Ancient whole-genome duplications (WGDs) characterize many large angiosperm lineages, including angiosperms themselves. Prominently, the core eudicot lineage accommodates 70% of all angiosperms and shares ancestral hexaploidy, termed \textit{gamma}. \textit{Gamma} arose via two WGDs that occurred early in eudicot history; however, the relative timing of these is unclear, largely due to the lack of high-quality genomes among early-diverging eudicots. Here, we provide complete genomes for \textit{Buxus sinica} (Buxales) and \textit{Tetracentron sinense} (Trochodendrales), representing the lineages most closely related to core eudicots. We show that \textit{Buxus} and \textit{Tetracentron} are both characterized by independent WGDs, resolve relationships among early-diverging eudicots and their respective genomes, and use the RAC-CROCHE pipeline to reconstruct ancestral genome structure at three key phylogenetic nodes of eudicot diversification. Our reconstructions indicate genome structure remained relatively stable during early eudicot diversification, and reject hypotheses of \textit{gamma} arising via inter-lineage hybridization between ancestral eudicot lineages, involving, instead, only stem lineage core eudicot ancestors.
Flowering plants (angiosperms), with nearly 400,000 species and a fossil record that dates to the Early Cretaceous, have a complex evolutionary history marked by early and rapid lineage divergences. Whole-genome duplication (WGD) events have also been frequent in angiosperms, and indeed all extant species are ancient polyploids descended from a common ancestor that experienced at least one WGD. Subsequent polyploidy events have been identified through ancient angiosperm phylogeny, often coinciding closely with the origin and/or radiation of major clades. Notably, the core eudicots, which evolved with an ancient hexaploid expansion, can be described as a gamma genome, and account for ~70% of extant angiosperm species. Moreover, a novel suite of floral features, 'whorled pentamery' with flower parts arranged in concentric whorls, evolved shortly after the origin of the core eudicots and could be genetically linked to this ancient hexaploidy event, e.g., through multiplications or rearrangements of floral transcriptional regulators. Such a causal relationship between the gamma genome and whorled pentamery, although still speculative, is consistent with the widely acknowledged role of gene and genome duplications providing the genetic raw material for evolutionary innovation.

The phylogenetic timing and mechanism of the gamma hexaploidy are currently unresolved. Hypotheses on the topic mostly envision a two-step process, in which the product of an initial WGD fused with a third genome in a second polyploidization, possibly via a wide cross after an extended period of random fractionation (loss of either copy of duplicated genomic regions following WGD) in the tetraploid intermediate. The breadth of this putative wide cross is also unclear and possibly includes extant early-diverging eudicot lineages. Alternatively, one of the gamma subgenomes may have been more resistant to fractionation, and all three subgenomes may have been joined rapidly in evolutionary time, perhaps in an autohexaploidy event. It has also been argued that the gamma hexaploidy derives from an initial tetraploidy shared by all eudicots. Further still, the lack of clear evidence of genome duplication outside of the core eudicots may be due to stochastic gene loss over more than 100 million years of independent evolution. Efforts to evaluate evolutionary scenarios of the gamma origins have been hampered by limited data and unsettled sister-group relationships to the core eudicots. Plastome sequence data support either Buxales [19,26,27] or Trochodendraceae [28,29] as immediate sisters to the core eudicots, while single-copy nuclear (SCN) genes from the core eudicots are shared by all eudicots. Further still, the lack of clear evidence of genome duplication outside of the core eudicots may be due to stochastic gene loss over more than 100 million years of independent evolution.

Phylogenetic positions of Buxus and Tetracentron. To reconstruct the branching sequence of the early eudicot radiation, we analyzed phylogenetic data sets for representative angiosperms composed of hundreds of BUSCO genes, the Angiosperms35 as a sister to all other living eudicot lineages, while the four subgenomes of Tetracentron appear to be highly rearranged at the chromosomal level (Fig. 1). The extent to which this structure reflects genome reshuffling, which is a prominent mechanism of post-polyploid diploidization (PPD) after WGDs, or artifacts of genome assembly, is unclear. In favor of PPD processes, the Tetracentron genome is appreciably downsized compared to its sister species, and the only other living member of Trochodendraceae, Trochodendron aralioides (0.9 versus 1.6 GB), which shares two WGDs with Tetracentron but exhibits more extensive blocks of inter-chromosomal synteny (Supplementary Fig. 1).

Flowering plants (angiosperms), with nearly 400,000 species and a fossil record that dates to the Early Cretaceous, have a complex evolutionary history marked by early and rapid lineage divergences. Whole-genome duplication (WGD) events have also been frequent in angiosperms, and indeed all extant species are ancient polyploids descended from a common ancestor that experienced at least one WGD. Subsequent polyploidy events have been identified through ancient angiosperm phylogeny, often coinciding closely with the origin and/or radiation of major clades. Notably, the core eudicots, which evolved with an ancient hexaploid expansion, can be described as a gamma genome, and account for ~70% of extant angiosperm species. Moreover, a novel suite of floral features, ‘whorled pentamery’ with flower parts arranged in concentric whorls, evolved shortly after the origin of the core eudicots and could be genetically linked to this ancient hexaploidy event, e.g., through multiplications or rearrangements of floral transcriptional regulators. Such a causal relationship between the gamma genome and whorled pentamery, although still speculative, is consistent with the widely acknowledged role of gene and genome duplications providing the genetic raw material for evolutionary innovation.

