nature plants

Article

https://doi.org/10.1038/s41477-023-01562-2

Subgenome dominance shapes novel gene evolution in the decaploid pitcher plant *Nepenthes gracilis*

In the format provided by the authors and unedited

List of Supplementary Materials:

Supplementary Texts 1–7

Supplementary Tables 1–18 (separate file)

Supplementary Figs. 1–29

Supplementary References

Supplementary Dataset (separate file on Dryad: https://doi.org/10.5061/dryad.xsj3tx9mj)

Supplementary Texts

Supplementary Text 1. No strong gene-tree signal for allopolyploidization between major Caryophyllales carnivorous plant lineages. Previous phylogenetic studies placed Nepenthaceae sister either to the Droseraceae[13,118–120] or to the Ancistrocladaceae-Dioncophyllaceae-Drosophyllaceae (ADD) clade[12,55,121–123]. This may be explained by the possibility that Nepenthaceae emerged by allopolyploidization between two lineages: one close to Droseraceae and another close to the ADD clade. To test this possibility, we analyzed the 17 eudicot species (Supplementary Table 6) using Grampa v1.3.1 (https://github.com/gwct/grampa), which accounts for individual gene tree topology to distinguish auto- and allopolyploidy events[124]. To obtain a high-confidence set of gene trees, including lineage-specific gene duplication events that provide the signal of WGDs, we extracted BUSCO genes with the eudicot dataset of OrthoDB v10 (eudicots_odb10), including those labeled as duplicated. Using the translated sequences aligned with MAFFT and TrimAl, the ML trees were reconstructed using IQ-TREE with the LG+R4 substitution model and were rooted by NOTUNG[125]. The analysis of the 2,326 gene trees using Grampa with *N. gracilis* assigned as H1 yielded the best multi-labeled tree, which infers an allopolyploidization between the ancestral lineages of Droseraceae and the ADD clade, with 21 mapped duplication events that were judged as supporting the sister relationship between *N. gracilis* gene(s) and Droseraceae gene(s) (Supplementary Table 17). However, manual inspection of the allopolyploidization-supporting 24 gene trees that contained multiple *N. gracilis* genes revealed prevalent *N. gracilis*-specific duplications and paralogy, or misplacement of *N. gracilis* genes. As a result, no analyzed gene trees actually showed the topology expected from the allopolyploidization scenario: i.e., ((*N. gracilis* gene 1, ADD genes), (*N. gracilis* gene 1, Droseraceae genes)).

Another aspect of the data that could potentially indicate a history of allopolyploidization is the topologies of phylogenies for genes that have returned to single-copy after such an event. Here, we took advantage of the advanced diploidization which has occurred since polyploidization events took place in *Nepenthes*: the set of BUSCO genes (eudicots_odb10) is largely single-copy in *Nepenthes* (Supplementary Fig. 3d). If *Nepenthes* involved allopolyploid hybridization between the ADD and Droseraceae lineages followed by re-diploidization, we would expect that single-copy genes place *Nepenthes* either as sister to the Droseraceae or sister to the ADD families (and their relative proportions possibly manifesting as coincident with subgenome dominance, see main text). Strong gene tree incongruence of this kind is indeed observed in the data (quartet support only ~0.45 in the ASTRAL species tree, Supplementary Fig. 3e). However, gene tree incongruence will not necessarily derive from hybridization but can also be caused by incomplete lineage sorting (ILS): i.e., rapid and nearly simultaneous speciation events among the three lineages (Nepenthaceae, Droseraceae, and ADD families). To distinguish between these alternatives, we applied Species Networks applying Quartets (SNaQ)[126], an ILS-aware (coalescent) method to infer phylogenetic networks on protein sequences of single-copy genes in *Nepenthes*. Since the inference of phylogenetic networks is computationally intensive, the number of taxa on a gene tree must be kept relatively small. We, therefore, pruned the available gene trees to subsets of nine taxa, and in the case that taxa were represented by more than one gene copy, these taxa were discarded entirely (treated as missing data; consistent with the method used for species tree inference in the main text; Supplementary Fig. 29). We evaluated ten taxon subsets including the same seven ingroup taxa (*Nepenthes*, *Triphyophyllum*, *Drosophyllum*, *Ancistrocladus*, *Dionaea*, *Aldrovanda*, and *Drosera*), and two randomly chosen outgroup taxa. Starting from a bifurcating topology inferred by ASTRAL v5.7.8[18], phylogenetic networks with zero and then one hybrid edge were inferred and optimized using the SNaQ approach with PhyloNetworks v0.15.3 (https://github.com/crsl4/PhyloNetworks.jl) with 24 parallel attempts. While network optimization with one hybrid edge by SNaQ was successful in all ten taxon sets, *Nepenthes* did not descend from a hybridization event in any of them. Instead, it was usually the Droseraceae lineage which was constructed as hybridized with a ghost lineage directly descending from the ingroup stem (seven taxon sets), or the ADD clade as descending from such an event (two taxon sets), or the ADD lineage as hybridized with the *Nepenthes* lineage (one taxon set). Although these analyses do not support the hypothesis that *Nepenthes* is an allopolyploid hybrid involving ancient Droseraceae and/or the ancient ADD clade, it must be noted that there is as yet no satisfying explanation for the observed gene tree incongruence. In particular, neither the bifurcating species trees nor the optimized phylogenetic

