Horticulture Research

Horticulture Research, 2024, **11**: uhae194
https://doi.org/10.1093/hr/uhae194

Article

Genomic data provides insights into the evolutionary history and adaptive differentiation of two tetraploid strawberries

Hanyang Lin[1,2], Luxi Chen[2], Chaonan Cai[2], Junxia Ma[2], Junmin Li[1,2,*], Tia-Lynn Ashman[3], Aaron Liston[4] and Ming Dong[5]

[1]School of Advanced Study, Taizhou University, Taizhou 318000, China
[2]Zhejiang Provincial Key Laboratory of Plant Evolutionary Ecology and Conservation, School of Life Sciences, Taizhou University, Taizhou 318000, China
[3]Department of Biological Sciences, University of Pittsburgh, Pittsburgh, PA 15260, USA
[4]Department of Botany and Plant Pathology, Oregon State University, Corvallis, OR 97331, USA
[5]Key Laboratory of Hangzhou City for Ecosystem Protection and Restoration, College of Life and Environmental Sciences, Hangzhou Normal University, Hangzhou 311121, China
[*] Corresponding author. E-mails: lijm@tzc.edu.cn; lijmtzc@126.com

Abstract

Over the decades, evolutionists and ecologists have shown intense interest in the role of polyploidization in plant evolution. Without clear knowledge of the diploid ancestor(s) of polyploids, we would not be able to answer fundamental ecological questions such as the evolution of niche differences between them or its underlying genetic basis. Here, we explored the evolutionary history of two *Fragaria* tetraploids, *Fragaria corymbosa* and *Fragaria moupinensis*. We *de novo* assembled five genomes including these two tetraploids and three diploid relatives. Based on multiple lines of evidence, we found no evidence of subgenomes in either of the two tetraploids, suggesting autopolyploid origins. We determined that *Fragaria chinensis* was the diploid ancestor of *F. corymbosa* while either an extinct species affinitive to *F. chinensis* or an unsampled population of *F. chinensis* could be the progenitor of *F. moupinensis*. Meanwhile, we found introgression signals between *F. chinensis* and *Fragaria pentaphylla*, leading to the genomic similarity between these two diploids. Compared to *F. chinensis*, gene families related to high ultraviolet (UV)-B and DNA repair were expanded, while those that responded towards abiotic and biotic stresses (such as salt stress, wounding, and various pathogens) were contracted in both tetraploids. Furthermore, the two tetraploids tended to down-regulate defense response genes but up-regulate UV-B response, DNA repairing, and cell division gene expression compared to *F. chinensis*. These findings may reflect adaptions toward high-altitude habitats. In summary, our work provides insights into the genome evolution of wild *Fragaria* tetraploids and opens up an avenue for future works to answer deeper evolutionary and ecological questions regarding the strawberry genus.

Introduction

Polyploidization (also known as whole-genome duplication) refers to the possession of more than two copies of each chromosome [1]. Polyploidization is widespread across eukaryotes, especially in plants [2, 3]. For instance, all extant species of vascular plants have experienced two ancient polyploidization events (ζ and ϵ), while monocots and dicots have undergone the other two (τ and γ, respectively) after their divergence [4]. Therefore, polyploidization is acknowledged to be one of the key factors promoting the evolution of higher plant genomes, the generation of new species, and the emergence of adaptive evolution [2, 5, 6].

Historically, some researchers have argued that polyploidization may lead to 'evolutionary dead-ends' [7]. Shortly after the polyploidization events, polyploids would exhibit lower rates of speciation and higher rates of extinction compared to their diploid relatives, further losing evolutionary potential [8]. However, mounting evidence has shown that polyploid species show stronger evolutionary advantages in the long term [6, 9]. Compared to their diploid relatives, polyploids tend to show stronger adaptability and a broader ecological niche, especially with respect to harsh environments (such as drought and extreme temperature) [10–13]. Nevertheless, due to the biased research interests (mainly cultivated plants, e.g., cotton [14]; sugarcane [15]; banana [16]) and the lack of sufficient genomic data, diploid ancestors of many wild polyploids await discovery. In addition, whether and how polyploids show differential responses to extreme habitats compared to their diploid relatives remains an open but intriguing question.

