**Supporting experimental procedures**

**Methods S1. Genome survey**

The 111.36 Gb short reads from the MGI-SEQ 2000 platform were used for genome survey analysis. The data were first cleaned using the HTQC (v1.92.310) (Yang et al., 2013). The GCE software (v1.0.0) (Liu et al., 2013) was used to count *k*-mers (*k* = 17) in the clean data, and in total 517401446 *k*-mer were obtained. *k*-mer depth-frequency distribution analysis (Figure S1) indicated two peaks with depths of 61 and 124, and the former was much greater than was the latter; thus, the peak with a depth of 61 was considered to be the main peak and the peak with a depth of 124 was considered to be resulted from whole-genome duplication; nearly no heterozygosity was found. The genome size was estimated to be 1.092 Gb, whose 71.43% is repetitive elements.

**Methods S2. Genome assembly**

The 133.33 Gb subreads from the Pacbio Sequal II platform, which is about 122 × *G. elata* genome size, were corrected, trimmed, and assembled using CANU (v1.9) (Koren et al., 2017) (genomeSize = 1.1g). Next, to correct the errors in the initial assembly, subreads were mapped to the initial assembly using pbmm2 (v1.2.1) (https://github.com/PacificBiosciences/pbmm2), and further polished using GCpp (v1.9.0) (https://github.com/PacificBiosciences/gcpp); three rounds of mapping and polishing were conducted. Thereafter, we used BWA (v0.7.17-r1188) (Li, 2013) to map the NGS short reads to the polished assembly, and the assembly was corrected by Pilon (v1.23) (Walker et al., 2014); this was repeated three times as well. Finally, a highly accurate assembly containing 457 contigs were obtained, whose N50 is 21.33 Mbp and the total length is 1.0428 Gb (Table S1). NGS short reads were mapped to the contigs using HISAT2 (Kim et al., 2019), and the results from mapping were used for calling SNPs and indels using SAMtools (v1.9) (Li et al., 2009) and BCFtools (v1.9) (Danecek et al., 2021); it was found that the final assembly has an error rate of 0.00094% and heterozygosity of 0.00388% (Table S1).

The original Hi-C data were cleaned using HTQC (v1.92.310) (Yang et al., 2013) and processed by the Juicer pipeline (v1.6) (Dudchenko et al., 2018); the data were further mapped to the contigs with BWA (v0.7.17-r1188) (Li, 2013), and the mapping data were used for inferring the Hi-C contact matrices using the Juicer (v1.6) (Dudchenko et al., 2018). Based on these matrices, the 3D-DNA pipeline (v180922) (Dudchenko et al., 2017) was employed to correct the contigs: when the Hi-C data disagreed with contigs, the target contigs were split into contig fragments. *G. elata* organelle-originated contig fragments were identified by BLAST (v2.4.0) (Altschul et al., 1990) search against the published *G. elata* organelle genomes (Yuan et al., 2018) and these contig fragments were removed from the nuclear genome assembly. GC content and mapping depths of NGS reads were calculated and contigs with extreme GC contents and very low read depths were considered to be contaminants and removed (Figure S2). Next, using the contact information inferred among contig fragments, scaffold data were finally obtained by processing with 3D-DNA (v180922) (Dudchenko et al., 2017). Hi-C contact matrices were displayed as heatmaps using JuiceBox (v1.6) (Dudchenko et al., 2018) to manually inspect the orientation and order of the contig fragments in the pseudo-chromosomes, and the errors were corrected manually. Finally, 18 *G. elata* pseudo-chromosomes were obtained and they were named Chr1 to Chr18, whose sizes are in a descending order (Figure 1). There were 9.45 Mb contig fragments which were not anchored to any pseudo-chromosomes, and they were randomly fused together, and this sequence was named the Chr0, which was only used for annotation but not for collinearity analysis.

