Early diversification and karyotype evolution of flowering plants

Pengchuan Sun
Sichuan University https://orcid.org/0000-0002-8999-4894
Yongzhi Yang
Lanzhou University https://orcid.org/0000-0001-6912-6718
Liang Liu
University of Georgia
Zhenxiang Xi
Sichuan University https://orcid.org/0000-0002-2851-5474
Hao Lin
Sichuan University
Landi Feng
Sichuan University
Jianxiang Ma
Lanzhou University
Hongyin Hu
Lanzhou University
Guiqi Bi
Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences
BeiBei Jiao
North China University of Science and Technology https://orcid.org/0000-0002-8755-6364
Shang Wang
Sichuan University
Jiao Yang
Lanzhou University
Wenjie Mu
Lanzhou University
Mingjia Zhu
Lanzhou University
Xiyin Wang
North China University of Science and Technology https://orcid.org/0000-0003-3454-0374
Zhenyue Wang
Lanzhou University
Charles Davis

Harvard University https://orcid.org/0000-0001-8747-1101

Jianquan Liu (liujq@nwipb.ac.cn)

Sichuan University https://orcid.org/0000-0002-4237-7418

Article

Keywords:

Posted Date: March 24th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1410884/v1

Abstract

Flowering plants (angiosperms) dominate our planet and sustain all life on Earth. However, evolutionary relationships among the angiosperm lineages that diverged early – Amborellales, Nymphaeales, Austrobaileyales and Mesangiospermae, which further comprises monocots and other four clades – have remained highly disputed likely because of their rapid diversification promoted by an ancestral polyploidization event. Here, we present high-quality chromosomal-level genome assemblies of two species – star anise (Illicium verum: 12.5 Gb with a large size) in the Austrobaileyales and calamus (Acorus gramineus) representing the sister lineage to all other extant monocots. These two genomes filled the final gaps of all genomic representatives for major angiosperm lineages. Our phylogenetic analyses of collinear genes support Amborellales and Nymphaeales as sister lineages and they, together with Austrobaileyales and monocots are successively sister to other Mesangiospermae clades. Based on chromosome-like synteny blocks shared between extant genomes, we constructed the ancestral angiosperm karyotype to be x=16. We independently established evolutionary relationships for all sampled species based on shared polyploidizations and chromosomal fusions from the common ancestral karyotype. This phylogenetic relationship is congruent with that from collinear genes, especially the finding that Amborellales and Nymphaeales shared a rare chromosomal fusion not seen in other angiosperms. These results advance our understanding and shed new lights on early diversification and karyotype evolution of flowering plants.

Full Text

Flowering plants (angiosperms) dominate our planet and comprise more than 350,000 extant species[1]. They support terrestrial ecosystems through cycling of oxygen, water and carbon and further provide nearly all necessities for human survivals, including foods, building materials, clothing and medicines. However, the rapid rise and early diversification of angiosperms has remained “an abominable mystery” since the time of Darwin[21] and estimates of the ancestral chromosome base number for angiosperms vary from x=5 to x=9, although x=7 is mostly likely[22]. Increasing molecular phylogenetic analyses of extant angiosperms have identified four major lineages: Amborellales, Nymphaeales and Austrobaileyales as the ANA grade and a core angiosperm lineage (Mesangiospermae) that includes the vast majority of the remaining flowering plant diversity[9-19]. Mesangiospermae is further comprised of five well-circumscribed clades: monocots, eudicots, magnoliids, Chloranthales, and Ceratophyllales[2]. However, evolutionary relationships between these lineages and Mesangiospermae clades remain highly debated or inconsistent based on different lines of evidence[3-18]. For example, Amborellales alone[9] or Amborellales and Nymphaeales together[10] have been identified as a sister to other angiosperms. In addition, within Mesangiospermae, monocots have been identified to be sister to eudicots, magnoliids or the remaining four clades[12-19]. These conflicting results may have been the result of the rapid diversification of early angiosperms promoted by one ancestral whole-genome duplication (WGD) (tetraploidization)[20]. Both incomplete lineage sorting (ILS) and hybridization may have been involved in radiative divergences between ancestors of these lineages and

clades within a short timescale[19]. In addition, further independent polyploidization including both WGD and whole-genome triplication (WGT, hexaploidization) for each lineage or clade during their later long evolutionary histories may have complicated such evolutionary relationships[19]. However, chromosome-level genome sequences undoubtedly outweigh other evidence for resolving these uncertainties in two ways. First, the highly collinear and orthologous genes with all copies retained after WGD or WGT can be extracted across representative genomes for phylogenetic analyses. These genes may be superior to other homologs for phylogenetic constructions, for example, the single-copy or clustering-identified homologous genes[14-17,19], for which paralogs with heterogeneous evolution after polyploidization may be randomly retained and ILS was modeled to occur more frequently during such phylogenetic analyses[23]. Second, the key and large genomic variations[24], which rarely occur but may be evolutionarily conserved, for example, chromosomal fusions following the ancestral polyploidization, may be retained in the extant genomes and can be used to reflect phylogenetic relationships. In the past decade, chromosome-level genomes have been published for most early-diverged angiosperm lineages and clades[9,14-16,19]. To date, however, no high-quality genome is available for Austrobaileyales in which most extant species have especially large genomes. There is also a lack of any chromosome-level genome for Acorales, a sister lineage to all other monocots[11,12]. The genomic gaps with respect to these two key representative nodes hamper our understanding of phylogenetic relationships and karyotype evolution of the early-diverged angiosperms.

