Robust identification of orthologous synteny with the *Orthology Index* and its applications in reconstructing the evolutionary history of plant genomes

Ren-Gang Zhang[1,2†], Hong-Yun Shang[1,2†], Min-Jie Zhou[1,2], Heng Shu[1,2], Kai-Hua Jia[3*], Yong-Peng Ma[1*]

[1] Yunnan Key Laboratory for Integrative Conservation of Plant Species with Extremely Small Populations / Key Laboratory for Plant Diversity and Biogeography of East Asia, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201 Yunnan, China.

[2] University of the Chinese Academy of Sciences, Beijing 100049 Beijing, China.

[3] Institute of Crop Germplasm Resources, Shandong Academy of Agricultural Sciences, Jinan 250100 Shandong, China.

[†] These authors contributed equally.

[*] Corresponding to: Yong-Peng Ma (mayongpeng@mail.kib.ac.cn)

Kai-Hua Jia (kaihuajia_saas@163.com)

Abstract

With the explosive growth of whole-genome datasets, accurate detection of orthologous synteny has become crucial for the reconstruction of evolutionary history based on these datasets. However, the methods of identifying orthologous synteny currently available for plants have great limitations: the methods are difficult to scale with varying polyploidy and the accurate removal of out-paralogy is challenging, given the high complexity of plant genomes. In this study, we developed a scalable and robust approach, the Orthology Index (OI), to accurately identify orthologous synteny by calculating the proportion of orthologs within syntenic blocks. Interestingly, our evaluation of a comprehensive dataset comprising nearly 100 known cases with diverse polyploidy and speciation events revealed that the technique is highly reliable in the identification of orthologous synteny, with an OI threshold value of 0.6 as a cutoff. This discovery highlights OI as a potentially universal criterion for the identification of orthologous synteny. In addition, we demonstrate its broad applications in reconstructing plant genome evolutionary histories, including inference of polyploidy, identification of reticulation, and phylogenomics. The index has been packaged in an all-in-one toolkit (freely available from https://github.com/zhangrengang/OrthoIndex) to facilitate its use in these applications. In conclusion, OI offers a robust, interpretable, and scalable approach for the automated identification of orthologous synteny, significantly expanding our analytical capabilities in plant evolutionary genomics.

Keywords: Orthologous synteny, Polyploidy, Reticulation, Phylogenomics, *Orthology Index*

Introduction

Reconstruction of the evolutionary histories of organisms, including inference of the tree/network of life, identification of polyploidy events, and placement of the events on the tree/network, usually relies on the orthologous relationships at gene, block, chromosome, subgenome and/or whole genome scales. In general, the more orthologous genes or loci involved in the reconstruction of evolutionary history, the more confidence can be placed in it. For example, phylogenomic reconstructions using genome-scale data have become the gold standard for understanding the evolution of lineages in the tree of life [1]. Orthologous synteny has been established as a proxy to generate the maximum number of orthologs, from the chromosomal to the subgenomic/genomic scales. Orthologous synteny has been used successfully in reconstructing evolutionary histories in, for example, the poppies [2] and the angiosperms as a whole [3]. However, accurately identification of orthologous synteny is still challenging, especially in flowering plants, where recurrent whole-genome duplications (WGDs) or polyploidizations have produced substantial numbers of syntenic paralogs, which can confuse the inference of the orthology [1, 4] and may mislead the reconstruction of evolutionary history. For example, the overlooked orthologous relationships in the synteny between two *Papaver* species resulted in an incorrect interpretation of the history of polyploidy in these species [2, 5, 6].

To date, two main strategies have been employed to identify orthologous synteny. One strategy is to filter the detected synteny with certain criteria, which are usually specific to each case and are therefore not scalable for large-scale datasets. One widely-used criterion is the synonymous substitution rate (*Ks*). Due to the whole genome duplications (WGDs; producing paralogs) and speciation (producing orthologs) events having occurred at different times in the past, *Ks* between syntenic gene pairs can be used to differentiate the ages of orthologous and paralogous synteny [e.g. 7, 8]. However, *Ks*-based methods are not always effective for distinguishing syntenic blocks from different events [9], and cannot be universally applied to the large-scale automated identification of orthologous syntenic blocks, because, as the events of different species occur at different times, the *Ks* values vary case by case, and because species have different substitution rates. A parameter, *homo*, has been proposed for use with WGDI, to extract the best homology (orthology) of syntenic blocks [10]. This parameter relies on another parameter, *multiple*, to define the top number of hits as best hits. Unfortunately, the parameter

cannot be universally applied to different cases with different syntenic depth ratios. Another tool, QUOTA-ALIGN [9] was developed to screen orthologous syntenic blocks under given constraints on syntenic depths (*QUOTA*). However, users are required to set the prior *QUOTA* parameter for cases of different evolutionary histories. For example, it is necessary to set *QUOTA* = 1:1 for *Arabidopsis thaliana* and *A. lyrata*, but to set *QUOTA* = 4:2 for *A. thaliana* and poplar [9]. This is also similarly required when setting the *multiple* parameter in WGDI (-c option) when extracting orthologous syntenic blocks.

The second alternative strategy for identifying orthologous synteny is the use of pre-inferred orthologs to call synteny, with tools such as MCScanX_h [11]. This strategy is scalable for large-scale datasets. However, hidden out-paralogs in the pre-inferred orthologs may result in out-paralogous syntenic blocks that need further removal with the first strategy. This problem can be frequently observed in previously published work, including certain studies investigating willow [12] and poplar [7] which used pre-inferred orthologs from OrthoFinder2 [13] and OrthoMCL [14], respectively, to call synteny. Moreover, the hidden orthologs (i.e. false negatives in ortholog inference) can reduce the efficiency of the subsequent detection of orthologous synteny in practice.

To address these issues and to screen orthologous synteny robustly for large-scale datasets, we developed a scalable approach called *Orthology Index* (*OI*). We combined the information regarding pre-inferred orthology (provided by OrthoFinder2 or analogs) and synteny (by WGDI or analogs) to calculate the *OI*, which represents the proportion of orthologous gene pairs within a syntenic block. We evaluated the efficacy of *OI* using a comprehensive dataset comprising nearly 100 well-documented cases with diverse WGD and speciation events. Our analysis demonstrated the high robustness and accuracy of *OI*, as well as its usefulness in evolutionary inference including polyploidy, reticulation and phylogenomics. We finally integrated the index into a toolkit (freely available from https://github.com/zhangrengang/OrthoIndex) to facilitate its use.

Results and Discussion

Identification of orthologous synteny using the Orthology Index

We aim to distinguish orthologous synteny from out-paralogy robustly for genomes of two species. The genomes should have shared a whole genome duplication event (WGD, or polyploidization;

producing out-paralogy) and a speciation event (producing orthology), and these events should have occurred within a short time period (denoted as ΔT, **Fig. 1A**). To evaluate our approach, we selected ninety well-documented plant species pairs that shared at least one polyploidy event (**Table S1**, **Figs. S1-90**), along with the typical poplar–willow case (**Fig. 1**), from the literature. The typical example is shown in **Fig. 1** with poplar (*Populus trichocarpa* as the representative) and willow (*Salix dunnii* as the representative), which shared two rounds of polyploidization. The first round was a paleohexaploidization event (γ event, whole genome triplication or WGT) common to the core eudicots at ~120 million years ago (Mya). The second was a lineage-specific tetraploidization event (WGD) shared by poplar and willow at ~60 Mya [15-18] (**Fig. 1A**). Following the recurrent polyploidization events, the two genera speciated later at ~50 Mya [16, 18] (**Fig. 1A**).

With the collinearity detector WGDI [10], syntenic blocks derived from the three evolutionary events (two polyploidization and one speciation) can be identified and distinguished visually using their distinct synonymous substitution rate (*Ks*) values and fragmentation levels (the more ancient the event, the more fragmented the blocks or greater the chromosomal rearrangements, and the higher the *Ks* value) (**Fig. 1B**). We denoted the three categories of syntenic blocks as WGT-SBs ($Ks \approx 1.5$, most fragmented, out-paralogous), WGD-SBs ($Ks \approx 0.27$, moderately fragmented, out-paralogous), and S-SBs ($Ks \approx 0.13$, least fragmented, orthologous) (**Fig. 1B**). The recent two *Ks* peaks derived from WGD-SBs (out-paralogy) and S-SBs (orthology) overlap (**Fig. 1B**) and are difficult to split with a unified cutoff.

The orthology inference method, OrhtoFinder2 [13], inferred a substantial number of orthologous genes (~ 30 %) that exhibit synteny in the out-paralogous blocks ($Ks \approx 0.27$; **Fig. 1C**), suggesting that they are hidden paralogs and would introduce out-paralogous synteny when imported to tools such as MCScanX_h [11]. This problem was observed in numerous instances where the ΔT was not substantial (e.g. **Figs. S1, 5, 6, 8, 10**), which could be attributed to biased gene loss or systematic errors [1]. Nevertheless, in the poplar–willow case, the peak ($Ks \approx 0.27$) of out-paralogs from the nearest WGD (WGD-SBs) was significantly reduced, while out-paralogs from the older WGT (WGT-SBs, $Ks \approx 0.13$) were almost entirely eliminated (**Fig. 1C**). This suggests that the level of accuracy of orthology inference by OrthoFinder2 is relatively high, although not entirely sufficient.

