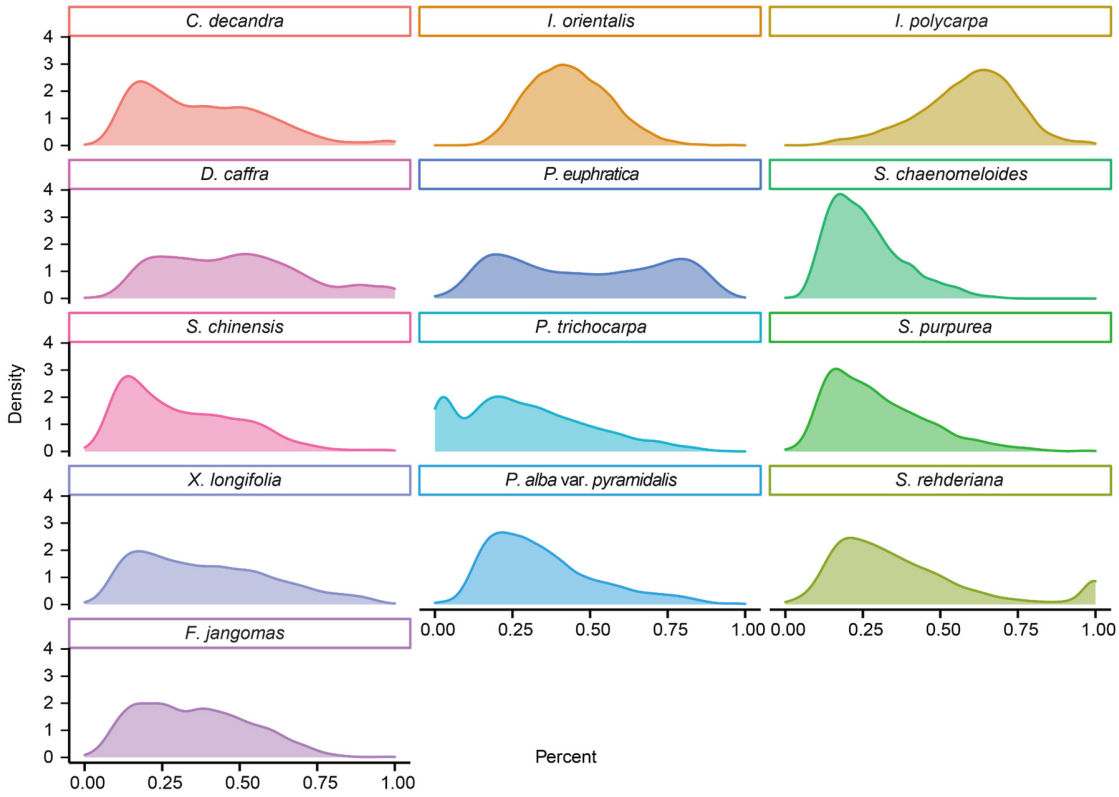
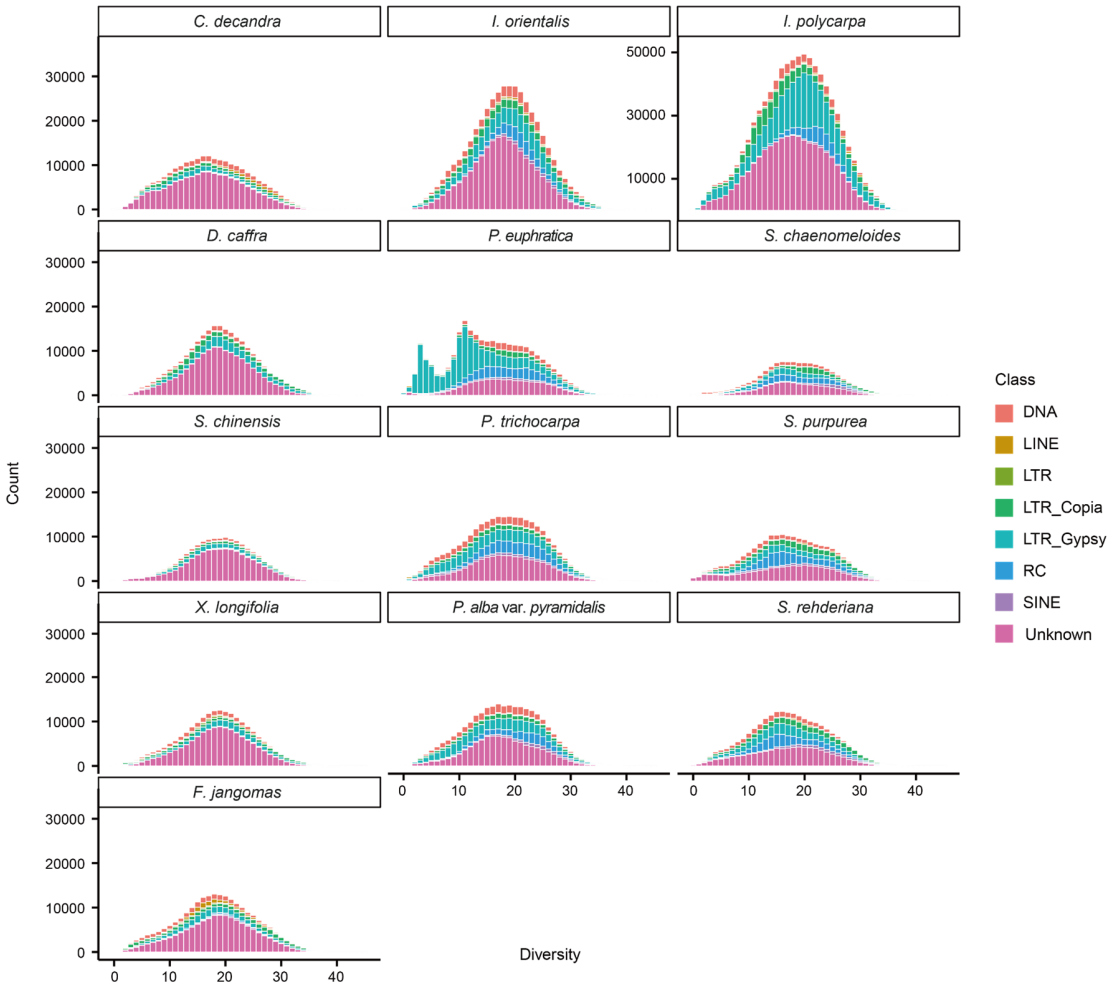
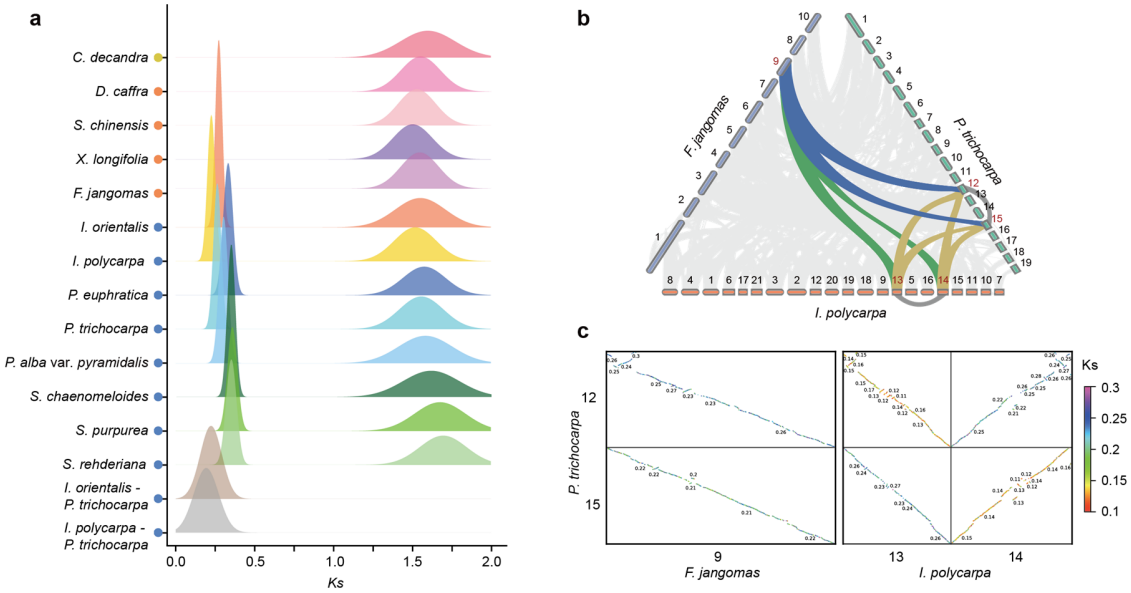


