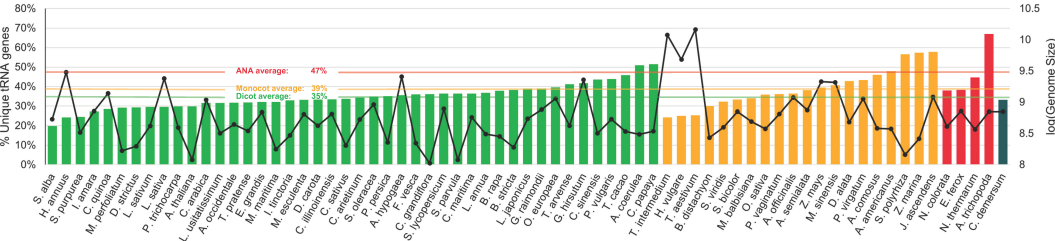


FIGURE 3 Percentage of unique tRNA gene sequences identified in the nuclear genomes of various plants. Each bar represents the proportion of unique tRNA gene sequences relative to the total number of tRNA genes within each genome. The bars are color coded according to plant lineages: green for eudicots, orange for monocots, red for ANA (Amborellales, Nymphaeales, and Austrobaileyales), and blue for *Ceratophyllum*. Additionally, a second y-axis displaying genome sizes is indicated by solid black lines. A horizontal line representing the average percentage for each major lineage is also included for reference.

TABLE 1 Mean intron lengths of tDNA^{eMet} and tDNA^{Tyr} in plant nuclear genomes.

Lineage	Mean intron length (bp)	
	tDNA ^{eMet}	tDNA ^{Tyr}
ANA	12.16	17.24
Dicot	11.49	14.54
Monocot	14.44	14.11

TABLE 2 Detected non-Met and non-Tyr intron-containing tDNAs in plant nuclear genomes.

Isotype	Genome(s) (no. of intron-containing tDNAs found)
tRNA ^{Ala}	None
tRNA ^{Gly}	<i>M. balbisiana</i> (1)
tRNA ^{Pro}	<i>P. virgatum</i> (2)
tRNA ^{Thr}	<i>L. annua</i> (1), <i>H. vulgare</i> (1), <i>T. intermedium</i> (3)
tRNA ^{Val}	<i>I. tinctoria</i> (1), <i>T. aestivum</i> (1), <i>T. intermedium</i> (1), <i>Z. mays</i> (1)
tRNA ^{Ser}	<i>C. grandiflora</i> (1), <i>P. vaginatum</i> (2)
tRNA ^{Arg}	<i>H. vulgare</i> (1), <i>A. trichopoda</i> (1)
tRNA ^{Leu}	<i>G. hirsutum</i> (1), <i>A. semialata</i> (1), <i>T. aestivum</i> (4), <i>T. intermedium</i> (3)
tRNA ^{Phe}	<i>G. hirsutum</i> (9), <i>G. raimondii</i> (3), <i>P. virgatum</i> (1), <i>T. intermedium</i> (1)
tRNA ^{Asn}	<i>I. amara</i> (2), <i>L. usitatissimum</i> (1)
tRNA ^{Lys}	<i>B. rapa</i> (1), <i>I. tinctoria</i> (2), <i>L. usitatissimum</i> (1), <i>T. aestivum</i> (3), <i>T. intermedium</i> (1)
tRNA ^{Asp}	None
tRNA ^{Glu}	<i>P. virgatum</i> (3)
tRNA ^{His}	None
tRNA ^{Gln}	<i>P. virgatum</i> (1)
tRNA ^{Ile}	<i>P. virgatum</i> (1)
tRNA ^{Cys}	<i>P. virgatum</i> (1), <i>A. trichopoda</i> (1)
tRNA ^{Trp}	<i>H. vulgare</i> (1)
tRNA ^{Sup}	<i>B. stricta</i> (1), <i>L. annua</i> (1), <i>M. sinensis</i> (1)

tRNA^{Sup} refers to the suppressor tRNA.

3.2 Nuclear tDNA regulatory regions

Previous analyses of plant tDNA sequences reveal the prevalence of several regulatory elements implicated in the proper recruitment of RNA polymerase III and its efficiency in transcribing nuclear plant tDNAs: an A-/T-rich upstream region (Choisne et al., 1998; Yukawa et al., 2000, 2013; Michaud et al., 2011), upstream TATA-box and CAA motifs (Choisne et al., 1998; Dieci et al., 2006; Michaud et al., 2011; Yukawa et al., 2011, 2013; Soprano et al., 2018), intragenic A and B box promoters (Yukawa et al., 2000, 2013; Michaud et al., 2011; Mitra et al., 2015; Soprano et al., 2018), and

downstream stretches of Ts for transcription termination (Yukawa et al., 2000; Arimbasseri and Maraia, 2015; Soprano et al., 2018). In our dataset, the 50 nucleotide sequences immediately upstream of the tDNAs are predominantly A-/T-rich (Supplementary Figure 4; Supplementary Table 4), and this A-/T-rich upstream region of tDNAs is not dictated by the A/T content of the genome (Supplementary Figures 4F–I). This A-/T-rich feature does not extend past the 50 nucleotides upstream of the tDNAs (Supplementary Figure 5).

Looking for regulatory elements in the sequences 50 bases upstream of the detected tDNAs revealed a modest percentage of tDNAs, at approximately 22%–32%, having at least one TATA-box motif, and a high percentage, at approximately 78%–82%, having at least one CAA motif (Table 3). Narrowing down on the first 10 nucleotides upstream of tDNAs, where CAA triplets usually are found to act as transcription initiation sites in *Arabidopsis* (Yukawa et al., 2011), reduces the percentages to approximately 36%–45% (Table 3). On the other hand, sequences 50 nucleotides downstream of all tDNAs revealed a high percentage, at approximately 67%–72%, of having at least one stretch of T residues at least four bases long (Table 4). Many of these tDNAs (39%–44%) also contain a “backup” stretch of T residues shortly after the first poly(T) stretch, a common characteristic found in eukaryotic tRNA genes (Braglia et al., 2005; Padilla-Mejía et al., 2009). The lengths of the poly(T) stretches are variable, the longest being 19, 26, and 23 bp for ANA, eudicots, and monocots, respectively (Supplementary Figure 6).

All tDNAs in the study contained A and B boxes within their coding regions, with varying consensus sequences depending on the tRNA isotype and lineage (Supplementary Files 2–3). For A boxes, there are generally conserved T and GG residues at the 5′ and 3′ positions, respectively. In contrast, for B boxes, there are generally conserved GG and CC residues at the 5′ and 3′ positions, respectively. Each tRNA isotype had varying internal A and B box sequences, but the internal sequences were generally conserved among lineages for each isotype. However, some A and B boxes had sequences vastly different from the consensus and are listed separately in Supplementary Table 5.

3.3 A single conserved tRNA^{Ala-AGC} species

A conserved tRNA^{Ala-AGC} species was detected in our genomic dataset (Supplementary Figure 7A; see Supplementary Figure 8 for consensus structures of other tRNA^{Ala} isoacceptors). Polymorphic tRNA^{Ala-AGC} sequences were also detected (Supplementary Figure 7B); thus, we also analyzed the evolution and structural conservation of all detected tRNA^{Ala-AGC} genes. Gene tree and species tree reconciliation via Notung (Chen et al., 2000; Zmasek and Eddy, 2001; Durand et al., 2006; Vernot et al., 2007; Stolzer et al., 2012; Darby et al., 2017) reveals that the evolution of tRNA^{Ala-AGC} in angiosperms is characterized by more gene losses than duplications (253 inferred gene duplications and 586 inferred gene losses; Supplementary File 4). The tRNA, cloverleaf stem, and variable loop lengths are generally conserved in the nuclear tRNA^{Ala-AGC} genes in plants (Figures 4A–F). Sequence covariation analysis reveals that the base pairing within each cloverleaf stem is

TABLE 3 Percentage of tRNA genes possessing upstream TATA-box and CAA motifs.

Lineage	tDNAs (%) with upstream TATA motifs				tDNAs (%) with upstream CAA motifs	
	=1 TATA	=2 TATA	>2 TATA	Total (≥1 TATA)	50 Bases upstream	10 Bases upstream
ANA	15.83 ± 4.95%	4.65 ± 1.08%	1.60 ± 0.87%	22.08 ± 6.13%	78.71 ± 1.76%	36.15 ± 2.21%
Dicot	20.36 ± 2.65%	6.97 ± 1.79%	4.56 ± 1.78%	31.89 ± 5.03%	82.29 ± 3.20%	45.40 ± 4.49%
Monocot	17.66 ± 4.31%	6.22 ± 3.08%	4.16 ± 3.38%	28.04 ± 9.74%	79.97 ± 4.98%	38.08 ± 5.70%

Sequences 50 bases upstream of tDNAs were searched for TATA-box motifs using PlantCARE (Lescot et al., 2002).

TABLE 4 Poly(T) termination signals found downstream of tRNA genes.

	ANA	Dicots	Monocots
Number of tDNAs	1,635	29,463	12,698
%tDNAs with poly(T)s	67.03%	72.77%	69.20%
%tDNAs with backup poly(T)s	39.14%	44.60%	44.32%
Mean length (bp)	5.02	5.20	5.44
Length (bp)	4–19	4–26	4–23

A tDNA contains a backup poly(T) signal if it has at least one additional stretch of T residues downstream of the first determined poly(T) signal.

not well conserved in tRNA^{Ala-AGC} (Figures 4G-I). In general, for all lineages, base pairs (represented by single arcs) show negative covariation, where should a base mutate in one of the stems, its paired base will not likely mutate to preserve the base pairing. An exception is the D-stem of monocot tRNA^{Ala-AGC} genes, whose base pairs or arcs exhibit positive covariation.

3.4 Nuclear tDNA clusters

We classified a group of tDNAs as a cluster if they have a density of at least three tDNAs per kilobase of a genomic region. The majority of eudicot genomes (40 out of 44) and only a modest percentage of monocot (13 out of 20) and ANA genomes (1 out of 4) contained at least one tDNA cluster using this criterion. The proportion of tRNA genes that are clustered is generally deficient among angiosperms (5% and 3% in eudicots and monocots, respectively), the highest being 20% in *Musa balbisiana*, followed by *A. thaliana* and *Isatis tinctoria* (19% and 16% clustered tDNAs, respectively). In the eudicot, monocot, and ANA lineages, 324, 103, and 2 tDNA clusters were identified, respectively. The following tDNA clusters were detected in our analysis: stretches of at least three tRNA^{Pro} (to as many as 10) found in *Ceratophyllum*, eudicots, and monocots; stretches of alternating tRNA^{Tyr} and tRNA^{Ser} found only in eudicots (Figure 5A); and a stretch of 28 tRNA^{Ile} found only in the monocot *Zea mays* (Figure 5B). Since these clusters may be linked to tRNA gene duplication, gene duplication events of tRNA^{Pro} and tRNA^{Ile} were inferred using Notung. Reconciliation of each tRNA gene tree with the species tree reveals that the tRNA^{Pro} and tRNA^{Ile} genes underwent 592 and 479 gene duplication events, respectively (Supplementary Files 5–6).

3.5 Organellar tDNA content, organization, and structure

In contrast to their nuclear counterparts, chloroplast and mitochondrial genomes show slight variation in their tDNA numbers. The tRNA isotype content of plastomes and mitogenomes also shows slight variation among the different plant lineages (Figure 6). The relative abundance of each isotype is almost uniform in all the surveyed chloroplast genomes, while it varies in all the surveyed mitogenomes. Apart from *A. coerulea*, all the surveyed plastomes lack a tRNA^{Lys} gene. Plastomes typically have 31–36 tDNAs regardless of lineage (except for *Cicer arietinum* and *A. coerulea*, with 25 and 41 chloroplast tDNAs, respectively). On the other hand, mitogenomes typically have 17–36 tDNAs and more variable tDNA content than the plastomes. The eudicot *Citrus sinensis* has 49 mitochondrial tDNAs.

Although *H. annuus* lacked nuclear tRNA^{Gly} genes (Figure 2B), one tRNA^{Gly-GCC} sequence was detected in its chloroplast and mitochondrial genomes. *S. alba*, which lacked a nuclear tRNA^{Asp} (Figure 2B), also had one detected tRNA^{Asp-GTC} in its chloroplast genome. While *S. alba* currently does not have an available mitochondrial genome, the closely related *Brassica rapa* (Supplementary Figure 1) also has one tRNA^{Asp-GTC} gene in its mitogenome.

The tRNA gene organization in the plastomes and mitogenomes reflects the evolutionary conservation of these organellar genomes. Plastomes of flowering plants show a relatively conserved tRNA gene organization, with some rearrangements in some species (Supplementary Files 7–9). Their mitogenomes, on the other hand, show little conservation in their tRNA gene organization (Supplementary Files 10–12).

Unlike their nuclear counterparts, sequences immediately upstream of organellar tDNAs do not exhibit a distinct, consistent pattern. Though chloroplast tDNAs still have predominantly A-/T-rich upstream sequences (Supplementary Figures 9 and 10), the same cannot be said about mitochondrial tDNAs, which exhibit much less conservation than chloroplast tDNAs (Supplementary Figures 11 and 12).

4 Discussion

A narrow range of nuclear tDNA numbers in angiosperms (500–600 tDNAs between five angiosperm genomes) had been previously reported (Michaud et al., 2011), and extending the

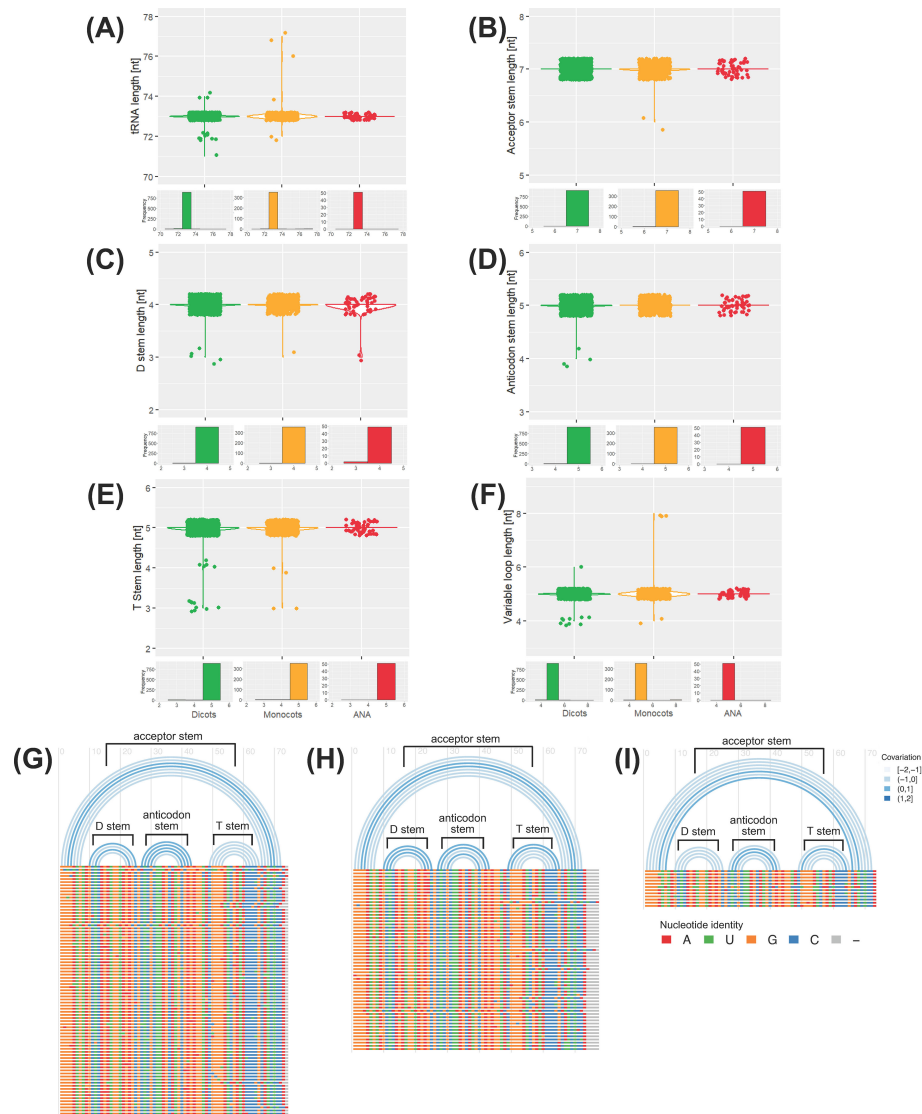


FIGURE 4

Conservation of the tRNA^{Ala-AGC} secondary structure. The distribution of lengths for various elements of the tRNA^{Ala-AGC} genes across different lineages is displayed: (A) tRNA, (B) acceptor stem, (C) D stem, (D) anticodon stem, (E) T stem, and (F) variable loop lengths for each lineage (green for eudicots, yellow for monocots, red for ANA). Structural representation of tRNA^{Ala-AGC} is also illustrated through arc diagrams for (G) eudicots, (H) monocots, and (I) ANA (Amborellales, Nymphaeales, and Austrobaileyales) generated using R-chie. Horizontal bars below the arcs (colored by nucleotide identity, bottom legend: A is red, U is green, G is orange, C is blue, and gray is a gap) represent the multiple sequence alignment of all unique tRNA^{Ala-AGC} genes of each lineage. Significant arcs corresponding to the different tRNA cloverleaf stems are labeled accordingly. The top legend for (G) to (I) indicates the covariation of the base pairing between the arches, where a negative and positive covariations indicates no conservation and conservation of base pairings, respectively.

coverage to 69 angiosperm genomes resulted in a broader range in the number of nuclear tDNAs that were detected (approximately 150–1,500 tDNAs; Figure 1A). This tDNA range is comparable to that reported by Bermudez-Santana et al. (2010), although their range also included tRNA pseudogenes (432–1,290 tDNAs between seven land plant genomes). In addition, the green algae *Volvox carteri* and *Chlamydomonas reinhardtii* were reported to have 1,051 (including tRNA pseudogenes) and 256 tDNAs, respectively (Bermudez-Santana et al., 2010; Michaud et al., 2011). Therefore, nuclear genomes from the green lineage can have tDNAs as few as 150 or as many as 1,500. This variation in plant nuclear tDNA numbers is relatively small compared to other eukaryotes.

Tetraodontiformes have approximately 700 tDNAs, while the related zebrafish, *Danio rerio*, has approximately 20,000 (Bermudez-Santana et al., 2010). Concurrently, in mammals, old-world monkeys and apes had 496–736 tDNAs, while cows and rats exceeded 100,000 tDNAs (Bermudez-Santana et al., 2010); a reannotation of the cow tRNAs showed that the majority of these putative tDNAs include tRNA-like sequences (Theologis et al., 2000; Tang et al., 2009). In nuclear eukaryotic genomes, the number of tDNAs can vary even within species of the same lineage or clade. Indeed, ANA, eudicot, and monocot genomes have varying numbers of nuclear tDNAs, and no lineage-specific pattern could be observed (Figure 1A).

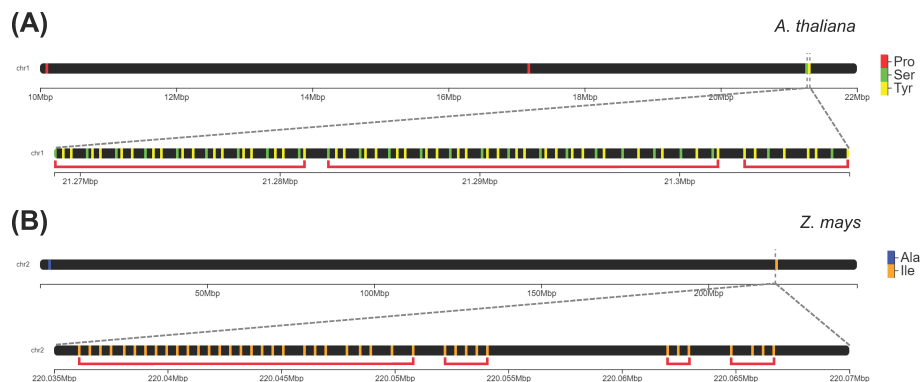


FIGURE 5

Extensive tRNA gene clusters identified in the genomes of eudicots and monocots. (A) In the genome of *Arabidopsis thaliana*, one cluster on Chromosome 1 consists of alternating tRNA^{Tyr} and tRNA^{Ser} genes. (B) In *Zea mays*, there are clusters on Chromosome 2 that are composed of tandem repeats of tRNA^{Ile} genes. Each red bracket indicates a distinct gene cluster.

The varying genome sizes in eukaryotes, including plants, could explain this variation in the number of tDNAs. While earlier studies suggested a strong correlation among plants, with *Arabidopsis* being an outlier (Bermudez-Santana et al., 2010; Michaud et al., 2011), our data showed a weak overall correlation in the 69 angiosperm genomes studied ($R^2 = 0.41$, p -value < 0.0001 ; Figure 1C), especially for the eudicot lineage, with an R-squared value of 0.29 (p -value = 0.0002). More recent studies have similarly reported a weak correlation among plants (Mohanta et al., 2020; Santos and Del-Bem, 2023). However, this was not the case for the monocot lineage, which exhibited a strong correlation ($R^2 = 0.79$, p -value < 0.0001 ; Figure 1C). A strong correlation between the monocot genome sizes and the number of tDNAs had been previously reported (Planta et al., 2022).

At least for the eudicot genomes, a likely explanation is related to the unique case of *Arabidopsis* (Michaud et al., 2011). A weak correlation between the number of tDNAs and genome size was initially shown in *A. thaliana*, with an R-squared value of 0.16. This correlation contrasted with the other analyzed plant genomes, which all had moderate to high R-squared values. Compared to four other angiosperms (*Medicago truncatula*, *Populus trichocarpa*, *Oryza sativa*, and *Brachypodium distachyon*) and one green alga (*C. reinhardtii*), *A. thaliana* had a higher number of tDNAs in each chromosome (Michaud et al., 2011). Except for *A. thaliana*, the other genomes had at most only two tDNAs per Mb of chromosome. Chromosomes 2–5 of *A. thaliana* had approximately four tDNAs per Mb, while Chromosome 1 had eight tDNAs per Mb of chromosome (Michaud et al., 2011). This unusually high number of tDNAs in Chromosome 1 of *A. thaliana* is largely due to the existence of two large tDNA clusters in this chromosome: tandem repeats of 27 tRNA^{Pro} and tandem repeats of 27 tRNA^{Tyr}–tRNA^{Tyr}–tRNA^{Ser} (Theologis et al., 2000). These clusters, indicative of gene duplications (Theologis et al., 2000; Bermudez-Santana et al., 2010), are likely the cause of the weak correlation between the tDNA number and genome size of *A. thaliana*. Indeed, removing the tRNA isotypes involved in the two identified clusters (tRNA^{Pro}, tRNA^{Ser}, and tRNA^{Tyr}) increased the R-squared value in *A. thaliana* from 0.16 to 0.70 (Michaud et al., 2011).

Similarly, the weak overall correlation found in the angiosperm genomes in this study might be explained by the prevalence of gene duplication events. This is likely the case, given that generally less than half of all tDNAs of each lineage were found to be unique (Figure 3). This may also explain the observation that plants, alongside vertebrates, appear to have higher tDNA count and redundancy compared to other organisms (Santos and Del-Bem, 2023). However, this does not explain why the monocots showed a strong correlation between tDNA number and genome size ($R^2 = 0.79$; Figure 1C), as opposed to weaker correlation observed in eudicot genomes ($R^2 = 0.29$). The key difference may lie in the existence of tDNA clusters, like the ones found in *A. thaliana*.

We considered tRNA genes to be clustered if at least three tDNAs were within 1 kb of each other. Using this criterion, 324 (in 40/44 genomes), 103 (in 13/20 genomes), and 2 (in 1/4 genomes) tDNA clusters were identified in eudicots, monocots, and ANA, respectively. Eudicots thus appear to have a stronger tendency toward gene duplication in the form of tDNA clustering compared to the other plant lineages, and this should explain the weaker correlation between tDNA numbers and genome sizes in eudicots compared to those in monocots. While ANA genomes appear to have a weak correlation like eudicots (Figure 1C), they had very few tDNA clusters. It is very likely that the linear regression model did not properly represent the correlation between ANA genome size and tRNA gene count due to the high p -value (0.7677). This may also be a result of our stricter criteria for tDNA clustering compared to other tRNA studies (Bermudez-Santana et al., 2010; Morgado and Vicente, 2019), which considered clusters as having at least two tDNAs within 1 kb of each other.

We identified tDNA clusters in Chromosome 1 of *A. thaliana*, similar to the two large clusters that were previously reported (Michaud et al., 2011) as follows: (i) consecutive tRNA^{Pro} clusters, adding up to 25 tandem repeats of tRNA^{Pro}, and (ii) consecutive tRNA^{Tyr}–tRNA^{Ser} clusters, comprising a long stretch of alternating tRNA^{Tyr} and tRNA^{Ser} genes. Unlike previously reported, these stretches of tRNA^{Tyr} and tRNA^{Ser} genes were not strictly tandem

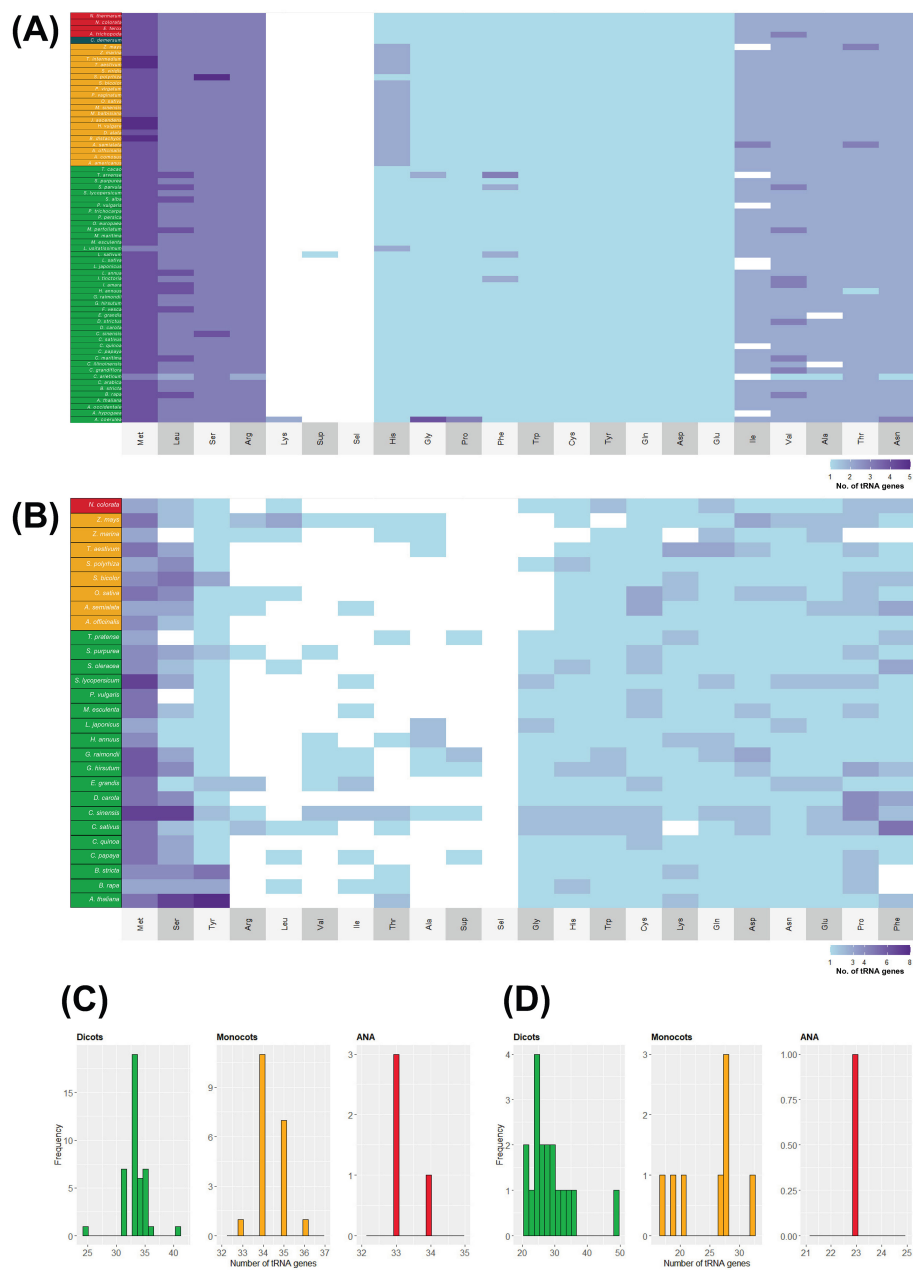


FIGURE 6 tRNA isotypes and gene numbers in plant organellar genomes. The heatmap illustrates the number of tRNA isotypes found in (A) chloroplast and (B) mitochondrial genomes of plants. Species names are color coded according to their lineage: green represents eudicots, orange denotes monocots, red indicates ANA (Amborellales, Nymphaeales, and Austrobaileyales), and blue signifies *Ceratophyllum*. Additionally, the distribution of tRNA gene counts is displayed for (C) chloroplast and (D) mitochondrial genomes.

repeats of the triplet tRNA^{Tyr}–tRNA^{Tyr}–tRNA^{Ser}. The difference in the size and order of these clusters compared to those found by Theologis et al. (2000) is likely due to the updated genome assembly for *A. thaliana*. These tRNA^{Pro} and tRNA^{Tyr}–tRNA^{Ser} clusters were also found in other plant genomes. Most eudicots (34 out of 44 genomes, including *A. thaliana*), a few monocots (6 out of 20 genomes), and *C. demersum* were also found to have stretches of tRNA^{Pro} genes. On the other hand, a long stretch of alternating tRNA^{Tyr} and tRNA^{Ser} genes was also found in eight other eudicot genomes (*Boechera stricta*,

Diptychocarpus strictus, *Iberis amara*, *I. tinctoria*, *Lunaria annua*, *Lepidium sativum*, *Malcolmia maritima*, and *Myagrum perfoliatum*). This tRNA^{Tyr}–tRNA^{Ser} tDNA cluster was not found in any other monocot or ANA genome. Another tDNA cluster detected is a tandem repeat of 28 tRNA^{Ile} found exclusively in Chromosome 2 of *Z. mays*. Among the clusters found in this study, this is the longest in size. Interestingly, this cluster is followed by three more clusters consisting purely of tRNA^{Ile} (5x tRNA^{Ile}, 3x tRNA^{Ile}, then 4x tRNA^{Ile}) within the same chromosome.

It remains to be seen whether these tDNA clusters serve any biological purpose. tDNA clusters are implicated in genome breakage resulting in genome rearrangement (Rienzi et al., 2009). They are also found to be involved in mobile genetic elements and horizontal gene transfer (Morgado and Vicente, 2019). tDNA clusters are likely dynamic and fragile genomic regions, and this inherent instability might be the reason for the evolution and prevalence of these tDNA clusters rather than being products of positive selection. Moreover, a study on the tDNA clusters of *Arabidopsis* shows that these clusters are predominantly methylated and transcriptionally repressed (Hummel et al., 2020). However, the case of tRNA^{Pro} clusters is intriguing given its frequency among the plant genomes studied.

Proline is found to have diverse roles in plants. They are involved in cell wall and plant growth (Kishor et al., 2015), but the more well-documented function of proline is related to plant stress. In response to different environmental stresses, e.g., drought or water loss, salt, metal, and pathogen attack, plants accumulate proline (Kishor et al., 2005; Verslues and Sharma, 2010; Patriarca et al., 2021; Vujanovic et al., 2022). Being an osmolyte, proline can maintain cellular metabolism and even reduce plant growth in stressful conditions (Maggio et al., 2002; Vujanovic et al., 2022). This physiological response of proline accumulation would involve tRNA^{Pro} activity and could thus be a reason behind the prevalence of tRNA^{Pro} clusters and duplications (Supplementary File 8). While these clusters might be initially repressed by methylation (Hummel et al., 2020), the plant stress response could induce the removal of these epigenetic marks, thereby increasing global tRNA^{Pro} transcription levels. To confirm this link, future studies are encouraged to look into the expression profile of these clustered tDNAs in plants. The potential biological functions of these tDNA clusters themselves may also be investigated further by future studies.

Another interesting observation is the apparent lack of certain tRNA isotypes in the nuclear genome of *H. annuus* and *S. alba*, even though their organellar counterparts are present. After further investigation, we found that prior to filtering via EukHigh ConfidenceFilter, *H. annuus* and *S. alba* had 117 tRNA^{Gly} and 82 tRNA^{Asp} predicted genes, respectively. None of these first-pass tRNA genes had an isotype score that met the cutoff for EukHigh ConfidenceFilter, which was 95 by default for these two isotypes. The tRNAscan-SE developers emphasized to only change the cutoff values with great caution, as they have already been tested on different large eukaryotic genomes (Chan and Lowe, 2019); thus, throughout our analysis, we opted to keep all default cutoff values unchanged. However, the fact that some of the first-pass tRNA^{Gly} and tRNA^{Asp} genes had scores that were very close to the cutoff value (as close as 94.5) indicates the need to reevaluate these score cutoffs.

To transcribe plant tRNAs, RNA polymerase III (Pol III) is recruited. One of the requirements for its recruitment is a TATA-binding protein (TBP), and the presence of TATA-box motifs upstream of plant tRNA genes is implicated in the efficiency of tRNA transcription (Dieci et al., 2006; Michaud et al., 2011). However, the proportion of angiosperm tDNAs containing such a

motif is strikingly low (Table 3). Previous studies have similarly reported the lack of TATA-box motifs upstream of many eukaryotic tDNAs (Hamada et al., 2001; Giuliadori et al., 2003; Dieci et al., 2006) as well as the little effect caused by the removal of TATA-box motifs in the transcription of plant tRNA^{Leu} genes (Choisne et al., 1998). For many Pol III-transcribed genes, TBP can be recruited without a specific TATA-like sequence. For these TATA-less genes, recruiting Pol III is instead facilitated by TFIIC, which binds the DNA via the A and B boxes and recruits TFIIB, which has a TBP as one of its subunits. TFIIB recruits Pol III (Choisne et al., 1998; Yukawa et al., 2000; Dieci et al., 2006). This suggests that while some plants prefer the TATA-mediated recruitment of TBP [e.g., *A. thaliana* (Choisne et al., 1998; Hamada et al., 2001)], it may not be preferred or deemed necessary by other organisms that lack conserved TATA-box motifs. Dieci et al. (2006) hinted that the difference between a TATA-box-dependent and a TATA-box-independent organism might be found in their respective transcription machinery. Notably, the intragenic A and B boxes bound by TFIIC were found in all detected nuclear tRNA genes (Supplementary Files 2 and 3). However, this can mainly be explained by the fact that the tRNA D- and T-loops are encoded within these boxes (Galli et al., 1981; Hofstetter et al., 1981; Turowski and Tollervey, 2016) and that the tRNAscan-SE program detects tRNA genes based on the presence of A and B box sequences (Lowe and Eddy, 1997).

The CAA motifs, on the other hand, were found in most angiosperm tDNAs between positions −1 and −50 bp (Table 3). Removal of these motifs upstream of plant tDNAs decreased *in vitro* expression levels of these tRNAs (Choisne et al., 1998; Yukawa et al., 2000). While previous studies reported functional CAA motifs to be between −1 and −10 bp in plant tDNAs (Yukawa et al., 2000, 2011; Michaud et al., 2011), more CAA motifs were found when the scope was extended up to −50 bp (Table 3). This suggests that transcription start sites (TSS) for many plant tDNAs may be further upstream than others.

The majority of angiosperm tDNAs contained at least one downstream stretch of T residues (Table 4), which is expected as it is considered an essential signal used by Pol III for transcription termination (Braglia et al., 2005; Arimbasseri and Maraia, 2015). In eukaryotic tRNAs, this poly(T) signal is commonly found to be approximately four to five bases long (Braglia et al., 2005). Aside from stretches of four to five T residues, there is also an abundance of poly(T) stretches that are 6 to 10 bases long, and those with extreme lengths—19, 26, and 23 bases—were found in the ANA, eudicot, and monocot tDNAs, respectively. While a significant percentage of angiosperm tDNAs do not contain a downstream poly(T) signal (Table 4), it is possible that increasing the coverage to 100 or more nucleotides downstream (instead of only 50) will locate more poly(T) signals, backup poly(T) signals, and other poly(T) signals of extreme and variable lengths.

Our results provide a comprehensive overview of the tRNA gene content, structure, and organization of nuclear and organellar angiosperm genomes, utilizing the recent abundance of genomic data enabled by next-generation sequencing technologies. This study can thus supplement further studies on plant tRNA gene

function and regulation. The specific function of these tRNA gene clusters and an explanation for the differences in the abundance of several regulatory motifs [e.g., TATA-boxes, CAA motifs, and poly (T) stretches] are some points that may be explored in the future.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/[Supplementary Material](#).

Author contributions

KM: Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. JP: Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was funded by the UP System Enhanced Creative Work and Research Grant (ECWRG-2021-2-8R) to JP.

References

- Alves, C. S., and Nogueira, F. T. S. (2021). Plant small RNA world growing bigger: tRNA-derived fragments, longstanding players in regulatory processes. *Front. Mol. Biosci.* 8. doi: 10.3389/fmolb.2021.638911/BIBTEX
- Anand, L., and Rodriguez Lopez, C. M. (2022). ChromoMap: an R package for interactive visualization of multi-omics data and annotation of chromosomes. *BMC Bioinf.* 23, 33. doi: 10.1186/s12859-021-04556-z
- Arimbasseri, A. G., and Maraia, R. J. (2015). Mechanism of transcription termination by RNA polymerase III utilizes a non-template strand sequence-specific signal element. *Mol. Cell* 58, 1124–1132. doi: 10.1016/j.molcel.2015.04.002
- Barter, R., and Yu, B. (2017). superheat: A graphical tool for exploring complex datasets using heatmaps. Available online at: <https://CRAN.R-project.org/package=superheat> (Accessed December 1, 2024).
- Bermudez-Santana, C., Attolini, C. S., Kirsten, T., Engelhardt, J., Prohaska, S. J., Steigle, S., et al. (2010). Genomic organization of eukaryotic tRNAs. *BMC Genomics* 11, 1–14. doi: 10.1186/1471-2164-11-270/FIGURES/7
- Bernhart, S. H., Hofacker, I. L., Will, S., Gruber, A. R., and Stadler, P. F. (2008). RNAalifold: improved consensus structure prediction for RNA alignments. *BMC Bioinf.* 9, 474. doi: 10.1186/1471-2105-9-474
- Braglia, P., Percudani, R., and Dieci, G. (2005). Sequence context effects on oligo(dT) termination signal recognition by *Saccharomyces cerevisiae* RNA polymerase III. *J. Biol. Chem.* 280, 19551–19562. doi: 10.1074/jbc.M412238200
- Capella-Gutiérrez, S., Silla-Martínez, J. M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973. doi: 10.1093/bioinformatics/btp348
- Chambers, J. M. (1992). "Linear Models," in *Statistical Models in S* (USA: Routledge).
- Chan, P. P., Lin, B. Y., Mak, A. J., and Lowe, T. M. (2021). tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. *Nucleic Acids Res.* 49, 9077–9096. doi: 10.1093/NAR/GKAB688
- Chan, P. P., and Lowe, T. M. (2019). tRNAscan-SE: Searching for tRNA genes in genomic sequences. *Methods Mol. Biol. (Clifton N.J.)* 1962, 1. doi: 10.1007/978-1-4939-9173-0_1
- Chao, J., Li, Z., Sun, Y., Aluko, O. O., Wu, X., Wang, Q., et al. (2021). MG2C: a user-friendly online tool for drawing genetic maps. *Mol. Horticulture* 1, 1–4. doi: 10.1186/S43897-021-00020-X
- Chen, C., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y., et al. (2020). TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* 13, 1194–1202. doi: 10.1016/j.molp.2020.06.009
- Chen, K., Durand, D., and Farach-Colton, M. (2000). NOTUNG: a program for dating gene duplications and optimizing gene family trees. *J. Comput. Biol.* 7, 429–447. doi: 10.1089/106652700750050871
- Chery, M., and Drouard, L. (2022). Plant tRNA functions beyond their major role in translation. *J. Exp. Botany* 74, 2352–2363. doi: 10.1093/JXB/ERAC483
- Choisne, N., Carneiro, V. T. C., Pelletier, G., and Small, I. (1998). Implication of 5'-flanking sequence elements in expression of a plant tRNA^{Leu} gene. *Plant Mol. Biol.* 36, 113–123. doi: 10.1023/A:1005988004924
- Cognat, V., Pawlak, G., Pflieger, D., and Drouard, L. (2022). PlantRNA 2.0: an updated database dedicated to tRNAs of photosynthetic eukaryotes. *Plant J.* 112, 1112–1119. doi: 10.1111/tjp.15997
- Corpet, F. (1988). Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res.* 16, 10881–10890. doi: 10.1093/NAR/16.22.10881
- Cozen, A. E., Quartley, E., Holmes, A. D., Hrabeta-Robinson, E., Phizicky, E. M., and Lowe, T. M. (2015). ARM-seq: AlkB-facilitated RNA methylation sequencing reveals a complex landscape of modified tRNA fragments. *Nat. Methods* 12, 879–884. doi: 10.1038/nmeth.3508
- Crick, F. H. (1966). Codon–anticodon pairing: the wobble hypothesis. *J. Mol. Biol.* 19, 548–555. doi: 10.1016/S0022-2836(66)80022-0

Acknowledgments

Data analysis was performed using the High-Performance Computing services of the DOST-ASTI Computing and Archiving Research Environment facility.

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.

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/fpls.2024.1486612/full#supplementary-material>

- Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004). WebLogo: a sequence logo generator. *Genome Res.* 14, 1188–1190. doi: 10.1101/GR.849004
- Darby, C. A., Stolzer, M., Ropp, P. J., Barker, D., and Durand, D. (2017). Xenolog classification. *Bioinformatics* 33, 640–649. doi: 10.1093/BIOINFORMATICS/BTW686
- Dieci, G., Yukawa, Y., Alzapiedi, M., Guffanti, E., Ferrari, R., Sugiura, M., et al. (2006). Distinct modes of TATA box utilization by the RNA polymerase III transcription machineries from budding yeast and higher plants. *Gene* 379, 12–25. doi: 10.1016/j.gene.2006.03.013
- Durand, D., Halldórsson, B. V., and Vernot, B. (2006). A hybrid micro-macroevolutionary approach to gene tree reconstruction. *J. Comput. Biol.* 13, 320–335. doi: 10.1089/cmb.2006.13.320
- Eigen, M., Lindemann, B. F., Tietze, M., Winkler-Oswatitsch, R., Dress, A., and von Haeseler, A. (1989). How old is the genetic code? Statistical geometry of tRNA provides an answer. *Science* 244, 673–679. doi: 10.1126/science.2497522
- Filailat, D. L., Ballerini, E. S., Mandáková, T., Aköz, G., Derieg, N. J., Schmutz, J., et al. (2018). The Aquilegia genome provides insight into adaptive radiation and reveals an extraordinarily polymorphic chromosome with a unique history. *eLife* 7, e36426. doi: 10.7554/eLife.36426
- Galli, G., Hofstetter, H., and Birnstiel, M. L. (1981). Two conserved sequence blocks within eukaryotic tRNA genes are major promoter elements. *Nature* 294, 626–631. doi: 10.1038/294626a0
- Giuliodori, S., Percudani, R., Braglia, P., Ferrari, R., Guffanti, E., Ottonello, S., et al. (2003). A composite upstream sequence motif potentiates tRNA gene transcription in yeast. *J. Mol. Biol.* 333, 1–20. doi: 10.1016/j.jmb.2003.08.016
- Givnish, T. J., Zuluaga, A., Spalink, D., Soto Gomez, M., Lam, V. K. Y., Saarela, J. M., et al. (2018). Monocot plastid phylogenomics, timeline, net rates of species diversification, the power of multi-gene analyses, and a functional model for the origin of monocots. *Am. J. Bot.* 105, 1888–1910. doi: 10.1002/ajb.2.1178
- 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
- Hamada, M., Huang, Y., Lowe, T. M., and Maraia, R. J. (2001). Widespread use of TATA elements in the core promoters for RNA polymerases III, II, and I in fission yeast. *Mol. Cell Biol.* 21, 6870–6881. doi: 10.1128/MCB.21.20.6870-6881.2001
- Hofstetter, H., Kressmann, A., and Birnstiel, M. L. (1981). A split promoter for a eucaryotic tRNA gene. *Cell* 24, 573–585. doi: 10.1016/0092-8674(81)90348-2
- Hummel, G., Berr, A., Graindorge, S., Cognat, V., Ubrig, E., Pflieger, D., et al. (2020). Epigenetic silencing of clustered tRNA genes in Arabidopsis. *Nucleic Acids Res.* 48, 10297–10312. doi: 10.1093/nar/gkaa766
- Institute for Theoretical Chemistry RNAfold web server. Available online at: <http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi> (Accessed January 28, 2024).
- Janssens, S. B., Couvreur, T. L. P., Mertens, A., Dauby, G., Dagallier, L. P. M. J., Abele, S. V., et al. (2020). A large-scale species level dated angiosperm phylogeny for evolutionary and ecological analyses. *Biodiversity Data J.* 8, e39677. doi: 10.3897/BDJ.8.E39677
- Katoh, K., and Toh, H. (2008). Improved accuracy of multiple ncRNA alignment by incorporating structural information into a MAFFT-based framework. *BMC Bioinf.* 9, 1–13. doi: 10.1186/1471-2105-9-212/TABLES/5
- Kishor, P. B. K., Kumari, P. H., Sunita, M. S. L., and Sreenivasulu, N. (2015). Role of proline in cell wall synthesis and plant development and its implications in plant ontogeny. *Front. Plant Sci.* 6, doi: 10.3389/FPLS.2015.00544/BIBTEX
- Kishor, P. B. K., Sangam, S., Amrutha, R. N., Laxmi, P. S., Naidu, K. R., Rao, K. R. S. S., et al. (2005). Regulation of proline biosynthesis, degradation, uptake and transport in higher plants: Its implications in plant growth and abiotic stress tolerance. *Curr. Sci.* 88, 424–438.
- Lai, D., Proctor, J. R., Zhu, J. Y. A., and Meyer, I. M. (2012). R-chic: a web server and R package for visualizing RNA secondary structures. *Nucleic Acids Res.* 40, e95. doi: 10.1093/NAR/GKS241
- Larsson, A. (2014). AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* 30, 3276–3278. doi: 10.1093/BIOINFORMATICS/BTU531
- Leitner, J., Retzer, K., Malenica, N., Bartkeviciute, R., Lucyshyn, D., Jäger, G., et al. (2015). Meta-regulation of Arabidopsis auxin responses depends on tRNA maturation. *Cell Rep.* 11, 516–526. doi: 10.1016/j.CELREP.2015.03.054
- Lescot, M., Déhais, P., Thijs, G., Marchal, K., Moreau, Y., Peer, Y. V. D., et al. (2002). PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res.* 30, 325. doi: 10.1093/NAR/30.1.325
- Lowe, T. M., and Eddy, S. R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25, 955–964. doi: 10.1093/nar/25.5.955
- Lucas, M. C., Pryszcz, L. P., Medina, R., Milenkovic, I., Camacho, N., Marchand, V., et al. (2024). Quantitative analysis of tRNA abundance and modifications by nanopore RNA sequencing. *Nat. Biotechnol.* 42, 72–86. doi: 10.1038/s41587-023-01743-6
- Madeira, F., Pearce, M., Tivey, A. R. N., Basutkar, P., Lee, J., Edbali, O., et al. (2022). Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Res.* 50, W276–W279. doi: 10.1093/NAR/GKAC240
- Maggio, A., Miyazaki, S., Veronese, P., Fujita, T., Ibeas, J. I., Damsz, B., et al. (2002). Does proline accumulation play an active role in stress-induced growth reduction? *Plant J.* 31, 699–712. doi: 10.1046/J.1365-3113.2002.01389.X
- Marechal-Drouard, L., Weil, J. H., and Dietrich, A. (1993). Transfer RNAs and Transfer RNA Genes in Plants. *Annu. Rev. Plant Biol.* 44, 13–32. doi: 10.1146/annurev.pp.44.060193.000305
- Michaud, M., Cognat, V., Duchêne, A. M., and Maréchal-Drouard, L. (2011). A global picture of tRNA genes in plant genomes. *Plant J.* 66, 80–93. doi: 10.1111/J.1365-3113.2011.04490.X
- Mitra, S., Samadder, A., Das, P., Das, S., and Chakrabarti, J. (2015). Eukaryotic tRNA paradox. *J. biomolecular structure dynamics* 33, 1–17. doi: 10.1080/07391102.2014.1003198
- Mohanta, T. K., and Bae, H. (2017). Analyses of genomic trna reveal presence of novel tRNAs in oryza sativa. *Front. Genet.* 8, doi: 10.3389/FGENE.2017.00090/FULL
- Mohanta, T. K., Mishra, A. K., Hashem, A., Abd-Allah, E. F., Khan, A. L., and Al-Harasi, A. (2020). Construction of anti-codon table of the plant kingdom and evolution of tRNA selenocysteine (tRNA^{Sec}). *BMC Genomics* 21, 804. doi: 10.1186/s12864-020-07216-3
- Mokhtar, M. M., and Allali, A. E. (2022). tRNAdb: Plant transfer RNA database. *PLoS One* 17, e0268904. doi: 10.1371/journal.pone.0268904
- Morgado, S., and Vicente, A. C. (2019). Exploring tRNA gene cluster in archaea. *Memórias do Instituto Oswaldo Cruz* 114, e180348. doi: 10.1590/0074-02760180348
- Padhiar, N. H., Katneni, U., Komar, A. A., Motorin, Y., and Kimchi-Sarfaty, C. (2024). Advances in methods for tRNA sequencing and quantification. *Trends Genet.* 40, 276–290. doi: 10.1016/j.tig.2023.11.001
- Padilla-Mejia, N. E., Florencio-Martinez, L. E., Figueroa-Angulo, E. E., Manning-Cela, R. G., Hernández-Rivas, R., Myler, P. J., et al. (2009). Gene organization and sequence analyses of transfer RNA genes in Trypanosomatid parasites. *BMC Genomics* 10, 232. doi: 10.1186/1471-2164-10-232
- Panstruga, R., and Spanu, P. (2024). Transfer RNA and ribosomal RNA fragments – emerging players in plant-microbe interactions. *New Phytol.* 241, 567–577. doi: 10.1111/nph.19409
- Park, E. J., and Kim, T. H. (2018). Fine-Tuning of Gene Expression by tRNA-Derived Fragments during Abiotic Stress Signal Transduction. *Int. J. Mol. Sci.* 19, 518. doi: 10.3390/IJMS19020518
- Patriarca, E. J., Cermola, F., D’Aniello, C., Fico, A., Guardiola, O., Cesare, D. D., et al. (2021). The multifaceted roles of proline in cell behavior. *Front. Cell Dev. Biol.* 9, 2236. doi: 10.3389/FCCELL.2021.728576/BIBTEX
- Percudani, R. (2001). Restricted wobble rules for eukaryotic genomes. *Trends Genet.* 17, 133–135. doi: 10.1016/s0168-9525(00)02208-3
- Phizicky, E. M., and Hopper, A. K. (2010). tRNA biology charges to the front. *Genes Dev.* 24, 1832–1860. doi: 10.1101/gad.1956510
- Pinkard, O., McFarland, S., Sweet, T., and Collier, J. (2020). Quantitative tRNA-sequencing uncovers metazoan tissue-specific tRNA regulation. *Nat. Commun.* 11, 4104. doi: 10.1038/s41467-020-17879-x
- Planta, J., Liang, Y.-Y., Xin, H., Chansler, M. T., Prather, L. A., Jiang, N., et al. (2022). Chromosome-scale genome assemblies and annotations for Poales species *Carex cristatella*, *Carex scoparia*, *Juncus effusus*, and *Juncus inflexus*. *G3 Genes/Genomes/Genetics* 12, jkac211. doi: 10.1093/g3journal/jkac211
- Quinlan, A. R., and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. doi: 10.1093/BIOINFORMATICS/BTQ033
- Rambaut, A. FigTree. Available online at: <http://tree.bio.ed.ac.uk/software/figtree/> (Accessed January 21, 2024).
- R Core Team (2021). R: A language and environment for statistical computing. Available online at: <https://www.R-project.org/> (Accessed December 1, 2024).
- Rienzi, S. C. D., Collingwood, D., Raghuraman, M. K., and Brewer, B. J. (2009). Fragile genomic sites are associated with origins of replication. *Genome Biol. Evol.* 1, 350–363. doi: 10.1093/GBE/EVP034
- Santos, F. B., and Del-Bem, L.-E. (2023). The evolution of tRNA copy number and repertoire in cellular life. *Genes* 14, 27. doi: 10.3390/genes14010027
- Sayers, E. W., Bolton, E. E., Brister, J. R., Canese, K., Chan, J., Comeau, D. C., et al. (2021). Database resources of the national center for biotechnology information. *Nucleic Acids Res.* 50, D20–D26. doi: 10.1093/nar/gkab1112
- Shigematsu, M., Honda, S., Lohrer, P., Telonis, A. G., Rigoutsos, I., and Kirino, Y. (2017). YAMAT-seq: an efficient method for high-throughput sequencing of mature transfer RNAs. *Nucleic Acids Res.* 45, e70. doi: 10.1093/nar/gkx005
- Soprano, A. S., Smetana, J. H. C., and Benedetti, C. E. (2018). Regulation of tRNA biogenesis in plants. Its link to plant growth and response to pathogens. *Biochim. Biophys. Acta (BBA) - Gene Regul. Mech.* 1861, 344–353. doi: 10.1016/j.BBAGRM.2017.12.004
- Stolzer, M., Lai, H., Xu, M., Sathaye, D., Vernot, B., and Durand, D. (2012). Inferring duplications, losses, transfers and incomplete lineage sorting with nonbinary species trees. *Bioinformatics* 28, i409–i415. doi: 10.1093/BIOINFORMATICS/BTS386
- Stöver, B. C., and Müller, K. F. (2010). TreeGraph 2: Combining and visualizing evidence from different phylogenetic analyses. *BMC Bioinf.* 11, 7. doi: 10.1186/1471-2105-11-7

- Tamura, K., Stecher, G., and Kumar, S. (2021). MEGA11: molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* 38, 3022–3027. doi: 10.1093/MOLBEV/MSAB120
- Tang, D. T. P., Glazov, E. A., McWilliam, S. M., Barris, W. C., and Dalrymple, B. P. (2009). Analysis of the complement and molecular evolution of tRNA genes in cow. *BMC Genomics* 10, 188. doi: 10.1186/1471-2164-10-188
- Theologis, A., Ecker, J. R., Palm, C. J., Federspiel, N. A., Kaul, S., White, O., et al. (2000). Sequence and analysis of chromosome 1 of the plant *Arabidopsis thaliana*. *Nature* 408, 816–820. doi: 10.1038/35048500
- Tourasse, N. J., and Darfeuille, F. (2020). Structural alignment and covariation analysis of RNA sequences. *Bio. Protoc.* 10, e3511. doi: 10.21769/BIOPROTOCOL.3511
- Trifinopoulos, J., Nguyen, L. T., von Haeseler, A., and Minh, B. Q. (2016). W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res.* 44, W232–W235. doi: 10.1093/NAR/GKW256
- Turowski, T. W., and Tollervey, D. (2016). Transcription by RNA polymerase III: insights into mechanism and regulation. *Biochem. Soc. Trans.* 44, 1367–1375. doi: 10.1042/BST20160062
- Usadel Lab Published plant genomes. Available online at: https://www.plabipd.de/plant_genomes_pa.ep (Accessed January 21, 2024).
- Vernot, B., Stolzer, M., Goldman, A., and Durand, D. (2007). Reconciliation with non-binary species trees. *Comput. Syst. Bioinformatics Conf.* 6, 441–452. doi: 10.1142/9781860948732_0044
- Verslues, P. E., and Sharma, S. (2010). Proline metabolism and its implications for plant-environment interaction. *Arabidopsis Book / Am. Soc. Plant Biologists* 8, e0140. doi: 10.1199/TAB.0140
- Vujanovic, S., Vujanovic, J., and Vujanovic, V. (2022). Microbiome-driven proline biogenesis in plants under stress: perspectives for balanced diet to minimize depression disorders in humans. *Microorganisms* 10, 2264. doi: 10.3390/MICROORGANISMS10112264
- Wang, C., Chen, W., Aili, M., Zhu, L., and Chen, Y. (2023). tRNA-derived small RNAs in plant response to biotic and abiotic stresses. *Front. Plant Sci.* 14. doi: 10.3389/fpls.2023.1131977
- Warren, J. M., Salinas-Giege, T., Hummel, G., Coots, N. L., Svendsen, J. M., Brown, K. C., et al. (2021). Combining tRNA sequencing methods to characterize plant tRNA expression and post-transcriptional modification. *RNA Biol.* 18, 64–78. doi: 10.1080/15476286.2020.1792089
- Wilkinson, G. N., and Rogers, C. E. (1973). Symbolic description of factorial models for analysis of variance. *J. R. Stat. Society Ser. C (Applied Statistics)* 22, 392–399. doi: 10.2307/2346786
- Yang, Y., Sun, P., Lv, L., Wang, D., Ru, D., Li, Y., et al. (2020). Prickly waterlily and rigid hornwort genomes shed light on early angiosperm evolution. *Nat. Plants* 6, 215–222. doi: 10.1038/s41477-020-0594-6
- Yukawa, Y., Akama, K., Noguchi, K., Komiya, M., and Sugiura, M. (2013). The context of transcription start site regions is crucial for transcription of a plant tRNA^{Lys} (UUU) gene group both in vitro and in vivo. *Gene* 512, 286–293. doi: 10.1016/j.gene.2012.10.022
- Yukawa, Y., Dieci, G., Alzapiedi, M., Hiraga, A., Hirai, K., Yamamoto, Y. Y., et al. (2011). A common sequence motif involved in selection of transcription start sites of *Arabidopsis* and budding yeast tRNA genes. *Genomics* 97, 166–172. doi: 10.1016/j.ygeno.2010.12.001
- Yukawa, Y., Sugita, M., Choisne, N., Small, I., and Sugiura, M. (2000). The TATA motif, the CAA motif and the poly(T) transcription termination motif are all important for transcription re-initiation on plant tRNA genes. *Plant J.* 22, 439–447. doi: 10.1046/j.1365-3113X.2000.00752.x
- Zheng, G., Qin, Y., Clark, W. C., Dai, Q., Yi, C., He, C., et al. (2015). Efficient and quantitative high-throughput tRNA sequencing. *Nat. Methods* 12, 835–837. doi: 10.1038/nmeth.3478
- Zmasek, C. M., and Eddy, S. R. (2001). ATV: display and manipulation of annotated phylogenetic trees. *Bioinformatics (Oxford, England)* 17, 383–384. doi: 10.1093/BIOINFORMATICS/17.4.383
- Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31, 3406–3415. doi: 10.1093/NAR/GKG595



OPEN ACCESS

EDITED BY

Santiago Martín-Bravo,
Universidad Pablo de Olavide, Spain

REVIEWED BY

Ricarda Riina,
CSIC, Spain
Patricio Pérez-Henríquez,
University of California, Los Angeles,
United States
Neng Wei,
Chinese Academy of Sciences (CAS), China

*CORRESPONDENCE

Božo Frajman

✉ bozo.frajman@uibk.ac.at

Michael H. J. Barfuss

✉ michael.h.j.barfuss@univie.ac.at

RECEIVED 10 January 2025

ACCEPTED 10 March 2025

PUBLISHED 14 April 2025

CITATION

Sharovikj Ivanova A, Schönschwetter P,
Kostadinovski M, Barfuss MHJ, Čušterevska R
and Frajman B (2025) Integrative taxonomy
reveals cryptic diversity within the *Euphorbia*
nicaeensis alliance (Euphorbiaceae) in the
central Balkan Peninsula.
Front. Plant Sci. 16:1558466.
doi: 10.3389/fpls.2025.1558466

COPYRIGHT

© 2025 Sharovikj Ivanova, Schönschwetter,
Kostadinovski, Barfuss, Čušterevska and
Frajman. 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.