networks with one hybrid edge do fit the observed gene trees well (QuartetNetworkGoodnessFit v0.4.0[127], all empirical P values < 10^{-5}).

To be thorough, we analyzed gene location on subgenomes. If allopolyploidy occurred between the distant Droseraceae and ADD lineages, genes on particular subgenomes should show the signal of particular phylogenetic affinity. A total of 33.4% (129/386) of the gene trees with the allopolyploidization signal (including those containing only one *N. gracilis* gene) were located on the dominant subgenome, which is comparable and not significantly different from the expectation with all genes (32.3%, 10,971/22,621, 418 gene models are on unanchored scaffolds and are thus excluded; χ^2 test, P = 0.79, χ^2 = 0.069). Therefore, our results do not support the possibility of the allopolyploidization events between the two distant lineages (i.e., ancestral Droseraceae and ancestral ADD clade). Still, there are phylogenetic signals that support the affinity of *N. gracilis* genes with Droseraceae, and this may be explained by ancient hybridization events before polyploidization. In this scenario, to explain the subgenomic frequencies of gene trees supporting alternative topologies, enough time would have been needed for allelic recombination to occur after hybridization and before polyploidization. Homoeologous exchange[128], if it did occur, would also contribute to gene shuffling between subgenomes.

Supplementary Text 2. Complexity of inferring polyploidization history. The data currently available is insufficient to propose a limited number of plausible evolutionary scenarios for polyploidization history from diploid ancestors to the extant decaploid species of *Nepenthes*. We inferred from trees containing dominant versus recessive BUSCO genes that the dominant subgenome represents its own distinct branch from the *Nepenthes*/ADD stem lineage; however the exact branching order for dominant vs. recessive is not consistent through all phylogenetic analyses performed (Supplementary Fig. 6). It is nonetheless clear that the last WGD should have established the modern 5-subgenome organization through either a 1 + 4 or 2 + 3 polyploidization event. Depending on the mode of the earlier WGD events, the former scenario can be further subdivided into 1 + (1 + 3) and 1 + (2 + 2). WGDs that generated ancestral tetraploids (with two subgenomes) and an octoploid (with four subgenomes) in the above scenarios can be either autopolyploid or allopolyploid. The latter adds further complexity to the history of reticulate evolution through the combination of parental species. As such, the polyploidization history of *Nepenthes* remains largely unknown. However, it is clear that the last WGD event was allopolyploidization in order to establish the odd-numbered subgenome organization.

Supplementary Text 3. Expression of syntelogs in the recessive subgenomes. In our study, we investigated gene expression patterns in the context of whole-genome duplications (WGDs) in *Nepenthes*. As shown in Fig. 2f, the overall average differences in gene expression levels between the dominant and recessive subgenomes are evident. However, many syntelog pairs exhibited higher expression levels in the recessive subgenomes (Fig. 2g), which may seem paradoxical. Such mosaic patterns of expression dominance are, however, known to occur in other polyploid species[128]. To better understand this phenomenon, it is crucial to consider the timescale of the WGD events in *Nepenthes*. Our analysis suggests that the *Nepenthes* WGD events are likely ancient (Supplementary Fig. 20c), as decaploidy is clearly constant within the genus based on its uniform chromosome number[15]. Moreover, phylogenetic analyses with divergence time estimation suggest that the crown group of *Nepenthes* is far younger than the age of its stem lineage and therefore its split from other genera[25]. Consequently, most syntelogs initially suppressed due to ancient subgenome dominance were likely lost over time, perhaps mostly before the current *Nepenthes* diversity evolved. In the immediate aftermath of WGDs, a sharp contrast in expression levels between dominant and recessive genomes has frequently been observed in other naturally occurring species as well as artificial polyploid induction experiments in otherwise diploid plant species[6,8]. With the *Nepenthes* genome having undergone substantial evolutionary change since its ancient WGDs, it is reasonable to assume that the subgenomes have completed their diploidization phase, as also indicated by the gene fractionation analysis (Fig. 2b,c). Therefore, the syntelogs observed in the present-day recessive subgenomes represent a small subset of survivors from the massive gene losses that occurred during ancient diploidization (Supplementary Fig. 7). This mosaic survival pattern could be attributed to processes such as neo- or sub-functionalization events[81] that