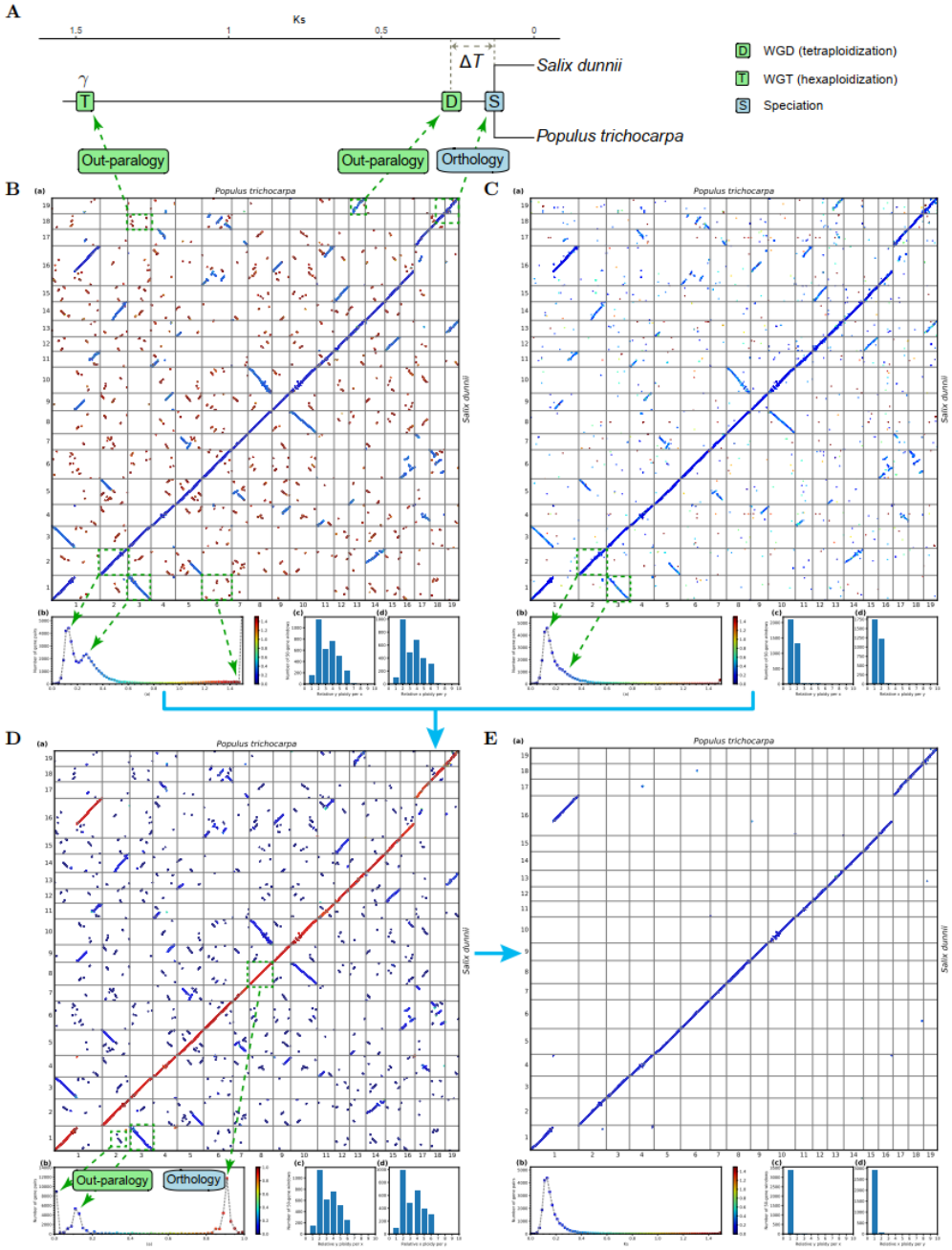


Fig. 1. The Illustration of the *Orthology Index* in identifying orthologous synteny: a typical case. A) Schematic evolutionary history of poplar and willow genomes, adapted from the literature [15-18]. **B)** *Ks*-colored dot plots showing synteny detected by WGDI, with an observable distinction of three categories of syntenic blocks derived from three evolutionary events (three peaks: *Ks* ≈ 1.5, *Ks* ≈ 0.27, and *Ks* ≈ 0.13). **C)** *Ks*-colored dot plots illustrating orthology inferred by OrthoFinder2, with an observable high proportion (~30 %) of hidden out-paralogs (*Ks* ≈ 0.27).

D) *Orthology Index* (*OI*)-colored dot plots: integrating synteny (**B**) and orthology (**C**), with polarized and scalable distinction of three categories of syntenic blocks (three peaks: $OI \approx 0$, $OI \approx 0.1$, and $OI \approx 0.9$). **E)** *Ks*-colored dot plots of synteny after applying an *OI* cutoff of 0.6, with clean 1:1 orthology as expected from the evolutionary history. **B-E** are plotted using the 'dotplot' subcommand with four subplots: a) dot plots with colored by *Ks* or *OI* (x-axis and y-axis, chromosomes of the two genomes; a dot indicates a homologous gene pair between the two genomes), b) histogram (with the same color map as the dot plots) of *Ks* or *OI* (x-axis, *Ks* or *OI*; y-axis, number of homologous gene pairs), c-d) synteny depth (relative ploidy) derived from 50-gene windows (x-axis, synteny depth; y-axis, number of windows). Examples of the three categories of syntenic blocks [referred to as WGT-SBs ($Ks \approx 1.5$, $OI \approx 0$), WGD-SBs ($Ks \approx 0.27$, $OI \approx 0.1$), and S-SBs ($Ks \approx 0.13$, $OI \approx 0.9$)] are highlighted with dashed squares. These are associated with the evolutionary events and peaks of *Ks* or *Orthology Index*, indicated by arrows, and labeled as 'Out-paralogy' or 'Orthology'. Additional visible cases of other lineages can be found in **Figs. S1-90** (summarized in **Table S1**).

Consequently, we combined the algorithmic advances of the two methods described above by introducing a straightforward index, referred to as the *Orthology Index* (*OI*), to determine the orthology of a syntenic block. Remarkably, the *OI* polarized orthology from out-paralogy with a very clear boundary (OI = ~0.3–0.7) (**Fig. 1D**), showing a much higher distinguishability than the *Ks*. The WGT-SBs exhibit an *OI* of nearly 0 (out-paralogous), and the WGD-SBs have an *OI* of approximately 0.1 (out-paralogous), whereas the S-SBs display an *OI* peak of approximately 0.9 (orthologous). We applied the index analysis to ninety study cases with diverse polyploidy and varying ΔT (**Table S1, Figs. S1-90**). The *OI* cutoff was then further adjusted to 0.6 to give a universal identification of orthology, according to the ranges of the polarized orthology – out-paralogy boundary calculated for these examined cases (**Fig. 1D, Figs. S1-90D**). Interestingly, we observed that when the *OI* cutoff value was set to 0.6, the index performed exceptionally and consistently well in identifying orthology accurately in nearly all instances (**Fig. 1E, Figs. S1-90E**). This finding highlights the efficiency and robustness of our index as a potentially universal criterion for identifying orthologous synteny. As for the poplar–willow case, a cutoff value of $OI = 0.6$ retained 12.2 % of the blocks and 47.6 % of the gene pairs as orthologs,

effectively retrieving a 1:1 orthology relationship (**Fig. 1E**) as expected from the evolutionary history (**Fig. 1A**). False positives (hidden paralogy) and false negatives (excessive removal of true orthologous synteny) were rarely observed in the examined cases (**Fig. 1E, Figs. S1-90E**), demonstrating the high robustness. Therefore, the index can be applied in automated pipelines for large-scale datasets (ranging from dozens to hundreds of genomes at present) with a unified threshold of *OI*. We integrated the index into a toolkit (https://github.com/zhangrengang/OrthoIndex) to facilitate downstream application.

Potential applications of *OI* in polyploidy inference: determining whether WGD events were shared among or specific to lineages

Distinguishing synteny resulting from orthology and out-paralogy is vital for the inference of the evolutionary history of polyploidy. We illustrate a simple model to explain the process. When two genomes exhibit a 2:2 ratio of synteny depth, there are two main hypotheses to consider: the genomes share a tetraploidization event (WGD), or they both underwent a lineage-specific tetraploidization event (WGD) separately (**Fig. 2A**). The *OI* offers a straightforward and visual method to test the two hypotheses with separated orthology and out-paralogy. If the genomes display a clear 1:1 orthology + 1:1 out-paralogy, similar to the poplar–willow case (**Fig. 1D**; ignoring the very ancient out-paralogy from the γ event), the first hypothesis is supported (**Fig. 2A, left panel**). On the contrary, if they exhibit a 2:2 orthology, the second hypothesis is supported (**Fig. 2A, right panel**). Our experience of studying this in dozens of known cases (**Figs. S1-90**) suggests that these inferences of polyploidy history from orthologous synteny patterns are reasonable and accurate.

Consequently, we applied this strategy to the controversial question of whether the Apiaceae and Araliaceae families (both within Apiales) share a tetraploidization event. There are two major opposing views on this: (i) the two families share a tetraploidization event, as evidenced by synteny and corrected *Ks* [19]; (ii) oppositely, the two families do not share the tetraploidization event, as inferred by synteny and *Ks*-based methods [20-22]. Here, we reexamined the polyploidy of the available genomes from the two families using the *OI*.