Integrative taxonomy reveals cryptic diversity within the *Euphorbia nicaeensis* alliance (Euphorbiaceae) in the central Balkan Peninsula

Angela Sharovikj Ivanova^{1,2}, Peter Schönschwetter²,
Mitko Kostadinovski¹, Michael H. J. Barfuss^{3*},
Renata Čušterevska¹ and Božo Frajman^{2*}

¹Institute of Biology, Faculty of Natural Sciences and Mathematics, Ss. Cyril and Methodius University Skopje, Skopje, North Macedonia, ²Department of Botany, University of Innsbruck, Innsbruck, Austria, ³Department of Botany and Biodiversity Research, University of Vienna, Vienna, Austria

The Eurasian steppes are the second-largest continuous biome on Earth. *Euphorbia glareosa*, a member of the *Euphorbia nicaeensis* alliance, is a widespread species in the macroclimatically defined zonal Pontic steppes and westerly and southerly adjacent extrazonal steppe outposts determined by local climatic, topographic, and edaphic conditions. In the extrazonal steppes, in particular within the Anatolian, Danubian, Thracian, and Pannonian grasslands, it is more or less continuously distributed, but with several disjunct occurrences in the central Balkan Peninsula, which is renowned for its high biodiversity. Several (infra)specific taxa have been recognised within *E. glareosa* s.l., but relationships among them remain elusive. We applied an integrative approach ranging from cytogenetics (relative genome size and ploidy estimation, chromosome counting) and morphometrics to phylogenetics (internal ribosomal transcribed spacer sequencing and amplified length polymorphism fingerprinting), with geographic focus on the central and eastern Balkan Peninsula. We inferred multiple polyploidisations within the group and complex phylogenetic patterns. We uncovered cryptic lineages in the central Balkan Peninsula, where the description of two new species, diploid *E. balcanica*, and tetraploid *E. skopjensis* was necessary. In addition, we revealed high diversity, partly related to polyploidisations, among the populations from the eastern Balkan and Pontic steppes, likely pertaining to different species. Finally, the main phylogeographic split within *E. glareosa* is between (1) Pannonian, central and eastern Balkan populations, and (2) the easternmost Balkan, Pontic, and Anatolian populations. Our results thus highlight the outstanding conservation value of the extrazonal European steppes that are not just an outpost of zonal Eurasian steppes. We also point to the remarkable biodiversity of the central and eastern Balkan Peninsula and the need for further in-depth studies of this biodiversity hotspot.

KEYWORDS

biodiversity, Balkan Peninsula, extrazonal steppes, new species, phylogeny, polyploidy, Pontic steppes, taxonomy

1 Introduction

The Eurasian steppes are the second-largest continuous biome on Earth spanning from Central and Eastern Europe (Pannonian and Pontic areas) to Central and northeastern Asia (Lal, 2004; Wesche et al., 2016; Kirschner et al., 2020). They represent various types of temperate grasslands (Coupland, 1993), which are shaped by strongly seasonal climates and cold winters (Peart, 2008). They are similar in several characteristics to the grasslands of Mediterranean mountain ranges, and several plant genera and species extend across both biomes (Hamasha et al., 2012; Stojilković et al., 2022). The Balkan Peninsula represents a link between extensive steppe areas in the East, and the Mediterranean Basin in the South and West, and thus a crossroad of floras from both areas (Tomović et al., 2014). With the exception of the northeasternmost Balkan Peninsula (Dobrudja), which is considered a part of the macroclimatically defined Eurasian zonal steppes, extrazonal steppe outposts determined by local climatic, topographic, and edaphic conditions are disjunctly distributed in the eastern and southern parts of the Balkans. Within the extrazonal steppes of the Balkan Peninsula, the Danubian and Thracian lowlands are covered with wide-ranging steppe vegetation, whereas smaller grassland areas embedded in a matrix of forest vegetation are characteristic of westerly adjacent areas (Wesche et al., 2016; Kirschner et al., 2020). The eastern and southern Balkans are also the region that, compared to the western and central parts, remained largely neglected in phylogenetic studies (Španiel and Rešetnik, 2022); the few existing studies dedicated to or including plants from these areas point to high intra- as well as interspecific diversity, e.g., in *Astragalus onobrychis* L. (Fabaceae; Závěská et al., 2019), *Aurinia saxatilis* (L.) Desv. (Brassicaceae; Rešetnik et al., 2022), *Cerastium decalvans* Schloss. & Vuk. (Caryophyllaceae; Niketić et al., 2022), *Cyanus tuberosus* (Vis.) Soják group (Asteraceae; Skokanová et al., 2019), *Sesleria rigida* Heuff. complex (Poaceae; Kuzmanović et al., 2013), and *Silene saxifraga* alliance (Caryophyllaceae; Đurović et al., 2017). In addition, description of several new species from the central parts of the Balkan Peninsula in the last decades (reviewed in Frajman et al., 2014) points to a lack of contemporary biodiversity studies in this area.

One of the widespread steppe species, more or less continuously distributed in the Pontic and Pannonian, as well as the Danubian and Thracian grasslands, with several disjunct occurrences in the central Balkan Peninsula (Bosnia and Herzegovina, North Macedonia, Serbia), as well as in Anatolia and the Caucasus, is *Euphorbia glareosa* Pall. ex M. Bieb. This species was considered conspecific with the Mediterranean *E. nicaeensis* All. by Radcliffe-Smith and Tutin (1968), but Stojilković et al. (2022) and Boschin et al. (2024) have shown that these taxa are clearly divergent, and three species should be recognised in the Mediterranean Basin as follows: the western Mediterranean *E. nicaeensis*, the central Mediterranean (Apennine Peninsula and north-western Balkan Peninsula) *E. japygica* Ten., and the western Balkan endemic *E. hercegovina* Beck. They all, along with *E. glareosa* and some species

distributed in the Irano-Turanian region, belong to the *E. nicaeensis* alliance (Stojilković et al., 2022).

On the other hand, *E. glareosa* was shown to constitute an assemblage of populations with different genome sizes and putatively different ploidies (Stojilković et al., 2022). This, along with pronounced morphological variability that led to the recognition of different taxa in the past (e.g., Prokhanov, 1949; Kuzmanov, 1979; Greuter et al., 1986; Geltman, 2009), suggests that *E. glareosa* is a species complex including multiple taxa rather than a single species. Stojilković et al. (2022) treated this assemblage as *E. glareosa* s.l., which includes seven taxa with unclear taxonomic status and relationships as follows: *E. cadrilateri* Prodan, *E. dobrogensis* Prodan, *E. glareosa*, *E. goldei* Prokh., *E. pannonica* Host., *E. stepposa* Zoz, and *E. volgensis* Krysth. These taxa were in the past partly treated as subspecies (e.g., Kuzmanov, 1979; Radcliffe-Smith, 1982; Greuter et al., 1986; Govaerts et al., 2000) or species (e.g., Prodan, 1936; Prokhanov, 1949; Geltman, 2009, 2020). We apply the name *E. glareosa* s.l. for all these taxa hereafter.

Although most populations of *E. glareosa* s.l. formed a monophyletic group in the phylogenetic trees based on the restriction site-associated DNA sequencing (RADseq) data, one population from the Skopje basin in North Macedonia was phylogenetically divergent. It appeared more closely related to the Mediterranean species of the *E. nicaeensis* alliance and Irano-Turanian *E. macroclada* Boiss. (Stojilković et al., 2022). This population had the highest relative genome size (RGS) of all studied populations, indicating its polyploid origin, which was assumed to explain its divergent phylogenetic position (Stojilković et al., 2022). In addition, within the *E. glareosa* clade, two populations from (1) Dobrudja and (2) Armenia, with higher genome size than most populations of *E. glareosa*, were phylogenetically divergent and sister to the other populations with weak support (Stojilković et al., 2022). Finally, isolated populations from northern Albania that were treated as *E. nicaeensis* (incl. *E. glareosa* s.l.) by Qosja et al. (1992) and Barina (2017) were not studied by Stojilković et al. (2022), who suggested that they might be more closely related to *E. japygica* than to *E. glareosa* s.l.

Given the high phylogenetic, RGS, and morphological variability within *E. glareosa* s.l., our aim is to disentangle the relationships among the populations currently classified as *E. glareosa* s.l., with geographic focus on the central and eastern Balkan Peninsula, including populations from north Albania and based on an extensive sampling in Bulgaria, North Macedonia, and Serbia. We sampled multiple populations that could pertain to different taxa of *E. glareosa* s.l. across this area and used an integrative approach to infer the relationships among them and to clarify their taxonomic status. More specifically, we (1) estimated the ploidy of all investigated populations via relative genome size (RGS) estimation with flow cytometry and chromosome counting of selected populations. (2) Using nuclear ITS sequences and amplified fragment length polymorphism (AFLP) fingerprinting, we inferred the origin of, and the phylogeographic differentiation within, the study species. Finally, (3) using multivariate morphometrics, we explored the

morphological differentiation, and (4) based on all data, we propose a revised taxonomic treatment. Since RGS and AFLP analyses revealed clear divergence of populations from the central Balkan Peninsula (Albania, Kosovo, North Macedonia) that were also morphologically divergent from *E. glareosa* s.l. (see Results), we describe them below as two new taxa, diploid *E. balcanica* and tetraploid *E. skopjensis*. For simplicity, we apply these names hereafter.

2 Materials and methods

2.1 Plant material

We collected plant material from *E. balcanica*, *E. skopjensis*, *E. glareosa* s.l., and outgroup taxa (leaves dried in silica gel, herbarium vouchers, and seeds) for molecular, morphometric, and karyological (RGS and chromosome number estimation) analyses in the field between 2006 and 2024. We studied a total of 11 populations of *E. balcanica*, 4 of *E. skopjensis*, and 62 of *E. glareosa* s.l. In total, 5 populations of *E. balcanica* and 3 of *E. skopjensis* were included in AFLP analyses, 4/2 in ITS analyses, all in RGS analyses, and 10/4 in morphometric analyses. For population 95 of *E. balcanica* and 105 of *E. skopjensis*, we counted the chromosomes. For *E. glareosa* s.l., 32 populations were included in ITS (16 from Stojilković et al., 2022), 51 (18) in RGS analyses, 37 in AFLP, and 83 (16) in morphometric

analyses, respectively, and for population 118, we counted the chromosomes (Figure 1; Supplementary Table S1).

2.2 Relative genome size and ploidy-level estimation

We measured the RGS using a CyFlow space flow cytometer (Partec, GmbH, Münster, Germany) following the procedure of Suda and Trávníček (2006). Nuclei from silica gel-dried leaf material of our samples (Supplementary Table S1) and fresh leaves of the reference standard *Bellis perennis* L. ($2C = 3.38$ pg; Schönswetter et al., 2007) were stained using 4',6-diamidino-2-phenylindole (DAPI). In cases where the peaks of the reference standard and the sample overlapped, *Pisum sativum* L. ($2C = 8.84$ pg; Greilhuber and Ebert, 1994) was used as secondary standard. In such cases, we first measured the RGS of both standards together, and then that of the secondary standard and the sample. We then recalculated the RGS of the sample and displayed it in relation to the primary standard. We recorded the relative fluorescence of 3,000 nuclei for three to five individuals per population and used FloMax software (Partec) to evaluate histograms and calculate coefficients of variation. The RGS was calculated as the ratio between the values of the mean relative fluorescence of the sample and the standard. We used the visualisation package “ggplot2” in RStudio 1.2.5019 (RStudio Team, 2022, version R-3.6.1) to produce box plots of RGS. In addition to *E. balcanica*, *E. skopjensis*, and *E.*

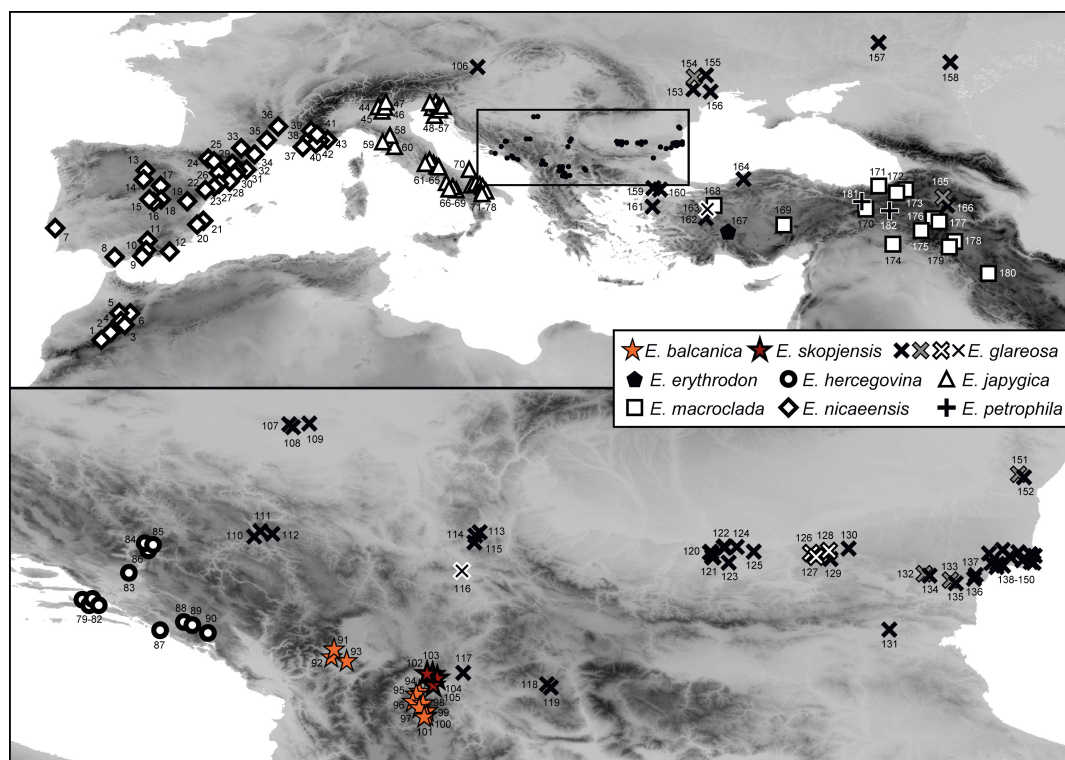


FIGURE 1

Populations of *Euphorbia balcanica* (populations 91–101), *E. skopjensis* (102–105), and other species from the *E. nicaeensis* alliance used in this study. The area within the frame in the upper panel is enlarged in the lower panel. Population numbers correspond to Supplementary Table S1. Different symbols of *E. glareosa* s.l. represent populations with different relative genome size/ploidy.

glareosa s.l., we included the RGS data of closely related *E. erythron* Boiss. & Heldr., *E. hercegovina*, *E. japygica*, *E. macroclada*, *E. nicaeensis*, and *E. petrophila* C. A. Mey. from Stojilković et al. (2022) and Boschín et al. (2024).

2.3 Chromosome number estimation

Seeds collected in the field were incubated at 4°C for 2 weeks. After removal of the caruncle, the seeds were sterilised with a mixture of bleach and water 1:3 for 10 min to inhibit fungal growth and then incubated on sterile paper in Petri dishes. Root tips of germinated seeds were pretreated with 0.002 M colchicine for 2 h at room temperature and then for 2 h at 4°C, fixed in Carnoy's solution (3:1 mixture of ethanol and acetic acid) for 24 h at 4°C and then stored in ethanol at −21°C.

Hydrolysis was performed in 5 N HCl for 60 min at room temperature. The tips were stained with Feulgen's reagent, kept in the dark for 2 h at room temperature and then rinsed with water. Slides were prepared by squashing the stained meristem in a drop of 45% acetic acid under the coverslip. Slides were then snap frozen, dehydrated in 96% ethanol for 5 min and air dried. Chromosomes were counted using a Nikon Eclipse 80i microscope; images were taken using a Canon 600D camera and processed using Canon EOS Utility software.

2.4 DNA extraction, ITS sequencing, and analysis of sequence data

Total genomic DNA extraction and ITS sequencing were performed as described by Frajman and Schönschetter (2011). Sequencing was performed at Eurofins Genomics (Ebersberg, Germany). Contigs were assembled, edited, and sequences aligned using Geneious Pro 5.5.9 (Kearse et al., 2012). Base polymorphisms were coded using NC-IUPAC ambiguity codes. Twenty-one ITS sequences were produced in this study. In addition, 60 ITS sequences of the outgroup taxa and 16 of *E. glareosa* s.l., as well as 1 of *E. skopjensis* were from Stojilković et al. (2022), and 8 of *E. japygica* were from Boschín et al. (2024). Genbank numbers are in Supplementary Table S1. Maximum parsimony (MP) and MP bootstrap (MPB) analyses were performed using PAUP v4.0b10 (Swofford, 2002) as described by Frajman et al. (2019). Bayesian analyses were performed using MrBayes 3.2.1 (Ronquist et al., 2012) using the HKY+Γ substitution model and settings such as in Frajman et al. (2019). We also created a NeighbourNet with ITS sequences of *E. balcanica*, *E. skopjensis*, and the most closely related taxa using SplitsTree4 12.3 (Huson and Bryant, 2006).

2.5 AFLP analyses

The AFLP procedure followed that of Vos et al. (1995) with modifications described by Cresti et al. (2019). In addition to 5 populations of *E. balcanica*, 3 of *E. skopjensis*, and 34 of *E. glareosa*

s.l., we included 1 population of *E. erythron*, 2 of *E. hercegovina*, 2 of *E. japygica*, 2 of *E. macroclada*, 2 of *E. nicaeensis*, and 1 of *E. petrophila* as outgroups, based on the study of Stojilković et al. (2022).

The three primers for selective PCR (fluorescent dye in brackets) were EcoRI (FAM)-ATG/MseI-CTT, EcoRI (VIC)-ACG/MseI-CAA, and EcoRI (NED)-ACC/MseI-CAG. Two microliters of the elution product was mixed with 10 µl of formamide and 0.1 µl of GeneScan ROX (ThermoFisher Scientific) and run on a 3130xl Genetic Analyzer (Applied Biosystems). A blank (DNA replaced by water) was included to test for systematic contamination, and 16 samples were used as replicates between the two PCR batches to evaluate the reproducibility of the method.

Electropherograms were analysed with Peak Scanner 1.0 (Applied Biosystems) using default peak detection parameters. Automated binning and scoring of AFLP fragments was performed using RawGeno 2.0-1 (Arrigo et al., 2009) for RStudio 2022.12.0 + 353 (RStudio Team, 2022) with the following settings: scoring range 75–400 bp, minimum intensity 100 RFUs, minimum bin width 1 bp, and maximum bin width 1.5 bp. Fragments with reproducibility <80% based on sample-replicate comparisons were eliminated. The error rate calculated in RawGeno based on 16 sample-replicate comparisons was 3.5%. Finally, after exclusion of nine individuals that failed to produce reliable fingerprints, a matrix of 167 individuals was generated. In addition, we produced a dataset of 114 individuals, including only *E. glareosa* s.l., using the same settings as above; in this case, the error rate was 2.9%.

A neighbour-joining (NJ) analysis based on Nei–Li genetic distances (Nei and Li, 1979) was performed and bootstrapped (2,000 pseudo-replicates) with TREECON 1.3b (Van de Peer and de Wachter, 1997) for both AFLP datasets. In the first dataset, we used *E. nicaeensis* for rooting, and, in the second dataset, the population 165 of *E. glareosa* that was the most early divergent in the NJ tree of the first dataset (see results). In addition, for the second dataset, including all populations of *E. glareosa* s.l., we used SplitsTree4 12.3 (Huson and Bryant, 2006) to create a NeighbourNet based on uncorrected P distances, as well as non-hierarchical K-means clustering (Hartigan and Wong, 1979) with a script by Arrigo et al. (2010) in RStudio 2022.12.0 + 353 (RStudio Team, 2022). In the last analysis 50,000 independent runs were performed (i.e., starting from random points) for each assumed value of K clusters ranging from 2 to 10. To select the best number of groups, the strategy proposed by Evanno et al. (2005) was used, and the proportions of individuals assigned to the K-means groups were plotted for populations on a map in ArcMAP 10.8.2 (ESRI, 2021).

2.6 Morphometric analyses

We performed morphometric analyses on 28 individuals from 10 populations of *E. balcanica* and 24 individuals from 4 populations of *E. skopjensis* as well as 83 individuals from 47 populations of *E. glareosa* s.l. In total, we measured 33 characters and calculated 15 ratios (Supplementary Table S2). Stem and leaf

characters were measured manually. All other characters (cyathium, fruit, and seed characters) were measured on images taken with an Olympus SZX9 stereomicroscope (Olympus GmbH, Hamburg, Germany) using the Olympus image analysis software analySIS pro. Our measurements were supplemented by morphometric data of 15 individuals from 15 populations of *E. glareosa* s.l., 20 individuals from 13 populations of *E. hercegovina*, and 1 individual of *E. skopjensis* from Stojilković et al. (2022), as well as 45 individuals from 33 populations of *E. japygica* from Bosch et al. (2024). A total of 14 individuals of *E. balcanica*, 3 of *E. glareosa* s.l., and 15 of *E. hercegovina* did not have fully developed fruits and seeds.

Statistical analyses were performed using SPSS 24.0 (IBM Corp., Armonk, NY), separately for (1) vegetative parts of the plants and cyathium characters, and (2) fruit and seed characters for (A) all abovementioned taxa as well as (B) for *E. balcanica* and *E. skopjensis*. Correlations between metric characters were tested using Pearson and Spearman correlation coefficients. In the vegetative dataset of all taxa (1A), the correlation coefficients exceeded 0.82 for the following character pairs: length of a middle stem leaf–distance from base to widest part of stem leaf, length of a ray leaf–distance from the base to the widest part of a ray leaf, depth of gland emargination–ratio depth of gland emargination/length of cyathial gland, and length of the longest terminal ray–length of the longest fertile axillary ray. In the vegetative dataset of *E. balcanica* and *E. skopjensis* (1B), the correlation coefficients exceeded 0.82 for the following character pairs: length of a middle stem leaf–distance from base to widest part of stem leaf, length of cyathial gland–ratio of length/width of a cyathial gland, length of the longest terminal ray–width of a raylet leaf, length of a raylet leaf–width of a raylet leaf, depth of gland emargination–ratio depth of gland emargination/length of cyathial gland, length of the longest terminal ray–length of a raylet leaf, length of a ray leaf–length of a raylet leaf, length of a ray leaf–width of a raylet leaf. The latter character of each character pair in all cases was excluded from further analyses. In the fruit and seed character sets, the correlation coefficients were lower than 0.82; therefore, all characters were kept for further analyses.

After standardisation to 0 mean and 1-U variance, principal component analysis (PCA) was performed. This was followed by discriminant analysis (DA). We also produced BoxPlots for the most differentiating characters. Finally, based on the morphometric data, we created taxon descriptions and an identification key. The metric values shown correspond to the 10th and 90th percentiles supplemented by extreme values in parentheses.

3 Results

3.1 Relative genome size and chromosome number

The RGS of *E. balcanica* was between 0.895 and 0.953 and that of *E. skopjensis* between 1.771 and 1.830 (Supplementary Table S1; Figure 2A). The first range corresponded to diploids with $2n = 18$

(population 95; Figure 2B) and the second to tetraploids with $2n = 36$ (population 105; Figure 2C). Within *E. glareosa* s.l., multiple RGS values were detected. The majority (41) of populations had RGS between 0.683 and 0.731. Among them was also population 118, for which we counted 18 chromosomes (Figure 2D). Populations 116 from Serbia and 162 from Turkey had RGS 1.401 and 1.436, respectively, which likely corresponds to tetraploids. In addition, several populations had deviating RGS values: those of populations 126–128 ranged between 0.592 and 0.601, those of populations 132, 151, and 165 between 1.047 and 1.146, and populations 133 and 154 had RGS 1.226 and 1.257, respectively (Figures 1, 2; Supplementary Table S1).

3.2 Phylogenetic relationships based on internal transcribed spacer sequences

The relationships in the ITS tree were poorly resolved (Figure 3A). Most accessions of *E. glareosa* s.l. were in a basal polytomy in a poorly supported clade (posterior probabilities, PP 0.61), including species of the *E. nicaeensis* alliance (sensu Stojilković et al., 2022). In this polytomy, there was another poorly supported clade (PP 0.62), including *E. balcanica*, *E. hercegovina*, *E. japygica*, *E. skopjensis*, as well as a clade (PP 0.99) with *E. erythron* and *E. macroclada*. In the NeighborNet (Figure 3B), *E. balcanica* and *E. skopjensis* had a central position, from which four main splits led to (1) *E. glareosa* s.l., (2) *E. erythron* and *E. macroclada*, (3) *E. nicaeensis*, and (4) *E. hercegovina*, *E. japygica*, and *E. nicaeensis*.

3.3 Phylogenetic relationships based on amplified fragment length polymorphisms

A total of 553 fragments were scored in 167 individuals of the complete dataset and 421 fragments in *E. glareosa* s.l. Of these, 59 and 45 fragments were excluded because they were present or absent in a single individual only. In the neighbour-joining (NJ) tree (Supplementary Figure S1; Figure 4), *E. hercegovina* and *E. japygica* formed a cluster (bootstrap support, BS 92%) sister to a cluster (BS 93%) containing all other accessions with the exception of *E. nicaeensis* that was used for rooting. In the latter cluster, *E. balcanica* and *E. skopjensis* (BS 96%) were sister to a cluster (51%) with the remaining taxa, corresponding to a cluster (57%) composed of *E. erythron*, *E. macroclada*, and *E. petrophila*, and a cluster (BS 68%) containing all accessions of *E. glareosa* s.l. The relationships between *E. balcanica* and *E. skopjensis* were unresolved as two individuals from population 103 of *E. skopjensis* were divergent from all other accessions that formed a cluster with BS 75%. The geographically disjunct populations of *E. balcanica* from Albania and Kosovo (BS 71%) were divergent from those of North Macedonia (BS <50%).

Within *E. glareosa* s.l., the relationships among the populations were mostly poorly resolved, and in general, only the terminal clusters, including individuals from the same populations, had high

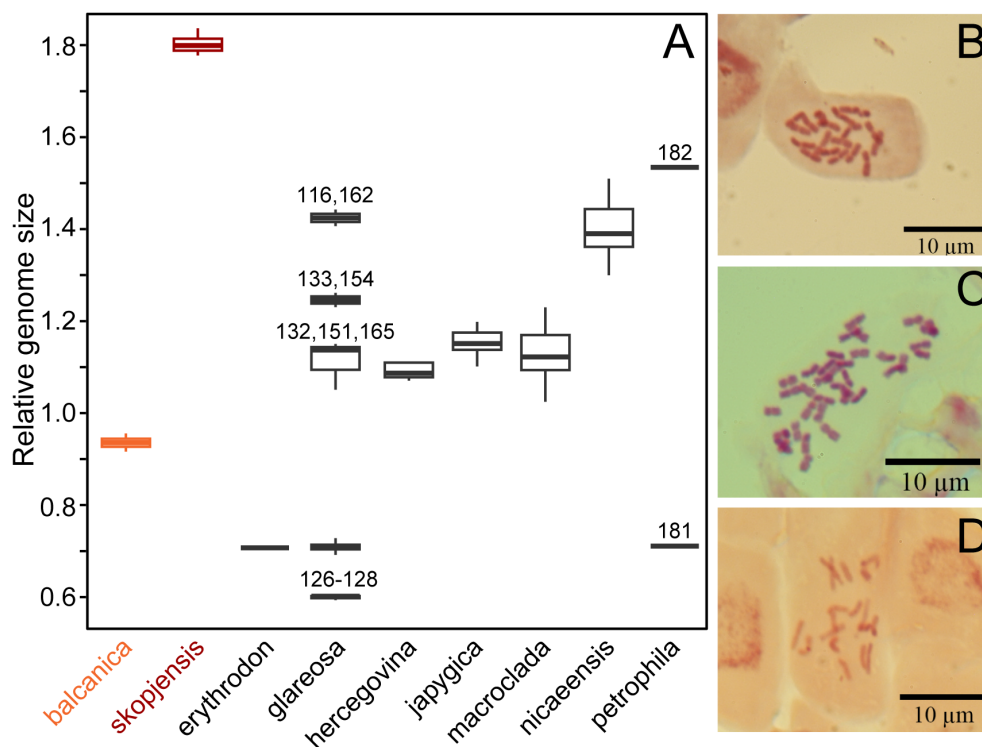


FIGURE 2

(A) Relative genome size (RGS) variation in *Euphorbia balcanica* and *E. skopjensis*, as well as other species from the *E. nicaeensis* alliance. Metaphase chromosome plates of (B) *E. balcanica* from population 95 with $2n = 18$, (C) *E. skopjensis* from population 105 with $2n = 36$, and (D) *E. glareosa* from population 118 with $2n = 18$.

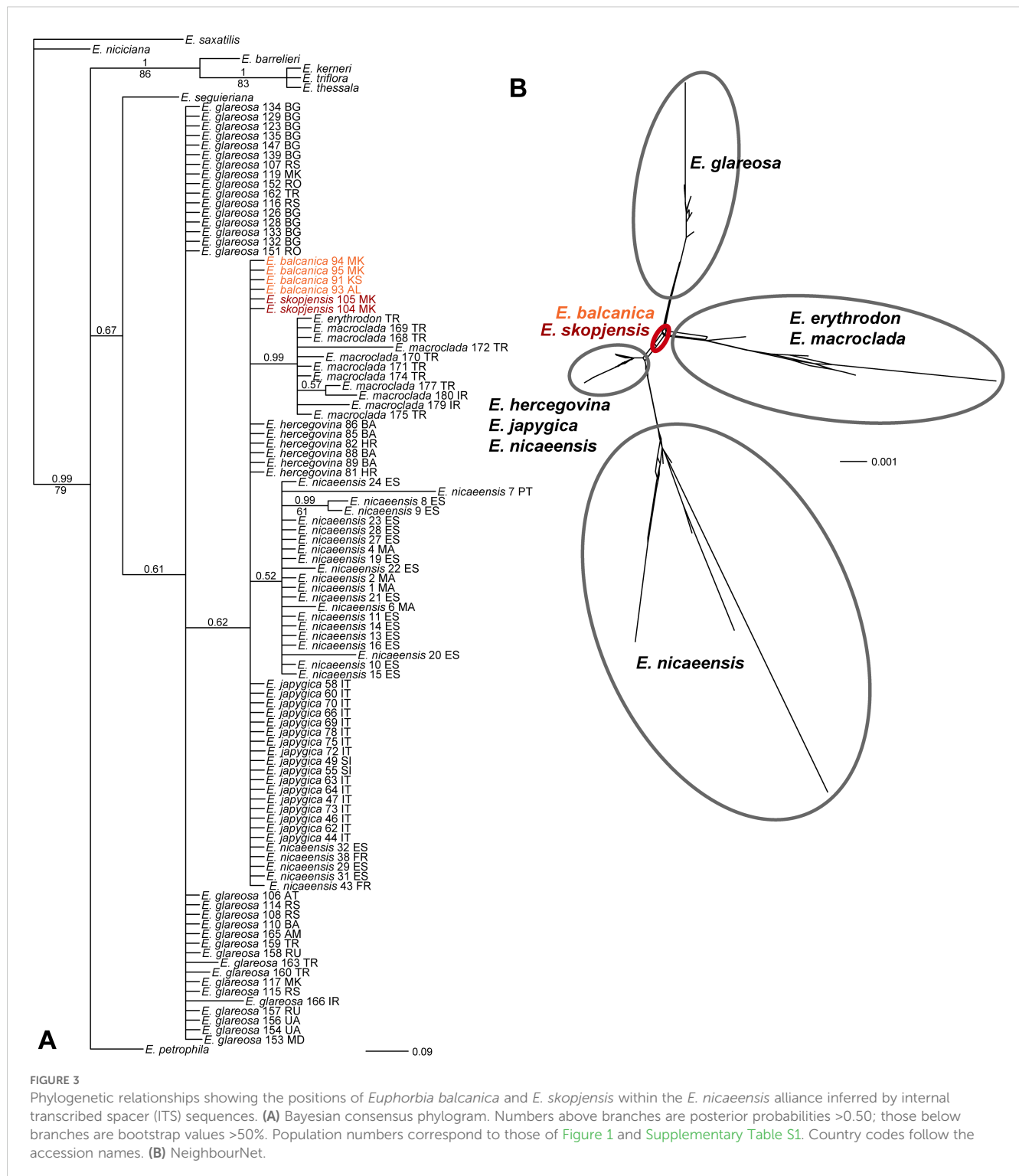
BS (Supplementary Figure S2). The most divergent was population 165 from Armenia (BS 100%), followed by population 151 from Romania, although with BS <50%; these two populations also had deviating RGS. In addition, the single cluster with high BS (BS 99%) included populations 126–128 from Bulgaria that all had the lowest RGS. The K-means analysis (Figure 5A) showed the best separation into two groups: the western populations from the Pannonian Basin and the central Balkan Peninsula (Western Group) were separated from the eastern populations from the Thracian and Pontic steppes as well as from Turkey and Armenia (Eastern Group). At $K = 3$ and $K = 4$, only the Eastern group was further divided into subclusters, although without any clear pattern. This structure was also reflected in the NeighbourNet (Figure 5B).

3.4 Morphological differentiation

For vegetative and cyathium characters, the principal component analysis (PCA) scatter plot, including all taxa (first three components explaining 26.2%, 11.5%, and 10.1% of the total variation), showed a weak separation trend of *E. hercegovina* and *E. skopjensis* along the first component, whereas *E. balcanica*, *E. glareosa* s.l., and *E. japygica* were intermediate, with considerable overlap among them (Figure 6A). The characters that contributed most to the separation along the first component were length of a

ray leaf, length of a raylet leaf, distance from the base to the widest part of a raylet leaf, stem width, and number of branchings of the longest terminal ray. The discriminant analysis (DA; Figure 6B) showed a similar pattern, although with a clearer differentiation of *E. hercegovina* and *E. skopjensis* from the other taxa along the first discriminant axis (42.3% of the total variation). The characters that contributed most to this separation were stem length, stem width, length of a middle stem leaf, length of a raylet leaf, width of a raylet leaf, ratio distance from the base to the widest part of a raylet leaf/length of a raylet leaf, and length of cyathial involucre. Along the second discriminant axis (35.9%), there was a separation trend of *E. glareosa* s.l. from *E. japygica* and *E. balcanica* that overlapped considerably. The characters that contributed most to this separation were length of a middle stem leaf, length of a raylet leaf, distance from the base to the widest part of a raylet leaf, ratio distance from the base to the widest part of a stem leaf/length of a stem leaf, ratio length/width of cyathial involucre, and number of branchings of the longest terminal ray.

Classificatory discriminant analysis classified 180 out of 200 individuals (90%) to the correct group. Of 20 incorrectly classified individuals, one (out of 24) of *E. skopjensis* was classified as *E. glareosa* s.l., and one as *E. japygica*; three individuals (out of 28) of *E. balcanica* were classified as *E. glareosa* s.l., and three as *E. japygica*; four individuals (out of 83) of *E. glareosa* s.l. were classified as *E. japygica*, two as *E. balcanica*, and one as *E. skopjensis*; one individual (out of 20) of *E. hercegovina* was



classified as *E. balcanica*; two (out of 45) individuals of *E. japygica* were classified as *E. balcanica* and two as *E. hercegovina*.

For fruit and seed characters, the PCA scatter plot including all taxa (first three components explaining 23.3%, 21.4%, and 12.5% of the total variation) showed a strong overlap among the taxa (Figure 6C). *Euphorbia balcanica* and *E. skopjensis* overlapped with *E. japygica*, whereas there was a separation trend of *E. glareosa* s.l. from *E. balcanica* and *E. skopjensis* along the second

principal component. The characters that contributed most to the separation along this component were fruit width, caruncle length and width, and seed width. The DA (Figure 6D) showed a similar pattern, although with a clearer pattern of differentiation of *E. glareosa* s.l. from other taxa along the first discriminant axis (47.5%). The characters that contributed most to this separation were fruit length, fruit width, ratio of fruit length/fruit width, seed width, and caruncle length. Along the second discriminant axis

(21.7%), there was a separation trend of *E. balcanica* from the other taxa, and of *E. hercegovina* from *E. skopjensis*. The characters that contributed most to this separation were distance from the base to the widest part of a seed, fruit width, fruit length, ratio distance from the base to the widest part of a seed/seed length, and seed width.

Classificatory discriminant analysis classified 116 out of 129 individuals (90%) to the correct group. Of 13 incorrectly classified individuals, one (out of 14) of *E. balcanica* was classified as *E. skopjensis*; two individuals (out of 21) of *E. skopjensis* were classified as *E. japygica* and one as *E. balcanica*; two individuals (out of 44) of

E. glareosa s.l. were classified as *E. japygica*, one as *E. skopjensis*, and one as *E. hercegovina*; one individual (out of 5) of *E. hercegovina* was classified as *E. japygica*; two (out of 45) individuals of *E. japygica* were classified as *E. hercegovina*, one as *E. balcanica*, and one as *E. skopjensis*.

For vegetative and cyathium characters, the PCA scatter plot, including only *E. balcanica* and *E. skopjensis* (first three components explaining 19.3%, 15.4%, and 12.1% of the total variation), showed a clear separation between them along the first component (Figure 7A). An exception was one individual from

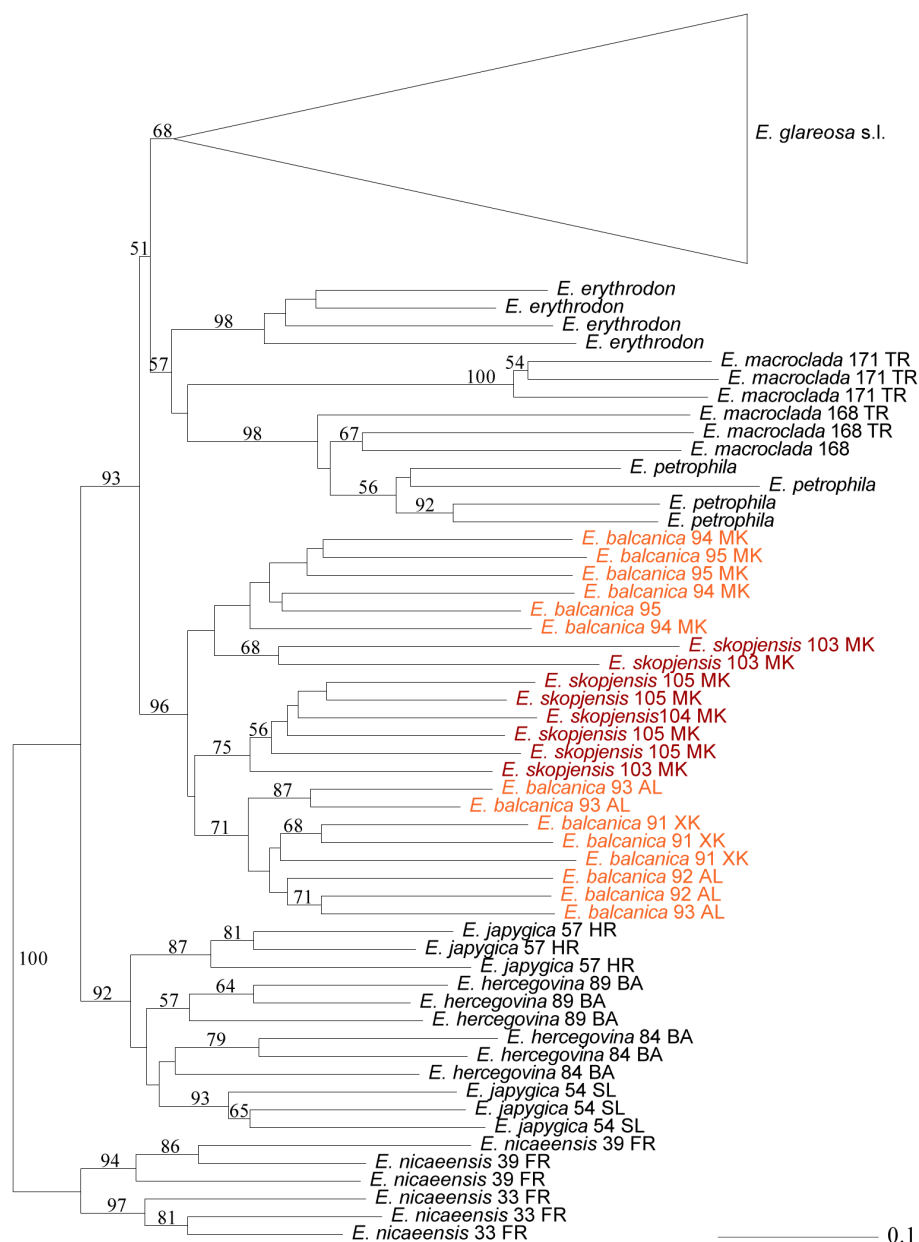


FIGURE 4

Neighbour Joining tree based on AFLP data showing the phylogenetic position of *Euphorbia balcanica* and *E. skopjensis* within the *E. nicaeensis* alliance. Numbers above branches are bootstrap values >50. Population numbers correspond to those of Figure 1 and Supplementary Table S1. The complete tree with all accessions of *E. glareosa* s.l. is in Supplementary Figure S1.

population 102 of *E. skopjensis* that was positioned among samples of *E. balcanica*. The characters that contributed most to this separation were length of (the longest) terminal ray, length of a ray leaf, length of (the longest) fertile axillary ray, and distance from

the base to the widest part of a ray leaf. In addition, the DA showed a clear differentiation between both species. The characters that contributed most to this separation were length of (the longest) terminal ray, length of a middle stem leaf, stem length, width of a

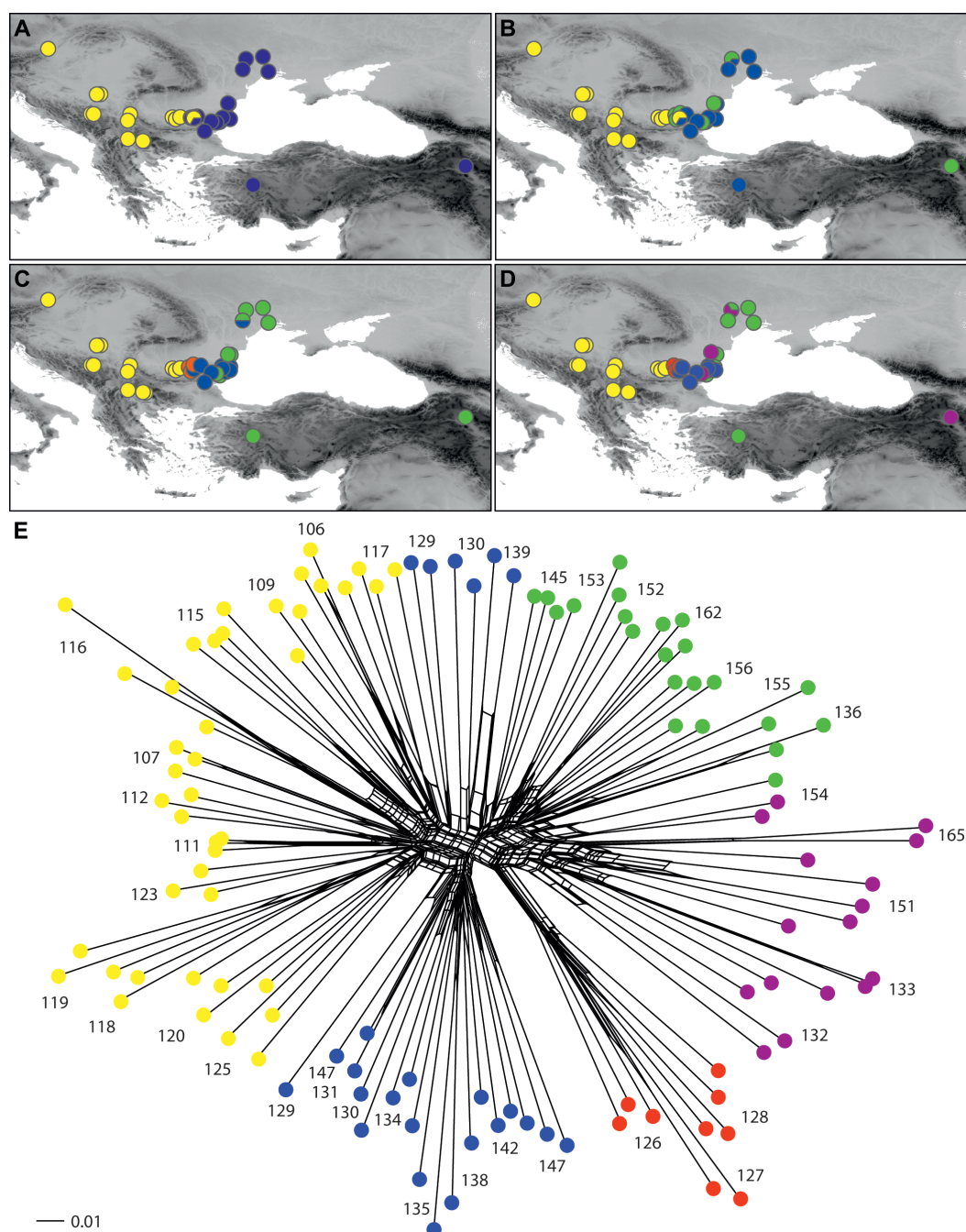


FIGURE 5

Phylogenetic relationships among populations of *Euphorbia glareosa* s.l. inferred by AFLP fingerprinting. (A–D) Geographical distribution of phylogroups inferred by non-hierarchical K-means clustering at K = 2 (optimal division), K = 3, K = 4, and K = 5. (E) NeighbourNet based on uncorrected P distances; coloured dots indicate the five groups inferred at K = 5 (as in D). Population numbers correspond to those of Figure 1 and Supplementary Table S1.

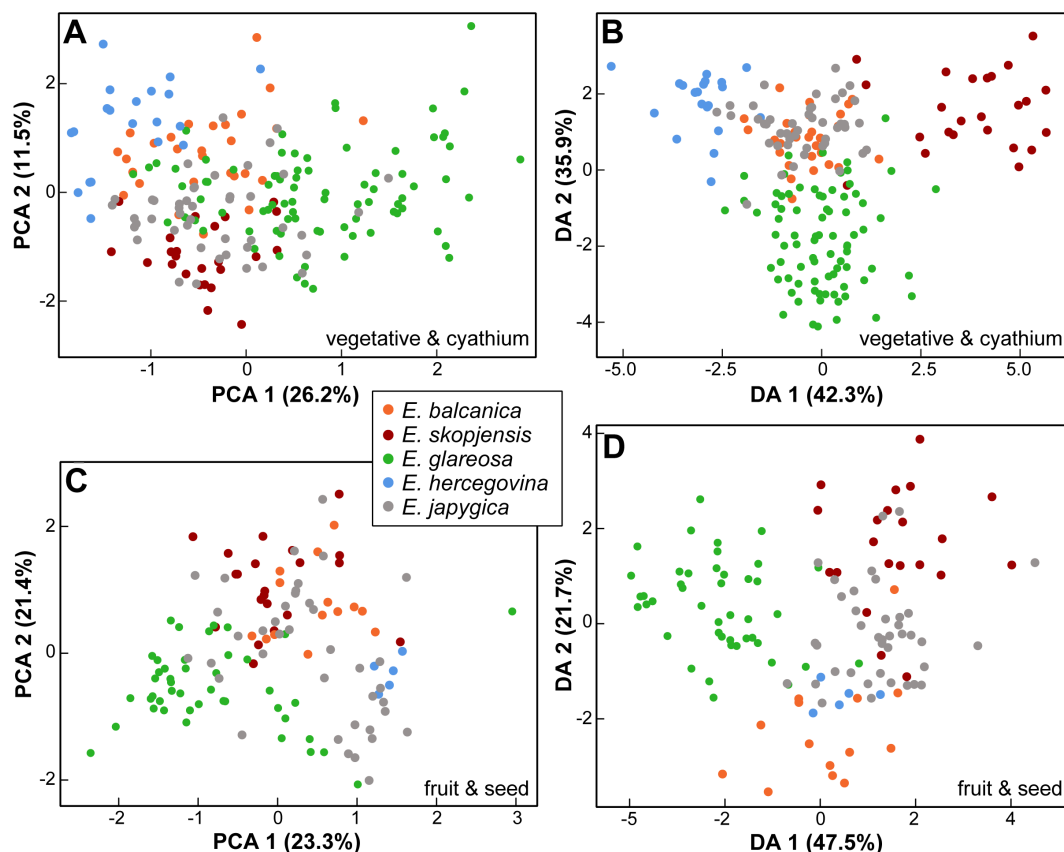


FIGURE 6

Morphological differentiation among *Euphorbia balcanica*, *E. skopjensis*, *E. glareosa* s.l., *E. hercegovina*, and *E. japygica*. (A) Principal component analysis (PCA) and (B) discriminant analyses (DA) based on vegetative and cyathium characters; (C) PCA and (D) DA based on fruit and seed characters.

raylet leaf, length of a raylet leaf, and the ratio length of cyathial involucre/width of cyathial involucre. Classificatory discriminant analysis classified all individuals to the correct group.

For fruit and seed characters, the PCA scatter plot, including only *E. balcanica* and *E. skopjensis* (first three components explaining 22.9%, 15.1%, and 14.0% of the total variation), showed a separation trend between them along the first component, although with considerable overlap (Figure 7C). The characters that contributed most to this separation were caruncle length, ratio of caruncle length/caruncle width, and ratio of seed length/seed width. In addition, the DA showed a clear differentiation between both species. The characters that contributed most to this separation were ratio of seed length/seed width, seed width, fruit width, and caruncle length. Classificatory discriminant analysis classified only a single individual of *E. skopjensis* incorrectly. Variation of selected characters that discriminate between *E. balcanica* and *E. skopjensis* as well as between them and closely related taxa is presented in boxplots (Supplementary Figure S3).

4 Discussion

4.1 Cryptic diversity in the central Balkan Peninsula and description of two new species

Our integrative taxonomic approach applying an array of complementary methods revealed cryptic diversity in the central part of the Balkan Peninsula rendering the description of two new species, *E. balcanica* and *E. skopjensis*, necessary. Divergence of a putatively polyploid population of *E. glareosa* s.l. from the Skopje Basin (North Macedonia) was previously revealed based on RADseq data (Stojilković et al., 2022). Our AFLP analyses (Figure 4) with extended geographic sampling confirmed that diploid populations from Albania, Kosovo, and North Macedonia, pertaining to *E. balcanica*, and tetraploid populations from Skopje Basin in North Macedonia, pertaining to *E. skopjensis*, form a lineage differentiated from all other constituents of the *E. nicaeensis* alliance.

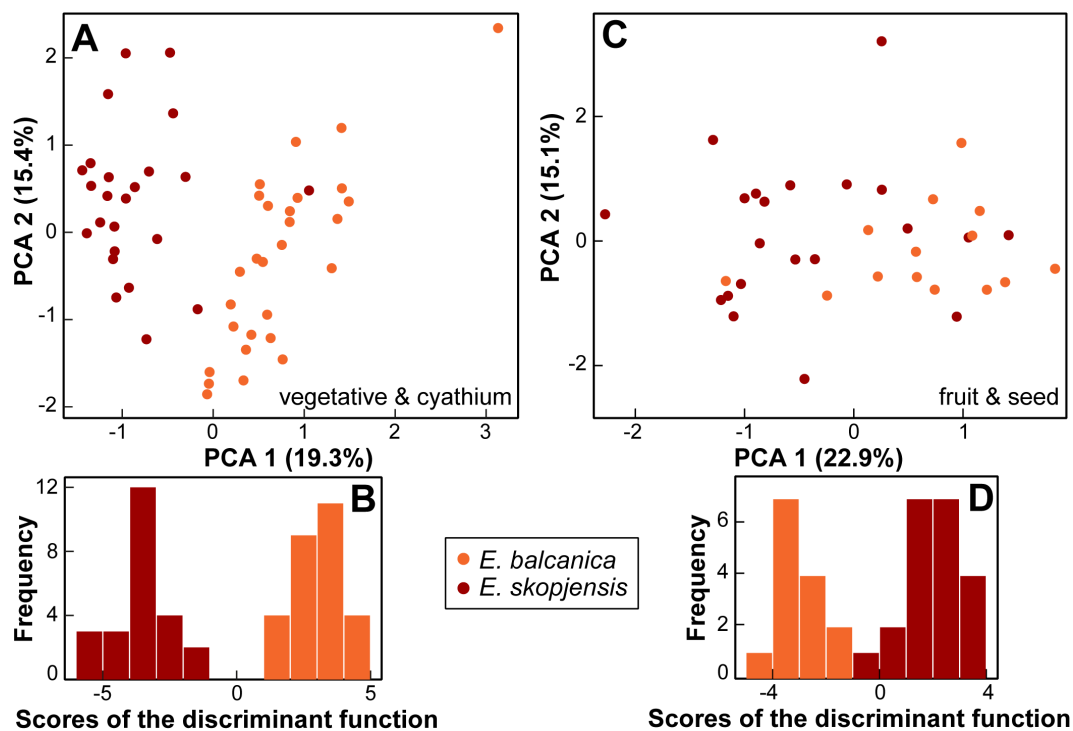


FIGURE 7

Morphological differentiation between *Euphorbia balcanica* and *E. skopjensis*. (A) Principal component analysis (PCA) and (B) discriminant analyses (DA) based on vegetative and cyathium characters; (C) PCA and (D) DA based on fruit and seed characters.