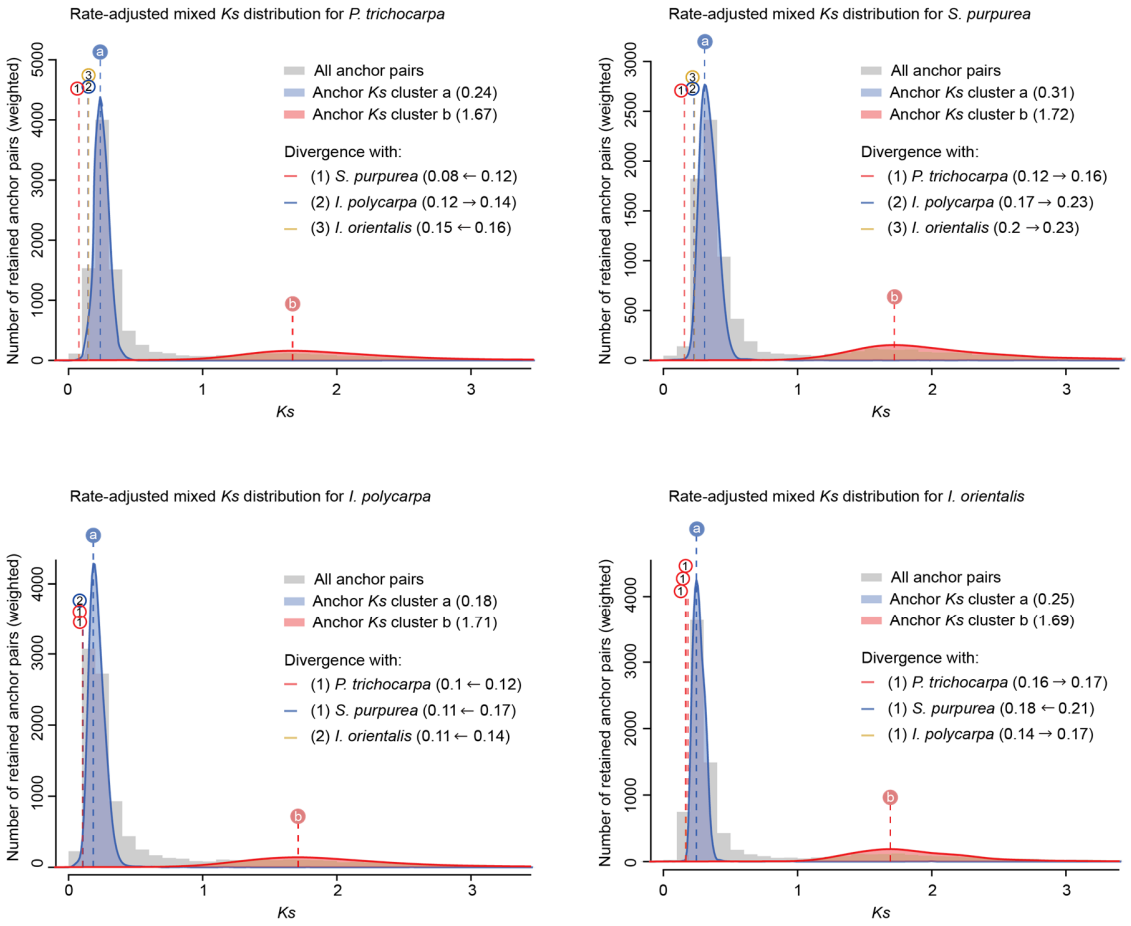
Supplementary Fig. 1. The distribution of repeat content in the bins of non-overlapping 100 Kb windows.



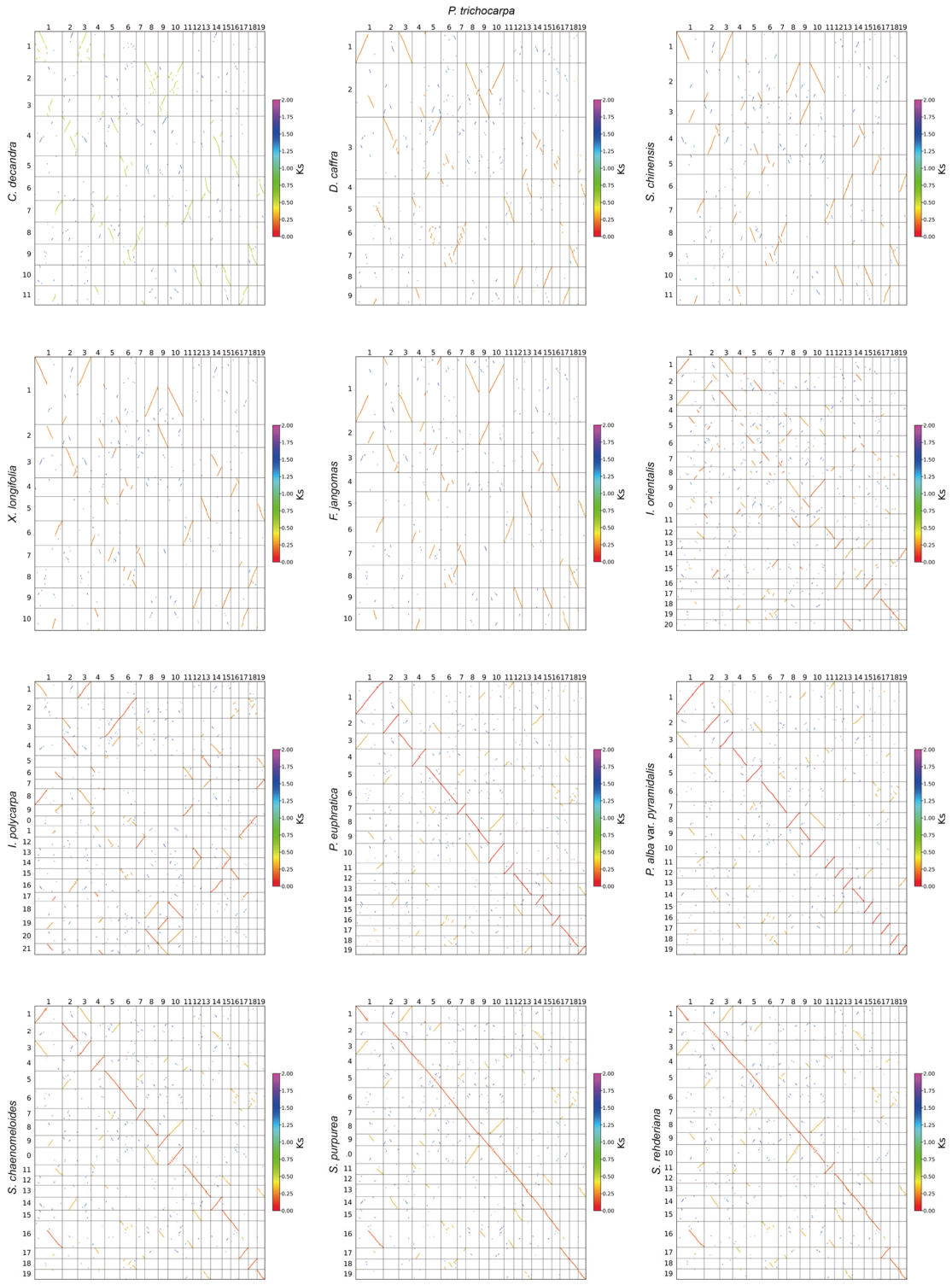
Supplementary Fig. 2. Evolutionary history of TE super-families in 13 Salicaceae species. Repeat landscape depicts the counts of repeat classes and the Kimura divergence from consensus sequences.



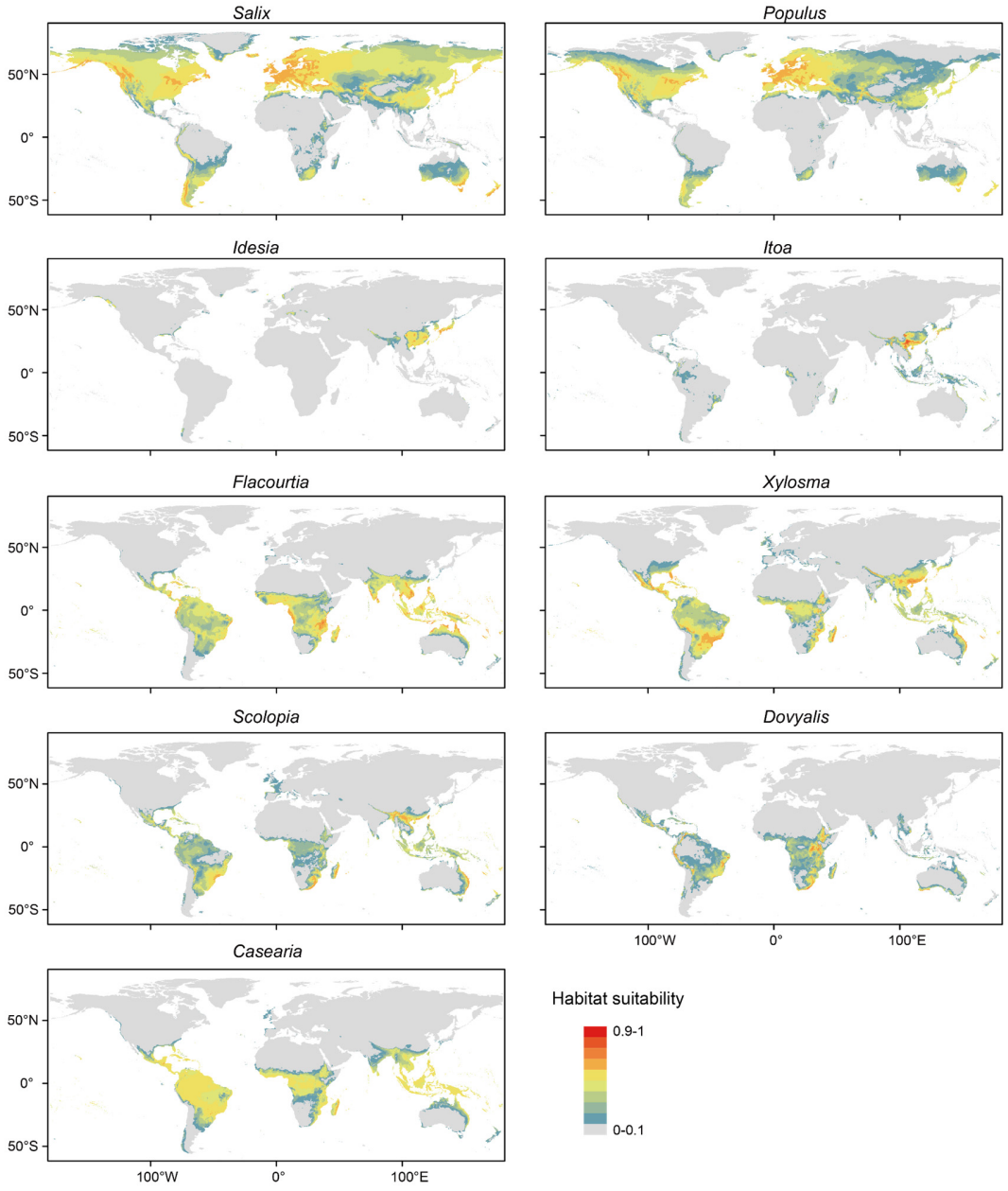
Supplementary Fig. 3. The whole-genome duplication history of species in the Salicaceae family. a, The distribution of synonymous substitutions (*K*s) of intragenomic and intergenomic synteny blocks. **b,** Synteny patterns between genomic regions from three species with highlighted instances of a 1:2:2 relationship involving Chr 9 in *F. jangomas*, Chr 12 and 15 in *P. trichocarpa*, and Chr13 and 14 in *I. polycarpa*. **c,** The Syntenic blocks among representative chromosomes of the three species in (b). The median *K*s value is displayed for each corresponding block.