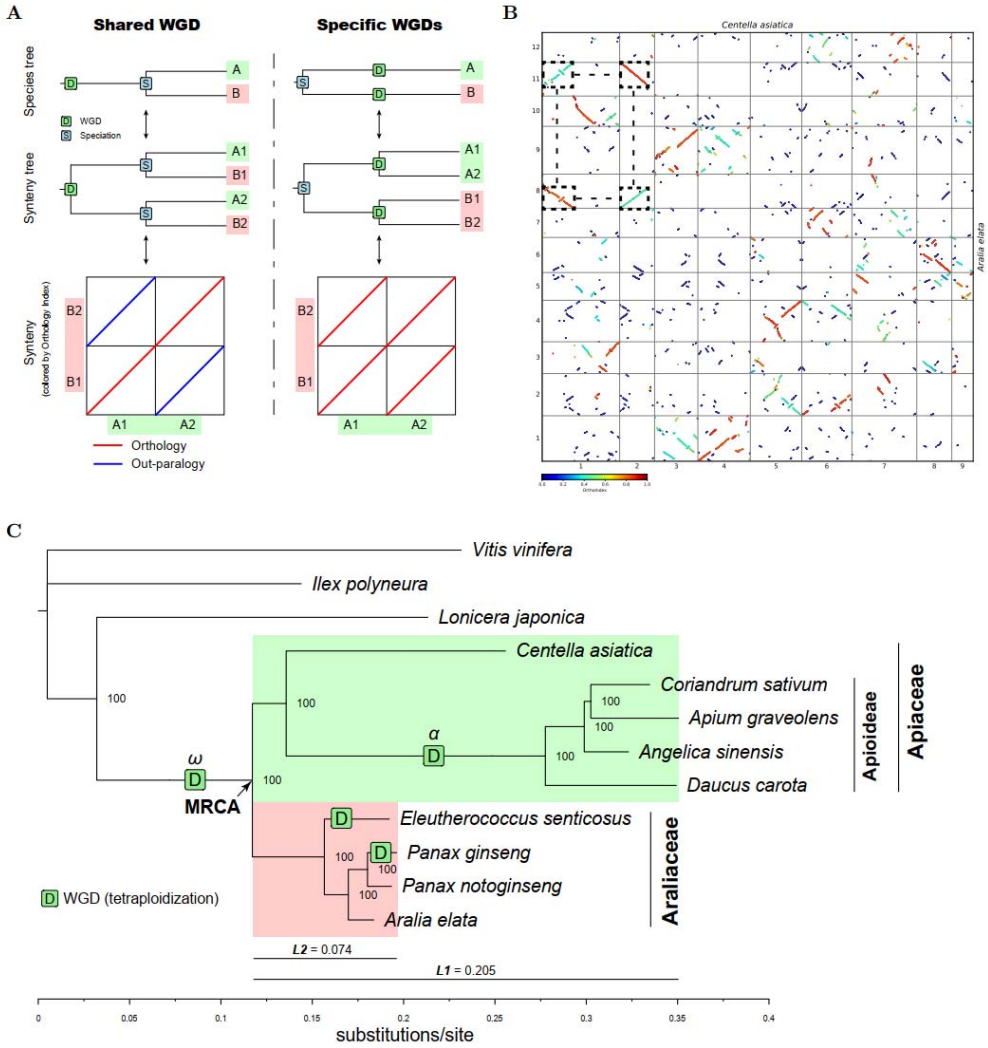


Fig. 2. Inference of polyploidy in the Apiaceae + Araliaceae genomes using the *Orthology Index*. A) Schematic illustration of determination of shared or lineage-specific polyploidy events using the orthologous syntenic relationships identified with the *Orthology Index*. Despite a similar 2:2 ratio of synteny depth, the two scenarios have distinct patterns of orthologous synteny (1:1 orthology vs. 2:2 orthology). Labels A and B indicate two species, and A1–2 and B1–2 indicate duplicated chromosomes or blocks from a WGD event. **B)** *Orthology Index*-colored dot plots indicating orthologous and out-paralogous synteny between genomes of *Centella asiatica* (Apiaceae) and *Aralia elata* (Araliaceae). A typical 1:1 orthology + 1:1 out-paralogy synteny pattern is highlighted by the dashed squares. **C)** Phylogeny reconstructed from genomes of species in the Apiaceae and the Araliaceae (Apiales) with labels indicating polyploidization events. *L1* and *L2* represent the average branch length (genetic distance) of the Apiaceae and the Araliaceae from

their most recent common ancestor (MRCA), respectively. Numbers at the nodes denote bootstrapping values (percentage). The maximum-likelihood phylogenetic tree was reconstructed using IQ-TREE2, based on concatenated codon alignments of 2,363 single-copy genes (with at most 20 % taxa missing). Additional evidence supporting the inferred polyploidization events can be found in **Figs. S91-97**.

Using *Centella asiatica* (Apiaceae) and *Aralia elata* (Araliaceae) as representatives, a clear 2:1 orthologous synteny depth with the grape (*Vitis vinifera*) genome is observed (**Fig. S91**), suggesting that both species have undergone a tetraploidization event following their divergence from grape. These inferences are in agreement with previous research [21, 22]. However, upon comparison between the two genomes, a clear 1:1 orthology + 1:1 out-paralogy pattern was highlighted by the *OI* (**Fig. 2B**), aligning with the first model (**Fig. 2A, left panel**) and thus suggesting a shared tetraploidization event (ω, WGD; **Fig. 2C**) that is inconsistent with the first view above [19]. This inference was corroborated with macro-synteny phylogenies (**Fig. S92**), demonstrating its reliability and confidence. Similar patterns were also observed in other combinations of genomes from the two families (e.g. *Centella asiatica* : *Eleutherococcus senticosus* = 1:2 orthology) (**Figs. S93-94**), further supporting the inference. Additionally, we noted a considerable disparity in branch length (or substitution rate, genetic distance) between the two families from their most recent common ancestor (MRCA), with the mean branch length of Apiaceae (*L1*) nearly three times than that of Araliaceae (*L2*) (**Fig. 2C**). This discrepancy can explain the unreasonable inferences drawn from uncorrected *Ks*-based methods [20-22], which assumed equal substitution rates for different lineages when estimating the timing of polyploidization and speciation events.

Furthermore, we deduced additional lineage-specific polyploidy events within the two families using the *OI*-based approach. We observed that all studied genomes from species in the Apioideae subfamily (including *Daucus carota*, *Angelica sinensis*, *Apium graveolens* and *Coriandrum sativum*) exhibit a 2:1 orthologous synteny depth when compared to the *C. asiatica* genome (**Fig. S95**), suggesting that they each experienced a tetraploidization (WGD) event subsequent to their divergence from *C. asiatica*. Moreover, any two genomes within this subfamily display distinct 1:1 orthology + 1:1 out-paralogy patterns (**Figs. S56, 96**), suggesting a

shared tetraploidization event (α), which is in line with previous research [19, 22]. Thus the α event was placed between the stem and crown of the Apioideae subfamily (**Fig. 2C**). In addition, employing the *OI*, we also inferred one species-specific tetraploidization event in both the *Panax ginseng* and *Eleutherococcus senticosus* genomes (**Fig. 2C, Figs. S57-59, 97**), consistent with some previous research [21, 22] but in conflict with the results from Song et al [19]. These inferences demonstrate the ability of the *OI* to infer the diverse polyploidy events explicitly and position them on the tree of life.

Potential applications of *OI* in the identification of reticulation

Reticulation, driven by allopolyploidization or hybridization, constitutes a significant factor in eukaryotic evolution, yielding novel phenotypes that facilitate ecological diversification and the occupation of new niches [23]. Numerous genomes originating from recent reticulation events have been documented (summarized in Jia et al [24]), encompassing critical cereals [25], fruits [26], vegetables [27], trees [7] and fish [28]. We evaluated the application of the *Orthology Index* in several well-documented cases, ranging from simple to complex evolutionary scenarios (**Fig. 3**). We conclude that straightforward reticulate speciation can be inferred solely from the *OI*-colored dot plots.

The cases of *Arabidopsis* [29] and *Arachis* [30] represent the simple hybridization or allopolyploidization scenarios. From the *OI*-colored dot plots, the two subgenomes of both the hybrid (*A. thaliana* × *A. renosa*) and tretaploid (*A. hypogaea*) show clear and separate orthologous relationships with their diploid progenitors (**Fig. 3A and 3C, Fig. S98-99**). Based on the orthologous relationships revealed by the *OI*, the hybridization events can be easily inferred (**Fig. 3B and 3D**).

The orthologous relationships between the genomes in complicated polyploid species complexes (*Triticum* and *Papaver*), are also clear from the *OI*-colored dot plots (**Fig. 3E and 3G, Fig. S100-101**). The two subgenomes of the tretaploid *T. turgidum* are orthologous to the two subgenomes of the hexaploid *T. aestivum* (**Fig. 3E, Fig. S100**), leading the inference that *T. turgidum* is the intermediate tretaploid progenitor of the hexaploid *T. aestivum* (**Fig. 3F**), in line with previous results [25]. Similarly, a reticulate allopolyploidization origin in two *Papaver* genomes (**Fig. 3H**) can be inferred from the *OI*-colored dot plots (**Fig. 3G, Fig. S101**). This

inference agrees with our previous work [2], however, the phylogenetic relationships between homoeologous subgenomes cannot be resolved directly from the *OI*-colored dot plots and require further evidence, such as chromosome or subgenome-scale phylogenies [2].