permitted these syntelogs to persist and possibly adopt new roles in the descendant modern *Nepenthes* genome.

Supplementary Text 4. Detection of the male-specific region of the Y-chromosome. In order to locate the MSY, we utilized a read depth-based analysis (Fig. 3a). This method, however, is inherently subject to noise due to various factors such as artifacts during library preparation, sequencing, and read mapping. This is especially pronounced in Restriction site-Associated DNA sequencing (RAD-seq) data, where the presence or absence of reads can be more stochastic. To address these challenges, we ensured uniform treatment protocols for both males and females throughout all laboratory and computational steps. A total of 21 unrelated individuals (Supplementary Table 1) were used for the analysis.

To establish statistical significance, we used a specific threshold derived from sex-based permutations with the read-depth data of each genomic window. This method is superior to a fixed genome-wide threshold as it considers the empirical and specific noise level of each window to generate the null distribution. A window was considered as a significant outlier only if the observed difference in male/female read depths (log_2-scaled ratio) exceeded the 99.9th percentile of the sex-permuted distribution for that specific window plus 0.1.

As shown in Fig. 3a, only one region with a few windows (window size 50 kb: chromosome 20: 7,650,000-8,750,000 bp; window size 100 kb: chromosome 20: 7,600,000-8,600,000 bp) has observed data (blue) that surpass the thresholds (grey). All other apparent peaks, such as those on chromosomes 7 and 37, fall within the sex-permuted null distributions (grey points exceeding blue points). These peaks appear because the noise level of read depth is higher in these windows. For example, the peak on chromosome 7 appears to encompass the centromere, a repeat-rich region where short-read mapping is notoriously challenging. However, these peaks are clearly dismissed by our significance test at the window sizes of 50 kb and 100 kb. We note that at the smallest window size (10 kb), some additional windows on several chromosomes passed the significance thresholds. We excluded these as false-positives because they did not intersect with results from the other window sizes and because the underlying read coverage data are naturally more coarse.

In addition to the mapping-based approach, we also utilized a second method based on male-specific *k*-mers as extracted from the raw sequencing reads without mapping (Methods; Supplementary Fig. 9). Only in a second step were these male-specific *k*-mers aligned to the genome. This analysis consistently identified a male-specific peak in the same region of chromosome 20. Meanwhile, minor peaks detected in the mapping-based approach, such as those on chromosomes 7 and 37, were not corroborated, further substantiating the robustness of our MSY detection.

Supplementary Text 5. Expression patterns of LEAFY in flowering plants. The FLORICAULA/LEAFY (*FLO*/*LFY*) gene, a critical regulator of flower development[47], typically forms a single-member gene family (i.e., no paralog) within angiosperm genomes[48], and exhibits expression in flower meristems and inflorescence meristems across a variety of angiosperms, including *Arabidopsis thaliana* (rosid, Brassicaceae)[129], *Ionopsidium acaule* (rosid, Brassicaceae)[130], *Pisum sativum* (rosid, Fabaceae)[131], *Carya cathayensis* (rosid, Juglandaceae)[132], *Vitis vinifera* (rosid, Vitaceae)[133], *Antirrhinum majus* (asterid, Plantaginaceae)[134], *Nicotiana tabacum* (asterid, Solanaceae)[135], *Solanum lycopersicum* (asterid, Solanaceae)[136], *Chrysanthemum lavandulifolium* (asterid, Asteraceae)[137], *Eschscholzia californica* (sister lineage to the core eudicots, Papaveraceae)[138], *Phalaenopsis hybrida* (monocot, Orchidaceae)[139], and *Zea mays* (monocot, Poaceae)[140]. In *Titanotrichum oldhamii* (asterid, Gesneriaceae), the reversal transition of reproductive to vegetative phases is accompanied by a great reduction of the *LFY* expression level in apical meristems[141]. It is also noteworthy that the single-copy *LFY* gene in *Carica papaya* (rosid, Caricaceae) is expressed in male, female, and hermaphroditic flowers, with marginally higher expression levels in unisexual flowers[142]. While *LFY* expression in other organs, such as leaves, vegetative meristems, and individual floral organs, tends to be lineage-specific, expression in flower meristems and inflorescence meristems remains largely conserved among flowering plants.