Euphorbia balcanica and *E. skopjensis* appear most closely related to Anatolian *E. erythron*, *E. macroclada*, and *E. petrophila*, as well as to Pannonian–Pontic–Anatolian *E. glareosa* s.l., but the relationships among these lineages are poorly resolved in the AFLP tree. This was similar in the case of RADseq data, where different analyses resulted in partly different topologies (Stojilković et al., 2022). In the same line, *E. balcanica* and *E. skopjensis* are intermediate among three groups in the NeighbourNet based on ITS sequences (Figure 3B), namely, (1) *E. hercegovina*, *E. japygica*, and *E. nicaeensis*, (2) *E. erythron* and *E. macroclada*, as well as (3) *E. glareosa* s.l. The RGS data and chromosome counts (Figure 2) further revealed that *E. balcanica* is diploid, with an RGS of 0.90–0.95 intermediate between that of most samples of *E. glareosa* s.l., *E. erythron*, and *E. petrophila* that have an RGS of approximately 0.70, and the samples of *E. hercegovina*, *E. japygica*, and *E. macroclada* (as well as some of *E. glareosa* s.l.; see below) that have an RGS between 1.00 and 1.20 (Figure 2A). Further, given that the populations of *E. skopjensis* are positioned in the same clade with *E. balcanica* in the NJ tree based on AFLP (Figure 4) and that the RGS of *E. skopjensis* is double of that of *E. balcanica* (Figure 2A), *E. skopjensis* is likely an autotetraploid that originated from *E. balcanica*.

The intermediate RGS and the central phylogenetic position in the ITS NeighbourNet of *E. balcanica* as well as poorly resolved relationships in the AFLP NJ tree suggest that this species, and thus also its tetraploid derivative *E. skopjensis*, might be of a hybrid origin, likely between geographically proximate *E. hercegovina* or *E. japygica* on one side, and *E. glareosa* s.l. on the other. This is

further supported by early divergence of *E. skopjensis* within the *E. nicaeensis* lineage in the maximum likelihood tree and its position between *E. hercegovina* and *E. japygica* in the species tree inferred with SNAPP, both based on RADseq data (Stojilković et al., 2022; results summarised in Supplementary Figure S4), which uncovered close relationships of *E. balcanica* and *E. skopjensis* to *E. hercegovina* and *E. japygica*. In addition, Bayesian clustering of RADseq data using fastSTRUCTURE indicated an admixed genetic pattern in *E. skopjensis*, with a share of its genome corresponding to the lineage including *E. hercegovina* and *E. japygica*, and a share derived from the lineage including *E. glareosa* s.l. (Stojilković et al., 2022; Supplementary Figure S4). The combination of genetic and RGS data are thus in favour of our hypothesis that the *E. balcanica*/*E. skopjensis* lineage might be of hybrid origin between Mediterranean and steppe lineages of the *E. nicaeensis* alliance.

Our morphometric analyses showed that there is a strong overlap in morphology between *E. balcanica* and closely related species, especially *E. japygica*, whereas *E. skopjensis* is much better differentiated based on vegetative and cyathium characters (Figure 6B; Supplementary Figure S3). On the other hand, *E. skopjensis* overlaps strongly with *E. japygica* in fruit and seed characters, whereas *E. balcanica* is better differentiated (Figure 6D). It was shown earlier (Stojilković et al., 2022) that morphological differentiation only partly follows evolutionary trajectories in the *E. nicaeensis* alliance, which resulted in taxonomic lumping of morphologically similar but evolutionarily divergent entities of this alliance in the past (e.g., Radcliffe-Smith and Tutin, 1968; Kuzmanov, 1979; Greuter et al., 1986; Govaerts et al., 2000). The discordant

patterns likely result from (1) adaptation to similar habitats within divergent phylogenetic lineages leading to similar morphology (e.g., between *E. nicaensis* and *E. glareosa* s.l.) or (2) from adaptation to divergent ecological niches within the same evolutionary lineages resulting in divergent morphology of closely related species (e.g., between *E. hercegovina* and *E. japygica*). The latter is also the case in the two newly described species that are morphologically clearly divergent (Figure 7), likely as a result of adaptation to divergent environments.

Euphorbia balcanica grows at higher altitudes in nutrient-poor habitats, such as gravelly grasslands or scrublands over serpentines (in Albania and Kosovo) or dolomite (in North Macedonia), which resemble the habitats of *E. hercegovina* and *E. japygica* (Stojilković et al., 2022; Boschin et al., 2024). On the other hand, *E. skopjensis* thrives at lower elevations in subruderal habitats, such as abandoned meadows or pastures with deeper, nutrient-rich soils, which is closer to the ecology of *E. glareosa* s.l. (Stojilković et al., 2022). This ecological and morphological divergence led to differential treatments of the populations of *E. balcanica* and *E. skopjensis* in the Flora of Macedonia (Micevski, 1998; Matevski et al., 2018). Whereas the former were treated as *E. hercegovina* (sub *E. barrelieri* Savi subsp. *hercegovina* (Beck) Kuzmanov), those of *E. skopjensis* were considered to belong to *E. nicaensis* (incl. *E. glareosa* s.l.). On the other hand, the populations of *E. balcanica* from Albania were thought to belong to *E. nicaensis* (incl. *E. glareosa* s.l.; Qosja et al., 1992; Barina, 2017).

4.2 Complex diversification of *Euphorbia glareosa* s.l.

Our extended sampling of *E. glareosa* s.l. compared to that of Stojilković et al. (2022) brings further evidence of complex diversification patterns within *E. glareosa* s.l. Complexity within this taxon was recognised already based on morphology, which led to recognition of different (infra)specific taxa (e.g., Prokhanov, 1949; Kuzmanov, 1979; Greuter et al., 1986; Geltman, 2009). In addition, Stojilković et al. (2022) have shown pronounced RGS variation within *E. glareosa* s.l. suggesting polyploidisations but also homoploid differentiation, as well as strong genetic differentiation among populations based on RADseq data.

Our RGS data, calibrated with a chromosome count of $2n = 2x = 18$ (Figure 2), indicated that most sampled populations are diploid with RGS between 0.683 and 0.731 (black cross in Figure 1) and pertain to *E. glareosa* s.str. In addition, populations 116 from Serbia and 162 from Turkey that have RGS 1.401 and 1.436 (black cross with white margin in Figure 1), respectively, are likely autotetraploids derived independently from geographically close diploid populations; they should thus also be treated as *E. glareosa* s.str. as they are also morphological similar to the diploids. The independent origin of the two tetraploid populations is supported by RGS (Figure 2A) and AFLP (Figure 5) data. The RGS of the diploid populations 113 and 115 that are geographically closest to the putatively tetraploid population 116 (all south-east Serbia) is 0.667 and 0.699, respectively, which is half of the RGS of

population 116. In addition, populations 115 and 116 are most closely related and share several common splits in the AFLP NeighbourNet (Figure 5B). On the other hand, the putatively tetraploid population 162 from Turkey is nested within the diploid populations 152 and 156 from the Pontic steppes in the AFLP NeighbourNet, which have RGS 0.726 and 0.718, respectively, i.e., half the RGS of population 162.

Phylogeographically, there is clear genetic divergence between the (1) Pannonian, central, and eastern Balkan populations, and (2) the easternmost Balkan, Pontic, and Anatolian populations of *E. glareosa* s.str. as indicated by the main AFLP split at $K = 2$ (Figure 5A). This split was evidenced also by the RADseq data, although based on a much scarcer geographic sampling (seven populations analysed; Stojilković et al., 2022) and corresponds to the border between zonal and extrazonal steppes (Wesche et al., 2016; Kirschner et al., 2020). Our study thus additionally highlights the outstanding conservation value of the extrazonal European steppes that are not just an outpost of zonal Eurasian steppes; many of their characteristic species evolved independently in isolation for most of their history (Kirschner et al., 2020).

Deviating RGS values (Figure 2; Supplementary Table S1) of populations 126–128 (RGS 0.592–0.601; white cross with black margin in Figure 1), were likely caused by other factors than polyploidisation. Alongside polyploidy, accumulation of retrotransposons and other repetitive elements is considered the main cause of GS increase in angiosperms (Pellicer et al., 2018). We observed also a partial morphological and ecological differentiation of these populations, which were growing on steeper, open gravelly, and stony grounds compared to *E. glareosa* s.str., which mostly grows on deeper soils. Our preliminary taxonomic assessment suggests that they likely belong to *E. dobrogensis*. These populations were also genetically clearly differentiated; they formed a well-supported (BS 99%) lineage in the AFLP NJ tree (Supplementary Figure S2) and the NeighbourNet (Figure 5E), and formed a separate cluster at $K = 4$ (Figure 5C).

On the other hand, the deviating RGS of populations 132, 151, and 165 (1.047–1.146), as well as of populations 133 and 154 (1.226, 1.257; grey cross with black margin in Figure 1), is likely connected both to polyploidisation and subsequent RGS differentiation. These populations were morphologically fairly variable and appeared similar to *E. dobrogensis* (populations 132, 151, and 154), *E. cadrilateri* (population 133), and *E. glareosa* s.str. (population 165). They all formed a separate AFLP cluster at $K = 5$ (Figure 5D), which was intermediate between the diploid populations 126–128 of *E. cf. dobrogensis* and the eastern populations of diploid *E. glareosa* s.str. in the AFLP NeighbourNet (Figure 5E). This, along with their deviating RGS and diverse morphology, suggests that they likely originated via polyploidisation, but their parental species might either have been diploid *E. dobrogensis* (i.e., autopolyploid origin), or *E. dobrogensis* and *E. glareosa* s.str. (i.e., allopolyploid origin). Alternatively, tetraploidisation of *E. dobrogensis* and subsequent hybridisation with tetraploid *E. glareosa* s.str. (not included in our limited sampling in the Pontic area) could have led to the same genetic and RGS patterns. All these populations were most early divergent

within *E. glareosa* s.l. in the AFLP NJ tree (BS <50%; [Supplementary Figure S2](#)), which is also in line with the tree based on RADseq ([Stojilković et al., 2022](#)). There, only populations 151 and 165 were included and resolved as sister to all other populations pertaining to *E. glareosa* s.str. In summary, the intricate patterns within *E. glareosa* s.l. call for further studies, especially in the east of the distribution range, i.e., the Pontic area and Anatolia, where the group appears to be most diverse.

4.3 Taxonomic treatment

Below, we provide a taxonomic treatment for the new species *E. balcanica* ([Figure 8A](#)) and *E. skopjensis* ([Figure 8B](#)), as well as a description of *E. glareosa* s.l. Despite the fact that some populations included in our study were preliminarily assigned to *E. dobrogensis* or *E. cadrilateri* (see Discussion above), we here refrain from taxonomic decisions about these two taxa; therefore, we do not

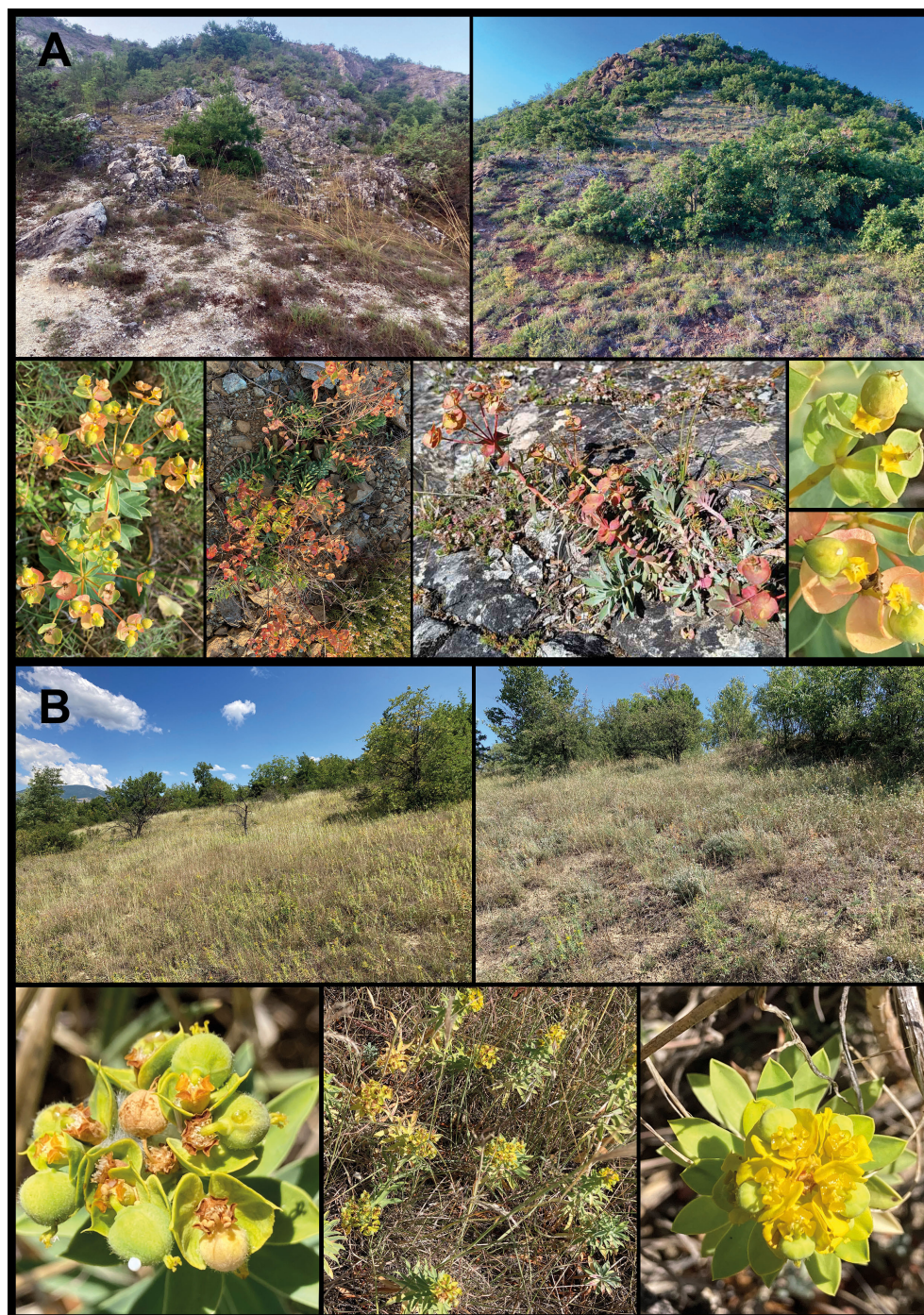


FIGURE 8
Euphorbia balcanica (A) and *E. skopjensis* (B) in their natural environments. Photos: Sharovikj Ivanova & Frajman.

list any heterotypic synonyms or infraspecific taxa of *E. glareosa* s.l. Further studies with extended geographic sampling in Anatolia and the Pontic areas are needed to clarify their taxonomic rank and status. Based on the available data, it appears reasonable to treat populations 126–128 with lower genome size as a distinct species, *E. dobrogensis*. In addition, the western AFLP cluster of *E. glareosa* s.str. pertains to *E. pannonica* and the eastern to *E. stepposa* [both to be treated as subspecies as suggested already by [Geltman \(2020\)](#)], but the relationships of the latter to *E. glareosa* s.str. (described from Crimea) as well as to the remaining taxa described from the Eurasian zonal steppes (*E. cadrilateri*, *E. goldei*, and *E. volgensis*) remain unresolved.

4.3.1 Identification key to species of the central and eastern Balkan Peninsula

- 1 Smaller decumbent plant, (8)11–29(38) cm high, with (1.0)1.5–3.0(4.0) mm-thick stems. Cauline leaves (1.2)1.4–2.9(3.6) × (0.3)0.4–0.7(1.1) cm, (2.2)2.7–4.6(5.1) times longer than wide. *Central Balkan Peninsula (but not Skopje Basin) on gravelly ground over serpentine or dolomite*.....*E. balcanica*
- 1* More robust plant, mostly erect, (5)16–46(60) cm high, with (1.0)1.8–3.9(4.8) mm-thick stems. Cauline leaves (0.8)1.9–4.5(5.7) × (0.3)0.5–1.2(1.9) cm, (0.9)2.9–5.6(7.7) times longer than wide. *Mostly on deeper soils, not over serpentine, rarely over dolomite*2
- 2 Terminal rays 3–6(9), the longest (0.5)1.0–2.7(3.6) cm long, once dichotomously branched. Fertile axillary rays (0)1–11(16). The longest fertile axillary ray (0.6)1.2–3.5(5.4)-cm long. Cauline leaves (1.5)2.4–5.0(5.7) × (0.4)0.5–1.0(1.4) cm, (3.2)3.8–6.3(7.4) times longer than wide. Ray leaves (0.6)0.7–1.2(1.9) × 0.5–1.2(1.3) cm, (0.8)0.9–1.8(2.7) times longer than wide. Nectarial glands sometimes with two horns. Seeds (2.1)2.2–2.8(3.0) × (1.2)1.6–2.0(2.2) mm. *Skopje Basin (North Macedonia)*.....*E. skopjensis*
- 2* Terminal rays (2)5–10(14), the longest (1.4)2.3–6.5(9.9) cm long, 1–3 times dichotomously branched. The longest fertile axillary ray (2.0)3.2–8.0(12.2) cm long. Cauline leaves (0.8)1.7–4.3(5.3) × (0.3)0.4–1.2(1.9) cm, (0.9)2.7–5.3(7.2) times longer than wide. Ray leaves (0.6)1.1–2.7(3.7) × (0.4)0.6–1.4(2.4) cm, (0.7)1.2–2.7(3.6) times longer than wide. Nectarial glands without horns. Seeds (1.6)1.9–2.5(3.4) × (1.1)1.3–1.7(1.8) mm. *Widespread in the Pannonian Plain, eastern Balkan Peninsula, Pontic and Anatolian steppes, Caucasus; rare in central Balkan Peninsula**E. glareosa*

Euphorbia balcanica Sharovikj Ivanova & Frajman, sp. nov.

Type: “Flora of North Macedonia, Dolneni, Debreška Krasta northeast of the village Debrešte, 682 m, 21°19′51″E, 41°29′16″N; gravelly pasture and scrubland over dolomite. Leg. A. Sharovikj Ivanova 18525, 15.06.2024” (holotype in W, isotypes in IB and MKNH).

Description: Glabrous and glaucous decumbent perennial, (8) 11–29(38) cm high, with (1.0)1.5–3.0(4.0) mm thick stems. Terminal rays (4)5–8(10), the longest (1.9)2.3–4.6(5.9) cm long, 1–2 times dichotomously branched. Fertile axillary rays 0–7(10), the longest (2.3)3.0–5.7(7.2) cm long. Cauline leaves pruinose–papillose, (1.2)1.4–2.9(3.6) × (0.3)0.4–0.7(1.1) cm, (2.2)2.7–4.6(5.1) times longer than wide, widest at (0.5)0.6–0.7(0.8) of their length, oblanceolate, with cuneate base and obtuse to mucronate apex; margin narrowly cartilaginous with dense, pointed papillae, minutely serrulate, especially toward the tips. Ray leaves broadly ovate to obovate, (0.7)0.9–1.5(2.5) × (0.5)0.6–1.3(1.7) cm, (0.8)1.0–2.0(2.3) times longer than wide, widest at (0.2)0.3–0.6 of their length. Raylet leaves broadly ovate–cordate, 0.6–1.2(1.5) × (0.7)0.9–1.5(1.7) cm, (0.5)0.6–0.9(1.0) times longer than wide, widest at (0.1)0.2–0.4(0.5) of their length, with shallowly cordate base and mucronate apex. Cyathial involucre campanulate, (1.9)2.3–3.1(4.0) × (1.2)1.6–2.5(3.0) mm, (0.9)1.0–1.6 times longer than wide. Cyathial lobes usually pubescent on the inner side. Nectarial glands broadly obovate–truncate to trapezoid or broadly semilunate, (0.5)0.7–1.8 × (0.9)1.2–1.7(2.0) mm, (0.4)0.5–1.4 times longer than wide, with 0–0.2(0.4) mm-deep emargination and mostly with two horns of different lengths. Fruits glabrous or pubescent, pruinose–papillose, broadly ovoid, (1.8)3.2–4.2(5.0) × (1.8)2.6–3.5(4.0) mm, (0.9)1.0–1.3 times longer than wide, styles (1.2)1.4–1.9 mm long. Seeds ovoid, smooth–papillose, brownish or greyish, (2.1)2.2–2.6(2.8) × (1.4)1.5–1.8(1.9) mm, (1.2)1.3–1.6 times longer than wide. Caruncle conical, (0.6)0.7–1.0 × (0.8)0.9–1.2(1.3) mm, (0.6)0.7–0.9(1.0) times longer than wide.

Distribution: Central-western Balkan Peninsula (Albania, Kosovo, and North Macedonia).

Habitat: Open rocky and gravelly, nutrient-poor grasslands and scrublands over serpentine (Albania, Kosovo) or dolomite (North Macedonia).

Etymology: We name this species after the Balkan Peninsula, where it is endemic in its central part.

Euphorbia skopjensis Sharovikj Ivanova & Frajman, sp. nov.

Type: “Flora of North Macedonia, Skopje Basin, north of the village Kondovo, 320 m, 21°18′44″E, 42°0′54″N; ruderalised pasture and road margin. Leg. A. Sharovikj Ivanova 18532, 24.06.2024” (holotype in W, isotypes in IB and MKNH).

Description: Glabrous and glaucous erect perennial, (16)18–46(51) cm high, with (1.9)2.3–3.2(4.5) mm thick stems. Terminal rays 3–6(9), the longest (0.5)1.0–2.7(3.6) cm long, once dichotomously branched. Fertile axillary rays (0)1–11(16), the longest (0.6)1.2–3.5(5.4) cm long. Cauline leaves pruinose–papillose, (1.5)2.4–5.0(5.7) × (0.4)0.5–1.0(1.4) cm, (3.2)3.8–6.3(7.4) times longer than wide, widest at (0.4)0.6–0.8 of their length, oblanceolate, with cuneate base and obtuse to mucronate apex; margin narrowly cartilaginous with dense, pointed papillae, minutely serrulate, especially toward the tips. Ray leaves elliptic to ovate, (0.6)0.7–1.2(1.9) × 0.5–1.2(1.3) cm, (0.8)0.9–1.8(2.7) times longer than wide, widest at (0.2)0.3–0.7 of their length. Raylet leaves broadly ovate–cordate, (0.4)0.5–0.8(1.1) × (0.4)0.6–1.1(1.4) cm, (0.5)0.6–0.9(1.0) times longer than wide, widest at (0.1)0.2–0.6(0.7) of their length, with shallowly cordate base and mucronate apex. Cyathial involucre campanulate, (1.3)2.6–3.6(3.9) × (1.5)1.7–

3.4(4.1) mm, (0.6)0.9–1.7(1.9) times longer than wide. Cyathial lobes usually pubescent on the inner side. Nectarial glands broadly obovate–truncate to trapezoid, rarely broadly semilunate, (0.5)0.6–1.1(1.3) × (1.0)1.1–1.8(2) mm, (0.4)0.5–0.7(0.8) times longer than wide, with 0–0.2(0.4) mm-deep emargination, sometimes with two horns of different lengths. Fruits glabrous or pubescent, pruinose–papillose, broadly ovoid, (2.5)3.2–4.6(5.0) × (2.7)2.8–4.3(4.9) mm, (0.8)0.9–1.2(1.6) times longer than wide, styles (1.0)1.3– to 2.0(2.4) mm long. Seeds ovoid, smooth–papillose, brownish or greyish, (2.1)2.2–2.8(3.0) × (1.2)1.6–2.0(2.2) mm, (1.0)1.1–1.6(1.9) times longer than wide. Caruncle conical, (0.5)0.6–0.9(1.0) × (0.7)0.8–1.2(1.4) mm, (0.5)0.7–0.9(1.0) times longer than wide.

Distribution: Skopje Basin in North Macedonia.

Habitat: Abandoned meadows and pastures, scrublands, often with semi-ruderal character.

Etymology: We have named the species after the city of Skopje as it is only known from this city and its surroundings.

Euphorbia glareosa Pall. ex M.Bieb. in Fl. Taur.-Caucas. 1: 373 (1808) ≡ *E. nicaeensis* subsp. *glareosa* (Pall. ex M.Bieb.) Radcl.-Sm. in Repert. Spec. Nov. Regni Veg. 79: 55 (1968) ≡ *Galarhoeus glareosus* (Pall. ex M.Bieb.) Prokh. in Trudy Kuibyshevsk. Bot. Sada 1: 40 (1941) ≡ *Tithymalus nicaeensis* subsp. *glareosus* (Pall. ex M.Bieb.) Soják in Čas. Nár. Mus., Odd. Přír. 152: 22 (1983).

Note: *Euphorbia glareosa* is here treated in the broader sense, including *E. cadrilateri* Prodan, *E. dobrogensis* Prodan, *E. pannonica* Host., and *E. stepposa* Zoz. The taxonomic status of these taxa as well as *E. goldei* Prokh. and *E. volgensis* Krysth. should be resolved in the future studies. For typification and possible synonyms, see also [Geltman \(2020\)](#).

Description: Glabrous, rarely pubescent, and glaucous erect, rarely decumbent perennial, (4)15–46(61) cm high, with (1.0)1.8–4.0(4.8) mm-thick stems. Terminal rays (2)5–10(14), the longest (1.4)2.3–6.5 (9.9) cm long, 1–3 times dichotomously branched. Fertile axillary rays (0)1–11(16), the longest (2.0)3.2–8.0(12.2) cm long. Cauline leaves pruinose–papillose, (0.8)1.7–4.3(5.3) × (0.3)0.4–1.2(1.9) cm, (0.9)2.7–5.3(7.2) times longer than wide, widest at (0.3)0.5–0.7(0.8) of their length, elliptic to oblanceolate, with cuneate base and broadly acuminate to mucronate apex; margin narrowly cartilaginous with dense, rounded, or pointed papillae, minutely serrulate, especially toward the tips. Ray leaves broadly ovate to obovate, (0.6)1.1–2.7 (3.7) × (0.4)0.6–1.4(2.4) cm, (0.7)1.2–2.7(3.6) times longer than wide, widest at (0.1)0.3–0.6(0.7) of their length. Raylet leaves broadly ovate–cordate, (0.6)0.7–1.4(1.9) × (0.8)0.9–1.6(2.5) cm, (0.5)0.7–1(1.4) times longer than wide, widest at (0.1)0.2–0.4(0.6) of their length, with shallowly cordate base and mucronate apex. Cyathial involucre campanulate, (1.8)2.3–4(4.9) × (0.9)1.5–2.3(2.7) mm, (1)1.3–2.3(3) times longer than wide. Cyathial lobes usually pubescent on the inner side. Nectarial glands broadly obovate–truncate to trapezoid or broadly semilunate, (0.5)0.6–1.0(1.6) × (0.6)1.0–1.6(2.2) mm, (0.4)0.5–0.8(0.9) times longer than wide, with 0–0.2(0.5) mm-deep emargination, without horns. Fruits glabrous or pubescent, pruinose–papillose, broadly ovoid, (2.4)2.8–4.5(5.5) × (2.1)2.6–3.3(3.5) mm, (0.8)0.9–1.6(1.9) times longer than wide, styles (0.8)1.2– to 2(2.2) mm long. Seeds ovoid, smooth–papillose, brownish or greyish, (1.6)1.9–2.5(3.4) × (1.1)1.3–1.7(1.8) mm, (1.2)1.3–1.6(2.0) times longer than wide.

Caruncle conical, (0.3)0.4–0.7(1.0) × (0.4)0.7–1.0(1.2) mm, (0.4)0.5–0.8(1.0) times longer than wide.

Distribution: Pannonian Basin, central and eastern Balkan Peninsula, Pontic area north of the Black Sea, Crimea, Asia Minor, Armenian Highlands.

Habitat: Grasslands and scrublands.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/[Supplementary Material](#).

Author contributions

AS: Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft, Writing – review & editing. PS: Methodology, Supervision, Validation, Writing – review & editing. MK: Investigation, Writing – review & editing. MB: Methodology, Writing – review & editing. RC: Funding acquisition, Investigation, Writing – review & editing. BF: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This study was financially supported by the Austrian Agency for International Cooperation in Education and Research (ÖAD-GmbH; grants MK 11/2022, MK 16/2024, and MULT 09/2023 to BF).

Acknowledgments

We thank all collectors listed in [Supplementary Table S1](#) for help with the collection of samples and Z. Barina, S. Đurović, N. Kuzmanović, A. Lambevskaja, V. Matevski, M. Niketić, K. Vasilev, and B. Zlatković for locality data. D. Pirkebner, M. Magauer, S. Vurmeska Cvetanoska, and C. Stoichevska helped with the lab work, N. Kuzmanović with the K-means analyses, and M. Magauer and E. Huber with the preparation of some figures. We are grateful to M. Bodner and his colleagues from the Botanical Garden of the University of Innsbruck for successfully cultivating our living collection of *Euphorbia*.

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.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

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/fpls.2025.1558466/full#supplementary-material>

References

- 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.1469-8137.2010.03282.x
- Arrigo, N., Tuszyński, J. W., Ehrlich, D., Gerdes, T., and Alvarez, N. (2009). Evaluating the impact of scoring parameters on the structure of intra-specific genetic variation using RawGeno, an R package for automating AFLP scoring. *BMC Bioinf.* 10, 1–14. doi: 10.1186/1471-2105-10-33
- Barina, Z. (2017). *Distribution atlas of vascular plants in Albania* (Budapest: Hungarian Natural History Museum).
- Boschin, M., Schönschetter, P., and Frajman, B. (2024). Genetic and morphological differentiation within *Euphorbia japygica* (Euphorbiaceae) suggests divergence of populations from the south-eastern Apennine Peninsula. *Bot. J. Linn. Soc.* 205, 38–54. doi: 10.1093/botlinnean/boad006
- Coupland, R. T. (1993). "Ecosystems of the world 8B. Natural grasslands," in *Eastern hemisphere and résumé* (Elsevier, Amsterdam).
- Cresti, L., Schönschetter, P., Peruzzi, L., Barfuss, M. H. J., and Frajman, B. (2019). Pleistocene survival in three Mediterranean refugia: Origin and diversification of the Italian endemic *Euphorbia gasparrinii* from the *E. verrucosa* alliance (Euphorbiaceae). *Bot. J. Linn. Soc.* 189, 262–280. doi: 10.1093/botlinnean/boy082
- Đurović, S., Schönschetter, P., Niketić, M., Tomović, G., and Frajman, B. (2017). Disentangling relationships among the members of the *Silene saxifraga* alliance (Caryophyllaceae): Phylogenetic structure is geographically rather than taxonomically segregated. *Taxon* 66, 343–364. doi: 10.12705/662.4
- ESRI (2021). ArcMAP: release 10. Environmental systems research institute. Redlands, CA: Environmental Systems Research Institute.
- 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
- Frajman, B., Pachschoell, C., and Schönschetter, P. (2014). Contributions to the knowledge of the flora of the Dinarides (Balkan Peninsula). *Phyton* 54, 27–46. doi: 10.12905/0380.phyton54(1)2014-0027
- Frajman, B., and Schönschetter, P. (2011). Giants and dwarfs: Molecular phylogenies reveal multiple origins of annual spurges within *Euphorbia* subg. *Esula*. *Mol. Phylogenet. Evol.* 61, 413–424. doi: 10.1016/j.ympev.2011.06.011
- Frajman, B., Závěská, E., Gamisch, A., Moser, T., The STEPNET Consortium, and Schönschetter, P. (2019). Integrating phylogenomics, phylogenetics, morphometrics, relative genome size, and ecological niche modelling disentangles the diversification of Eurasian *Euphorbia seguieriana* s.l. (Euphorbiaceae). *Mol. Phylogenet. Evol.* 134, 238–252. doi: 10.1016/j.ympev.2018.10.046
- Geltman, D. V. (2009). Spurges (*Euphorbia* L., Euphorbiaceae) of the boreal Eurasia. I. Section *Paralias* Dumort [In Russian with English summary. *Novosti Sist. Vyssh. Rast.* 41, 166–191.
- Geltman, D. V. (2020). A synopsis of *Euphorbia* (Euphorbiaceae) for the Caucasus. *Novosti Sist. Vyssh. Rast.* 51, 43–78. doi: 10.31111/novitates/2020.51.43
- Govaerts, R., Frodin, D. G., and Radcliffe-Smith, A. (2000). *Checklist and bibliography of Euphorbiaceae 2* (Kew: Royal Botanic Gardens).
- Greilhuber, J., and Ebert, I. (1994). Genome size variation in *Pisum sativum*. *Genome* 37, 646–655. doi: 10.1139/g94-092
- Greuter, W., Burdet, H. M., and Long, G. (1986). *Med-Checklist 3. Palermo: Med-Checklist trust of OPTIMA*.
- Hamasha, H. N., von Hagen, B., and Röser, M. (2012). *Stipa* (Poaceae) and allies in the Old World: Molecular phylogenetics realigns genus circumscription and gives evidence on the origin of American and Australian lineages. *Plant Syst. Evol.* 298, 351–367. doi: 10.1007/s00606-011-0549-5
- Hartigan, J. A., and Wong, M. A. (1979). Algorithm AS 136: A k-means clustering algorithm. *J. R. Stat. Soc. C-App.* 28, 100–108. doi: 10.2307/2346830
- Huson, D. H., and Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* 23, 254–267. doi: 10.1093/molbev/msj030
- 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
- Kirschner, P., Závěská, E., Gamisch, A., Hilpold, A., Trucchi, E., Paun, O., et al. (2020). Long-term isolation of European steppe outposts boosts the biome's conservation value. *Nat. Commun.* 11, 1–10. doi: 10.1038/s41467-020-15620-2
- Kuzmanov, B. (1979). "Euphorbia", in *Flora reipublicae popularis bulgaricae* 7. Ed. B. Kuzmanov (Academia scientiarum Bulgaricae, Sofia), 118–177.
- Kuzmanović, N., Comanescu, P., Frajman, B., Lazarević, M., Paun, O., Schönschetter, P., et al. (2013). Genetic, cytological and morphological differentiation within the Balkan-Carpathian *Sesleria rigida* sensu *Fl. Eur.* (Poaceae): A taxonomically intricate tetraploid-octoploid complex. *Taxon* 62, 458–472. doi: 10.12705/623.13
- Lal, R. (2004). Carbon sequestration in soils of Central Asia. *Land Degrad. Dev.* 15, 563–572. doi: 10.1002/ldr.624
- Matevski, V., Čarni, A., Čusterevska, R., Kostadinovski, M., and Mucina, L. (2018). Syntaxonomy and biogeography of dry grasslands on calcareous substrates in the central and southern Balkans. *Appl. Veg. Sci.* 21, 488–513. doi: 10.1111/avsc.12374
- Micevski, K. (1998). "Flora na Republika Makedonija (The flora of the Republic of Macedonia) 1-4," in *Skopje: Macedonian Academy of Sciences and Arts*, 781–1113.
- Nei, M., and Li, W. H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *P. Natl. A. Sci.* 76, 5269–5273. doi: 10.1073/pnas.76.10.5269
- Niketić, M., Đurović, S. Z., Tomović, G., Schönschetter, P., and Frajman, B. (2022). Diversification within ploidy-variable Balkan endemic *Cerastium decalvans* (Caryophyllaceae) reconstructed based on genetic, morphological and ecological evidence. *Bot. J. Linn. Soc.* 199, 578–608. doi: 10.1093/botlinnean/boab037
- Pearl, B. (2008). "Life in a working landscape: Towards a conservation strategy for the world's temperate grasslands: Compendium of regional templates on the status of temperate grasslands," in *Temperate grasslands conservation initiative*. Vancouver: IUCN/WCPA.
- Pellicer, J., Hidalgo, O., Dodsworth, S., and Leitch, I. J. (2018). Genome size diversity and its impact on the evolution of land plants. *Genes* 9, 88. doi: 10.3390/genes9020088
- Prodan, I. (1936). Conspectul florei dobrogei. Partea a II-a. *Bulet. Acad. Înalte Stud. Agron. Cluj* 6, 202–259.
- Prokhanov, Y. I. (1949). "Genus 856. Euphorbia L.," in *Flora SSSR*. Eds. B. K. Shishkin and E. G. Bobrov (Akademii Nauk SSSR, Moskva-Leningrad), 233–378.
- Qosja, X., Paparisto, K., Demiri, M., and Vangjeli, J. (1992). *Flora e shqipërisë 2 [The flora of Albania 2]* (Tirana: Academy of Science of Albania).
- Radcliffe-Smith, A. (1982). "Euphorbia L.," in *Flora of Turkey*, vol. 7. Ed. P. H. Davis (Edinburgh University Press, Edinburgh), 571–630.
- Radcliffe-Smith, A., and Tutin, T. G. (1968). "Euphorbia L.," in *Flora europaea 2*. Eds. T. G. Tutin, V. H. Heywood, D. M. Moore, D. H. Valentine, S. M. Walters and D. A. Webb (Cambridge University Press, Cambridge), 213–226.
- Rešetnik, I., Závěská, E., Grgurev, M., Bogdanović, S., Bartolić, P., and Frajman, B. (2022). Stability in the south, turbulence toward the north: evolutionary history of *Aurinia saxatilis* (Brassicaceae) revealed by phylogenomic and climatic modelling data. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.822331
- Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., et al. (2012). MrBayes 3.2: Efficient bayesian phylogenetic inference and model

choice across a large model space. *Syst. Biol.* 61, 539–542. doi: 10.1093/sysbio/sys029

RStudio Team (2022). *RStudio: Integrated development for R* (Boston, MA: RStudio PBC). Available at: <http://www.rstudio.com/> (Accessed May 12 2024).

Schönswetter, P., Suda, J., Popp, M., Weiss-Schneeweiss, H., and Brochmann, C. (2007). Circumpolar phylogeography of *Juncus biglumis* (Juncaceae) inferred from AFLP fingerprints, cpDNA sequences, nuclear DNA content and chromosome numbers. *Mol. Phylogenet. Evol.* 42, 92–103. doi: 10.1016/j.ympev.2006.06.016

Skokanová, K., Hodálová, I., Mereďa, P., Slovák, M., and Kučera, J. (2019). The *Cyanus tuberosus* group (Asteraceae) in the Balkans: biological entities require correct names. *Plant Syst. Evol.* 305, 569–596. doi: 10.1007/s00606-019-01576-4

Španiel, S., and Rešetnik, I. (2022). Plant phylogeography of the Balkan Peninsula: Spatiotemporal patterns and processes. *Plant Syst. Evol.* 308, 38. doi: 10.1007/s00606-021-01756-3

Stojilković, V., Závěská, E., and Frajman, B. (2022). From western Asia to the Mediterranean Basin: diversification of the widespread *Euphorbia nicaeensis* alliance (Euphorbiaceae). *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.815379

Suda, J., and Trávníček, P. (2006). Estimation of relative nuclear DNA content in dehydrated plant tissues by flow cytometry. *Curr. Protoc. Cytom.* 38, 7.30.1–7.30.14. doi: 10.1002/0471142956.cy0730s38

Swofford, D. L. (2002). *Phylogenetic analysis using parsimony* (Sunderland, MA: Sinauer Associate). doi: 10.1111/j.0014-3820.2002.tb00191.x

Tomović, G., Niketić, M., Lakušić, D., Randelović, V., and Stevanović, V. (2014). Balkan endemic plants in Central Serbia and Kosovo regions: Distribution patterns, ecological characteristics, and centres of diversity. *Bot. J. Linn. Soc.* 176, 173–202. doi: 10.1111/boj.12197

Van de Peer, Y., and de Wachter, R. (1997). Construction of evolutionary distance trees with TREECON for Windows: Accounting for variation in nucleotide substitution rate among sites. *Bioinformatics.* 13, 227–230. doi: 10.1093/bioinformatics/13.3.227

Vos, P., Hogers, R., Bleeker, M., Reijans, M., Van de Lee, T., Hornes, M., et al. (1995). AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res.* 23, 4407–4414. doi: 10.1093/nar/23.21.4407

Wesche, K., Ambarlı, D., Kamp, J., Török, P., Treiber, J., and Dengler, J. (2016). The Palearctic steppe biome: a new synthesis. *Biodivers. Conserv.* 25, 2197–2231. doi: 10.1007/s10531-016-1214-7

Závěská, E., Maylandt, C., Paun, O., Bertel, C., Frajman, B., Schönswetter, P., et al. (2019). Multiple auto- and allopolyploidisations marked the Pleistocene history of the widespread Eurasian steppe plant *Astragalus onobrychis* (Fabaceae). *Mol. Phylogenet. Evol.* 139, 106572. doi: 10.1016/j.ympev.2019.106572



OPEN ACCESS

EDITED BY

Enrique Maguilla,
Universidad Pablo de Olavide, Spain

REVIEWED BY

Rowan J. Schley,
University of Exeter, United Kingdom
Elvira Hörandl,
University of Göttingen, Germany

*CORRESPONDENCE

Angélica Gallego-Narbón
✉ Gallego-NarbonA@si.edu

RECEIVED 17 March 2025

ACCEPTED 23 May 2025

PUBLISHED 13 June 2025

CITATION

Gallego-Narbón A, Johnson G,
Fernández-Mazuecos M, Wen J and
Valcárcel V (2025) Ancient polyploidization
events influence the evolution of the
ginseng family (Araliaceae).
Front. Plant Sci. 16:1595321.
doi: 10.3389/fpls.2025.1595321

COPYRIGHT

© 2025 Gallego-Narbón, Johnson,
Fernández-Mazuecos, Wen and Valcárcel. 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.

Ancient polyploidization events influence the evolution of the ginseng family (Araliaceae)

Angélica Gallego-Narbón ^{1,2*}, Gabriel Johnson ¹,
Mario Fernández-Mazuecos ^{2,3,4}, Jun Wen ¹
and Virginia Valcárcel ^{2,3}

¹Department of Botany, National Museum of Natural History, Smithsonian Institution, Washington, DC, United States, ²Departamento de Biología, Universidad Autónoma de Madrid, Madrid, Spain, ³Centro de Investigación en Biodiversidad y Cambio Global (CIBC-UAM), Universidad Autónoma de Madrid, Madrid, Spain, ⁴Departamento de Biodiversidad y Conservación, Real Jardín Botánico (RJB), CSIC, Madrid, Spain

Introduction: Whole genome duplication events (WGDs) have been recognized as major drivers of evolution in plants, especially when they involve hybridization (allopolyploidization). In this study we evaluated if WGDs acted as evolutionary forces at the origin and early divergence of the Asian Palmate group (AsPG) of the plant family Araliaceae. This clade encompasses most of the generic and species diversity as well as most of the polyploids of the family, and a role of hybridization in its origin has been suggested.

Methods: In order to test this hypothesis, we obtained nuclear and plastid time-calibrated phylogenomic trees including 80% of Araliaceae genera (37 genera, 237 species) using the Hyb-Seq approach. The role of WGDs in the early evolution of the AsPG was tested using ancestral chromosome number reconstructions based on chromosome counts for 62% of the sampled genera, while recent polyploidization events were explored by inferring ploidy of the sequenced species from allelic frequencies.

Results: Phylogenetic analyses of nuclear and plastid sequences provided highly resolved but incongruent topologies consistent with ancient hybridization not only for the origin of the AsPG, but also in the second most highly diverse clade of the family. Our ancestral chromosome number reconstructions supported that one or two WGDs preceded the origin of two of the three main clades of Araliaceae (AsPG and *Polyscias-Pseudopanax*), which could have acted as background variables necessary for the posterior diversification of these lineages. Ploidy inference based on allelic frequencies provided signal of recent polyploidization in the AsPG and the third main clade of Araliaceae (*Aralia-Panax*).

Conclusions: In summary, WGDs are linked to the origin of the main clades of the Araliaceae family, but the drivers of the strong diversification of the AsPG remain an open question.

KEYWORDS

Araliaceae, Asian Palmate group, chromosome evolution, Hyb-Seq, whole genome duplication, polyploidy, hybridization

1 Introduction

Chromosomes can evolve in a series of ways that range from structural rearrangements like duplications, deletions, inversions, and translocations among individual chromosomes to partial genome duplication and polyploidization or whole genome duplication (WGD, Bado et al., 2015). These changes are common across different plant groups and have evolutionary significance (i.e., they can lead to an increase in speciation or extinction) depending on the characteristics of the lineage involved and its different adaptations to the surrounding environment (Stebbins, 1971). The footprint of WGD events is widespread across angiosperms (Soltis and Soltis, 2009), and polyploidy is currently recognized as one of the main drivers of speciation in plants (Valdés-Flórido et al., 2023). WGD can occur as a result of the fusion of unreduced diploid gametes within the same species (autopolyploidy) or by means of hybridization between different species that is associated with genome doubling (allopolyploidy; Landis et al., 2018). While WGDs, and especially allopolyploidization events, were considered an evolutionary dead-end in the past (Stebbins, 1971), its evolutionary potential is now widely recognized (Soltis et al., 2015). In fact, there is increasing evidence that WGD was a trigger of diversification in several species-rich lineages of angiosperms (Schranz et al., 2012; Tank et al., 2015; Landis et al., 2018). Allopolyploidization events could be especially relevant for speciation, given that they often provide genetic and morphological novelty (Qiu et al., 2020). In this regard, multiple gene copies from duplicate genomes can lead to the neofunctionalization of homologues to develop new functions in molecular pathways and allow greater genomic plasticity (Tao et al., 2021).

The ginseng family (Araliaceae, Apiales) constitutes an ideal study case to assess the role of chromosome number changes and WGDs in evolution. This diverse family encompasses approximately 46 genera and 1700 species, with a widespread distribution across five continents but mostly occupying tropical and subtropical areas in Asia, the Americas and Oceania (Wen et al., 2001; Nicolas and Plunkett, 2014; Valcárcel and Wen, 2019). The family is characterized by chromosome numbers from $n=12$ to $n=96$, with $n=12$ and $n=24$ as the numbers present in most genera, and polyploids present in several lineages (e.g.: *Aralia*, *Hedera*, and *Panax*; Wen and Zimmer, 1996; Vargas et al., 1999; Yi et al., 2004; Zhang et al., 2025). In addition, the signal of an ancient WGD has been detected in the early evolution of Apiales (Landis et al., 2018), which could be related to the high diversity currently found in the most diversified families Apiaceae and Araliaceae. Altogether, the chromosome counts and ploidy variability in the family and the presence of ancient WGDs in the history of Apiales suggest that polyploidization could be an important force in the evolution of Araliaceae. Yi et al. (2004) compiled chromosome counts in the family and observed that $n=12$ was the most common chromosome number in most Araliaceae lineages, while $n=24$ was the most common number in the Asian Palmate group (AsPG; Wen et al., 2001; Plunkett et al., 2004a), a clade that encompasses around 50% of the generic and species diversity of the family. These results pointed to an early WGD related to the origin of the AsPG. Gallego-Narbón et al. (2022) tested this hypothesis on highly resolved

nuclear and plastid trees of these genera by performing ancestral chromosome number reconstructions which revealed a WGD at the origin of the AsPG. The authors additionally inferred a hybridization event involved in the origin of the AsPG and hypothesized that the early evolution of the AsPG was driven by an allopolyploidization event. However, broader taxon sampling and higher phylogenetic resolution are necessary to test this hypothesis. Therefore, the placement of the detected WGD in the evolutionary history of Araliaceae and its impact on the early divergence of AsPG genera could not be assessed.

In this study, we aim to assess the relevance of chromosome evolution and ploidy changes in the origin and subsequent evolution of the Asian Palmate group of Araliaceae. We hypothesized that an ancient allopolyploidization event underlays the origin of the clade. Additionally, given the evidence of polyploidization in several Araliaceae lineages, we considered that WGDs may have also been relevant in the origin and subsequent evolution of the other two main clades of Araliaceae (*Polyscias-Pseudopanax* and *Aralia-Panax*; Wen, 2001; Plunkett et al., 2004b). These hypotheses were explored by: 1) disentangling the evolutionary relationships in the family Araliaceae using nuclear and plastid time-calibrated phylogenies with an extensive taxon sampling (96% of AsPG genera and 80% of Araliaceae genera), 2) assessing the link between WGDs and speciation events in the early evolution of Araliaceae and determining their role in the evolution of the most speciose clades based on ancestral chromosome number reconstructions, and 3) evaluating the signal of recent polyploidization in the evolution of Araliaceae.

2 Methods

2.1 Species sampling and library preparation

A total of 239 samples were included in this study, comprising 238 Araliaceae samples (237 species) and *Mackinlaya schlechteri* (Harms) Philipson (Apiaceae) as the outgroup (Supplementary Data 1). We represented the three main clades of Araliaceae (AsPG: 169 taxa, 168 species, *Aralia-Panax*: 28 species, and *Polyscias-Pseudopanax*: 11 species) and seven of the eleven early-diverging lineages (greater *Raukaua*: 13 species; *Cussonia-Seemanaralia*: six species; *Osmoxylon*: three species; two species of *Harmsiopanax* and *Hydrocotyle*, and one species of *Neocussonia-Astropanax* and *Trachymene*). Our sampling covered 22 AsPG genera (96% of AsPG genera) and 37 Araliaceae genera (80% of Araliaceae genera). These percentages are based on the taxonomic treatment of the family published by Plunkett et al. (2018) and subsequent taxonomic publications (Lowry et al., 2019, 2020; Fiaschi et al., 2020; Plunkett et al., 2021).

The raw reads of 73 species were obtained from previous studies: 63 species from Gallego-Narbón et al. (2022) and two species from the 1000 Plant Transcriptomes Initiative (www.onekp.com), obtained using an Illumina HiSeq platform sequencing 150 bp paired-end reads, and 8 species from Shee

et al. (2020) obtained using an Illumina MiSeq platform sequencing 250 bp paired-end reads. Genomic data for the remaining 166 taxa were newly generated using both silica-dried materials (132 samples) and herbarium specimens from the United States National Herbarium (US, 34 samples). DNA extraction, library preparation, and target enrichment followed the Hyb-Seq protocol used by Gallego-Narbón et al. (2022), except for the number of samples per indexed pool (12 for silica-dried material and 10 for herbarium material). We used a Araliaceae-specific bait set targeting 936 nuclear exons designed based on two genomes and two transcriptomes representing the three main Araliaceae lineages (AsPG, *Aralia-Panax* and *Polyscias-Pseudopanax*; Gallego-Narbón et al., 2022) synthesized by Daicel Arbor BioSciences (Ann Arbor MI, USA). Paired-end sequencing with 150-bp reads was performed on the Illumina HiSeq 4000 platform at Novogene (Sacramento, CA, USA).

2.2 Sequence assembly and alignment

Resulting reads were trimmed using Trimmomatic 0.39 (Bolger et al., 2014) following the parameters in Gallego-Narbón et al. (2022). Trimmed paired files were assembled using HybPiper 2.1.1 (Johnson et al., 2016) by mapping the reads to the bait targets using BWA (Li and Durbin, 2009) and assembling the contigs with SPAdes (Bankevich et al., 2012). For the nuclear data, exons were assembled for the bait-set loci and flanking regions were retrieved with the “run_intronerate” tool. HybPiper statistics were obtained and paralogy was assessed with the “paralog_retriever” tool. Plastid loci were assembled by mapping reads to a reference plastome (*Eleutherococcus senticosus* (Rupr. & Maxim.) Maxim, GenBank accession JN637765) as described in Gallego-Narbón et al. (2022). Species with low recovery that showed clearly incorrect phylogenetic placements (i.e., recovered in a different genus) in preliminary analyses were not included in the final alignments. Alignments were performed for each gene separately using MAFFT 7.475 with standard parameters (Katoh and Standley, 2013), and those with < 10 species were excluded from further analyses. After removal of the paralogs detected with the “paralog_retriever” HybPiper tool, the nuclear matrix was concatenated (*NuTot*) and the plastid gene alignments were concatenated (*CpTot*) in Geneious Prime 5.1.7 (Kearse et al., 2012). Statistics for these concatenated matrices were generated with AMAS 1.0 (Borowiec, 2016).

2.3 Phylogenomic analyses

Concatenation-based and coalescent-based analyses were performed. For concatenation-based analyses, we applied a maximum likelihood (ML) approach using RAXML-HPC v.8.2.10 (Stamatakis, 2014) on the concatenated matrices under a GTRCAT evolutionary model with 1,000 fast bootstrap replicates using one partition per locus. RAXML generates reduced matrices by excluding sites that contain only undetermined values prior to the analysis. These matrices were also used as the input for Bayesian

Inference (BI) analyses. A BI approach was also applied exclusively to the *CpTot* matrix in ExaBayes v.1.4.1 (Aberer et al., 2014) using a parsimony starting tree, and two parallel MCMC runs with two coupled chains for 1.2 million generations under a GTR+G evolutionary model. A burn-in of 25% was used and convergence was assessed using Tracer 1.7.2 (Rambaut et al., 2018). A BI analysis was not run using the *NuTot* matrix because of computing limitations. To run coalescent-based analyses, 1,000 fast bootstrap replicates were generated with RAXML under a GTRCAT model for each nuclear locus alignment to obtain gene trees. These were then processed with ASTRAL 5.6.2 with default settings (Zhang et al., 2018) to obtain a coalescent-based tree. Local posterior probabilities (PP_{local}) and quartet scores were retrieved as measures of branch support and gene tree discordance for the coalescent-based tree. The quartet scores indicate the proportion of gene trees that support the main topology (Q1), the main alternative (Q2), and the second alternative (Q3) for each quartet connecting four descendant lineages. Low Q1 values are indicative of high gene tree discordance. In this scenario, similar Q2 and Q3 values can be interpreted as a signal of incomplete lineage sorting (ILS), while different Q2 and Q3 values are interpreted as a signal of hybridization (Schumer et al., 2016).

2.4 Phylogenetic dating

Divergence times were estimated using a penalized likelihood approach as implemented in treePL (Smith and O'Meara, 2012), with our ML nuclear and plastid trees separately as the input. We used two calibration points: (1) the 95% highest posterior density intervals obtained by Magallón et al. (2015) for the crown node of Araliaceae (49.28–72.9 Ma), and (2) the age of the oldest known macrofossil of *Hedera* (23.0–39.9 Ma, Rim, 1994; Kong, 2000) for the divergence between *Hedera* and *Merrillioanax*. The three-step calibration was started with a preliminary calibration with the “prime” option to establish the optimal parameter values. Secondly, the optimal parameter values obtained from the previous calibration were used with a random sample cross-validation with 200,000 iterations for penalized likelihood and 5,000 iterations for cross-validation. This second calibration provided the best value for the smoothing parameter, which was used in the third and final calibration. The resulting nuclear and plastid time-calibrated trees were used for downstream analyses. As treePL does not directly provide uncertainty ranges, we performed time calibration for each of the 1,000 bootstrapped trees generated by RAXML analyses of nuclear and plastid matrices with the parameters obtained for the ML trees as described above. Age statistics for the nodes were summarized in TreeAnnotator v.1.10 (Drummond et al., 2012).

2.5 Chromosome evolution and chromosomal cladogenesis

To study the role of ancient WGDs in the evolution of Araliaceae, we used ChromEvol 2.0 (Mayrose et al., 2010; Glick

and Mayrose, 2014) and ChromoSSE (Freyman and Höhna, 2018) in RevBayes v.1.0.12 (Höhna et al., 2016). Given that the species sampling was highly incomplete for most of the genera both for molecular data and for chromosome counts, chromosomal variation during the evolution of each genus could not be analyzed at the species level. Considering this limitation and our interest in exploring the role of ancient WGDs in the early evolution of Araliaceae, ancestral reconstructions at the genus level were performed. Chromosome counts compiled by Gallego-Narbón et al. (2022) for the Araliaceae family were complemented with additional counts of newly sampled genera from the Chromosome Counts Database (CCDB, Rice et al., 2015), the Index to Plant Chromosome Numbers database (IPCN, Goldblatt and Johnson, 2006), and Yi et al. (2004). This newly compiled cytological database contained 143 counts from $n=9$ to $n=96$ representing 24 Araliaceae genera (62% of sampled genera, Supplementary Data 2). When a genus presented several different chromosome numbers (eight genera), the frequency of each chromosome number was estimated based on the proportion of counts for each number, considering the total number of species with chromosome counts per genus (number of species with a certain chromosome number for a genus/total number of species with chromosome counts for that genus). Generic lineages whose chromosome number was not available (15) were treated as missing data as specified in the ChromEvol guidelines. The resulting chromosome numbers or proportions of each number per genus were used to reconstruct ancestral chromosome numbers based on the nuclear and plastid time-calibrated trees pruned to include a single species per genus, except for the non-monophyletic genera for which one tip per lineage was retained (40 samples in total, including *Mackinlaya* as the outgroup). ChromEvol tested ten different models of chromosome evolution and estimated ancestral chromosome numbers, whole genome duplication (WGD), half genome duplication, and dispoloidy events across the phylogeny. The best nuclear and plastid models of chromosome evolution were selected according to their AIC values.