The phylogenetic timing and mechanism of the gamma hexaploidy are currently unresolved. Hypotheses on the topic mostly envision a two-step process, in which the product of an initial WGD fused with a third genome in a second polyploidization, possibly via a wide cross after an extended period of random fractionation (loss of either copy of duplicated genomic regions following WGD) in the tetraploid intermediate. The breadth of this putative wide cross is also unclear and possibly includes extant early-diverging eudicot lineages. Alternatively, one of the gamma subgenomes may have been more resistant to fractionation, and all three subgenomes may have been joined rapidly in evolutionary time, perhaps in an autohexaploidy event. It has also been argued that the gamma hexaploidy derives from an initial tetraploidy shared by all eudicots. Further still, the lack of clear evidence of genome duplication outside of the core eudicots may be due to stochastic gene loss over more than 100 million years of independent evolution. Efforts to evaluate evolutionary scenarios of the gamma origins have been hampered by limited data and unsettled sister-group relationships to the core eudicots. Plastome sequence data support either Buxales [19,26,27] or Trochodendraceae [28,29] as immediate sisters to the core eudicots, while single-copy nuclear (SCN) genes from the core eudicots are shared by all eudicots. Further still, the lack of clear evidence of genome duplication outside of the core eudicots may be due to stochastic gene loss over more than 100 million years of independent evolution.

Phylogenetic positions of Buxus and Tetracentron. To reconstruct the branching sequence of the early eudicot radiation, we analyzed phylogenetic data sets for representative angiosperms composed of hundreds of BUSCO genes, the Angiosperms35 as a sister to all other living eudicot lineages, while the four subgenomes of Tetracentron appear to be highly rearranged at the chromosomal level (Fig. 1). The extent to which this structure reflects genome reshuffling, which is a prominent mechanism of post-polyploid diploidization (PPD) after WGDs, or artifacts of genome assembly, is unclear. In favor of PPD processes, the Tetracentron genome is appreciably downsized compared to its sister species, and the only other living member of Trochodendraceae, Trochodendron aralioides (0.9 versus 1.6 GB), which shares two WGDs with Tetracentron but exhibits more extensive blocks of inter-chromosomal synteny (Supplementary Fig. 1).

Flowering plants (angiosperms), with nearly 400,000 species and a fossil record that dates to the Early Cretaceous, have a complex evolutionary history marked by early and rapid lineage divergences. Whole-genome duplication (WGD) events have also been frequent in angiosperms, and indeed all extant species are ancient polyploids descended from a common ancestor that experienced at least one WGD. Subsequent polyploidy events have been identified through ancient angiosperm phylogeny, often coinciding closely with the origin and/or radiation of major clades. Notably, the core eudicots, which evolved with an ancient hexaploid expansion, can be described as a gamma genome, and account for ~70% of extant angiosperm species. Moreover, a novel suite of floral features, ‘whorled pentamery’ with flower parts arranged in concentric whorls, evolved shortly after the origin of the core eudicots and could be genetically linked to this ancient hexaploidy event, e.g., through multiplications or rearrangements of floral transcriptional regulators. Such a causal relationship between the gamma genome and whorled pentamery, although still speculative, is consistent with the widely acknowledged role of gene and genome duplications providing the genetic raw material for evolutionary innovation.

The phylogenetic timing and mechanism of the gamma hexaploidy are currently unresolved. Hypotheses on the topic mostly envision a two-step process, in which the product of an initial WGD fused with a third genome in a second polyploidization, possibly via a wide cross after an extended period of random fractionation (loss of either copy of duplicated genomic regions following WGD) in the tetraploid intermediate. The breadth of this putative wide cross is also unclear and possibly includes extant early-diverging eudicot lineages. Alternatively, one of the gamma subgenomes may have been more resistant to fractionation, and all three subgenomes may have been joined rapidly in evolutionary time, perhaps in an autohexaploidy event. It has also been argued that the gamma hexaploidy derives from an initial tetraploidy shared by all eudicots. Further still, the lack of clear evidence of genome duplication outside of the core eudicots may be due to stochastic gene loss over more than 100 million years of independent evolution. Efforts to evaluate evolutionary scenarios of the gamma origins have been hampered by limited data and unsettled sister-group relationships to the core eudicots. Plastome sequence data support either Buxales [19,26,27] or Trochodendraceae [28,29] as immediate sisters to the core eudicots, while single-copy nuclear (SCN) genes from the core eudicots are shared by all eudicots. Further still, the lack of clear evidence of genome duplication outside of the core eudicots may be due to stochastic gene loss over more than 100 million years of independent evolution.