Harboring about 25 species with diverse ploidy levels (ranging from diploid [2×], to tetraploid [4×] and above [5×, 6×, 8×, 10×]), strawberries (*Fragaria* L., Rosaceae) are not only important food plants but also serve as a promising system to study the above-mentioned issues [17, 18]. To date, the diploid ancestors of the octoploid cultivated strawberry (*Fragaria* × *ananassa*) have been intensively studied [19–26]. Lines of evidence indicate that one *F.* × *ananassa* subgenome is most closely related to the extant

Received: 14 March 2024; **Accepted:** 5 July 2024; **Published:** 11 July 2024; **Corrected and Typeset:** 1 September 2024
© The Author(s) 2024. Published by Oxford University Press on behalf of Nanjing Agricultural University. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from https://academic.oup.com/hr/article/11/9/uhae194/7712263 by guest on 12 February 2026

process' category were focused and only those with Fisher's exact test (with Benjamini and Hochberg FDR correction) p-value <0.05 were kept (hereafter).

Differentially expressed genes between tetraploids and their relatives

To explore whether the pattern of DEGs was partly correlated with gene family expansions or contractions between the tetraploids and their relatives, we used RNA sequencing data generated in aiding genome annotation (four tissues; Table S6, see online supplementary material) to investigate DEGs between *F. corymbosa* and *F. chinensis*, and between *F. moupinensis* and *F. chinensis*. Clean reads were mapped to the primary haplotype of each tetraploid respectively using HISAT2 (v. 2.2.1) [83]. Then, we analysed the expression levels of genes using StringTie (v. 2.2.1) [84]. The read count information was extracted using the prepDE.py script available in StringTie. Next, the DEGs were identified using the R package DESeq2 (v. 1.34.0) [85] based on the criteria of $|\log_2 FC| > 1$ and $p_{adj} < 0.05$. Subsequently, these putative DEGs were annotated and enriched against *A. thaliana* using KOBAS-i. The GO terms under the 'biological process' category were focused.

Acknowledgements

This work was funded by the National Natural Science Foundation of China (No. 31261120580 to M.D. and J.L.) and the National Science Foundation (DEB 12410006 to T.-L.A.; DEB 1241217 to A.L.). The authors thank Lingsong Chen, Minghua Song, Tong Chen, Xiaoyan Wang, Beifen Yang, Jinzhi Li, and Zhaokui Du for the collection of samples.

Author contributions

M.D., J.L., T.-L.A., and A.L. initiated the project. H.L. and J.L. wrote and finalized the manuscript. H.L., L.C., C.C., and J.M. designed and performed most of the experiments. T.-L.A., A.L., and M.D. provided assistance and suggestions. All authors discussed the results and commented on the manuscript.

Data availability

The assembled genome sequences and annotation files are available from the figshare repository (https://doi.org/10.6084/m9.figshare.25999996).

Conflict of interest statement

The authors declare no competing interests.

Supplementary data

Supplementary data is available at *Horticulture Research* online.