Afterwards, the same inputs were used for a ChromoSSE analysis (Freyman and Höhna, 2018), as implemented in RevBayes v.1.0.12 (Höhna et al., 2016). Using this analysis, we assessed whether the chromosome changes detected early in the evolution of Araliaceae were temporally correlated with speciation events, or if they occurred along branches independently of the speciation events recovered in the phylogeny. To construct our ChromoSSE nuclear and plastid models, the parameters of chromosome evolution were restricted to those included in the best ChromEvol models (chromosome gain constant rate, chromosome loss constant rate, and WGD rate) and speciation rates associated with no chromosome change, chromosome gain, chromosome loss, and WGD were incorporated. Taxon sampling probability was adjusted to 0.8 (percentage of Araliaceae genera sampled). ChromoSSE does not consider the possibility of having multiple chromosome counts for each tip of the phylogeny or the presence of missing data. Therefore, we took a conservative approach to this issue, by using the most frequent count per genus as the input for the eight genera exhibiting multiple values,

and the chromosome number inferred by ChromEvol for genera without available chromosome counts. The analyses were run with 5,500 iterations discarding the first 500 as burn-in. We ensured the effective sample sizes (ESS) of all the parameters were higher than 200 for all replicates as suggested by Freyman and Höhna (2018). The ChromoSSE results were visualized using the R package RevGadgets (Tribble et al., 2022).

2.6 Assessment of recent polyploidy based on Hyb-Seq data

Recent polyploidy events in Araliaceae were assessed using the pipeline nQuire (Weiß et al., 2018), which estimates ploidy level from target-enriched sequence data without chromosome counts or cytometric data (Viruel et al., 2019, 2023). This allowed the estimation of the ploidy levels for all the species included in the complete nuclear and plastid trees, even those without chromosome counts. This methodology estimates the allelic ratios of the SNPs of a species and compares them with those expected for different ploidy levels, allowing the estimation of the ploidy level for diploids, triploids, and tetraploids, and is especially useful to detect the signal of recent polyploidization events (Viruel et al., 2019, 2023). The trimmed reads for all samples included in the final phylogenetic trees were mapped to a reference using exclusively loci not identified as paralogs by HybPiper. This reference was used to run nQuire, using the “denoise” tool to remove basal noise that could hinder the identification of allelic frequency patterns following the pipeline of Viruel et al. (2019). Ploidy level determination was based on several criteria according to Weiß et al. (2018) and Viruel et al. (2019). The best model for each species was selected by comparing the $\Delta\log$ of the three models (2x, 3x and 4x) with the free model. The model of best fit must exhibit an elevated R^2 and y-y slope, as well as low standard error and Norm SSR to ensure sufficient adjustment. When no model exhibited a $R^2 \geq 0.1$, then the species did not fit any of the three models, which could be due to either poor data quality or alternative ploidy scenarios, such as higher levels of polyploidy. To further explore the data, we calculated allelic frequencies from the nQuire denoised results and obtained several plots per sample following Viruel et al. (2019): histograms of allelic frequencies per sample, boxplots of allelic frequencies per sample and per SNP, distribution of the allelic ratios per SNP, and density plots of allelic ratio values. Mean, median, and proportions of allelic ratios <2 per sample were calculated. Median values <2 were considered as suggestive of diploidy while those >2 as indicative of polyploidy.

3 Results

3.1 Sequence capture and alignment

Sequences were recovered for 100% of targeted nuclear regions (936 regions) and 90% of targeted plastid regions (262 regions). After removing samples with low recovery and incorrect

phylogenetic placement, 227 samples were included in the *NuTot* alignment and 225 in the *CpTot* alignment. The proportion of recovered length for targeted regions was generally high (average of 86.2% for nuclear regions and 81% for plastid regions). The concatenated nuclear alignment (*NuTot*) included 156 loci and 562,329 bp after paralog removal, while the plastid concatenated alignment (*CpTot*) was 177,454 bp long. Additional information on Hyb-Piper statistics, capture efficiency, paralog retrieval, and alignment statistics can be found in [Supplementary Data 3](#).

3.2 Phylogenetic relationships and dating

The concatenation-based nuclear analysis ([Figures 1, 2](#)) showed high support for most phylogenetic relationships among Araliaceae genera and all phylogenetic relationships among the AsPG genera. The dating analysis based on the nuclear phylogeny ([Supplementary Figure 1A](#)) recovered a crown age in the late Cretaceous or early Paleocene (72.9 [63.8–72.9] Ma) for Araliaceae. The *Hydrocotyle-Trachymene* clade exhibited the earliest divergence within Araliaceae, followed by *Harmsioplanax* (Paleocene to early Eocene, 49.4–63.4 Ma), which was sister to a clade including the rest of Araliaceae lineages. This clade was formed by two main clades: one including the clades *Aralia-Panax*, *Polyscias-Pseudopanax s.s.* and greater *Raukaua* in a polytomy, and another clade including the AsPG-*Osmoxylon* clade sister to the *Cussonia-Astropanax* clade. The divergence of *Aralia-Panax*, *Polyscias-Pseudopanax s.s.*, and the greater *Raukaua* clades occurred between the early and the late Eocene (36.9–53.8 Ma), with crown ages from the early to the late Eocene for greater *Raukaua* (33.8–49.9 Ma), from the late Eocene to the Oligocene for *Polyscias-Pseudopanax s.s.* (27.6–35.9 Ma), and from the Oligocene to the early Miocene for *Aralia-Panax* (18.8–24.2 Ma). On the other hand, the AsPG and *Osmoxylon* diverged between the late Eocene to the Oligocene (31.7–38.3 Ma), and the *Cussonia-Astropanax* clade diverged between the early to the late Eocene (34.9–44.4 Ma) with a crown age between the late Eocene to the Oligocene (27.2–36.6 Ma). The AsPG crown age was placed in the Oligocene (29.8–33.0 Ma). The earliest-diverging lineage within the AsPG was *Oplopanax*, followed by the *Heptapleurum-Tetrapanax* clade, which was sister to a clade including the rest of the AsPG lineages. Within this clade we recovered a clade with *Brassaiopsis-Trevesia* and *Fatsia-Oreopanax* clades as sisters to each other, which was sister to a clade including the *Kalopanax-Macropanax* and *Dendropanax-Gamblea* clades as consecutive sisters of a clade formed by the *Sciodaphyllum-Frodinia* clade as sister to the *Hedera-Merrillioipanax* clade. Ten of the sampled AsPG genera diverged during the Oligocene (*Dendropanax*, *Eleutherococcus*, *Fatsia*, *Hedera*, *Heptapleurum*, *Heteropanax*, *Kalopanax*, *Merrillioipanax*, *Oplopanax* and *Tetrapanax*), while the rest diverged during the Miocene.

The topology displayed by the coalescent-based nuclear phylogeny ([Supplementary Figure 2](#)) was mostly compatible with that of the concatenation-based phylogeny except for the placement of *Harmsioplanax*, which appeared as sister to the *Hydrocotyle-*

Trachymene clade. This reconstruction maintained high support for recent relationships but displayed lower support for deep relationships as a result of high gene tree discordance, with quartet scores congruent with a pattern of ILS. The coalescent-based nuclear phylogeny recovered an internal polytomy of four Araliaceae lineages: the clade formed by *Harmsioplanax* as sister to the *Hydrocotyle-Trachymene* clade, the greater *Raukaua* clade, the clade formed by *Aralia-Panax* and *Polyscias-Pseudopanax s.s.* clades, and a clade including the *Cussonia-Astropanax* clade and *Osmoxylon* as consecutive sisters of the AsPG. Within the AsPG we recovered an internal polytomy involving *Oplopanax*, *Heptapleurum-Tetrapanax*, and the clade including the remaining AsPG lineages. The latter exhibited a polytomy comprising five lineages (*Dendropanax-Gamblea* clade, *Hedera-Merrillioipanax* clade, *Kalopanax-Macropanax* clade, *Sciodaphyllum-Frodinia* clade, and a clade formed by the *Brassaiopsis-Trevesia* and *Fatsia-Oreopanax* clades).

The plastid phylogeny ([Figures 3, 4](#)) also recovered the main Araliaceae clades and those within the AsPG, and resolved the evolutionary relationships among them. However, the relationships among these clades were highly incongruent between nuclear and plastid phylogenies ([Figures 1–4; Supplementary Figure 2](#)). The *Hydrocotyle-Trachymene* clade was sister to *Harmsioplanax*, forming a clade that diverged during the late Cretaceous (67.5–72.4 Ma, [Supplementary Figure 1B](#)), sister to a clade including the rest of Araliaceae lineages. Within this highly diverse clade, *Astropanax* diverged during the Paleocene to Early Eocene (41.6–67.6 Ma), followed by the greater *Raukaua* clade, which diverged from its sister clade between the late Eocene and the Oligocene (31.8–37.4 Ma). The greater *Raukaua* clade was sister to a clade including two main clades: the AsPG as sister to a clade including *Cussonia* and *Polyscias-Pseudopanax s.s.* (Oligocene crown age, 30.4–33.1 Ma), and a clade formed by *Aralia-Panax* as sister to *Osmoxylon* (Oligocene crown age, 27.5–30.6 Ma). The AsPG, *Polyscias-Pseudopanax s.s.*, and *Aralia-Panax* exhibited crown ages during the Oligocene (25.7–29.7 Ma), between the Oligocene and early Miocene (22.1–25.9 Ma), and during the Early Miocene (19.8–22.6 Ma), respectively. Within the AsPG, *Oplopanax*, *Heptapleurum-Tetrapanax*, *Sciodaphyllum-Frodinia* and *Dendropanax-Gamblea* were consecutive sisters of a clade including two clades: *Kalopanax-Macropanax* as sister to *Brassaiopsis-Trevesia* and *Hedera-Merrillioipanax* as sister to *Fatsia-Oreopanax*.

Our nuclear and plastid phylogenies indicate that several Araliaceae genera are paraphyletic (*Aralia* and *Raukaua*, with *Panax* and *Pseudopanax lessonii* nested within them, respectively) or polyphyletic (*Dendropanax* and *Pseudopanax*). In this regard, *Dendropanax hainanensis* and *D. lancifolius* form a well-supported clade independent from the remaining *Dendropanax* species and sister to the *Chengioplanax-Gamblea* clade (*Dendropanax hainanensis-lancifolius* lineage). *Pseudopanax s.s.* formed a clade with *Meryta* sister to *Polyscias* (*Polyscias-Pseudopanax s.s.* clade), whereas the American *Pseudopanax* species formed a clade with *Cephalalaria* that had an early divergence in the greater *Raukaua* clade, and the New Zealander species *P. lessonii* was nested in *Raukaua s.s.*

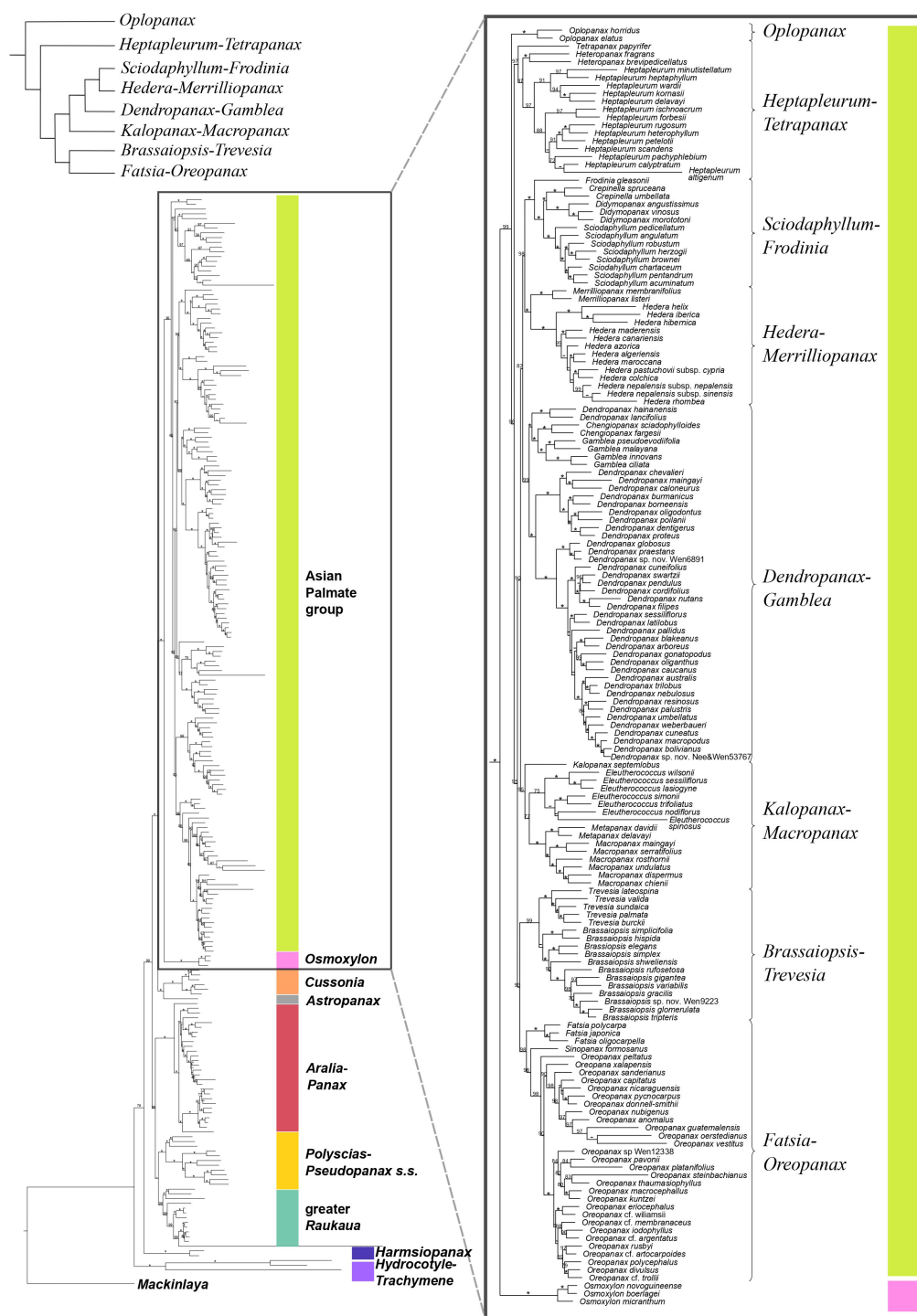


FIGURE 1

Concatenation-based maximum-likelihood phylogeny of Araliaceae based on the nuclear DNA matrix (*NuTot*) generated through Hyb-Seq, with a detailed view of the phylogenetic relationships among major clades of the Asian Palmate group (ASPG). Numbers above branches are percentage bootstrap support (BS) values, with asterisks indicating BS values equal to 100. Hyphens indicate unsupported branches.

3.3 Ancestral polyploidy in the evolution of Araliaceae

The best-fitting model explaining chromosome evolution was “DysDup” for both nuclear (AIC=40.2499, log-likelihood=-

17.1250) and plastid (AIC=40.3929, log-likelihood=-17.1965) analyses (Supplementary Data 4). This model considers ascending and descending dysploidy as well as WGD events.

The estimated ancestral chromosome number for Araliaceae was $n=12$ for both nuclear and plastid reconstructions (posterior

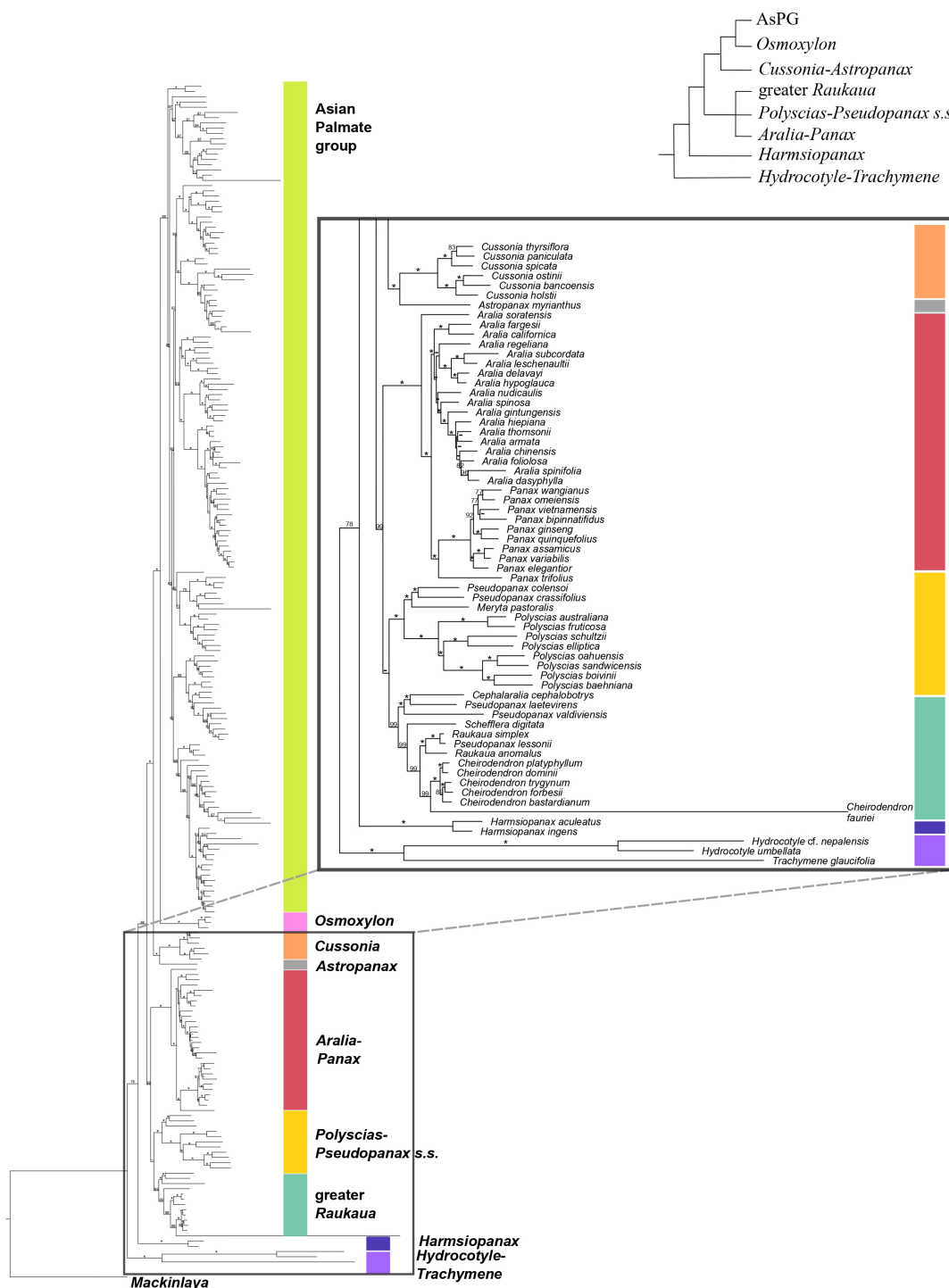


FIGURE 2

Concatenation-based maximum-likelihood phylogeny of Araliaceae based on the nuclear DNA matrix (*NuTot*) generated through Hyb-Seq, with a detailed view of the early phylogenetic relationships among major clades of Araliaceae. Numbers above branches are percentage bootstrap support (BS) values, with asterisks indicating BS values equal to 100. Hyphens indicate unsupported branches.

probabilities of 0.98 and 0.99, respectively; Figure 5) and the estimated chromosome numbers for all ancestral nodes were $n=12$ or $n=24$ in both reconstructions (12 nodes with $n=12$ and 27 with $n=24$ for the nuclear phylogeny; 13 nodes with $n=12$ and 26 with $n=24$ for the plastid phylogeny). In addition, both the nuclear

and plastid reconstructions recovered a descending dispoloidy event for *Trachymene* (from $n=12$ to $n=11$). According to the nuclear reconstruction, at least four WGD events occurred during the evolutionary history of Araliaceae (4.22 expected events), with two of them in internal branches and representing transitions

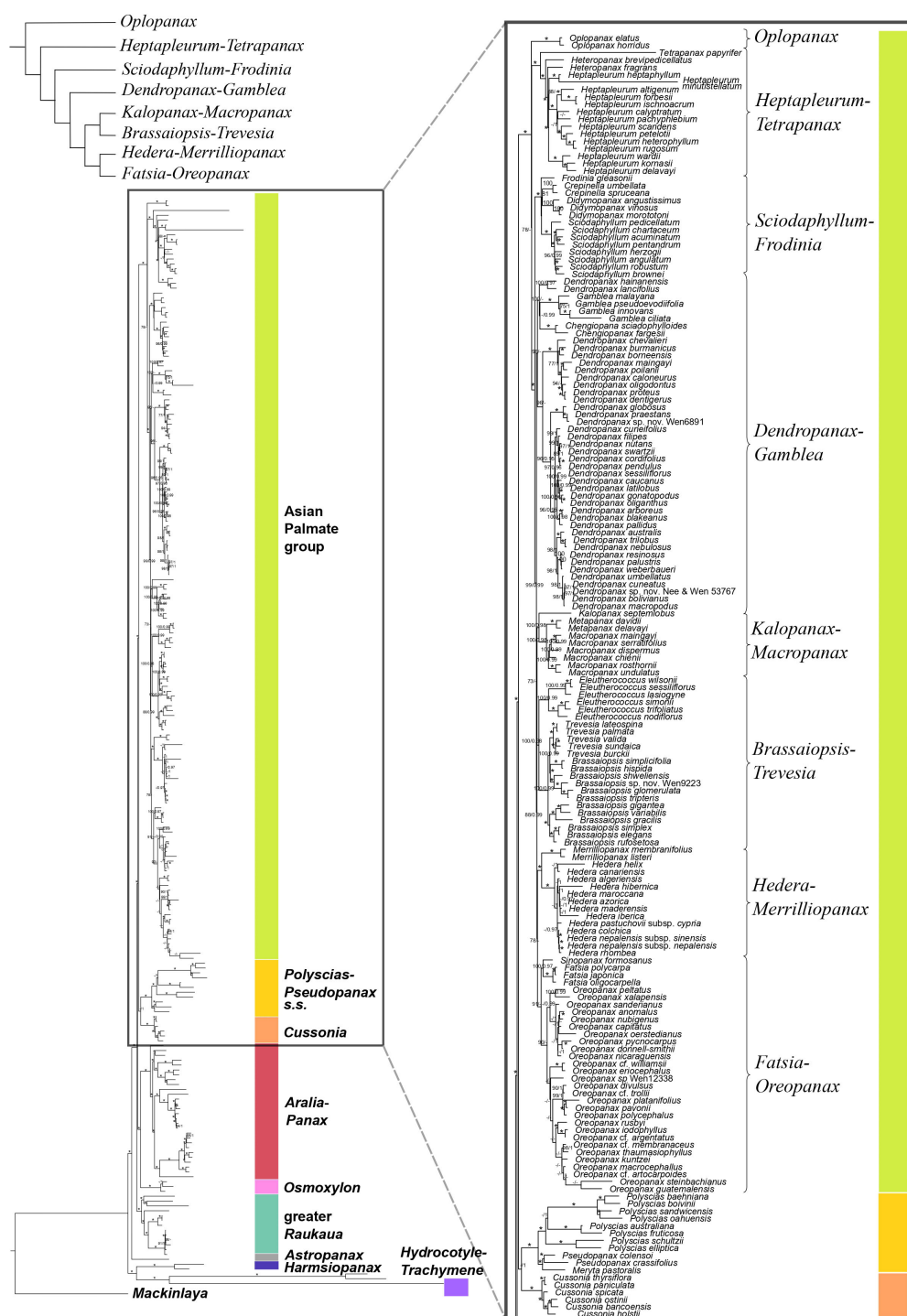


FIGURE 3

Concatenation-based maximum-likelihood phylogeny of Araliaceae based on the plastid DNA matrix (CpTot) generated through Hyb-Seq, with a detailed view of the phylogenetic relationships among major clades of the Asian Palmate group (AsPG). Numbers above branches are percentage bootstrap support (BS) and Bayesian posterior probability (PP) values, with asterisks indicating maximum support (BS = 100, PP = 1.0). Hyphens indicate unsupported branches.

from $n=12$ to $n=24$ (Figure 5A). One WGD event was retrieved at the base of the clade formed by the AsPG, *Osmoxylon*, *Cussonia* and *Astropanax* (N55, expected transitions=0.78). The second ancient WGD event was recovered at the base of the *Polyscias*-*Pseudopanax* s.s. clade (N52, expected transitions=0.61). The probability of WGD

events was also relatively elevated for *Mackinlaya* (expected transitions=0.32) and *Hydrocotyle* (expected transitions=0.30) and the base of the *Pseudopanax* s.s.-*Meryta* clade (N53, expected transitions=0.25). On the other hand, the plastid reconstruction (Figure 5B) supported at least two WGD events during the

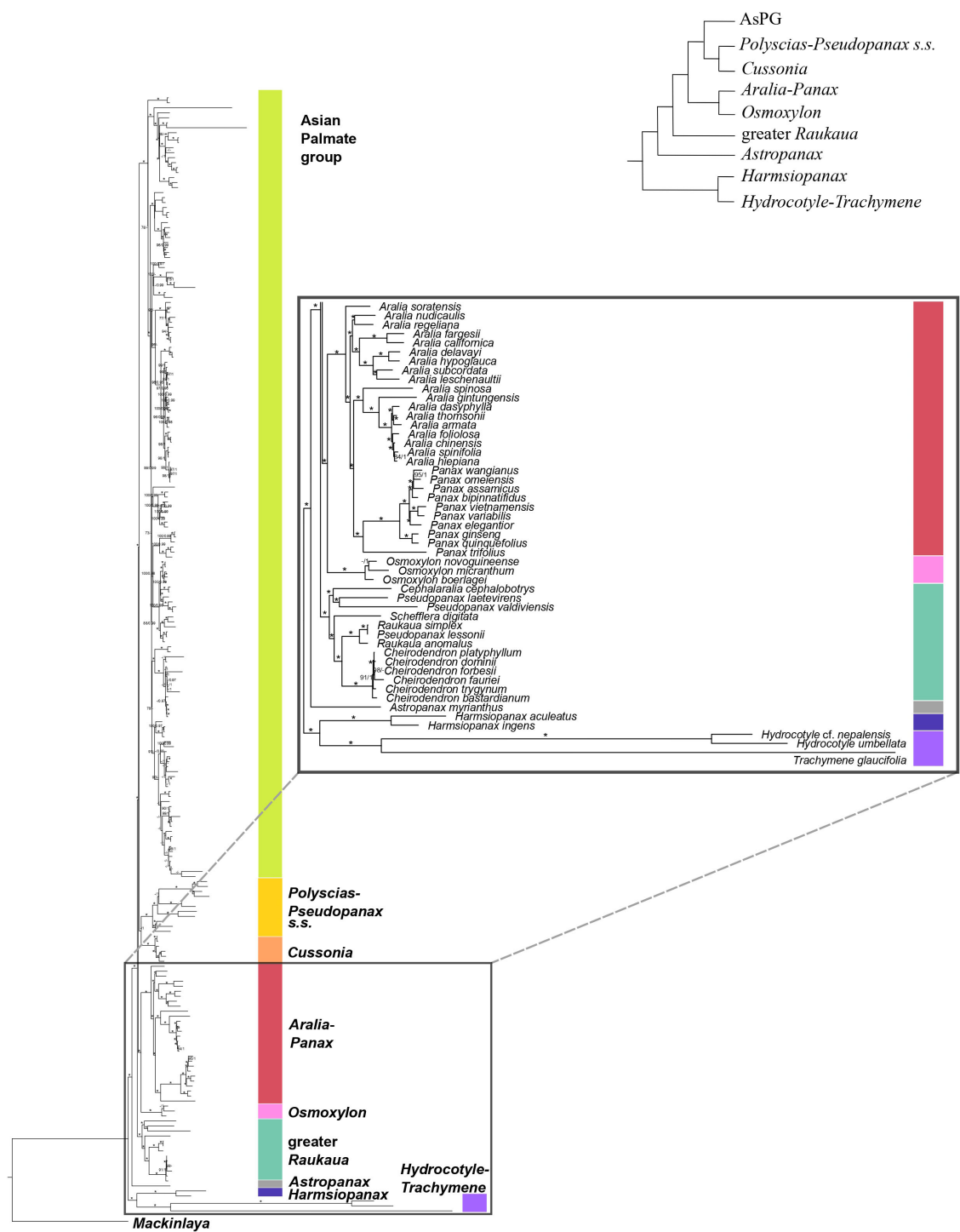
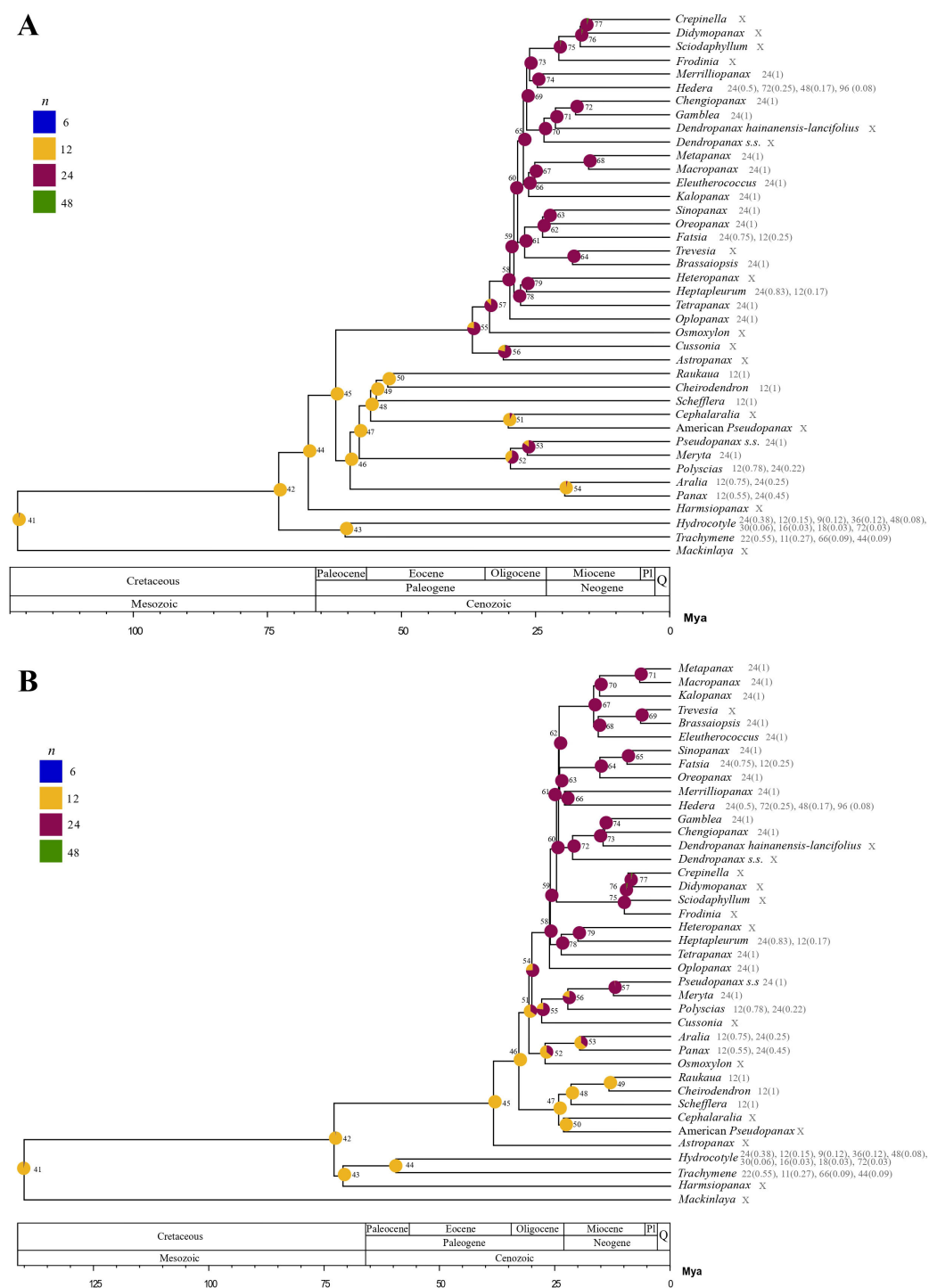


FIGURE 4
Concatenation-based maximum-likelihood phylogeny of Araliaceae based on the plastid DNA matrix (CpTot) generated through Hyb-Seq, with a detailed view of the early phylogenetic relationships among major clades of Araliaceae. Numbers above branches are percentage bootstrap support (BS) and Bayesian posterior probability (PP) values, with asterisks indicating maximum support (BS = 100, PP = 1.0). Hyphens indicate unsupported branches.

evolutionary history of Araliaceae (2.82 expected events), one of them in an internal branch. This WGD event was recovered at the base of the clade including the AsPG, *Polyscias-Pseudopanax* s.s. and *Cussonia* (N54, expected transitions=0.40) and represents a change from $n=12$ to $n=24$. The probability of WGD events was also relatively elevated for *Macklinaya* (expected transitions=0.32) and *Hydrocotyle* (expected transitions=0.26).

Regarding the nature of chromosome number changes, the distribution of the parameters in ChromoSSE models supported that chromosome number changes were mostly unrelated to



speciation (Figure 6; Supplementary Figure 3, Supplementary Data 5). For chromosome change events, the highest rates were obtained for the polyploidization parameter, while for the parameters assessing the relationship between chromosome number changes

and speciation the highest rates were found for no chromosome changes, several orders of magnitude higher than the rest of the rates. These results indicated that chromosome number changes were mostly asynchronous with speciation in Araliaceae. The

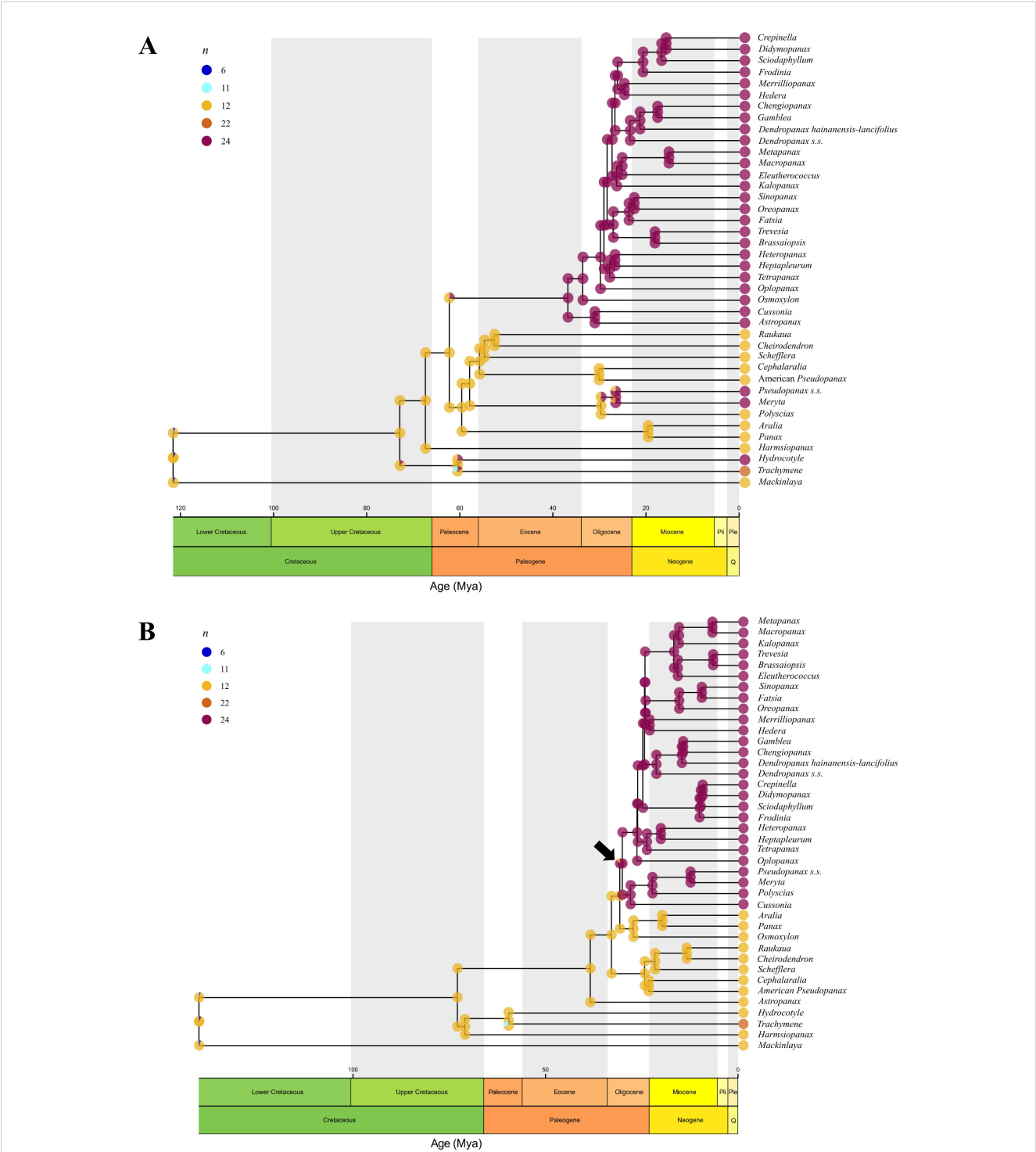


FIGURE 6
Ancestral chromosome number reconstructions for Araliaceae based on ChromoSSE models for nuclear (A) and plastid (B) time-calibrated phylogenies. Ancestral probabilities for each chromosome number are provided as pie charts with different colors for each ancestral number. Only chromosome numbers with inferred ancestral probabilities above zero in at least one ancestral node are shown in the legend. Chromosome counts and their proportions are provided for each genus.

ChromoSSE nuclear model (Figure 6A) recovered two ancient WGDs unrelated with speciation. The ancient WGD event at the base of the clade formed by the AsPG, *Osmoxylon*, *Cussonia* and *Astropanax*, recovered using ChromEvol, was recovered again by ChromoSSE. However, the ancient WGD event at the base of the

Polyscias-*Pseudopanax* s.s. clade recovered with ChromEvol was instead recovered within that clade at the base of the *Meryta*-*Pseudopanax* s.s. clade in the ChromoSSE nuclear reconstruction. Our plastid ChromoSSE model recovered a single WGD event that was associated with the origin of the clade formed by the AsPG,

Polyscias-Pseudopanax s.s. and *Cussonia* (Figure 6B), congruently with ChromEvol results (Figure 5B).

3.4 Assessment of recent ploidy events within the evolution of Araliaceae genera

Eighty-eight percent of the analyzed species conformed to one of the ploidy levels tested with nQuire (2x, 3x or 4x; 199 species, Supplementary Data 6). Most of the samples were characterized as diploids (174 taxa), while 23 were characterized as tetraploids and two were inferred as triploids (*Hedera colchica* and *Aralia subcordata*). The inferred tetraploids were scattered across different Araliaceae clades, with most of them recovered for *Aralia-Panax* (12 species in two genera) and the AsPG (7 species in two genera). Median allelic frequencies ranged from 1.16 to 1.77 for diploids, 1.78 to 2.11 for triploids, and 1.30 to 2.95 for tetraploids (Supplementary Data 7). The proportion of diploid loci (allelic ratio <2) ranged from 64.1 to 95.6% in diploids, 36.3 to 62.3% in triploids and 11.0 to 60.7% in tetraploids. Exploratory plots of allelic frequencies for each sample are provided in Supplementary Figure 4.

4 Discussion

4.1 Phylogenomics increases the resolution of the AsPG and Araliaceae backbones

Until recently, the knowledge of early evolutionary relationships in Araliaceae was based on studies using a small number of nuclear and plastid DNA regions (Wen et al., 2001; Plunkett et al., 2004a, b, 2020; Yi et al., 2004; Nicolas and Plunkett, 2009; Mitchell et al., 2012; Valcárcel et al., 2014; Li and Wen, 2016). In addition, phylogenetic studies have been frequently focused on the most diverse clades (AsPG, *Polyscias-Pseudopanax* s.s. and *Aralia-Panax*), while the lineages with low species diversity have usually been underrepresented. As a result of this limited genetic data and biased generic sampling, the early evolutionary patterns in Araliaceae remain poorly understood. The studies conducted so far at the family level consistently pointed to an early divergence of the genera *Hydrocotyle*, *Trachymene* and/or *Harmsioplanax* as sisters to a clade including the core Araliaceae (Plunkett et al., 2004a, b; Yi et al., 2004; Nicolas and Plunkett, 2009; Mitchell et al., 2012; Valcárcel et al., 2014; Li and Wen, 2016), and they recovered a large polytomy at the base of the core Araliaceae. The recent application of the Hyb-Seq technique to sequence hundreds of loci along with higher generic sampling (65% of Araliaceae genera; Gallego-Narbón et al., 2022), recovered well resolved phylogenetic trees. These revealed incongruent relationships when comparing the plastid and nuclear trees, suggesting hybridization events were involved in the origin of the core Araliaceae, as well as in the origin of two species-rich clades (AsPG and the *Aralia-Panax* clade). After improving the generic sampling to 80% of Araliaceae genera, the results of this study confirmed the previous findings in Gallego-

Narbón et al. (2022). On the one hand, the early divergence of the *Hydrocotyle-Trachymene* clade was confirmed, as well as the incongruent placement of *Harmsioplanax*, that either appears as sister to this clade or to the core Araliaceae. On the other hand, this study recovered (i) the AsPG as sister to *Osmoxylon* (Figure 1) or to the clade formed by *Polyscias-Pseudopanax* and *Cussonia* (Figure 3), (ii) the *Aralia-Panax* clade as sister to *Polyscias-Pseudopanax* or to *Osmoxylon*, and (iii) *Astropanax* as sister to *Cussonia*, forming a clade sister to the AsPG-*Osmoxylon* clade or as the earliest-diverging clade within the core. Altogether, these results preclude resolving the early evolutionary relationships in Araliaceae due to extensive phylogenetic incongruence between gene trees (Supplementary Figure 2) and between nuclear and plastid data (Figures 1, 2 vs 3, 4). Given that, for most of the nodes, gene tree incongruence supports a pattern of ILS (Schumer et al., 2016), our results provide evidence of early radiation in the evolution of Araliaceae. In addition, previous hybridization analyses in Araliaceae provided evidence of hybridization events involving the AsPG and other Araliaceae clades (*Astropanax* and greater *Raukaua*, Gallego-Narbón et al., 2022). Considering our results in this context, we interpret that they provide evidence of inter-lineage hybridization at a deep evolutionary level in the family with a subsequent evolutionary radiation, similar to the pattern previously inferred for the early evolution of its largest clade, the AsPG (Gallego-Narbón et al., 2022).

The present results improved the phylogenetic resolution in the AsPG. Previous research based on a limited number of loci recovered an internal polytomy involving most of the main clades of the AsPG (Lowry et al., 2004; Plunkett et al., 2004a; Li and Wen, 2013, 2016; Valcárcel et al., 2014). Subsequent studies based on plastid genomes with limited lineage sampling improved the resolution in comparison with Sanger-based studies (Valcárcel and Wen, 2019; Kang et al., 2023), but the deep relationships within the AsPG remained unresolved. The use of Hyb-Seq led to a significant improvement in the resolution, with only one internal node unresolved for the nuclear-based phylogeny of the AsPG and two internal nodes for the plastid-based phylogeny (Gallego-Narbón et al., 2022). Here, the increased generic and species sampling enabled the resolution of all the internal nodes in both nuclear and plastid trees when concatenation-based methods were applied (Figures 1, 3). Both trees supported an early divergence of *Oplopanax* followed by *Heptapleurum-Tetrapanax* as sister to a clade including the rest of the AsPG lineages. This topology was supported in previous studies based on nuclear loci (Gallego-Narbón et al., 2022). Research using Sanger sequencing of plastid DNA also suggested this topology (Valcárcel et al., 2014; Li and Wen, 2016; Plunkett et al., 2020), but only Valcárcel and Wen (2019) obtained enough support for these early relationships in the AsPG, which were confirmed here. In addition, our nuclear tree helped clarify the sister relationships of two clades (*Hedera-Merrillioplanax* and *Kalopanax-Macropanax*), that, according to Gallego-Narbón et al. (2022), were part of a polytomy alongside a clade including the rest of the lineages (*Dendropanax-Gamblea* clade, *Sciadaphyllum-Frodinia* clade, and the clade formed by *Brassaiopsis-Trevesia* and *Fatsia-Oreopanax* clades). Specifically,

the present nuclear reconstruction (Figure 1) supports *Hedera-Merrillioanax* as sister to the *Sciodaphyllum-Frodinia* clade, with *Dendropanax-Gamblea* and *Kalopanax-Macropanax* as consecutive sisters of this clade. However, all these clades are part of a polytomy according to the coalescence-based analysis (Supplementary Figure 2), with a distribution of quartet scores congruent with a pattern of ILS (equal values of Q2 and Q3; Schumer et al., 2016). These results provide further support for the pattern of an evolutionary radiation during the early evolution of the AsPG that has been discussed in previous literature (Valcárcel and Wen, 2019; Gallego-Narbón et al., 2022). In addition, our plastid tree (Figures 3, 4) is topologically incongruent with both nuclear trees (Figures 1, 2; Supplementary Figure 2), which we interpret as evidence of the hybridization events identified in Gallego-Narbón et al. (2022) for the early evolution of the AsPG.

4.2 Polyploidization and evolution in Araliaceae

The ancestral chromosome number reconstructions consistently recovered WGDs around the Eocene-Oligocene that coincided in time with the origin of two of the most species-rich lineages (Figures 5, 6). According to the chromosome reconstruction based on the nuclear tree (Figure 5A), one ancient WGD occurred at the base of the clade formed by AsPG-*Osmoxylon* and *Cussonia-Astropanax* and a second one at the base of the *Polyscias-Pseudopanax s.s.* lineage. However, according to the reconstruction based on the plastid tree, only one ancient WGD occurred around this time, placed at the base of a clade including the AsPG, *Polyscias-Pseudopanax s.s.* and *Cussonia* (Figure 5B). The link between these WGDs and speciation is unclear. According to the nuclear reconstruction the ancient WGD that preceded the origin of the AsPG cannot be linked to speciation (Figure 6A) while the WGD detected in the plastid reconstruction prior to the origin of the AsPG is linked to speciation (Figure 6B). The incongruence in the number and placement of WGD events between nuclear and plastid reconstructions was expected, considering the extensive incongruence we detected between the two topologies for these clades (Figures 1, 2 vs 3, 4). Despite these incongruent results, the two reconstructions (nuclear and plastid) agreed on the recovery of ancient WGDs predating the origin of the same lineages, that currently include polyploids and display the highest number of species (AsPG, *Polyscias-Pseudopanax s.s.* and *Cussonia*). We interpret the ancient WGD predating the origin of the AsPG as a background factor for the early radiation of this group. This follows the rationale described by Bouchenak-Khelladi et al. (2015) that the factors present before diversification bursts provide the necessary conditions for future evolutionary radiations, which is related to the WGD radiation time-lag model set by Schranz et al. (2012). These authors identified a consistent pattern of ancient WGDs in angiosperms leading to species-rich clades sister to species-poor clades, suggesting that there is a one-node lag between the acquisition of the novelties associated to WGDs and the subsequent radiations. Later on, this hypothesis was expanded by

the identification of a general pattern in which the lag between ancient WGDs and the subsequent diversification bursts generally involved several nodes (Landis et al., 2018). The results of the present study are congruent with this time-lag hypothesis, as the ancient WGDs we recovered in Araliaceae predated by one or two-nodes the origin of two of the three most species-rich clades of Araliaceae (AsPG and *Polyscias-Pseudopanax s.s.*), that together encompass approximately 70% of the species diversity of the Araliaceae family (Plunkett and Lowry, 2010; Valcárcel and Wen, 2019).

Given the extensive pattern of hybridization found in those Araliaceae clades where hybridization has been studied in more detail (e.g., AsPG, Gallego-Narbón et al., 2022; *Aralia-Panax*, Liu et al., 2023; Zhang et al., 2025), and the consistent pattern of genomic incongruence affecting the divergence of the main clades of Araliaceae obtained in this study, we consider that ancient WGDs could be the result of hybridization events (allopolyploidy). Gallego-Narbón et al. (2022) hypothesized that the WGD associated with the origin of the AsPG involved the genera *Polyscias* and *Osmoxylon*. We propose that an ancestor of the African genera of Araliaceae (*Astropanax* and *Cussonia*) was involved in the WGD at the origin of Araliaceae. While our nuclear trees supported the close relationship between the *Astropanax-Cussonia* and the AsPG-*Osmoxylon* clade (Figures 1, 2; Supplementary Figure 2), the plastid analysis supports that *Astropanax* is an early-diverging lineage in the Araliaceae core clade and *Cussonia* forms a clade with *Polyscias-Pseudopanax s.s.* (Figures 3, 4). Gallego-Narbón et al. (2022) performed phylogenetic network analyses for the family Araliaceae, with all networks supporting a hybrid origin for the AsPG. Most of the networks supported *Osmoxylon* as the major sister of the AsPG and *Astropanax* as the minor sister, which would mean that an ancestor of *Osmoxylon* and an ancestor of *Astropanax* were the parental lineages involved in the hybrid origin of the AsPG. However, these analyses generally supported that there was an additional hybridization event involved in the origin of the AsPG, which means that several hybridization events could be involved in the origin of the AsPG. Further research providing additional chromosome counts and phylogenetic trees with increased species sampling across the family Araliaceae are required to precisely determine the placement of allopolyploidization events in the evolution of Araliaceae, and their role in the origin and diversification of the AsPG.

4.3 Allelic frequencies and whole genome duplications

The use of allelic frequencies derived from Hyb-Seq data provides a novel approach to infer ploidy levels without requiring fresh material (Weiß et al., 2018). This methodology is promising, and several pipelines have been released in recent years, but their performance has not yet been intensively explored (e.g., nQuire, Weiß et al., 2018; and nQuack, Gaynor et al., 2024). In our study, the use of allelic frequencies derived from Hyb-Seq data provided further insights into the distribution of different ploidy levels across

the family Araliaceae and provided evidence of recent polyploidization events. This methodology allowed the detection of polyploids across the three main lineages of Araliaceae (AsPG, *Polyscias-Pseudopanax* s.s. and *Aralia-Panax*; [Supplementary Figure 4](#) and [Supplementary Data 6](#) and 7), which agrees with previous information based on chromosome counts ([Yi et al., 2004](#)). The most relevant result was the detection of abundant tetraploids in the *Aralia-Panax* clade, which supports the relevance of recent polyploidization events in the evolution of this clade. However, the correlation between ploidy levels derived from allelic frequencies and those obtained from direct ploidy measurements has not been extensively tested in previous studies and the combination of this methodology with direct measurements is still recommended ([Viruel et al., 2019](#); [Gaynor et al., 2024](#)).

In fact, while all existing evidence provided by chromosome counts and ancestral reconstructions supported the polyploid origin of the AsPG ([Yi et al., 2004](#); [Gallego-Narbón et al., 2022](#); [Figures 5, 6](#)), our analysis indicates that most of the sampled AsPG species show a pattern of allelic frequencies compatible with diploidy ([Supplementary Data 6](#)). These results showcase that ploidy inference based on allelic frequencies is not suitable to identify the signal of ancient polyploidization events. This pattern was already pointed out by the authors of the methodology, who emphasized the usefulness of allelic frequencies to study polyploidization at an intraspecific scale ([Weiß et al., 2018](#)). Our inference of ancient polyploidization events ([Figures 5, 6](#)) and extensive detection of diploids based on allelic frequencies in the AsPG ([Supplementary Data 6](#)) suggest that this clade might have undergone a process of diploidization, which is common after allopolyploidization in angiosperms ([Leitch and Bennett, 2004](#)).

It is also remarkable that we found several incongruences between the ploidy level inferred from allelic frequencies and that obtained from chromosome counts for those genera whose ploidy levels have been studied more extensively. In this regard, chromosome counts of *Hedera* have been used to study ploidy for the twelve species of the genus ([Vargas et al., 1999](#)). The use of allelic frequencies allowed the accurate inference of diploidy for five species (*H. azorica*, *H. canariensis*, *H. helix*, *H. maroccana* and *H. nepalensis*) and species known to be hexaploid (*H. iberica*, *H. maderensis*, *H. pastuchovii*) were incompatible with diploidy, triploidy, or tetraploidy, as expected ([Supplementary Data 6](#)). However, the ploidy levels inferred for the four remaining species of the genus were incorrect, with a diploid and a tetraploid species not fitting into any of the tested models (*H. rhombea* and *H. algeriensis*, respectively), an octoploid characterized as triploid (*H. colchica*) and a tetraploid (*H. hibernica*) showing allelic frequencies typical of a diploid. Previous research on the plant genus *Dioscorea* showed that allelic frequencies of recent autotetraploids may be comparable to those of diploids ([Viruel et al., 2019, 2023](#)). However, this is not the case for *H. hibernica*, which is known to be an allotetraploid ([Vargas et al., 1999](#)). The incongruence between chromosome counts and allelic frequencies may be the result of the extensive hybridization that has been reported for *Hedera* based on molecular data ([Vargas et al., 1999](#); [Ackerfield and Wen, 2003](#)), as

hybridization can influence the pattern of allelic frequencies ([Gaynor et al., 2024](#)). In addition, the *Hedera* polyploids whose ploidy level was incorrectly inferred generally exhibited early divergences during the evolutionary history of *Hedera* according to previous research ([Gallego-Narbón et al., 2023](#)), which provides further support for the application of ploidy inference techniques based on allelic frequencies exclusively for species of recent origin. Nevertheless, we acknowledge that new methodologies for inferring ploidy from allelic frequencies, published during the development of our study and improving upon nQuire ([Gaynor et al., 2024](#)), are promising for addressing some of the limitations of our approach.

In addition, we observed an extensive overlap in the median allelic frequencies between species categorized as tetraploids and those categorized as diploids by the models ([Supplementary Data 7](#)). This overlap is not observed for the proportion of diploid loci (allelic ratio <2) and the distributions of allelic frequencies are clearly different between diploids and tetraploids ([Supplementary Figure 4](#)). In this regard, model parameters ([Supplementary Data 6](#)) must be interpreted together with the median and distribution of allelic frequencies and the percentage of diploid loci ([Supplementary Data 7](#) and [Supplementary Figure 4](#)) for a more accurate assessment of ploidy, especially for those samples in which different model parameters support different ploidy models. Despite these limitations, our results still support the relevance of WGDs not only in the early evolutionary history of the family, but also in the evolutionary history of several genera across different clades of Araliaceae. Additionally, they provide a preliminary assessment of ploidy in the family that should be evaluated in further detail in future research incorporating novel ploidy inference methodologies and direct ploidy measurements.

5 Conclusions

We obtained the first nuclear and plastid phylogenomic trees of the Asian Palmate group of Araliaceae with full internal support as a result of a high taxonomic and genetic sampling. Our ancestral chromosome number reconstructions based on these phylogenies provide further support for the occurrence of a WGD event related to the origin of the AsPG. According to the reconstruction based on the plastid tree, this polyploidization event not only preceded the origin of the AsPG but also the origin of the *Polyscias-Pseudopanax* s.s. clade. The use of information provided by allele frequencies allowed the detection of several recent polyploids in Araliaceae, especially for the *Aralia-Panax* clade. However, this methodology failed to accurately detect the signal of ancient polyploidization and recent polyploidization in lineages with extensive hybridization (e.g., *Hedera*). Therefore, the application of direct methods to measure ploidy is still recommended in such cases. In summary, our results support that ancient WGD events were involved in the origin of two of the three most diversified lineages of Araliaceae (AsPG and *Polyscias-Pseudopanax* s.s.) and polyploidization was involved in the recent evolutionary history of the third most diversified lineage

(*Aralia-Panax* clade). Further research expanding the taxon sampling and providing additional cytogenetic data will be necessary to disentangle the spatiotemporal context of polyploidization events and their role in the diversification of the family Araliaceae.

Data availability statement

The raw reads (FASTQ files) used in this study are available in the NCBI Sequence Read Archive (SRA) database (Bioproject ID: PRJNA841627).