Phylogenetic positions of Buxus and Tetracentron. To reconstruct the branching sequence of the early eudicot radiation, we analyzed phylogenetic data sets for representative angiosperms composed of hundreds of BUSCO genes, the Angiosperms35 as a sister to all other living eudicot lineages, while the four subgenomes of Tetracentron appear to be highly rearranged at the chromosomal level (Fig. 1). The extent to which this structure reflects genome reshuffling, which is a prominent mechanism of post-polyploid diploidization (PPD) after WGDs, or artifacts of genome assembly, is unclear. In favor of PPD processes, the Tetracentron genome is appreciably downsized compared to its sister species, and the only other living member of Trochodendraceae, Trochodendron aralioides (0.9 versus 1.6 GB), which shares two WGDs with Tetracentron but exhibits more extensive blocks of inter-chromosomal synteny (Supplementary Fig. 1).
Concatenated data sets of the SCN loci, whether analyzed in Maximum Likelihood (Fig. 2a, right panel, Supplementary Fig. 2b) or Bayesian Inference (Supplementary Fig. 4) frameworks, recover Buxales alone as the core eudicot sister group, with Trochodendrales as sister to this Buxales +core eudicot clade. Although this branching sequence receives maximal statistical support in both Maximum Likelihood (bootstrap) and Bayesian Inference (posterior probability) analyses, incomplete lineage sorting (ILS) is a potential confounding factor in phylogenetic analyses of concatenated data sets in the face of rapid radiations46, as is the case for the eudicots. Indeed, the quartet-support values associated with the Buxales+Trochodendrales clade in the coalescence tree indicate considerable gene tree discordance with respect to the positions of these taxa. Further exploration of conflicts affecting the eudicot clade, visualized as a cladogram of gene trees (Fig. 2b), however, reveals that ~30% of the gene trees support the Buxales+Trochodendrales as the core eudicot sister group (Supplementary Data 4). We also estimated the branching sequence of early-diverging eudicots using the 'Trees in the Peaks' method, which reconstructs speciation and polyploidization events from Ks and similarity score distributions of syntenic homologs47,48 (Fig. 2c). This method, which requires that ancestral Ks and similarity scores and/or their ranges must precede (greater Ks or lower similarity) or overlap those in the descendants, was applied to evaluate each of all possible binary rooted phylogenies. The only branching sequence that satisfies these conditions is one in which Buxales and Trochodendrales are collectively sister to the core eudicots. Specifically, the peak Ks value of syntenic orthologs that diverged via the Buxus/Tetracentron speciation is younger than those derived from the phylogenetic divergence of Vitis (a core eudicot) from Buxus or from Tetracentron.

**Phylogenomics of eudicot subgenomes.** Synteny-guided phylogenomic analyses of eudicot subgenomes were conducted to assess the several hypothesized scenarios for the origin of gamma hexaploidy (Fig. 3). Pairwise analyses of inter-genomic collinearity (macrosynteny) and fractionation patterns identify extensive regions of early-diverging eudicot genomes shared with the gamma-derived hexaploid genome of Vitis, and each other (Supplementary Figs. 5–9). The ratios of syntenic depths (the number of times a genomic region is syntenic to regions in another genome) for the respective species. Thus, we see 2:3 syntenic depth between Buxus and Vitis, and 4:3 syntenic depth between Tetracentron and Vitis, while Tetracentron to Buxus is 4:2 in syntenic depth. Likewise, as previously reported, Aquilegia and Nelumbo each exhibit 2:3 syntenic depth with Vitis, and 2:2 with each other. Collectively, these macrosyntenic alignments approximate the modern distribution of the seven

---

**Fig. 1 Habit and genome assembly features of Buxus and Tetracentron.** Flowering branch of Buxus sinica (a) courtesy of PiPi; and leafy shoot of Tetracentron sinense (b) courtesy of Daderot. Hi-C contact heatmaps, intragenomic syntenic with syntenic blocks colored according to the Ks scale, and Circos plots for Buxus (c) and Tetracentron (d). Concentric tracks in the Circos plots, from innermost outwards, show gene, Copia, and Gypsy retrotransposon densities per 1 Mb, and chromosomes, while ribbons connect inter-chromosomal syntenic regions. Source data underlying Fig. 1c, d are provided as a Source data file.
Fig. 2 Phylogenetic relations of Buxus and Tetracentron. a Phylograms depicting the coalescent solution of individual Maximum Likelihood (ML) gene trees (left) and partitioned ML analysis of a supermatrix of nucleotide sequence alignments (right). Node labels indicate quartet (coalescence) and bootstrap (supermatrix) support values, and orange stars highlight the positions of Buxales and Trochodendrales in the two trees. b Cloudogram of 763 SCN gene trees illustrating discordance surrounding the deep branches of eudicot phylogeny. The most frequent trees are blue, the next most frequent red, the third most frequent green, and the rest are dark green. c Ks (left) and Similarity (right) distributions showing peaks (arrows) that stem from WGD (top) and speciation events (bottom), respectively. Source data underlying Fig. 2b, c are provided as a Source data file.
Fig. 3 Synteny and phylogenomics of eudicot subgenomes. a Macrosyntenic alignments of early-diverged eudicots against Vitis with tracking of genomic positions by color-coded syntenic blocks representing the seven ancestral eudicot chromosomes. b Coalescence-based phylogenies of syntelogs derived from duplication events affecting the seven ancestral eudicot chromosomes. Green, red, blue, purple, yellow, aqua, and brown tracks highlight positions of ancestral chromosomes 1 through 7, respectively. Branch labels are posterior probabilities. c Schematic reconstruction of ancient eudicot WGD history. Differently color-filled circles label putative independent duplication events and stars highlight the two gamma WGDs in which the third genome is donated to the initial tetraploid (green star) from an extinct lineage to form the hexaploid (yellow star). Source data underlying Fig. 3a are provided as a Source data file.
ancestral eudicot chromosomes (Fig. 3a, Supplementary Data 5, and see below), the evolutionary histories of which we have estimated through phylogenetic analyses of 1932 gene trees populated with 15872 genes (Fig. 3b). For example, syntenic blocks descended from ancestral chromosome 4 (purple tracks in Fig. 3a) occupy regions of Vitis chromosomes 5, 7, and 14, as well as portions of chromosomes 2 and 5 of Aquilegia, 1 and 2 of Nelumbo, 5 and 6 of Buxus, and 1, 2, 3, and 18 of Tetracentron. Microsynteny (gene level) alignments within these major synteny blocks comprise 235 homologous loci and a total of 1837 syntenic (genes derived from the same ancestral genomic region) useful for inferring the evolutionary history of ancestral chromosome 4 (see Supplementary Data 5 for the modern distribution and representation of each ancestral chromosome). The coalescent solution of phylogenetic trees for these 235 loci shows that duplicated blocks of ancestral chromosome 4 now present in Aquilegia, Nelumbo, Buxus, and Tetracentron constitute lineage-specific clades (Fig. 3b), indicating that ancestral chromosome 4 was duplicated independently in each of the respective stem lineages of these four modern genomes. Indeed, the duplicated blocks of all seven ancestral chromosomes in Aquilegia, Nelumbo, Buxus, and Tetracentron constitute lineage-specific groupings (Fig. 3b), providing consensus that their respective WGDs are independent events and, importantly, exclusively involved genome donors that belonged to their respective clades, i.e., their stem lineage ancestors.