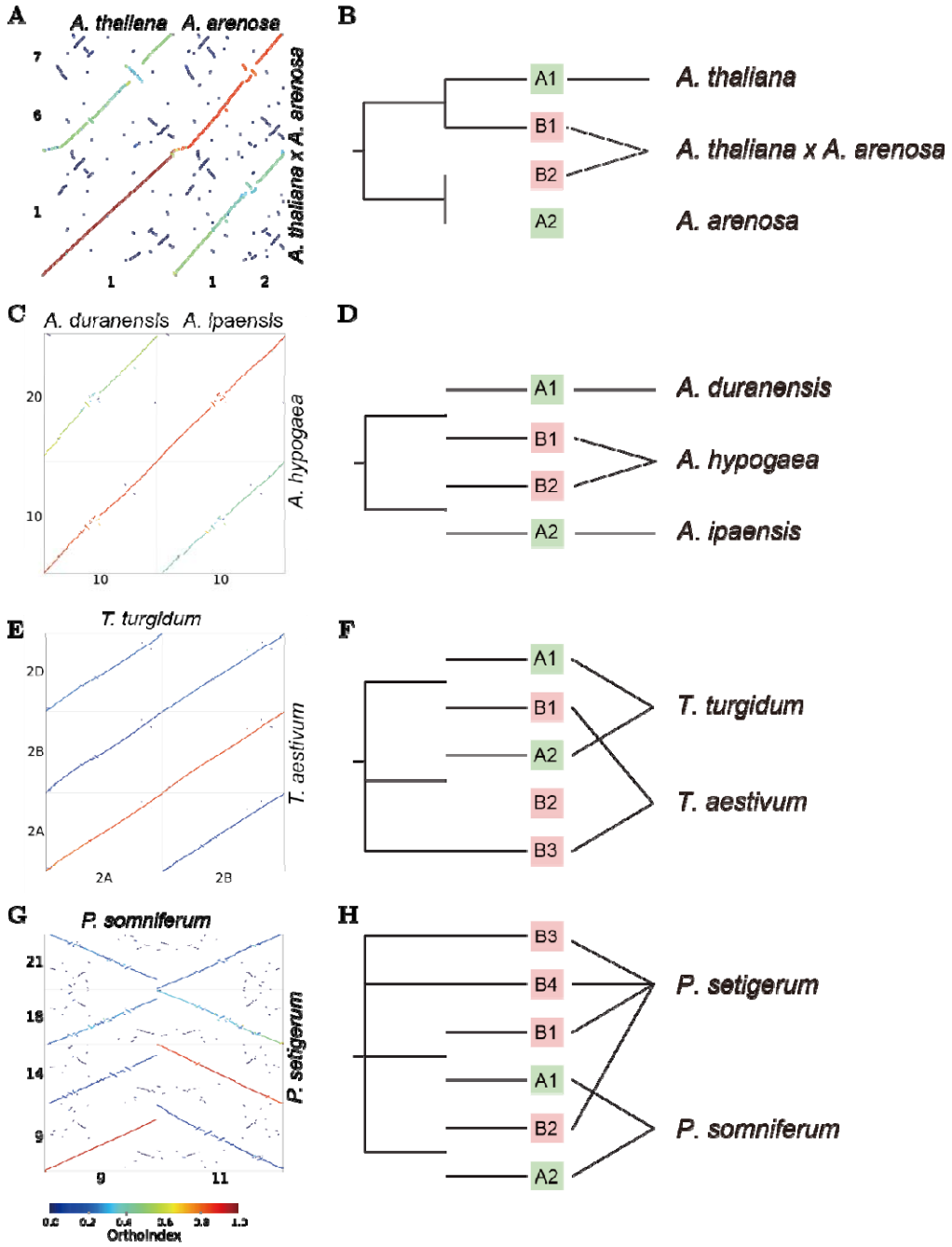


Fig. 3. **Examples of reticulation inferences based on the *Orthology Index*. A-B**) *Orthology Index*-colored dot plots (**A**) of the genomes of *A. thaliana* + *A. arenosa* and their hybrid *A. thaliana* × *A. arenosa*, and the inference (**B**; in dendrogram form) from the orthology evidence.

C-D) *Orthology Index*-colored dot plots (**C**) of the genomes of the tretaploid *A. hypogaea* and its diploid progenitors, and the inference (**D**; in dendrogram form) from the orthology evidence. **E-F**) *Orthology Index*-colored dot plots (**E**) of the genomes of the hexaploid *T. aestivum* and its intermediate tretaploid *T. turgidum*, and the inference (**F**; in dendrogram form) from the orthology evidence. **G-H**) *Orthology Index*-colored dot plots (**G**) of the genomes of the neo-octoploid *P. setigerum* and its intermediate tretaploid *P. somniferum*, and the inference (**H**; in dendrogram form) from the orthology evidence. Only one set of representative homoeologous chromosomes is shown in the dot plots, and the dot plots with full set of chromosomes can be found in **Figs. S98-101**.

Potential applications of *OI* in phylogenomics

Accurate inference of orthology plays a crucial role in the estimation of species trees, and the pseudo-orthologs (i.e. hidden paralogs) derived from WGD and gene loss can mislead species tree inference greatly under some circumstances [31]. As demonstrated in various cases above (**Figs. 1, S1-90**), the *Orthology Index* exhibited a high level of accuracy in identifying syntenic orthology (or syntelogs) and can therefore minimize the detrimental influence. Consequently, the results can be directly applied to species tree reconstruction. We used the example of the core eudicots to showcase the application. It is accepted that all core eudicots share a paleohexaploidization event (γ event, WGT), meanwhile no two orders within core eudicots share an additional polyploidy event [32]. Therefore, it is appropriate to use the *Orthology Index* to remove the out-paralogy produced from the γ event, and to better resolve the phylogenetic relationships among the orders of the core eudicots, which remain poorly resolved for some orders within the core eudicots, such as the Celastrales–Oxalidales–Malpighiales (COM) clade [33]. Here we utilized the genome-scale syntenic orthologs inferred by the *Orthology Index* to reconstruct a backbone phylogeny of the core eudicots, aiming to minimize the detrimental influence of the γ event.

We collected a high-quality genomic dataset that covers 28 (70 %) of the 40 orders and 98 (33 %) of the 298 families of core eudicots treated in APG IV [33]. We then applied the *OI* to identify syntenic orthologs, which resulted in the identification of 54,322 syntenic orthogroups (SOGs). After filtering, 12,277 multi-copy and 5,154 single-copy SOGs were retrieved, allowing for up to 40 % taxa missing (**Fig. 4A**). This imbalance between the numbers of multi-copy and

single-copy SOGs is attributable to lineage-specific polyploidy events within the core eudicots. As a result of these polyploidy events, the occupancy of single-copy SOGs showed significant decreases in species with high relative ploidy (i.e. orthologous syntenic depth to the grape genome) (**Fig 4B**). Nevertheless, the order-level species tree topologies based on the two gene sets were identical and strongly supported with high posterior probabilities (**Fig. 4C)**, although the tree based on multi-copy SOGs was more robust, with equal or higher posterior probabilities at nearly all nodes (**Fig. 4C)** and there were slight differences in the positions of a few of the species (**Figs. S102-103**).

We observed large incongruences between our phylogeny and the APG IV [33], as nearly half of the orders covered in this study have inconsistent phylogenetic positions (**Fig. 4C)**. For example, the COM clade is not monophyletic and is placed with the malvids, in contrast to that from APG IV (**Fig. 4C)**. However, recent phylogenomics/phylotranscriptomic results are consistent with most of our findings [e.g. 34-40]. For example, the phylogenetic positions of the Fagales, Rosales, Celastrales–Oxalidales–Malpighiales, Myrtales, Cornales and Ericales (**Fig. 4C)** are consistent with the genome-scale phylogenomics based on coalescent-based analysis of 482 single-copy nuclear orthologous sequences [38]. In addition, the phylogenetic positions of the Santalales (sister to superrosids), Aquifoliales (sister to Garryales), and Crossosomatales (sister to fabids + remaining malvids) (**Fig. 4C)** are consistent with the phylogenetic inference based on coalescent tree analysis of 410 single-copy gene families extracted from transcriptome and genome data [35].

Our results further provide interesting phylogenomic insights for the core eudicots. We discovered that Vitales is sister to Saxifragales with high support (**Fig. 4C)**. Additionally, the Vitales + Saxifragales clade is found to be the sister group of the remaining rosids (**Fig. 4C)**, rejecting the hypothesis that Vitales are sister to Saxifragales and all other taxa in the superrosids clade [34-36, 38, 39]. We also found that the Zygophyllales were sister to the malvids clade with high support (**Fig. 4C)**, inconsistent with its placement as sister of the Malvales [35, 39]. Considering that our analyses involved ten thousand syntenic orthologous gene families, and that we have minimized the detrimental effects of the shared γ event, we believe that our results are likely to be more robust than these mentioned above in the depiction of the real tree of life of the core eudicots. Nevertheless, our analyses were limited by taxon sampling (lack of high-quality

genomes), which is expected to be resolved in the near future with the continuing developments and efforts in the field of genome sequencing [41].