References

1. Soltis PS, Marchant DB, Van de Peer Y. *et al.* Polyploidy and genome evolution in plants. *Curr Opin Genet Dev*. 2015;**35**:119–25
2. Jiao Y, Wickett NJ, Ayyampalayam S. *et al.* Ancestral polyploidy in seed plants and angiosperms. *Nature*. 2011;**473**:97–100
3. Van de Peer Y, Mizrachi E, Marchal K. The evolutionary significance of polyploidy. *Nat Rev Genet*. 2017;**18**:411–24
4. Alix K, Gérard PR, Schwarzacher T. *et al.* Polyploidy and interspecific hybridization: partners for adaptation, speciation and evolution in plants. *Ann Bot*. 2017;**120**:183–94
5. Ge S. A review of recent studies of plant systematics and evolution in China. *Biodivers Sci*. 2022;**30**:22385
6. Van de Peer Y, Ashman T-L, Soltis PS. *et al.* Polyploidy: an evolutionary and ecological force in stressful times. *Plant Cell*. 2021;**33**:11–26
7. Arrigo N, Barker MS. Rarely successful polyploids and their legacy in plant genomes. *Curr Opin Plant Biol*. 2012;**15**:140–6
8. Mayrose I, Zhan SH, Rothfels CJ. *et al.* Recently formed polyploid plants diversify at lower rates. *Science*. 2011;**333**:1257–7
9. Soltis DE, Segovia-Salcedo MC, Jordon-Thaden I. *et al.* Are polyploids really evolutionary dead-ends (again)? A critical reappraisal of Mayrose et al. (2011). *New Phytol*. 2014;**202**:1105–17
10. Abbott RJ, Brochmann C. History and evolution of the arctic flora: in the footsteps of Eric Hultén. *Mol Ecol*. 2003;**12**:299–313
11. Baniaga AE, Marx HE, Arrigo N. *et al.* Polyploid plants have faster rates of multivariate niche differentiation than their diploid relatives. *Ecol Lett*. 2020;**23**:68–78
12. Pandit MK, Pocock MJO, Kunin WE. Ploidy influences rarity and invasiveness in plants. *J Ecol*. 2011;**99**:1108–15
13. Rice A, Šmarda P, Novosolov M. *et al.* The global biogeography of polyploid plants. *Nat Ecol Evol*. 2019;**3**:265–73
14. Paterson AH, Wendel JF, Gundlach H. *et al.* Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres. *Nature*. 2012;**492**:423–7
15. Wang T, Wang B, Hua X. *et al.* A complete gap-free diploid genome in *Saccharum* complex and the genomic footprints of evolution in the highly polyploid *Saccharum* genus. *Nat Plants*. 2023;**9**:554–71
16. Li X, Yu S, Cheng Z. *et al.* Origin and evolution of the triploid cultivated banana genome. *Nat Genet*. 2024;**56**:136–42
17. Lei JJ, Xue L, Guo RX. *et al.* The *Fragaria* species native to China and their geographical distribution. *Acta Hortic*. 2017;**1156**:37–46
18. Liston A, Cronn R, Ashman T. *Fragaria*: a genus with deep historical roots and ripe for evolutionary and ecological insights. *Am J Bot*. 2014;**101**:1686–99
19. Edger PP, McKain MR, Yocca AE. *et al.* Reply to: revisiting the origin of octoploid strawberry. *Nat Genet*. 2020;**52**:5–7
20. Edger PP, Poorten TJ, VanBuren R. *et al.* Origin and evolution of the octoploid strawberry genome. *Nat Genet*. 2019;**51**:541–7
21. Feng C, Wang J, Harris AJ. *et al.* Tracing the diploid ancestry of the cultivated octoploid strawberry. *Mol Biol Evol*. 2021;**38**:478–85
22. Jin X, Du H, Zhu C. *et al.* Haplotype-resolved genomes of wild octoploid progenitors illuminate genomic diversifications from wild relatives to cultivated strawberry. *Nat Plants*. 2023;**9**:1252–66
23. Liston A. On the origin of strawberries. *Nat Plants*. 2023;**9**:1176–7
24. Liston A, Wei N, Tennessen JA. *et al.* Revisiting the origin of octoploid strawberry. *Nat Genet*. 2020;**52**:2–4
25. Session AM, Rokhsar DS. Transposon signatures of allopolyploid genome evolution. *Nat Commun*. 2023;**14**:3180
26. Zhou Y, Xiong J, Shu Z. *et al.* The telomere-to-telomere genome of *Fragaria vesca* reveals the genomic evolution of *Fragaria* and the origin of cultivated octoploid strawberry. *Hortic Res*. 2023;**10**:uhad027
27. Kamneva OK, Syring J, Liston A. *et al.* Evaluating allopolyploid origins in strawberries (*Fragaria*) using haplotypes generated from target capture sequencing. *BMC Evol Biol*. 2017;**17**:180
28. Feng C, Wang J, Liston A. *et al.* Recombination variation shapes phylogeny and introgression in wild diploid strawberries. *Mol Biol Evol*. 2023;**40**:msad049

Downloaded from https://academic.oup.com/hr/article/11/9/uhae194/7712263 by guest on 12 February 2026