Phylogenetic alliances of the seven ancestral chromosomes occupying the modern, gamma-derived, Vitis genome are less clear. Of the three copies of ancestral chromosome 4, the syntenic blocks preserved on Vitis chromosome 5 and 7 form a well-supported sister group, but the block on Vitis chromosome 14 is placed as an earlier branch, albeit with low support. Vitis-specific clades were also not recovered for ancestral chromosomes 1 and 3, although again without high statistical support for non-monophyly. However, triplicated copies of ancestral chromosomes 2, 5, 6, and 7 in the Vitis genome group together as each other’s closest relatives. Although clade support is strong only for the copies of ancestral chromosome 7 currently preserved on Vitis chromosomes 6, 8, and 13, the phylogenies of these four sets of genomic regions suggest they uniquely share a common ancestor, one that evolved separately from the other, earlier-diverged, eudicot lineages. Altogether, we recover Vitis-specific groupings for duplicates of four of the ancestral eudicot chromosome, albeit as a well-supported clade only once. The relationships of the other three ancestral chromosomes may best be described as phylogenetically unresolved. Importantly, these findings are inconsistent with evolutionary scenarios of gamma formation through an extremely wide cross between a core eudicot and an early-diverging eudicot lineage, as has been previously proposed22. An initial tetraploidy event in the common ancestor of the eudicots24 is also inconsistent with our finding that paralogous genomic blocks in Aquilegia, and all other basal eudicots, constitute lineage-specific clades. The only evolutionary scenario consistent with our analyses is one in which gamma hexaploidy exclusively involved stem lineage ancestors of extant core eudicot species as genome donors. As such, if hexaploidy was attained via a two-step process of sequential WGDs, the third of the gamma genomes must have been donated from a now extinct lineage that branched off the core eudicot ancestral line before the initial tetraploidy event (Fig. 3c).

Ancient genomes. The independence of each of the WGD events associated with each of the early-diverging eudicot lineages implies unduplicated ancestral genomes leading all the way from the ancestral angiosperm up to gamma and the core eudicots. We explore this key inference through ancestral genome reconstruction. We reconstructed ancestral genomes at three nodes of the eudicot phylogeny (Fig. 4a): the common ancestor of the core eudicot clade (ancestor 3), two sequentially older nodes ancestral also to Buxus and Tetracentron (ancestor 2), and Nelumbo (ancestor 1).

All three of these ancestral genomes are reconstructed as seven putative protochromosomes, each with between 700 and 1600 protegenes, totaling more than 8000 protegenes, arranged in their ancestral order (Fig. 4b). Our ancestral genome reconstructions include ~2000 more (ca. 25%) ordered protegenes than previous reconstructions of an ancestral eudicot genome49,50. To understand the early evolution of eudicot genome structure, we partitioned the modern eudicot chromosomes into sets of syntenic regions and painted each of these according to its corresponding protochromosomes (Fig. 4c; Supplementary Fig. 10). These projections relate modern eudicot genomes to successive ancestral precursors and provide insights into the relative timing of any structural changes during eudicot genome evolution. Projections of the three ancestral genome reconstructions onto Vitis chromosomes (Fig. 4c) are globally similar, indicating genome structure remained relatively stable during early eudicot diversification. Inconsistent with the hypothesis of one ancestral eudicot tetraploidy24, these projections indicate that fusion of the two ancient chromosome now combined in Vitis chromosome 7 and Aquilegia chromosome 5 (juxtaposed purple and green blocks in Fig. 4c and Supplementary Fig. 10, respectively) did not occur prior to the origin of the eudicot ancestor. Were this the case, both sections of these Vitis and Aquilegia chromosomes would be painted with a common color representing one ancestral chromosome whose ‘chimeric’ origin would be invisible to our methods. Instead, these, and other, chromosomal fusions appear to be independent, lineage-specific events that post-date ancestral genome arrangements. Several other genomic rearrangements, as measured by the ‘choppiness’ of chromosomal paintings (Supplementary Data 6), emerge from our reconstructions. In the case of Vitis, the modern genome has accumulated 41 inter-chromosomal exchanges relative to ancestors 1 and 2, and 31 after ancestor 3. The reduced number of inter-chromosomal exchanges indicates greater similarity of Vitis to the core eudicot ancestor (ancestor 3) relative to the more ancient ancestors 1 and 2. A similar reduction of inter-chromosomal exchanges, from 67 (relative to ancestor 2) to 56 (relative to ancestor 3), was also observed for Amaranthus tuberculatus, the other core eudicot genome in our analyses. As such, we can reject the occurrence of any single WGD in the eudicot stem lineage and instead firmly resolve independent WGDs in each modern eudicot lineage, including the core eudicots with their unique gamma hexaploid structure.

