Revisiting the origin of octoploid strawberry

Aaron Liston1*, Na Wei2, Jacob A. Tennessen3, Junmin Li4, Ming Dong5 and Tia-Lynn Ashman2

ARISING FROM: P. Edger et al. Nature Genetics https://doi.org/10.1038/s41588-019-0356-4 (2019)

The cultivated strawberry (Fragaria × ananassa) is an octoploid, and the identity of its four subgenomes has long been a mystery. In their recent publication on the strawberry genome, Edger et al.1 present a new hypothesis: each subgenome originated from a different extant diploid progenitor, and the hexaploid species Fragaria moschata was a direct ancestor. We reanalyzed the four octoploid subgenomes in a phylogenomic context, and our results support only two extant diploid progenitors; we also found no support for F. moschata as a direct ancestor. We argue that the tree-searching algorithm of Edger et al.1 is potentially biased against accepting extinct or unsampled progenitors, a weakness that could affect the resolution of polyploid genomic architecture.

Publication by Edger et al.1 of the chromosome-scale genome assembly of the octoploid strawberry Fragaria × ananassa cultivar ‘Camarosa’ represents a major scientific advance and provides a foundational resource for this important cultivated plant. Correctly identifying its diploid progenitors is important for understanding and predicting the responses of polyploid plants to climate change and associated environmental stress. Since the phylogenetic hypotheses proposed by Edger et al.1 (Fig. 1a) conflict with several recent studies2–4 (Supplementary Table 1), we conducted two new analyses: a chromosome-scale phylogenomic analysis of the four Fragaria × ananassa subgenomes (Figs. 1 and 2, Supplementary Table 2 and Extended Data Fig. 1) and a phylogenetic analysis of genetic linkage-mapped loci in F. moschata (Extended Data Figs. 2 and 3).

In our results (Figs. 1c and 2), the octoploid subgenomes Camarosa vesca and Camarosa inunae (A and B, respectively) generally are sister to their eponymous diploids F. vesca (red) and F. inunae (blue). In contrast, subgenomes Camarosa nipponica and Camarosa viridis (C and D, respectively) predominantly share a most recent common ancestor (clade) with F. inunae (light blue).

Edger et al.1 also hypothesized that the hexaploid species F. moschata “may be evolutionary intermediate between the diploids and wild octoploid species.” This predicts that the hexaploid’s three subgenomes correspond to the three Asian diploid ancestors (Fig. 1a). We find no support for this hypothesis (Supplementary Tables 3 and 4 and Extended Data Figs. 2 and 3). Instead, its progenitors appear restricted to the ‘vesca clade’ (F. vesca, F. mandshurica and F. bucharica).

We further evaluated the two phylogenetic hypotheses of Edger et al.1 using topology tests and found neither evidence of F. nipponica and F. viridis ancestry in the octoploid (Supplementary Table 3) nor support for the hexaploid as an ancestor to the octoploid strawberries (Supplementary Table 4).

We believe that the phylogenetic analysis of subgenomes tree-searching algorithm (PhyDS) developed by Edger et al.1 was responsible for their identification of F. nipponica and F. viridis ancestry in the octoploids. Their analysis was based on 8,405 individual gene trees compiled from the annotation of the newly assembled octoploid genome and 31 transcriptomes from 12 diploid Fragaria species. In each gene tree, when multiple octoploid genes resolved as sister to the same diploid, these were treated as in-paralogs and ignored. This forces each of the four octoploid subgenomes to have a different diploid ancestor, an assumption at odds with most classical genetic hypotheses for the octoploid ancestry (reviewed in ref. 1), as well as the results of molecular phylogenetic analyses of Fragaria11–13.

In effect, Edger et al.1 considered only gene trees as informative when genes from a single octoploid and diploid comprise an exclusive clade. This is an effective strategy for subgenomes A and B, where a high proportion of PhyDS-resolved gene trees (their supplementary table 8) and 100-kb windows (labeled ‘sister’ in Fig. 1c) resolve these with F. vesca and F. inunae, respectively. However, the approach fails for subgenomes C and D, where a smaller percentage of PhyDS-resolved gene trees (their supplementary table 8), and very few 100-kb windows (Figs. 1c and 2), resolve an exclusive clade relationship between these subgenomes and F. nipponica and F. viridis, respectively.