Author contributions

AG-N: Investigation, Visualization, Writing – original draft, Writing – review & editing, Formal Analysis, Conceptualization. GJ: Writing – review & editing, Investigation. MF-M: Writing – review & editing, Formal Analysis. JW: Writing – review & editing, Investigation, Resources. VV: Writing – original draft, Conceptualization, Funding acquisition, Resources, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This study was supported by the Spanish Ministry of Economy, Industry, and Competitiveness (CGL2017-87198-P), the Spanish Ministry of Science and Innovation (PID2019-106840GA-C22 and PID2022-140985NB-C22), the Laboratories of Analytical Biology of the Smithsonian National Museum of Natural History (<https://ror.org/05b8c0r92>), and the Smithsonian Institution DNA Barcode Network (SIBN). A. Gallego-Narbón was supported by the program “Contratos predoctorales para Formación de Personal Investigador FPI-UAM” of Universidad Autónoma de Madrid (FPI-UAM 2018) and the Smithsonian Institution Fellowship program (Graduate Student Fellowship 2022, Peter Buck Postdoctoral Fellowship 2024). M. Fernández-Mazuecos was supported by a Ramón y Cajal Fellowship (RYC2022-036418-I), funded by the Spanish Ministry of Science and Innovation (MCIN/AEI/10.13039/501100011033) and the European Social Fund Plus (FSE+).

References

- Aberer, A. J., Kobert, K., and Stamatakis, A. (2014). ExaBayes: massively parallel bayesian tree inference for the whole-genome era. *Mol. Biol. Evol.* 31, 2553–2556. doi: 10.1093/molbev/msu236
- Ackerfield, J., and Wen, J. (2003). Evolution of *Hedera* (the ivy genus, Araliaceae): insights from chloroplast DNA data. *Int. J. Plant Sci.* 164, 593–602. doi: 10.1086/375423
- Bado, S., Forster, B. P., Nielen, S., Ali, A. M., Lagoda, P. J. L., Till, B. J., et al. (2015). “Plant mutation breeding: current progress and future assessment,” in *Plant breeding reviews: volume 39*. Ed. J. Janick (Hoboken, New Jersey: Wiley), 23–88. doi: 10.1002/9781119107743.ch02
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, S., et al. (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021
- 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
- Borowiec, M. L. (2016). AMAS: a fast tool for alignment manipulation and computing of summary statistics. *PeerJ* 4, e1660. doi: 10.7717/peerj.1660
- Bouchenak-Khelladi, Y., Onstein, R. E., Xing, Y., Schwery, O., and Linder, H. P. (2015). On the complexity of triggering evolutionary radiations. *New Phytol.* 207, 313–326. doi: 10.1111/nph.13331
- Drummond, A. J., Suchard, M. A., Xie, D., and Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* 29, 1969–1973. doi: 10.1093/molbev/mss075
- Fiaschi, P., Lowry, P. P., and Plunkett, G. M. (2020). Studies in Neotropical Araliaceae. III. Resurrection of the New World genus *Didymopanax* Decne. &

Acknowledgments

The authors are grateful to A. Talavera and S. Winslow for their support in laboratory tasks, and to A. Otero, L. Valente and I. Masa-Iranzo for their advice on statistical analyses. All the laboratory work was conducted in and with the support of the Laboratories of Analytical Biology (<https://ror.org/05b8c0r92>) of the Smithsonian National Museum of Natural History.

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 RS declared a past co-authorship with the author MF-M to the handling editor.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

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/fpls.2025.1595321/full#supplementary-material>

- Planch., previously included in *Schefflera* (Araliaceae). *Brittonia* 72, 16–22. doi: 10.1007/s12228-019-09604-w
- Freyman, W. A., and Höhna, S. (2018). Cladogenetic and anagenetic models of chromosome number evolution: a bayesian model averaging approach. *Syst. Biol.* 67, 195–215. doi: 10.1093/sysbio/syx065
- Gallego-Narbón, A., Alonso, A., Válcárcel, V., and Fernández-Mazuecos, M. (2023). Repeated asynchronous evolution of single-species endemics of ivies (*Hedera* L.) in Macaronesian archipelagos. *J. Biogeogr.* 50, 1763–1777. doi: 10.1111/jbi.14690
- Gallego-Narbón, A., Wen, J., Liu, J., and Válcárcel, V. (2022). Hybridization and genome duplication for early evolutionary success in the Asian Palmate group of Araliaceae. *J. Syst. Evol.* 60, 1303–1318. doi: 10.1111/jse.12906
- Gaynor, M. L., Landis, J. B., O'Connor, T. K., Laport, R. G., Doyle, J. J., Soltis, D. E., et al. (2024). nQuack: An R package for predicting ploidal level from sequence data using site-based heterozygosity. *Appl. Plant Sci.* 12, e11606. doi: 10.1002/aps3.11606
- Glick, L., and Mayrose, I. (2014). ChromEvol: assessing the pattern of chromosome number evolution and the inference of polyploidy along a phylogeny. *Mol. Biol. Evol.* 31, 1914–1922. doi: 10.1093/molbev/msu122
- Goldblatt, P., and Johnson, D. E. (2006). "Index to plant chromosome numbers 2001–2003," in *Monographs in systematic botany from the missouri botanical garden: volume 106* (St. Louis, Missouri), 1–242.
- Höhna, S., Landis, M. J., Heath, T. A., Boussau, B., Lartillot, N., Moore, B. R., et al. (2016). RevBayes: bayesian phylogenetic inference using graphical models and an interactive model-specification language. *Syst. Biol.* 65, 726–736. doi: 10.1093/sysbio/syw021
- Johnson, M. G., Gardner, E. M., Liu, Y., Medina, R., Goffinet, B., Shaw, A. J., et al. (2016). HybPiper: extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. *Appl. Plant Sci.* 4, 1600016. doi: 10.3732/apps.1600016
- Kang, J.-S., Giang, V. N. L., Park, H.-S., Park, Y. S., Cho, W., Nguyen, V. B., et al. (2023). Evolution of the Araliaceae family involved rapid diversification of the Asian Palmate group and *Hydrocotyle* specific mutational pressure. *Sci. Rep.* 13, 22325. doi: 10.1038/s41598-023-49830-7
- Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010
- 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
- Kong, W. S. (2000). Vegetational history of the Korean peninsula. *Glob. Ecol. Biogeogr.* 9, 391–402. doi: 10.1046/j.1365-2699.2000.00203.x
- Landis, J. B., Soltis, D. E., Li, Z., Marx, H. E., Barker, M. S., Tank, D. C., et al. (2018). Impact of whole-genome duplication events on diversification rates in angiosperms. *Am. J. Bot.* 105, 348–363. doi: 10.1002/ajb2.1060
- Leitch, I. J., and Bennett, M. D. (2004). Genome downsizing in polyploid plants. *Biol. J. Linn. Soc.* 82, 651–663. doi: 10.1111/j.1095-8312.2004.00349.x
- 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, R., and Wen, J. (2013). Phylogeny and biogeography of *Dendropanax* (Araliaceae), an amphi-pacific disjunct genus between tropical/subtropical Asia and the Neotropics. *Syst. Bot.* 38, 536–551. doi: 10.1600/036364413X666606
- Li, R., and Wen, J. (2016). Phylogeny and diversification of Chinese Araliaceae based on nuclear and plastid DNA sequence data. *J. Syst. Evol.* 54, 453–467. doi: 10.1111/jse.12196
- Liu, J., Nie, Z.-L., Ren, C., Su, C., and Wen, J. (2023). Phylogenomics of *Aralia* sect. *Aralia* (Araliaceae): signals of hybridization and insights into its species delimitations and intercontinental biogeography. *Mol. Phylogenet. Evol.* 181, 107727. doi: 10.1016/j.ympev.2023.107727
- Lowry, P. P., Plunkett, G. M., Mora, M. M., Cano, A., Fiaschi, P., Frodin, D. G., et al. (2020). Studies in Neotropical Araliaceae. I. Resurrection of the genus *Sciodaphyllum* P. Browne to accommodate most New World species previously included in *Schefflera* J. R. Forst. & G. Forst. *Brittonia* 72, 1–15. doi: 10.1007/s12228-019-09593-w
- Lowry, P. P., Plunkett, G. M., and Neill, D. A. (2019). Studies in Neotropical Araliaceae. II. Resurrection of the Neotropical genus *Crepinella* for aClade of New World species previously included in *Schefflera* (Araliaceae). *Novon* 27, 253–261. doi: 10.3417/2019510
- Lowry, P. P., Plunkett, G. M., Wen, J., Van Wyk, B.-E., and Tilney, P. M. (2004). Generic relationships in Araliaceae: looking into the crystal ball. *S. Afr. J. Bot.* 70, 382–392. doi: 10.1016/S0254-6299(15)30221-0
- Magallón, S., Gómez-Acevedo, S., Sánchez-Reyes, L. L., and Hernández-Hernández, T. (2015). A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. *New Phytol.* 207, 437–453. doi: 10.1111/nph.13264
- Mayrose, I., Barker, M. S., and Otto, S. P. (2010). Probabilistic models of chromosome number evolution and the inference of polyploidy. *Syst. Biol.* 59, 132–144. doi: 10.1093/sysbio/syp083
- Mitchell, A., Li, R., Brown, J. W., Schönberger, I., and Wen, J. (2012). Ancient divergence and biogeography of *Raukaua* (Araliaceae) and close relatives in the southern hemisphere. *Aust. Syst. Bot.* 25, 432. doi: 10.1071/SB12020
- Nicolas, A. N., and Plunkett, G. M. (2009). The demise of subfamily Hydrocotylaceae (Apiaceae) and the re-alignment of its genera across the entire order Apiales. *Mol. Phylogenet. Evol.* 53, 134–151. doi: 10.1016/j.ympev.2009.06.010
- Nicolas, A. N., and Plunkett, G. M. (2014). Diversification times and biogeographic patterns in Apiales. *Bot. Rev.* 80, 30–58. doi: 10.1007/s12229-014-9132-4
- Plunkett, G. M., Chandler, G. T., Lowry, P. P., Pinney, S. M., Sprenkle, T. S., van Wyk, B.-E., et al. (2004b). Recent advances in understanding Apiales and a revised classification. *S. Afr. J. Bot.* 70, 371–381. doi: 10.1016/S0254-6299(15)30220-9
- Plunkett, G. M., and Lowry, P. P. (2010). Paraphyly and polyphyly in *Polyscias* sensu lato: molecular evidence and the case for re-circumscribing the "pinnate genera" of Araliaceae. *Plant Div. Evol.* 128, 23–54. doi: 10.1127/1869-6155/2010/0128-0002
- Plunkett, G. M., Lowry, P. P., Fiaschi, P., Frodin, D. G., and Nicolas, A. N. (2020). Phylogeny, biogeography, and morphological evolution among and within the Neotropical and Asian clades of *Schefflera* (Cham, Switzerland: Araliaceae). *TAXON* 68, 1278–1313. doi: 10.1002/tax.12177
- Plunkett, G. M., Lowry, P. P., and Neill, D. A. (2021). Studies in Neotropical Araliaceae. VII. Two new genera, *Cephalopanax* and *Frodinia*, to accommodate the remaining species of Neotropical *Schefflera*. *Brittonia* 73, 251–261. doi: 10.1007/s12228-021-09660-1
- Plunkett, G. M., Wen, J., and Lowry, P. P. (2004a). Intrafamilial classifications and characters in Araliaceae: Insights from the phylogenetic analysis of nuclear (ITS) and plastid (trnL-trnF) sequence data. *Plant Syst. Evol.* 245, 1–39. doi: 10.1007/s00606-003-0101-3
- Plunkett, G. M., Wen, J., Lowry, P. P., Mitchell, A. D., Henwood, M. J., and Fiaschi, P. (2018). "Araliaceae: Araliaceae Jussieu, Gen. Pl.: 217 (1789), nom. cons. Hydrocotylaceae Bercht. & J. Presl, (1820), nom. cons," in *Flowering plants. Eudicots*. Eds. J. W. Kadereit and V. Bittrich (Springer International Publishing), 413–446. doi: 10.1007/978-3-319-93605-5_4
- Qiu, T., Liu, Z., and Liu, B. (2020). The effects of hybridization and genome doubling in plant evolution via allopolyploidy. *Mol. Biol. Rep.* 47, 5549–5558. doi: 10.1007/s11033-020-05597-y
- Rambaut, A., Drummond, A. J., Xie, D., Baele, G., and Suchard, M. A. (2018). Posterior summarization in bayesian phylogenetics using Tracer 1.7. *Syst. Biol.* 67, 901–904. doi: 10.1093/sysbio/syy032
- Rice, A., Glick, L., Abadi, S., Einhorn, M., Kopelman, N. M., Salman-Minkov, A., et al. (2015). The chromosome counts database (CCDB) - a community resource of plant chromosome numbers. *New Phytol.* 206, 19–26. doi: 10.1111/nph.13191
- Rim, K. H. (1994). *Fossils of North Korea* (Pyongyang: Science and Technology Press).
- Schranz, E. M., Mohammadin, S., and Edger, P. P. (2012). Ancient whole genome duplications, novelty and diversification: the WGD radiation lag-time model. *Curr. Opin. Plant Biol.* 15, 147–153. doi: 10.1016/j.pbi.2012.03.011
- Schumer, M., Cui, R., Powell, D. L., Rosenthal, G. G., and Andolfatto, P. (2016). Ancient hybridization and genomic stabilization in a swordtail fish. *Mol. Ecol.* 25, 2661–2679. doi: 10.1111/mec.13602
- Shee, Z. Q., Frodin, D. G., Cámara-Leret, R., and Pokorny, L. (2020). Reconstructing the complex evolutionary history of the Papuan *Schefflera* radiation through herbariomics. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.00258
- Smith, S. A., and O'Meara, B. C. (2012). TreePL: divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics* 28, 2689–2690. doi: 10.1093/bioinformatics/bts492
- Soltis, P. S., Marchant, D. B., Van De Peer, Y., and Soltis, D. E. (2015). Polyploidy and genome evolution in plants. *Curr. Opin. Genet. Dev.* 35, 119–125. doi: 10.1016/j.gde.2015.11.003
- Soltis, P. S., and Soltis, D. E. (2009). The role of hybridization in plant speciation. *Annu. Rev. Plant Biol.* 60, 561–588. doi: 10.1146/annurev.arplant.043008.092039
- Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033
- Stebbins, G. L. (1971). *Chromosomal evolution in higher plants* (London: Arnold).
- Tank, D. C., Eastman, J. M., Pennell, M. W., Soltis, P. S., Soltis, D. E., Hinchliff, C. E., et al. (2015). Nested radiations and the pulse of angiosperm diversification: increased diversification rates often follow whole genome duplications. *New Phytol.* 207, 454–467. doi: 10.1111/nph.13491
- Tao, X., Li, M., Zhao, T., Feng, S., Zhang, H., Wang, L., et al. (2021). Neofunctionalization of a polyploidization-activated cotton long intergenic non-coding RNA *DAN1* during drought stress regulation. *Plant Physiol.* 186, 2152–2168. doi: 10.1093/plphys/kiab179
- Tribble, C. M., Freyman, W. A., Landis, M. J., Lim, J. Y., Barido-Sottani, J., Kopperud, B. T., et al. (2022). RevGadgets: An R package for visualizing Bayesian phylogenetic analyses from RevBayes. *Methods Ecol. Evol.* 13, 314–323. doi: 10.1111/2041-210X.13750

- Valcárcel, V., Fiz-Palacios, O., and Wen, J. (2014). The origin of the early differentiation of ivies (*Hedera* L.) and the radiation of the Asian Palmate group (Araliaceae). *Mol. Phylogenet. Evol.* 70, 492–503. doi: 10.1016/j.ympev.2013.10.016
- Valcárcel, V., and Wen, J. (2019). Chloroplast phylogenomic data support Eocene amphipacific early radiation for the Asian Palmate core Araliaceae. *J. Syst. Evol.* 57, 547–560. doi: 10.1111/jse.12522
- Valdés-Flórida, A., Tan, L., Maguilla, E., Simón-Porcar, V. I., Zhou, Y.-H., Arroyo, J., et al. (2023). Drivers of diversification in *Linum* (Linaceae) by means of chromosome evolution: correlations with biogeography, breeding system and habit. *Ann. Bot.* 132, 949–962. doi: 10.1093/aob/mcad139
- Vargas, P., McAllister, H. A., Morton, C., Jury, S. L., and Wilkinson, M. J. (1999). Polyploid speciation in *Hedera* (Araliaceae): phylogenetic and biogeographic insights based on chromosome counts and ITS sequences. *Pl. Syst. Evol.* 219, 165–179. doi: 10.1007/BF00985577
- Viruel, J., Conejero, M., Hidalgo, O., Pokorný, L., Powell, R. F., Forest, F., et al. (2019). A target capture-based method to estimate ploidy from herbarium specimens. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00937
- Viruel, J., Hidalgo, O., Pokorný, L., Forest, F., Gravendeel, B., Wilkin, P., et al. (2023). “A bioinformatic pipeline to estimate ploidy level from target capture sequence data obtained from herbarium specimens,” in *Plant cytogenetics and cytogenomics*. Eds. T. Heitkam and S. Garcia (Springer US, New York, NY), 115–126. doi: 10.1007/978-1-0716-3226-0_5
- Weiß, C. L., Pais, M., Cano, L. M., Kamoun, S., and Burbano, H. A. (2018). nQuire: a statistical framework for ploidy estimation using next generation sequencing. *BMC Bioinform.* 19, 122. doi: 10.1186/s12859-018-2128-z
- Wen, J. (2001). Evolution of the *Aralia-Panax* complex (Araliaceae) as inferred from nuclear ribosomal ITS sequences. *Edinb. J. Bot.* 58, 243–257. doi: 10.1017/S0960428601000610
- Wen, J., Plunkett, G. M., Mitchell, A. D., and Wagstaff, S. J. (2001). The evolution of Araliaceae: a phylogenetic analysis based on ITS sequences of nuclear ribosomal DNA. *Syst. Bot.* 26, 144–167.
- Wen, J., and Zimmer, E. A. (1996). Phylogeny and biogeography of *Panax* L. (the ginseng genus, Araliaceae): inferences from ITS sequences of nuclear ribosomal DNA. *Mol. Phylogenet. Evol.* 6, 167–177. doi: 10.1006/mpev.1996.0069
- Yi, T., Lowry, P. P., Plunkett, G. M., and Wen, J. (2004). Chromosomal evolution in Araliaceae and close relatives. *Taxon* 53, 987–1005. doi: 10.2307/4135565
- Zhang, M. H., Nie, Z. L., Fairbanks, R. A., Liu, J., Literman, R., Johnson, G., et al. (2025). Phylogenomic insights into species relationships, reticulate evolution, and biogeographic diversification of the ginseng genus *Panax* (Araliaceae), with an emphasis on the diversification in the Himalayan-Hengduan Mountains. *J. Syst. Evol.* 63, 99–114. doi: 10.1111/jse.13138
- Zhang, C., Rabiee, M., Sayyari, E., and Mirarab, S. (2018). ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinform.* 19, 153. doi: 10.1186/s12859-018-2129-y



OPEN ACCESS

EDITED BY

Mathieu Rouard,
Alliance Bioversity International and CIAT,
France

REVIEWED BY

Marcelo De Bello Cioffi,
Federal University of São Carlos, Brazil
Giovana Torres,
Universidade Federal de Lavras, Brazil
Michiel Van Bel,
Ghent University, Belgium

*CORRESPONDENCE

Carmen Benítez-Benítez
✉ cbenitez2@us.es

RECEIVED 05 April 2025

ACCEPTED 11 June 2025

PUBLISHED 22 August 2025

CITATION

Benítez-Benítez C, Mohan AV,
Sánchez-Villegas R, Gómez-Ramos I,
Valdés-Florido A, Lucek K, Slovák M,
Kolář F, Leitch IJ, Luceño M, Larridon I,
Carta A, Cuscó-Borràs J, Maguilla E,
Montero-Ramírez A, Meirmans PG,
Scott AD, Martín-Bravo S and
Escudero M (2025) Bridging micro
and macroevolution: insights from
chromosomal dynamics in plants.
Front. Plant Sci. 16:1606450.
doi: 10.3389/fpls.2025.1606450

COPYRIGHT

© 2025 Benítez-Benítez, Mohan,
Sánchez-Villegas, Gómez-Ramos,
Valdés-Florido, Lucek, Slovák, Kolář, Leitch,
Luceño, Larridon, Carta, Cuscó-Borràs,
Maguilla, Montero-Ramírez, Meirmans, Scott,
Martín-Bravo and Escudero. This is an open-
access article distributed under the terms of
the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). 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.

Bridging micro and macroevolution: insights from chromosomal dynamics in plants

Carmen Benítez-Benítez^{1*}, Ashwini V. Mohan²,
Rogelio Sánchez-Villegas³, Inés Gómez-Ramos¹,
Ana Valdés-Florido¹, Kay Lucek², Marek Slovák^{4,5}, Filip Kolář^{4,6},
Ilia J. Leitch⁷, Modesto Luceño³, Isabel Larridon⁷,
Angelino Carta^{8,9}, Joan Cuscó-Borràs¹, Enrique Maguilla³,
Alegria Montero-Ramírez¹, Patrick G. Meirmans¹⁰,
Alison Dawn Scott¹¹, Santiago Martín-Bravo³
and Marcial Escudero¹

¹Botany Area, Department of Plant Biology and Ecology, Faculty of Biology, University of Seville, Seville, Spain, ²Biodiversity Genomics Laboratory, Institute of Biology, University of Neuchâtel, Neuchâtel, Switzerland, ³Botany Area, Department of Molecular Biology and Biochemical Engineering, Universidad Pablo de Olavide, Seville, Spain, ⁴Department of Botany, Faculty of Science, Charles University, Prague, Czechia, ⁵Department of Microbial and Plant Interactions, Plant Science and Biodiversity Centre, Slovak Academy of Sciences, Institute of Botany, Bratislava, Slovakia, ⁶Czech Academy of Sciences, Institute of Botany, Průhonice, Czechia, ⁷Royal Botanic Gardens, Kew, Richmond, United Kingdom, ⁸Department of Biology, Botany Unit, University of Pisa, Pisa, Italy, ⁹Centre for Climate Change Impact (CIRSEC), University of Pisa, Pisa, Italy, ¹⁰Institute for Biodiversity and Ecosystem Dynamics (IBED), Faculty of Science, University of Amsterdam, Amsterdam, Netherlands, ¹¹Department of Chromosome Biology, Max Planck Institute for Plant Breeding Research, Cologne, Germany

Understanding the relationship between macro- and microevolutionary processes and their delimitation remains a challenge. This review focuses on the role of chromosomal rearrangements in plant population differentiation and lineage diversification resulting in speciation, helping bridge the gap between macro- and microevolution through chromosomal evolution. We focus on angiosperms, a group that comprises the majority of extant plant species diversity and exhibits the largest chromosomal and genomic variations. Here, we address the following questions: Are macroevolutionary patterns of chromosome evolution the result of accumulated microevolutionary changes, or do chromosomal dynamics drive larger shifts along the speciation continuum? At the macroevolutionary level, we investigated the association between karyotype diversity and diversification rates using evidence from comparative genomics, chromosomal evolution modelling across phylogenies, and the association with several traits across different angiosperm lineages. At the microevolutionary level, we explore if different karyotypes are linked to morphological changes and population genetic differentiation in the same lineages. Polyploidy (autopolyploidy and allopolyploidy) and dysploidy are known drivers of speciation, with karyotypic differences often leading to reproductive barriers. We found that dysploidy, involving gains and losses of single chromosomes with no significant change in overall content of the genome, appears to be relatively more frequent and persistent across macroevolutionary histories than polyploidy. Additionally, chromosomal rearrangements that do not entail change in chromosome number, such as

insertions, deletions, inversions, and duplications of chromosome fragments, as well as translocations between chromosomes, are increasingly recognized for their role in local adaptation and speciation. We argue that there is more evidence linking chromosomal rearrangements with genetic and morphological trait differentiation at microevolutionary scales than at macroevolutionary ones. Our findings highlight the importance of selection across evolutionary scales, where certain chromosomal dynamics become fixed over macroevolutionary time. Consequently, at microevolutionary scales, chromosome rearrangements are frequent and diverse, serving as key drivers of plant diversification and adaptation by providing a pool of variation from which beneficial chromosomal changes can be selected and fixed by evolutionary forces.

KEYWORDS

angiosperms, chromosome, dysploidy, evolution, polyploidy, speciation

Introduction

The role of chromosomal alterations in evolution: a brief overview

Understanding the relative importances of different chromosomal dynamics across macro- and microevolutionary processes remains a challenge in the field of evolutionary biology (Lucek et al., 2023). However, research has advanced significantly due to the unprecedented amount and accuracy of genomic datasets being available. This review focuses on chromosomal evolution in plants, especially angiosperms, from both micro- and macroevolution perspectives. It spans processes at the intraspecific level (population dynamics), cladogenetic events (lineage diversification and speciation), to broad patterns observed across large number of species (macroevolution). Specifically, by focusing on angiosperms, a group that comprises the majority of extant plant species diversity (Benton et al., 2021) and exhibiting the largest variation in chromosomal and genomic assembly across living organisms (Escudero and Wendel, 2020), we take into consideration the impact of different chromosomal rearrangements (CRs) in the process of plant population differentiation, lineage diversification and speciation. While recent technological and methodological advances allow the detection of CRs at a much broader scope, their evolutionary implications are still elusive (reviewed in Lucek et al., 2023). While Lucek et al. (2023) reviewed the implications of several types of CRs, many CRs that are especially widespread in plants such as CRs resulting from polyploidization have not been considered. Chromosomal rearrangements associated with diversification (Carta and Escudero, 2023) do not seem to be primarily driven by ecological adaptation, instead, they may arise from mutational processes or intrinsic genetic conflicts (Maheshwari and Barbash, 2011).

Particularly, factors that limit the exchange of genetic material are crucial in understanding how chromosomal dynamics can result in adaptation and speciation. Chromosomal rearrangements can reduce gene flow by affecting recombination, but these alone do not fully account for most models of chromosomal speciation (Rieseberg, 2001a). One reason for this is that the restriction of gene flow across a large chromosomal block is often insufficient to drive the speciation process on its own, unless the rearrangement involves key regions that contribute to reproductive isolation. Such key regions can be a single gene or a set of linked genes, depending on the organism, e.g. inversions of crucial chromosomal regions have been shown to drive speciation in plants (Huang and Rieseberg, 2020). However, a synergistic interaction between isolation genes (those contributing to reproductive barriers) and CRs can enhance the likelihood of models of speciation, especially when the CRs are only weakly underdominant (Rieseberg, 2001a). Chromosomal rearrangements can cause problems in chromosome pairing during meiosis, leading to reduced fertility in hybrids by producing unbalanced gametes (Rieseberg, 2001a; Stathos and Fishman, 2014). This reduction acts as a partial reproductive barrier, decreasing the probability of successful gene exchange between populations with different karyotypes and CRs. Thus, while CRs may not always be the primary driver of speciation, their direct or indirect interactions with other reproductive isolating mechanisms could contribute to playing a significant role in the process.

Cytogenetic mutations can alter the number of chromosomes, their composition, the order of the genetic material within them, or interactions between chromosomes. Chromosomal rearrangements may be classified into two main categories: those that lead to a change in chromosome number and those that result in structural changes, i.e. within a chromosome (see Box 1). Regarding the former, polyploidy involves the duplication of one or more

chromosome sets. From CRs that impact chromosome number, polyploidy has received the most attention, even though chromosome gains and losses are highly prevalent among angiosperms (Grant, 1981; Otto and Whitton, 2000; Rice et al., 2015; Stebbins, 1971). In particular, whole genome duplication (WGD) plays an important role in plant evolution at different temporal scales, with profound effects from molecular to ecological levels (Stebbins, 1971; Grant, 1981), including the restoration of fertility after hybridization (Charron et al., 2019). This phenomenon also has significant effects on gene expression, often resulting in epigenetically induced gene silencing (Osborn et al., 2003). It is particularly important, as many of the world's crops (especially those essential to global food production, such as wheat (*Triticum aestivum* L.), maize (*Zea mays* L.) and potato (*Solanum tuberosum* L.)) have a polyploid origin (Adams and Wendel, 2005; Tate et al., 2005).

Polyploids can also return back to a diploid state over time through diploidization, which comprise a diverse set of molecular processes leading to gene losses, genome downsizing and chromosomal fusions (Mandáková and Lysak, 2018). These processes can facilitate the stabilization of polyploids and enhance their ability to adapt to new environments. Such processes have potential implications for the evolutionary history of species (e.g., distribution patterns, life history and ecological adaptations). In angiosperms, diploidization may lead to species with relatively low chromosome numbers and small genome sizes (Luo et al., 2009; Mandáková et al., 2010a). While many plants are classified as diploids, the increasing amounts of sequenced genomes data are now revealing that many are diploidized polyploids. In fact, it is now recognized that all angiosperms have undergone at least one polyploidy event (often two or more) during their evolutionary history (Jiao et al., 2011; Van de Peer et al., 2017; Chen et al., 2024). Although this was not definitively proven cytogenetically or from a microevolutionary point of view, it is apparent from a macroevolutionary perspective that diploidization has played a crucial role in the evolution of vascular plants, particularly in angiosperms, as seen in model cases such as Brassicaceae (Mandáková et al., 2016; Mandáková and Lysak, 2018).

Apart from polyploidy, another kind of CR causing changes in chromosome number is dysploidy (Guerra, 2016). This mechanism produces changes in the karyotype configuration but results in no significant changes in DNA content (Heilborn, 1924; Luceño and Guerra, 1996; Escudero et al., 2014). This is especially frequent in plants with holocentric chromosomes, i.e. chromosomes that lack a single centromere but have centromeric regions spread across their chromosomes (Da Silva et al., 2017; Escudero et al., 2018; Márquez-Corro et al., 2019; Johnen et al., 2020). However, the potential impact of dysploidy on rates of diversification (henceforth encompassing both speciation and extinction processes) has not been studied in detail because it is challenging to link chromosome number evolution and species diversification at a macroevolutionary scale (but see Tribble et al., 2025 for an example of association between dysploidy and diversification rates). Interestingly, karyotype changes arising from dysploidy are thought to persist longer over time than those from polyploidy (Escudero et al., 2014; Sader et al., 2019).

Furthermore, there are CRs that do not entail changes in chromosome number, but imply structural changes. These structural variants (SVs) are known to cause chromosomal disorders, affecting mainly the phenotype and size of chromosomes (Lysak and Schubert, 2012; see more details in Box 1). A large amount of SVs has been observed at the whole-genome level between individuals belonging to related species (Lysak et al., 2006; Seymour et al., 2014). However, recurrent rearrangements with similar breakpoints, sizes and genomic context can also be shared by unrelated individuals (Mandáková and Lysak, 2008; Carvalho and Lupski, 2016).

Box 1. Chromosomal rearrangements entailing change in chromosome number vs. structural changes within a chromosome

A whole genome duplication (WGD) event entails abrupt multiplication of chromosomal sets, resulting in a polyploid individual. This process can generally be categorized into two types, although in reality, there is a continuum between them (Figure 1): (i) autopolyploidy, which involves the multiplication of chromosomes within the same species, usually involving the fusion of one or two unreduced gametes, and which may result in rapid reproductive isolation between individuals with different ploidy levels (Stebbins, 1947; Soltis and Soltis, 1989; Servick et al., 2015), and (ii) allopolyploidy, which involves the combination of chromosome sets from different parental species via hybridization between different species followed by polyploidy (Barrier et al., 1999; Rodionov, 2023; Van der Heijden et al., 2024). Autopolyploidy is very common in plants (Parisod et al., 2010; Scarrow et al., 2020) but has received less attention compared to allopolyploidy. This is because it was long believed to have little impact on plant divergence due to its formation through genome duplication only, without the involvement of hybridization. However, other advances in plant molecular biology suggest that both auto- and allopolyploidy have significant roles for evolutionary adaptation and subsequent divergence of plant species (Barker et al., 2016). Autopolyploidy, allopolyploidy, and dysploidy are the most important CRs recognized in the evolutionary history of plants (Mandáková and Lysak, 2018; Mandáková et al., 2018).

Dysploidy involves gains (ascending dysploidy via chromosome fission) or losses (descending dysploidy via chromosome fusion) of single chromosomes (Mayrose et al., 2010; Escudero et al., 2014) and appears to be relatively frequent and with longer persistence during evolutionary history than polyploidy (Escudero et al., 2014; Carta et al., 2020). While aneuploidy involves chromosome duplications and losses that result in changes in DNA content (Figure 1), such a condition is strongly selected against and tends to have little evolutionary persistence (Escudero et al., 2014). Diploidization is the process of converting a polyploid back into a diploid one. This is also a significant process in the evolutionary history of polyploids, as it can facilitate their adaptation and environmental establishment (Zhong et al., 2022; Huang et al., 2023), thereby contributing to their

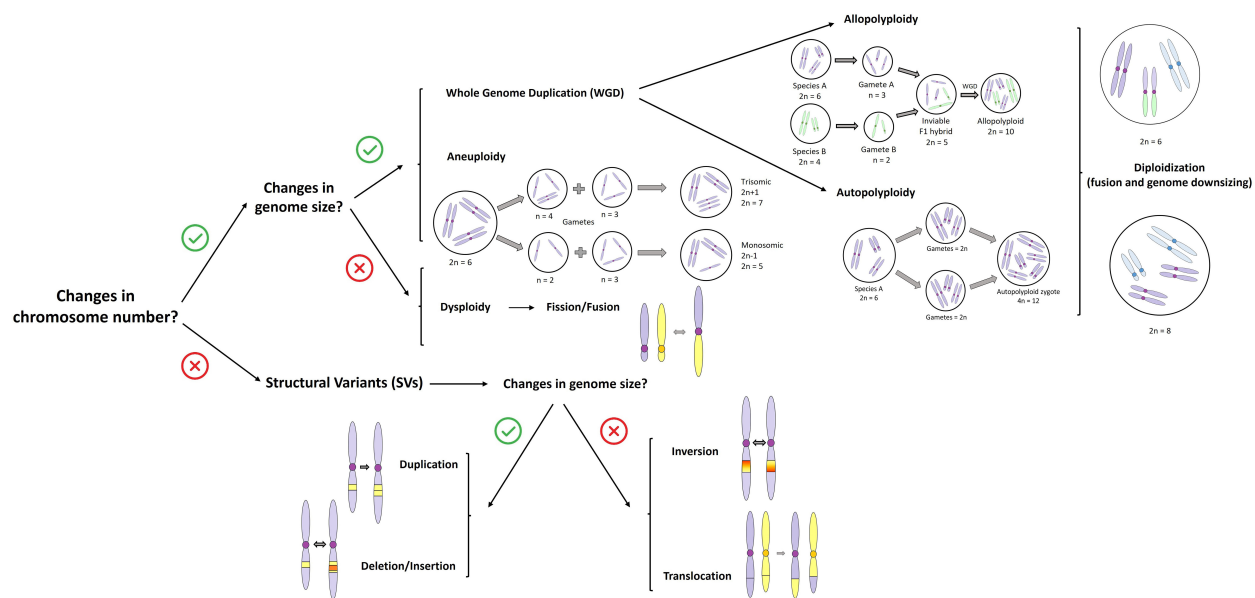


FIGURE 1

Summary of the main types of chromosomal rearrangements (CRs) that do (or not) entail changes in chromosome number (>50bp following Berdan et al., 2024). The figure shows the main pathway within allopolyploidy to produce a polyploid hybrid. For more cases of allopolyploid hybridization see in detail Hegarty and Hiscock (2008). The blue color in chromosomes indicates complex rearrangements that result in genome size reduction in the diploidization process.

evolutionary success by eliminating redundant genetic material or resolving meiotic irregularities (Figure 1).

Otherwise, structural variants can arise from various molecular mechanisms, such as DNA replication, DNA repair, and recombination processes. These processes can give rise to duplications or deletions of chromosome fragments, as well as translocations and, especially, inversions. While the two latter typically do not involve gain or loss of genetic material, they do rearrange gene order along the chromosome (Figure 1). Despite this, such rearrangements might still disrupt coding or regulatory sequences and alter chromatin structure. Consequently, they provide a mechanistic basis for how CRs may act as drivers of evolution.

Methods and models to study chromosomal rearrangements

Several methodologies are available to study chromosome number evolution, as well as CRs and their breakpoints. The earliest approach for investigating changes in chromosome number involved optical cytogenetic techniques to count chromosomes from mitotic or meiotic cells and karyotype them (Guerra, 2008). Many genomic disorders caused by SVs were initially uncovered by these early cytogenetic methods. For instance, the classical protocol for detecting inversions were based on observations of the strength of linkage of hybrids between different strains, showing an inverted order of genes with respect to a reference strain (Sturtevant, 1921; Dobzhansky and Sturtevant, 1938).

Most recently, the progress in molecular cytogenetic techniques combined with high throughput DNA sequencing, has enabled the rapid and precise detection of CRs across the genome at increasingly high levels of resolution. Such results are now increasingly being combined with insights gained from chromosome level whole genome assemblies to reveal the nature of the DNA sequences including single nucleotide polymorphisms (SNPs) associated with CRs (Le Scouarnec and Gribble, 2012). More advanced cytogenetic techniques based on *in situ* hybridization methods, such as FISH (fluorescence *in situ* hybridization) and GISH (genomic *in situ* hybridization), are widely used and effective for investigating chromosomal evolution. These methods enable us to visualize CRs as well as changes in chromosome number introduced through aneuploidy, dysploidy and polyploidy (Chester et al., 2010; Jiang, 2019).

The era of whole genome sequencing has revolutionized the ability to detect genomic rearrangements with unprecedented precision. By utilizing the order and spacing of genomic regions, these rearrangements can be quantified through synteny analyses, providing ever deeper insights into genome evolution and structure (Tang et al., 2008). Among the latest computational innovations, tools like SyRI (Goel et al., 2019) allow the detailed identification of CRs by comparing whole-genome assemblies. These results are visualized using Plotsr implemented as a Python package (Goel and Schneeberger, 2022), a tool that graphically represents synteny and rearrangement patterns across genomes, enabling the exploration of structural differences.

Complementing these genomic approaches, the combination of karyotypic and cytogenetic data (chromosome number and DNA content) is being examined within a phylogenetic framework that

accounts for non-independence using comparative methods and models of chromosome evolution (Costa et al., 2017). The use of probabilistic models such as ChromEvol (Mayrose et al., 2010; Glick and Mayrose, 2014) that allow inferences of chromosome number changes across molecular phylogenies is becoming more frequent. More recently, the development of comparative phylogenetic frameworks, i.e. ChromoSSE (Freyman and Höhna, 2018), further enables researchers to determine the degree to which CRs, including chromosomal fusions, fissions and WGD, are anagenetic or cladogenetic (in this second case there is an association between CRs and diversification). Modern comparative phylogenetic methods, and more specifically models developed to study chromosome number variation across phylogenetic trees together are increasingly enabling us to uncover the key roles played by karyotypic changes in the evolution of plants.

Chromosomal rearrangements and their role in the diversification of land plants

The study of plant evolution has long focused on understanding polyploidy and CRs, both of which play an important role in shaping genome structure. While angiosperms have been extensively studied in this context, other vascular plant groups non-commercially important as food crops, such as gymnosperms or ferns, are now gaining attention due to recent technical development and the declining cost of DNA-sequencing, which have made genome data more widely available. Advances in sequencing technologies and analytical pipelines have provided valuable insights into genome size evolution and the mechanisms driving their evolution.

Genomic studies have shown that angiosperm evolution is rich in WGD events (Paterson et al., 2012), with each subsequent polyploid event layered upon the genomic remnants of earlier rounds of polyploidization events (Wendel, 2015; Carta et al., 2020; Escudero and Wendel, 2020). Thus, chromosome doubling has played an important role in the diversification of many genera of angiosperms (Wendel et al., 2018; Barker et al., 2024). Studies on fern genomes have highlighted that recurrent WGD events without subsequent diploidization and reduction in genome size may explain several key genomic characteristics (Clark et al., 2016; Kinoshita and Barker, 2024). In contrast to angiosperms, ferns exhibit a significantly higher number of chromosomes, likely driven by a greater number of meiotic events, which contributes to increased rates of polyploid speciation (Klekowski and Baker, 1966; Carta et al., 2020; Zheng et al., 2024). This suggests that the diploidization process following polyploidy is less strong or prevalent in ferns, for which the average rate of chromosome loss is estimated to be about half the rate of angiosperms (Zheng et al., 2024). Thus, this lower estimated rate of chromosome loss among ferns is consistent with their typically higher number of chromosomes, compared with angiosperms (Klekowski and Baker, 1966; Carta et al., 2020; Zheng et al., 2024). In addition, and in contrast to angiosperms and gymnosperms, a clear positive correlation between genome size and chromosome number has also

been found in ferns with larger genomes having more chromosomes (Barker, 2013; Leitch and Leitch, 2013).

Genomic and chromosomal phylogenetic analyses have shown that angiosperms have the highest rates of ancient WGD and dysploidy among vascular plants, while ferns seem to experience multiple rounds of polyploid speciation events followed by gene silencing but not chromosome losses (Haufler, 1987; Zheng et al., 2024). Particularly, genome size reconstruction studies across angiosperms suggest that their ancestral genome size was very small compared to their plant relatives, gymnosperms and ferns (Leitch and Bennett, 1997, 2002; Soltis et al., 2013; Clark et al., 2016; Pellicer et al., 2018). This pattern is consistent across most major clades of flowering plants, including both monocots and eudicots (Leitch and Bennett, 1997; Wendel, 2015; Escudero and Wendel, 2020). Although angiosperms and gymnosperms may be subject to similar selection pressures for genome size reduction (e.g., nutrient limitations, drought stress), only angiosperms appear to have the molecular mechanisms necessary to achieve significant decreases in genome size (Michael, 2014; Ezoe and Seki, 2024).

The evidence from the ancestral node shared between angiosperms and gymnosperms suggests that genome duplication did not occur during the initial emergence of angiosperms but may have happened later (Wendel, 2015; Carta et al., 2020; Escudero and Wendel, 2020). Within angiosperms, monocots exhibit a clear pattern of repeated genome duplication throughout the diversification of various genera, while the common ancestor of eudicots is considered to have undergone a whole genome triplication (Jiao et al., 2012). Polyploidy has also recurred in many lineages that have diversified more recently within this group (e.g., *Brassica*, *Solanum*; Wendel, 2015; and at the base of many families including Asteraceae, Fabaceae, Brassicaceae, Ranunculaceae; Qiao et al., 2019). These cycles of polyploidy tend to repeat over timescales ranging from thousands to millions of years (Wendel, 2015; Escudero and Wendel, 2020). As outlined above, polyploidy is, at least in part, reversible. Over time, it is often followed by extensive CRs, reductions in chromosome number, and large-scale losses of both repetitive sequences and duplicated genes, ultimately leading to genome downsizing (Leitch and Leitch, 2008). This diploidization phenomenon involves diverse processes that result in descendants behaving cytogenetically as typical diploids, while still retaining vestigial evidence of past polyploidy events within their genomes (Wendel, 2015; Carta et al., 2020; Escudero and Wendel, 2020). In addition, during diploidization, one of the two genomes is preferentially retained and exhibits higher gene expression levels, as widely observed across angiosperm lineages (Cheng et al., 2012; Wendel, 2015).

Despite the advances reported above, our understanding of CRs in the process of plant population differentiation and lineage diversification in ferns and gymnosperms remains more limited, not allowing a detailed overview as for angiosperms. In light of this, our review provides new insights into the mechanisms underlying the transition from macro- to microevolutionary processes focusing on angiosperms, contributing to a deeper understanding of evolutionary dynamics across both scales. To this end, we addressed six questions (Q1-Q6) allowing us to present the topic

in a systematic way. Specifically, at the macroevolutionary level, we investigate the association between karyotype diversity and rates of diversification, discussing comparative genomics and chromosomal evolution modeling across phylogenies (Q1-Q2). We also explore how CRs correlate with several traits across different angiosperm lineages (Q3). At the microevolutionary level, we examine how different karyotypes (including differences in ploidy level, chromosome number and structure) are linked to geographic, environmental, and phenotypic changes (including anatomical and physiological shifts; Q4-Q5). Additionally, we investigated how population genetic differentiation through allo- and autopolyploidy may promote the formation of new genetic combinations (Q6). Understanding the evolutionary processes leading to intraspecific chromosomal diversification is crucial, as such divergence can lead not only to reproductive incompatibilities between species but also to diversification and speciation processes within a given species. Ultimately, the goal of this review is to address whether patterns of chromosomal macroevolution reported in the literature are the result of singular chromosomal changes accumulating within microevolutionary timeframes, or alternatively, whether some chromosomal changes are able to cause a larger shift in the speciation continuum compared to others.

Q1: How have patterns of CRs shaped the recent evolutionary history of angiosperms?

Understanding how evolutionary trends have changed through time is challenging, as it involves the interaction of several factors. The most important among these challenges are the ability to distinguish true evolutionary patterns (actual evolutionary events) from those influenced by sampling bias towards the present (i.e., only extant representatives of diverse clades) and the unequal extinction rates across different groups of plants.

The macroevolutionary study of chromosomal changes through evolutionary history utilizing comparative phylogenetic frameworks is no different. These studies focus on karyotypic changes that lead to changes in chromosome number, through poly- and dysploidy, as chromosome counts are widely available for many plants (see Figure 1 in Carta and Escudero, 2023; The Index to Plant Chromosome Numbers: Goldblatt and Johnson, 2011; Goldblatt and Lowry, 2011; The Chromosome Counts Database: Rice et al., 2015; Rice and Mayrose, 2023). In contrast, genomic or detailed karyotypic data needed for detecting other types of CRs, such as inversions or translocations, have only started to be generated in the last two to three decades.

The two major processes underpinning chromosome losses and gains (i.e. aneuploidy and dysploidy) have different impacts on lineage diversification. While aneuploidy leads to unstable lineages that persist for shorter periods over evolutionary time (under 1–5 million years; Yona et al., 2012), dysploidy has a less drastic impact, resulting in more stable lineages throughout time (Otto, 2007). Most, if not all, of the chromosome gains and losses that have been inferred in phylogenies of angiosperms are caused by dysploidy (Escudero et al., 2014; Carta et al.,

2020). For instance, in the genus *Carex* L. (Cyperaceae), one of the most diverse genera of vascular plants in terms of species richness and chromosome number diversity, there is a tendency for dysploidy to be the main underlying mechanism responsible for chromosome evolution (Hipp et al., 2009).

Multiple studies have found that rates of polyploidy have increased in more recent evolutionary time in angiosperms (Carta et al., 2020; Escudero et al., 2014; see in Figure 2 data from Zhan et al., 2021). In the case of dysploidy, the rates are more constant and slightly accelerate towards the present (Escudero et al., 2014; see in Figure 2). Data suggest that there are no significant differences between descending and ascending dysploidy rates (Figure 3). Here, we show the rate of dysploidy and polyploidy against time using the data from Zhan et al. (2021), who inferred rates of polyploidy and dysploidy using ChromEvol (Glick and Mayrose, 2014; Mayrose et al., 2010) from phylogenetic (Qian and Jin, 2016; Zanne et al., 2014) and chromosome count data (Rice et al., 2015) for 30,000 taxa representing 46 orders and 147 families of angiosperms. Their analyses show that both polyploidy and dysploidy increase exponentially with time, but the increase in rate is much greater for polyploidy (Figure 2). The evolution of CRs through time is yet to be explored for gymnosperms and ferns, which could be especially interesting in the latter as they have an even greater chromosome number variation than angiosperms (Leitch and Leitch, 2012) and similar rates of paleopolyploidy (Li et al., 2024).

This acceleration towards the present in CRs is probably partially caused by a detection bias, as reconstructing genomes further back in time requires increasingly extensive genomic data to remain accurate. Chromosome rearrangement inference from chromosome counts is less reliable with increasing phylogenetic depth as homoplasies become more frequent, especially within groups with high karyological instability (Mayrose and Lysak, 2021; Weiss-Schneeweiss and Schneeweiss, 2013). Methods based on either genetic (gene duplicate analysis and paralogue divergence) or genomic data (synteny analysis), are more reliable at higher depths than those based on chromosome counts, even allowing detection of other CRs (Wang et al., 2021). As more genomic data become available for a wider range of taxa, studies of CR events in angiosperm history will become more accurate. One of the most recent studies addressing this issue was performed by McKibben et al. (2024). The authors studied synonymous divergences of paralogs (Ks) and orthologs, along with syntenic analysis on 462 genomes distributed throughout the angiosperm phylogenetic tree. Their analyses inferred multiple ancient WGD events and concluded that most extant species have experienced, on average, three WGD events, while also detected an increase in WGD towards the tips of the phylogenetic tree.

While a bias towards more recent events may contribute to the apparent acceleration in rates of CR, it does not account for the higher pattern of acceleration observed in polyploids compared to dysploids (Figure 2). This bias would be expected to influence both types of CRs, and potentially even more strongly for polyploidy because it is generally easier to detect than dysploidy. Thus, the discrepancy in acceleration rates suggests additional factors are at play. One hypothesis shows that the detection of polyploidization in

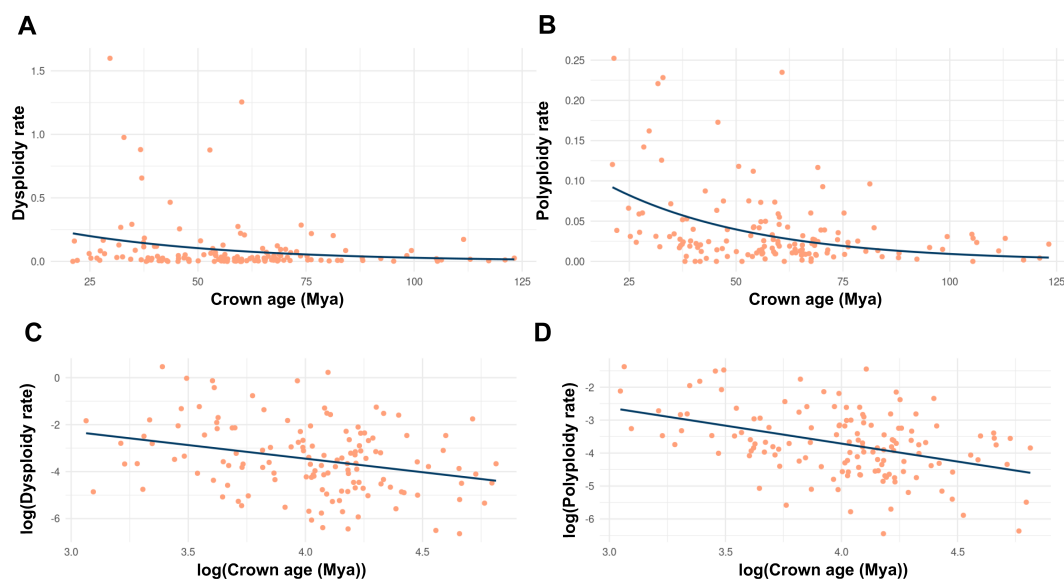


FIGURE 2

Timescale-dependent patterns of chromosome rearrangements, based on phylogenetic and chromosome count data from [Zhan et al. \(2021\)](#), who inferred the relationships between polyploidy, dysploidy with lineage diversification by combining chromosome number data with a time-calibrated mega-phylogeny, assembling clade-level datasets for 30,000 taxa representing 46 orders and 147 families of angiosperms. Each dot represents a clade for which polyploidy and dysploidy rates were estimated using ChromEvol. (A, C) for dysploidy, (B, D) for polyploidy, (A, B) for non-transformed data and (C, D) for log-transformed data.

the past is more difficult because they stay “polyploid” for a shorter time due to the activity of diploidization processes leading to the rearrangement of multiple sets of parental chromosomes in their polyploid ancestors ([Zhang et al., 2019](#)). Considering the amount of WGD throughout the evolution of angiosperms ([McKibben et al., 2024](#)), extant species would be expected to have a much higher number of chromosomes and larger genomes than what is currently observed. This discrepancy underscores the impact of diploidization in shaping angiosperm genomes ([Wang et al., 2021](#)). Most species undergo a process of diploidization after WGD, which impacts at a genomic, epigenomic and proteomic level ([Dodsworth et al., 2016](#); [Wendel et al., 2016](#)). This process often results in a reduction in genome size and a decrease in chromosome number through dysploidy, sometimes reaching a number of chromosomes equal, or even lower, than the original diploid number ([Li et al., 2021](#); [Mandáková and Lysak, 2018](#)). Different factors have been proposed to explain the so-called “large genome constraint hypothesis”, which is primarily linked to genome size. These include limitations in phosphorus (P) and nitrogen (N) availability, constraints related to life cycle duration, narrow ecological tolerances, or small population sizes ([Knight et al., 2005](#); [Carta and Peruzzi, 2016](#); [Guignard et al., 2016](#); [Wang et al., 2021](#)). Also, there is strong evidence that supports a lower diversification rate for neopolyploids in both ferns and angiosperms ([Mayrose et al., 2010](#)). Diploidization mechanisms remain to be studied in gymnosperms and ferns ([Li et al., 2021](#)) in order to understand how it impacts CR rates. A study in the hexaploid coast redwood (*Sequoia sempervirens*), one of the few polyploid species in conifers (see other examples in [Farhat et al., 2019; 2022](#)), showed very low diploidization rates ([Scott et al., 2016](#)). This finding may help to

explain the evolutionary success of polyploid lineages and a limited chromosome number variation in gymnosperms. Ferns propose a more interesting, yet unexplored, study system for the impact of diploidization on CR rates. Unlike angiosperms, ferns undergo diploidization primarily through gene deletion and pseudogenization rather than gene loss ([Haufler, 1987](#); [Zheng et al., 2024](#)), while still maintaining a high rate of diploidization ([Haufler and Soltis, 1986](#); [Haufler, 1987](#); [Wolf et al., 1987](#)).

Descending dysploidy is one of the most frequent routes of diploidization contributing to the reversion towards functional diploidy of polyploid angiosperms ([Mandáková and Lysak, 2018](#)). Therefore, descending dysploidy is expected to accelerate after WGD, while this acceleration is not present in species undergoing increasing dysploidy. The latter is often driven by chromosomal fragmentation but does not directly impact genome size significantly, unless accompanied by changes in DNA content. To differentiate the decreasing dysploidy associated with diploidization, we compared decreasing and increasing dysploidy rates to test if they were significantly different ([Figure 3](#)). However, we did not observe a difference, suggesting that the decreasing dysploidy associated with subsequent diploidization is not as easily detected as ancient WGD ([Figure 2](#)). This suggests that ancient WGD and the associated chromosome number reductions (decreasing dysploidy) during diploidization are difficult to detect, as polyploids may undergo rapid diploidization, which may limit their diversification potential in the polyploid state. Further support for this comes from models predicting 3–4 rounds of WGD in angiosperms, with high rates of diploidization alongside lower polyploid diversification rates compared to diploids ([Barker et al., 2016; 2024](#)). This instability is not observed following dysploidy, as

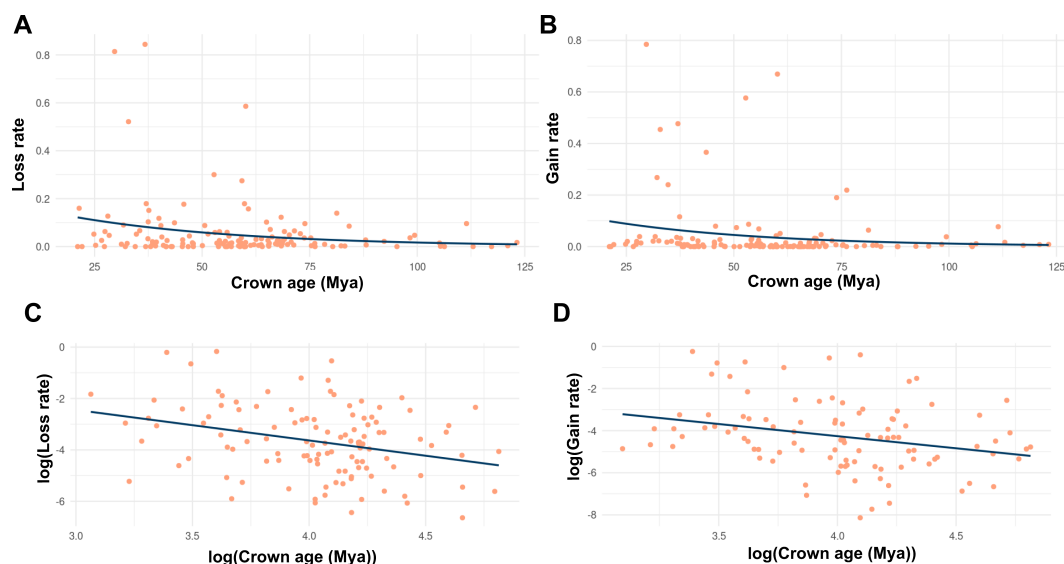


FIGURE 3

Timescale-dependent patterns of chromosome rearrangements based on phylogenetic and chromosome count data from Zhan et al. (2021), who inferred the relationships between descending (chromosome loss) or ascending (chromosome gain) dysploidy with lineage diversification by combining chromosome number data with a time-calibrated mega-phylogeny, assembling clade-level datasets for 30,000 taxa representing 46 orders and 147 families of angiosperms. Each dot represents a clade for which descending or ascending dysploidy rates were estimated using ChromEvol. (A, C) for descending dysploidy (chromosome loss), (B, D) for ascending dysploidy (chromosome gain), (A, B) for non-transformed data and (C, D) for log-transformed data.

its rate appears to remain more constant throughout angiosperm evolutionary time (Escudero et al., 2014; Figure 2). Such findings support the hypothesis that dysploidy, unlike polyploidy, is not as disadvantageous in generating long-term persisting lineages and does not entail significant changes in DNA content (Escudero et al., 2014). Overall, this highlights the importance of further exploring the impact of dysploidy in evolution, which has traditionally received less attention compared to polyploidy, despite its potential influence on chromosomal evolution and species diversification at both macro- and microevolutionary scales. Even less is known about the role of dysploidy in other plant lineages. Descending dysploidy associated with diploidization has been demonstrated to be lower in ferns than in angiosperms even for similar rates of paleopolyploidy (Haufler, 1987; Li et al., 2024; Zheng et al., 2024), leading to high chromosome numbers (Leitch and Leitch, 2012).