Our Buxus and Tetracentron genome assemblies have facilitated rigorous assessments of alternative hypothesized scenarios for the origin of gamma, a key hexaploidy associated with a major event in the history of terrestrial life, the origin of core eudicots, which comprise the vast majority of flowering plants. We have presented and analyzed several lines of evidence, including Ks distributions, genomic synteny, fractionation bias, phylogenomics, and ancestral genome reconstruction, that bear relevance to the phylogenetic and WGD history of the early-diverging eudicot angiosperms. These analyses reconstruct the sequential branching order of the initial eudicot radiation and show that each of the early-diverging eudicot lineages is characterized by its own independent duplication event(s). We find no evidence to support hypotheses that a single polyploidy event might have been formative for eudicot diversification as a whole. Instead, our analyses place gamma hexaploidy on the stem lineage of core eudicots and rule out a role for other living early-diverging
Fig. 4 Ancestral eudicot genomes. a Schematic phylogeny of the eudicot clade depicting extant and ancestral genomes (nodes 1–3) examined here. Images Credits: Vitis, Bob Nichols; Amaranthus, Patrick Alexander; Buxus, PiPi; Tetracentron, Daderot; Nelumbo, Engin Akyurt; Aquilegia, Ejohnsonboulder; all available in the Public Domain via Wikimedia Commons. b Protogene content of ancestral genomes. c Heatmaps of conserved synteny supporting the delimitation of seven protochromosomes in each ancestral genome (left panel), and Vitis chromosomes painted according to the protochromosomes of the three diploid ancestral genomes (right panel). Source data underlying Fig. 4c are provided as a Source data file.
eudicots as genome donors, a possibility that was consistent with the results of previous analyses\textsuperscript{13–15,22}. Without a single, linking WGD common to all eudicots, an argument that one polyploidy event may have helped spur the massive eudicot diversification (via adaptive, alternative deployments of duplicate genes), even following a time lag, is not supported by our data. Instead, each independent WGD among the early-diverging eudicot lineages, other than \textit{gamma}, underlies relatively species-poor lineages that show limited fossil or living evidence for extensive radiation. Thus, with the genomes of all living early-diverging eudicot lineages now examined for a possible genomic contribution to \textit{gamma}, the origin of \textit{gamma} remains another abominable angiosperm mystery despite intensive study.

Methods

DNA extraction, sequencing, and assembly. \textit{Buxus} \textit{sinica} and \textit{Tetracentron} \textit{sinense} tissues were extracted from individuals cultivated at the University of Wisconsin-Madison (accession no. UW 136) and the University of Washington Arboretum, Seattle (accession no. 385–62), respectively. Genome sizes for these accessions were estimated using flow cytometry with BD CellQuest Pro software (Supplementary Data 7) by the Benaroya Research Institute (Seattle, WA). High-molecular-weight eugenic DNA was isolated from young leaf tissue and purified by nuclei-preparation and cetyltrimethylammonium bromide (CTAB) DNA extraction methods. Briefly, leaf tissue was ground to a fine powder under liquid nitrogen and mixed with nuclear isolation buffer (15 mM Tris, 10 mM EDTA, 130 mM KCl, 20 mM NaCl, 1 mM Spermine, 1 mM Spermidine, 8% PVP-10, 0.1% Triton X-100, and 0.1% mercaptoethanol), passed sequentially through 100 and 40 µm mesh filters, treated with 1% Triton X-100, and centrifuged at 2000 × g for 10 min at 4 °C to pellet the nuclei. The pellet resuspended for 1 h at 65 °C in lysis buffer (100 mM Tris-HCl, 100 mM NaCl, 50 mM EDTA, 2% CTAB, 1% PEG 6000), and high-molecular-weight DNA was isolated from the lysate via 241:chloroform/isoamyl alcohol and purified with the Qiagen Genomic kit. SMRTbell 20-kb libraries were generated and sequenced on the PacBio RSII platform to ~160x genome coverage. In addition, Hi-C libraries were prepared and sequenced to coverage depths of ~40x by Phase Genomics (Seattle, WA). PacBio reads were assembled using the pb-assemble suite of programs which includes the FALCON and FALCON-unzip assembly pipeline and performs contig phasing and polishing\textsuperscript{35}. The polished assemblies were deduplicated with Purge Haplotigs\textsuperscript{50} and scaffolded using Proximity Guided Assembly (PGA) and Hi-C reads by Phase Genomics (Seattle, WA).

RNA-seq data. Transcriptome assemblies were produced for \textit{Buxus} \textit{sinica} and \textit{Tetracentron} \textit{sinense} to aid annotation of their genome assemblies. We also produced transcriptomes for six early-diverging eudicots (\textit{Buxus} \textit{semperfervens}, \textit{Melosma dillenifolia}, \textit{Nelumbo lutea}, \textit{Sabia emarginata}, \textit{Sabia sempervirens}, \textit{Trochodendron aralioides}), as well as the core eudicot (\textit{Vitis}, \textit{Aquilegia}, \textit{Buxus}, \textit{Nelumbo}, \textit{Vitis} \textit{swinhonei}, \textit{Tribocladon aralioides}), to improve taxon sampling in phylogenetic analyses. Paired-end RNA-seq libraries were constructed from polyAselected total RNA extracted from floral and/or leaf tissues (Supplementary Data 8), and sequenced using the Illumina HiSeq 3000 system. Reads were trimmed with Trimmomatic\textsuperscript{31} and assembled using Trinity\textsuperscript{52}. Coding DNA (CDS) and protein sequences were predicted with TransDecoder (http://transdecoder.github.io).