Chromosome-level genomes of two key angiosperms

We report here the chromosome-level genome assemblies of *Illicium verum* representing Austrobaileyales and *Acorus gramineus* representing the earliest-diverged monocot family (Acoraceae), derived using the PacBio HiFi technologies and Hi-C approaches. The assembled *I. verum* genome is 12.5 Gb (contig N50 of 10.68 Mb) and 46.6% could be anchored onto 14 pseudo-chromosomes (Extended Data Fig. 1 and Extended Data Table 1). The genome size of this species is obviously higher than that of any other species from the early-diverged angiosperm lineages[9,14-16,19]. The extensive amplification of repetitive elements in the recent past has resulted in the extreme increase in genome size of this species (Supplementary Note 1). For *A. gramineus*, the assembled genome was 381 Mb (contig N50 of 35.6 Mb) and 100% could be anchored onto 12 pseudo-chromosomes (Extended Data Fig. 1 and Extended Data Table 1). Both genomes showed high contiguity, completeness and accuracy (Supplementary Note 2). We further predicted 47,041 and 23,942 protein-coding genes in *I. verum* and *A. gramineus*, respectively, and both showed a high BUSCO score (> 92.9%).

Polyploidization and phylogenetic analyses

We used chromosome-level genomes from three ANA-grade lineages and five Mesangiospermae clades for further analyses. A total of 17 species were selected, three ANA-grade species, seven monocots, two eudicots, three magnoliids, *Ceratophyllum demersum, Chloranthus sessilifolius*, and five species used as outgroups (Supplementary Table 2). As polyploidization events are common in angiosperms, in particular one common WGD occurred before the origin of angiosperms[20], we first examined the respective polyploidization histories of each selected angiosperm species based on gene collinearity analyses. Both *I. verum* and *A. gramineus* experienced an independent WGD event (Extended Data Figs. 2f and Supplementary Note 4.1). The distribution of synonymous substitutions per synonymous site (Ks) was further employed to date the occurrence of these two evolutionary events. Two obvious Ks peaks at 0.57 and 0.54 represent 62-70 and 59-67 million years ago (Ma) based on the evolutionary rate correction in *I. verum* and *A. gramineus*, respectively, close to the mass extinction of the Cretaceous-Paleogene stage (Supplemental Note 4.2). Further independent or shared WGDs or WGTs were confirmed for every other species except *Am. trichopoda*[9] and *Ar. fimbriata*[17].

We used the *Am. trichopoda* genome without extra WGD as a reference to assign the different copies of each species originating from polyploidy as different subgenomes based on synteny analyses. We used all collinear genes that were still retained after WGD or WGT for phylogenetic analyses. Only genes that had a collinear relationship in all three ANA species and at least one eudicot, one monocot, one magnoliid ortholog, and a best hit in at least one outgroup species were retained. A total of 1235 collinear genes were retrieved to construct the species tree based on the coalescent method, and the results supported that Amborellales and Nymphaeales were sisters and together were a sister group to the other species, including Austrobaileyales and Mesangiospermae. Gene trees also showed largely discordant topologies (Supplemental Note 5.1), as found previously based on the other single-copy or clustering-identified homologous genes[14-17,19]. Internal branch lengths between three ANA-grade species were very short, especially between *Am. trichopoda* and *N. colorata* on both coalescent species tree (Fig. 1) and concatenated tree (Supplemental Note 5.1). These findings together reflect rapid radiation of three early-diverged angiosperm lineages (the ANA grade). Within Mesangiospermae, coalescent analyses (Fig. 1) suggested that monocots were sister to the other four clades, which comprised two respective sister groups, magnoliids+Chloranthales and Ceratophyllales+eudicots. However, tree topologies of these five clades were also discordant based on different collinear genes and internal branch lengths between them were short (Fig. 1, Supplemental Note 5.1) because of the radiative diversification[14-17,19]. However, within the monocots, two species, *A. gramineus* (Acoraceae) and *S. polyrhiza* (Alismataceae), are always successively sister to other monocots. In addition, this synteny-based tree also clearly mirrored the shared or independent polyploidization histories of these species, for example, the τ event was shared by most monocots except for Acorales and Alismatales. Within magnoliids, Magnoliales and Laurales shared one WGD. Finally, we also recovered internal branch lengths and discordant topologies for three early-diverged lineages and five Mesangiospermae clades by single-copy gene datasets, similarly indicating rapid radiation of these lineages and clades (Supplementary Note 5.2).

Ancestral angiosperm karyotype, chromosomal fusion, and evolutionary relationship

The protochromosomes and ancestral karyotype for a single lineage can be reconstructed for both plants and animals and the resultant chromosomal evolution is widely used for inferring or confirming phylogenetic relationships[24-30]. The accuracy of such a karyotype reconstruction relies critically on the representative chromosome-level genomes[30]. In addition, diploidization diversification after an ancestral polyploidization event usually involves chromosomal fusions[27,28]. Such a genomic signature may be retained in the diploidized species for a long time[24]. The protochromosomes can be entirely nested within one fused chromosome of the extant genomes as one 'intact block' or retained as one independent chromosome even after species-specific WGD or WGT[24] (Supplementary Table 20). The chromosomal fusions involve three basic patterns: reciprocally translocated chromosome arms (RTA), end-end joining (EEJ) and nested chromosome fusion (NCF)[25-28,31] (Fig. 2, Extended Data Fig. 2 and Fig. 3a) (Extended Data Fig. 3a). We extracted protochromosomes and constructed the ancestral angiosperm karyotype (AAK) based on the shared chromosome-like 'intact blocks' or independent chromosomes between multiple species. We selected extant chromosome-level genomes of 17 representative species from the early-diverged lineages and identified 16 intact protochromosomes (Fig.3), each of which was retained as the independent chromosomes or chromosome-like 'synteny blocks' across at least three species (Supplementary Note 6.3). The chromosomes of all extant genomes were composed of these protochromosomes through repeated chromosomal fusions of the three types (Fig.2) and further fissures with syntenic alignments. This AAK, constituting 16 protochromosomes, differs from the one derived previously[30] mainly because of adding high-quality genomes for more representative lineages and clades and adopting a different method for reconstructing the ancestral karyotype.