The rationale of Edger et al.1 for treating in-paralogs in this way was to avoid errors when homoeologous exchange has ‘replaced’ the syntelog of one subgenome with another, as illustrated in their supplementary fig. 8c. According to their supplementary table 10, 11.4% of the genome has experienced homeologous exchange, suggesting that 90% of the syntelog gene trees should correctly resolve subgenome ancestry. In contrast, only around 3% of syntelogs across the genome resulted in a subgenome assignment by PhyDS. This indicates that their in-paralog exclusion criterion was too strict. Our results suggest that the great majority of trees rejected by PhyDS resolved F. inunae as the diploid progenitor of three subgenomes (Fig. 1b, Extended Data Fig. 1 and Supplementary Table 5). This topology matches the examples of ‘incorrectly identified as progenitor’ in their supplementary fig. 8c, but in our view reflects the most likely evolutionary scenario for the origin of the octoploid strawberries.

Our phylogenomic approach found that 12.5% of the genome has experienced homoeologous exchange (Figs. 1c and 2 and Supplementary Table 6), similar to the estimate by Edger et al. of 11.4%. Although none of the 2,191 trees from 100-kb windows across the genome (Supplementary Table 5) matches the hypothesis of Edger et al. (Fig. 1a), a small number do resolve subgenome C or D with the diploid species F. nipponica and F. viridis, respectively (Fig. 1c and Supplementary Table 6). However, a similar number...

1Department of Botany and Plant Pathology, Oregon State University, Corvallis, OR, USA. 2Department of Biological Sciences, University of Pittsburgh, Pittsburgh, PA, USA. 3Department of Immunology and Infectious Diseases, Harvard T. H. Chan School of Public Health, Boston, MA, USA. 4Zhejiang Provincial Key Laboratory of Plant Evolutionary Ecology and Conservation, Taizhou University, Taizhou, China. 5Key Laboratory of Hangzhou City for Ecosystem Protection and Restoration, College of Life and Environmental Sciences, Hangzhou Normal University, Hangzhou, China. e-mail: aaron.liston@oregonstate.edu
Fig. 1 | Phylogenomics of octoploid subgenomes. a, Summary of the phylogenetic hypothesis of Edger et al. F. moschata subgenomes are indicated by asterisks. Bold lines represent the topology constraint we applied in the Shimodaira–Hasegawa test of the hypothesis of Edger et al. versus our phylogenomic results (Supplementary Table 3). b, Maximum-likelihood estimate of phylogeny for base chromosome 1; all nodes have 100% bootstrap support. This topology is shared by five of the seven chromosomes (Extended Data Fig. 1). c, Summary of phylogenetic positions of the four octoploid subgenomes in 2,191 windows of 10 kb across seven base chromosomes.

Fig. 2 | Phylogenetic positions of the four octoploid subgenomes in maximum-likelihood estimates of phylogeny in 2,191 windows of 100 kb across the seven base chromosomes. Phylogenetic matrices averaged 89,077 nucleotides per 100-kb window (Supplementary Table 5). For each window, we recorded the diploid species sharing the most recent common ancestor with each octoploid subgenome, and we further noted when this diploid was ‘sister’ to that subgenome. If the most recent common ancestor diploid was not sister to that subgenome, we labeled these as ‘clade’. For details, see Supplementary note, Methods.
of trees resolve the opposite species, or *F. nilgerrensis*, as the most closely related diploid. We suspect that these results can be attributed to incomplete lineage sorting.

In 81.4% of the 2,011 trees where subgenome A (Camarosa vesca) was resolved with the *F. vesca* clade, its most recent common ancestor or sister taxon was *F. vesca* subsp. *bracteata* (Supplementary Table 6), native to northwest North America. This is consistent with the PhyDS results of Edger et al., and with most previous studies. Likewise, our results and previous studies resolve subgenome B (Camarosa iinumae) with *F. iinumae* from Japan. Subgenomes C and D are most commonly resolved as sister to each other (Fig. 1b and Extended Data Fig. 1) and may represent an autotetraploid ancestor, although this requires cytogenetic confirmation. *Fragaria iinumae* is the closest diploid progenitor of these two subgenomes (Figs. 1c and 2), and may have originated with an extinct species or an unsampled population of *F. iinumae* (for example, from Sakhalin Island).

In their reply, Edger et al. suggest that, by conducting whole-genome alignment for phylogenetic analysis, we have introduced 'phylogenetic noise'. We concur that there is certainly some incorrectly aligned sequence in our matrix. However, the very consistent overall results across seven chromosomes and 2,191 phylogenetic analyses indicate to us that misalignment is producing a small amount of random signal that does not perturb what we hypothesize to be the actual evolutionary history of the octoploid strawberry.