Annotation. Genomes were annotated using the MAKER pipeline\textsuperscript{53}. De novo transcriptome assemblies for \textit{Buxus} and \textit{Tetracentron}, along with proteomes for four publicly available eudicot genomes—\textit{Arabidopsis thaliana}, \textit{Aquilegia coerulea}, \textit{Nelumbo nucifera}, and \textit{Vitis vinifera} (Supplementary Data 8)—were provided as evidence. Custom repeat libraries for genome masking were produced according to the MAKER\textsuperscript{P} advanced protocol\textsuperscript{54} using LTRharvest\textsuperscript{55}, LTRdigest\textsuperscript{56}, MITE-Hunter\textsuperscript{57}, RepeatModeler\textsuperscript{58}, and RepeatMasker\textsuperscript{59}. Gene models were predicted from the masked assemblies using SNAP\textsuperscript{60} and Augustus\textsuperscript{61} ab initio predictors from the masked assemblies using the SNAP\textsuperscript{60} and Augustus\textsuperscript{61} ab initio predictors after three rounds of training on interim high-quality (AED < 0.25; length > 50 amino acids) and BUSCO gene models, respectively.

Phylogenetic analyses. Three phylogenetic data sets were compiled from translated transcriptomes or genome-annotated proteomes for 40 angiosperms (Supplementary Data 8). Conserved single-copy land plant genes were identified by BUSCO\textsuperscript{36} analyses with the embryophytaodb10 data set, orthologs of the Angiosperm clade collected by ARA, orthologs of the AngioDicot clade, and orthologs of the AngioPicea clade as queries. The BUSCOs were compared with Amborella \textit{trichopoda} proteins, and orthogroups were circumscribed by Orthofinder\textsuperscript{44}. For all data sets, protein sequences were aligned using MAFFT\textsuperscript{26} and converted to codon alignments using PAL2NAL\textsuperscript{53}, which were refined in three successive rounds of sequence filtering and trimming using trimAl\textsuperscript{42}. Initially, sequences with less than 50% refinement overlap of their length were removed to discard any potentially spurious homologs. The passing sequences were next trimmed with trimAl\textsuperscript{19}’s heuristic automatic method (automated1) and filtered again as above to remove sequences that might contribute extensive missing data to the phylogenetic matrix. Alignments with fewer than 4 sequences, and missing representatives of both Buxales and Trochodendrales were discarded. After all filtering steps, 1248 BUSCOs, 346 Angiosperms353 loci, and 2573 orthogroups were retained for phylogenetic analyses. Maximum likelihood (ML) trees for the single-copy data sets were inferred from alignments of individual loci as well as concatenations of these, produced with RAxML\textsuperscript{90} with the GTR + gamma model of nucleotide evolution and 1000 bootstrap replicates. Concatenated alignments were analyzed using a partition scheme that defines individual genes as units for parameter optimization. Partitioned Bayesian Inference analyses were run with MrBayes with the GTR + I + G model for all partitions. Two independent parallel runs of four Metropolis-coupled Monte Carlo Markov Chains were run for 10 million generations with sampling every 1000 generations. Majority rule consensus trees and posterior probabilities of bipartitions were computed after discarding the first 25% of the sampled trees as burn-in. Orthogroup trees were inferred with IQTree\textsuperscript{70} with the best substitution model selected from among those implemented in RAxML and 1000 ultrafast bootstrap replicates. ASTRAIL\textsuperscript{11} and ASTRAIL-Pro\textsuperscript{99} were used to infer the species trees from single- and multi-copy gene trees, respectively, under the multi-species coalescent. DensiTree\textsuperscript{70} was used for visualizations of discordance among a subset of single-copy gene trees without missing taxa.

Comparative genomics of polyploidy. CoGe’s SynMap and FractBias programs were used to infer pan species and fractions. FractBias analyses were conducted using all genes in the target genomes and syntenic depth settings in accordance with ploidy levels of respective genomes, as revealed by SynMap plots. All analyses can be regenerated on the CoGe platform (see Code availability below). For synteny-guided phylogenomic analyses, inter-genomic alignments were produced and screened to identify all syntenic homologs divided into seven pools in accordance with the major synteny blocks conserved across these eudicot genome (as identified by SynMap and FractBias mappings, and which correspond with ancestral eudicot chromosomes). Unique identifiers for individual loci were replaced by ‘Species, chromosome’ codes to create comparable phylogenetic matrices and trees for coalescence-based phylogenetic analyses as outlined above.