29. Staudt G. Strawberry biogeography, genetics and systematics. *Acta Hortic.* 2009;**842**:71–84
30. Johnson AL, Govindarajulu R, Ashman T-L. Bioclimatic evaluation of geographical range in *Fragaria* (Rosaceae): consequences of variation in breeding system, ploidy and species age. *Bot J Linn Soc.* 2014;**176**:99–114
31. Wan J-Z, Chen L-X, Gao S. *et al.* Ecological niche shift between diploid and tetraploid plants of *Fragaria* (Rosaceae) in China. *South African J Bot.* 2019;**121**:68–75
32. Qiao Q, Edger PP, Xue L. *et al.* Evolutionary history and pan-genome dynamics of strawberry (*Fragaria* spp.). *Proc Natl Acad Sci.* 2021;**118**:e2105431118
33. Liu B, Davis TM. Conservation and loss of ribosomal RNA gene sites in diploid and polyploid *Fragaria* (Rosaceae). *BMC Plant Biol.* 2011;**11**:157
34. Zhang L, Han Y, Li Z. Advances in cytogenetics of *Fragaria*. *J Jiangsu Norm Univ Sci Ed.* 2020;**38**:65–9
35. Qiao Q, Cao Q, Zhang R. *et al.* Genomic analyses provide insights into sex differentiation of tetraploid strawberry (*Fragaria moupinensis*). *Plant Biotechnol J.* 2024;**22**:1552–65
36. Wu S, Wang Y, Wang Z. *et al.* Species divergence with gene flow and hybrid speciation on the Qinghai–Tibet plateau. *New Phytol.* 2022;**234**:392–404
37. Hu R, Li X, Hu Y. *et al.* Adaptive evolution of the enigmatic *Takakia* now facing climate change in Tibet. *Cell.* 2023;**186**:3558–3576.e17
38. Tsukaya H, Tsuge T. Morphological adaptation of inflorescences in plants that develop at low temperatures in early spring: the convergent evolution of "downy plants". *Plant Biol.* 2001;**3**:536–43
39. Kumar S, Suyal DC, Yadav A. *et al.* Microbial diversity and soil physiochemical characteristic of higher altitude. *PLoS One.* 2019;**14**:e0213844
40. Pellissier L, Roger A, Bilat J. *et al.* High elevation *Plantago lanceolata* plants are less resistant to herbivory than their low elevation conspecifics: is it just temperature? *Ecography.* 2014;**37**:950–9
41. Rasmann S, Buri A, Gallot-Lavallée M. *et al.* Differential allocation and deployment of direct and indirect defences by *Vicia sepium* along elevation gradients. *J Ecol.* 2014;**102**:930–8
42. Zhu Y, Zhang Y, Yan S. *et al.* Viral community structure and functional potential vary with lifestyle and altitude in soils of Mt. Everest. *Environ Int.* 2023;**178**:108055
43. Zhang X, Kuang T, Dong W. *et al.* Genomic convergence underlying high-altitude adaptation in alpine plants. *J Integr Plant Biol.* 2023;**65**:1620–35
44. Nathewet P, Yanagi T, Hummer KE. *et al.* Karyotype analysis in wild diploid, tetraploid and hexaploid strawberries, *Fragaria* (Rosaceae). *Cytologia.* 2009;**74**:355–64
45. Li R, Zhu H, Ruan J. *et al.* *De novo* assembly of human genomes with massively parallel short read sequencing. *Genome Res.* 2010;**20**:265–72
46. Chin C-S, Peluso P, Sedlazeck FJ. *et al.* Phased diploid genome assembly with single-molecule real-time sequencing. *Nat Methods.* 2016;**13**:1050–4
47. Hu J, Fan J, Sun Z. *et al.* NextPolish: a fast and efficient genome polishing tool for long-read assembly. *Bioinformatics.* 2020;**36**:2253–5
48. Langmead B, Salzberg SL. Fast gapped-read alignment with bowtie 2. *Nat Methods.* 2012;**9**:357–9
49. Zhang X, Zhang S, Zhao Q. *et al.* Assembly of allele-aware, chromosomal-scale autopolyploid genomes based on Hi-C data. *Nat Plants.* 2019;**5**:833–45
50. Robinson JT, Turner D, Durand NC. *et al.* Juicebox.Js provides a cloud-based visualization system for hi-C data. *Cell Syst.* 2018;**6**:256–258.e1
51. Manni M, Berkeley MR, Seppey M. *et al.* BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol Biol Evol.* 2021;**38**:4647–54
52. Wang Y, Tang H, DeBarry JD. *et al.* MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 2012;**40**:e49–9
53. Ranallo-Benavidez TR, Jaron KS, Schatz MC. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat Commun.* 2020;**11**:1432
54. Wang L, Sun X, Peng Y. *et al.* Genomic insights into the origin, adaptive evolution, and herbicide resistance of *Leptochloa chinensis*, a devastating tetraploid weedy grass in rice fields. *Mol Plant.* 2022;**15**:1045–58
55. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing, 2021
56. Hibrand Saint-Oyant L, Ruttink T, Hamama L. *et al.* A high-quality genome sequence of *Rosa chinensis* to elucidate ornamental traits. *Nat Plants.* 2018;**4**:473–84
57. VanBuren R, Wai CM, Colle M. *et al.* A near complete, chromosome-scale assembly of the black raspberry (*Rubus occidentalis*) genome. *Gigascience.* 2018;**7**:giy094
58. Ireland HS, Wu C, Deng CH. *et al.* The *Gillenia trifoliata* genome reveals dynamics correlated with growth and reproduction in Rosaceae. *Hortic Res.* 2021;**8**:233
59. Khan A, Carey SB, Serrano A. *et al.* A phased, chromosome-scale genome of 'Honeycrisp' apple (*Malus domestica*). *Gigabyte.* 2022;**2022**:1–15
60. Zheng T, Li P, Zhuo X. *et al.* The chromosome-level genome provides insight into the molecular mechanism underlying the tortuous-branch phenotype of *Prunus mume*. *New Phytol.* 2022;**235**:141–56
61. Zhang A, Zhou H, Jiang X. *et al.* The draft genome of a flat peach (*Prunus persica* L. cv. '124 pan') provides insights into its good fruit flavor traits. *Plan Theory.* 2021;**10**:538
62. Linsmith G, Rombauts S, Montanari S. *et al.* Pseudo-chromosome–length genome assembly of a double haploid "Bartlett" pear (*Pyrus communis* L.). *Gigascience.* 2019;**8**:giz138
63. Cheng C, Krishnakumar V, Chan AP. *et al.* Araport11: a complete reannotation of the *Arabidopsis thaliana* reference genome. *Plant J.* 2017;**89**:789–804
64. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 2019;**20**:238
65. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat Methods.* 2015;**12**:59–60
66. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 2013;**30**:772–80
67. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics.* 2009;**25**:1972–3
68. Hoang DT, Chernomor O, von Haeseler A. *et al.* UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol.* 2018;**35**:518–22
69. Minh BQ, Schmidt HA, Chernomor O. *et al.* IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol.* 2020;**37**:1530–4
70. Kalyaanamoorthy S, Minh BQ, Wong TKF. *et al.* ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods.* 2017;**14**:587–9
71. Zhang C, Rabiee M, Sayyari E. *et al.* ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics.* 2018;**19**:153