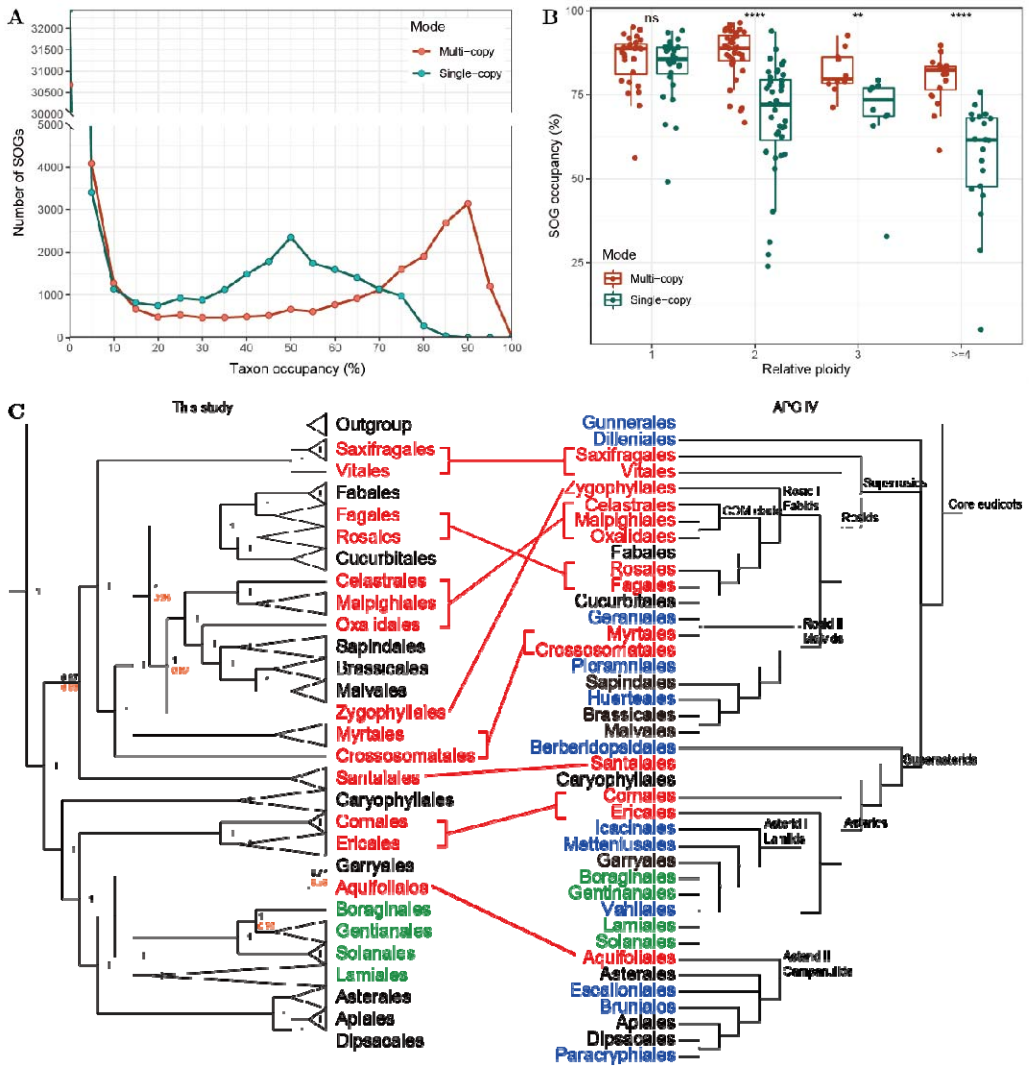


Fig. 4. An example (core eudicots) of phylogenomics based on the *Orthology Index*. **A**) The number of multi-copy and single-copy syntenic orthogroups (SOGs) with different taxon occupancy. **B**) The SOG occupancy in species with different relative ploidy (i.e. orthologous syntenic depth to the grape genome) with at most 40 % taxa missing. Each point represents one species. ns, $p > 0.05$; **, $p <= 0.01$; ****, $p <= 0.0001$; Wilcoxon test. **C**) Phylogenetic relationships within the core eudicots reconstructed in this study versus in APG IV. Conflicting positions are marked in red; unresolved relationships in APG IV are marked in green, and orders not covered in this study are marked in blue. The numbers at the nodes are posterior probabilities from ASTRAL, with the black representing those from the multi-copy SOGs and orange

representing those from the single-copy SOGs (omitted for equal values). Further details of the two trees reconstructed in this study can be found in **Figs. S102-103**.

Limitations

The *Orthology Index* may not perform well in extremely complex cases. For instance, when ΔT is notably small (e.g. radiation following polyploidization in a few generations), *OI* may find it difficult to distinguish between out-paralogy and orthology. This method is also limited in some scenarios when ortholog inference and/or synteny detection is limited. For example, synteny is known not to be conserved in distantly-related lineages [42], suggesting that this method should not be applied in cases involving lineages that are distantly related (e.g. gymnosperms–angiosperms). Additionally, fragmented assemblies, as well as mis-assemblies, can disrupt synteny and subsequently reduce the efficiency of the index. However, this is likely to cease to be a concern in the near future with the fast development of sequencing and assembly techniques.

Conclusions

In summary, we present a human-interpretable and machine-actionable approach to distinguish orthology from out-paralogy for syntenic blocks. The approach can identify orthologous synteny robustly, as validated with nearly 100 representative cases. We have demonstrated the broad and valuable applications of this approach to the reconstruction of evolutionary history in plant genomes, including reconstruction of the tree/network of life, and identification of and placing of polyploidy events on the tree/network. This approach will extend our analytical capacity in evolutionary genomics and might reduce the misleading data generated using some traditional methods.

Methods

Data collection and pre-processing

The genomic data were obtained from public databases or from the corresponding authors, as detailed in **Table S2**.

An all-versus-all BLAST search of protein sequences for each species was conducted pairwise using DIAMOND v0.9.24 [43]. Orthologous relationships were inferred using

OrthoFinder v2.3.1 (Emms and Kelly 2019) (parameters: -M msa). Syntenic/collinear blocks were identified with the `-icl` option of WGDI [10] v0.6.2 (default parameters). The synonymous substitution rate (*Ks*) was calculated for homologous gene pairs using the `-ks` option of WGDI (default parameters).

Distinguishing orthologous and out-paralogous syntenic blocks using the Orthology Index

We propose an index, named the *Orthology Index*, that can distinguish the syntenic blocks of orthology from out-paralogy. The index (*OI*) is defined as:

$$OI = \frac{n}{m},$$

where *m* represents the total number of collinear gene pairs in a pre-inferred collinearity block, and *n* denotes the number of collinear gene pairs pre-inferred as orthologs. Thus, the index represents the proportion of orthologous syntenic gene pairs within a block. It ranges from 0 to 1 and can differentiate orthology from out-paralogy. A higher index value suggests a higher likelihood of orthology, whereas a lower index value suggests a higher likelihood of out-paralogy. From our experience with many different cases, the peak with the highest index can be generally considered to be orthologous (Figs. S1–90).

Using this index as a foundation, we developed user-friendly all-in-one programs for visualization and downstream analyses. The subcommand 'dotplot' enables visualization and evaluation of synteny, with the dots colored by the index or *K*s values. The subcommand 'filter' retrieves orthologous blocks by discarding all blocks with less than a default index value of 0.6. The subcommand 'cluster' groups orthologous syntenic genes into syntenic orthogroups (SOGs), by constructing an orthologous syntenic graph and applying the Markov Cluster (MCL) algorithm [44] to perform graph clustering and break weak links. The subcommand 'outgroup' retrieves syntenic orthologs from outgroups that lack WGDs shared with ingroups. The subcommand 'phylo' reconstructs multi-copy or single-copy gene trees, by aligning protein sequences with MAFFT v7.481 [45], converting protein alignments to codon alignments with PAL2NAL v14 [46], trimming alignments with trimAl v1.2 [47] (parameter: -automated1) and reconstructing maximum-likelihood trees with IQ-TREE v2.2.0.3 [48]. These gene trees serve as input to infer a

species tree with the coalescence-based method ASTRAL-Pro v1.10.1.3 [49]. The default threshold for missing taxa is set to 40 %, according to a recent study [50].

This tool is implemented in Python3 and supports synteny outputs from state-of-the-art synteny detectors, including MCscan/JCVI [17], MCscanX [11] and WGDI [10], as well as orthology outputs from OrthoFinder2 [13], OrthoMCL [14] and other tools upon request. The tool can be easily installed using the conda environment or the Apptainer/Singularity container system [51], and can be seamlessly integrated into other pipelines. The source code is accessible on GitHub (https://github.com/zhangrengang/OrthoIndex).

Code availability

The codes and typical examples used in this study can be found on GitHub (https://github.com/zhangrengang/OrthoIndex).

Data availability

The codon alignments and gene trees of the core eudicots are available from Figshare (https://doi.org/10.6084/m9.figshare.24174930).

Acknowledgments

We thank Professors Ying-Xiong Qiu, Xiao-Ming Song, Quan-Jun Hu, Tao Zhou, and Zhao-Ying Liu for generously sharing their genomic data, and the other researchers who have already released their genomic data publicly.

Author contributions

RGZ and YPM conceived and designed the study; RGZ, HYS and HS programmed and tested the code; RGZ collected and analyzed the data; RGZ and HYS prepared figures; KHJ, RGZ and HYS drafted the manuscript; YPM revised the manuscript; all authors approved the final manuscript.

Competing interests

The authors declare no competing interests.

Supplementary Figures and Tables

Figs. S1-90. ***Orthology Index* in genome evolution of species X and species Y.** Refer to **Fig. 1** for detailed descriptions.

Fig. S91. *Orthology Index*-colored dot plots showing orthologous syntenic relationships of *Centella asiatica* : *Vitis vinifera* (2:1 orthologous synteny depth ratio), *Aralia elata* : *Vitis* vinifera (2:1), *Angelica sinensis* : *Vitis* vinifera (4:1), and *Panax ginseng* : *Vitis* vinifera (4:1).

Fig. S92. Macro-synteny phylogenies inferred based on concatenated 1:2:2 (*Vitis* vinifera : *Aralia* elata : *Centella* asiatica) orthologous syntenic genes. The leaf labels on the trees are chromosome numbers. *n*, number of orthologous syntenic genes involved. Numbers at nodes represent the bootstrap values (percentage).

Fig. S93. *Orthology Index*-colored dot plots showing orthologous syntenic relationships between *Centella* asiatica and the Araliaceae (*Centella* asiatica : *Aralia* elata = 1:1, *Centella* asiatica : *Eleutherococcus* senticosus = 1:2, *Centella* asiatica : *Panax ginseng* = 1:2, *Centella* asiatica : *Panax notoginseng* = 1:1).

Fig. S94. *Orthology Index*-colored dot plots showing orthologous syntenic relationships between *Aralia* elata and the Apioideae (all 1:2).

Fig. S95. *Orthology Index*-colored dot plots showing orthologous syntenic relationships between *Centella* asiatica and the Apioideae (all 1:2).

Fig. S96. *Orthology Index*-colored dot plots showing orthologous syntenic relationships within the Apioideae (all 1:1).

Fig. S97. *Orthology Index*-colored dot plots showing orthologous syntenic relationships within the Araliaceae (*Panax* ginseng : *Panax* notoginseng = 2:1, *Aralia* elata : *Panax* notoginseng = 1:1, *Eleutherococcus* senticosus : *Panax* notoginseng = 2:1). The other orthologous syntenic depth ratios (*Aralia elata* : *Eleutherococcus* senticosus = 1:2, *Aralia elata* : *Panax* ginseng = 1:2, *Eleutherococcus senticosus* : *Panax* ginseng = 2:2) can be found in Fig. S57-59.

Fig. S98. *Orthology Index*-colored dot plots showing orthologous syntenic relationships between *Arabidopsis thaliana* × *A. arenosa* and *A. thaliana* + *A. arenosa*.

Fig. S99. *Orthology Index*-colored dot plots showing orthologous syntenic relationships between *Arachis hypogaea* and *A. duranensis* + *A. ipaensis*.

Fig. S100. *Orthology Index*-colored dot plots showing orthologous syntenic relationships between *Triticum turgidum* and *T. aestivum*.

Fig. S101. *Orthology Index*-colored dot plots showing orthologous syntenic relationships between *Papaver somniferum* and *P. setigerum*.

Fig. S102. Phylogenetic relationships within the core eudicots based on 12,277 multi-copy SOGs. The numbers at the nodes are posterior probabilities from ASTRAL. Bar, 3.0 coalescent units.