Ancestral genomes. To build the three ancestral genomes indicated in Fig. 4, we used the RACCOCHRE pipeline\textsuperscript{35}. Briefly, RACCOCHRE uses all the syntenically varied gene lists of four eudicot genomes as single- and multi-copy orthogroups, the eudicot genomes based on the principle that a gene homologous (orthologous or paralogous) with any gene in a family must also be a member of that family. For each genome, RACCOCHRE extracts a set of ‘generalized’ adjacencies, namely all oriented pairs of genes within the same window containing seven consecutive genes. The pairs are represented by the non-adjacent ends of the two genes. The genes in these pairs are then labeled according to the gene families to which they belong. Each ancestor node has three incident branches, partitioning the tree into three subtrees defined by the one incoming edge (its ancestor) and two outgoing edges (its descendants). If an adjacency is found anywhere in any of the genomes in two or three of these subtrees, that adjacency is considered a candidate adjacency. With candidate adjacencies weighted as 2 or 3 according to the number of occurrences in subtrees, a maximum weight matching (MWM) of gene ends constructs the highest weight sets of compatible contiguous adjacencies (ancestral contigs). A gene end can only be matched to one end of another gene, so that these ancestral contigs are guaranteed to be linearly, or very occasionally circularly, ordered. Inversion breakpoints within windows of seven consecutive genes will preserve common adjacencies between two genomes, but not reading directions within the window. Common adjacencies are our primary concern, so we do not use reading direction information in MWM. Circular contigs were linearized by breaking an adjacency of lowest weight. The ancestral contigs from MWM solutions were then aligned to chromosomes of modern genomes, and co-occurring contigs were clustered to assemble ancestral chromosomes. A complete-linkage clustering was applied to the correlations of contigs’ co-occurrence to assemble ancestral chromosomes\textsuperscript{72}. To aid in future studies of the genomic organization of gene function, a GO-term enrichment analysis of the members of each gene family was implemented to produce a functional annotation for the inferred ancestral genes. The functional annotations of ancestral gene sets can be downloaded from https://gt.cst.usask.ca/buxus/buxus-tetra.

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

Data availability

All raw sequence reads used in this study have been deposited in NCBI under the BioProject accession numbers PRJNA489075, PRJNA547721, and PRJNA548936. In addition, the Buxus and Tetracentron genome assemblies, associated annotation files, and predicted CDS and protein sequences, along with all phylogenetic data sets analyzed here, and ancestral genome reconstructions have been deposited in the Dryad Digital
References

1. Govaerts, R. How many species of seed plants are there? TAXON 50, 1085–1090 (2001).

2. Friis, E. M., Pedersen, K. R. & Crane, P. R. Cretaceous angiosperm flowers: Innovation and evolution in plant reproduction. Palaeogeogr., Palaeoclimatol., Palaeoecol. 232, 251–293 (2006).

3. Magallon, S., Gomez-Acevedo, S., Sanchez-Reyes, L. L. & Hernandez-Hernandez, T. A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. N. Phytol. 207, 437–453 (2015).

4. Jiao, Y. et al. Ancestral polyploidy in seed plants and angiosperms. Nature 473, 97–100 (2011).

5. Amborella Genome Project. The Amborella genome and the evolution of flowering plants. Science 342, 12141089 (2013).

6. Schranz, M. E., Mohammadin, S. & Edger, P. P. Ancient whole genome duplications, novelty and diversification: the WGD Radiation Lag-Time Model. Curr. Opin. Plant Biol. 15, 147–152 (2012).

7. Vanneste, K., Maere, S. & Peer, deY. Tangled up in two: a burst of genome duplications at the end of the Cretaceous and the consequences for plant evolution. Philos. Trans. R. Soc. B 369, 20130353 (2014).

8. Tank, D. C. et al. Nested radiations and the pulse of angiosperm diversification: increased diversification rates often follow whole genome duplications. N. Phytologist 207, 454–467 (2015).

9. Soltis, P. S. & Soltis, D. E. Ancient WGD events as drivers of key innovations in angiosperms. Curr. Opin. Plant Biol. 30, 159–165 (2016).

10. Landis, J. B. et al. Impact of whole-genome duplication events on diversification rates in angiosperms. Am. J. Bot. 105, 348–363 (2018).

11. Cantino, P. D. et al. Towards a phylogenetic nomenclature of Tracheophyta. Int. J. Plant Sci. 171, 397–411 (2010).

12. Jaillon, O. et al. The grapevine genome sequence suggests ancestral genome structure and evolutionary innovation in major subclades. Science 321, 176–180 (2008).

13. Tang, H. et al. Unraveling ancient hexaploidy through multiply-aligned genome analysis of IR gene content evolution. Mol. Phylogenet. Evol. 449, 207–207 (2015).

14. Vekemans, D. et al. Gamma paleohexaploidy in the stem-lineage of core eudicots. Mol. Phylogenet. Evol. 73, 413–413 (2014).

15. Soltis, D. E. et al. Angiosperm phylogeny: 17 genes, 640 taxa. Evolution by Gene Duplication (Springer-Verlag, 1970).

16. Moore, M. J., Soltis, P. S., Bell, C. D., Burleigh, J. G. & Soltis, D. E. Phylogenetic analysis of 83 plastid genes further resolves the early radiation of eudicots. Proc. Natl Acad. Sci. USA 107, 4622–4628 (2010).

17. Sun, Y. et al. Phylogenetic and structural analyses of 18 complete plastomes across nearly all families of early-diverging eudicots, including an angiosperm-wide analysis of IR gene content evolution. Mol. Phylogenet. Evol. 96, 93–101 (2016).

18. Leebens-Mack, J. H. et al. One thousand plant transcriptomes and the phylogenomics of green plants. Nature 574, 679–685 (2019).

19. Filiault, D. L. et al. The Aquilegia genome provides insight into adaptive radiation and reveals an extraordinarily polymorphic chromosome with a unique history. Elife 7, e36426 (2018).

20. Strik, J. S., Hinsinger, D. D., Zhang, F. & Cao, K. Trochodendron aralioides, the first chromosome-level draft genome in Trochodendrales and a valuable resource for basal eudicot research. Gigasch 8, giz136 (2019).