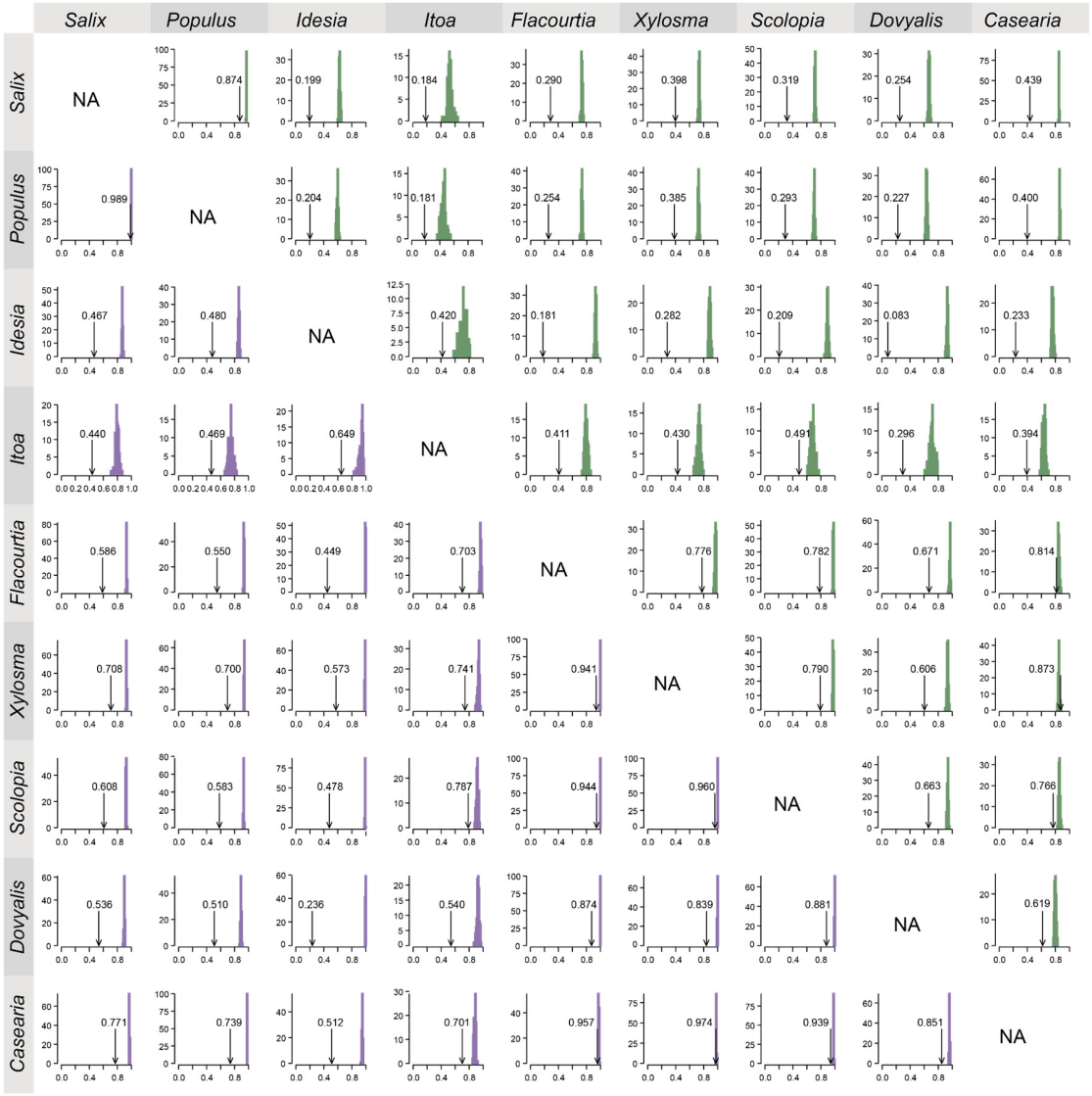
Supplementary Fig. 4. Rate-adjusted mixed paralog-ortholog *Ks* distributions for Clade I species. The colored dashed lines labeled with letters (a, b) indicate WGD events, where 'a' represents the recent 'salicoid' WGD event and 'b' represents the core-eudicot-common γ event. The colored dashed lines labeled with numbers (1, 2, 3) represent the divergence of the focal species from other species. The numbers and arrows in parentheses in the legend indicate the *Ks* values before and after substitution rate adjustment by ksrates.



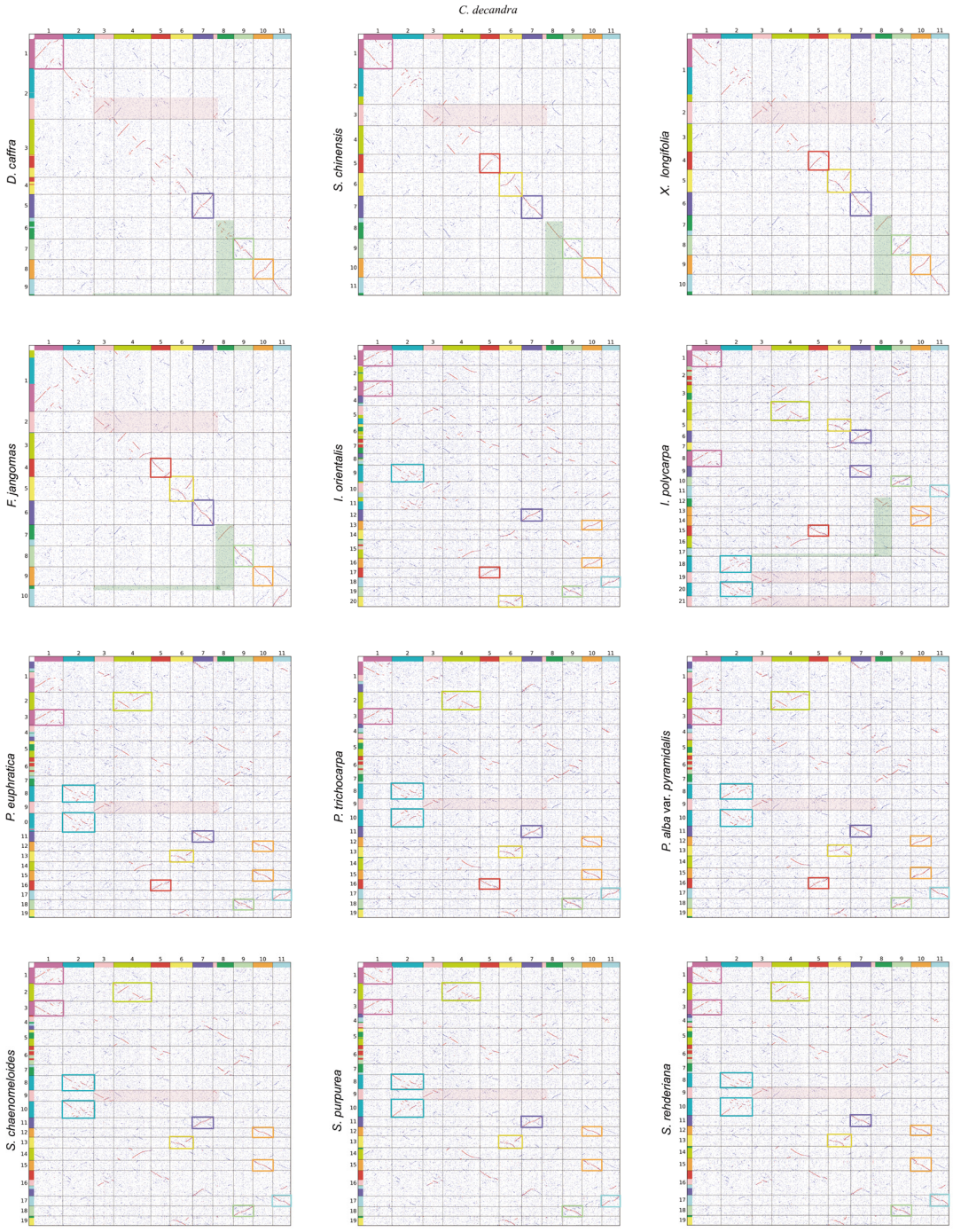
Supplementary Fig. 5. Syntenic block dotplot between Salicaceae species and *P. trichocarpa*. All the blocks contain ≥ 5 syntenic gene pairs. The color representing the mean *Ks* values of each syntenic block.