We determined evolutionary relationships of extant genomes based on three principles: (1) the shared WGD or WGT, (2) the shared chromosomal fusion types from the AAK and (3) the shortest evolutionary pathway from the AAK. We independently and tentatively established evolutionary relationships for 17 genomes and illustrated chromosomal fusions in each node (Fig. 3). This relationship is basically consistent with that inferred from collinear genes, especially between Amborellales, Nymphaeales, Austrobaileyales, monocots and four other Mesangiospermae clades (Fig. 1). For example, from the AAK for all angiosperms after the common polyploidization event, we found that one chromosome (Chr1=AAK1+2) of *Am. trichopoda* (Amborellales) was formed by AAK1 and AAK2 through NCF. This NCF event, or the resultant AAK1+2 chromosome at the same genomic location and breaking point, was only shared by two species (*N. colorata* and *Euryale ferox*) of Nymphaeales (Extended Data Figs. 4). Such a chromosomal event is probably derived from the common ancestor, therefore supporting the sister relationship of these two lineages as inferred previously (Fig.1). In their ancestral node, 16 protochromosomes of AAK were therefore reduced to 15 through this fusion. Two species of Nymphaeales retained more because of the later species-specific polyploidizations.

In the other direction, we identified that the fusion of AAK7 and AAK16 through the end-end joining (EEJ), formed one chromosome (AAK7+16) before the divergence of Austrobaileyales and Mesangiospermae, with a similar reduction of x=16 to x=15 chromosomes. In *I. verum* (Austrobaileyales), this chromosome became two (Chrs 5 and 13) because of specific WGD (Extended Data Figs. 5). In the common ancestor of Mesangiospermae, we identified that two AAK chromosomes (AAK1 and AAK9) further fused into one more chromosome through RTA. However, both AAK1+9 and AAK7+16 chromosomes were fragmented at the same positions in the remaining Mesangiospermae species, with the exception of the monocots (Fig. 3 and Extended Data Figs. 6). In addition, these fragmented chromosomes were able to further fuse with other protochromosomes during the following karyotype evolution in these Mesangiospermae clades. For instance, a part of the ancestral AAK7+16 was found to be connected with a part of AAK14 in one chromosome of *Ar. fimbriata* (Extended Data Fig. 7 and Supplementary Note 6.4).

In contrast, the AAK1+9 and AAK7+16 chromosomes were retained intact in all extant monocots although with further fusions with other protochromosomes in more recently diverged species. In addition to these two chromosomes, the earliest-diverged *A. gramineus* retained more Mesangiospermae primitive chromosomes even after species-specific WGD and further chromosomal fusions. However, the common ancestor of the remaining monocots (ancestral monocot karyotype except for Acoraceae, AMKA) experienced repeated chromosomal fusions involving 12 primitive Mesangiospermae chromosomes, reducing the haploid chromosome number from x=14 to x=8 (Fig.2 and Supplementary Note 6.5). For example, AAK7+16 further fused with AAK14 through EEJ to produce an AAK7+16+14 chromosome. This connection is similar to that found in *Ar. fimbriata*[17] but with a totally different evolutionary history (Extended Data Fig. 7). After the divergence of *S. polyrhiza* (Alismatales), the τ tetraploidization (WGD) and two further AMKA chromosomal fusions (AMKA1+7 and AMKA1+8) (Extended Data Figs. 9b,c) resulted in one ancestral karyotype with x=14 for the remaining monocots. Four retained primitive chromosomes (AMKA2+7 and AMKA2+8) from this karyotype further fused into two chromosomes respectively before the σ hexaploidization (WGT) in the common ancestors of *A. comosus* and *Oryza sativa* (Extended Data Figs. 10b). This genomic signature was obviously retained in the later polyploidization events, for example, six shares after WGT in *O. sativa*. Therefore, such strong signatures within the chromosomal fusions and structural variations could be used as genomic markers to accurately identify the polyploidy levels of the lately evolved monocots including banana[32] and orchid[33] and even correct the wrong genome assembly (Extended Data Figs. 10).

Discussion

Our coalescent phylogenetic analyses of the accurately collinear genes including all paralogs due to lineage-specific polyploidizations support Amborellales and Nymphaeales are sisters and they together with Austrobaileyales and monocots are successively sister to other Mesangiospermae clades. This result differs from most previous phylogenetic analyses that Amborellales is sister to the other angiosperms based on single-copy or clustering-identified homologous genes[9-11,13,14]. However, our phylogenetic analyses of each collinear gene also suggest the widespread discordant tree topologies and

short internal branch lengths in differentiating three lineages of the ANA grade, Amborellales, Nymphaeales and Austrobaileyales and five Mesangiospermae clades (Fig. 1). Therefore, both statistical errors and ILS could not be totally avoided during reconstructing phylogenetic relationships of these lineages and clades with rapid radiations based on collinear genes. The evolutionary relationships between them and alternative divergence patterns remain to be examined based on homologous genes.

It is interesting that we identified a shared chromosomal fusion event by Amborellales and Nymphaeales, which may further support their sister relationship. Chromosomal fusions, as the major macrostructural genomic changes, may be homoplasious or reversed during the long evolutionary history of diverse plants and animals[34,35]. However, the rapid radiation of only three early angiosperm lineages within a very short timescale, as indicated by various analyses, seems to reduce such a probability for the recovered shared NCF event by Amborellales and Nymphaeales having occurred although we could not exclude it totally. In addition, this event may have occurred prior to their common ancestral stem with Austrobaileyales together and have been retained as one ancestral rather than synapomorphic trait for Amborellales and Nymphaeales. Under this scenario, Amborellales would have diverged first, followed by Nymphaeales, as found by most phylogenetic analyses[11-14], but together with this NCF signature. However, one reverse evolution and another chromosomal EEJ fusion have to be employed for explaining chromosomal fusion of Austrobaileyales and Mesangiospermae. Such a non-parsimonious scenario seems to conflict with the mostly assumed hypothesis for phylogenetic analyses on all morphological traits and sometimes sequence variations[36]. In addition, ancient polymorphisms in macrostructural genomic variations in the ancestral species may also lead to ILS when using such a trait for phylogenetic inference. However, chromosomal fusions usually result in rapid speciation with reproductive isolation[27,28,37,38], therefore reducing such a possibility. Despite this, further evidence is needed to exclude this likeliness during rapid radiation of three ANA lineages.