In conclusion, the increase in CR rates in more recent angiosperms is influenced by both biological processes and methodological biases. While sampling biases favor the detection of recent chromosomal changes, true evolutionary mechanisms, particularly polyploidy, also contribute to this tendency. Polyploidy rates increase more rapidly than dysploidy, where ancient WGD followed by diploidization are common. Dysploidy, while more constant over time, shows a slight acceleration toward the present. Models like ChromoSSE have been developed to reconstruct chromosomal evolution by integrating chromosomal changes, speciation, and extinction rates. These models offer valuable insights into how CRs, such as dysploidy and polyploidy, influence diversification across lineages. However, the detection of these patterns is greatly affected by the availability of data (phylogenetic and chromosome number data). Incorporating more comprehensive

datasets in future studies, not only within angiosperms but expanding to the rest of plants, may provide a clearer understanding of the great differences between the main plant groups. Further studies should focus on chromosome evolution in ferns, in order to understand how the differences in diploidization mechanisms have shaped chromosome evolution in ferns and angiosperms, both groups with high polyploidy rates.

Q2: Do chromosome changes driving diversification occur prior to cladogenesis?

Chromosomal rearrangements have the potential to drive speciation by reducing gene flow between divergent populations (Berdan et al., 2024; Lucek et al., 2023). Specifically, chromosomal changes may occur at the speciation event, either by directly initiating a cladogenetic process or by reinforcing speciation that was already initiated or completed by other geographical or ecological drivers or other genomic mutations independent of CR. Allopolyploid speciation is an example of chromosomal changes and cladogenesis coinciding (see section related to allopolyploidy). Currently, two primary models explain how CRs may contribute to reproductive isolation and speciation (Rieseberg, 2001a): hybrid dysfunction and the suppression of recombination.

The classical models are based on hybrid dysfunction and hypothesize that hybrids resulting from crosses between two different chromosomal races have reduced fitness (Coyne and Orr, 2004). This reduced fitness can lead to strong selection against hybrids, primarily because newly arising CRs are often underdominant. While strongly

underdominant rearrangements are unlikely to reach fixation, those with weaker underdominance may become fixed but usually create only shallow barriers to reproductive isolation, making them unlikely to drive speciation (Rieseberg, 2001a; Faria and Navarro, 2010). This model of chromosomal speciation suggests that chromosomal changes occur at, or just prior to a cladogenetic event, as such changes are expected to establish significant barriers to gene flow that may result in speciation. In plants, ploidy changes are generally expected to lead to hybrid dysfunction (Escudero et al., 2016b; 2024) because for example, a cross between a diploid and a tetraploid typically produces an inviable or sterile triploid (Köhler et al., 2010). However, there are also counter examples where gene flow occurs between different ploidy levels through at least partially fertile transitional cytotypes. This gene flow can facilitate heteroploid gene transfer, contribute to adaptation via adaptive introgression and even lead to the *de novo* formation of a new polyploid (Čertner et al., 2017; Kolář et al., 2017; Peskoller et al., 2021, reviewed by Brown et al., 2024; Bartolić et al., 2024). This might lead to a slowing down of the genomic separation of given cytotypes and hampering diversification and speciation as well. Besides, there are many cases of intraspecific variation in ploidy levels within the same species (e.g., *Hieracium* subgenus *Pilosella*, Suda et al., 2007; *Elettaria cardamomum*, Anjali et al., 2016; *Phragmites australis*, Wang et al., 2024), suggesting that auto-polyploidization does not necessarily drive rapid speciation. Regarding intraspecific genetic variation, ploidy level in populations may represent (i) different genetic groups (Balao et al., 2010), (ii) only a partial correspondence with the genetic clustering (Vanrell et al., 2024), or (iii) a complete mismatch indicating a lack of genetic differentiation among different ploidy levels within a species (Chumová et al., 2024; Kauai et al., 2024). Given the limited number of ancient polyploidization events inferred for plants, it seems that most of these frequent intraspecific polyploidy variations do not persist. Otherwise, we would observe many more ancient polyploidization events in extant species. This conclusion is congruent with recent polyploids showing often lower diversification rates than their diploid progenitors (Mayrose et al., 2010). This aligns with reports suggesting that polyploidization from a macroevolutionary viewpoint is an evolutionary “dead end” since polyploids exhibit higher rates of extinction than their diploid relatives (Arrigo and Barker, 2012; Mayrose et al., 2014; Shu et al., 2022). In contrast, there are also examples of autopolyploid cytotypes that have undergone speciation processes being ultimately recognized as independent species (Fernández et al., 2022). Other alternative views suggest that there is no significant association between shifts in diversification rates and ancient polyploidization (Landis et al., 2018).

The second line of theory emphasizes the role of CRs for recombination, whereby CRs become fixed through natural selection as they suppress recombination in locally advantageous groups of genes, known as supergenes (Ayala and Coluzzi, 2005). By acting as barrier loci, they create genomic regions that facilitate the maintenance of beneficial combinations of alleles within populations, even in the presence of gene flow and can ultimately promote reproductive isolation and speciation. The suppression-recombination model of chromosomal speciation predicts that chromosomal changes (inversions, dysploidy, deletions/insertions, duplications, and reciprocal translocations) that may affect recombination rates will

occur before a cladogenetic event, where over time, locally adapted alleles accumulate, eventually resulting in a cladogenetic event. Under this model, we predict genome variation among populations of the same species, a variation that is indeed observed (Rieseberg, 2001a). Furthermore, the speciation process can occur concurrently with gene flow between different karyotypes, indicating that speciation is not an instantaneous process but often a rather gradual one influenced by ongoing evolutionary forces (Rieseberg, 2001a; Ravinet et al., 2017).

Both theoretical frameworks suggest that variation in CRs among populations facilitates reproductive isolation, thereby enhancing the potential for speciation (Lucek et al., 2023). In addition, these two models of chromosomal speciation are not mutually exclusive: some CRs may on one hand reduce gene flow between different karyotypes, resulting in partial hybrid dysfunction, and on the other hand suppress recombination. Together, these factors may eventually lead to cladogenesis. In this context, CRs that act as barriers to gene flow are predicted to occur either before the speciation process is complete or afterward, preventing interspecific gene flow, e.g. during secondary contact (Faria et al., 2011; Berdan et al., 2024). Finally, there are some CRs that are not or less important for speciation and are likely to become extinct over time (Lucek et al., 2023) or could be retained through a process similar to incomplete lineage sorting, where the origin of a CR predates the speciation event.

In a phylogenetic framework, the hybrid dysfunction type model is consistent with cladogenesis, whereas the suppression of recombination type model is more consistent with anagenesis (Lucek et al., 2022). However, only few phylogenetic approaches exist to model chromosomal evolution (Mayrose and Lysak, 2021). The joint modeling of chromosome evolution, speciation, and extinction is implemented in ChromoSSE (Freyman and Höhna, 2018). In this model, chromosomal changes may occur anagenetically (along the branches of the phylogeny) or cladogenetically (at the time of speciation). From the few examples for plants that implement this model (Freyman and Höhna, 2018; Valdés-Florido et al., 2023, 2024b; Tribble et al., 2025) a clear pattern emerges: the vast majority of chromosomal changes happen anagenetically, while only a small percentage occur cladogenetically. In summary, although chromosomal changes can occur around the speciation event, most changes occur between cladogenetic events. Additionally, it is thought that only a small percentage of chromosomal changes are able to survive through the filter of speciation, with most eventually being lost or becoming extinct.

Ultimately, chromosomal changes can indeed occur at the time of speciation, either initiating or reinforcing the process of cladogenesis (see more details in Box 2). In the hybrid dysfunction model, CRs such as polyploidy create reproductive barriers by reducing hybrid fitness. In contrast, the recombination suppression model proposes that chromosomal changes, like inversions or translocations, reduce recombination allowing locally adapted alleles to accumulate potentially driving speciation. In most cases, chromosomal changes do not persist long-term, as only a small percentage survive the speciation process, with most becoming extinct. Alternatively, CRs can also exist without affecting phenotypes or physiological functions,

remaining neutral and simply persisting. Therefore, while CRs can contribute to speciation, the majority of these changes occur anagenetically, with relatively few occurring directly at the moment of cladogenesis.

Box 2. From chromosomal assortative mating at population level to chromosomal cladogenesis across the phylogeny of holocentric true sedges

The theory of chromosomal speciation primarily assumes that chromosomes are monocentric, meaning they contain a single centromeric region where all kinetochores are concentrated for spindle attachment during mitosis and meiosis (Escudero et al., 2016a). However, approximately 15–20% of extant eukaryotes, spanning 19 different animal and plant lineages, possess holocentric chromosomes. These chromosomes are characterized by holocentromeres: small, centromere-like regions dispersed along their entire chromosome length, rather than a single localized centromere (Escudero et al., 2016a). In holocentric species, segmental rearrangements may not lead to the same segregation issues during cell division as seen in monocentric species (Lucek et al., 2022). For example, in monocentric species, chromosomal fission can result in segments lacking a centromere, making them prone to loss during meiosis, while fusion events may create chromosomes with two centromeres, leading to segregation errors (Figure 4).

In monocentric species, when fission occurs, the chromosome fragment lacking a centromere is typically lost. In contrast, in holocentric species, fragmented chromosome sections can retain kinetochore function due to the distribution of centromere-like structures along the entire chromosome, allowing these fragments to be preserved (Lucek et al., 2022).

True sedges (*Carex*) belong to the holocentric sedge family Cyperaceae, one of the most diverse plant groups, comprising approximately 5,700 species (Larridon et al., 2021). The remarkable diversification of *Carex* is closely linked to its extensive variation in chromosome numbers ($2n = 10$ –132), which has primarily evolved through chromosomal fusions and fissions, rather than polyploidization (Roalson, 2008). This unique evolutionary trajectory has positioned *Carex* as a model system for studying the dynamics of holocentric chromosomes and the mechanisms of chromosomal speciation, providing insights at both micro- and macroevolutionary levels. In *Carex* species, striking chromosome-number polymorphism is frequently observed, even within populations or individual plants (Whitkus, 1988; Luceño and Castroviejo, 1991; Escudero et al., 2013a, b; 2024). For example, *Carex scoparia* Schkuhr ex Willd. exhibits a range of $2n = 56$ to $2n = 70$ (Escudero et al., 2013b), *C. laevigata* Sm. ranges from $2n = 69$ to $2n = 84$ (Luceño and Castroviejo, 1991; Escudero et al., 2013a; Márquez-Corro et al., 2023), and *C. helodes* Link varies from $2n = 68$ to $2n = 75$ (Escudero et al., 2024). Experimental evidence from artificial crosses between cytotypes indicates that reproductive isolation intensifies as CRs accumulate, resulting in increasingly severe hybrid seed germination dysfunction (Figure 5; Escudero et al., 2016b; Whitkus, 1988). Nonetheless, individuals with differing chromosome numbers can often reproduce and exchange alleles—directly, if only minor CRs are involved, or indirectly (via individuals with intermediate karyotypes) in cases of major chromosomal differences—maintaining gene flow across chromosomal boundaries and species coherence (Escudero et al., 2013b). This suggests that while small chromosomal differences are insufficient to establish reproductive barriers, the accumulation of CRs can drive reproductive isolation over time (Hipp et al., 2009). The inferred isolation driven by gene flow and the accumulation of CRs in true sedges may significantly shape the genetic structure of populations. Hipp et al. (2010) found that both geographic distance and the number of karyotype rearrangements between populations influence the rate of

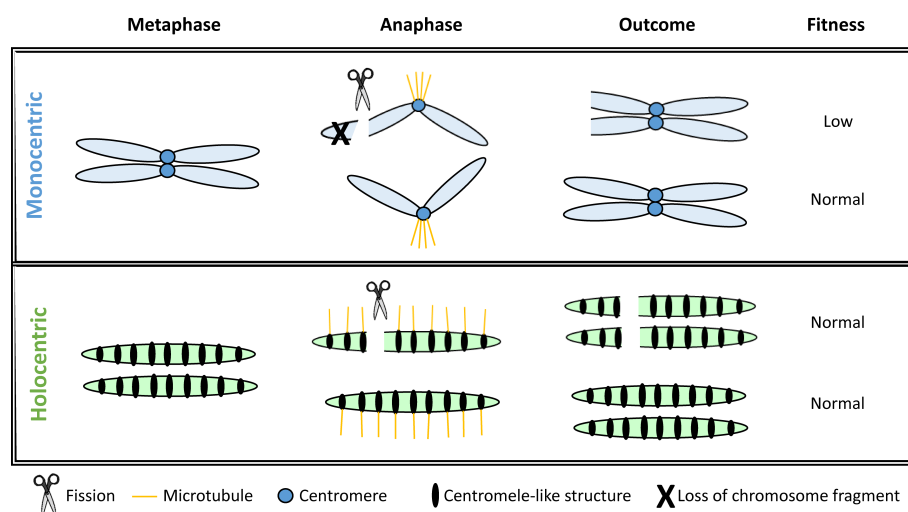


FIGURE 4

Modified from Lucek et al. (2022). Comparison of the outcomes of chromosome fission events during cell division for mono- and holocentric species.

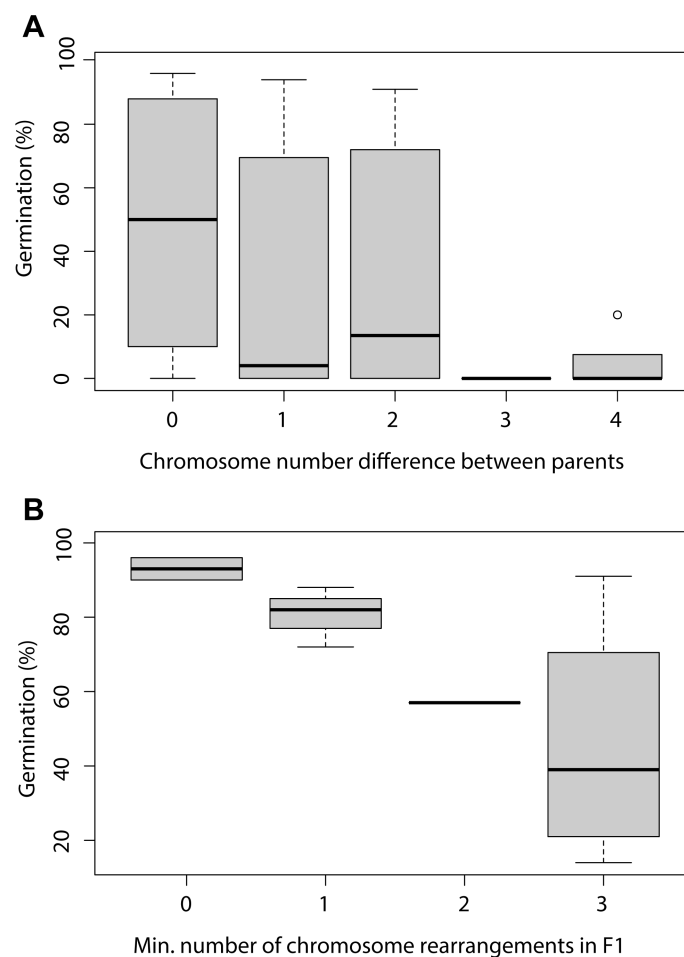


FIGURE 5

Modified from Escudero et al. (2016b). Boxplots showing seed germination percentages for (A) the offspring of artificial crosses between parent pairs with zero, one, two, three, and four chromosomal differences (N=33); and (B) F1 hybrids with zero, one, two, and three chromosomal irregularities (with the minimum number of irregularities considered) (N=11).

gene flow in *C. scoparia*. A similar conclusion was reached at a finer evolutionary scale by Escudero et al. (2013a). Interestingly, this pattern also seems to apply at higher evolutionary levels, where the time to species coalescence is directly proportional to chromosomal variation within species in *Carex* sect. *Spirostachyae* (Drejer) L. H. Bailey (Escudero et al., 2010).

The impact of dysploidy on *Carex* diversification has been previously explored using QuaSSE, which models chromosome number as a continuous trait (Márquez-Corro et al., 2021). Tribble et al. (2025) have been the first to jointly model chromosome number changes and diversification using a specialized model for chromosome evolution—the ChromoHiSSE model. This is a version of ChromoSSE (Freyman and Höhna, 2018) that accounts for hidden states, allowing the rates of chromosome number changes and their association with cladogenesis to vary across the phylogeny. Their results reveal an association between higher speciation rates and dysploidy in certain parts of the true sedge phylogeny, despite heterogeneity in the diversification process. In some clades, gains and losses in chromosome number drive diversification (hidden state i), while in other regions of the tree, these changes have the opposite effect (hidden

state ii, see Figure 6). Furthermore, although dysploidy does not lead to higher speciation rates across the entire phylogeny, it strongly drives speciation in specific clades. Moreover, as indicated before, the vast majority of the CRs happen anagenetically along the branches of the phylogeny and only a small percentage of them are cladogenetic.

Tribble et al. (2025) proposed that the discrepancies in the impact of dysploidy on cladogenesis (with most changes happening anagenetically) may be linked to the unique nature of holocentric chromosomes. In this context, a single dysploidy event may not be sufficient on its own to induce reproductive isolation (Whitkus, 1988; Hipp et al., 2010; Escudero et al., 2016b; Lucek et al., 2022). However, the accumulation of CRs within a lineage over time could eventually establish a reproductive barrier, thereby driving speciation (Whitkus, 1988; Escudero et al., 2016b). This idea supports the hybrid dysfunction/recombination suppression model of chromosomal speciation, a central hypothesis discussed by Lucek et al. (2022). One possibility is that the accumulation of chromosomal changes eventually leads to reproductive isolation, with a “last straw” dysploidy event acting as the final trigger for speciation (the last-straw hypothesis in Tribble et al., 2025).

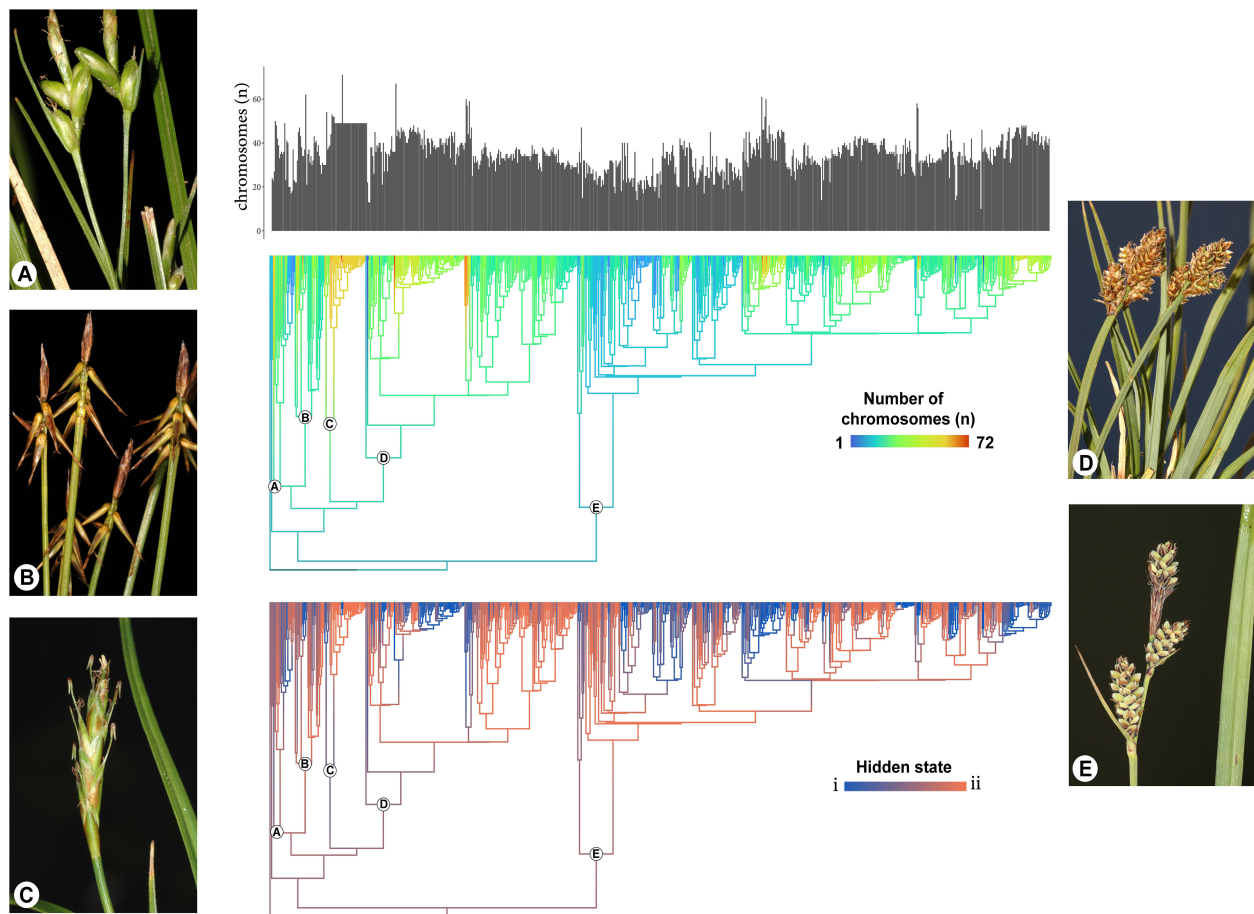


FIGURE 6

Modified from Tribble et al. (2025). Reconstruction of chromosome numbers and hidden states on the *Carex* phylogeny. At the top, the distribution of haploid chromosome numbers for all extant taxa included in the analysis. In the middle, the reconstructed evolution of chromosome numbers along phylogenetic branches. Warmer colors represent a higher number of chromosomes. At the bottom, the reconstructed evolution of the hidden states along phylogenetic branches. Blue color indicates strong statistical support for state i (cladogenesis driven by chromosomal changes), red color indicates strong support for state ii (cladogenesis is independent of chromosomal changes), and intermediary colors represent uncertainty in the estimates. Subgenera in phylogeny are labeled (A = *Psyllophorae*, B = *Euthyceras*, C = *Uncinia*, D = *Vignea*, and E = *Carex*). Photos display examples of species for each of the subgenera (A = *C. oedipostyla* Duval-Jouvé, B = *C. microglochin* Wahlenb., C = *C. meridensis* (Steyerm.) J.R. Starr, D = *C. lucennoiberica* Maguilla and M. Escudero, and E = *C. adelostoma* V.I.Krecz.). Photo credits to M. Luceño.

Another possibility is that rearrangements in certain genomic regions are more stable than others, and the specific locations where fission or fusion occurs within the genome determine the evolutionary impact of dysploidy (Tribble et al., 2025).

Q3: Do bursts of phenotype evolution, chromosome evolution, and speciation occur at the same time?

Changes in physical traits, in the chromosomes, and speciation can occur simultaneously in plants, but their interplay is complex and influenced by multiple factors. Intraspecific phenotypic and allelic changes are driven by natural selection, genetic drift or adaptation to local biotic and abiotic factors. Duplications, inversions, or translocations can result in gene expression changes that lead to significant phenotypic effects (Wray et al., 2003). As a consequence, these rearrangements can also lead to new linkage relationships or the

formation of new genes. It is important to highlight that few breakpoints have been characterized for inversions with clear phenotypic effects (Hoffmann and Rieseberg, 2008; Guan et al., 2021; Chen et al., 2024). Previous studies have demonstrated that chromosomal inversions have putatively evolved as a response to environmental conditions because they were associated with morphological traits and showed increased fitness in adapted environments (Lowry and Willis, 2010; Lee et al., 2016). For instance, strong karyotype differences between closely related Mediterranean orchid species that also share pollinators have shown that CRs play an important role in reducing hybrid fitness and maintaining reproductive isolation (Cozzolino and Scopece, 2008). Otherwise, it is widely known that chromosomal deletions can have significant phenotypic consequences, since dominant alleles can be deleted, exposing recessive alleles in heterozygosity (Huettel et al., 2008). The phenotypic effects derived from other SVs have been less studied in plants. However, comparative genomic mapping has begun to facilitate their identification and annotation techniques have allowed

the identification of possible SVs (including gene presence/absence and copy number variations) responsible for phenotypic traits (Huang and Rieseberg, 2020; Zhang et al., 2020). These kinds of studies have mainly focused on crops, identifying how SV impacts in genes with agronomic value (Yuan et al., 2021). Since crop gene pools are often derived from multiple species, sequencing and assembly efforts are put into all the species within the genus of interest. This has led to the development of super-pangenomes, which enable the detection of conserved and diverged genomic regions, as well as their frequencies within populations (Zhao et al., 2020). The relevance of pangenomics has grown significantly with the availability of high-quality genomes assemblies from multiple cultivars, especially in agriculturally important crops (Zhao et al., 2018).

Otherwise, WGD are common in plants, with extensive impacts on gene expression, cellular function, and organism phenotype. Polyploids can display differences in floral traits (Balao et al., 2011; McCarthy et al., 2015), chemical scents (Vereecken et al., 2010; Jersáková et al., 2010), and flowering phenology (Schranz and Osborn, 2000; Pegoraro et al., 2019). Such phenotypic changes evolved immediately after polyploidization, and it may have served to establish and stabilize novel cytotypes (Oswald and Nuismer, 2011; Clo and Kolář, 2021). Thus, polyploidy often results in reproductive isolation, leading to rapid speciation because new polyploid individuals may not be able to reproduce with their diploid progenitors. These new species frequently exhibit novel phenotypic traits as a consequence of changes in gene expression caused by the increased chromosome number (Chen, 2007; Balao et al., 2011; Basit and Lim, 2024). Phenotypic and morphological changes known to be induced by polyploidy are those related to variation in flower number and flowering time (Schranz and Osborn, 2000), plant structure, or alterations in plant physiology under stress tolerance (Cohen et al., 2013; Van de Peer et al., 2021; Turcotte et al., 2024). Polyploidy may for instance contribute to higher tolerance to nutrient-poor soils and resistance to stressful environments such as drought, cold or pathogens (Levin, 2002; Sader et al., 2019). In fact, a common phenomenon in polyploid species is the “gigas effect”, which results in increased cell sizes and overall plant features in comparison with their diploid parents (Stebbins, 1971; Soltis et al., 2014). In addition, CRs can generate genetic diversity through evolutionary changes, since bursts of lineage splitting in plants often result in adaptive radiation (Parent et al., 2020). In these cases, the accumulation of genomic changes leads to rapid phenotypic evolution promoted by genetic variation and accelerated evolution, particularly under changing environmental conditions. These phenotypic changes are not gradual but instead occur in bursts, often linked to speciation, which frequently occurs simultaneously. Consequently, long periods of evolutionary stasis could be interrupted by short and rapid bursts of evolutionary change linked to chromosomal events (Stebbins, 1971; Levin, 2002; Lysak et al., 2006). However, although most of intraspecific polyploidy variations do not persist over time (see above section), some studies have revealed that species with recent polyploid origins may undergo rapid speciation and significant phenotypic divergence (Fehrer et al., 2022). Floral evolution in the genus *Calochortus* Pursh (Liliaceae) represents a

case of radiation, where selection for adaptation to diverse local habitats drives the specialization of flowers to various pollinators. This contrasts with adaptive radiation, which typically involves selection for specific pollinators within a single habitat (Patterson and Givnish, 2004). The ability to reproduce can also be directly linked to polyploidization, and typical patterns of cytotype distribution have been found in different studies (Krak et al., 2013; Valdés-Flórido et al., 2024a). On the other hand, in isolated environments like Hawaiian Islands, plant species often display bursts of phenotypic diversity and chromosomal evolution as a consequence of rapid adaptation resulting in speciation events (Barrier et al., 1999; Bellinger et al., 2022). Thus, polyploidization is a process that assists speciation and diversification into new areas, being able to entail evolution of reproductive strategies. However, sometimes polyploidization could be related to the loss of a trait, such as it occurs with the loss of heterostyly for the two major families that present this trait (Primulaceae and Rubiaceae; Guggisberg et al., 2006; Naiki, 2012).

Aneuploidy has a very drastic impact on genetic dosage (Birchler et al., 2007; Birchler and Veitia, 2007) and is most commonly deleterious, while dysploidy is much more widespread with a high impact on plant evolution (Escudero et al., 2012). While dysploidy does not involve significant changes in DNA content, it can also have an impact on phenotype through structural rearrangements. At macroevolutionary scales, some studies highlight that gain or loss in the number of chromosomes can be associated with novel morphological features and influence diversification processes (Farminhão et al., 2021; Sader et al., 2019). For instance, recent research has shown that ascending dysploidy together with genome size expansion correlates both with larger flowers and higher diversification rates in the subgenus *Passiflora* L., suggesting a positive selection towards bigger genome sizes through morphological/ecological changes (Sader et al., 2019). The recurrent and parallel evolution of the same dysploid cytotype in the genus *Soldanella* L. has consistently resulted in speciation events (Slovák et al., 2023; Rurik et al., 2024). Similarly, Farminhão et al. (2021) found a correlation between dysploidy events and the evolution of leaflessness in the *Dendrophylax-Microcoelia* clade of angraecoids (Orchidaceae) with an eventful karyotypic history dominated by descending dysploidy, although the underlying mechanisms remain unexplored. No increases in net diversification rates could be related to chromosome number changes with the predominance of karyotypic stasis. However, species experiencing shifts in chromosome number appear to show parallel evolution of some phenotypic structures, leaflessness, and changes in floral color (Farminhão et al., 2021).

In summary, bursts of phenotypic and chromosomal evolution can occur simultaneously with speciation, but their relationship is complex due to the timing and interplay between these processes being highly dependent on evolutionary forces and environmental factors (see Box 3 for a case study in the genus *Linum*). On the one hand, CRs, such as inversions or duplications, can lead to changes in traits like morphology, fitness, or reproductive isolation. On the other hand, polyploidy often leads to rapid phenotypic evolution and speciation due to changes in chromosome number, resulting in traits such as altered flowering times or increased environmental tolerance. Adaptive radiation often triggers

rapid speciation and significant phenotypic divergence, but the persistence of these changes can vary since some chromosomal changes may become extinct over time, while others promote long-term diversification. Additionally, shifts in chromosome number, such as dysploidy or aneuploidy, can also contribute to phenotypic diversity and speciation, although their effects vary depending on the context.

Box 3. Exploring biogeographic and ecological trait correlations in chromosome evolution: A case study in the genus *Linum*

The genus *Linum* L. (Linaceae) exhibits high rates of chromosomal evolution, primarily driven by polyploidy and dysploidy events. Despite this, only a limited number of chromosomal speciation events have been inferred across the whole phylogeny (Valdés-Flórido et al., 2023). Specifically, five chromosomal speciation events were inferred, involving both ascending and descending dysploidy, along with two polyploid speciation events. These findings support the higher contribution of anagenetic events compared to cladogenetic ones (Figure 7A). In particular, species within the genus *Linum* are mostly diploid in the Palearctic region, being the ancestral area of distribution (Maguilla et al., 2021), whereas polyploid species are more common in regions outside this area. Rates of both ascending and descending dysploidy are higher in colonized areas, while polyploidization events are more frequent in the genus' original distribution range (Valdés-Flórido et al., 2023; Figure 7B). The model thus supports differing rates of chromosomal evolution between the source area and colonized regions. Interestingly, the elevated rates of dysploidy observed in colonized areas may be associated with *in situ* speciation events. This study also reveals a relationship between chromosome number and plant life history (annual vs. perennial). While most species are perennial, the rates of polyploidy are higher in annual species than in perennials, even though polyploidy has traditionally been associated with perennial life forms (Stebbins, 1971). This unexpected result may be explained by polyploidization events occurring in terminal short branches of some species. Besides, the woodiness and non-clonal nature of perennial species in *Linum* may account for this discrepancy, as polyploidy may not be associated with perenniality per se but rather with clonality (Van Drunen and Husband, 2019). Conversely, rates of descending dysploidy are significantly higher in perennial species.

Q4: Does dispersal into new geographical areas or local environmental changes coincide with chromosome evolution and speciation?

One of the best-known drivers of speciation is geographic isolation, where populations of the same species become separated as a consequence of space in the presence or absence of geographical and

ecological barriers (Dobzhansky, 1951; Coyne and Orr, 2004). Geographical isolation can provide the conditions for chromosome changes to accumulate either through drift or selection and without the diluting effects of gene flow from other populations. These chromosomal changes further enhance population differentiation in morphology, ecology, pre-and/or post-zygotic barriers, cumulatively or individually giving rise to reproductive isolation and eventually speciation (Levin, 2002; Scopece et al., 2010). This indicates that intrinsic postzygotic mechanisms may trigger polymorphism among allopatric conspecific plant populations. In contrast, a study found no links between chromosome transitions and the major diversification events associated with ecological events in the temperate grasses (Pooideae) (Pimentel et al., 2017). However, we should consider confounding factors in such macroevolutionary studies, such as the effect of undetected polyploidization followed by diploidization processes or genome sampling bias due to the use of few genetic markers, which does not negate the existence of chromosomal changes. In addition, chromosome evolution and cladogenetic processes were not modeled together which may lead to biased results if chromosomal changes are, in fact, affecting cladogenesis.

Changes in the environment (e.g., altitude, temperature) can also trigger adaptive responses through chromosome evolution (chromosomal inversion in adaptation; Huang and Rieseberg, 2020). Moreover, the occurrence of polyploidy in the tree of life also seems to correlate with periods of environmental change (Van de Peer et al., 2017). For instance, polyploidy can cause variation in plant functional traits and generate individuals that can adapt and exploit new environmental niches (Wan et al., 2020) and can facilitate adaptive response to harsh environmental conditions (Alix et al., 2017). Specifically, environmental stress has been proposed to foster the production of unreduced gametes, which are the main drivers of polyploidization in angiosperms (Bretagnolle and Thompson, 1995; Levin, 2002). The formation of diploid pollen grains has been promoted by low temperatures in the genera *Solanum* L. (Solanaceae), *Datura* L. (Solanaceae), *Oenothera* L. (Onagraceae), or *Epilobium* L. (Onagraceae) (e.g., McHale, 1983; Alsamir et al., 2021; Krakos et al., 2022). However, not only do low temperatures enhance the production of unreduced gametes, high temperature environments also have the potential to increase ploidy levels as seen in the genera *Rosa* L. (Rosaceae) (Pécirix et al., 2011; Crespel et al., 2015) and *Populus* L. (Salicaceae) (Wang et al., 2017). Other environmental factors, such as temperature fluctuations, low nutrient stress, or the presence of parasites and viruses have similarly been reported to promote the formation of unreduced gametes (Levin, 2002). These strategies are consistent with a broader adaptability and ecological tolerance and higher invasive potential of polyploids than their diploid relatives (Pandit et al., 2011; Te Beest et al., 2012; Marks et al., 2024).

Geographical and environmental pressures (or only one of them) may also occur simultaneously, driving chromosomal changes that indirectly promote speciation (Coyne and Orr, 2004). In addition, chromosome evolution has been strongly linked to biogeography in angiosperms (Rice et al., 2015), with polyploidization showing significant evolutionary implications, including the potential for range expansion (Te Beest et al., 2012; Soltis et al., 2015). For

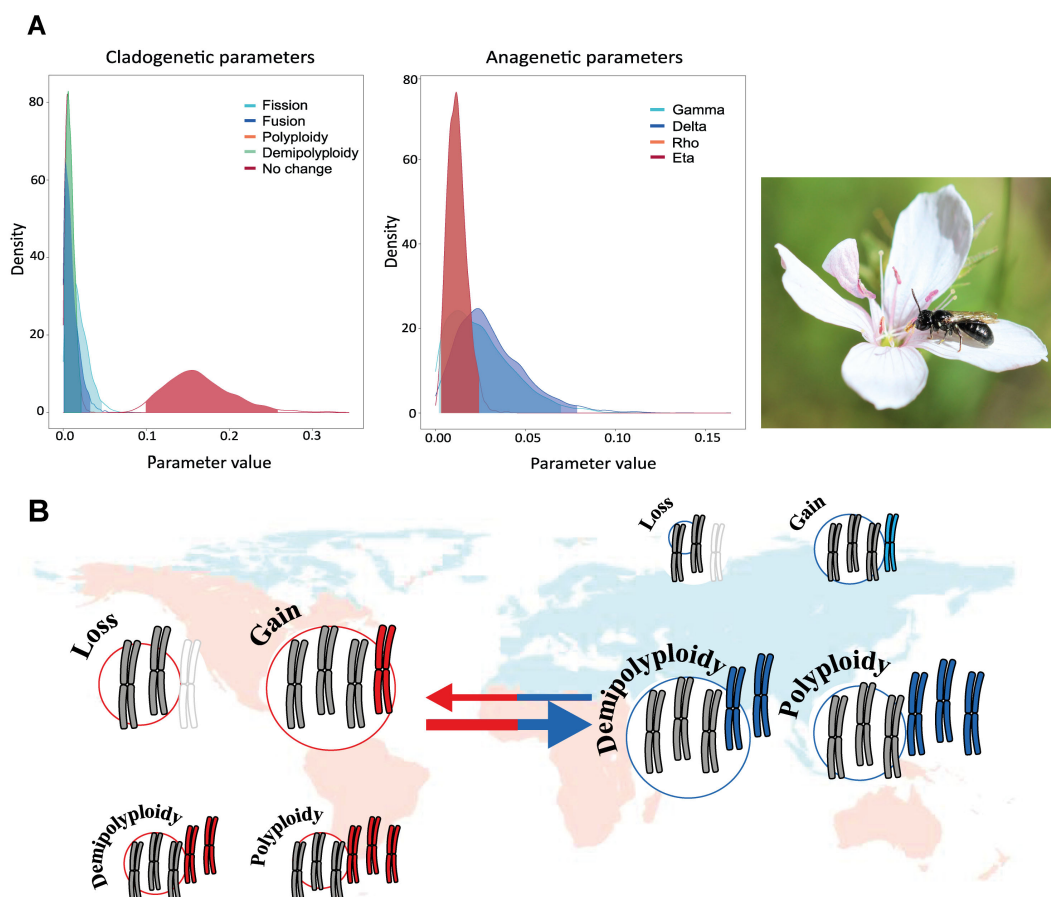


FIGURE 7

Modified from Valdés-Florió et al. (2023). (A) Posterior probability densities of the estimated clado- and anagenetic parameters using ChromoSSE in the *Linum* phylogeny. The x-axis displays the parameter value, and the y-axis indicates the posterior probability density of each value. The species in the photo is *L. tenuifolium* Schousb. Photo credits to B. Arroyo. Gamma represents the rate of chromosome gains, Delta the rate of chromosome losses, Rho the rate of polyploidization, and Eta corresponds to demipolyploidization, permitting the multiplication of the number of chromosomes by 1.5. (B) Correlation between chromosome number evolution and biogeography. Arrow and circle diameter is proportional to the rate.

instance, in the genus *Panax* L. (Araliaceae) it has been demonstrated that the ancient and recent WGDs along with geographical and ecological isolations might have together contributed to the diversification of this genus, suggesting that distinct selection pressures appear to have acted during the genus' evolutionary history (Shi et al., 2015). In the genus *Centaureum* (L.) Hill polyploid species may have an optimal climatic niche related to harsher environments (Valdés-Florió et al., 2024b; see Box 4). However, this cytotype adaptation is not always linked to speciation, as cytotypes can coexist within a single species as seen in the case of cryptic invasion of polyploid *Centaurea stoebe* L. expanding into the range of its diploid relative in Europe (Rosche et al., 2025).

Taking into consideration the above, geographic and environmental changes can trigger chromosome evolution, which can either coincide or accumulate post exposure together putatively driven by the severity of the changes experienced by plants. Geographic isolation is a key driver of this process, as it allows for chromosomal changes to undergo fixation in populations without gene flow. These chromosomal changes, including inversions and polyploidy, can contribute to further differentiation in traits, such as morphology

and ecology (stress factors like temperature fluctuations or low nutrients availability), fostering reproductive isolation and speciation. In some cases, both geographic isolation and environmental pressures work together, while in other instances, environmental factors alone can influence chromosome evolution and diversification.

Box 4. Chromosome evolution and climatic adaptation in *Centaureum*

Some studies in the genus *Centaureum* (Gentianaceae) have examined the interplay between biogeography, climatic niche, and polyploid evolution. One of them revealed that diploid species primarily occupy the ancestral area at the Mediterranean Basin, while polyploids have successfully expanded into northern temperate regions as well as southern and eastern arid regions (Maguilla et al., 2021). Applying ChromoSSE to infer chromosome number evolution across the genus highlights several important patterns. Although a significant number of cladogenesis events are associated with polyploidization events, most cladogenetic events

do not correspond to chromosomal changes. Anagenetic changes are associated with both dysploidy and polyploid events (Figure 8A; Valdés-Flórido et al., 2024b). Polyploid speciation was inferred at both ancestral and more recent nodes and branches, while dysploidy events predominantly occur along terminal branches (Valdés-Flórido et al., 2024b). Most transitions from diploid to tetraploid appear to be associated with transitions from drier, warmer to colder, wetter climatic niches, as well as the expansions from southern to northern distribution ranges (Figure 8B). In contrast, transitions leading to the hexaploids coincide with transitions from temperate to warmer and drier climatic niches at the southern distribution limit of the genus (Figure 8B; Maguilla et al., 2021; Valdés-Flórido et al., 2024b). These findings suggest a strong link between polyploidization and climatic adaptation in the mostly Mediterranean *Centaureum* genus, with specific polyploid levels corresponding to distinct ecological niches and geographic distribution within its range. Although polyploidization itself does not necessarily drive dispersal within this genus, it appears to enhance the likelihood of establishment and persistence in newly colonized areas (Maguilla et al., 2021). Therefore, while geological barriers likely play a role in the speciation process of *Centaureum*, the observed pattern of niche

expansion of polyploids may reduce competitive pressures and improve lineage survival.

Q5: Do ecological interactions leave signatures on chromosome macroevolution?

Ecological interactions describe the diverse mechanisms through which organisms influence each other's survival, reproduction success, and distribution within ecosystems. These interactions can occur between individuals of the same species (intraspecific) or between different species (interspecific) (Schoener, 1990), and play an important role in evolution, as they can act as selective forces during speciation (Thompson, 2009). From a microevolutionary perspective, empirical evidence supports the idea that ecological interactions have acted as selective forces on certain CRs (Burak et al., 2018). However, from a macroevolutionary point of view, there are not many cases where such interactions have left detectable signature in the speciation patterns among plant groups.

Among the aforementioned CRs, polyploidy yet again, is the most widely studied in the context of ecological interactions (Thompson

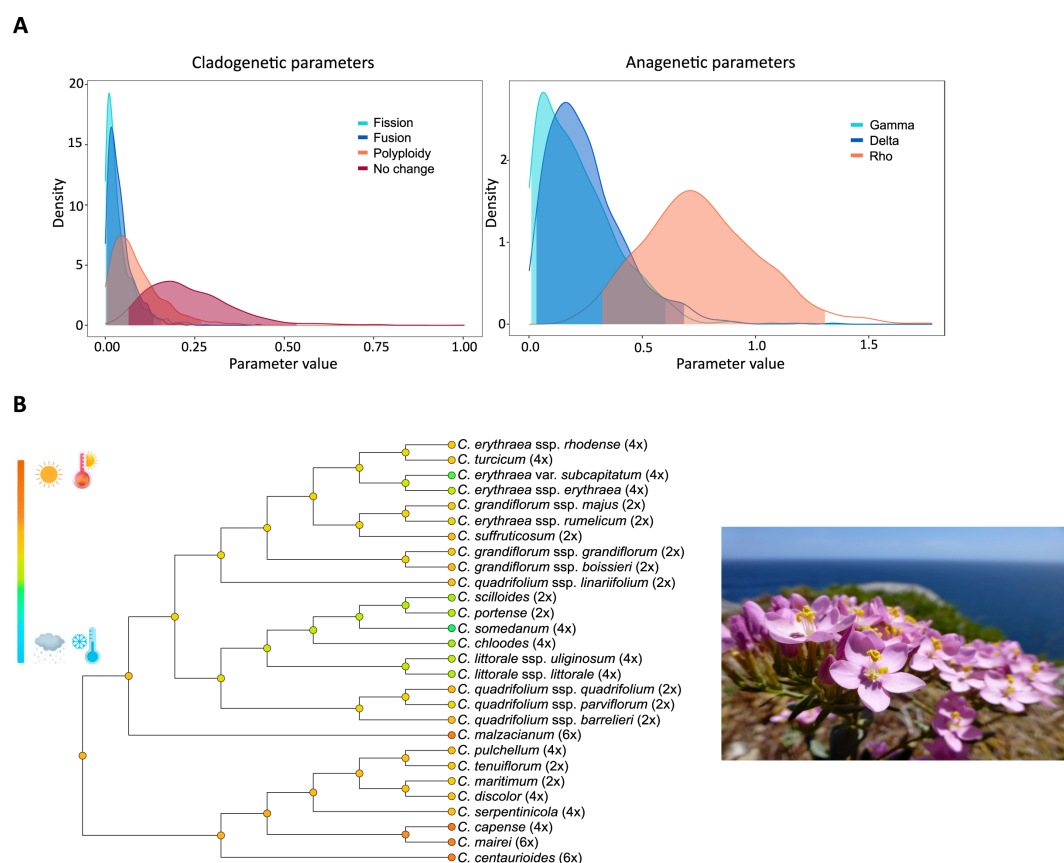


FIGURE 8

Modified from Valdés-Flórido et al. (2024b). (A) Posterior probability densities of the estimated clado- and anagenetic parameters using ChromoSSE. The x-axis displays the parameter value, and the y-axis indicates the posterior probability density of each value. Gamma represents the rate of chromosome gains, Delta the rate of chromosome losses, and Rho the rate of polyploidization. (B) On the left climatic niche characterization of *Centaureum*. Colors in the nodes of the phylogeny correspond to the mean value of the climatic values used for the study. The species in the photo is *C. grandiflorum* ssp. *majus* (Hoffmanns. and Link) Díaz Lifante. Photo credits to S. Castro.

et al., 2004; Segraves and Anneberg, 2016). Polyploidization has the potential to create genetically isolated entities with divergent genetic and phenotypic traits that can shape the interactions of plants with other organisms, and likewise, these organisms can act as selective forces in stabilizing new polyploid races (Segraves and Anneberg, 2016). For species that rely on animal-mediated pollination, pollinators can contribute to the reproductive isolation of polyploids from their diploid relatives through assortative mating (Rezende et al., 2020). With the exception of the genus *Nicotiana* L. (Solanaceae), where pollinators have significantly influenced macroevolutionary patterns of speciation through floral color selection (McCarthy et al., 2015), clear examples of pollinators shaping diversification patterns in polyploid lineages are rare. In fact, the results of a current study, which examined whether neo-polyploidization in *Arabidopsis arenosa* (L.) Lawalrée led to changes in flower size that might influence pollinator behavior, found no evidence for assortative mating due to polyploidization. Instead, it was observed that polyploidization facilitated pollen exchange between different ploidy levels (Schmickl et al., 2024). Additionally, polyploids may evolve new defense mechanisms against herbivores, such as the production of new secondary chemical compounds (Orians, 2000; Edger et al., 2015; see more details in Box 5) or an expanded host range (Nuismer and Thompson, 2001; Arvanitis et al., 2010), which could also influence patterns of speciation in polyploid lineages and their diploid relatives. One of the most compelling examples of how key innovations

responses to herbivory can shape macroevolutionary speciation patterns in plants through gene and WGD as described by Edger et al. (2015). In this scenario, the evolutionary arms race between members of the order Brassicales and pierid butterflies has played a significant role in driving the diversification rates of both groups (Figure 9). The evolutionary interplay between the two organismal groups has led to co-evolutionary dynamics, where plants evolve new chemical defenses against herbivory, while the butterflies develop mechanisms to overcome these defenses. Changes in the secondary chemistry of polyploids can also influence their interactions with other organisms, offering potential protection from parasites and pathogens (Burdon and Marshall, 1981; Vleugels et al., 2013) or disrupting relationships with mutualistic fungi, possibly resulting in fewer associations (Gundel et al., 2014; Franco et al., 2015). However, the underlying mechanisms remain complex and no documented cases have demonstrated a lasting impact on the macroevolutionary patterns of speciation in polyploid races. Lastly, an increase in genetic variability provided by polyploidization can enhance the environmental tolerance of polyploids, making them more competitive and prone to invasions (Pandit et al., 2011; Thébault et al., 2011; Te Beest et al., 2012; Cheng et al., 2021; Moura et al., 2021). This phenomenon is exemplified by certain species of the genus *Spartina* Schreb. (Poaceae), where meso- and neo-polyploid events have enabled them to be more competitive in stressful habitats by introducing novel regulatory patterns in gene expression (Giraud et al., 2021).

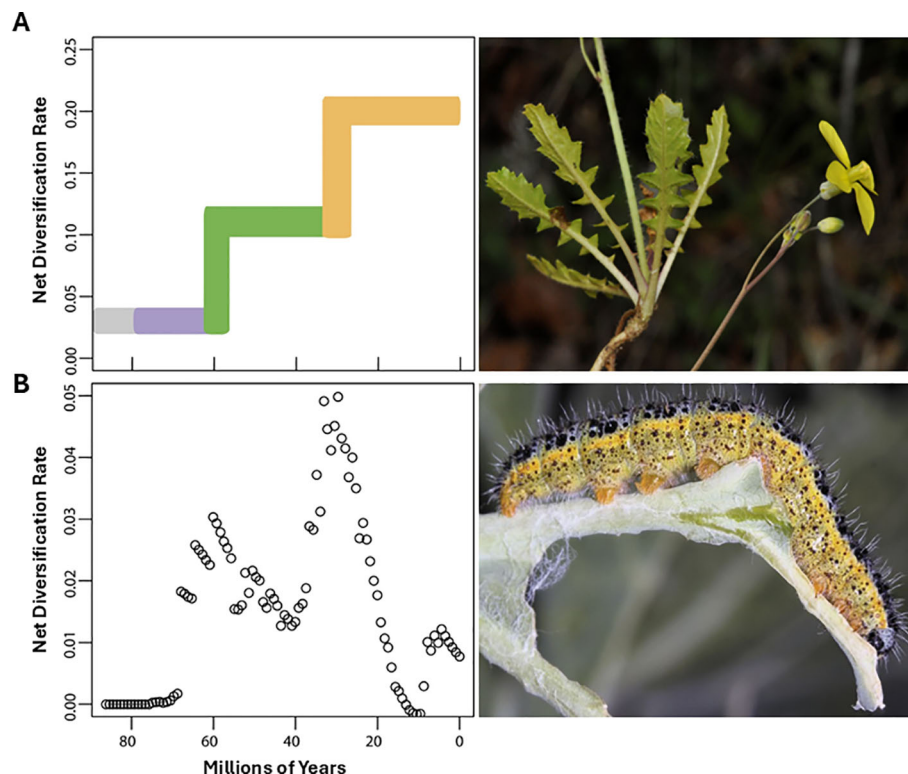


FIGURE 9

Modified from Edger et al. (2015). (A) Shifts in diversification rates during Brassicales Bromhead evolution. Colors indicate the emergence of indolic glucosinolates (purple), methionine derived glucosinolates (green), and novel structural elaborations in Brassicaceae Burnett lineage (orange). (B) Diversification of Pierinae butterflies during the same period. Time estimates are shown at the bottom. Photo credits: *Brassica barrelieri* (L.) Janka (top) by M. Luceño, and *Pieris brassicae* L. feeding on a *Brassica* species (bottom) by Edger et al.

However, not only polyploidy but also dysploidy events may also have an impact in ecological interactions. A higher chromosome number may lead to higher recombination rates (Nijalingappa, 1974; Bell, 1982; Nokkala et al., 2004; Escudero et al., 2012). Higher recombination rates provide more evolutionary potential, which is advantageous in highly competitive communities with temporarily predictable environments (Koella, 1993; Escudero et al., 2013a; Escudero and Hipp, 2013). Holocentric chromosomes distribute centromeric activity along their length, unlike the single centromere of monocentric chromosomes (Márquez-Corro et al., 2018). This structure enables fragmented chromosomes to retain kinetochore function, reducing segregation errors during meiosis and potentially facilitating chromosomal fissions and fusions without compromising genome stability. In a highly competitive environment, this genetic flexibility can offer an advantage. Conversely, low rates of recombination should be positively selected in unstable and low competition communities where a pioneering strategy could be successful (Stebbins, 1958; Grant, 1958; Bell, 1982; Koella, 1993). Accordingly, in high competition environments, higher numbers of chromosomes are expected in holocentric species. This hypothesis has been tested at the microevolutionary level within *Carex laevigata* group populations (sect. *Spirostachyae*, Escudero et al., 2013a). The results indicated that chromosome numbers are indeed higher in lowland ancestral areas where competition is more intense. However, when the same hypothesis was extended to the whole genus *Carex*, the relationship between chromosome number and competition was less clear (Bell, 1982; Escudero et al., 2012). This indicates that other evolutionary or ecological factors could have shaped chromosome number evolution at a broader scale within this group.

Chromosomal rearrangements that do not involve changes in chromosome number can also play a role in adaptation by reducing recombination between favorable combinations of alleles (Kirkpatrick and Barton, 2006; Lowry and Willis, 2010). This reduced recombination contributes to speciation in a similar way - by suppressing recombination between local adapted alleles and those causing assortative mating (Trickett and Butlin, 1994). One of the examples is the differentiation of monkeyflowers (*Erythranthe guttata* G.L.Nesom, Phrymaceae) populations in two ecotypes, where a large inversion has been shown to affect the growth form (Lowry and Willis, 2010), herbivore resistance through secondary compounds synthesis (Kollar et al., 2024), as well as also causing assortative mating by allochory in flowering time (Lowry and Willis, 2010). At a broader evolutionary scale, in this same genus, another inversion has been associated with differences in corolla length and flower color contributing to both prezygotic and postzygotic isolation of a sister species pair: *Erythranthe lewisii* (Pursh) G.L.Nesom and N.S.Fraga and *Erythranthe cardinalis* (Douglas ex Benth.) Spach (Fishman et al., 2013). Besides, Johnson et al. (2009) described SVs as having an effect on ecological interactions in plants at the macroevolutionary level. Under this scenario, some species from the Onagraceae family display Permanent Translocation Heterozygosity (PTH), a characteristic that prevents pairing of homologous chromosomes in meiosis and thus recombination. Moreover, most of these species are self-fertilizing, resulting in offspring that are genetically identical

to the parent. The study found that species with PTH are more susceptible to generalist herbivores, suggesting that these may have reduced defenses against herbivory, likely due to their limited genetic variability.

Overall, ecological interactions leave a detectable signal on chromosome macroevolution, although the relationship is complex and not always straightforward. Evidence shows that ecological interactions, such as pollinator preferences, herbivory, and competition, can act as selective forces on CRs and influence speciation patterns in a microevolutionary and macroevolutionary scale.

Box 5. Coevolutionary arms-race: gene and genome duplications driving diversification in plants and herbivores

One of the primary drivers of life's diversity on Earth is coevolution between organisms that maintain close ecological interactions (Thompson, 2009). The mutual pressures that these organisms exert on each other can act as strong selective forces, driving the emergence of biodiversity. Moreover, as seen in this section, key innovations resulting from these selective pressures can be facilitated by CRs, creating new adaptive potential. Edger et al. (2015) shows how key innovations in response to herbivory, driven by gene and whole genome duplications, can influence macroevolutionary speciation patterns.