Supplementary Text 6. Potential for analyzing male-specific genes in *Nepenthes* by gene transfer to genetically tractable plants. Our analyses have identified the male-specific presence and expression

of *LFY-Y* in *Nepenthes*. However, the functional evidence supporting *LFY-Y* as a primary candidate for the sex determinant gene remains limited. The lack of reliable, stable genetic transformation techniques for *Nepenthes* restricts opportunities for functional validation within the species under study. An alternative approach involves introducing *Nepenthes LFY-Y* into model plants to assess whether it alters the morphology of hermaphroditic flowers. This approach assumes that *LFY-Y* can regulate a set of downstream genes that are sufficiently similar between the donor species (*N. gracilis*) and the recipient species (e.g., *Arabidopsis thaliana*). This strategy has been widely employed and has successfully facilitated the identification of sex determination genes in several angiosperm species. For instance, a Y-linked CLAVATA3-like gene in *Silene latifolia*, when introduced to *A. thaliana* under the control of its native promoter, inhibits pistil growth, thereby providing evidence for its role as a female suppressor[143]. Similarly, the Y-linked gene Shy Girl (*SyGI*) in *Actinidia chinensis* induces female sterility when introduced to *Nicotiana tabacum* using its native promoter[144]. Another Y-linked gene from *Actinidia*, Friendly Boy (*FrBy*), has been shown to complement the male sterility phenotype of ortholog knockout lines in *N. tabacum*[145]. In the context of *Nepenthes*, we have detected genetic changes in both *cis*-regulatory elements (Fig. 3f) and protein-coding sequences (Supplementary Fig. 15) in *LFY-Y*. Therefore, it would be beneficial to compare the effects of genetic transfers on both promoter and coding sequences. This could potentially be achieved in the future through promoter swapping experiments between *LFY-A* and *LFY-Y* conducted in model plants.

Supplementary Text 7. Evolution of dioecy in Caryophyllales. Previous research has identified 400 dioecious species that represent 12 independent dioecious lineages within Caryophyllales[146]. In the case of *Spinacia oleracea* (Amaranthaceae), a Y-specific *NRT1/PTR6.4* gene encoding a major facilitator superfamily protein has been identified as a strong candidate for its single-factor sex determination[147]. *Silene latifolia* (Caryophyllaceae), on the other hand, is believed to have evolved dioecy via the two-factor model, which postulates the presence of separate sterility mutations for male and female functions, and a gene encoding a CLAVATA3-like secreted peptide hormone has been identified as the female suppressor[143]. Our findings in *Nepenthes*, along with these studies, suggest that the Caryophyllales, like other plant lineages, has a complex history of dioecy evolution, with largely lineage-specific patterns in sex determination systems and identities of sex determinants.

Supplementary Tables

Supplementary Table 1. DNA sequencing.

Supplementary Table 2. Genome assembly statistics.

Supplementary Table 3. Repetitive elements in the male assembly annotated by RepeatMasker.

Supplementary Table 4. RNA sequencing.

Supplementary Table 5. Annotation and expression of *Nepenthes* genes. The *P* values for differential expression were calculated using two-sided quasi-likelihood *F*-tests. Corrections for multiple testing were made using the Benjamini-Hochberg FDR method.