Fig. S103. Phylogenetic relationships within the core eudicots based on 5,154 single-copy SOGs. The numbers at the nodes are posterior probabilities from ASTRAL. Bar, 3.0 coalescent units.

Table S1. Cases with shared polyploidization event(s).

Table S2. Genomic data used in this study (accessed before 2023-06-29).

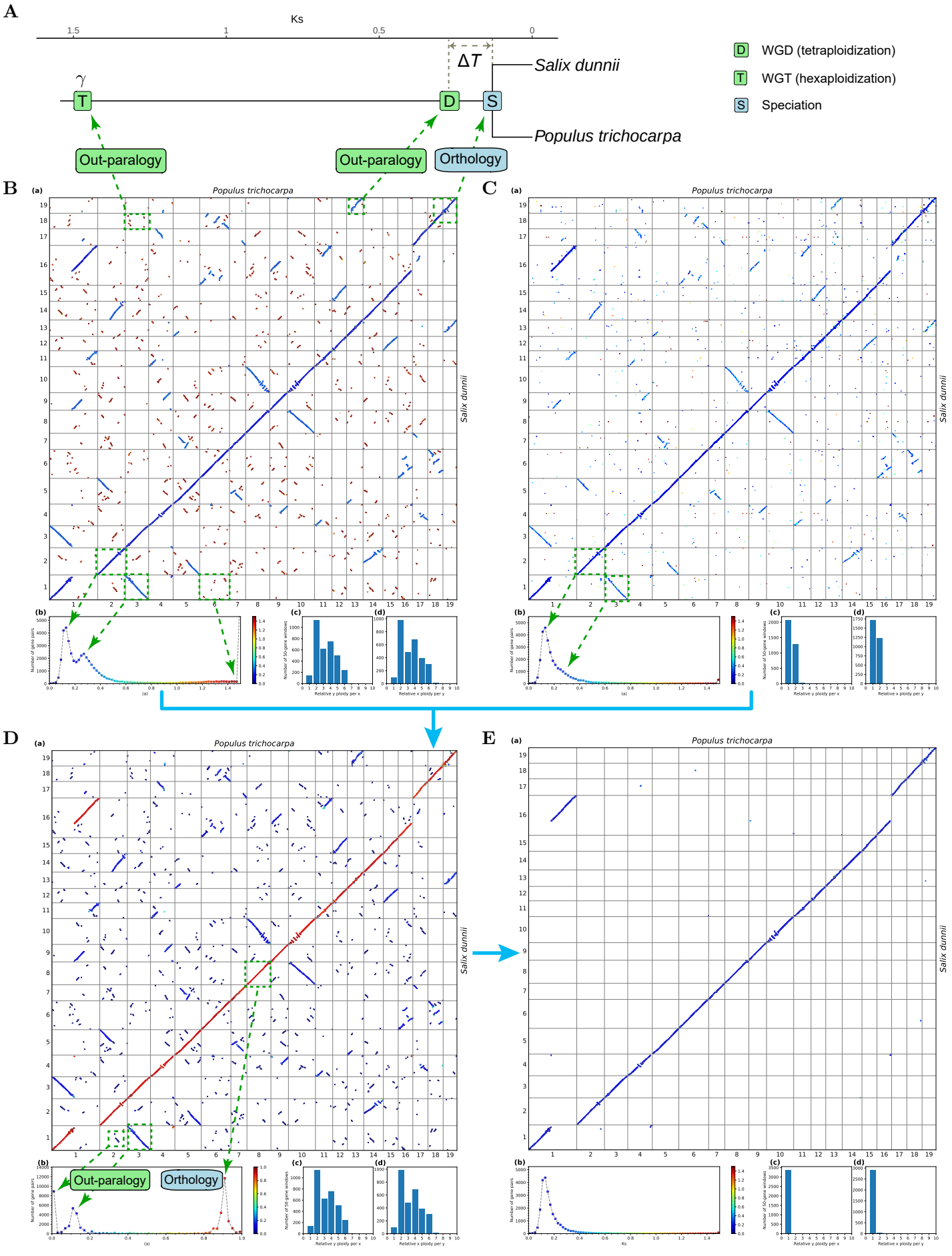
References

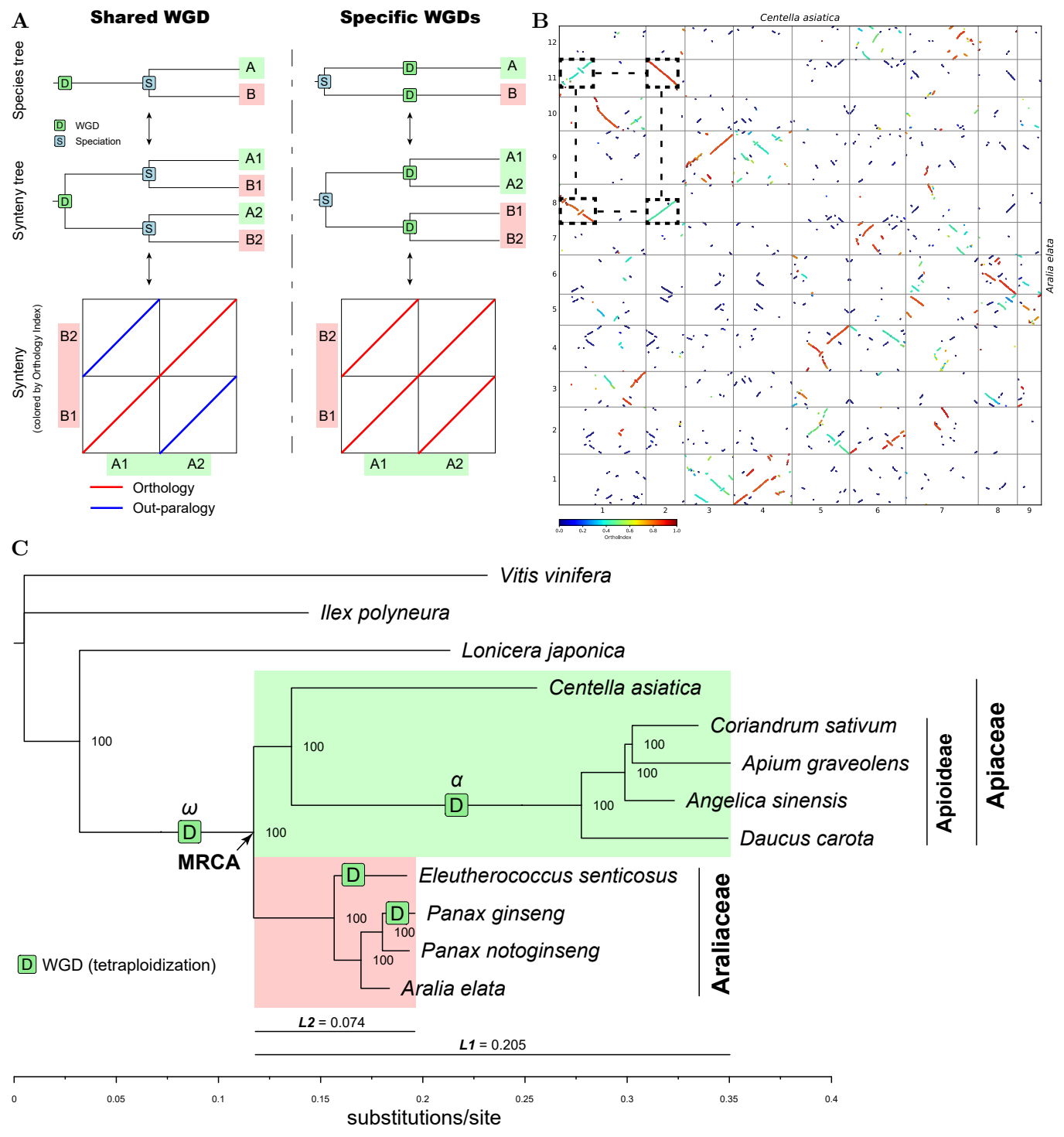
1. Steenwyk JL, Li Y, Zhou X, Shen X, Rokas A. Incongruence in the phylogenomics era. Nat Rev Genet. 2023.
2. Zhang R, Lu C, Li G, Lv J, Wang L, Wang Z, Chen Z, Liu D, Zhao Y, Shi T *et al.* Subgenome-aware analyses suggest a reticulate allopolyploidization origin in three *Papaver* genomes. Nat Commun. 2023; 14:2204.
3. Ma J, Sun P, Wang D, Wang Z, Yang J, Li Y, Mu W, Xu R, Wu Y, Dong C *et al*. The *Chloranthus sessilifolius* genome provides insight into early diversification of angiosperms. Nat Commun. 2021; 12:6929.
4. Guo C, Luo Y, Gao LM, Yi TS, Li HT, Yang JB, Li DZ. Phylogenomics and the flowering plant tree of life. J Integr Plant Biol. 2023; 65:299–323.
5. Yang X, Gao S, Guo L, Wang B, Jia Y, Zhou J, Che Y, Jia P, Lin J, Xu T *et al*. Three chromosome-scale *Papaver* genomes reveal punctuated patchwork evolution of the morphinan and noscapine biosynthesis pathway. Nat Commun. 2021; 12:6030.
6. Yang X, Gao S, Xu T, Wang B, Jia Y, Ye K. Reply to "Subgenome-aware analyses suggest a reticulate allopolyploidization origin in three *Papaver* genomes". Nat Commun. 2023; 14:2203.
7. An X, Gao K, Chen Z, Li J, Yang X, Yang X, Zhou J, Guo T, Zhao T, Huang S *et al*. High quality haplotype-resolved genome assemblies of *Populus tomentosa* Carr., a stabilized interspecific hybrid species that is widespread in Asia. Mol Ecol Resour. 2021; 22:786–802.
8. Zhang T, Qiao Q, Du X, Zhang X, Hou Y, Wei X, Sun C, Zhang R, Yun Q, Crabbe MJC *et al.* Cultivated hawthorn (*Crataegus pinnatifida* var. *major*) genome sheds light on the evolution of Maleae (apple tribe). J Integr Plant Biol. 2022; 64:1487–501.
9. Tang H, Lyons E, Pedersen B, Schnable JC, Paterson AH, Freeling M. Screening synteny blocks in pairwise genome comparisons through integer programming. BMC Bioinformatics. 2011; 12:102.
10. Sun P, Jiao B, Yang Y, Shan L, Li T, Li X, Xi Z, Wang X, Liu J. WGDI: A user-friendly toolkit for evolutionary analyses of whole-genome duplications and ancestral karyotypes. Mol Plant. 2022; 15:1841–51.
11. Wang Y, Tang H, DeBarry JD, Tan X, Li J, Wang X, Lee TH, Jin H, Marler B, Guo H *et al.* MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. 2012; 40:e49.
12. Chen J, Huang Y, Brachi B, Yun Q, Zhang W, Lu W, Li H, Li W, Sun X, Wang G *et al.* Genome-wide analysis of Cushion willow provides insights into alpine plant divergence in a biodiversity hotspot. Nat Commun. 2019; 10:5230.
13. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 2019; 20:238.
14. Li L, Stoeckert CJ, Roos DS. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. 2003; 13:2178–89.
15. Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science. 2006; 313:1596–604.
16. Dai X, Hu Q, Cai Q, Feng K, Ye N, Tuskan GA, Milne R, Chen Y, Wan Z, Wang Z. The willow genome and divergent evolution from poplar after the common genome duplication. Cell Res.