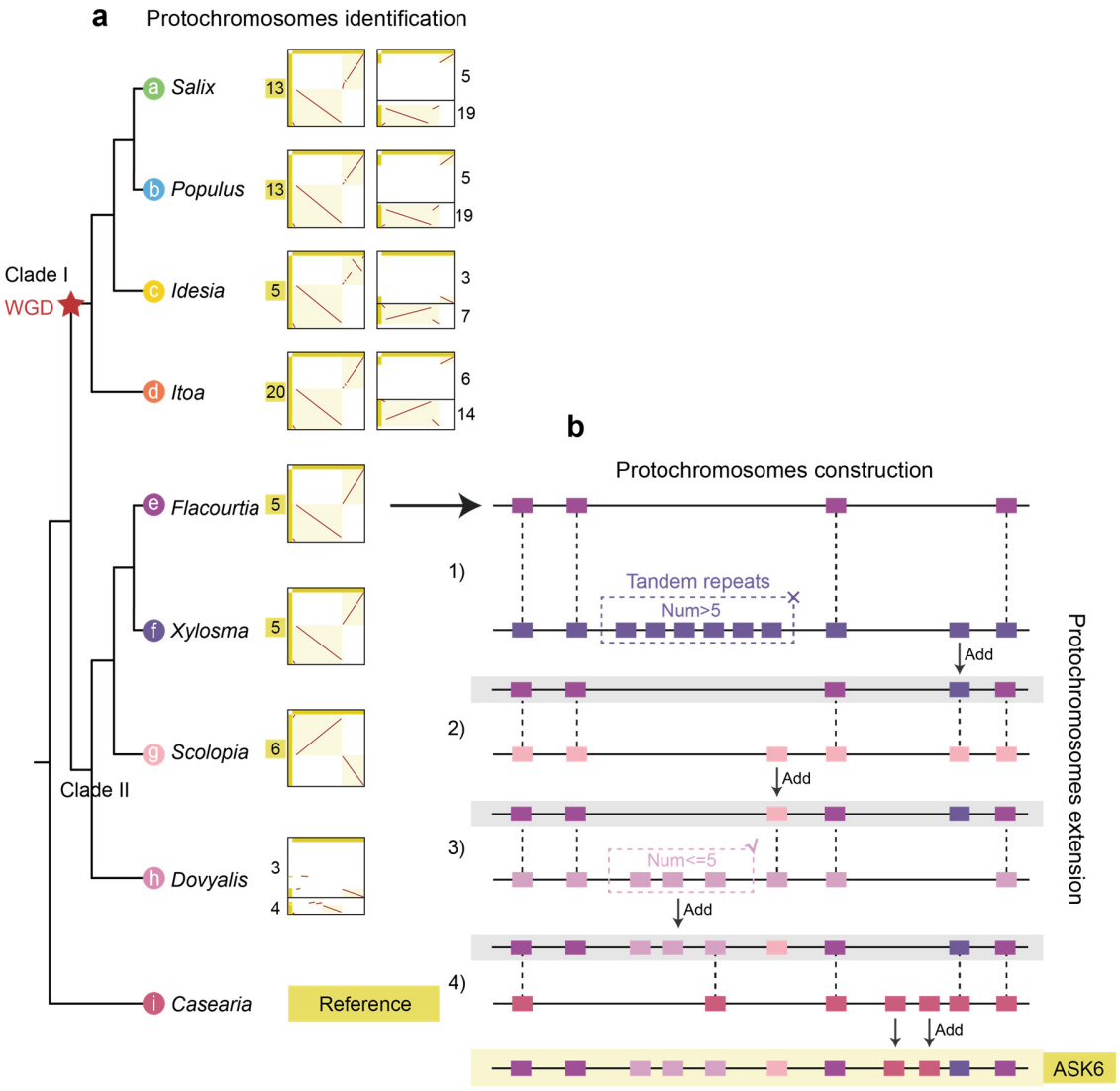
Supplementary Fig. 6. The present predicted distribution of several genera based on ecological niche modelling using MaxEnt. The color represents the current habitat suitability, with warmer color indicating the high suitability degree for species presence.



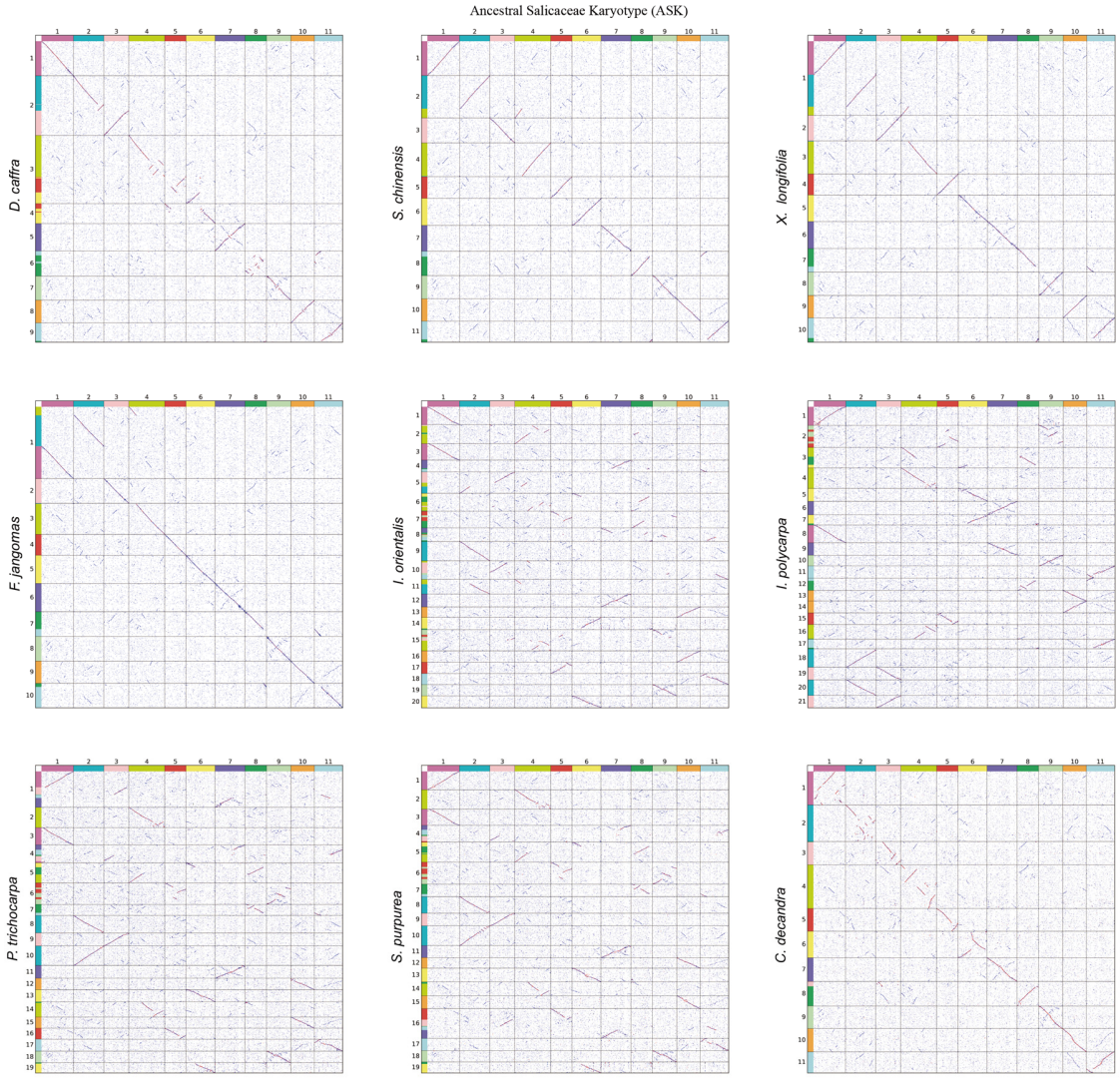
Supplementary Fig. 7. Results of the identity test for pairwise comparisons of different genera. Black arrows refer to the actual niche overlap (Schoener's D and Hellinger's I), Bars show the result of identity test with 100 replicates by using ENMtools. Schoener's D and Hellinger's I are at the top (green) and bottom (purple) of the matrix, respectively. See Supplementary Table 7 for values of Hellinger's I, Schoener's D, and P-values.