Downloaded from https://academic.oup.com/hr/article/11/9/uhae194/7712263 by guest on 12 February 2026

72. Langdon QK, Peris D, Kyle B. *et al.* sppIDer: a species identification tool to investigate hybrid genomes with high-throughput sequencing. *Mol Biol Evol*. 2018;**35**:2835–49
73. Chen S, Zhou Y, Chen Y. *et al.* Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 2018;**34**:i884–90
74. Patterson N, Moorjani P, Luo Y. *et al.* Ancient admixture in human history. *Genetics*. 2012;**192**:1065–93
75. Li H, Durbin R. Fast and accurate short read alignment with burrows–wheeler transform. *Bioinformatics*. 2009;**25**:1754–60
76. Danecek P, Bonfield JK, Liddle J. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience*. 2021;**10**:giab008
77. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*. 2011;**27**:2987–93
78. Danecek P, Auton A, Abecasis G. *et al.* The variant call format and VCFtools. *Bioinformatics*. 2011;**27**:2156–8
79. Mendes FK, Vanderpool D, Fulton B. *et al.* CAFE 5 models variation in evolutionary rates among gene families. *Bioinformatics*. 2021;**36**:5516–8
80. Sanderson MJ. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics*. 2003;**19**:301–2
81. Shen W, Le S, Li Y. *et al.* SeqKit: a cross-platform and ultrafast toolkit for fasta/q file manipulation. *PLoS One*. 2016;**11**:e0163962
82. Bu D, Luo H, Huo P. *et al.* KOBAS-i: intelligent prioritization and exploratory visualization of biological functions for gene enrichment analysis. *Nucleic Acids Res*. 2021;**49**:W317–25
83. Kim D, Paggi JM, Park C. *et al.* Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol*. 2019;**37**:907–15
84. Pertea M, Kim D, Pertea GM. *et al.* Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat Protoc*. 2016;**11**:1650–67
85. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;**15**:550

Downloaded from https://academic.oup.com/hr/article/11/9/uhae194/7712263 by guest on 12 February 2026