**Methods S3. Annotation of genome features**

**Repetitive elements.** Multiple bioinformatic pipelines were used to annotate repetitive elements. For ribosomal RNAs, we used the Arabidopsis 5S rRNA (NCBI NR\_139969.1) and 45S rRNA (NCBI X52322.1) as queries to search for 5S and 45S rRNAs in *G. elata*; Snoscan (v1.0) (Lowe and Eddy, 1999) was used to identify snoRNAs (default parameters); MITEs (miniature inverted-repeat transposable elements) were search using the MITE-Hunter (Han and Wessler, 2010) with default parameters; for identification of centromeric repetitive sequences we used Tandem Repeats Finder (v4.09) (Benson, 1999) (parameter: 1 1 2 80 5 200 2000 -d -h) to identified all tandem repeats, which were further filtered to obtain tandem repeats whose total length was the largest with a ~ 200-bp length per unit; for retrotransposons and DNA transposons, the EDTA pipeline (v1.8.3) (Ou et al., 2019) was employed, which integrates the LTR\_FINDER (Xu and Wang, 2007), LTRharvest (Ellinghaus et al., 2008), LTR\_retriever (Ou and Jiang, 2018), Generic Repeat Finder (Shi and Liang, 2019), TIR-learner (Su et al., 2019), HelitronScanner (Xiong et al., 2014), and TEsorter (Zhang et al., 2019) and performs ab inito annotation of genome-wide transposon elements. After all the repetitive elements were obtained, they were combined with the Dfam database (https://dfam.org/), which were fed to RepeatMasker (Smit et al.), outputting the annotation of the complete repetitive elements (Table S3).

**tRNA annotation.** We used tRNAscan (v2.0.5) (Lowe and Eddy, 1997) to annotate tRNAs following the default parameters.

**Annotation of protein-coding genes.** To annotate the protein-coding genes in *G. elata*, four steps were used to guide the annotation. 1) Augustus (v3.4.0) (Keller et al., 2011) was trained with the MAKER\_P (Campbell et al., 2014) pipeline using *G. elata* RNA-seq data (Table S2) and the protein-coding genes in the genomes of *Arabidopsis thaliana*, *Oryza sativa*, *Asparagus officinalis*, *Apostasia shenzhenica*, *Dendrobium catenatum*, and *Phalaenopsis equestris* and those in the previously published *G. elata* genome (Yuan et al., 2018) . After three rounds of training of Augustus, *G. elata*’ de novo annotation of protein-coding genes was obtained. 2) The RNA-seq data were mapped to *G. elata* unmasked genome using TopHat2 (v2.1.1) (Trapnell et al., 2012), and the mapped reads were assembled to gene structures by Cufflinks (v2.2.1) (Trapnell et al., 2012); TRINITY (v2.8.5) (Grabherr et al., 2011) was used to de novo assemble the RNA-seq reads and thereafter PASA2 (Haas et al., 2008) was used to map the assembled contigs to the *G. elata* genome and annotate the gene structures. These annotation results were merged. 3) The protein-coding genes in the closely related species and the previously published *G. elata* genome were aligned to the unmasked *G. elata* genome using GenBlastA (v1.0.1) (She et al., 2009), and GeneWise2 (v2.4.1) (Birney et al., 2004) was employed to generate gene structures. 4) The gene structures obtained from the above three steps were combined by EVidenceModeler (v1.1.1) (Haas et al., 2008) to create the final version of annotation, in which the transposons were not filtered out, aiming to minimize false positives in gene loss analysis. In this manner, 21115 protein-coding genes were identified in the *G. elata* genome (Table S3). To evaluate the completeness of annotation, BUSCO4 (v4.1.0) (Simão et al., 2015) analysis was performed using the embryophyte dataset. Under the genome mode, the completeness of conserved genes was 70.80% and the missing ratio was 25.50%; under the protein mode, BUSCO4 analysis on the annotated proteins indicated that the conserved genes’ completeness was 76.40% and the missing ratio was 21.60% (Table S4).

**Methods S4. Gene family expansion and contraction**

In order to analyze the gene families in Orchidaceae (Asparagales), we selected at least one species from each order of monocotyledonous plants, which have been sequenced with good quality. These species are *G. elata*, *Phalaenopsis equestris*, *Dendrobium catenatum*, *Vanilla planifolia*, *Apostasia shenzhenica*, and *Asparagus officinalis* (Asparagales), *Ananas comosus*, *Oryza sativa*, and *Zea mays* (Poales), *Cocos nucifera* (Arecales), *Xerophyta viscosa* (Pandanales), and *Dioscorea alata* (Dioscoreales). Three ancient species *Cinnamomum kanehirae* (Laurales), *Liriodendron chinense* (Magnoliales), and *Amborella trichopoda* (Amborellales) were also included as the outgroups. The protein sequences from all species were firstly all-vs-all aligned using DIAMOND (v0.9.24.125) (Buchfink et al., 2021). The E-value of each hit was transformed into a W index ranging from 0 to 100 using the following formula.