Polyploidy is increasingly recognized for its role in generating novel diversity in organismal evolution, and the differential retention of diploid progenitors’ genes is central to this process. In particular, these diploid progenitors have each experienced a unique set of environments, pathogens and other challenges to survival and so, before coming together in a polyploid genome, each progenitor independently evolved ‘answers’ to a diverse array of abiotic and biotic conditions. Understanding the biology of these diploid species can inform polyploid plant adaptation to changing climate and associated environmental stress. For these reasons, it is critical to correctly identify the diploid progenitors of polyploids.

**Online content**

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-019-0543-3.

Received: 16 April 2019; Accepted: 4 November 2019; Published online: 16 December 2019

**References**

1. Edger, P. P. et al. Origin and evolution of the octoploid strawberry genome. *Nat. Genet.* **51**, 541–547 (2019).
2. Tennessen, J. A., Govindarajulu, R., Ashman, T.-L. & Liston, A. Evolutionary origins and dynamics of octoploid strawberry subgenomes revealed by dense targeted capture linkage maps. *Genome Biol. Evol.* **6**, 3295–3313 (2014).
3. Kamneva, O. K., Syring, J., Liston, A. & Rosenberg, N. A. Evaluating allopolyploid origins in strawberries (*Fragaria*) using haplotypes generated from target capture sequencing. *BMC Evol. Biol.* **17**, 180 (2017).
4. Yang, Y. & Davis, T. M. A new perspective on polyploid *Fragaria* (strawberry) genome composition based on large-scale, multi-locus phylogenetic analysis. *Genome Biol. Evol.* **9**, 3433–3448 (2017).
5. Liston, A., Cronn, R. & Ashman, T.-L. *Fragaria*: a genus with deep historical roots and ripe for evolutionary and ecological insights. *Am. J. Bot.* **101**, 1686–1699 (2014).
6. Njuguna, W., Liston, A., Cronn, R., Ashman, T.-L. & Basil, N. Insights into phylogeny, sex function and age of *Fragaria* based on whole chloroplast genome sequencing. *Mol. Phylogenet. Evol.* **66**, 17–29 (2012).
7. Rousseau-Gueutin, M. et al. Tracking the evolutionary history of polyploidy in *Fragaria* L. (strawberry): new insights from phylogenetic analyses of low-copy nuclear genes. *Mol. Phylogenet. Evol.* **51**, 515–530 (2009).
8. Wei, N., Tennessen, J. A., Liston, A. & Ashman, T.-L. Present-day sympatry belies the evolutionary origin of a high-order polyploid. *New Phytol.* **216**, 279–290 (2017).
9. Edger, P. P. et al. Reply to: Revisiting the origin of octoploid strawberry. *Nat. Genet.* https://doi.org/10.1038/s41588-019-0544-2 (2019).
10. Van de Peer, Y., Mizrachi, E. & Marchal, K. The evolutionary significance of polyploidy. *Nat. Rev. Genet.* **18**, 411–424 (2017).
11. Wei, N., Cronn, R., Liston, A. & Ashman, T.-L. Functional trait divergence and trait plasticity confer polyploid advantage in heterogeneous environments. *New Phytol.* **221**, 2286–2297 (2019).

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2019
Methods
Sequence reads derived from seven diploid *Fragaria* species, the outgroup *Potentilla micrantha*\(^2\) and the four *Fragaria ×ananassa* Camarosa subgenomes were individually mapped to the diploid *Fragaria vesca* v.4.1 genome assembly\(^3\), and subsequently combined into multiple sequence alignments for seven chromosomes and 2,191 100-kb windows for phylogenetic analysis with RAxML (v.8.2.12)\(^4\). An F\(_1\) cross of the hexaploid *F. moschata* was used for genetic linkage mapping and phylogenetic assignment of subgenomes with POLiMAPS (v.1.1)\(^5\). Hypothesis testing of alternative phylogenetic topologies was conducted with the Shimodaira–Hasegawa test in RAxML. Methods are fully described in the Supplementary note, Methods.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability
The sequence alignments and phylogenetic trees are available on Dryad: [https://doi.org/10.5061/dryad.ncjsxksqj](https://doi.org/10.5061/dryad.ncjsxksqj). The raw sequence data are available in the Sequence Read Archive under NCBI BioProject nos. PRJNA577462 and PRJNA576199. Biological samples are available from the authors on request.