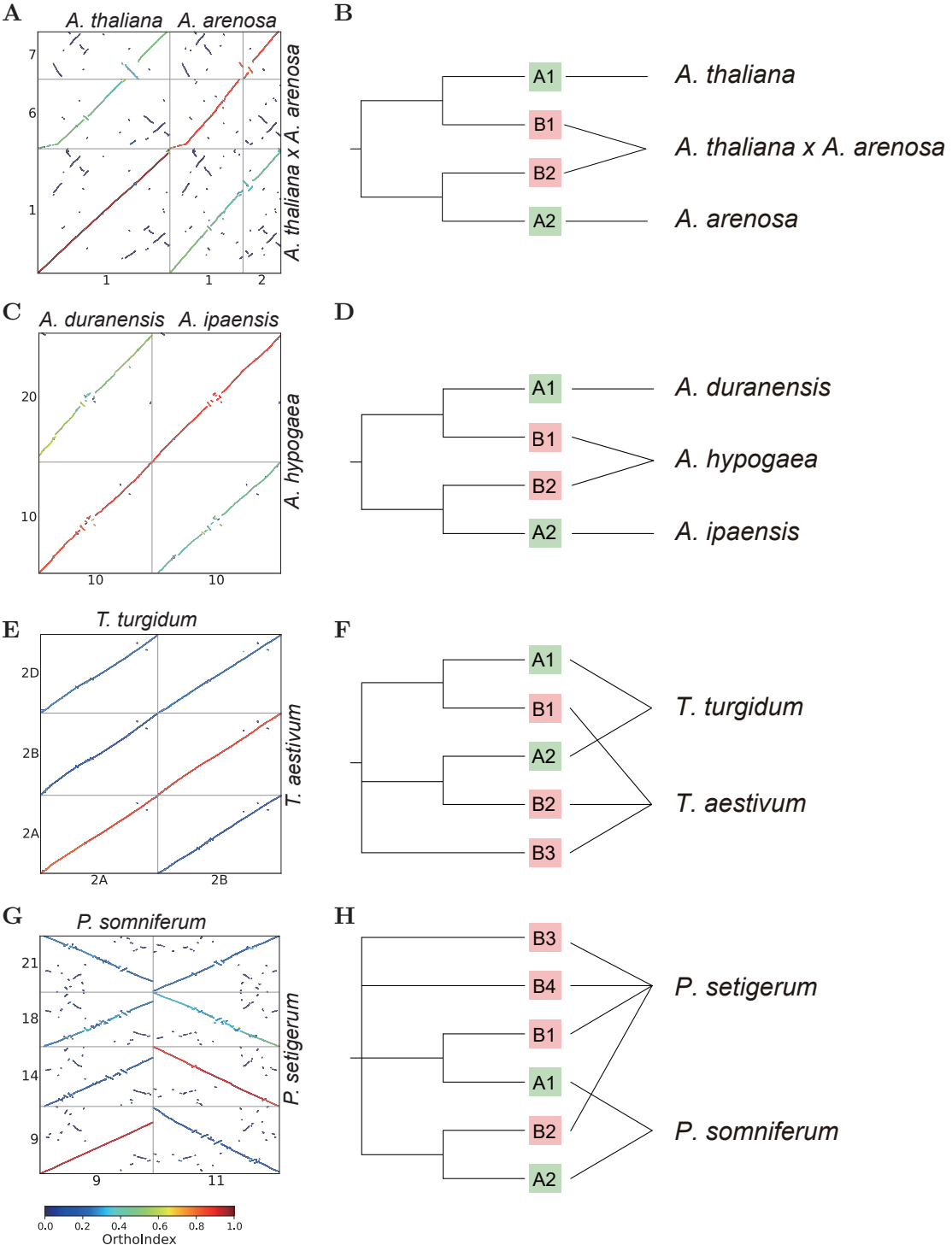
2014; 24:1274.
17. Tang H, Bowers JE, Wang X, Ming R, Alam M, Paterson AH. Synteny and collinearity in plant genomes. Science. 2008; 320:486–8.
18. He L, Jia KH, Zhang RG, Wang Y, Shi TL, Li ZC, Zeng SW, Cai XJ, Wagner ND, Hörandl E *et al*. Chromosome□scale assembly of the genome of *Salix dunnii* reveals a male□heterogametic sex determination system on chromosome 7. Mol Ecol Resour. 2021; 21:1966–82.
19. Song X, Sun P, Yuan J, Gong K, Li N, Meng F, Zhang Z, Li X, Hu J, Wang J *et al.* The celery genome sequence reveals sequential paleo-polyploidizations, karyotype evolution and resistance gene reduction in apiales. Plant Biotechnol J. 2021; 19:731–44.
20. Jiang Z, Tu L, Yang W, Zhang Y, Hu T, Ma B, Lu Y, Cui X, Gao J, Wu X *et al.* The chromosome-level reference genome assembly for *Panax notoginseng* and insights into ginsenoside biosynthesis. Plant Commun. 2021; 2:100113.
21. Wang Y, Zhang H, Ri HC, An Z, Wang X, Zhou J, Zheng D, Wu H, Wang P, Yang J *et al.* Deletion and tandem duplications of biosynthetic genes drive the diversity of triterpenoids in *Aralia elata*. Nat Commun. 2022; 13:2224.
22. Yang Z, Chen S, Wang S, Hu Y, Zhang G, Dong Y, Yang S, Miao J, Chen W, Sheng J. Chromosomal□scale genome assembly of *Eleutherococcus senticosus* provides insights into chromosome evolution in Araliaceae. Mol Ecol Resour. 2021; 21:2204–20.
23. Otto SP, Whitton J. Polyploid incidence and evolution. Annu Rev Genet. 2000; 34:401–37.
24. Jia K, Wang Z, Wang L, Li G, Zhang W, Wang X, Xu F, Jiao S, Zhou S, Liu H *et al*. SubPhaser: a robust allopolyploid subgenome phasing method based on subgenome-specific *k*-mers. New Phytol. 2022; 235:801–9.
25. Marcussen T, Sandve SR, Heier L, Spannagl M, Pfeifer M, Jakobsen KS, Wulff BBH, Steuernagel B, Mayer KFX, Olsen OA *et al.* Ancient hybridizations among the ancestral genomes of bread wheat. Science. 2014; 345:1250092.
26. Edger PP, Poorten TJ, VanBuren R, Hardigan MA, Colle M, McKain MR, Smith RD, Teresi SJ, Nelson ADL, Wai CM *et al.* Origin and evolution of the octoploid strawberry genome. Nat Genet. 2019; 51:541–7.
27. Chalhoub B, Denoeud F, Liu S, Parkin IA, Tang H, Wang X, Chiquet J, Belcram H, Tong C, Samans B *et al*. Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. Science. 2014; 345:950–3.
28. Li JT, Wang Q, Huang YM, Li QS, Cui MS, Dong ZJ, Wang HW, Yu JH, Zhao YJ, Yang CR *et al.* Parallel subgenome structure and divergent expression evolution of allo-tetraploid common carp and goldfish. Nat Genet. 2021; 53:1493–503.
29. Jiang X, Song Q, Ye W, Chen ZJ. Concerted genomic and epigenomic changes accompany stabilization of *Arabidopsis* allopolyploids. Nat Ecol Evol. 2021; 5:1382–93.
30. Zhuang W, Chen H, Yang M, Wang J, Pandey MK, Zhang C, Chang W, Zhang L, Zhang X, Tang R *et al.* The genome of cultivated peanut provides insight into legume karyotypes, polyploid evolution and crop domestication. Nat Genet. 2019; 51:865–76.
31. Xiong H, Wang D, Shao C, Yang X, Yang J, Ma T, Davis CC, Liu L, Xi Z. Species tree estimation and the impact of gene loss following whole-genome duplication. Syst Biol. 2022; 71:1348–61.
32. Van de Peer Y, Mizrachi E, Marchal K. The evolutionary significance of polyploidy. Nat Rev Genet. 2017; 18:411–24.