Next, the W indices were passed to hcluster (v0.5.1) (Vilella et al., 2009) as the weight matrices to build gene families with the parameters “-m 170 -w 0 -s 0.34 -O”. A total of 30274 gene families covering 92.8% of the total genes were identified. For each gene family, the numbers of genes in each species and the topology of species tree were fed to a home-made pipeline based on BadiRate (v1.35) (Librado et al., 2012) to infer whether there are statistically significant events of expansion or contraction at each node of the species tree. This pipeline used the gain-and-death (GD) model in BadiRate, a program that estimates gene family turnover rates based on likelihood. Firstly, we used the free model in BadiRate to estimate the sizes of the ancestral gene families in all nodes of the species tree. For the branches whose gene families did not experience family size changes, they were set to be the background branches, which have the same family turnover rate, and a null hypothesis model was built based on these settings and the likelihood was estimated. For branches that experienced gene family size changes, an alternative hypothesis model for each branch was built by forcing the given branch to follow the turnover rate of the background branches. Size changes were considered to be significant, if AIC (alternative hypothesis) - AIC (null hypothesis) > 2, in which AIC (Akaike’s information criterion) (Burnham and Anderson, 2002) was computed from the likelihood and numbers of parameters in each model. Finally, the statistically supported expansion/contraction of all gene families in all nodes of the species tree were obtained (Table S6).

To directly compare sizes of gene families between species, we computed the *F*-indices (Sun et al., 2018), which describe the size differences of conserved gene families in all autotrophs (abovementioned plants except orchids). *F*-indices range from 0 to 1, and when *Fij* = 0.5, in species *i*, the gene number in gene family *j* equals to the average size of this gene family in all species; if *Fij* = 0, there are no genes in family *j* in species *i*, and if *Fij* = 1, it indicates that only species *i*, but not the others, harbors the gene family *j*. Thereafter, a box plot of *F*-indices of the families (data are listed in Table S6) that are conserved in autotrophs was constructed (Figure 2).

**Methods S5. Identification of orthologous introns**

Orthologs in five orchid species, *A. shenzhenica*, *V. planifolia*, *P. equestris*, *D. catenatum*, and *G. elata*, and nine closely related monocots (Table S7) were identified using OrthoFinder (v2.3.11) (Emms and Kelly, 2019). The protein sequences of orthologs in each orthogroup were aligned using CLUSTAL\_W2 (v2.1) (Thompson et al., 1994). The matrixes containing alignments and gene structural information were further given to GenePainter (v2.0) (Muhlhausen et al., 2015) to obtain the relative position information of introns. Introns which are found at least in two orchids and five out of the nine monocots were considered to be orthologous introns, and they were further used for comparison of lengths.

**Methods S6. Comparison of gene loss in Orchids and *C. australis*, *Striga asiatica*, and *Sapria himalayana***

Three orchids *Apostasia shenzhenica*, *Vanilla planifolia*, and *G. elata*, which are initial and full mycoheterotrophic, and parasitic plants *Striga asiatica*, *C. australis*, and *Sapria himalayana* and 38 autotrophic angiosperm species (Table S5) were included in the analysis to obtain gene families, and the conserved gene families were further compared to reveal gene loss patterns in these mycoheterotrophic and parasitic plants.

First, we clustered protein sequences of all gene models in every species using OrthoFinder (v2.3.11) (Emms and Kelly, 2019), and thus, 50486 orthogroups were constructed. To further screen for well conserved orthogroups (WCOs) in both monocots and eudicots, the orthogroups fulfilling the following criteria were defined as WCOs: a WCO should be obtained from at least three species in each of the following groups, autotrophic monocots, rosids, and asterids from the abovementioned 38 autotrophic angiosperms. We obtained 10196 WCOs after the screening. All WCOs were inspected, and if any of the mycoheterotrophic or parasitic plants lost the ortholog, while the ortholog exists in the most closely related autotrophic species, this specific ortholog is considered to be lost in this mycoheterotrophic or parasitic plant. To minimize the errors of gene loss identification caused by annotation mistakes, all the sequences from the other species were assigned as queries to align with the intergenic regions in the genome(s) of species whose corresponding ortholog(s) is(are) lost using the genBlastA (v1.0.1) (She et al., 2009); if genBlastA analysis resulted in positive hits in the intergenic regions, GeneWise2 (v2.4.1) (Birney et al., 2004) was next used to annotate the gene structures of these regions. The validity of the retrieved genes was examined by phylogenetic analysis, and the valid ones were incorporated back to the respective WCOs.