21. Liu, P.-L. et al. The Tetraclonium genome provides insight into the early evolution of eudicots and the formation of vessel elements. Genome Biol. 21, 291 (2020).

22. XU, Q., Jin, L., Zheng, C., Leebens-Mack, J. & Sankoff, D. in Lecture Notes in Bioinformatics Vol. 12868 (2021).

23. Chiu, C.-S. et al. Phased diploid genome assembly with single-molecule real-time sequencing. Nat. Methods 13, 1050–1054 (2016).

24. Ghraye, J. et al. Integrating Hi-C links with assembly graphs for chromosome-scale assembly. PLoS Comput. Biol. 15, e1007237 (2019).

25. Yang, X., Lu, S. & Peng, H. Cytological studies on the eastern Asian family Amborellaceae. Bot. J. Linn. Soc. 158, 363–385 (2008).

26. Neale, D. B., Herrmans, L. & Leul, X. & Van Huylenbroeck, J. Genetic relationships in European and Asian Buxus species based on AFLP markers, genome sizes and chromosome numbers. Plant Syst. Evol. 293, 1–11 (2011).

27. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31, 3210–3212 (2015).

28. Sepevry, M., Manui, M. & Zdobnov, E. M. in Gene Prediction: Methods and Protocols (ed. Kollmar, M.) 227–245 (Springer, 2019).

29. Ratter, J. A. & Milne, C. in Notes from the Royal Botanic Garden Edinburgh (UK) (1976).

30. Dodsworth, S., Chase, M. W. & Leitch, A. R. Is post-polypliodization diploidization the key to the evolutionary success of angiosperms?. Botanical J. Linn. Soc. 180, 1–5 (2016).

31. Waterhouse, R. M. et al. BUSCO applications from quality assessments to gene phylogenetics and phylogenomics. Mol. Biol. Evol. 35, 543–548 (2018).

32. Johnson, M. G. et al. A universal probe set for targeted sequencing of 353 nuclear genes from any flowering plant designed using k-medoids clustering. Syst. Biol. 68, 594–606 (2019).

33. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 20, 238 (2019).

34. Mirarab, S. & Warnow, T. ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. Bioinformatics 31, 144–152 (2015).

35. Sankoff, D., Zheng, C., Lyons, E. & Tang, H. in Algorithms for Computational Biology (eds. Botón-Fernández, M., Martín-Vide, C., Santander-Jiménez, S. & Vega-Rodríguez, M. A.) 3–14 (Springer International Publishing, 2016).

36. Sankoff, D. & Zheng, C. in Comparative Genomics: Methods and Protocols (eds. Setubal, J. C., Stoye, J. & Stadler, P. F.) 291–315 (Springer, 2018).

37. Murat, F., Armero, A., Pont, C., Klop, C. & Salse, J. Reconstructing the genome of the most recent common ancestor of flowering plants. Nat. Genet. 49, 490–496 (2017).

38. Roach, M. J., Schmidt, S. A. & Borneman, R. A. Purge Haplotigs: allelic contig reassembly for third-gen diploid genome assemblies. BMC Bioinforma. 19, 460 (2018).

39. Bolger, A. M., Lohse, M. & Usadel, B. Trimomatic: a flexible trimmer for illumina sequence data. Bioinformatics 30, 2114–2120 (2014).

40. Grabherr, M. G. et al. Trinity: reconstructing a full-length transcriptome without a genome from RNA-seq data. Nat. Biotechnol. 29, 644–652 (2011).

41. Holt, C. & Yandell, M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinforma. 12, 491 (2011).

42. Campbell, M. S. et al. MAKER-P: a tool kit for the rapid creation, management, and quality control of plant genome annotations. Plant Physiol. 164, 513–524 (2014).

43. Ellingham, D., Kurtz, S. & Willhoefu, U. LTRHarvest, an efficient and flexible software for de novo detection of LTR retrotransposons. BMC Bioinforma. 9, 18 (2008).

44. Steinbiss, S., Willhoefu, U., Gremme, G. & Kurtz, S. Fine-grained annotation and classification of de novo predicted LTR retrotransposons. Nucleic Acids Res. 37, 7002–7013 (2009).

45. Han, Y. & Wessler, S. R. MITe-Hunter: a program for discovering miniature inverted-repeat transposable elements from genomic sequences. Nucleic Acids Res. 38, e199 (2010).
Acknowledgements

This work was supported by National Science Foundation grants DEB-1455601 to A.S.C., DEB-1457440 to D.G.H., DEB-2030871 to V.A.A., and Discovery grants to L.J. and to D.S. from the Natural Sciences and Engineering Research Council of Canada. D.S. holds the Canada Research Chair in Mathematical Genomics. We thank Brent Berger, Ray Larson, and Veronica Di Stilio for aid in plant collection and DNA extraction and Hanqi Ye for bioinformatics discussion.

Author contributions

A.S.C., D.E.S., D.G.H., D.S., P.S.S., and V.A.A. conceived and designed the study. A.S.C. generated the whole-genome and transcriptome assemblies, performed phylogenetic and comparative genomics analyses, and drafted the primary manuscript. D.S., L.J., and Q.X. generated and analyzed the ancestral genome reconstructions. Y.Z., E.C., and V.A.A. analyzed data. S.J. provided data. Additional text and discussion were provided by D.E.S., D.G.H., D.S., L.J., J.Z., P.S.S., and V.A.A. All authors approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-022-28312-w.

Correspondence and requests for materials should be addressed to Andre S. Chanderbali.

Peer review information Nature Communications thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

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

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2022