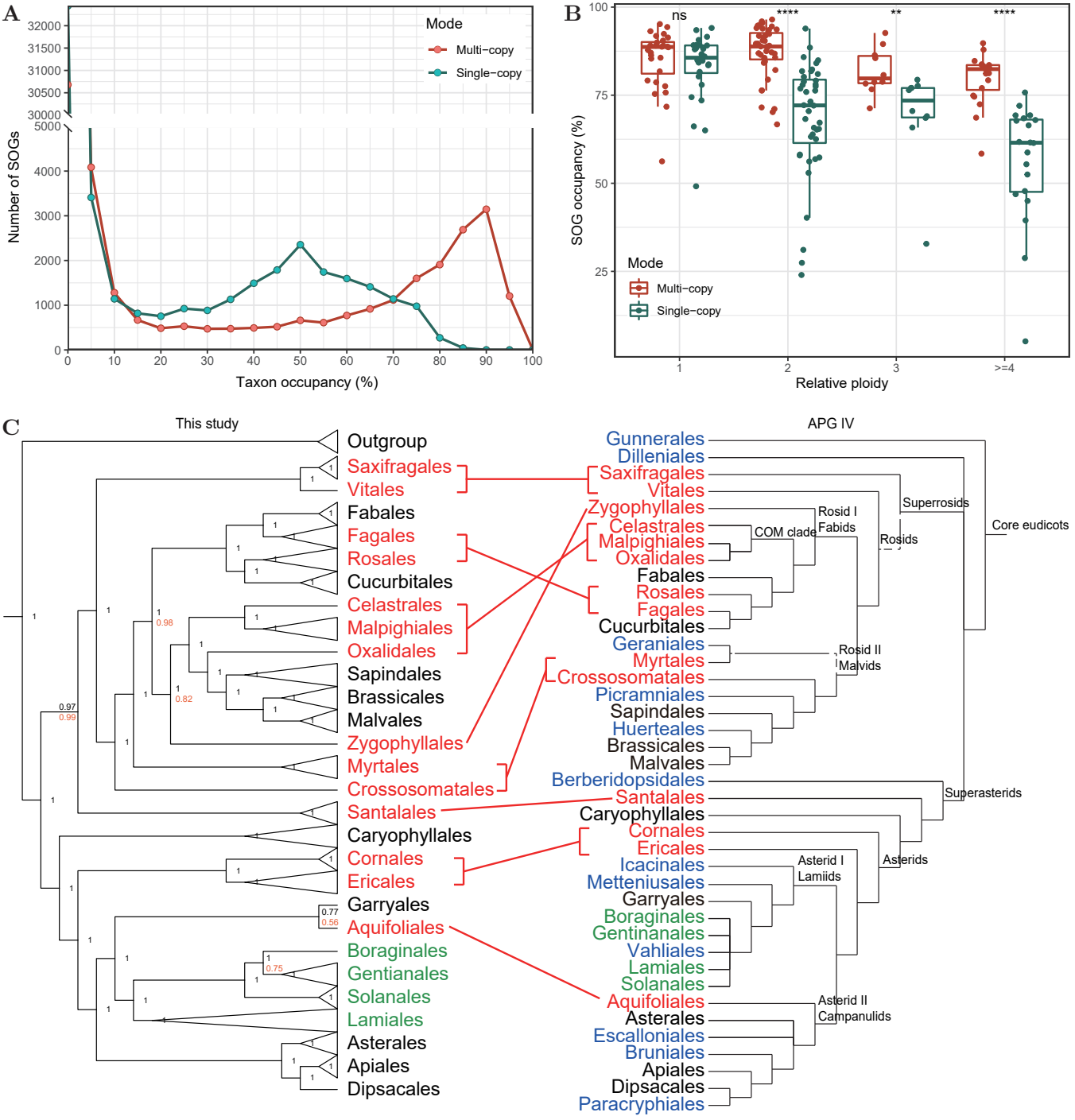
33. The Angiosperm Phylogeny Group. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. Bot J Linn Soc. 2016; 181:1–20.
34. Zeng L, Zhang N, Zhang Q, Endress PK, Huang J, Ma H. Resolution of deep eudicot phylogeny and their temporal diversification using nuclear genes from transcriptomic and genomic datasets. New Phytol. 2017; 214:1338.
35. Leebens-Mack JH, Barker MS, Carpenter EJ, Deyholos MK, Gitzendanner MA, Graham SW, Grosse I, Li Z, Melkonian M, Mirarab S *et al*. One thousand plant transcriptomes and the phylogenomics of green plants. Nature. 2019; 574:679–85.
36. Yang L, Su D, Chang X, Foster CSP, Sun L, Huang C, Zhou X, Zeng L, Ma H, Zhong B. Phylogenomic insights into deep phylogeny of angiosperms based on broad nuclear gene sampling. Plant Commun. 2020; 1:100027.
37. Zhang C, Zhang T, Luebert F, Xiang Y, Huang C, Hu Y, Rees M, Frohlich MW, Qi J, Weigend M *et al*. Asterid phylogenomics/phylotranscriptomics uncover morphological evolutionary histories and support phylogenetic placement for numerous whole genome duplications. Mol Biol Evol. 2020:msaa160.
38. Hu H, Sun P, Yang Y, Ma J, Liu J. Genome□scale angiosperm phylogenies based on nuclear, plastome, and mitochondrial datasets. J Integr Plant Biol. 2023; 65:1479–89.
39. Liu L, Chen M, Folk RA, Wang M, Zhao T, Shang F, Soltis DE, Li P. Phylogenomic and syntenic data demonstrate complex evolutionary processes in early radiation of the rosids. Mol Ecol Resour. 2023.
40. Baker WJ, Bailey P, Barber V, Barker A, Bellot S, Bishop D, Botigué LR, Brewer G, Carruthers T, Clarkson JJ *et al*. A comprehensive phylogenomic platform for exploring the angiosperm tree of life. Syst Biol. 2022; 71:301–19.
41. Lewin HA, Richards S, Lieberman Aiden E, Allende ML, Archibald JM, Bálint M, Barker KB, Baumgartner B, Belov K, Bertorelle G *et al*. The Earth BioGenome Project 2020: Starting the clock. Proc Natl Acad Sci U S A. 2022; 119:e2115635118.
42. Kristensen DM, Wolf YI, Mushegian AR, Koonin EV. Computational methods for Gene Orthology inference. Brief Bioinform. 2011; 12:379–91.
43. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat Methods. 2015; 12:59–60.
44. van Dongen S. Graph clustering via a discrete uncoupling process. SIAM J Matrix Anal Appl. 2008; 30:121–41.
45. Standley DM, Katoh K. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013; 30:772–80.
46. Suyama M, Torrents D, Bork P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res. 2006; 34:W609–12.
47. Capella-Gutierrez S, Silla-Martinez J, Gabaldon T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009; 25:1972–3.
48. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. 2020; 37:1530–4.
49. Zhang C, Mirarab S. ASTRAL-Pro 2: ultrafast species tree reconstruction from multi-copy gene family trees. Bioinformatics. 2022; 38:4949–50.

50. Morel B, Williams TA, Stamatakis A, Schwartz R. Asteroid: a new algorithm to infer species trees from gene trees under high proportions of missing data. Bioinformatics. 2023; 39:btac832.
51. Kurtzer GM, Sochat V, Bauer MW. Singularity: Scientific containers for mobility of compute. PLoS One. 2017; 12:e177459.


A
Ks
1.5
1
0.5
0
ΔT
Salix dunnii
γ
T
D
S
Populus trichocarpa
D WGD (tetraploidization)
T WGT (hexaploidization)
S Speciation
Out-paralogy
Out-paralogy
Orthology
B
C
D
E
Populus trichocarpa
Salix dunnii
Out-paralogy
Orthology


A
Shared WGD
Specific WGDs
Species tree
Synteny tree
Synteny
(colored by Orthology Index)
D WGD
S Speciation
A
B
A1
B1
A2
B2
Orthology
Out-paralogy
B
Centella asiatica
Aralia elata
Orthoindex
C
Vitis vinifera
Ilex polyneura
Lonicera japonica
Centella asiatica
Coriandrum sativum
Apium graveolens
Angelica sinensis
Daucus carota
Apioideae
Apiaceae
Eleutherococcus senticosus
Panax ginseng
Panax notoginseng
Aralia elata
Araliaceae
ω
α
MRCA
D WGD (tetraploidization)
L2 = 0.074
L1 = 0.205
substitutions/site


A
A. thaliana
A. arenosa
A. thaliana x A. arenosa
7
6
1
1
1
2
B
A1
B1
B2
A2
A. thaliana
A. thaliana x A. arenosa
A. arenosa
C
A. duranensis
A. ipaensis
A. hypogaea
20
10
10
10
D
A1
B1
B2
A2
A. duranensis
A. hypogaea
A. ipaensis
E
T. turgidum
T. aestivum
2D
2B
2A
2A
2B
F
A1
B1
A2
B2
B3
T. turgidum
T. aestivum
G
P. somniferum
P. setigerum
21
18
14
9
9
11
0.0
0.2
0.4
0.6
0.8
1.0
OrthoIndex
H
B3
B4
B1
A1
B2
A2
P. setigerum
P. somniferum


A
Mode
Multi−copy
Single−copy
Number of SOGs
Taxon occupancy (%)
B
ns
****
**
****
SOG occupancy (%)
Relative ploidy
>=4
C
This study
Outgroup
Saxifragales
Vitales
Fabales
Fagales
Rosales
Cucurbitales
Celastrales
Malpighiales
Oxalidales
Sapindales
Brassicales
Malvales
Zygophyllales
Myrtales
Crossosomatales
Santalales
Caryophyllales
Cornales
Ericales
Garryales
Aquifoliales
Boraginales
Gentianales
Solanales
Lamiales
Asterales
Apiales
Dipsacales
APG IV
Gunnerales
Dilleniales
Saxifragales
Vitales
Zygophyllales
Celastrales
Malpighiales
Oxalidales
Fabales
Rosales
Fagales
Cucurbitales
Geraniales
Myrtales
Crossosomatales
Picramniales
Sapindales
Huerteales
Brassicales
Malvales
Berberidopsidales
Santalales
Caryophyllales
Cornales
Ericales
Icacinales
Metteniusales
Garryales
Boraginales
Gentinanales
Vahliales
Lamiales
Solanales
Aquifoliales
Asterales
Escalloniales
Bruniales
Apiales
Dipsacales
Paracryphiales
Rosid I
Fabids
COM clade
Rosid II
Malvids
Superrosids
Rosids
Core eudicots
Superasterids
Asterids
Asterid I
Lamiids
Asterid II
Campanulids