Finally, in 10196 WCOs, we found that *Apostasia shenzhenica*, *Vanilla planifolia*, *G. elata*, *Striga asiatica*, *C. australis*, and *Sapria himalayana* lost 463 (4.52%), 528 (5.16%), 1211 (11.83%), 402 (3.93%), 1012 (9.89%), and 3604 (35.21%) orthogroups, respectively (Table S8).

Venn diagram analysis was done to compare the lost genes in mycoheterotrophic and parasitic plants (Figure S5). WCOs in different sets of the Venn diagrams were set to be the foreground, and all the 10196 WCOs were sets to be the background (the function of a given WCO is defined by the function of all the genes in the WCO), and topGO (v 3.12) (Alexa and Rahnenfuhrer, 2020) was used to perform function enrichment analysis (classic Fisher’s Exact Test p-value <= 0.05).

**Methods S7. Identification of Arabidopsis orthologs in *G. elata*, *C. australis*, and their reference plants**

For comparing the genes loss in *G. elata* and *C. australis* on the level of individual genes, we carefully identified the Arabidopsis orthologs in *G. elata*, *C. australis*, and their close relatives. To decrease the complication resulted from complex tree topologies, we only used a minimal number of required species, which are the monocots *Xerophyta viscosa* (Xvi), *Asparagus officinalis* (Aof), *Apostasia shenzhenica* (Ash), *G. elata* (Gel), *Vanilla planifolia* (Vpl), *Cocos nucifera* (Cnu), *Ananas comosus* (Acom), *Ensete ventricosum* (Eve), and *Oryza sativa* (Osa) and the eudicots *Aquilegia coerulea* (Aco), *Arabidopsis thaliana* (Ath), *Coffea canephora* (Cca), *Mimulus guttatus* (Mgu), *Solanum lycopersicum* (Sly), *C. australis* (Cau), and *Ipomoea nil* (Ini) and two outgroup species *Piper nigrum* (Pni) and *Amborella trichopoda* (Atr). All protein sequences from these 18 species were clustered with OrthoFinder (v2.3.11) (Emms and Kelly, 2019), resulting in 33099 orthogroups. Within each orthogroup, all sequences were subjected to multiple sequences alignment using CLUSTAL\_W2 (v2.1) (Thompson et al., 1994), and the resulting matrix was used for constructing the maximum likelihood tree with FastTree (v2.1) (Price et al., 2010); the tree was further converted to a rooted tree based on the gene tree rooting algorithm which we adopted from OrthoFinder (v2.3.11) (Emms and Kelly, 2019). All the rooted gene trees were examined to collapse the clades with less than 70% bootstrap values. Two terms were used in the next analysis: Minimal eudicot conserved clades, each of which must contain Ath and three from Cca, Mgu, Sly, Ini, or Aco, and minimal monocot/eudicot conserved clades, each of which must contain Ath and Osa, and three from Cca, Mgu, Sly, Ini, or Aco, and three from Xvi, Aof, Cnu, Acom, or Eve. On each gene tree we search for the minimal eudicot conserved clades and the minimal monocot/eudicot conserved clades. Within a minimal eudicot conserved clade, if  and Arabidopsis coexist, then these *C. australis* and Arabidopsis gene(s) are orthologs. Similarly, within a minimal monocot/eudicot conserved clade, if *G. elata* and Arabidopsis coexist, these *G. elata* and Arabidopsis gene(s) are orthologs. Arabidopsis genes that do not fall into minimal eudicot conserved clades or monocot/eudicot conserved clades, they are considered to be non-conserved genes. Finally, among the 27620 Arabidopsis genes, 1) 19358 were found to be conserved in both monocots and eudicots, and among them, 16117 have orthologs in *G. elata*; 2) 20859 are conserved in eudicots, and among which, 16590 have orthologs in *C. australis* (data are shown in Table S9).