The ancestral karyotype of all angiosperms with 16 protochromosomes (x=16) that we reconstructed here seems to suggest that it was derived from the more primitive haploid chromosome number x=8 through the ancestral WGD. Such a chromosome base number differs from previous suggestions x=7[22]. However, it is consistent with the extant chromosome numbers of three early-diverged lineages Amborellales (x=15), Nymphaeales (x=15) and Austrobaileyales (x=15) because of the chromosomal fusions during the subsequent diploidization after the common WGD. As noted before, both Amborellales and Nymphaeales (x=15) share one rarely evolved chromosomal fusion, suggesting their direct origin from the ancestral karyotype x=16 and both together as the earliest diverged sister group to the remaining angiosperms. Our results based on chromosomal evolution further support that the monocot lineage was sister to the other four Mesangiospermae clades[13], rather than sister to eudicots or magnoliids[12,14-18]. In addition, we also found that more protochromosome-like blocks were retained in the calamus genome, suggesting the early divergence of the calamus family from the other monocots. This finding parallels to the similar conclusion based on the genome sequence of another congeneric species[40] (Supplementary Note 6.4). These results shed sight into the early evolution of angiosperms and monocots and will, we hope, inspire future researchers into genome and karyotype evolution in other angiosperm lineages

through reconstructing ancestral karyotypes. Our results also indicate that it is better to employ both Amborellales and Nymphaeales together as the earliest diverged lineages to trace the evolution of both traits and genes in flowering plants.

Declarations

Reporting summary

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

Data availability

Newly generated genome assembly and annotation are available (https://figshare.com/s/5894193691c9a047fb34). The main custom scripts have been deposited in Github (https://github.com/SunPengChuan/Angiosperm-karyotype-evolution).

Acknowledgements

We thank Profs. Jim Leebens-Mack (University of Georgia, USA) and Yves Van de Peer (Ghent University, Belgium) for constructive comments on the manuscript. This work was supported equally by the Strategic Priority Research Program of Chinese Academy of Sciences (XDB31000000) and the Second Tibetan Plateau Scientific Expedition and Research (STEP) program (2019QZKK0502), and further by National Key Research and Development Program of China (2017YFC0505203), by the National Natural Science Foundation of China (grant numbers 31590821, 91731301 and 31561123001), Fundamental Research Funds for the Central Universities (YJ201936, SCU2019D013 and 2020SCUNL20). and National High-Level Talents Special Support Plans.

Author contributions

J.L. designed the experiments and coordinated the project. J.L. and Y.Y. performed fieldwork and collected samples. P.S., Y.Y., Z.X., J.M., G.B., H.L. and L.F. performed the assembly of the two genomes. P.S., Y.Y., J.M., J.Y., W.M., M.Z. and H.H. carried out the repeat and gene annotations. P.S., Y.Y., Z.X., J.M., H.L. and L.F. performed the polyploidization analysis. Y.Y., P.S. and Z.X. carried out the gene family analysis and the phylogenomic analysis. P.S., X.W., B.J., S.W., Z.W. and L.F. performed the karyotype analysis. P.S., Y.Y., L.L., H.L. and L.F. finished karyotypic evolution and phylogenetic analyses. J.L., P.S., C.D. and Y.Y. wrote and edited the manuscript. All of the authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper.

References

1 Kress, W. J. *et al.* Green plant genomes: What we know in an era of rapidly expanding opportunities. *Proceedings of the National Academy of Sciences* **119** (2022).

2 Chase, M. W. *et al.* An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Botanical Journal of the Linnean Society* **181**, 1-20 (2016).

3 Crane, P. R., Friis, E. M. & Pedersen, K. R. The origin and early diversification of angiosperms. *Nature* **374**, 27-33, doi:10.1038/374027a0 (1995).

4 Qiu, Y.-L. *et al.* The earliest angiosperms: evidence from mitochondrial, plastid and nuclear genomes. *Nature* **402**, 404-407, doi:10.1038/46536 (1999).

5 Williams, J. H. & Friedman, W. E. Identification of diploid endosperm in an early angiosperm lineage. *Nature* **415**, 522-526, doi:10.1038/415522a (2002).

6 Friedman, W. E. Embryological evidence for developmental lability during early angiosperm evolution. *Nature* **441**, 337-340, doi:10.1038/nature04690 (2006).

7 Friedman, W. E. Hydatellaceae are water lilies with gymnospermous tendencies. *Nature* **453**, 94-97, doi:10.1038/nature06733 (2008).

8 Goremykin, V. V. *et al.* The Evolutionary Root of Flowering Plants. *Syst. Biol.* **62**, 50-61, doi:10.1093/sysbio/sys070 (2012).

9 Albert, V. A. *et al.* The Amborella genome and the evolution of flowering plants. *Science* **342**, 1241089 (2013).

10 Xi, Z., Liu, L., Rest, J. S. & Davis, C. C. Coalescent versus concatenation methods and the placement of Amborella as sister to water lilies. *Syst. Biol.* **63**, 919-932 (2014).

11 Leebens-Mack, J. H. *et al.* One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* **574**, 679-685, doi:10.1038/s41586-019-1693-2 (2019).

12 Li, H.-T. *et al.* Origin of angiosperms and the puzzle of the Jurassic gap. *Nat. Plants* **5**, 461-470, doi:10.1038/s41477-019-0421-0 (2019).

13 Yang, L. *et al.* Phylogenomic insights into deep phylogeny of angiosperms based on broad nuclear gene sampling. *Plant communications* **1**, 100027 (2020).