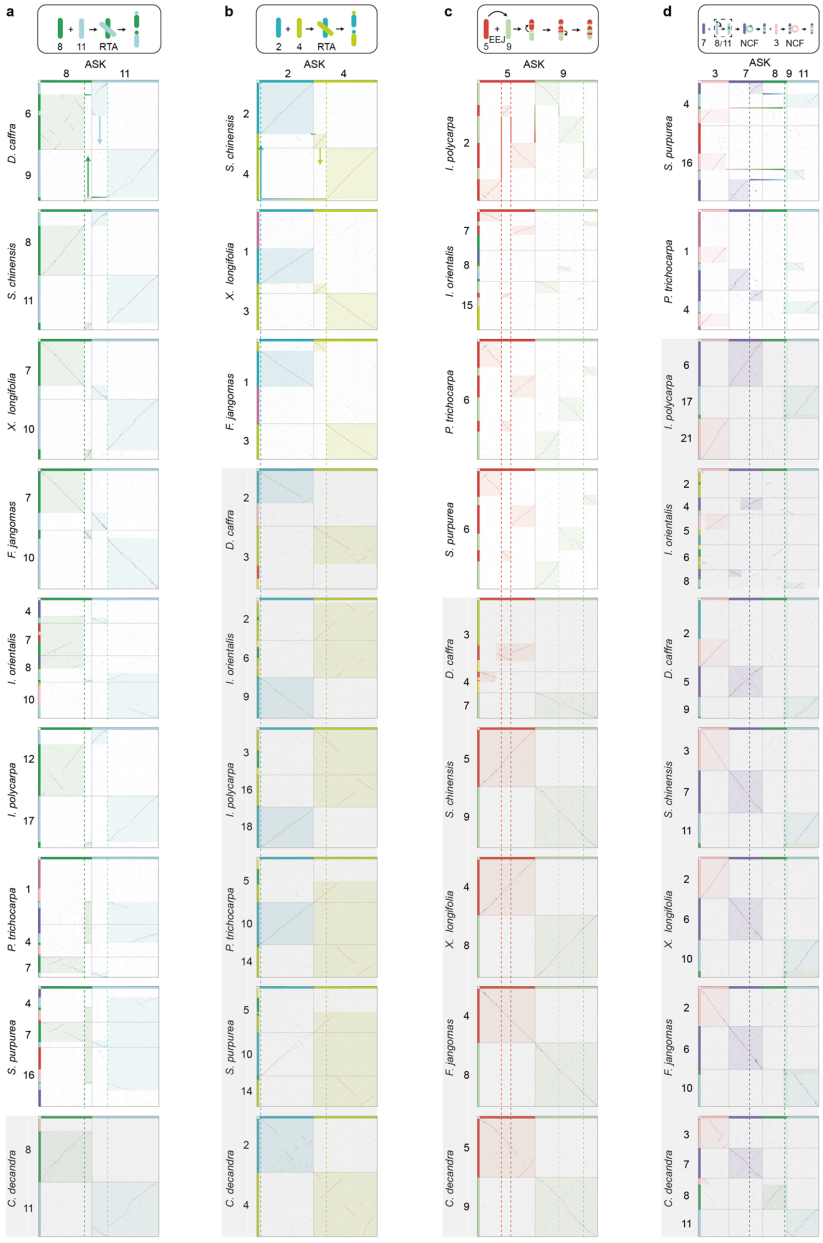
Supplementary Fig. 8. Synteny plot between Salicaceae species and *C. decandra*. Different colors represent 11 protochromosomes. The boxes show independent chromosomes. Chr1, 2, 4, 5, 6, 7, 9, 10 and 11 of *C. decandra* exist as independent chromosomes in at least three genera, corresponding to ASK1, 2, 4, 5, 6, 7, 9, 10 and 11, respectively. The shaded areas show the fusion of ASK3 and ASK8.



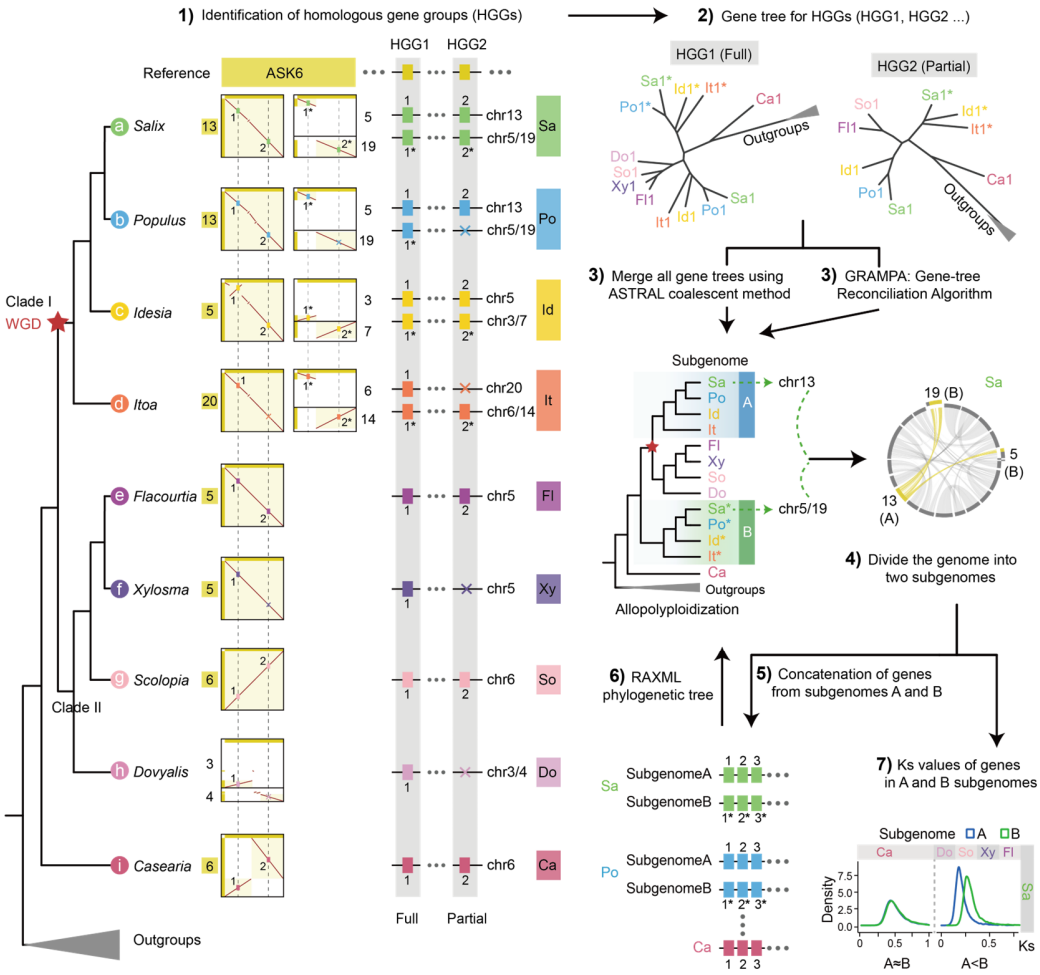
Supplementary Fig. 9. Schematic diagram of protochromosome identification, construction and extension. a, Taking ASK6 as an example, the synteny between Casearia as a reference and the remaining species is shown, with the numbering of independent chromosomes highlighted in yellow. **b,** One independent chromosome (Chr5 of *Flacourtia*) was selected as the initial protochromosome and aligned with the genomes of the other Clade II species and *Casearia* one by one (steps 1-4). Any redundant genes between two collinear genes, not exceeding five in number, were added to the protochromosome. The gray shading represents temporary protochromosomes generated during intermediate steps, ultimately resulting in the extended protochromosome.



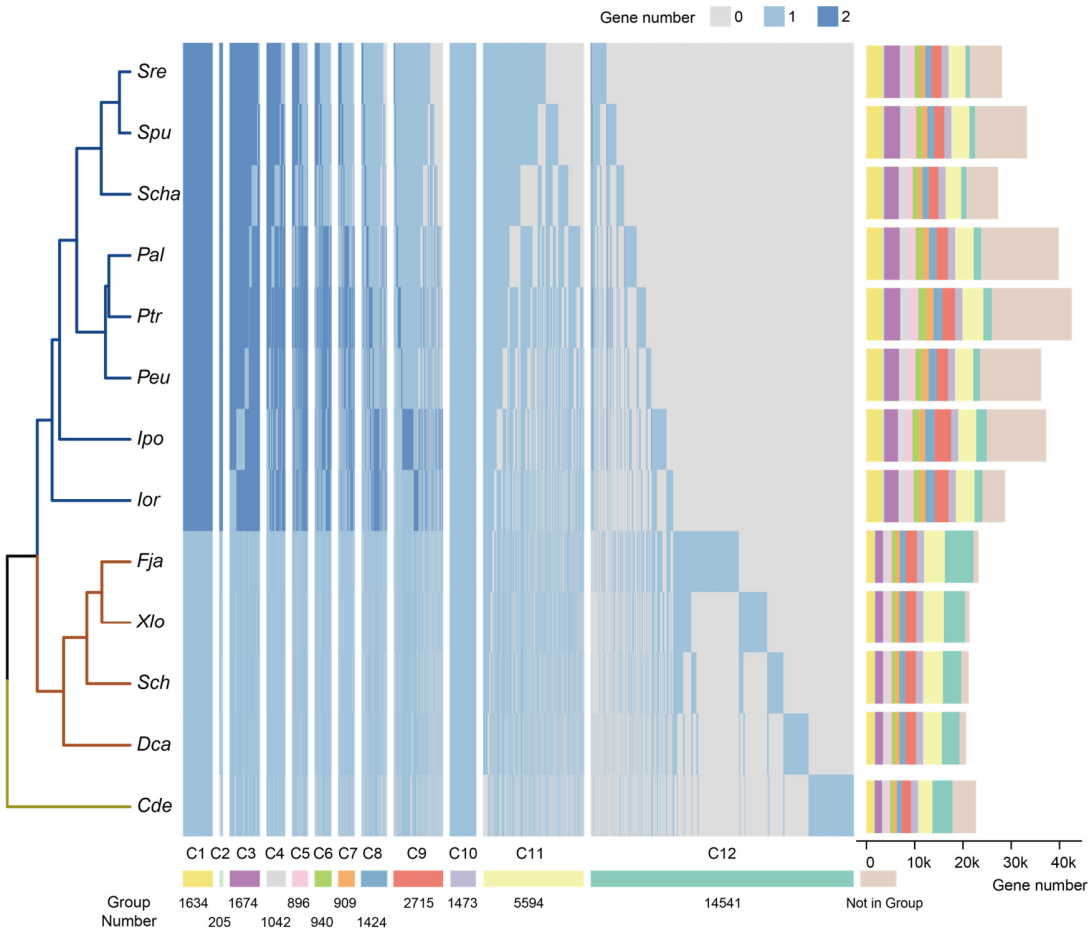
Supplementary Fig. 10. Synteny plot between the inferred 11 ASK protochromosomes and the genomes of different genera within Salicaceae. Different colors represent 11 protochromosomes.