**References**

**Alexa, A. and Rahnenfuhrer, J.** (2020) topGO: Enrichment Analysis for Gene Ontology. *R package version 2.42.0.*

**Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J.** (1990) Basic local alignment search tool. *Journal of Molecular Biology*, **215**, 403-410. <https://doi.org/10.1016/S0022-2836(05)80360-2>

**Benson, G.** (1999) Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research*, **27**, 573-580. <https://doi.org/10.1093/nar/27.2.573>

**Birney, E., Clamp, M. and Durbin, R.** (2004) GeneWise and Genomewise. *Genome Research*, **14**, 988-995. <https://doi.org/10.1101/gr.1865504>

**Buchfink, B., Reuter, K. and Drost, H.G.** (2021) Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nature Methods*, **18**, 366-368. <https://doi.org/10.1038/s41592-021-01101-x>

**Burnham, K.P. and Anderson, D.R.** (2002) *Model selection and multi-model inference: A practical information-theoretic approach*. New York: Springer.

**Campbell, M.S., Law, M., Holt, C., Stein, J.C., Moghe, G.D., Hufnagel, D.E., et al.** (2014) MAKER-P: A tool kit for the rapid creation, management, and quality control of plant genome annotations. *Plant Physiology*, **164**, 513-524. <https://doi.org/10.1104/pp.113.230144>

**Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., et al.** (2021) Twelve years of SAMtools and BCFtools. *GigaScience*, **10**. <https://doi.org/10.1093/gigascience/giab008>

**Dudchenko, O., Batra, S.S., Omer, A.D., Nyquist, S.K., Hoeger, M., Durand, N.C., et al.** (2017) De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science*, **356**, 92-95. <https://doi.org/10.1126/science.aal3327>

**Dudchenko, O., Shamim, M.S., Batra, S., Durand, N.C., Musial, N.T., Mostofa, R., et al.** (2018) The Juicebox Assembly Tools module facilitates de novo assembly of mammalian genomes with chromosome-length scaffolds for under $1000. <https://doi.org/http://dx.doi.org/10.1101/254797>

**Ellinghaus, D., Kurtz, S. and Willhoeft, U.** (2008) LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinformatics*, **9**, 18. <https://doi.org/10.1186/1471-2105-9-18>

**Emms, D.M. and Kelly, S.** (2019) OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology*, **20**, 238. <https://doi.org/10.1186/s13059-019-1832-y>

**Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., et al.** (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, **29**, 644-652. <https://doi.org/10.1038/nbt.1883>

**Haas, B.J., Salzberg, S.L., Zhu, W., Pertea, M., Allen, J.E., Orvis, J., et al.** (2008) Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biology*, **9**, R7. <https://doi.org/10.1186/gb-2008-9-1-r7>

**Han, Y. and Wessler, S.R.** (2010) MITE-Hunter: a program for discovering miniature inverted-repeat transposable elements from genomic sequences. *Nucleic Acids Research*, **38**, e199. <https://doi.org/10.1093/nar/gkq862>

**Keller, O., Kollmar, M., Stanke, M. and Waack, S.** (2011) A novel hybrid gene prediction method employing protein multiple sequence alignments. *Bioinformatics*, **27**, 757-763. <https://doi.org/10.1093/bioinformatics/btr010>

**Kim, D., Paggi, J.M., Park, C., Bennett, C. and Salzberg, S.L.** (2019) Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*, **37**, 907-915. <https://doi.org/10.1038/s41587-019-0201-4>

**Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H. and Phillippy, A.M.** (2017) Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Research*, **27**, 722-736. <https://doi.org/10.1101/gr.215087.116>

**Li, H.** (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint*. <https://doi.org/arXiv:1303.3997>

**Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al.** (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078-2079. <https://doi.org/10.1093/bioinformatics/btp352>

**Librado, P., Vieira, F.G. and Rozas, J.** (2012) BadiRate: estimating family turnover rates by likelihood-based methods. *Bioinformatics*, **28**, 279-281. <https://doi.org/10.1093/bioinformatics/btr623>

**Liu, B., Shi, Y., Yuan, J., Hu, X., Zhang, H., Li, N., et al.** (2013) Estimation of genomic characteristics by analyzing k-mer frequency in de novo genome projects. <https://doi.org/arXiv:1308.2012>

**Lowe, T.M. and Eddy, S.R.** (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research*, **25**, 955-964. <https://doi.org/10.1093/nar/25.5.955>

**Lowe, T.M. and Eddy, S.R.** (1999) A computational screen for methylation guide snoRNAs in yeast. *Science*, **283**, 1168-1171. <https://doi.org/10.1126/science.283.5405.1168>