14 Zhang, L. *et al.* The water lily genome and the early evolution of flowering plants. *Nature* **577**, 79-84, doi:10.1038/s41586-019-1852-5 (2020).

15 Guo, X. *et al.* Chloranthus genome provides insights into the early diversification of angiosperms. *Nature Communications* **12**, 6930, doi:10.1038/s41467-021-26922-4 (2021).

16 Ma, J. *et al.* The Chloranthus sessilifolius genome provides insight into early diversification of angiosperms. *Nature Communications* **12**, 6929, doi:10.1038/s41467-021-26931-3 (2021).

17 Qin, L. *et al.* Insights into angiosperm evolution, floral development and chemical biosynthesis from the Aristolochia fimbriata genome. *Nat. Plants*, doi:10.1038/s41477-021-00990-2 (2021).

18 Soltis, P. S., Soltis, D. E. & Chase, M. W. Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology. *Nature* **402**, 402-404, doi:10.1038/46528 (1999).

19 Yang, Y. *et al.* Prickly waterlily and rigid hornwort genomes shed light on early angiosperm evolution. *Nat. Plants* **6**, 215-222 (2020).

20 Jiao, Y. *et al.* Ancestral polyploidy in seed plants and angiosperms. *Nature* **473**, 97-100, doi:10.1038/nature09916 (2011).

21 Friedman, W. E. The meaning of Darwin's “abominable mystery”. *American Journal of Botany* **96**, 5-21 (2009).

22 Carta, A., Bedini, G. & Peruzzi, L. A deep dive into the ancestral chromosome number and genome size of flowering plants. *New Phytologist* **228**, 1097-1106, doi:https://doi.org/10.1111/nph.16668 (2020).

23 Yan, Z., Smith, M. L., Du, P., Hahn, M. W. & Nakhleh, L. Species Tree Inference Methods Intended to Deal with Incomplete Lineage Sorting Are Robust to the Presence of Paralogs. *Syst. Biol.*, doi:10.1093/sysbio/syab056 (2021).

24 Rokas, A. & Holland, P. W. Rare genomic changes as a tool for phylogenetics. *Trends in ecology & evolution* **15**, 454-459 (2000).

25 Hill, J. *et al.* Unprecedented reorganization of holocentric chromosomes provides insights into the enigma of lepidopteran chromosome evolution. *Science Advances* **5**, eaau3648, doi:doi:10.1126/sciadv.aau3648 (2019).

26 Graphodatsky, A. S., Trifonov, V. A. & Stanyon, R. The genome diversity and karyotype evolution of mammals. *Molecular Cytogenetics* **4**, 22, doi:10.1186/1755-8166-4-22 (2011).

27 Stebbins, G. L. Chromosomal evolution in higher plants. *Chromosomal evolution in higher plants.* (1971).

28 Grant, V. *Plant speciation*. Columbia university press, (1981).

29 Simakov, O. *et al.* Deeply conserved synteny resolves early events in vertebrate evolution. *Nature Ecology & Evolution* **4**, 820-830, doi:10.1038/s41559-020-1156-z (2020).

30 Murat, F., Armero, A., Pont, C., Klopp, C. & Salse, J. Reconstructing the genome of the most recent common ancestor of flowering plants. *Nat. Genet.* **49**, 490-496, doi:10.1038/ng.3813 (2017).

31 WANG, Z. & WANG, X. Evolutionary genomics model of chromosome number reduction and B chromosome production. *SCIENTIA SINICA Vitae* **50**, 524-537 (2020).

32 D'Hont, A. *et al.* The banana (Musa acuminata) genome and the evolution of monocotyledonous plants. *Nature* **488**, 213-217, doi:10.1038/nature11241 (2012).

33 Zhang, Y. *et al.* Chromosome-scale assembly of the Dendrobium chrysotoxum genome enhances the understanding of orchid evolution. *Horticulture Research* **8**, 183, doi:10.1038/s41438-021-00621-z (2021).

34 Boore, J. L. The use of genome-level characters for phylogenetic reconstruction. *Trends in Ecology & Evolution* **21**, 439-446 (2006).

35 Simakov, O. *et al.* Deeply conserved synteny and the evolution of metazoan chromosomes. *Science Advances* **8**, eabi5884, doi:doi:10.1126/sciadv.abi5884 (2022).

36 Avise, J. C. *Evolutionary pathways in nature: a phylogenetic approach*. (Cambridge University Press, 2006).

37 Yin, Y. *et al.* Molecular mechanisms and topological consequences of drastic chromosomal rearrangements of muntjac deer. *Nature Communications* **12**, 6858, doi:10.1038/s41467-021-27091-0 (2021).

38 Luo, J., Sun, X., Cormack, B. P. & Boeke, J. D. Karyotype engineering by chromosome fusion leads to reproductive isolation in yeast. *Nature* **560**, 392-396, doi:10.1038/s41586-018-0374-x (2018).

39 Acorus americanus v1.1, DOE-JGI, http://phytozome-next.jgi.doe.gov/.

Methods

Polyploidization and phylogenetic analysis

In synteny analyses, collinear genes were identified with the parameter '-icl' of WGDI within each genome and between genomes, and the collinear gene dotplots were used to calculate the syntenic ratios between different species to confirm the polyploidy level of each species. Frequencies of synonymous substitutions per synonymous site (Ks values) between collinear genes were estimated using the Nei-Gojobori approach as implemented in PAML through WGDI (-ks). The median Ks values of each block were selected to perform Ks peak fitting by WGDI (-pf). We then determined the shared and species-specific polyploidizations (WGD or WGT). We further divided the synteny blocks into different subgenomes according to the Ks distribution and the syntenic depth ratios within and between genomes.