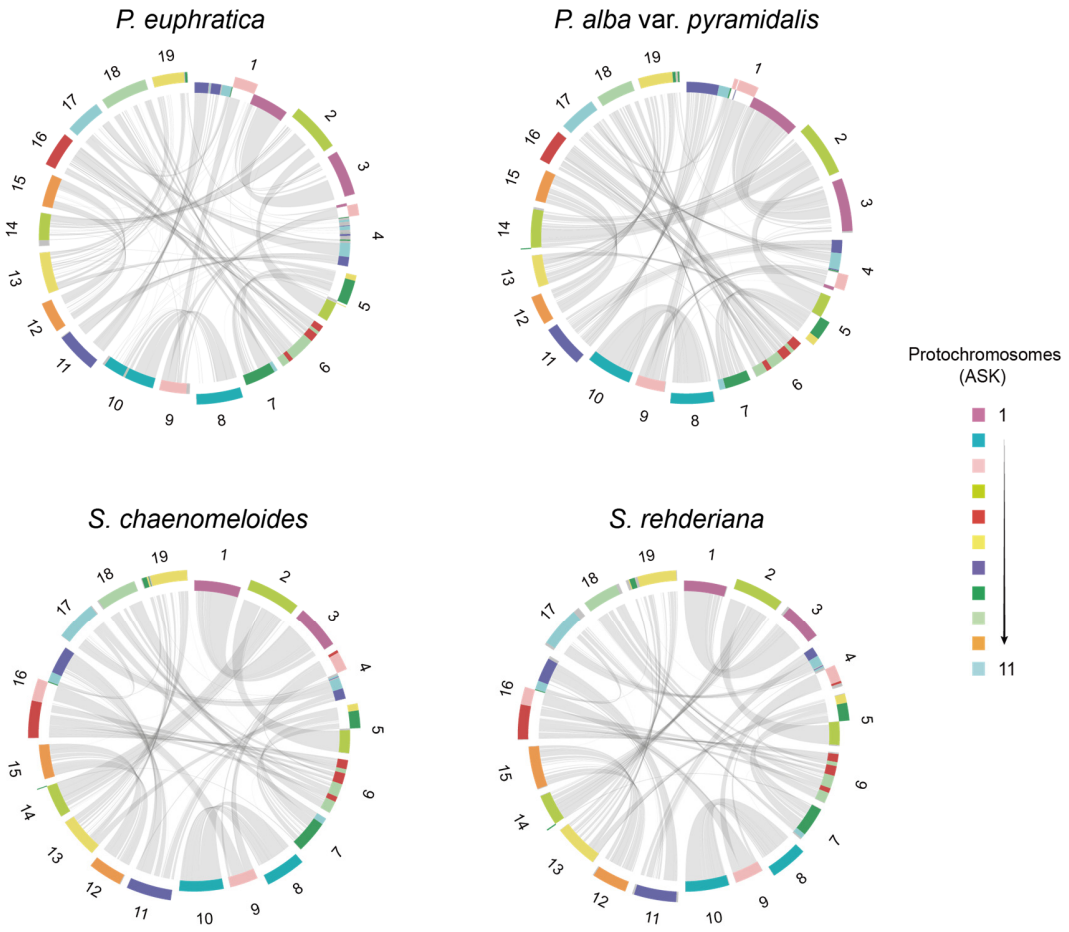
Supplementary Fig. 11. Local synteny dot plots between ASK protochromosomes and extant genomes associated with the chromosomal fusion events at the four nodes in Fig. 2. a, RTA event between protochromosome 8 and 11 in Clade II. **b,** RTA event between protochromosome 2 and 4 in Clade II. **c,** EEJ event between protochromosome 5 and 9 in Clade I. **d,** NCF events involving protochromosomes 3, 7, and 11 in Clade I. The first dotplot in each event illustrates the pattern of change in detail. The dashed lines indicate fusion breakpoints (a, b, d) or inversion breakpoints (c). Species without fusion events are shaded in gray.



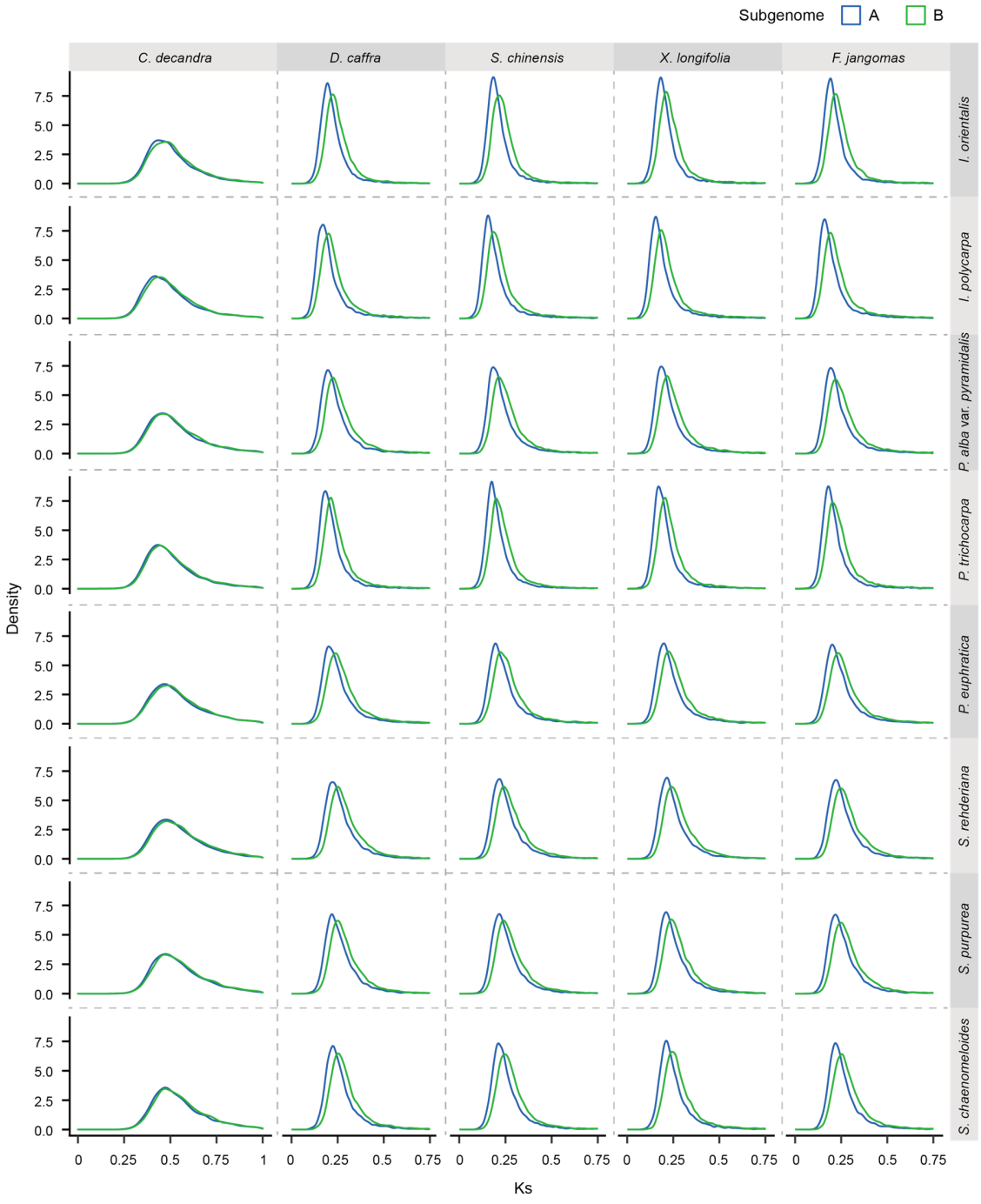
Supplementary Fig. 12. Workflow of allopolyploidization and subgenome identification. Taking protochromosome 6 (ASK6) as an example: 1) Based on gene collinearity between 13 genomes and ASK6, 1:2 HGGs were identified between Clade II / *Casearia* and Clade I, including Full- and Partial-HGGs. 2) Gene trees were constructed for each HGG using RAxML. 3) Gene trees of each protochromosome were merged using ASTRAL and the polyploidization type was inferred based on the proportions of different topologies at key nodes. GRAMPA was also used to infer the polyploidization pattern, showing consistent topology with ASTRAL tree. Genes closer to Clade II were classified as the A subgenome, and the other paralogs belonged to the B subgenome. 4) Chromosomes of Clade I species were divided into A and B subgenomes based on gene positions. 5) Genes in different subgenomes were concatenated, and 6) a tree was constructed using RAxML, which also obtained a consistent topology. 7) *Ks* comparisons between A / B subgenomes and other species, further validated allopolyploidization.