Code availability
Custom perl and R scripts used to summarize and display the phylogenomic results are available at [https://github.com/jacobtennessen/FragariaGenome](https://github.com/jacobtennessen/FragariaGenome).

References
12. Buti, M. et al. The genome sequence and transcriptome of *Potentilla micrantha* and their comparison to *Fragaria vesca* (the woodland strawberry). *Gigascience* 7, giy010 (2018).
13. Edger, P. P. et al. Single-molecule sequencing and optical mapping yields an improved genome of woodland strawberry (*Fragaria vesca*) with chromosome-scale contiguity. *Gigascience* 7, gix124 (2018).
14. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).

Acknowledgements
The authors thank L. Longway and R. Cronn for laboratory assistance, K. Weitemier and T. Jennings for assistance with targeted sequence capture and K. Schuller, E. Jiang, A. Freundlich and the University of Pittsburgh glasshouse staff for plant cultivation. The OSU Center for Genome Research and Biocomputing conducted Illumina sequencing and provided computational support. This work was funded by the US National Science Foundation (no. DEB 1241217 to A.L. and no. DEB 1241096 to T.-L.A.) and the National Natural Science Foundation of China (no. 31261120580 to M.D. and J.L.).

Author contributions
*Fragaria moschata* linkage mapping was conducted by N.W. and T.-L.A. Phylogenetic analyses were performed by A.L., N.W. and J.T. Figures were prepared by J.T. and N.W. Sample collection and DNA sequencing in China was led by J.L. and M.D. All authors contributed to the writing and reviewing of the manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Extended data is available for this paper at [https://doi.org/10.1038/s41588-019-0543-3](https://doi.org/10.1038/s41588-019-0543-3). Supplementary information is available for this paper at [https://doi.org/10.1038/s41588-019-0543-3](https://doi.org/10.1038/s41588-019-0543-3). Correspondence and requests for materials should be addressed to A.L. Reprints and permissions information is available at www.nature.com/reprints.
Extended Data Fig. 1 | Maximum likelihood estimate of phylogeny for base chromosomes 2-7. Bootstrap values for all nodes are 100%, unless indicated.
Extended Data Fig. 2 | Maximum-likelihood phylogenies of base chromosomes 1-7 involving the maternal linkage groups (LGs) of the hexaploid Fragaria moschata. These LGs are named after species (‘Fmos’), chromosome (I-VII), maternal map (m) and LG number (1-3). Some chromosomes (e.g. 1, 5 and 6) have incomplete F. moschata LG numbers. Numbers associated with branches are ML bootstrap support values from 100 replicates.
Extended Data Fig. 3 | Maximum-likelihood phylogenies of base chromosomes 1–7 involving the paternal linkage groups (LGs) of the hexaploid *Fragaria moschata*. These LGs are named after species (‘Fmos’), chromosome (I–VII), paternal map (p) and LG number (1–3). Chromosome 1 has incomplete *F. moschata* LG numbers. Numbers associated with branches are ML bootstrap support values from 100 replicates.
Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
- Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted
- Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection The commercial DNA sequencing platform used in this study is fully described.

Data analysis All commercial and custom software used in this study for data analysis are fully described including specifying versions used.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The sequence alignments and phylogenetic trees are available on Dryad at https://doi.org/10.5061/dryad.ncjsxks0j. The raw sequence data are available in the Sequence Read Archive under NCBI BioProjects PRJNA577462 and PRJNA576199. Biological samples are available from the authors upon request.
Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences  
- Behavioural & social sciences  
- Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

| Sample size | Single individuals of six diploid Fragaria species were sequenced. Parents and 46 progeny of a Fragaria moschata artificial cross were sequenced. See Supplementary Table 2 for sample details. |
| Data exclusions | No data were excluded from analyses. |
| Replication | Bootstrapping for phylogenetic analyses were replicated 100 times. |
| Randomization | Randomization was not needed for this study. |
| Blinding | Blinding is not possible for phylogenomic analysis. |

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

| Materials & experimental systems | Methods |
| n/a | Involved in the study |
| ☑ | Antibodies |
| ☑ | Eukaryotic cell lines |
| ☑ | Palaeontology |
| ☑ | Animals and other organisms |
| ☑ | Human research participants |
| ☑ | Clinical data |
| n/a | Involved in the study |
| ☑ | ChiP-seq |
| ☑ | Flow cytometry |
| ☑ | MRI-based neuroimaging |