For example, if species A experienced two independent WGDs compared to species O, species A therefore had four subgenomes, which were named A1, A2, A3, and A4. We adjusted the order of subgenomes according to the recent WGD to ensure that their phylogenetic relationships matched ((A1, A2), (A3, A4)). Then, we recorded the subgenome regions on the chromosomes and used WGDI (-pc, -a) to obtain the hierarchical gene list.

The hierarchical gene lists were used to infer maximum likelihood (ML) trees by IQ-TREE with automatic selection of the best-fit substitution model (-m MFP) through WGDI (-at). ASTRAL is able to calculate the frequencies of collinear genes trees that support independent WGD occurring in each paired species. So, we used ASTRAL-III v.5.7.7 with the parameter of "-t 16" to construct the coalescent tree and estimate branch support.

Identification of all protochromosomes for all angiosperms and ancestral angiosperm karyotype (AAK)

Diploidization after polyploidization usually occurs through chromosomal fusions. However, protochromosomes may be ultimately retained intact in the fused or independent chromosomes in the diploidized offspring lineages. The chromosomal fusions usually follow three basic patterns (Figs. 3a, Supplementary Note 6.1): reciprocally translocated chromosome arms (RTA), end-end joining (EEJ) and nested chromosome fusion (NCF). If one protochromosome nests within the fused chromosome through such a pattern, it can be still considered as the 'intact' protochromosome in the extant genomes. In addition, repeated fusions and fissions of one protochromosome should produce multiple syntenic blocks, which could be further identified in the extant genomes. Due to the role of telomeres in protecting chromosome integrity, chromosomes with telomeres should be firstly considered to be intact. All chromosomes assembled by default for the extant genomes have telomeres. Therefore, all protochromosomes comprising the AAK should be present in the extant genomes of the basal angiosperms and could be extracted across multiple representative species.

We used WGDI with the parameter "-d" to align 17 genomes and plotted homologous dotplots within each genome and between genomes. We searched for chromosome-like 'synteny blocks' with telomeres (including independent chromosomes) across all genomes using the *Am. trichopoda* genome without WGD as a reference. Only one chromosome-like intact 'synteny block' (independent chromosome or being retained as one of three chromosomal fusions) identified for at least three species was assumed as the protochromosome. We used the most intact one, for example, one independent chromosome in one species, across all extant genomes to represent this protochromosome. When this 'synteny block' was extracted from one fused chromosome of one extant genome, we connected the other parts together as one entire chromosome or kept the remaining part as one chromosome if this 'synteny block' was at one end. When such a 'protochromosome' was fragmented into multiple 'syntenic' parts, which were fully

deleted and the other parts were further connected together as one ‘integrated’ chromosome for the next circle of the protochromosome identification. We started the second search cycle for the next protochromosome using the same approach. All protochromosomes were assumed to have been fully recovered when no more genomic blocks were found for all species. Ultimately, we extracted 16 protochromosomes, which were together assumed to comprise the AAK, after which no more genomic blocks were found for each extant genome. We numbered these protochromosomes with different colors.

Karyotype changes from the AAK and construction of evolutionary relationships

We first compared each of 17 extant genomes with the AAK and determined its karyotype composition from the protochromosomes based on the collinearity fragments using WGDI with the parameter “-km”. Then, we compared the permutations and combinations of protochromosome color patches and inferred all chromosomal fusions and evolutionary patterns.

We clustered the species with the common WGD or WGT as the respective monophyletic groups. We manually determined evolutionary relationships for all 17 species based on the shared chromosomal fusion types with the ordered changes from the AAK with the shortest evolutionary steps, as assumed for most phylogenetic analyses. We illustrated all chromosomal fusions on the ancestral node. We left unsolved relationships for some species because we could not find shared chromosomal fusion types due to a lack of sufficient representative chromosome-level genomes.

Extended Data Table 1

Extended Data Table 1 is not available with this version.