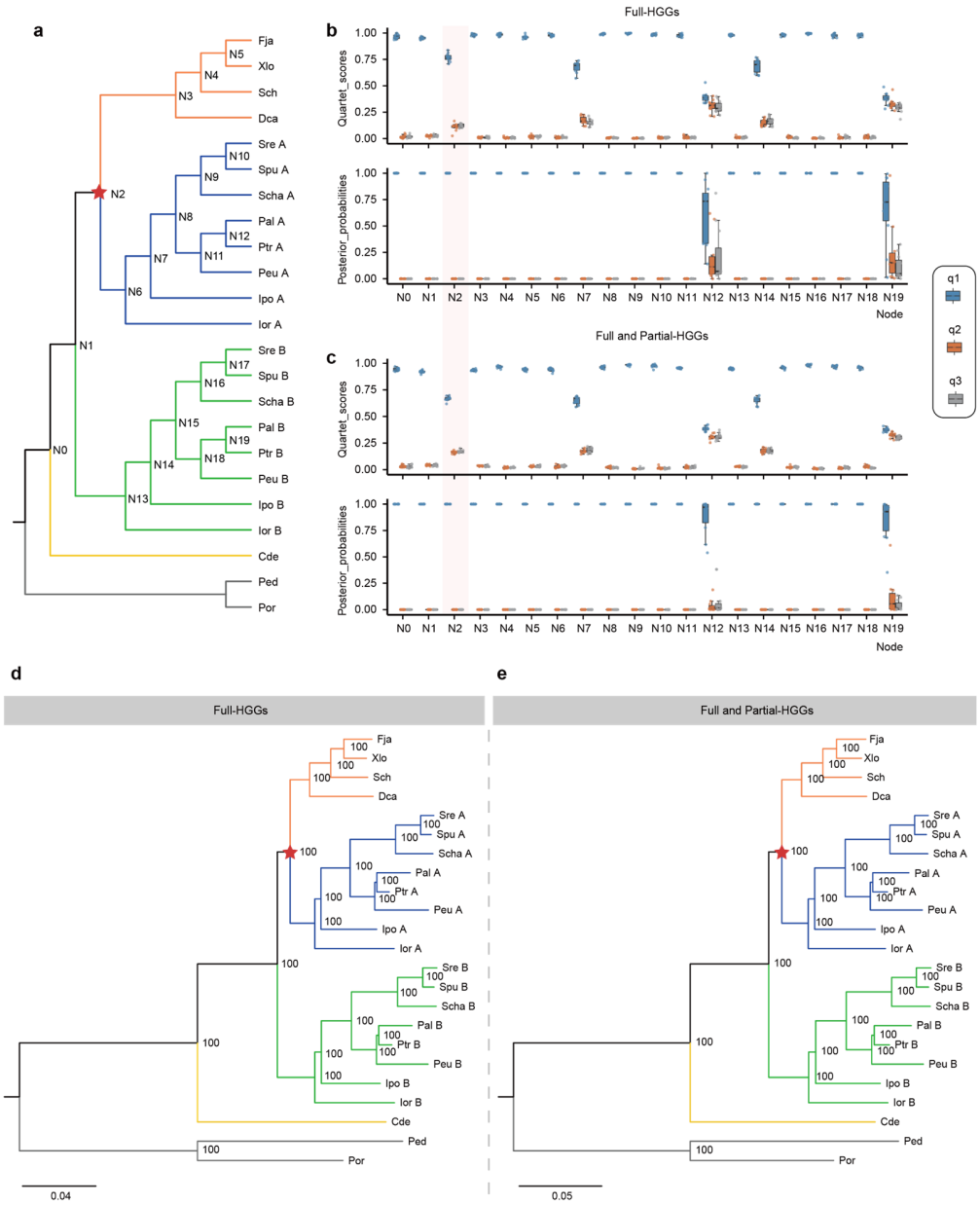
Supplementary Fig. 13. Homologous gene groups (HGGs) identified in the studied species. On the left is the phylogenetic tree of the 13 species studied. The middle heatmap shows an overview of the number of genes per species in each HGGs. C1 (Full-HGGs): All Clade I species have two gene copies, and the rest of the branch species have one copy. C2-C9 (Partial-HGGs): 8-1 species in Clade I have two gene copies, and at least two species of the remaining branches have one gene copy. C10: All species have one copy. C11: At least 3 species in Clade I have one copy, and the rest of the branches have at least two species with one copy. C12 represents other scenarios. Genes not in the group are those without homologous relationships with the reconstructed protochromosomes. On the right is the number of genes per HGGs in each species.



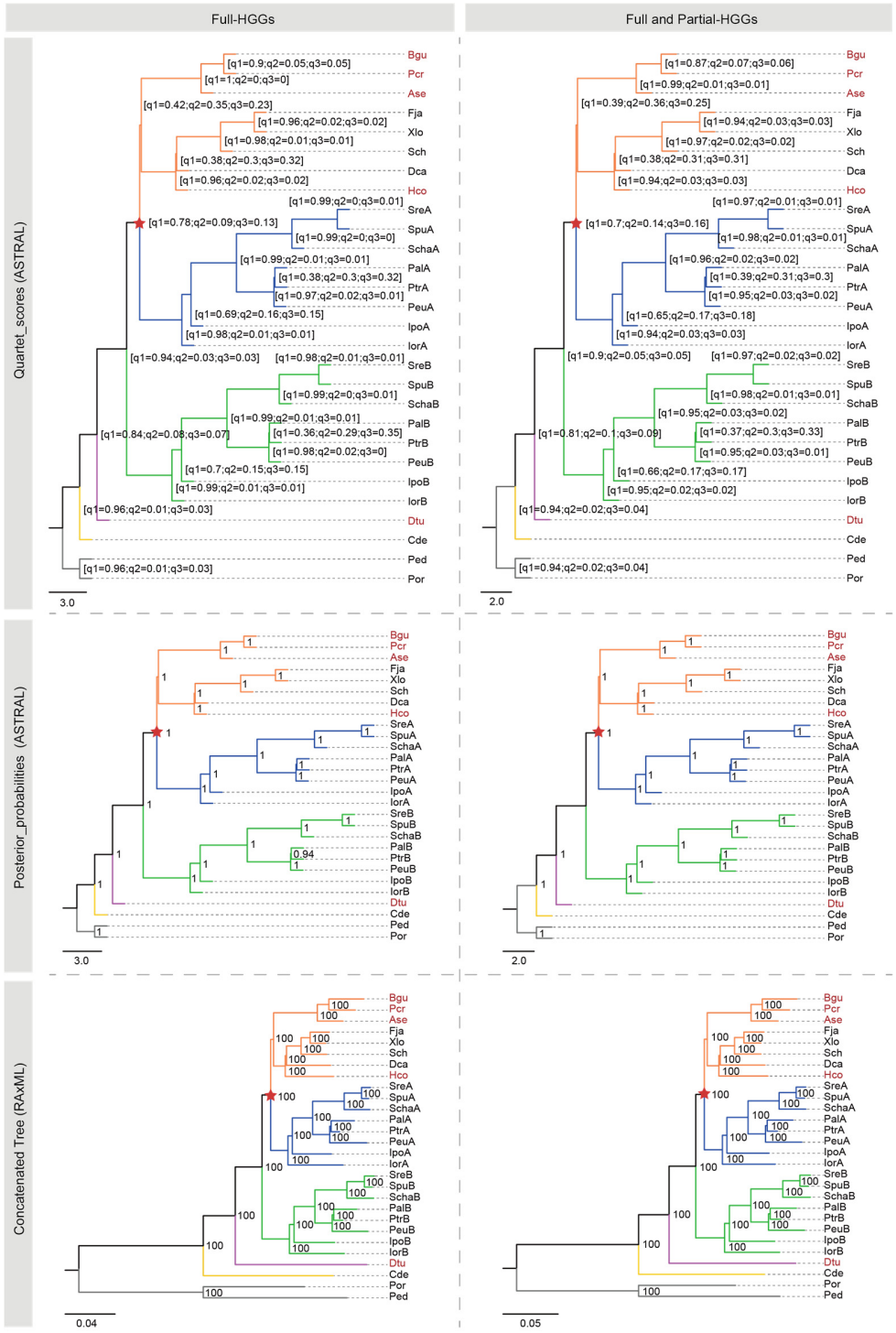
Supplementary Fig. 14. The collinearity between the two subgenomes of *P. euphratica*, *P. alba var. pyramidalis*, *S. chaenomeloides* and *S. rehderiana*. The inner circle is the A subgenome and the outer circle is the B subgenome. The color of each chromosome indicates its ancestral origin.