Plants from the order Brassicales can produce glucosinolates, a secondary metabolite which upon tissue damage are transformed into toxins, making them harmful to their main herbivores, the caterpillars of *Pierinae* Swainson butterflies. In contrast, these caterpillars have developed a mechanism to detoxify glucosinolates by utilizing a gene that encodes a nitrile-specifier protein (NSP), which converts these compounds into inert metabolites.

The acquisition of key innovations by plants and butterflies has unfolded in several stages. When Brassicales arose approximately 92 Mya, they could only synthesize glucosinolates from phenylalanine and branched-chain amino acids. The complexity of these chemical compounds diversified after a WGD approximately 77.5 Mya. This event enabled the production of indolic glucosinolates from tryptophan through the neofunctionalization of duplicated genes involved in glucosinolate synthesis. Adaptation of pierid butterflies to Brassicales occurred approximately 68 Mya and was facilitated by the evolution of glucosinolate detoxification. This key innovation significantly increased the diversification rate of the herbivores. Further complexity arose when additional gene duplications in the ancestors of Capparaceae and Cleomaceae plant families enabled the synthesis of methionine-derived glucosinolates, again via gene neofunctionalization, contributing to the diversification of Brassicales. However, the evolution of different copies of the NSP gene in *Pierinae* butterflies allowed them to adapt to these new compounds, as these copies developed functional differences against glucosinolates detoxification, which helped them overcome the plant defenses and enabled further colonization and diversification. The final phase of diversification took place around 32 Mya with the emergence of the Brassicaceae family. A subsequent WGD event

enabled the origin and retention of genes involved in glucosinolate synthesis, driving the remarkable diversification of Brassicaceae. After their appearance, two different lineages of Pierinae butterflies (*Anthocharidini* Scudder and *Pierini* Swainson) colonized Brassicaceae. Both lineages utilized different copies of the NSP gene, each with functional adaptations to detoxify glucosinolates, and these colonization events coincided again with a significant increase in diversification rates.

Q6: What is the role of hybridization-induced chromosomal changes in bridging microevolutionary processes with macroevolutionary patterns?

Hybridization serves as a critical evolutionary mechanism that plays a significant role in shaping diversification and speciation processes in plants (Soltis and Soltis, 2009; Taylor and Larson, 2019). It can play a creative role in species evolution, having both positive and negative effects, contributing to species evolution by triggering hybrid speciation (Abbott et al., 2013; Buerkle et al., 2000; Taylor and Larson, 2019), facilitating adaptive introgression (Suarez-Gonzalez et al., 2018), and even fueling adaptive radiations (Barrier et al., 1999; Mandáková et al., 2010b; Stankowski and Streisfeld, 2015). However, hybridization can also result in negative consequences, such as complete hybrid sterility, extensive introgression that merges previously separated gene pools, thereby hindering speciation and potentially leading to the extinction of parental taxa (Kearns et al., 2018; Slovák et al., 2023; Todesco et al., 2016). From a karyotypic perspective, the effects of hybridization depend on the compatibility of parental genomes, with ploidy level being a key factor. Hybridization can occur without changes in chromosome number (homoploid hybridization) or may involve WGD (allopolyploidization) (Ramsey and Schemske, 1998; Nieto Feliner et al., 2020; see a case study in Box 6 for details on both mechanisms).

Allopolyploidization promotes speciation by integrating distinct parental genomes and establishing thus immediate reproductive barriers (Ramsey and Schemske, 1998; Pikaard, 2001; Mallet, 2007; Qiu et al., 2020). A major evolutionary advantage of allopolyploids is their increased heterozygosity, which confers hybrid vigor while overcoming some challenges of homoploid hybridization such as (sub)genomic incompatibility (Comai, 2005; Qiu et al., 2020). This can drive phenotypic changes and adaptations, potentially leading to further lineage diversification (Tayalé and Parisod, 2013; Mutti et al., 2017). However, allopolyploidization also presents remarkable challenges that can affect the viability and establishment of new allopolyploids, with potential negative implications for speciation. Newly formed allopolyploids face extrinsic challenges, such as population bottlenecks (Novikova et al., 2017) and competition with parental species (Fowler and Levin, 1984), as well as intrinsic challenges, including complications in chromosome segregation (Bomblies et al., 2016) and changes in the genome structure (Leitch and

Leitch, 2008). As WGD significantly reduces or eliminates homeologous recombination in the hybrid genome, potential incompatibilities between divergent parental subgenomes cannot be effectively purged, which may pose significant challenges for newly formed allopolyploids (Rieseberg, 2001b; Qiu et al., 2020). As a result, the genomes of newly evolved allopolyploids must undergo continuous alterations at the genetic, epigenetic, transcriptomic, and proteomic levels during the early stages of establishment to achieve genomic stabilization (Chelaifa et al., 2010; Chester et al., 2012; Edger et al., 2017; Qiu et al., 2020). These genomic changes are typically absent in older allopolyploid lineages (Burns et al., 2021), indicating that structural and expression plasticity of genomes in newly formed allopolyploids is crucial for their stabilization and integration, with the timing of these changes varying among species (Wendel et al., 2018; Burns et al., 2021).

The association between allopolyploidy and diversification processes is frequently linked to biological traits; however, even greater attention has been given to adaptations within ecological and biogeographical contexts. One key aspect of establishment and subsequent evolution of allopolyploids is their niche breadth and shifts relative to their diploid progenitors. In contrast to previous assumptions (Stebbins, 1984), recent comparative studies have shown that ecological niche shifts in allopolyploids, relative to their diploid progenitors, are highly variable, exhibiting patterns of expansion, contraction, intermediacy, and novelty (Blaine Marchant et al., 2016; Parisod and Broennimann, 2016). Mata et al. (2023) further found no consistent differences in the distribution ranges or habitat types of allopolyploids in relation to extreme conditions; allopolyploid species do not necessarily occupy more extreme environments or broader geographic ranges compared to their diploid progenitors. The significant overlap observed between the niches and distribution ranges of allopolyploids and their progenitors suggests that these niches are largely shaped by the climatic and geographical characteristics of the parental species. However, biotic and microclimatic factors likely play a significant role in the establishment of allopolyploids (Blaine Marchant et al., 2016; Griffiths et al., 2019; Akiyama et al., 2020; Mata et al., 2023). Reevaluating the classical view that allopolyploids predominantly thrive in deglaciated temperate habitats (Stebbins, 1984) reveals contradictions considering recent findings. Indeed, an increasing number of studies emphasize the evolution of allopolyploids in the Mediterranean region, where they often exhibit restricted distribution ranges and specialized ecological niches (López-González et al., 2021; Šlenker et al., 2021; Kantor et al., 2023). It seems that the ecological dynamics of allopolyploids may be more complex than previously recognized, as their successful establishment in these environments suggests adaptive strategies that enable them to thrive despite often limited geographical distributions.

The critical question of whether allopolyploidy facilitates species diversification, especially at the macroevolutionary level through mechanisms such as species radiation, remains inadequately understood. Nevertheless, studies on allopolyploids in the tribe Microlepidieae (Brassicaceae) Al-Shehbaz and approximately 50 taxa of *Nicotiana* sect. *Suaveolentes* Goodsp. (Solanaceae) suggest that all taxa in both groups are derived from a single allopolyploid ancestor,

from which they diversified and radiated across the Southern Hemisphere (Clarkson et al., 2004; Kelly et al., 2013; Mandáková et al., 2017; Chase et al., 2023). In both cases, diploidization has led to the evolution of diverse dysploid lineages, facilitating further diversification processes (Mandáková et al., 2017; Chase et al., 2023). Similarly, a study by Tomlin et al. (2024) uncovered the allopolyploid origin of monophyletic Hawaiian mints (Lamiaceae), which are derived from North American ancestors of the genus *Stachys* L. While the authors did not directly test the impact of allopolyploidy on the diversification and radiation of Hawaiian mints, they proposed that this allopolyploid ancestry could provide a genomic substrate for morphological differentiation within the lineage and potentially foster evolutionary radiation in the rapidly evolving Hawaiian landscape (Tomlin et al., 2024). In contrast, Estep et al. (2014) found that while one-third of species in the grass tribe Andropogoneae Dumort. (Poaceae) are allopolyploids, diversification did not precede the allopolyploidization event and does not correlate with subsequent speciation bursts. The emergence of Andropogoneae species in the Late Miocene coincides with the expansion of major C4 grasslands, and although allopolyploidy remains a significant mode of speciation within this tribe, its role in diversification is less clear. Furthermore, additional studies indicate that allopolyploidy may catalyze diversification and even radiation, although these hypotheses still need to be tested (Barrier et al., 1999; Julca et al., 2018).

In conclusion, allopolyploidy plays a significant role in rapid speciation by generating new species that exhibit hybrid vigor and enhanced ecological potential (Mallet, 2007). While it facilitates evolutionary diversification, the relationship between chromosomal evolution and ecological adaptation is variable and influenced by the ecological requirements of the progenitors. Although some studies suggest that allopolyploidy can trigger lineage diversification or rapid radiations, further comprehensive empirical investigations are needed to support these findings and allow for broader generalizations.

Homoploid hybrid speciation (HHS), in contrast to allopolyploid speciation, occurs without WGD, resulting in hybrid species that retain the same chromosome number as their parental species (Stebbins, 1958; Mallet, 2007; Soltis and Soltis, 2009; Abbott et al., 2010; Schumer et al., 2014). Unlike allopolyploidy, HHS does not immediately generate reproductive barriers. Instead, hybrid sterility, which restricts gene flow between hybrids and progenitors, must evolve through genetic incompatibilities (genic sterility) or CRs (chromosomal sterility) (Stebbins, 1958; Abbott et al., 2010; Yakimowski and Rieseberg, 2014). In the absence of reproductive barriers, homoploid hybrids are susceptible to backcrossing with their parental species, which can lead to introgression, the formation of hybrid zones, reinforcement, or even genetic assimilation (Soltis and Soltis, 2009; Todesco et al., 2016; Aguillon et al., 2022). An alternative mechanism for the stabilization and establishment of homoploid hybrid species involves extrinsic factors, such as spatial isolation from parental taxa, often accompanied by ecological divergence (Grant, 1981; Gross and Rieseberg, 2005; Abbott et al., 2010; Yakimowski and Rieseberg, 2014). Chromosomal rearrangements are widely recognized as a key factor for establishing intrinsic reproductive barriers between homoploid hybrids and their parental species, a

phenomenon explained by the recombinational model of HHS (Stebbins, 1957; Grant, 1958, 1981; Buerkle et al., 2000). According to this model, parental species exhibit at least two independent CRs, resulting in reduced gamete viability in F1 hybrids due to heterozygosity. Subsequently, homozygous recombinants in the F2 generation may restore self-fertility while remaining incompatible with the parental species, potentially enabling sympatric speciation. This recombinational model demonstrates how hybrid lineages can achieve reproductive isolation from their parental species in sympatry, suggesting it as a likely pathway for initiating HHS. However, extrinsic reproductive barriers, such as ecological divergence between hybrids and parental species, may further influence the success of HHS (Buerkle et al., 2000).

Demonstrating HHS driven by chromosomal or genetic mechanisms, however, is not straightforward and requires a comprehensive investigation that integrates (cyto)genomic, ecological, and experimental approaches (Seehausen, 2004; Mallet, 2007; Schumer et al., 2014). This evidence includes reproductive isolation from parental species, documentation of past hybridization events, and confirmation that isolating mechanisms have emerged as a result of hybridization (Schumer et al., 2014). Therefore, evidence of HHS driven by CRs has only been documented in a limited number of plant systems (Rieseberg et al., 1995; Archibald et al., 2005; Wu and Tanksley, 2010; Yakimowski and Rieseberg, 2014; Ostevik et al., 2020). Nonetheless, the genus *Helianthus* L. (sunflowers) serves as the most iconic model system for understanding HHS, where CRs play a crucial role in establishing chromosomal sterility and driving speciation (Rieseberg et al., 1995; Ostevik et al., 2020; Todesco et al., 2020). The dynamic chromosomal evolution in sunflowers, driven by rearrangements, has facilitated rapid diversification, primarily through intrachromosomal inversions and interchromosomal translocations (Rieseberg et al., 1995; Ostevik et al., 2020). Both types of rearrangements are common in plant genome and karyotype evolution (Weiss-Schneeweiss and Schneeweiss, 2013), although intrachromosomal rearrangements tend to occur more frequently than interchromosomal ones in plant evolution (Wu and Tanksley, 2010; Ostevik et al., 2020). In addition, CRs that induce hybrid sterility appear to be strongly linked to an annual life strategy. These rearrangements are more frequent in annuals, which undergo more meiotic events per generation, thereby accelerating chromosomal mutation rates (Archibald et al., 2005; Owens and Rieseberg, 2014; Yakimowski and Rieseberg, 2014). While there is growing evidence supporting interspecific hybridization prior to adaptive radiations in both plants and animals (Seehausen, 2004; Meier et al., 2017; Stankowski and Streisfeld, 2015; Svardal et al., 2020; Skopalíková et al., 2023), documented cases of ancient homoploid hybridization preceding lineage diversification and radiation in plants remain scarce (Pease et al., 2016; Liu et al., 2017). Nevertheless, HHS occurs through genetic or chromosomal sterility. If the latter is involved, the extent to which CRs contribute to the homoploid speciation in the ancestors of these radiations remains unclear. In conclusion, the role of CRs in establishing reproductive barriers against closely related or more distantly related species during HHS-induced diversifications, and particularly radiation, remains largely unexplored. We propose that

investigating the role of HHS and particularly the influence of CRs as a mechanism of speciation in plant diversification, presents significant opportunities for future research.

Box 6. Evolutionary drivers and consequences of autopolyploidy vs. allopolyploidy: the case of *Arabidopsis*

Arabidopsis Heynh. is a leading model for plant genetics and physiology, primarily due to the well-studied species *A. thaliana* (L.) Heynh. However, the genus *Arabidopsis* is broader beyond the selfer *A. thaliana*, encompassing approximately six predominantly outcrossing diploid species of varying ecological niche and distribution area, with its center of diversity in Europe (Koch, 2019). In addition to sporadic reports of autopolyploid *A. thaliana* accessions (Bomblied and Madlung, 2014), there are two species encompassing established autotetraploid lineages: *A. arenosa* (Arnold et al., 2015; Kolář et al., 2016; Monnahan et al., 2019) and *A. lyrata* (L.) O'Kane & Al-Shehbaz (Schmickl et al., 2010; Bohutínská et al., 2024), and two allopolyploid species resulting from interspecific hybridization within the genus: *A. kamchatica* (Fisch. ex DC.) K. Shimizu & Kudoh (Paape et al., 2018; Kolesnikova et al., 2023) and *A. suecica* (Fr.) Norrl. (Novikova et al., 2017; Burns et al., 2021). The recent origin of both auto- and allopolyploids (Pleistocene; Novikova et al., 2018) within the same well-characterized genus has provided valuable opportunities to address general questions regarding the origins, post-WGD diversity dynamics, and the evolutionary significance of WGD and hybridization.

Interestingly, several distinct evolutionary features have emerged that differentiate these two types of polyploidy. Firstly, while autopolyploidy arises and remains exclusively in outcrossing lineages, allopolyploid origin is exclusively linked to a rapid transition toward selfing (Novikova et al., 2017, 2023; Monnahan et al., 2019; Kolesnikova et al., 2023). Secondly, post-WGD evolution appears to play an important role in shaping both adaptive and deleterious genetic variation in autopolyploids (Yant et al., 2013; Monnahan et al., 2019; Bohutínská et al., 2024; Vlček et al., 2025). In contrast, genetic variation inherited from diploid ancestors significantly contributes to the diversity of allopolyploids, underlying their patterns of breeding system (Novikova et al., 2018, 2023; Kolesnikova et al., 2023) and genome-wide diversity (Paape et al., 2018). Finally, the origin of polyploidy seems to be constrained by the availability of (pre) adaptive variation, which would enable rapid post-WGD adaptation to the challenges imposed by the transition to polyploidy (Figure 10). Some species have undergone none (*A. halleri* (L.) O'Kane & Al-Shehbaz and rare species) or few (single-WGD origin in *A. arenosa*; Arnold et al., 2015; Monnahan et al., 2019) WGD events, while others are prone to recurrent polyploidization (*A. lyrata*; Kolesnikova et al., 2023; Bohutínská et al., 2024; Scott et al., 2024). Moreover, the origin of allopolyploids seems to be constrained by the spatio-temporal availability of a self-compatible ancestral diploid lineage (Burns et al., 2021, 2024; Kolesnikova et al., 2023), which may enable further

silencing of functional S-alleles from the outcrossing parent (Novikova et al., 2017). As expected in polyploid speciation, WGD imposes a strong postzygotic reproductive barrier between ploidies (Morgan et al., 2021). However, despite this triploid block, interploidy gene flow toward autotetraploids occurs in several natural ploidy contact zones (Jørgensen et al., 2011; Monnahan et al., 2019). Notably, WGD also opens up the possibility for further interspecific hybridization, including adaptive introgression, as seen in the case of post-WGD gene flow between autotetraploids of *A. arenosa* and *A. lyrata* (Marburger et al., 2019; Schmickl and Yant, 2021; Bohutínská et al., 2024; Scott et al., 2024). Thus, WGD in *Arabidopsis* not only acts as a speciation trigger (origin of polyploid lineages and species), but also as a factor dissolving species boundaries between previously reproductively isolated species (Lafon-Placette et al., 2017).

In summary, *Arabidopsis* shows that natural WGD transitions may go through multiple evolutionary trajectories, with their direction critically dependent on the mode of polyploid origin, ancestral variation, and the breeding system of the founding lineage(s). Investigating other systems in a comparative manner will provide insights into the generality of these findings and may further improve our understanding of possible transition cases between the two extreme modes of polyploidy, such as segmental allopolyploids.

Discussion

Challenges and outlooks in understanding CRs and evolution

Phylogenetic trees, which are used to reconstruct the timeline of cladogenetic events, are typically built using genetic divergence data, such as nucleotide substitution models, and/or morphological data. However, CRs and karyotypes are currently analyzed within pre-constructed phylogenetic trees, despite CRs being known to drive speciation through macroevolutionary dynamics. If insufficient time has passed since diversification, genetic divergence at the sequence level may not have accumulated sufficiently, leading to variability in estimates of divergence times depending on the phylogenetic datasets used. Even though large-scale CR and genomic architectural data were considered difficult to obtain a few years ago, their potential utility in reconstructing evolutionary relationships has long been recognized (Boore, 2006). Therefore, the inference of timeframes for cladogenetic events could be flawed if CRs were used as inputs for building phylogenetic trees. There may be a disconnection between sequence divergence, morphological divergence, and structural genomic divergence (including all forms of CRs), each of which could yield distinct phylogenetic outcomes.

Several methods have been developed to incorporate whole genome architecture data into phylogenetic reconstruction (Moret and Warnow, 2005; Lin et al., 2012). However, despite the increasing availability of whole genomes, these approaches have yet to gain widespread adoption in the field of evolutionary biology. It remains to be seen whether CRs will eventually be used to build phylogenies and/or resolve complex phylogenetic placement issues,

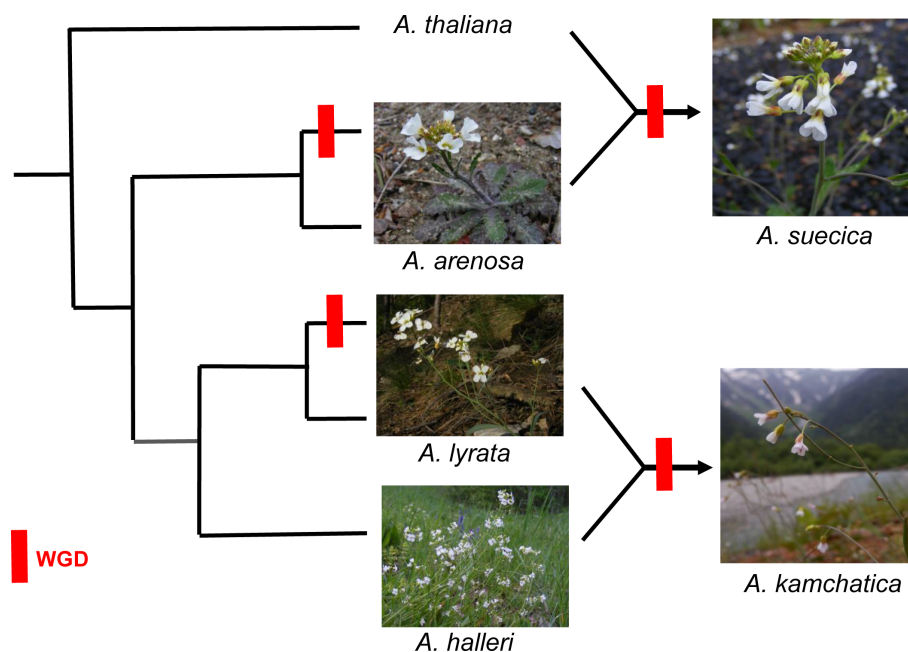


FIGURE 10

A tree showing the evolutionary history of six *Arabidopsis* species. The red bar labeled "WGD" indicates a Whole-Genome Duplication event that occurred in the lineage leading to *A. suecica* (Fr.) Norrl. and *A. kamchatica* (Fisch. ex DC.) K. Shimizu & Kudoh, and the polyploid lineages of *A. arenosa* (L.) Lawalrée and *A. lyrata* (L.) O'Kane & Al-Shehbaz. Photo credits to F. Kolář.

and whether these methods will yield findings that differ from those produced by current sequence-based approaches.

Unraveling missing pieces in CRs and evolution

Bridging the gap between micro- and macroevolution remains an outstanding goal in evolutionary biology. While macroevolutionary patterns and processes can be primarily inferred using indirect measures, future experimental work could shed light on the causality of the CRs as discussed in this review. The availability of long-read sequence data now allows us to establish pangenomes involving different scales, i.e. varieties, breeds, lineages to different species. Such pangenome approaches provide a comprehensive map of CR diversity at an unprecedented scope. Pan-genomes analyses have revealed that many CRs are polymorphic within species and are often linked to adaptation or domestication traits (e.g., Li et al., 2022; Jin et al., 2023; Jayakodi et al., 2024). Therefore, adopting a pan-genomic perspective is essential for a more robust assessment of the evolutionary significance of these CRs. Moreover, genome editing tools now enable direct testing of evolutionary implications of CRs for trait evolution, adaptation, and species diversification.

Chromosomal rearrangements and changes in ploidy are particularly important types of structural changes, as they not only affect genome structure at the genetic level but can also reshape the three-dimensional (3D) organization of the genome within a cell (Kumar et al., 2021). The 3D genome conformation is

hierarchically packaged DNA at multiple levels to facilitate gene regulation and expression within a cell, with features such as chromatin loops and self-interacting genomic regions, known as Topologically Associated Domains (TADs), helping to organize interactions within chromosomes. Beyond these TADs, chromosomes are organized into distinct chromosomal territories, which are spatial compartments where both intra- and inter-chromosomal regions interact. Although TADs are not prominent in plants (commonly found in mammals and fruit flies; Dixon et al., 2012; Hou et al., 2012), TAD-like boundaries have been identified in plant genomes (Ouyang et al., 2020), indicating that 3D chromatin organization plays a role in genome function across different organisms. Furthermore, 3D genome chromatin states in plant genomes can be active, suppressed, or silenced through specific histone modifications, DNA methylation, and functioning of specific enzymes (Ouyang et al., 2020). Thus, alterations in the 3D genome structure, driven by CRs can impact gene expression, chromatin accessibility, and recombination patterns - processes that are directly relevant to species diversification. For instance, in some angiosperms chromosomes are arranged along a telomere to centromere axis (e.g. in common wheat *Triticum aestivum* L.; Hoencamp et al., 2021). By comparing the 3D genome structures of each subgenome in tetraploid cotton (*Gossypium hirsutum* L. and *G. barbadense* L.) with their respective diploid progenitors, it was found that genome polyploidization has influenced significant changes in genome organization (Wang et al., 2018). Specifically, polyploidization has driven the switching of active (A) and inactive (B) chromatin compartments and led to the reorganization of TADs

(Pei et al., 2021). Thus, changes in the 3D genome can in turn affect gene expression, lead to a loss of chromatin accessibility, suppress recombination, and even may result in reproductive isolation (Álvarez-González et al., 2022). One mechanism through which the change in chromosomal 3D conformations can alter transcription is through the loss or gain of chromatin accessibility (Li et al., 2023). Together, these processes could lead to population differentiation, promote divergent adaptation and ultimately lead to speciation (Mohan et al., 2024). It suggests that CRs not only facilitate adaptation at the microevolutionary level but may also contribute to long-term macroevolutionary trends. Alternatively, sex chromosomes may fuse with autosomes, or new sex chromosomes may evolve, giving rise to reproductive strategies that can further facilitate CRs within the genome (Ming et al., 2011).

Thus, integrating insights from 3D genome organization with pangenomic studies of CRs provides a powerful approach to bridging micro- and macroevolutionary processes. By understanding how CRs and ploidy changes influence genome structure and function, we can begin to unravel the complex genetic mechanisms driving species diversification across different timescales. Furthermore, the interplay between CRs and epigenetic modifications, such as DNA methylation, adds another layer of complexity to genome function. These epigenetic changes, along with 3D genome conformation, may be playing a crucial role in regulating gene expression, recombination patterns, and, ultimately in shaping evolutionary species divergence.

Author contributions

CB-B: Writing – original draft, Visualization, Resources, Supervision, Conceptualization, Investigation, Validation, Writing – review & editing. AM: Investigation, Supervision, Writing – original draft, Writing – review & editing. RS-V: Writing – review & editing, Writing – original draft, Investigation, Conceptualization. IG-R: Writing – review & editing, Supervision, Writing – original draft, Conceptualization. AV-F: Writing – original draft, Writing – review & editing, Conceptualization, Supervision. KL: Writing – review & editing, Investigation, Supervision. MS: Writing – original draft, Supervision, Writing – review & editing, Investigation. FK: Investigation, Writing – review & editing, Writing – original draft, Supervision. IJL: Supervision, Writing – review & editing. ML: Supervision, Validation, Writing – review & editing. IL: Writing – review & editing, Supervision, Validation. AC: Supervision, Validation, Writing – review & editing. JC-B: Conceptualization, Writing – review & editing. EM: Conceptualization, Writing – review & editing. AM-R: Conceptualization, Writing – review & editing. PM: Supervision, Writing – review & editing. AS: Validation, Writing – review & editing. SM-B: Conceptualization, Validation, Writing – review & editing, Supervision. ME: Writing – review & editing, Writing – original draft, Resources, Investigation, Conceptualization, Validation, Funding acquisition, Supervision.

Funding

The author(s) declare financial support was received for the research and/or publication of this article. The present work was funded by the Spanish Government and FEDER funds (European Commission) through the granted project PID2021-122715NB-I00. CB-B was supported by a Postdoctoral Fellowship Program grant (JDC2022-048955-I) at the Ministry of Science, Innovation and Universities (MICIU/AEI/10.13039/501100011033) and by the European Union NextGenerationEU/PRTR. ME was supported by a Leonardo fBBVA grant (LEO23-2-9663-CMT-CMA-11). AM was supported by the Fondation Pierre Mercier pour la Science awarded to KL. KL was further supported by the Swiss National Science Foundation grants 202869 and 220868. SM-B, RS-V and ML were supported by the Andalusian regional government (Spain) through project ProyExcel_00125.

Acknowledgments

The authors thank B. Arroyo and S. Castro for providing photographs of the study systems; as well as CM. Tribble, PP. Edger, C. Pires and CW. Wheat for providing the original version of Figures 6, 9. Thanks to the researchers and colleagues from Royal Kew Gardens for supporting the first author during her postdoctoral stay, where much of this review was conducted.

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 author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

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.

References

- Abbott, R., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J. E., Bierne, N., et al. (2013). Hybridization and speciation. *J. Evolutionary Biol.* 26, 229–246. doi: 10.1111/j.1420-9101.2012.02599.x
- Abbott, R. J., Hegarty, M. J., Hiscock, S. J., and Brennan, A. C. (2010). Homoploid hybrid speciation in action. *Taxon* 59, 1375–1386. doi: 10.1002/tax.595005
- Adams, K. L., and Wendel, J. F. (2005). Polyploidy and genome evolution in plants. *Curr. Opin. Plant Biol.* 8, 135–141. doi: 10.1016/j.pbi.2005.01.001
- Aguillon, S. M., Dodge, T. O., Preising, G. A., and Schumer, M. (2022). Introgression. *Curr. Biol.* 32, 865–868. doi: 10.1016/j.cub.2022.07.004
- Akiyama, R., Sun, J., Hatakeyama, M., Lischer, H. E. L., Briskine, R. V., Hay, A., et al. (2020). Fine-scale empirical data on niche divergence and homeolog expression patterns in an allopolyploid and its diploid progenitor species. *New Phytol.* 229, 3587–3601. doi: 10.1111/nph.17101
- Alix, K., Gérard, P. R., Schwarzacher, T., and Heslop-Harrison, J. S. (2017). Polyploidy and interspecific hybridization: partners for adaptation, speciation and evolution in plants. *Ann. Bot.* 120, 183–194. doi: 10.1093/aob/mcx079
- Alsamir, M., Mahmood, T., Trethowan, R., and Ahmad, N. (2021). An overview of heat stress in tomato (*Solanum lycopersicum* L.). *Saudi J. Biol. Sci. Mar.* 28, 1654–1663. doi: 10.1016/j.sjbs.2020.11.088
- Álvarez-González, L., Burden, F., Doddamani, D., Malinverni, R., Leach, E., Marín-García, C., et al. (2022). 3D chromatin remodelling in the germ line modulates genome evolutionary plasticity. *Nat. Commun.* 13, 2608. doi: 10.1038/s41467-022-30296-6
- Anjali, N., Ganga, K. M., Nadiya, F., Shefeek, S., and Sabu, K. K. (2016). Intraspecific variations in cardamom (*Elettaria cardamomum* Maton): assessment of genomic diversity by flow cytometry, cytological studies and ISSR analysis. *SpringerPlus* 5, 1560. doi: 10.1186/s40064-016-3226-x
- Archibald, J. K., Mort, M. E., Crawford, D. J., and Kelly, J. K. (2005). Life history affects the evolution of reproductive isolation among species of *Ceroisopsis* (Asteraceae). *Evolution* 59, 2362–2369. doi: 10.1111/j.0014-3820.2005.tb00946.x
- Arnold, B., Kim, S.-T., and Bomblies, K. (2015). Single geographic origin of a widespread autotetraploid *Arabidopsis arenosa* lineage followed by interploidy admixture. *Mol. Biol. Evol.* 32, 1382–1395. doi: 10.1093/molbev/msv089
- Arrigo, N., and Barker, M. S. (2012). Rarely successful polyploids and their legacy in plant genomes. *Curr. Opin. Plant Biol.* 15, 140–146. doi: 10.1016/j.pbi.2012.03.010
- Arvanitis, L., Wiklund, C., Münzbergova, Z., Dahlgren, J. P., and Ehrlén, J. (2010). Novel antagonistic interactions associated with plant polyploidization influence trait selection and habitat preference. *Ecol. Lett.* 13, 330–337. doi: 10.1111/j.1461-0248.2009.01429.x
- Ayala, F. J., and Coluzzi, M. (2005). Chromosome speciation: humans, *Drosophila*, and mosquitoes. *Proc. Natl. Acad. Sci. U.S.A.* 102 Suppl 1, 6535–6542. doi: 10.1073/pnas.0501847102
- Balao, F., Herrera, J., and Talavera, S. (2011). Phenotypic consequences of polyploidy and genome size at the microevolutionary scale: A multivariate morphological approach. *New Phytol.* 192, 256–265. doi: 10.1111/j.1469-8137.2011.03787.x
- Balao F., Valente L. M., Vargas P., Herrera J., Talavera S. (2010). Radiative evolution of polyploid races of the Iberian carnation *Dianthus broteri* (Caryophyllaceae). *New Phytol.* 187, 542–551. doi: 10.1111/j.1469-8137.2010.03280.x
- Barker M. S. (2013). “Karyotype and genome evolution in pteridophytes,” in *Plant genome diversity*, vol. 2. Eds. Greilhuber J., Doležal J., Wendel J. F. (Springer, Vienna, Austria), 245–253.
- Barker M. S., Arrigo N., Baniaga A. E., Zheng L., Levin D. A. (2016). On the relative abundance of autopolyploids and allopolyploids. *New Phytol.* 210, 391–398. doi: 10.1111/nph.13698
- Barker M. S., Jiao Y., Glennon, K. L. (2024). Doubling down on polyploid discoveries: Global advances in genomics and ecological impacts of polyploidy. *Am. J. Bot.* 111, e16395. doi: 10.1002/ajb2.16395
- Barrier M., Baldwin B. G., Robichaux R. H., Purugganan M. D. (1999). Interspecific hybrid ancestry of a plant adaptive radiation: allopolyploidy of the Hawaiian silversword alliance (Asteraceae) inferred from floral homeotic gene duplications. *Mol. Bio. Evol.* 16, 8, 1105–1113. doi: 10.1093/oxfordjournals.molbev.a026200
- Bartolić P., Morgan E. J., Padilla-García N., Kolář F. (2024). Ploidy as a leaky reproductive barrier: mechanisms, rates and evolutionary significance of interploidy gene flow. *Ann. Bot.* 134, 537–550. doi: 10.1093/aob/mcae096
- Basit A., Lim K.-B. (2024). Systematic approach of polyploidy as an evolutionary genetic and genomic phenomenon in horticultural crops. *Plant Sci.* 348, 112236. doi: 10.1016/j.plantsci.2024.112236
- Bell G. (1982). *The masterpiece of nature: The evolution and genetics of sexuality* (USA: University of California Press).
- Bellinger M. R., Datlof E. M., Selph K. E., Gallaher T. J., Knope M. L. (2022). A genome for *bidens hawaiiensis*: A member of a hexaploid hawaiian plant adaptive radiation. *J. Hered.* 113, 205–214. doi: 10.1093/jhered/esab077
- Benton M. J., Wilf P., Sauquet H. (2021). The angiosperm terrestrial revolution and the origins of modern biodiversity. *New Phytol.* 233, 2017–2035. doi: 10.1111/nph.17822
- Berdan E. L., Aubier T. G., Cozzolino S., Faria R., Feder J. L., Giménez M. D., et al. (2024). Structural variants and speciation: multiple processes at play. *Cold Spring Harb. Perspect. Biol.* 16, a041446. doi: 10.1101/cshperspect.a041446
- Birchler J. A., Veitia R. A. (2007). The gene balance hypothesis: from classical genetics to modern genomics. *Plant Cell* 19, 395–402. doi: 10.1105/tpc.106.049338
- Birchler J. A., Yao H., Chudalayandi S. (2007). Biological consequences of dosage dependent gene regulatory systems. *Biochim. Biophys. Acta Gene Struct. Expr.* 1769, 422–428. doi: 10.1016/j.bbaexp.2006.12.002
- Blaine Marchant D., Soltis D. E., Soltis P. S. (2016). Patterns of abiotic niche shifts in allopolyploids relative to their progenitors. *New Phytol.* 212, 708–718. doi: 10.1111/nph.14069
- Bohutinská M., Petříková E., Booker T. R., Cobo C. V., Vlček J., Šrámková G., et al. (2024). Polyploids broadly generate novel haplotypes from trans-specific variation in *Arabidopsis arenosa* and *Arabidopsis lyrata*. *PLoS Genet.* 20, e1011521. doi: 10.1371/journal.pgen.1011521
- Bomblies K., Jones G., Franklin C., Zickler D., Kleckner N. (2016). The challenge of evolving stable polyploidy: could an increase in ‘crossover interference distance’ play a central role? *Chromosoma* 125, 287–300. doi: 10.1007/s00412-015-0571-4
- Bomblies K., Madlung A. (2014). Polyploidy in the *arabidopsis* genus. *Chromosome Res.* 22, 117–134. doi: 10.1007/s10577-014-9416-x
- Boore J. L. (2006). The use of genome-level characters for phylogenetic reconstruction. *Trends Ecol. Evol.* 21, 439–446. doi: 10.1016/j.tree.2006.05.009
- Bretagnolle F., Thompson J. D. (1995). Tansley review no. 78. Gametes with the stomatic chromosome number: mechanisms of their formation and role in the evolution of autopolyploid plants. *New Phytol.* 129, 1–22. doi: 10.1111/j.1469-8137.1995.tb03005.x
- Brown M. R., Abbott R. J., Twyford A. D. (2024). The emerging importance of cross-ploidy hybridisation and introgression. *Mol. Ecol.* 33, e17315. doi: 10.1111/mec.17315
- Buerkle C. A., Morris R. J., Asmussen M. A., Rieseberg L. H. (2000). The likelihood of homoploid hybrid speciation. *Heredity* 84, 441–451. doi: 10.1046/j.1365-2540.2000.00680.x
- Burak M. K., Monk J. D., Schmitz O. J. (2018). Eco-evolutionary dynamics: the predator-prey adaptive play and the ecological theater. *Yale J. Biol. Med.* 91, 481–489.
- Burdon J. J., Marshall D. R. (1981). Inter- and intraspecific diversity in the disease response of *Glycine* species to the leaf-rust fungus *Phakopsora pachyrhizi*. *J. Ecol.* 69, 381–390. doi: 10.2307/2259674
- Burns R., Kulkarni A., Glushkevich A., Kolesnikova U. K., Kolář F., Scott A. D., et al. (2024). Diploid origins, adaptation to polyploidy, and the beginning of rediploidization in allotetraploid *Arabidopsis suecica*. *BioRxiv*. doi: 10.1101/2024.12.06.627142
- Burns R., Mandáková T., Gunis J., Soto-Jiménez L. M., Liu C., Lysak M. A., et al. (2021). Gradual evolution of allopolyploidy in *Arabidopsis suecica*. *Nat. Ecol. Evol.* 5, 1367–1381. doi: 10.1038/s41559-021-01525-w
- Carta A., Bedini G., Peruzzi L. (2020). A deep dive into the ancestral chromosome number and genome size of flowering plants. *New Phytol.* 228, 1097–1106. doi: 10.1111/nph.16668
- Carta A., Escudero M. (2023). Karyotypic diversity: a neglected trait to explain angiosperm diversification? *Evol.* 77, pp.1158–1164. doi: 10.1093/evolut/qpaa014
- Carta A., Peruzzi L. (2016). Testing the large genome constraint hypothesis: plant traits, habitat and climate seasonality in Liliaceae. *New Phytol.* 210, 709–716. doi: 10.1111/nph.13769
- Carvalho C. M. B., Lupski J. R. (2016). Mechanisms underlying structural variant formation in genomic disorders. *Nat. Rev. Genet.* 17, 224–238. doi: 10.1038/nrg.2015.25
- Čertner M., Fenclová E., Kúr P., Kolář F., Koutecký P., Krahulcová A., et al. (2017). Evolutionary dynamics of mixed-ploidy populations in an annual herb: dispersal, local persistence and recurrent origins of polyploids. *Ann. Bot.* 120, 303–315. doi: 10.1093/aob/mcx032
- Charron G., Marsit S., Hénault M., Martin H., Landry C. R. (2019). Spontaneous whole-genome duplication restores fertility in interspecific hybrids. *Nat. Commun.* 10, 4126. doi: 10.1038/s41467-019-12041-8
- Chase M. W., Samuel R., Leitch A. R., Guignard M. S., Conran J. G., Nollé F., et al. (2023). Down, then up: non-parallel genome size changes and a descending chromosome series in a recent radiation of the Australian allotetraploid plant species, *Nicotiana* section. *Suaevolentes* (*Solanaceae*). *Ann. Bot.* 131, 123–142. doi: 10.1093/aob/mcac006
- Chelaifa H., Monnier A., Ainouche M. (2010). Transcriptomic changes following recent natural hybridization and allopolyploidy in the salt marsh species *Spartina × townsendii* and *Spartina anglica* (Poaceae). *New Phytol.* 186, 161–174. doi: 10.1111/j.1469-8137.2010.03179.x
- Chen Z. J. (2007). Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. *Annu. Rev. Plant Biol.* 58, 377–406. doi: 10.1146/annurev.arplant.58.032806.103835
- Chen J., Zhong Y., Zou P., Ni J., Liu Y., Dai S., et al. (2024). Identification of Genomic Regions Associated with Differences in Flowering Time and Inflorescence Architecture between *Melastoma candidum* and *M. normale*. *Int. J. Mol. Sci.* 25, 10250. doi: 10.3390/ijms251910250

- Cheng J., Li J., Zhang Z., Lu H., Chen G., Yao B., et al. (2021). Autopolyploidy-driven range expansion of a temperate-originated plant to pan-tropic under global change. *Ecol. Monogr.* 91, e01445. doi: 10.1002/ecm.1445
- Cheng F., Wu J., Fang L., Sun S., Liu B., Lin K., et al. (2012). Biased gene fractionation and dominant gene expression among the subgenomes of *Brassica rapa*. *PLoS One* 7, e36442. doi: 10.1371/journal.pone.0036442
- Chester M., Gallagher J. P., Symonds V. V., Cruz da Silva A. V., Mavrodiev E. V., Leitch A. R., et al. (2012). Extensive chromosomal variation in a recently formed natural allopolyploid species. *Tragopogon miscellus* (Asteraceae). *PNAS* 109, 1176–1181. doi: 10.1073/pnas.1112041109
- Chester M., Leitch A. R., Soltis P. S., Soltis D. E. (2010). Review of the application of modern cytogenetic methods (FISH/GISH) to the study of reticulation (Polyploidy/hybridisation). *Genes* 1, 166–192. doi: 10.3390/genes1020166
- Chumová Z., Monier Z., Šemberová K., Havlíčková E., Euston-Brown D., Muasya A., et al. (2024). Diploid and tetraploid cytotypes of the flagship Cape species *Dicerothamnium rhinocerotis* (Asteraceae): variation in distribution, ecological niche, morphology and genetics. *Ann. Bot.* 133, 851–870. doi: 10.1093/aob/mcad084
- Clark J., Hidalgo O., Pellicer J., Liu H., Marquardt J., Robert Y., et al. (2016). Genome evolution of ferns: evidence for relative stasis of genome size across the fern phylogeny. *New Phytol.* 210, 1072–1082. doi: 10.1111/nph.13833
- Clarkson J. J., Knapp S., Garcia V., Olmstead R. G., Chase M. W. (2004). Phylogenetic relationships in Nicotiana (Solanaceae) inferred from multiple plastid DNA regions. *Mol. Phylogenet. Evol.* 33, 75–90. doi: 10.1016/j.ympev.2004.05.002
- Clo J., Kolář F. (2021). Short- and long-term consequences of genome doubling: a meta-analysis. *Am. J. Bot.* 108, 2315–2322. doi: 10.1002/ajb2.1759
- Cohen H., Fait A., Tel-Zur N. (2013). Morphological, cytological and metabolic consequences of autopolyploidization in *hylocereus* (Cactaceae) species. *BMC Plant Biol.* 13, 173. doi: 10.1186/1471-2229-13-173
- Comai L. (2005). The advantages and disadvantages of being polyploid. *Nat. Rev. Genet.* 6, 836–846. doi: 10.1038/nrg1711
- Costa L., Oliveira A., Carvalho-Sobrinho J., Souza G. (2017). Comparative cytological analyses reveal karyotype variability related to biogeographic and species richness patterns in Bombacoideae (Malvaceae). *Plant Syst. Evol.* 303, 1131–1144. doi: 10.1007/s00606-017-1427-6
- Coyne J. A., Orr H. A. (2004). *Speciation* (Sunderland: Sinauer).
- Cozzolino S., Scopece G. (2008). Specificity in pollination and consequences for postmating reproductive isolation in deceptive Mediterranean orchids. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 363, 3037–3046. doi: 10.1098/rstb.2008.0079
- Crespel L., Le Bras C., Relion D., Roman H., Morel P. (2015). Effect of high temperature on the production of 2n pollen grains in diploid roses and obtaining tetraploids via unilateral polyploidization. *Plant Breed.* 134, 356–364. doi: 10.1111/pbr.12271
- Da Silva C. R. M., Souza T. B., Trevisan R., González-Elizondo M. S., Torezan J. M. D., de Souza R. F., et al. (2017). Genome differentiation, natural hybridisation and taxonomic relationships among *Eleocharis viridans*, *E. niederleinii* and *E. ramboana* (Cyperaceae). *Austral Syst. Bot.* 30, 183–195. doi: 10.1071/SB17002
- Dixon J. R., Selvaraj S., Yue F., Kim A., Li Y., Shen Y., et al. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485, 376–380. doi: 10.1038/nature11082
- Dobzhansky T. (1951). *Genetics and the origin of species*. 3rd Ed (New York: Columbia University Press).
- Dobzhansky T., Sturtevant A. H. (1938). Inversions in the chromosomes of *Drosophila pseudoobscura*. *Genetics* 23, 28–64. doi: 10.1093/genetics/23.1.28
- Dodsworth S., Chase M. W., Leitch A. R. (2016). Is post-polyploidization diploidization the key to the evolutionary success of angiosperms? *Bot. J. Linn. Soc.* 180, 1–5. doi: 10.1111/boj.12357
- Edger P. P., Heidel-Fischer H. M., Bekaert M., Rota J., Glöckner G., Platts A. E., et al. (2015). The butterfly plant arms-race escalated by gene and genome duplications. *PNAS* 112, 8362–8366. doi: 10.1073/pnas.1503926112
- Edger P. P., Smith R., McKain M. R., Cooley A. M., Vallejo-Marin M., Yuan Y., et al. (2017). Subgenome dominance in an interspecific hybrid, synthetic allopolyploid, and a 140-year-old naturally established neo-allopolyploid monkeyflower. *Plant Cell* 29, 2150–2167. doi: 10.1105/tpc.17.00010
- Escudero M., Arroyo J., Sánchez-Ramírez S., Jordano P. (2024). Founder events and subsequent genetic bottlenecks underlie karyotype evolution in the Ibero-North African endemic *Carex helodes*. *Ann. Bot.* 133, 871–882. doi: 10.1093/aob/mcad087
- Escudero M., Hahn M., Brown B., Lueders K., Hipp A. (2016b). Chromosomal rearrangements in holocentric organisms lead to reproductive isolation by hybrid dysfunction: the correlation between karyotype rearrangements and germination rates in sedges. *Am. J. Bot.* 103, 1529–1536. doi: 10.3732/ajb.1600051
- Escudero M., Hahn M., Hipp A. L. (2018). RAD-seq linkage mapping and patterns of segregation distortion in sedges: meiosis as a driver of karyotypic evolution in organisms with holocentric chromosomes. *J. Evol. Biol.* 31, 833–843. doi: 10.1111/jeb.13267
- Escudero M., Hipp A. (2013). Shifts in diversification rates and clade ages explain species richness in higher-level sedge taxa (Cyperaceae). *Am. J. Bot.* 100, 2403–2411. doi: 10.3732/ajb.1300162
- Escudero M., Hipp A. L., Hansen T. F., Voje K. L., Luceño M. (2012). Selection and inertia in the evolution of holocentric chromosomes in sedges (Carex, Cyperaceae). *New Phytol.* 195, 237–247. doi: 10.1111/j.1469-8137.2012.04137.x
- Escudero M., Hipp A. L., Luceño M. (2010). Karyotype stability and predictors of chromosome number variation in sedges: A study in *Carex* section *Spirostachyae* (Cyperaceae). *Mol. Phylogenet. Evol.* 57, 353–363. doi: 10.1016/j.ympev.2010.07.009
- Escudero M., Maguilla E., Luceño M. (2013a). Selection by climatic regime and neutral evolutionary processes in holocentric chromosomes (*Carex* gr. *laevigata*: Cyperaceae): A microevolutionary approach. *Perspect. Plant Ecol. Evol. Syst.* 15, 118–129. doi: 10.1016/j.ppees.2013.01.001
- Escudero M., Márquez-Corro J. I., Hipp A. L. (2016a). The phylogenetic origins and evolutionary history of holocentric chromosomes. *Syst. Bot.* 41, 580–585. doi: 10.1600/036364416X692442
- Escudero M., Martín-Bravo S., Mayrose I., Fernández-Mazuecos M., Fiz-Palacios O., Hipp A. L., et al. (2014). Karyotypic Changes through Dysploidy Persist Longer over Evolutionary Time than Polyploid Changes. *PLoS One* 9, e85266. doi: 10.1371/journal.pone.0085266
- Escudero M., Weber J. A., Hipp A. L. (2013b). Species coherence in the face of karyotype diversification in holocentric organisms: the case of a cytogenetically variable sedge (*Carex scoparia*, Cyperaceae). *Ann. Bot.* 112, 515–526. doi: 10.1093/aob/mct119
- Escudero M., Wendel J. F. (2020). The grand sweep of chromosomal evolution in angiosperms. *New Phytol.* 228, 805–808. doi: 10.1111/nph.16802
- Estep M. C., McKain M. R., Diaz D. V., Zhong J. S., Hodge J. G., Hodkinson T. R., et al. (2014). Allopolyploidy, diversification, and the Miocene grassland expansion. *Proc. Natl. Acad. Sci. U.S.A.* 111, 15149–15154. doi: 10.1073/pnas.1404177111
- Ezoe A., Seki M. (2024). Exploring the complexity of genome size reduction in angiosperms. *Plant Mol. Biol.* 114, 121. doi: 10.1007/s11103-024-01518-w
- Farhat P., Hidalgo O., Robert T., Siljak-Yakovlev S., Leitch I. J., Adams R. P., et al. (2019). Polyploidy in the conifer genus *Juniperus*: An unexpectedly high rate. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00676
- Farhat P., Siljak-Yakovlev S., Hidalgo O., Rushforth K., Bartel J. A., Valentin N., et al. (2022). Polyploidy in Cupressaceae: Discovery of a new naturally occurring tetraploid, *Xanthocyparis vietnamensis*. *J. Syst. Evol.* 60, 824–834. doi: 10.1111/jse.12751
- Faria R., Navarro A. (2010). Chromosomal speciation revisited: rearranging theory with pieces of evidence. *Trends Ecol. Evol.* 25, 660–669. doi: 10.1016/j.tree.2010.07.008
- Faria R., Neto S., Noor M., Navarro A. (2011). *Role of Natural Selection in Chromosomal Speciation* in eLS, (Ed.). doi: 10.1002/9780470015902.a0022850
- Farminhão J. N. M., Verlynde S., Kaymak E., Droissart V., Simo-Droissart M., Collobert G., et al. (2021). Rapid radiation of angraeoids (Orchidaceae, Angraecinae) in tropical Africa characterised by multiple karyotypic shifts under major environmental instability. *Mol. Phylogenet. Evol.* 159, 107105. doi: 10.1016/j.ympev.2021.107105
- Fehrer J., Bertrand Y. J. K., Hartmann M., Caklová P., Josefiová J., Bräutigam S., et al. (2022). Multigene phylogeny of native american hawkweeds (*Hieracium* subgen. *Chionoracium*, cichorieae, asteraceae): origin, speciation patterns, and migration routes. *Plants (Basel)* 11, 2584. doi: 10.3390/plants11192584
- Fernández P., Hidalgo O., Juan A., Leitch I. J., Leitch A. R., Palazzesi L., et al. (2022). Genome Insights into Autopolyploid Evolution: A Case Study in *Senecio doronicum* (Asteraceae) from the Southern Alps. *Plants (Basel)* 11, 1235. doi: 10.3390/plants11091235
- Fishman L., Stathos A., Beardsley P. M., Williams C. F., Hill J. P. (2013). Chromosomal rearrangements and the genetics of reproductive barriers in *Mimulus* (monkey flowers). *Evol.* 67, 2547–2560. doi: 10.1111/evo.12154
- Fowler N. L., Levin D. A. (1984). Ecological constraints on the establishment of a novel polyploid in competition with its diploid progenitor. *Amer. Nat.* 124, 703–711. doi: 10.1086/284307
- Franco M. F., Colabelli M. N., Petigrosso L. R., De Battista J. P., Echeverría M. M. (2015). Evaluation of infection with endophytes in seeds of forage species with different levels of ploidy. *N. Z. J. Agric. Res.* 58, 181–189. doi: 10.1080/00288233.2015.1011283
- Freyman W. A., Höhna S. (2018). Cladogenetic and anagenetic models of chromosome number evolution: a Bayesian model averaging approach. *Syst. Biol.* 67, 195–215. doi: 10.1093/sysbio/syx065
- Giraud D., Lima O., Rousseau-Gueutin M., Salmon A., Ainouche M. (2021). Gene and transposable element expression evolution following recent and past polyploidy events in *spartina* (Poaceae). *Front. Genet.* 12. doi: 10.3389/fgene.2021.589160
- Glick L., Mayrose I. (2014). ChromEvol: assessing the pattern of chromosome number evolution and the inference of polyploidy along a phylogeny. *Mol. Biol. Evol.* 31, 1914–1922. doi: 10.1093/molbev/msu122
- Goel M., Schneeberger K. (2022). Plots: visualizing structural similarities and rearrangements between multiple genomes. *Bioinformatics* 38, 2922–2926. doi: 10.1093/bioinformatics/btac196
- Goel M., Sun H., Jiao W. B., Schneeberger K. (2019). SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. *Genome Biol.* 20, 277. doi: 10.1186/s13059-019-1911-0
- Goldblatt P., Johnson D. E. (2010). The Index to Plant Chromosome Numbers. *Ann. Missouri Bot. Gard.* 98, 226–227. doi: 10.3417/2011027