Figures

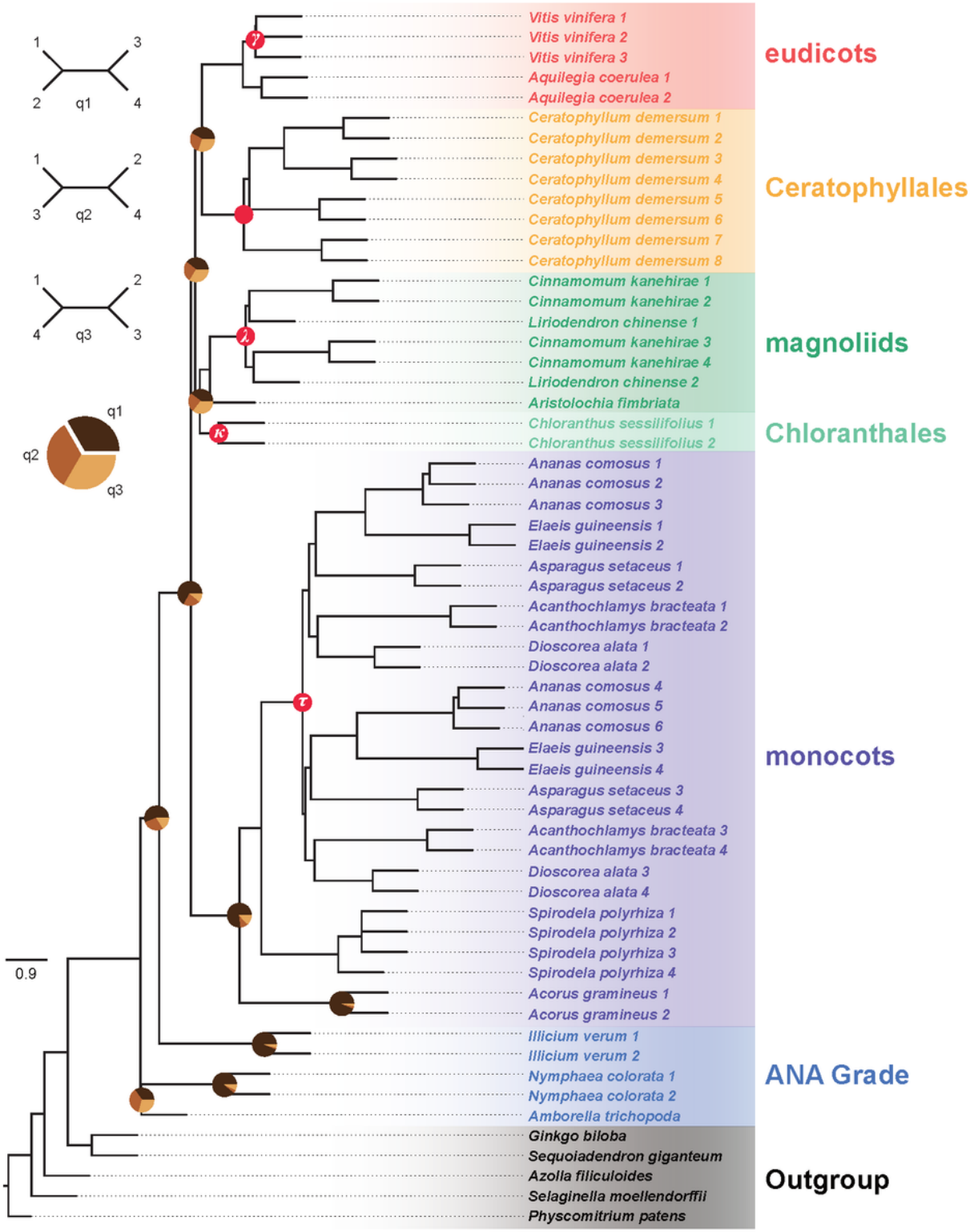

Figure 1

Phylogenetic relationships between the early-diverged angiosperm lineages and clades. The tree was constructed based on the collinearly retained genes after lineage-specific polyploidization events. The frequency of three topologies (q1–q3, sector representations in different colors) around focal internal branches of ASTRAL species trees in colinear genes.