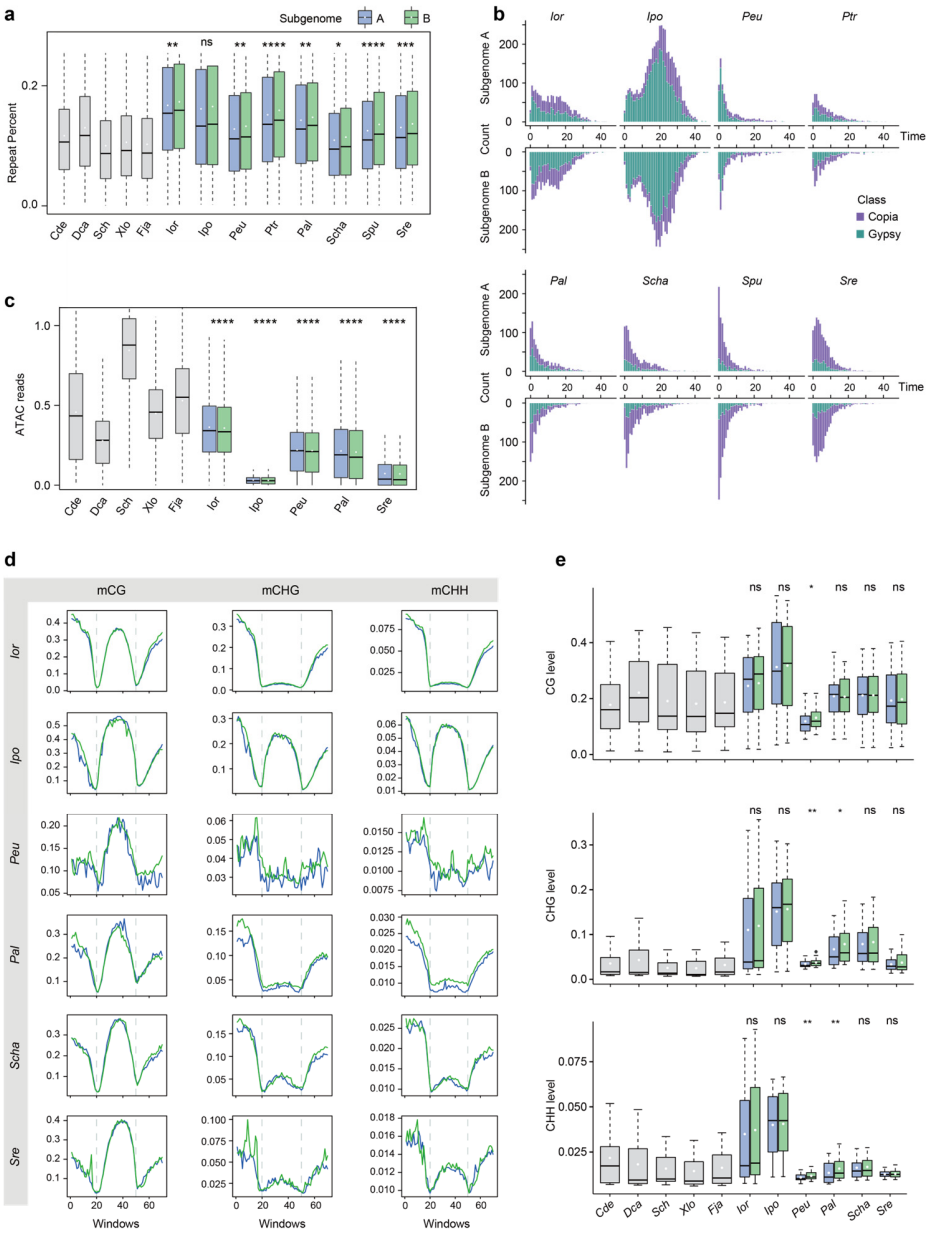
Supplementary Fig. 15. The *Ks* distribution between the A/B subgenomes of the eight species in Clade I and 4 species in Clade II, as well as *C. decandra*. The A and B subgenomes are represented in blue and green, respectively.



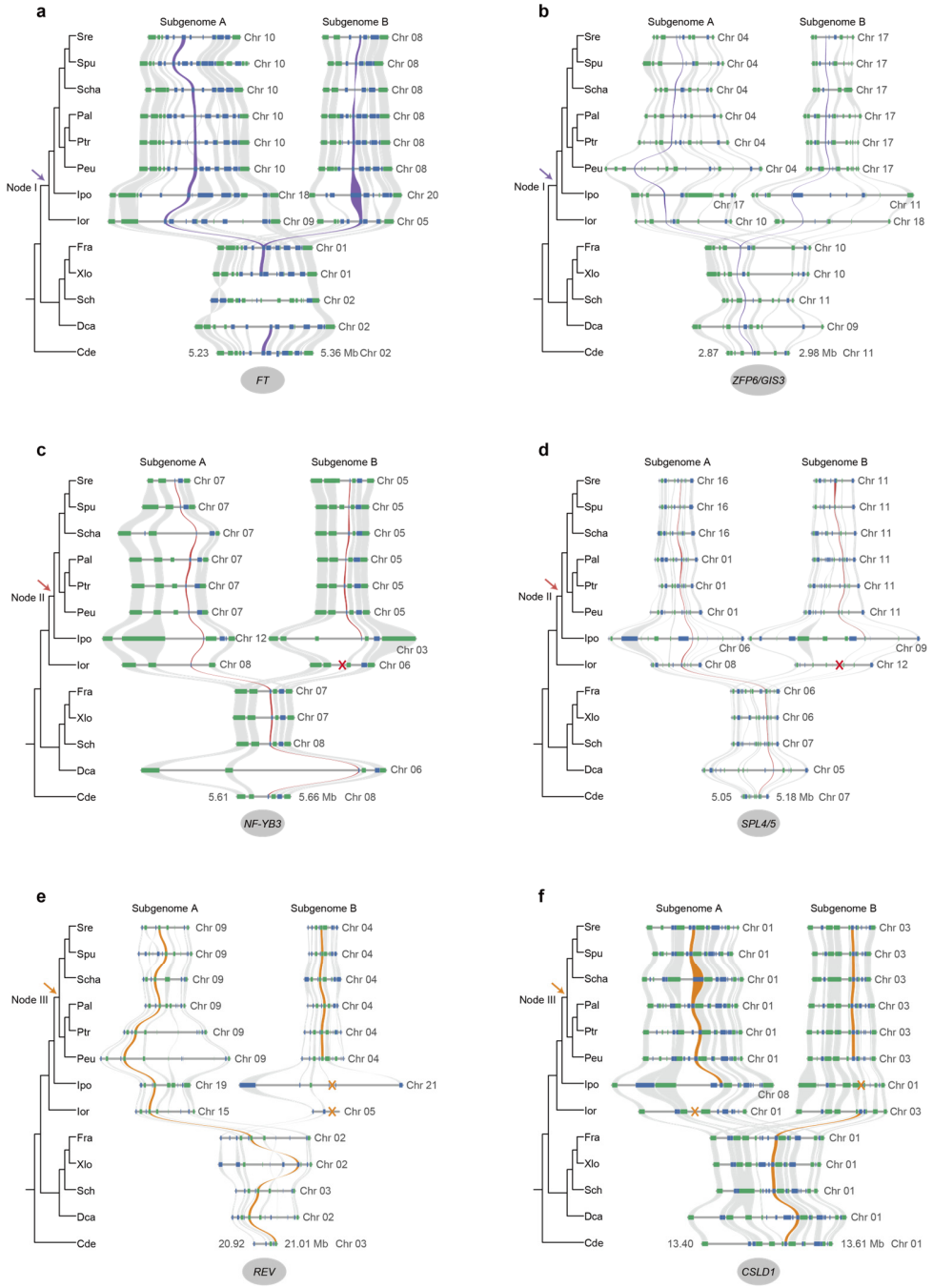
Supplementary Fig. 16. The support of phylogenetic trees constructed by coalescent and concatenated methods. a-c, The frequency and support for three topologies at each node of Astral trees constructed with 11 protochromosomes as references. **a,** Each node label in the evolutionary tree. The quartet scores and posterior probabilities for the three topologies at each node of the astral tree inferred using Full-HGGs **(b)** and Full-/Partial-HGGs **(c)** datasets. **d-e,** Phylogenetic tree inferred by concatenated method (RAxML) using Full-HGGs **(d)** and Full-/Partial-HGGs **(e)** datasets. It shares the same topology as the Astral tree, with a bootstrap value of 100 at each node.



Supplementary Fig. 17. Coalescent (Astral) and concatenated (RAxML) trees inferred using the Full-HGGs and Full-/Partial-HGGs dataset after adding 5 species. Bgu: *Banara guianensis*, Pcr: *Prockia crucis*, Ase: *Azara serrata*, Hco: *Homalium cochinchinense*, Dtu: *Dianyuea turbinata.*



Supplementary Fig. 18. The repetitive sequences and epigenetic features of the two subgenomes of Clade I species, along with the remaining species. This includes the content of repetitive sequences **(a)** and methylation levels **(d-e)** in genes and surrounding regions, as well as chromatin accessibility **(c)** in the upstream regions of genes. The A, B subgenomes and the remaining species are represented in blue, green, and gray, respectively. **b,** The distribution of insertion times for two types of full-length LTRs in the A and B subgenomes of Clade I species. The x and y axes represent insertion time and Count respectively. Abbreviations for all species correspond to Fig. 4b.



Supplementary Fig. 19. Microsynteny visualization of example duplicated gene pairs specifically retained in four genera (Node I), three genera (Node II), and two genera (Node III). **a-b,** two copies were retained in *Itoa*, *Idesia*, *Populus*, and *Salix*. **c-d,** Only one copy of the duplicated gene was retained in *Itoa*, while two copies were retained in *Idesia*, *Populus*, and *Salix*. **e-f,** Only one copy of the duplicated gene was retained in *Itoa* and *Idesia*, while two copies were retained in *Populus* and *Salix*.