- Goldblatt P., Lowry P. P. (2011). The index to plant chromosome numbers (Ipcn): three decades of publication by the missouri botanical garden come to an end. *Ann. Missouri Botanical Garden* 98, 226–227. doi: 10.3417/2011027
- Grant V. (1958). “The regulation of recombination in plants,” in *Exchange of genetic material: mechanisms and consequences, cold spring harbor symposium on quantitative biology*, vol. 23. (Cold Spring Harbor, NY, USA), 337–363.
- Grant V. (1981). *Plant speciation*. 2nd edn (NY, USA: Columbia University Press).
- Griffiths A. G., Moraga R., Tausen M., Gupta V., Bilton T. P., Campbell, et al. (2019). Breaking free: the genomics of allopolyploidy-facilitated niche expansion in white clover. *Plant Cell* 31, 1466–1487. doi: 10.1105/tpc.18.00606
- Gross B. L., Rieseberg L. H. (2005). The ecological genetics of homoploid hybrid speciation. *J. Heredity* 96, 241–252. doi: 10.1093/jhered/esi026
- Guan J., Xu Y., Yu Y., Fu J., Ren F., Guo J., et al. (2021). Genome structure variation analyses of peach reveal population dynamics and a 1.67 Mb causal inversion for fruit shape. *Genome Biol.* 22, 13. doi: 10.1186/s13059-020-02239-1
- Guerra M. (2008). Chromosome numbers in plant cytotoxicity: concepts and implications. *Cytogenet. Genome Res.* 120, 339–350. doi: 10.1159/000121083
- Guerra M. (2016). Agmatoploidy and symploidy: a critical review. *Genet. Mol. Biol.* 39, 492–496. doi: 10.1590/1678-4685-GMB-2016-0103
- Guggisberg A., Mansion G., Kelso S., Conti E. (2006). Evolution of biogeographic patterns, ploidy levels, and breeding systems in a diploid–polyploid species complex of *Primula*. *New Phytol.* 171, 617–632. doi: 10.1111/j.1469-8137.2006.01722.x
- Guignard M. S., Nichols R. A., Knell R. J., Macdonald A., Romila C. A., Trimmer M., et al. (2016). Genome size and ploidy influence angiosperm species' biomass under nitrogen and phosphorus limitation. *New Phytol.* 210, 1195–1206. doi: 10.1111/nph.13881
- Gundel P. E., Dirihan S., Helander M., Zabalgoatzea I., Väre H., Saikkonen K. (2014). Systemic fungal endophytes and ploidy level in *Festuca vivipara* populations in North European Islands. *Pl. Syst. Evol.* 300, 1683–1691. doi: 10.1007/s00606-014-0994-z
- Haufler C. H. (1987). Electrophoresis is modifying our concepts of evolution in homosporous pteridophytes. *Am. J. Bot.* 74, 953–966. doi: 10.1002/j.1537-2197.1987.tb08700.x
- Haufler C. H., Soltis D. E. (1986). Genetic evidence suggests that homosporous ferns with high chromosome numbers are diploid. *Proc. Nat. Acad. Sci. U.S.A.* 83, 4389–4393. doi: 10.1073/pnas.83.12.4389
- Hegarty M. J., Hiscock S. J. (2008). Genomic clues to the evolutionary success of review polyploid plants. *Curr. Biol.* 18, R435–R444. doi: 10.1016/j.cub.2008.03.043
- Heilborn O. (1924). Chromosome numbers and dimensions, species formation and phylogeny in the genus *Carex*. *Hereditas* 5, 128–216. doi: 10.1111/j.1601-5223.1924.tb03128.x
- Hipp A. L., Rothrock P. E., Roalson E. H. (2009). The evolution of chromosome arrangements in *Carex* (Cyperaceae). *Bot. Rev.* 75, 96–109. doi: 10.1007/s12229-008-9022-8
- Hipp A. L., Rothrock P. E., Whitkus R., Weber J. A. (2010). Chromosomes tell half of the story: the correlation between karyotype rearrangements and genetic diversity in sedges, a group with holocentric chromosomes. *Mol. Evol.* 19, 3124–3138. doi: 10.1111/j.1365-294X.2010.04741.x
- Hoencamp C., Dudchenko O., Elbatsh A. M., Brahmachari S., Raaijmakers J. A., van Schaik T., et al. (2021). 3D genomics across the tree of life reveals condensin II as a determinant of architecture type. *Science* 372, 984–989. doi: 10.1126/science.abe2218
- Hoffmann A. A., Rieseberg L. H. (2008). Revisiting the impact of inversions in evolution: from population genetic markers to drivers of adaptive shifts and speciation? *Annu. Rev. Ecol. Syst.* 39, 21–42. doi: 10.1146/annurev.ecolsys.39.110707.173532
- Hou C., Li L., Qin Z. S., Corces V. G. (2012). Gene density, transcription, and insulators contribute to the partition of the *Drosophila* genome into physical domains. *Mol. Cell* 48, 471–484. doi: 10.1016/j.molcel.2012.08.031
- Huang Y., Guo X., Zhang K., Mandáková T., Cheng F., Lysak M. A. (2023). The meso-octoploid *Heliophila variabilis* genome sheds a new light on the impact of polyploidization and diploidization on the diversity of the Cape flora. *Plant J.* 116, 446, 466. doi: 10.1111/tpj.16383
- Huang K., Rieseberg L. H. (2020). Frequency, origins, and evolutionary role of chromosomal inversions in plants. *Front. P. Sci.* 11. doi: 10.3389/fpls.2020.00296
- Huetzel B., Kreil D. P., Matzke M., Matzke A. J. (2008). Effects of aneuploidy on genome structure, expression, and interphase organization in *Arabidopsis thaliana*. *PLoS Genet.* 4, e1000226. doi: 10.1371/journal.pgen.1000226
- Jayakodi M., Lu Q., Pidon H., Rabanus-Wallace M. T., Bayer M., Lux T., et al. (2024). Structural variation in the pangenome of wild and domesticated barley. *Nature* 636, 654–662. doi: 10.1038/s41586-024-08187-1
- Jersáková J., Castro S., Sonk N., Milchreit K., Schödelbauerová I., Tolasch T., et al. (2010). Absence of pollinator-mediated premating barriers in mixed-ploidy populations of *Gymnadenia conopsea* s.l. (Orchidaceae). *Evol. Ecol.* 24, 1199–1218. doi: 10.1007/s10682-010-9356-7
- Jiang J. (2019). Fluorescence in situ hybridization in plants: recent developments and future applications. *Chromosome Res.* 27, 153–165. doi: 10.1007/s10577-019-09607-z
- Jiao Y., Leebens-Mack J., Ayyampalayam S., Bowers J. E., McKain M. R., McNeal J., et al. (2012). A genome triplication associated with early diversification of the core eudicots. *Genome Biol.* 26, R3. doi: 10.1186/gb-2012-13-1-r3
- Jiao Y., Wickett N., 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
- Jin S., Han Z., Hu Y., Si Z., Dai F., He L., et al. (2023). Structural variation (SV)-based pan-genome and GWAS Reveal the Impacts of SVs on the speciation and diversification of allotetraploid cottons. *Mol. Plant* 16, 678–693. doi: 10.1016/j.molp.2023.02.004
- Johnen L., de Souza T. B., Rocha D. M., Parteka L. M., González-Elizondo M. S., Trevisan R., et al. (2020). Allopolyploidy and genomic differentiation in holocentric species of the *Eleocharis montana* complex (Cyperaceae). *Plant Syst. Evol.* 306, 39. doi: 10.1007/s00606-020-01666-8
- Johnson M. T. J., Smith S. D., Rausher M. D. (2009). Plant sex and the evolution of plant defenses against herbivores. *Proc. Natl. Acad. Sci. U.S.A.* 106, 18079–18084. doi: 10.1073/pnas.0904695106
- Jørgensen M. H., Ehrich D., Schmickl R., Koch M. A., Brysting A. K. (2011). Interspecific and interploidal gene flow in Central European *Arabidopsis* (Brassicaceae). *BMC Evol. Biol.* 11, 346. doi: 10.1186/1471-2148-11-346
- Julca I., Marcet-Houben M., Vargas P., Gabaldón T. (2018). Phylogenomics of the olive tree (*Olea europaea*) reveals the relative contribution of ancient allo- and autopolyploidization events. *BMC Biol.* 16, 15. doi: 10.1186/s12915-018-0482-y
- Kantor A., Kučera J., Šlenker M., Breidly J., Dönmez A. A., Marhold K., et al. (2023). Evolution of hygrophytic plant species in the Anatolia–Caucasus region: insights from phylogenomic analyses of *Cardamine* perennials. *Ann. Bot.* 131, 585–600. doi: 10.1093/aob/mcad008
- Kauai F., Bafort Q., Mortier F., Montagu M. V., Bonte D., Van de Peer Y. (2024). Interspecific transfer of genetic information through polyploid bridges. *PNAS* 121, (21), e2400018121. doi: 10.1073/pnas.2400018121
- Kearns A. M., Restani M., Szabo I., Schröder-Nielsen A., Kim J. A., Richardson H. M., et al. (2018). Genomic evidence of speciation reversal in ravens. *Nat. Commun.* 9, 906. doi: 10.1038/s41467-018-03294-w
- Kelly L. J., Leitch A. R., Clarkson J. J., Knapp S., Chase M. W. (2013). Reconstructing the complex evolutionary origin of wild allopolyploid tobaccos (*Nicotiana* sect. *Suaevolentes*). *Evol.* 76, 80–94. doi: 10.1111/j.1558-5646.2012.01748.x
- Kinoshian S., Barker M. S. (2024). Meiotic drive and genome evolution in vascular land plants. *EcoevoRxiv*. doi: 10.32942/X2RD0K
- Kirkpatrick M., Barton N. (2006). Chromosome inversions, local adaptation and speciation. *Genetics* 173, 419–434. doi: 10.1534/genetics.105.047985
- Klekowski E. J. Jr., Baker H. G. (1966). Evolutionary significance of polyploidy in the pteridophyta. *Sci.* 153, 305–307. doi: 10.1126/science.153.3733.305
- Knight C. A., Molinari N., Petrov D. A. (2005). The large genome constraint hypothesis: evolution, ecology and phenotype. *Ann. Bot.* 95, 177–190. doi: 10.1093/aob/mci011
- Koch M. A. (2019). The plant model system *Arabidopsis* set in an evolutionary, systematic, and spatio-temporal context. *J. Exp. Bot.* 70, 55–67. doi: 10.1093/jxb/ery340
- Koella J. C. (1993). Ecological correlates of chiasma frequency and recombination index of plants. *Biol. J. Linn. Soc.* 48, 227–238. doi: 10.1111/j.1095-8312.1993.tb00889.x
- Köhler C., Scheid O. M., Erilova A. (2010). The impact of the triploid block on the origin and evolution of polyploid plants. *Trends Genet.* 26, 142–148. doi: 10.1016/j.tig.2009.12.006
- Kolář F., Čertner M., Suda J., Šchönschetter P., Husband B. C. (2017). Mixed-ploidy species: progress and opportunities in polyploid research. *Trends Plant Sci.* 22, 1041–1055. doi: 10.1016/j.tplants.2017.09.011
- Kolář F., Lučanová M., Závěská E., Fuxová G., Mandáková T., Španiel S., et al. (2016). Ecological segregation does not drive the intricate parapatric distribution of diploid and tetraploid cytotypes of the *Arabidopsis arenosa* group (Brassicaceae). *Biol. J. Linn. Soc. Lond.* 119, 673–688. doi: 10.1111/bij.12479
- Kolesnikova U. K., Scott A. D., Van de Velde J. D., Burns R., Tikhomirov N. P., Pfordt U., et al. (2023). Transition to Self-compatibility Associated With Dominant S-allele in a Diploid Siberian Progenitor of Allotetraploid *Arabidopsis kamchatica* Revealed by *Arabidopsis lyrata* Genomes. *Mol. Biol. Evol.* 40, msad122. doi: 10.1093/molbev/msad122
- Kollar L. M., Stanley L. E., Raju S. K. K., Lowry D. B., Niederhuth C. E. (2024). The evolutionary dynamics of locally adaptive chromosome inversions in *Mimulus guttatus*. *bioRxiv*. doi: 10.1101/2023.12.06.570460
- Krak K., Čaklová P., Chrtek J., Fehrer J. (2013). Reconstruction of phylogenetic relationships in a highly reticulate group with deep coalescence and recent speciation (Hieracium, Asteraceae). *Hered.* 110, 138–151. doi: 10.1038/hdy.2012.100
- Krakos K. N., Johnson M. G., Hoch P. C., Wagner W. L., Huang P., Raven P. H. (2022). Molecular phylogenetics reveals multiple transitions to self-compatibility in a primary subclade of *oenothera* (Onagraceae). *Ann. Missouri Bot. Gard* 107, 405–421. doi: 10.3417/2022735
- Kumar S., Kaur S., Seem K., Kumar S., Mohapatra T. (2021). Understanding 3D genome organization and its effect on transcriptional gene regulation under environmental stress in plant: A chromatin perspective. *Front. Cell Dev. Biol.* 9. doi: 10.3389/fcell.2021.774719
- Lafon-Placette C., Johannessen I. M., Hornslien K. S., Ali M. F., Bjerkan K. N., Bramsiepe J., et al. (2017). Endosperm-based hybridization barriers explain the pattern

of gene flow between *Arabidopsis lyrata* and *Arabidopsis arenosa* in Central Europe. *PNAS* 114, E1027–E1035. doi: 10.1073/pnas.1615123114

Landis J. B., Soltis D. E., Li Z., Marx H. E., Barker M. S., Tank D. C., et al. (2018). Impact of whole-genome duplication events on diversification rates in angiosperms. *Am. J. Bot.* 105, 348–363. doi: 10.1002/ajb2.1060

Larridon I., Zuntini A. R., Léveillé-Bourret É., Barrett R. L., Starr J. R., Muasya A. M., et al. (2021). A new classification of Cyperaceae (Poales) supported by phylogenomic data. *J. Syst. Evol.* 59, 852–895. doi: 10.1111/jse.12757

Lee Y. W., Fishman L., Kelly J. K., Willis J. H. (2016). A segregating inversion generates fitness variation in yellow monkeyflower (*Mimulus guttatus*). *Genetics* 202, 4, 1, 1473–1484. doi: 10.1534/genetics.115.183566

Leitch I. J., Bennett M. D. (1997). Polyploidy in angiosperms. *Trends Plant Sci.* 12, 470–476. doi: 10.1016/S1360-1385(97)01154-0

Leitch I. J., Hanson L. (2002). DNA C-values in seven families fill phylogenetic gaps in the basal angiosperms. *Bot. J. Linn.* 140, 175–179. doi: 10.1046/j.1095-8339.2002.00096.x

Leitch A., Leitch I. J. (2008). Genomic plasticity and the diversity of polyploid plants. *Science* 320, 48–483. doi: 10.1126/science.1153585

Leitch A. R., Leitch I. J. (2012). Ecological and genetic factors linked to contrasting genome dynamics in seed plants. *New Phytol.* 194, 629–646. doi: 10.1111/j.1469-8137.2012.04105.x

Leitch I. J., Leitch A. R. (2013). “Genome size diversity and evolution in land plants,” in *Plant genome diversity, vol. 2, physical structure, behaviour and evolution of plant genomes*. Eds. Leitch I. J., Greilhuber J., Doležel J., Wendel J. F. (Springer-Verlag, Wien, Austria), 307–322.

Le Scouarnec S., Gribble S. (2012). Characterising chromosome rearrangements: recent technical advances in molecular cytogenetics. *Heredity* 108, 75–85. doi: 10.1038/hdy.2011.100

Levin D. A. (2002). *The role of chromosomal change in plant evolution* (New York, NY, USA: OUP).

Li Z., Kinosis S. P., Zhan S. H., Barker M. S. (2024). Ancient polyploidy and low rate of chromosome loss explain the high chromosome numbers of homosporous ferns. *bioRxiv*. doi: 10.1101/2024.09.23.614530

Li Z., McKibben M. T. W., Finch G. S., Blischak P. D., Sutherland B. L., Barker M. S. (2021). Patterns and processes of diploidization in land plants. *Annu. Rev. Plant Biol.* 72, 387–410. doi: 10.1146/annurev-arplant-050718-100344

Li H., Wang S., Chai S., Yang Z., Zhang Q., Xin H., et al. (2022). Graph-based pan-genome reveals structural and sequence variations related to agronomic traits and domestication in cucumber. *Nat. Commun.* 13, 682. doi: 10.1038/s41467-022-28362-0

Li X., Wang J., Yu Y., Li G., Wang J., Li C., et al. (2023). Genomic rearrangements and evolutionary changes in 3D chromatin topologies in the cotton tribe (Gossypieae). *BMC Biol.* 21, 56. doi: 10.1186/s12915-023-01560-y

Lin Y., Rajan V., Moret B. M. (2012). Bootstrapping phylogenies inferred from rearrangement data. *Algorithms Mol. Biol.* 7, 21. doi: 10.1186/1748-7188-7-21

Liu Y., Li D., Zhang Q., Song C., Zhong C., Zhang X., et al. (2017). Rapid radiations of both kiwifruit hybrid lineages and their parents shed light on a two-layer mode of species diversification. *New Phytol.* 215, 877–890. doi: 10.1111/nph.14607

López-González N., Bobo-Pinilla J., Padilla-García N., Loureiro J., Castro S., Rojas-Andrés B. M., et al. (2021). Genetic similarities versus morphological resemblance: unraveling a polyploid complex in a Mediterranean biodiversity hotspot. *Mol. Phylogenet. Evol.* 155, 107006. doi: 10.1016/j.ympev.2020.107006

Lowry D. B., Willis J. H. (2010). A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation. *PLoS Biol.* 8, e1000500. doi: 10.1371/journal.pbio.1000500

Lucek K., Augustijn H., Escudero M. (2022). A holocentric twist to chromosomal speciation? *Trends Ecol. Evol.* 37, 655–662. doi: 10.1016/j.tree.2022.04.002

Lucek K., Giménez M. D., Joron M., Rafajlovic M., Searle J. B., Walden N., et al. (2023). The impact of chromosomal rearrangements in Speciation: From Micro to Macroevolution. *Cold Spring Harb. Perspect. Biol.* 15, a041447. doi: 10.1101/cshperspect.a041447

Luceño M., Castroviejo S. (1991). Agmatoploidy in *Carex laevigata* (Cyperaceae). Fusion and fission of chromosomes as the mechanism of cytogenetic evolution in Iberian populations. *Pl. Syst. Evol.* 177, 149–159. doi: 10.1007/BF00937952

Luceño M., Guerra M. (1996). Numerical variations in species exhibiting holocentric chromosomes: a nomenclatural proposal. *Caryologia* 49, 301–309. doi: 10.1080/00087114.1996.10797374

Luo M. C., Deal K. R., Akhunov E. D., Akhunova A. R., Anderson O. D., Anderson J. A., et al. (2009). Genome comparisons reveal a dominant mechanism of chromosome number reduction in grasses and accelerated genome evolution in Triticeae. *Proc. Natl. Acad. Sci. U.S.A.* 106, 15780–15785. doi: 10.1073/pnas.0908195106

Lysak M. A., Berr A., Pecinka A., Schmidt R., McBreen K., Schubert I. (2006). Mechanisms of chromosome number reduction in *Arabidopsis thaliana* and related Brassicaceae species. *Proc. Natl. Acad. Sci. U.S.A.* 103, 5224–5229. doi: 10.1073/pnas.0510791103

Lysak M. A., Schubert I. (2012). “Mechanisms of chromosome rearrangements,” in *Plant Genome Diversity, vol. 2*. Eds. Greilhuber J., Doležel J., Wendel J. (Springer, Vienna). doi: 10.1007/978-3-7091-1160-4_9

Mauguilla E., Escudero M., Ruiz-Martín J., Arroyo J. (2021). Origin and diversification of flax and their relationship with heterostyly across the range. *J. Biogeog.* 48, 1994–2007. doi: 10.1111/jbi.14129

Maheshwari S., Barbash D. A. (2011). The genetics of hybrid incompatibilities. *Annu. Rev. Genet.* 45, 331–355. doi: 10.1146/annurev-genet-110410-132514

Mallet J. (2007). Hybrid speciation. *Nature* 446, 279–283. doi: 10.1038/nature05706

Mandáková T., Gloss A. D., Whiteman N. K., Lysak M. A. (2016). How diploidization turned a tetraploid into a pseudotriploid. *Am. J. Bot.* 103, 1187–1196. doi: 10.3732/ajb.1500452

Mandáková T., Guo X., Özudoğlu B., Mummenhoff K., Lysak M. A. (2018). Hybridization-facilitated genome merger and repeated chromosome fusion after 8 million years. *TPJ* 96, 748–760. doi: 10.1111/tpj.14065

Mandáková T., Heenan P. B., Lysak M. A. (2010a). Island species radiation and karyotypic stasis in Pachycladon allopolyploids. *BMC Evol. Biol.* 10, 367. doi: 10.1186/1471-2148-10-367

Mandáková T., Joly S., Krzywinski M., Mummenhoff K., Lysak M. A. (2010b). Fast diploidization in close mesopolyploid relatives of *arabidopsis*. *Plant Cell* 22, 2277–2290. doi: 10.1105/tpc.110.074526

Mandáková T., Lysak M. A. (2008). Chromosomal phylogeny and karyotype evolution in x=7 crucifer species (Brassicaceae). *Plant Cell* 20, 2559–2570. doi: 10.1105/tpc.108.062166

Mandáková T., Lysak M. A. (2018). Post-polyploid diploidization and diversification through dysploid changes. *Plant Biol.* 42, 55–65. doi: 10.1016/j.pbi.2018.03.001

Mandáková T., Pouch M., Harmanová K., Zhan S. H., Mayrose I., Lysak M. A. (2017). Multispeed genome diploidization and diversification following ancient allopolyploidization. *Mol. Ecol.* 26, 6445–6462. doi: 10.1111/mec.14379

Marburger S., Monahan P., Seear P. J., Martin S. H., Koch J., Pajananen P., et al. (2019). Interspecific introgression mediates adaptation to whole genome duplication. *Nat. Commun.* 10, 1–11. doi: 10.1038/s41467-019-13159-5

Marks R. A., Delgado P., Makonya G. M., Cooper K., VanBuren R., Farrant J. M. (2024). Higher order polyploids exhibit enhanced desiccation tolerance in the grass *Microchloa caffra*. *J. Exp. Bot.* 75, 3612–3623. doi: 10.1093/jxb/erae126

Márquez-Corro J. I., Escudero M., Luceño M. (2018). Do holocentric chromosomes represent an evolutionary advantage? A study of paired analyses of diversification rates of lineages with holocentric chromosomes and their monocentric closest relatives. *Chromosome Res.* 26, 139–152. doi: 10.1007/s10577-017-9566-8

Márquez-Corro J. I., Martín-Bravo S., Blanco-Pastor J. L., Luceño M., and Escudero M. (2024). The holocentric chromosome microevolution: From phylogeographic patterns to genomic associations with environmental gradients. *Mol. Ecol.* 33, e17156. doi: 10.1111/mec.17156

Márquez-Corro J. I., Martín-Bravo S., Jiménez-Mejías P., Hipp A. L., Spalink D., Naczi R. F., et al. (2021). Macroevolutionary insights into sedges (*Carex*: Cyperaceae): The effects of rapid chromosome number evolution on 490 lineage diversification. *J. Syst. Evol.* 59, 776–790. doi: 10.1111/jse.12730

Márquez-Corro J. I., Martín-Bravo S., Pedrosa-Harand A., Hipp A. L., Luceño M., Escudero M. (2019). *Karyotype Evolution In Holocentric Organisms* (Chichester: eLS. John Wiley and Sons, Ltd). doi: 10.1002/9780470015902.a0028758

Mata J. K., Martin S. L., Smith T. W. (2023). Global biodiversity data suggest allopolyploid plants do not occupy larger ranges or harsher conditions compared with their progenitors. *Ecol. Evol.* 13, e10231. doi: 10.1002/eece3.10231

Mayrose I., Barker M. S., Otto S. P. (2010). Probabilistic models of chromosome number evolution and the inference of polyploidy. *Syst. Biol.* 59, 132–144. doi: 10.1093/sysbio/syp083

Mayrose I., Lysak M. A. (2021). The evolution of chromosome numbers: mechanistic models and experimental approaches. *Genome Biol. Evol.* 13, evaa220. doi: 10.1093/gbe/evaa220

Mayrose I., Zhan S. H., Rothfels C. J., Arrigo N., Barker M. S., Rieseberg L. H., et al. (2014). Methods for studying polyploid diversification and the dead end hypothesis: A reply to Soltis et al. *New Phytol.* 206, 27–35. doi: 10.1111/nph.13192

McCarthy E. W., Arnold S. E., Chittka L., Le Comber S. C., Verity R., Dodsworth S., et al. (2015). The effect of polyploidy and hybridization on the evolution of floral colour in Nicotiana (Solanaceae). *Ann. Bot.* 115, 1117–1131. doi: 10.1093/aob/mcv048

McHale N. A. (1983). Environmental induction of high frequency 2n pollen formation in diploid Solanum. *Can. J. Genet. Cytol.* 25, 609–615. doi: 10.1139/g83-091

McKibben M. T. W., Finch G., Barker M. S. (2024). Species-tree topology impacts the inference of ancient whole-genome duplications across the angiosperm phylogeny. *Am. J. Bot.* 111, e16378. doi: 10.1002/ajb2.16378

Meier J. L., Marques D. A., Mwaiko S., Wagner C. E., Excoffier L., Seehausen O. (2017). Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nat. Commun.* 8, 14363. doi: 10.1038/ncomms14363

Michael T. P. (2014). Plant genome size variation: bloating and purging DNA. *Brief Funct. Genomics* 13, 308–317. doi: 10.1093/bfpg/elu005

Ming R., Bendahmane A., Renner S. S. (2011). Sex chromosomes in land plants. *Annu. Rev. Plant Biol.* 62, 485–514. doi: 10.1146/annurev-arplant-042110-103914

Mohan A. V., Escuer P., Cornet C., Lucek K. (2024). A three-dimensional genomics view for speciation research. *Trends Genet.* 40, 638–641. doi: 10.1016/j.tig.2024.05.009

- Monnahan P., Kolář F., Baduel P., Sailer C., Koch J., Horvath R., et al. (2019). Pervasive population genomic consequences of genome duplication in *Arabidopsis arenosa*. *Nat. Ecol. Evol.* 3, 457. doi: 10.1038/s41559-019-0807-4
- Moret B., and Warnow T. (2005). "Advances in phylogeny reconstruction from gene order and content data," in *Methods In Enzymology, Molecular Evolution: Producing The Biochemical Data, Part B*. Vol. 395. Ed. Roalson Z. (Amsterdam: Elsevier), 673–700.
- Morgan E. J., Čertner M., Lučanová M., Deniz U., Kubíková K., Venon A., et al. (2021). Disentangling the components of triploid block and its fitness consequences in natural diploid–tetraploid contact zones of *Arabidopsis arenosa*. *New Phytol.* 232, 1449–1462. doi: 10.1111/nph.17357
- Moura R. F., Queiroga D., Vilela E., Moraes A. P. (2021). Polyploidy and high environmental tolerance increase the invasive success of plants. *J. Plant Res.* 134, 105–114. doi: 10.1007/s10265-020-01236-6
- Mutti J. S., Bhullar R. K., Gill K. S. (2017). Evolution of gene expression balance among homeologs of natural polyploids. *G3: Genes Genomes Genet.* 7, 1225–1237. doi: 10.1534/g3.116.038711
- Naiki A. (2012). Heterostyly and the possibility of its breakdown by polyploidization. *Plant Species Biol.* 27, 3–29. doi: 10.1111/j.1442-1984.2011.00363.x
- Nieto Feliner G., Casacuberta J., Wendel F. J. (2020). Genomics of evolutionary novelty in hybrids and polyploids. *Front. Genet.* 11. doi: 10.3389/fgene.2020.00792
- Nijalingappa B. H. M. (1974). Cytological studies in *Fimbristylis paupercula* Boeck. (Cyperaceae). *Caryologia* 27, 379–384.
- Nokkala S., Kuznetsova V. G., Maryanska-Nadachowska A., Nokkala C. (2004). Holocentric chromosomes in meiosis. I. Restriction of the number of chiasmata in bivalents. *Chromosome Res.* 12, 733–739. doi: 10.1023/B:CHRO.0000045797.74375.70
- Novikova P. Y., Hohmann N., Van de Peer Y. (2018). Polyploid *Arabidopsis* species originated around recent glaciation maxima. *Curr. Opin. Plant Biol.* 42, 8–15. doi: 10.1016/j.pbi.2018.01.005
- Novikova P. Y., Tsuchimatsu T., Simon S., Nizhynska V., Voronin V., Burns R., et al. (2017). Genome sequencing reveals the origin of the allotetraploid *Arabidopsis suecica*. *Mol. Biol. Evol.* 34, 957–968. doi: 10.1093/molbev/msw299
- Novikova P., Yu, Kolesnikova U. K., Scott A. D. (2023). Ancestral self-compatibility facilitates the establishment of allopolyploids in Brassicaceae. *Plant Reprod.* 36, 125–138. doi: 10.1007/s00497-022-00451-6
- Nuismer S. L., Thompson J. N. (2001). Plant polyploidy and non-uniform effects on insect herbivores. *Proc. R. Soc. Lond. B.* 268, 1937–1940. doi: 10.1098/rspb.2001.1760
- Orians C. M. (2000). The effects of hybridization in plants on secondary chemistry: implications for the ecology and evolution of plant–herbivore interactions. *Am. J. Bot.* 87, 1749–1756. doi: 10.2307/2656824
- Osborn T. C., Pires J. C., Birchler J. A., Auger D. L., Chen Z. J., Lee H. S., et al. (2003). Understanding mechanisms of novel gene expression in polyploids. *Trends Genet.* 19, 141–147. doi: 10.1016/S0168-9525(03)00015-5
- Ostevik K. L., Samuk K., Rieseberg L. H. (2020). Ancestral reconstruction of karyotypes reveals an exceptional rate of nonrandom chromosomal evolution in sunflowers. *Genetics* 214, 1031–1045. doi: 10.1534/genetics.120.303026
- Oswald B. P., Nuismer S. L. (2011). A unified model of autopolyploid establishment and evolution. *Am. Nat.* 178, 687–700. doi: 10.1086/662673
- Otto S. P. (2007). The evolutionary consequences of polyploidy. *Cell* 131, 452–462. doi: 10.1016/j.cell.2007.10.022
- Otto S. P., Whitton J. (2000). Polyploid incidence and evolution. *Annu. Rev. Genet.* 34, 401–437. doi: 10.1146/annurev.genet.34.1.401
- Ouyang W., Xiong D., Li G., Li X. (2020). Unraveling the 3D genome architecture in plants: present and future. *Mol. Plant* 13, 1676–1693. doi: 10.1016/j.molp.2020.10.002
- Owens G. L., Rieseberg L. H. (2014). Hybrid incompatibility is acquired faster in annual than in perennial species of sunflower and tarweed. *Evol.* 68, 893–900. doi: 10.1111/evo.12297
- Paape T., Briskine R. V., Halstead-Nussloch G., Lischer H. E. L., Shimizu-Inatsugi R., Hatakeyama M., et al. (2018). Patterns of polymorphism and selection in the subgenomes of the allopolyploid *Arabidopsis kamchatica*. *Nat. Commun.* 9, 3909. doi: 10.1038/s41467-018-06108-1
- Pandit M. K., Pocock M. J. O., Kunin W. E. (2011). Ploidy influences rarity and invasiveness in plants. *J. Ecol.* 99, 1108–1115. doi: 10.1111/j.1365-2745.2011.01838.x
- Parent C., Patton A., Pfennig K., Rubinoff D., Schluter D., Seehausen O., et al. (2020). Comparing adaptive radiations across space, time, and taxa. *J. Hered.* 111, 1–20. doi: 10.1093/jhered/esz064
- Parisod C., Broennimann O. (2016). Towards unified hypotheses of the impact of polyploidy on ecological niches. *New Phytol.* 212, 540–542. doi: 10.1111/nph.14133
- Parisod C., Holderegger R., Brochmann C. (2010). Evolutionary consequences of autopolyploidy. *New Phytol.* 186, 5–17. doi: 10.1111/j.1469-8137.2009.03142.x
- Paterson A. H., Wang X., Li J., Tang H. (2012). "Ancient and recent polyploidy in monocots," in *Polyploidy and genome evolution*. Eds. Soltis P. S., Soltis D. E. (Springer, Berlin, Germany), 93–108.
- Patterson T. B., Givnish T. J. (2004). Geographic cohesion, chromosomal evolution, parallel adaptive radiations, and consequent floral adaptations in *Calochortus* (Calochortaceae): evidence from a cpDNA phylogeny. *New Phytol.* 161, 253–264. doi: 10.1046/j.1469-8137.2003.00951.x
- Pease J. B., Haak D. C., Hahn M. W., Moyle L. C. (2016). Phylogenomics reveals three sources of adaptive variation during a rapid radiation. *PLoS Biol.* 14, e1002379. doi: 10.1371/journal.pbio.1002379
- Pécir Y., Rallo G., Folzer H., Cigna M., Gudin S., Bris M. L. (2011). Polyploidization mechanisms: temperature environment can induce diploid gamete formation in *Rosa* sp. *J. Exp. Bot.* 62, 3587–3597. doi: 10.1093/jxb/err052
- Pegoraro L., Vos J. M., Cozzolino S., Scopece G. (2019). Shift in flowering time allows diploid and autotetraploid *Anacamptis pyramidalis* (Orchidaceae) to coexist by reducing competition for pollinators. *Bot. J. Linn. Soc.* 191, 274–284. doi: 10.1093/botlinnean/boz043
- Pei L., Li G., Lindsey K., Zhang X., Wang M. (2021). Plant 3D genomics: the exploration and application of chromatin organization. *New Phytol.* 230, 1772–1786. doi: 10.1111/nph.17262
- Pellicer J., Hidalgo O., Dodsworth S., Leitch I. J. (2018). Genome size diversity and its impact on the evolution of land plants. *Genes* 9, 88. doi: 10.3390/genes9020088
- Peskoller A., Silbernagl L., Hülber K., Sonnleitner M., and Schönswetter P. (2021). Do pentaploid hybrids mediate gene flow between tetraploid *Senecio disjunctus* and hexaploid *S. carniolicus* s. str. (*S. carniolicus* aggregate, Asteraceae)? *Alp. Bot.* 131, 151–160. doi: 10.1007/s00035-021-00254-x
- Pikaard C. S. (2001). Genomic change and gene silencing in polyploids. *Trend Genet.* 17, 675–677. doi: 10.1016/S0168-9525(01)02545-8
- Pimentel M., Escudero M., Sahuquillo E., Minaya M.Á., Catalán P. (2017). Are diversification rates and chromosome evolution in the temperate grasses (Pooideae) associated with major environmental changes in the Oligocene-Miocene? *PeerJ.* 5, e3815. doi: 10.7717/peerj.3815
- Qian H., Jin Y. (2016). An updated megaphylogeny of plants, a tool for generating plant phylogenies and an analysis of phylogenetic community structure. *J. Plant Ecol.* 9, 233–239. doi: 10.1093/jpe/rtv047
- Qiao X., Li Q., Yin H., Qi K., Li L., Wang R., et al. (2019). Gene duplication and evolution in recurring polyploidization–diploidization cycles in plants. *Genome Biol.* 20, 38. doi: 10.1186/s13059-019-1650-2
- Qiu T., Liu Z., Liu B. (2020). The effects of hybridization and genome doubling in plant evolution via allopolyploidy. *Mol. Biol. Rep.* 47, 5549–5558. doi: 10.1007/s11033-020-05597-y
- Ramsey J., Schemske D. W. (1998). Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annu. Rev. Ecol. Syst.* 29, 467–501. doi: 10.1146/annurev.ecolsys.29.1.467
- Ravinet M., Faria R., Butlin R. K., Galindo J., Bierre N., Rafajlović M., et al. (2017). Interpreting the genomic landscape of speciation: A road map for finding barriers to gene flow. *J. Evol. Biol.* 30, 1450–1477. doi: 10.1111/jeb.13047
- Rezende L., Suzigan J., Amorim F. W., Moraes A. P. (2020). Can plant hybridization and polyploidy lead to pollinator shift? *Acta Bot. Bras.* 34, 229–242. doi: 10.1590/0102-33062020abb0025
- Rice A., Glick L., Abadi S., Einhorn M., Kopelman N. M., Salman-Minkov A., et al. (2015). The Chromosome Counts Database (CCDB) – a community resource of plant chromosome numbers. *New Phytol.* 206, 19–26. doi: 10.1111/nph.13191
- Rice A., Mayrose I. (2023). "The chromosome counts database (CCDB)," in *Plant Genomic and Cytogenetic Databases*, vol. 2703. Eds. Garcia S., Nualart N. (Humana, New York). doi: 10.1007/978-1-0716-3389-2_10
- Rieseberg L. H. (2001a). Chromosomal rearrangements and speciation. *Trends Ecol. Evol.* 16, 351–358. doi: 10.1016/S01695347(01)02187-5
- Rieseberg L. H. (2001b). Polyploid evolution: Keeping the peace at genomic reunions. *Curr. Biol.* 11, R925–R928. doi: 10.1016/S0960-9822(01)00556-5
- Rieseberg L. H., Van Fossen C., Desrochers A. (1995). Hybrid speciation accompanied by genomic reorganization in wild sunflowers. *Nature* 375, 313–316. doi: 10.1038/375313a0
- Roalson E. H. (2008). A synopsis of chromosome number variation in the Cyperaceae. *Bot. Rev.* 74, 209–393. doi: 10.1007/s12229-008-9011-y
- Rodionov A. V. (2023). Eupolyploidy as a mode in plant speciation. *Russ. J. Genet.* 59, 419–443. doi: 10.1134/S1022795423050113
- Rosche C., Broennimann O., Novikov A., Mrázová V., Boiko G. V., Danihelka J., et al. (2025). Herbarium specimens reveal a cryptic invasion of polyploid *Centaurea stoebe* in Europe. *New Phytol.* 245, 392–405. doi: 10.1111/nph.20212
- Rurik I., Melicháková A., Gbúrová Štubová E., Kučera J., Kochjarová J., Paun O., et al. (2024). Homoplastic versus xenoplastic evolution: exploring the emergence of key intrinsic and extrinsic traits in the montane genus *Soldanella* (Primulaceae). *Plant J.* 118, 753–765. doi: 10.1111/tjp.16630
- Sader M. A., Amorim B. S., Costa L., Souza G., Pedrosa-Harand A. (2019). The role of chromosome changes in the diversification of *Passiflora* L. (Passifloraceae). *Syst. Biodivers.* 17, 7–21. doi: 10.1080/14772000.2018.1546777
- Scarow M., Wang Y., Sun G. (2020). Molecular regulatory mechanisms underlying the adaptability of polyploid plants. *Biol. Rev.* 96, 394–407. doi: 10.1111/bvr.12661
- Schmickl R., Jørgensen M. H., Brysting A. K., Koch M. A. (2010). The evolutionary history of the *Arabidopsis lyrata* complex: a hybrid in the amph-Beringian area closes a large distribution gap and builds up a genetic barrier. *BMC Evol. Biol.* 10, 98. doi: 10.1186/1471-2148-10-98

- Schmickl R., Vallejo Marín M., Hojka J., Gorospe J. M., Haghighatnia M. J., İltas Ö., et al. (2024). Polyploidy-induced floral changes lead to unexpected pollinator behavior in *Arabidopsis arenosa*. *Oikos* 2024, e10267. doi: 10.1111/oik.10267
- Schmickl R., Yant L. (2021). Adaptive introgression: how polyploidy reshapes gene flow landscapes. *New Phytol.* 230, 457–461. doi: 10.1111/nph.17204
- Schoener T. W. (1990). "Ecological interactions," in *Analytical Biogeography*. Eds. Myers A. A., Giller P. S. (Springer, Dordrecht). doi: 10.1007/978-94-009-0435-4_9
- Schranz M. E., Osborn T. C. (2000). Novel flowering time variation in the resynthesized polyploid *Brassica napus*. *J. Hered.* 91, 242–246. doi: 10.1093/jhered/91.3.242
- Schumer M., Rosenthal G. G., Andolfatto P. (2014). How common is homoploid hybrid speciation? *Evol.* 68, 1553–1560. doi: 10.1111/evo.12399
- Scopece G., Lecer C., Widmer A., Cozzolino S. (2010). Polymorphism of postmating reproductive isolation within plant species. *Taxon* 59, 1367–1374. doi: 10.2307/20774034
- Scott A. D., Kolesnikova U., Glushkevich A., Steinmann L., Tikhomirov N., Pfordt U., et al. (2024). Multiple polyploidizations in *Arabidopsis lyrata* stabilized by long-range adaptive introgression across Eurasia. *BioRxiv*. doi: 10.1101/2024.08.27.609292
- Scott A. D., Stenz N. W. M., Ingvarsson P. K., Baum D. A. (2016). Whole genome duplication in coast redwood (*Sequoia sempervirens*) and its implications for explaining the rarity of polyploidy in conifers. *New Phytol.* 211, 186–193. doi: 10.1111/nph.13930
- Seehausen O. (2004). Hybridization and adaptive radiation. *Trends Ecol. Evol.* 19, 198–207. doi: 10.1016/j.tree.2004.01.003
- Segraves K. A., Anneberg T. J. (2016). Species interaction and plant polyploidy. *Am. J. Bot.* 103, 1326–1335. doi: 10.3732/ajb.1500529
- Servick S., Visger C. J., Gitzendanner M. A., Soltis P. S., Soltis D. E. (2015). Population genetic variation, geographic structure, and multiple origins of autopolyploidy in *Galax urceolata*. *Am. J. Bot.* 102, 973–982. doi: 10.3732/ajb.1400554
- Seymour D. K., Koenig D., Hagmann J., Becker C., Weigel D. (2014). Evolution of DNA methylation patterns in the brassicaceae is driven by differences in genome organization. *PLoS Genet.* 10, e1004785. doi: 10.1371/journal.pgen.1004785
- Shi F. X., Li M. R., Li Y. L., Jiang P., Zhang C., Pan Y. Z., et al. (2015). The impacts of polyploidy, geographic and ecological isolations on the diversification of *Panax* (Araliaceae). *BMC Plant Biol.* 15, 297. doi: 10.1186/s12870-015-0669-0
- Shu J.-P., Wang H., Shen H., Wang R.-J., Fu Q., Wang Y.-D., et al. (2022). Phylogenomic analysis reconstructed the order matoniales from paleopolyploidy veil. *Plants* 11, 1529. doi: 10.3390/plants11121529
- Skopaliková J., Leong-Skornicková J., Šída O., Newman M., Chumová Z., Fér T., et al. (2023). Ancient hybridization in *Curcuma* (Zingiberaceae) – accelerator or brake in lineage diversification? *Plant J.* 116, 773–785. doi: 10.1111/tpj.16408
- Šlenker M., Kantor A., Marhold K., Schmickl R., Mandáková T., Lysak M. A., et al. (2021). Allele sorting as a novel approach to resolving the origin of allotetraploids using hybrid-Seq data: A case study of the balkan mountain endemic *cardamine barbaraoides*. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.659275
- Slovák M., Melichárková A., Gburová Štubňová E., Kučera J., Mandáková T., Smyčka J., et al. (2023). Pervasive introgression during rapid diversification of the european mountain genus *soldanella* (L.) (Primulaceae). *Syst. Biol.* 72, 491–504. doi: 10.1093/sysbio/syaa071
- Soltis P. S., Liu X., Marchant D. B., Visger C. J., Soltis D. E. (2014). Polyploidy and novelty: Gottlieb's legacy. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 369, 20130351. doi: 10.1098/rstb.2013.0351
- Soltis P. S., Ma D. B., Van de Peer Y., Soltis D. E. (2015). Polyploidy and genome evolution in plants. *Curr. Opin. Genet. Dev.* 35, 119–125. doi: 10.1016/j.gde.2015.11.003
- Soltis D. E., Soltis P. S. (1989). Genetic consequences of autopolyploidy in *Tolmiea* (Saxifragaceae). *Evol.* 43, 586–594. doi: 10.1111/j.1558-5646.1989.tb04254.x
- Soltis P. S., Soltis D. E. (2009). The role of hybridization in plant speciation. *Annu. Rev. Plant Biol.* 60, 561–588. doi: 10.1146/annurev.arplant.043008.092039
- Soltis D. E., Soltis P. S., Bennett M. D., Leitch I. J. (2013). Evolution of genome size in the angiosperms. *Am. J. Bot.* 90, 1596–1603. doi: 10.3732/ajb.90.11.1596
- Stankowski S., Streisfeld M. A. (2015). Introgressive hybridization facilitates adaptive divergence in a recent radiation of monkeyflowers. *Proc. R. Soc. B: Biol. Sci.* 282, 20151666. doi: 10.1098/rspb.2015.1666
- Stathos A. M., Fishman L. (2014). Chromosomal rearrangements directly cause underdominant F1 pollen sterility in *Mimulus lewisii*-*M. cardinalis* hybrids. *Evol.* 68, 3109–3119. doi: 10.1111/evo.12503
- Stebbins G. L. (1947). Types of polyploids: their classification and significance. *Adv. Genet.* 1, 403–429. doi: 10.1016/S0065-2660(08)60490-3
- Stebbins G. L. (1957). The hybrid origin of microspecies in the *Elymus glaucus* complex. *Cytologia Suppl.* 36, 336–340.
- Stebbins G. L. (1958). The inviability, weakness, and sterility of interspecific hybrids. *Adv. Genet.* 9, 147–215. doi: 10.1016/S0065-2660(08)60162-5
- Stebbins G. L. (1971). *Chromosomal evolution in higher plants* (London, UK: Edward Arnold).
- Stebbins G. L. (1984). Mosaic evolution, mosaic selection and angiosperm phylogeny. *Bot. Jour. Linn. Soc.* 88, 149–164. doi: 10.1111/j.1095-8339.1984.tb01568.x
- Sturtevant A. H. (1921). A case of rearrangement of genes in *Drosophila*. *Proc. Nat. Acad. Sci.* 7, 235–237. doi: 10.1073/pnas.7.8.235
- Suarez-Gonzalez A., Lexer C., Cronk Q. C. B. (2018). Adaptive introgression: A plant perspective. *Biol. Lett.* 14, 20170688. doi: 10.1098/rsbl.2017.0688
- Suda J., Krahulcová A., Trávníček P., Rosenbaumová R., Peckert T., Krahulec F. (2007). Genome size variation and species relationships in *hieracium* sub-genus *pilosella* (Asteraceae) as inferred by flow cytometry. *Ann. Bot.* 100, 323–335. doi: 10.1093/aob/mcm218
- Svardal H., Quah F. X., Malinsky M., Ngatunga B. P., Miska E. A., Salzburger W., et al. (2020). Ancestral hybridization facilitated species diversification in the Lake Malawi cichlid fish adaptive radiation. *Mol. Biol. Evol.* 37, 1100–1113. doi: 10.1093/molbev/msz294
- Tang H., Bowers J. E., Wang X., Ming R., Alam M., Paterson A. H. (2008). Synteny and collinearity in plant genomes. *Science* 320, 486–488. doi: 10.1126/science.1153917
- Tate J. A., Douglas E. S., Soltis P. S. (2005). "Polyploidy in plants," in *The evolution of the genome*. Ed. Gregory, T. R. (San Diego: Elsevier Academic Press), 371–426.
- Tayalé A., Parisod C. (2013). Natural pathways to polyploidy in plants and consequences for genome reorganization. *Cytogenet. Genome Res.* 140, 79–96. doi: 10.1159/000351318
- Taylor S. A., Larson E. L. (2019). Insights from genomes into the evolutionary importance and prevalence of hybridization in nature. *Nat. Ecol. Evol.* 3, 170–177. doi: 10.1038/s41559-018-0777-y
- Te Beest M., Le Roux J. J., Richardson D. M., Brysting A. K., Suda J., Kubešová M., et al. (2012). The more the better? The role of polyploidy in facilitating plant invasions. *Ann. Bot.* 109, 19–45. doi: 10.1093/aob/mcr277
- Thébault A., Gillet F., Müller-Schärer H., Buttler A. (2011). Polyploidy and invasion success: trait trade-offs in native and introduced cytotypes of two Asteraceae species. *Plant Ecol.* 212, 315–325. doi: 10.1007/s11258-010-9824-8
- Thompson J. N. (2009). The coevolving web of life. *Am. Nat.* 173, 125–140. doi: 10.1086/595752
- Thompson J. N., Nuismer S. L., Merg. K. (2004). Plant polyploidy and the evolutionary ecology of plant/animal interactions. *Biol. J. Linn. Soc.* 82, 511–519. doi: 10.1111/j.1095-8312.2004.00338.x
- Todesco M., Owens G. L., Bercovich N., Legeré J. S., Soudi S., Burge D. O., et al. (2020). Massive haplotypes underlie ecotypic differentiation in sunflowers. *Nature* 584, 602–607. doi: 10.1038/s41586-020-2467-6
- Todesco M., Pascual M. A., Owens G. L., Ostevik K. L., Moyers B. T., Hübner S., et al. (2016). Hybridization and extinction. *Evol. Appl.* 9, 892–908. doi: 10.1111/eva.12367
- Tomlin C. M., Rajaraman S., Sebesta J. T., Scheen A.-C., Bendiksby M., Wen Low Y., et al. (2024). Allopolyploid origin and diversification of the Hawaiian endemic mints. *Nat. Commun.* 15, 3109. doi: 10.1038/s41467-024-47247-y
- Tribble C. M., Márquez-Corro J. I., May M. R., Hipp A. L., Escudero M., Zenil-Ferguson R. (2025). Macroevolutionary inference of complex modes of chromosomal speciation in a cosmopolitan plant lineage. *New Phytol.* 24, 2350–2361. doi: 10.1111/nph.20353
- Trickett A. J., Butlin R. K. (1994). Recombination suppressors and the evolution of new species. *Heredity* 73, 339–345. doi: 10.1038/hdy.1994.180
- Turcotte M. M., Kaufmann N., Wagner K. L., Zallek T. A., Ashman T.-L. (2024). Neopolyploidy increases stress tolerance and reduces fitness plasticity across multiple urban pollutants: support for the "general-purpose" genotype hypothesis. *Evol. Lett.* 8, 416–426. doi: 10.1093/evlett/grad072
- Valdés-Florido A., González-Toral C., Maguilla E., Cires E., Díaz-Lifante Z., Andrés-Camacho C., et al. (2024a). Polyploidy and hybridization in the Mediterranean: unraveling the evolutionary history of *Centaurea* (Gentianaceae). *Ann. Bot.* 134, 247–262. doi: 10.1093/aob/mcae066
- Valdés-Florido A., Tan L., Maguilla E., Simón-Porcar V. I., Zhou Y.-H., Arroyo J., et al. (2023). Drivers of diversification in *Linum* (Linaceae) by means of chromosome evolution: correlations with biogeography, breeding system and habit. *Ann. Bot.* 132, 949–961. doi: 10.1093/aob/mcad139
- Valdés-Florido A., Valcárcel V., Maguilla E., Díaz-Lifante Z., Andrés-Camacho C., Zeltner L., et al. (2024b). The interplay between climatic niche evolution, polyploidy and reproductive traits explains plant speciation in the Mediterranean Basin: a case study in *Centaurea* (Gentianaceae). *Front. Plant Sci.* 15. doi: 10.3389/fpls.2024.1439985
- Van de Peer Y., Ashman T. L., Soltis P. S., Soltis D. E. (2021). Polyploidy: an evolutionary and ecological force in stressful times. *Plant Cell* 33, 11–26. doi: 10.1093/plcell/koaa015
- Van de Peer Y., Mizrahi E., Marchal K. (2017). The evolutionary significance of polyploidy. *Nat. Rev. Genet.* 18, 411–424. doi: 10.1038/nrg.2017.26
- Van der Heijden E., Näsvalk K., Seixas F. A., Becerra Nobre C. E., Campos D. Maia A., Salazar-Carrion P., et al. (2024). Genomics of Neotropical biodiversity indicators: two butterfly radiations with rampant chromosomal rearrangements and hybridisation. *BioRxiv*. doi: 10.1101/2024.07.07.602206
- Van Drunen W. E., Husband B. C. (2019). Evolutionary associations between polyploidy, clonal reproduction, and perenniality in the angiosperms. *New Phytol.* 224, 1266–1277. doi: 10.1111/nph.15999
- Vanrell M. A., Novaes L. R., Afonso A., Arroyo J., Simón-Porcar V. (2024). Ecological correlates of population genetics in *Linum suffruticosum*, an heterostylous polyploid and taxonomic complex endemic to the Western Mediterranean Basin. *Acta Bot. Plants* 16, plae027. doi: 10.1093/aobpla/plae027

- Vereecken N. J., Cozzolino S., Schiestl F. P. (2010). Hybrid floral scent novelty drives pollinator shift in sexually deceptive orchids. *BMC Evol. Biol.* 10, 103. doi: 10.1186/1471-2148-10-103
- Vlček J., Hämälä T., Cobo C. V., Curran E., Šrámková G., Slotte T., et al. (2025). Whole-genome duplication increases genetic diversity and load in outcrossing *Arabidopsis*. *BiorXiv*. doi: 10.1101/2025.01.12.632621
- Vleugels T., Cnops G., Van Bockstaele E. (2013). Screening for resistance to clover rot (*Sclerotinia* spp.) among a diverse collection of red clover populations (*Trifolium pratense* L.). *Euphytica*. 194, 371–382. doi: 10.1007/s10681-013-0949-4
- Wan J., Oduor A. M., Pouteau R., Wang B., Chen L., Yang B., et al. (2020). Can polyploidy confer invasive plants with a wider climatic tolerance? A test using *Solidago canadensis*. *Ecol. Evol.* 10, 5617–5630. doi: 10.1002/ece3.6303
- Wang J., Li D., Shang F., Kang X. (2017). High temperature-induced production of unreduced pollen and its cytological effects in *Populus*. *Sci. Rep.* 7, 5281. doi: 10.1038/s41598-017-05661-x
- Wang C., Liu L., Yin M., Eller F., Brix H., Wang T., et al. (2024). Genome-wide analysis tracks the emergence of intraspecific polyploids in *Phragmites australis*. *NPJ Biodivers.* 3, 29. doi: 10.1038/s44185-024-00060-8
- Wang X., Morton J. A., Pellicer J., Leitch I. J., Leitch A. R. (2021). Genome downsizing after polyploidy: mechanisms, rates and selection pressures. *Plant J.* 107, 1003–1015. doi: 10.1111/tj.15363
- Wang M., Wang P., Lin M., Ye Z., Li G., Tu L., et al. (2018). Evolutionary dynamics of 3D genome architecture following polyploidization in cotton. *Nat. Plants* 4, 90–97. doi: 10.1038/s41477-017-0096-3
- Weiss-Schneeweiss H., Schneeweiss G. M. (2013). “Karyotype diversity and evolutionary trends in angiosperms,” in *Plant Genome Diversity vol.2: Physical Structure, Behaviour and Evolution of Plant Genomes* vol. 2. Eds. Greilhuber J., Dolezel J., Wendel J. F. (Springer, Vienna), 209–230.
- Wendel J. F. (2015). The wondrous cycles of polyploidy in plants. *Am. J. Bot.* 102, 1753–1756. doi: 10.3732/ajb.1500320
- Wendel J. F., Jackson S. A., Meyers B. C., Wing R. A. (2016). Evolution of plant genome architecture. *Genome Biol.* 17, 1–14. doi: 10.1186/s13059-016-0908-1
- Wendel J. F., Lisch D., Hu G., Mason A. S. (2018). The long and short of doubling down: polyploidy, epigenetics, and the temporal dynamics of genome fractionation. *Curr. Opin. Genet. Dev.* 49, 1–7. doi: 10.1016/j.gde.2018.01.004
- Whitkus R. (1988). Experimental hybridizations among chromosome races of *Carex pachystachya* and the related species *C. macloviana* and *C. preslii* (Cyperaceae). *Syst. Bot.* 13, 146–153. doi: 10.2307/2419251
- Wolf P. G., Haufler C. H., Sheffield E. (1987). Electrophoretic evidence for genetic diploidy in the bracken fern (*Pteridium aquilinum*). *Science* 236, 947–949. doi: 10.1126/science.236.4804.947
- Wray G. A., Hahn M. W., Abouheif E., Balhoff J. P., Pizer M., Rockman M. V., et al. (2003). The evolution of transcriptional regulation in eukaryotes. *Mol. Biol. Evol.* 20, 1377–1419. doi: 10.1093/molbev/msg140
- Wu F., Tanksley S. D. (2010). Chromosomal evolution in the plant family Solanaceae. *BMC Genomics* 11, 182. doi: 10.1186/1471-2164-11-182
- Yakimowski S. B., Rieseberg L. H. (2014). The role of homoploid hybridization in evolution: A century of studies synthesizing genetics and ecology. *Am. J. Bot.* 101, 1247–1258. doi: 10.3732/ajb.1400201
- Yant L., Hollister J. D., Wright K. M., Arnold B. J., Higgins J. D., Franklin F. C. H., et al. (2013). Meiotic adaptation to genome duplication in *Arabidopsis arenosa*. *Curr. Biol.* 23, 2151–2156. doi: 10.1016/j.cub.2013.08.059
- Yona A. H., Manor Y. S., Herbst R. H., Romano G. H., Mitchell A. M., Kupiec M., et al. (2012). Chromosomal duplication is a transient evolutionary solution to stress. *Proc. Natl. Acad. Sci.* 109, 21010–21015. doi: 10.1073/pnas.1211150109
- Yuan Y., Bayer P. E., Batley J., Edwards D. (2021). Current status of structural variation studies in plants. *Plant Biotechnol. J.* 19, 2153–2163. doi: 10.1111/pbi.13646
- Zanne A. E., Tank D. C., Cornwell W. K., Eastman J. M., Smith S. A., FitzJohn R. G., et al. (2014). Three keys to the radiation of angiosperms into freezing environments. *Nature*. 506, 89–92. doi: 10.1038/nature12872
- Zhan S., Otto S., Barker M. (2021). Broad variation in rates of polyploidy and dysploidy across flowering plants is correlated with lineage diversification. *BioRxiv*. doi: 10.1101/2021.03.30.436382
- Zhang K., Wang X., Cheng P. (2019). Plant polyploidy: origin, evolution, and its influence on crop domestication. *Hortic. Plant Journ.* 5, 231–239. doi: 10.1016/j.hpj.2019.11.003
- Zhang X., Wang G., Zhang S., Chen S., Wang Y., Wen P., et al. (2020). Genomes of the banyan tree and pollinator wasp provide insights into fig-wasp coevolution. *Cell*. 183, 875–889.e17. doi: 10.1016/j.cell.2020.09.043
- Zhao J., Bayer P. E., Ruperao P., Saxena R. K., Khan A. W., Golicz A. A., et al. (2020). Trait associations in the pangenome of pigeon pea (*Cajanus cajan*). *Plant Biotechnol. J.* 18, 1946–1954. doi: 10.1111/pbi.13354
- Zhao J., Feng Q., Lu H., Li Y., Wang A., Tian Q., et al. (2018). Pan-genome analysis highlights the extent of genomic variation in cultivated and wild rice. *Nat. Genet.* 50, 278–284. doi: 10.1038/s41588-018-0041-z
- Zheng L., Kinosian S. P., Zhan S. H., Barker M. S. (2024). Ancient polyploid and low rate of chromosome loss explain the high chromosome numbers of homosporous ferns. *BioRxiv*. doi: 10.1101/2024.09.23.614530
- Zhong S., Li B., Chen W., Wang L., Guan J., Wang Q., et al. (2022). The chromosome-level genome of *Akebia trifoliata* as an important resource to study plant evolution and environmental adaptation in the Cretaceous. *Plant J.* 112, 1316–1330. doi: 10.1111/tj.16011

Frontiers in Plant Science

Cultivates the science of plant biology and its applications

The most cited plant science journal, which advances our understanding of plant biology for sustainable food security, functional ecosystems and human health.

Discover the latest Research Topics

[See more →](#)

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact