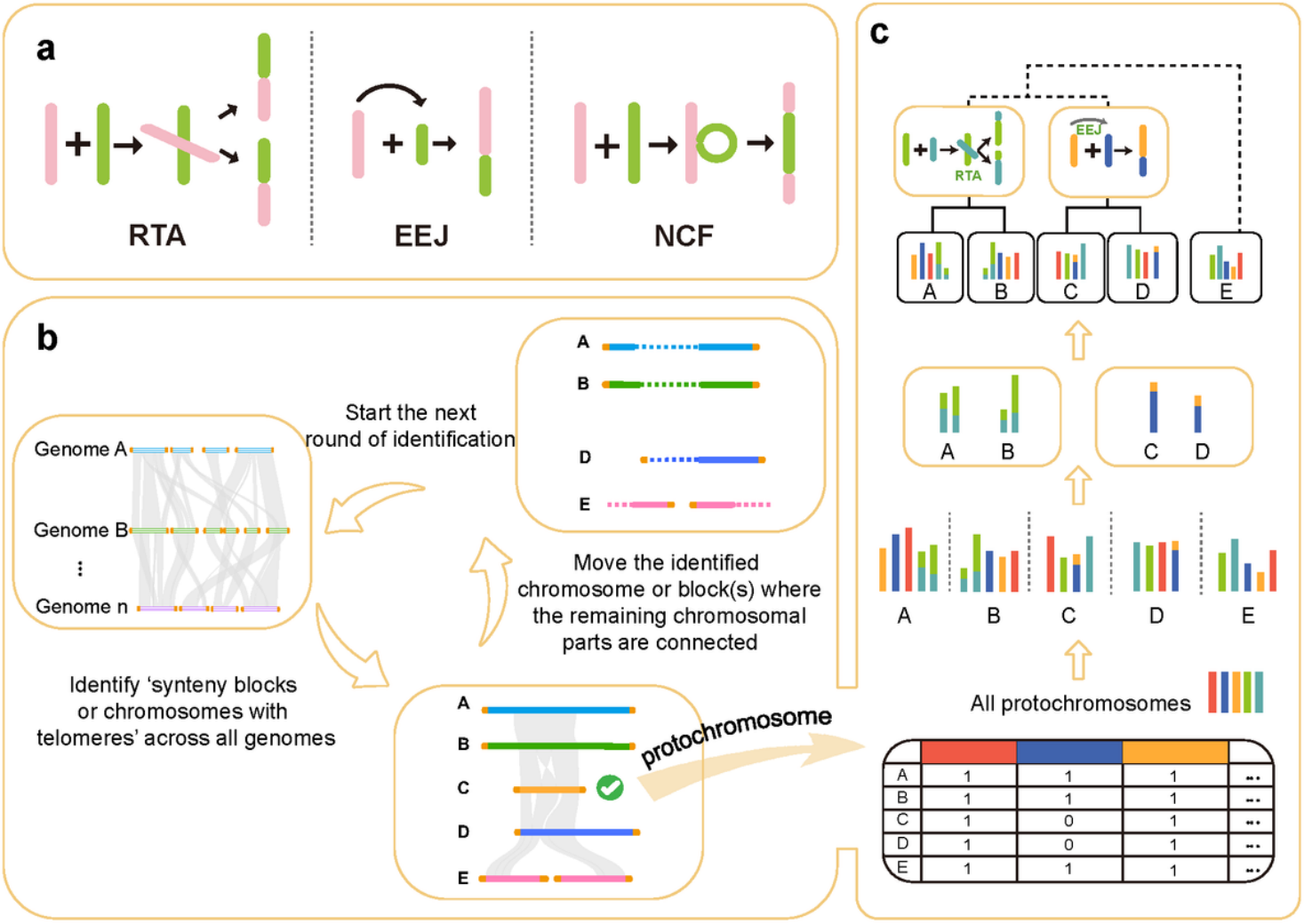

Figure 2

The methods to identify ancestral karyotypes and construct evolutionary relationships. a, Karyotype and aneuploidy evolution during diploidization after a common polyploidization event involving three basic types of chromosomal fusion: reciprocally translocated chromosome arms (RTA), end-end joining (EEJ) and nested chromosome fusion (NCF). B, Four steps were undertaken to construct the ancestral karyotype. First, identifying chromosome-like 'synteny blocks' and small syntenic blocks across all sampled genomes. Second, exploring the shared 'synteny block' with telomeres (chromosome-like) across extant genomes and extracting the most intact of these (for example, one chromosome) as one protochromosome. Third, deleting this shared block and its syntenic small blocks across all extant genomes and connecting the remaining parts together as 'entire chromosomes'. Fourth, starting more rounds of 'synteny exploration' to extract all protochromosomes until no genomic block was left for each extant genome. C. All protochromosomes of the ancestral karyotype are compared with each extant genome and the karyotypic composition from the protochromosomes is determined. The karyotypic changes are further inferred and used for constructing evolutionary relationships.

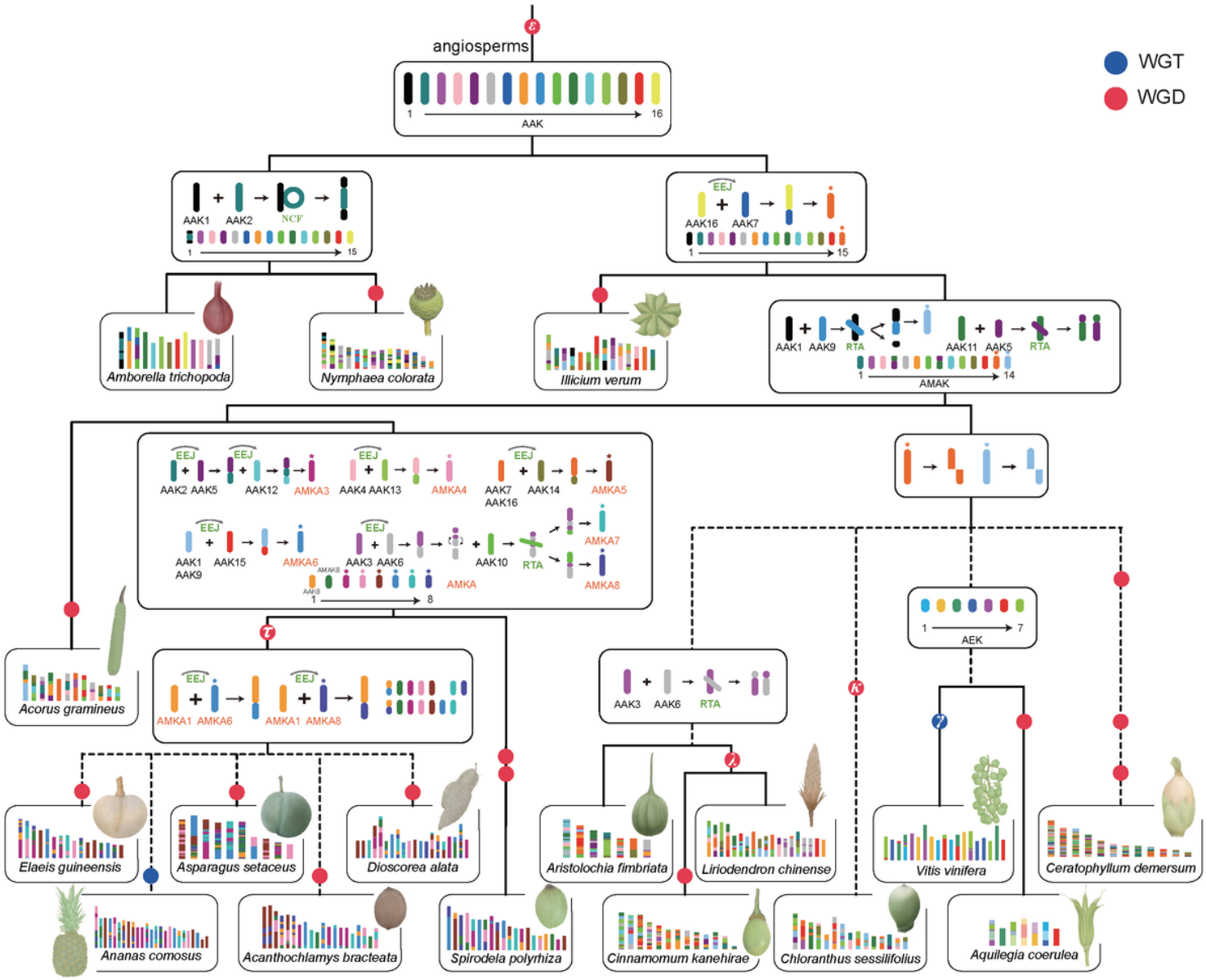

Figure 3

Evolutionary relationship between early flowering plants, constructed independently based on the shared polyploidizations and shared chromosomal fusions from the ancestral angiosperm karyotype (AAK). The reconstructed AAK for all flowering plants comprises 16 protochromosomes (x=16) after the

polyploidy event. The evolutionary relationships were constructed based on the shared polyploidy events and chromosome fusions from the AAK with the shortest steps. All shared karyotypic changes are marked on the ancestral tree nodes. Polyploidization events are shown as dots (red, tetraploidization or whole-genome duplication (WGD); blue, hexaploidization or whole-genome triplication (WGT)). The dotted lines indicate the unsolved relationships due to the lack of sufficient representative chromosome-level genomes of the closely related species. AEK ancestral eudicots karyotype. AMKA: ancestral monocot karyotype except for Acoraceae.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryNotes.docx
- SupplementaryTables.xlsx
- ExtendedDataFigures.docx
- Fig2.mp4