Supplementary methods

Whole genome sequencing

For reference genome sequencing, we collected fresh leaves from an adult plant of tetraploid *Trapa natans* (2n = 4x = 96) grown in Hangzhou Botanical Garden, Zhejiang, China. Genomic DNA was extracted from fresh leaves using DNAsecure Plant Kit (Tiangen Biotech, Beijing, China) according to the manufacturer's protocol. We sequenced and assembled the reference genome using a hybrid approach that combined Pacific Biosciences SMRT sequencing (PacBio), 10X Genomics linked read sequencing (10X Genomics), Illumina sequencing and a Hi-C chromatin interaction map. A 20-kb insert size library was constructed with a SMRTbell Template Prep Kit (PacBio) and subsequently sequenced with a total of 12 SMRT cells on the PacBio Sequel platform. A total of ~58.66 Gb of PacBio sequence data with an N50 read length of 17,389 bp were generated (Table S1). Meanwhile, DNA fragments longer than 50 kb were used to construct a 10X Genomics Chromium Genome linked read library with a Chromium controller instrument and reagents (10X Genomics), followed by sequencing on an Illumina HiSeq X Ten platform (Illumina, NEB, USA) with paired-end reads 150 bp in length. A total of ~72.48 Gb of 10X linked read data were generated (Table S1). For Illumina sequencing, a short-insert (350 bp) genomic library was performed using NEBNext Ultra DNA Library Prep Kit for Illumina, and sequenced on the Illumina HiSeq X Ten platform. Approximately 63.11 Gb of short read Illumina data were generated (Table S1). Besides, a Hi-C library was created following the procedures described previously (Shaw, 2010), and sequenced on an Illumina HiSeq X Ten platform with paired-end, 150 bp reads, ultimately producing 58.12 Gb Hi-C data (Table S1). Finally, total RNA was extracted from five tissues (including roots, stems, leaves, flowers and fruits) using the RNAprep Pure Plant Kit (Tiangen Biotech, Beijing, China). RNA-Seq libraries were then constructed using the TruSeq RNA Library Kit (Illumina, CA, USA), followed by sequencing on the Illumina HiSeq X Ten platform, and generating 150-bp paired-end reads. These RNA-Seq data were used for gene structure prediction.

Phylogenetic analysis and divergence time estimation of the two sub-genomes

Orthogroups were retrieved from the two sub-genomes and 11 other species including *A. thaliana*, *Citrus sinensis*, *Cucumis melo*, *Eucalyptus grandis*, *Glycine max*, *Juglans regia*, *Populus trichocarpa*, *Prunus persica*, *P. granatum*, *Theobroma cacao* and *Vitis vinifera* using the program ORTHOMCL (Li *et al*., 2003), with *Oryza sativa* used as the outgroup species. The single-copy orthologs were aligned using MUSCLE v.3.8.3 (Edgar, 2004) and then concatenated into a super-gene alignment matrix. The alignment was trimmed with GBLOCKS (Castresana, 2000). Maximum likelihood (ML) analyses were conducted using RAXML-HPC v.8.2.8 (Stamatakis, 2014) with 1,000 bootstrap replicates. The best model and parameter settings were chosen according to the Akaike Information Criterion (AIC) using JMODELTEST v.2.1.4 (Posada and Buckley, 2004).

Divergence times were estimated by the program MCMCTREE in PAML v.4.7a (Yang, 2007), with the parameters: burn-in = 10,000, sample number = 100,000, sample frequency = 2. The following six divergence times were obtained from the TIMETREE database (http://www.timetree.org/) and used for calibrations (in million years ago, Ma): i) *O. sativa* and *V. vinifera* (115–308 Ma), ii) *V. vinifera* and *P. trichocarpa* (107–135 Ma), iii) *E. grandis* and *P. persica* (98–117 Ma), iv) *A. thaliana* and *P. persica* (98–117 Ma), v) *T. cacao and P. trichocarpa* (98–117 Ma) and vi) *G. max* and *C. melo* (95–135 Ma).

References

Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution* **17**: 540–552.

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**: 1792–1797.

Li L, Stoeckert CJ, Roos DS (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Research* **13**: 2178–2189.

Posada D, Buckley TR (2004) Model selection and model averaging in phylogenetics: advantages of akaike information criterion and bayesian approaches over likelihood ratio tests. *Systematic Biology* **53**: 793–808.

Shaw PJ (2010) Mapping chromatin conformation. *F1000 Biology Reports* **2**: 18.

Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**: 1312–1313.

Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* **24**: 1586–1591.