**Muhlhausen, S., Hellkamp, M. and Kollmar, M.** (2015) GenePainter v. 2.0 resolves the taxonomic distribution of intron positions. *Bioinformatics*, **31**, 1302-1304. <https://doi.org/10.1093/bioinformatics/btu798>

**Ou, S. and Jiang, N.** (2018) LTR\_retriever: A Highly accurate and sensitive program for identification of long terminal repeat retrotransposons. *Plant Physiology*, **176**, 1410-1422. <https://doi.org/10.1104/pp.17.01310>

**Ou, S., Su, W., Liao, Y., Chougule, K., Agda, J.R.A., Hellinga, A.J., et al.** (2019) Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biology*, **20**, 275. <https://doi.org/10.1186/s13059-019-1905-y>

**Price, M.N., Dehal, P.S. and Arkin, A.P.** (2010) FastTree 2 — approximately maximum-likelihood trees for large alignments. *PLoS One*, **5**, e9490. <https://doi.org/10.1371/journal.pone.0009490>

**She, R., Chu, J.S., Wang, K., Pei, J. and Chen, N.** (2009) GenBlastA: enabling BLAST to identify homologous gene sequences. *Genome Research*, **19**, 143-149. <https://doi.org/10.1101/gr.082081.108>

**Shi, J. and Liang, C.** (2019) Generic repeat finder: A high-sensitivity tool for genome-wide *de novo* repeat detection. *Plant Physiology*, **180**, 1803-1815. <https://doi.org/10.1104/pp.19.00386>

**Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V. and Zdobnov, E.M.** (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, **31**, 3210-3212. <https://doi.org/10.1093/bioinformatics/btv351>

**Smit, A.F.A., Hubley, R. and Green, P.** RepeatMasker.

**Su, W., Gu, X. and Peterson, T.** (2019) TIR-learner, a new ensemble method for TIR transposable element annotation, provides evidence for abundant new transposable elements in the maize genome. *Molecular Plant*, **12**, 447-460. <https://doi.org/10.1016/j.molp.2019.02.008>

**Sun, G., Xu, Y., Liu, H., Sun, T., Zhang, J., Hettenhausen, C., et al.** (2018) Large-scale gene losses underlie the genome evolution of parasitic plant *Cuscuta australis*. *Nature Communications*, **9**, 2683. <https://doi.org/10.1038/s41467-018-04721-8>

**Thompson, J.D., Higgins, D.G. and Gibson, T.J.** (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, **22**, 4673-4680. <https://doi.org/10.1093/nar/22.22.4673>

**Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., et al.** (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature protocols*, **7**, 562-578. <https://doi.org/10.1038/nprot.2012.016>

**Vilella, A.J., Severin, J., Ureta-Vidal, A., Heng, L., Durbin, R. and Birney, E.** (2009) EnsemblCompara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates. *Genome Research*, **19**, 327-335. <https://doi.org/10.1101/gr.073585.107>

**Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., et al.** (2014) Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One*, **9**, e112963. <https://doi.org/10.1371/journal.pone.0112963>

**Xiong, W., He, L., Lai, J., Dooner, H.K. and Du, C.** (2014) HelitronScanner uncovers a large overlooked cache of Helitron transposons in many plant genomes. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, 10263-10268. <https://doi.org/10.1073/pnas.1410068111>

**Xu, Z. and Wang, H.** (2007) LTR\_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Research*, **35**, W265-268. <https://doi.org/10.1093/nar/gkm286>

**Yang, X., Liu, D., Liu, F., Wu, J., Zou, J., Xiao, X., et al.** (2013) HTQC: a fast quality control toolkit for Illumina sequencing data. *BMC Bioinformatics*, **14**, 33. <https://doi.org/10.1186/1471-2105-14-33>

**Yuan, Y., Jin, X., Liu, J., Zhao, X., Zhou, J., Wang, X., et al.** (2018) The *Gastrodia elata* genome provides insights into plant adaptation to heterotrophy. *Nature Communications*, **9**, 1615. <https://doi.org/10.1038/s41467-018-03423-5>

**Zhang, R.-G., Wang, Z.-X., Ou, S. and Li, G.-Y.** (2019) TEsorter: lineage-level classification of transposable elements using conserved protein domains.[*https://doi.org/10.1101/800177*](https://doi.org/10.1101/800177). <https://doi.org/> <https://doi.org/10.1101/800177>