- gene_id: gene ID. One gene per line.
- orthogroup: Orthofinder-based orthogroup ID to which the gene is assigned.
- sprot_best: UniProt ID of DIAMOND BLASTP best hit.
- sprot_alias: Alias ID (if any) of sprot_best.
- sprot_coverage: DIAMOND BLASTP hit coverage. Q and H indicate the coordinates in queries and hits, respectively.
- sprot_identity: The percentage of sequence identity in sprot_coverage.
- sprot_evalue: DIAMOND BLASTP E-value of sprot_best.
- sprot_recnam: Gene name of sprot_best.
- signalp_start: Signal peptide start coordinate predicted by SignalP.
- signalp_start: Signal peptide end coordinate predicted by SignalP.
- signalp_score: SignalP score.
- tmhmm_expa: Expected number of amino acids in transmembrane helices predicted by TMHMM.
- tmhmm_predhel: Number of predicted transmembrane helices by N-best.
- tmhmm_topology: Topology predicted by N-best. i = inner, o = outer.
- kegg_gene: KEGG gene ID matched to sprot_best.
- kegg_orthology: KEGG orthology ID of kegg_gene.
- go_ids: GO IDs annotated to sprot_best.
- go_aspects: ordered GO aspects of go_ids.
- go_terms: ordered GO terms of go_ids.
- busco_id: BUSCO gene ID.
- busco_status: BUSCO status.
- busco_sequence: BUSCO hit coordinate of gene_id.
- busco_score: BUSCO score.
- busco_length: BUSCO gene length.
- busco_orthodb_url: OrthoDB URL of the BUSCO gene.
- busco_description: Gene description of the BUSCO gene.
- feature_size: Size of coding sequences (bp).
- num_intron: Number of introns.
- chromosome: Chromosome, scaffold, or contig on which the gene locates.
- start: Start coordinate of the gene.
- end: End coordinate of the gene.
- strand: Strand of the gene.
- TISSUE_REPLICATE: Expression level (Log_2 TMM-FPKM).
- is_enough_tmmfpkm: Whether the maximum TMM-FPKM value is greater or equal to 0.1.
- max_tissue: Tissue with the maximum value of average TMM-FPKM.
- tau: Yanai's tau for the tissue specificity of gene expression obtained from TMM-FPKM values.
- is_enough_tau: Whether the tau value is greater or equal to 0.67.
- fdr_TISSUE: False discovery rate of the TISSUE versus TISSUEFed comparison by edgeR.
- is_enough_tissue_tmmfpkm_TISSUE: Whether the TMM-FPKM value is greater or equal to 0.1 in the TISSUE or TISSUEFed samples.

- logFC_TISSUE: log_2 fold change values (TISSUEFed/TISSUE) estimated by edgeR.
- logCPM_TISSUE: Log_2 average of CPM values over compared samples (TISSUE and TISSUEFed).
- F_TISSUE: F statistics of the TISSUE versus TISSUEFed comparison by edgeR.
- PValue_TISSUE: Uncorrected *P* value of the TISSUE versus TISSUEFed comparison by edgeR.

Supplementary Table 6. Plant genomes for evolutionary analyses.

Supplementary Table 7. JASPAR motifs detected in *LFY* promoters. *P* values were calculated to assess the likelihood of observing the given motif in random sequences. To account for multiple comparisons, adjustments were made using the Benjamini-Hochberg FDR method.

Supplementary Table 8. GO enrichment analysis of the genes up-regulated in the digestive zone. GO terms that are annotated to ≥3 genes and have an FDR value of <0.05 are highlighted in grey. *P* values were calculated using two-sided Fisher's exact tests. Corrections for multiple testing were made using the Benjamini-Hochberg FDR method.

Supplementary Table 9. GO enrichment analysis of the genes down-regulated in the digestive zone. GO terms that are annotated to ≥3 genes and have an FDR value of <0.05 are highlighted in grey. *P* values were calculated using two-sided Fisher's exact tests. Corrections for multiple testing were made using the Benjamini-Hochberg FDR method.

Supplementary Table 10. GO enrichment analysis of the genes up-regulated in the tendril. GO terms that are annotated to ≥3 genes and have an FDR value of <0.05 are highlighted in grey. *P* values were calculated using two-sided Fisher's exact tests. Corrections for multiple testing were made using the Benjamini-Hochberg FDR method.

Supplementary Table 11. GO enrichment analysis of the genes down-regulated in the tendril. GO terms that are annotated to ≥3 genes and have an FDR value of <0.05 are highlighted in grey. *P* values were calculated using two-sided Fisher's exact tests. Corrections for multiple testing were made using the Benjamini-Hochberg FDR method.

Supplementary Table 12. GO enrichment analysis of the syntelogs on the dominant subgenome. *P* values were calculated using two-sided Fisher's exact tests. Corrections for multiple testing were made using the Benjamini-Hochberg FDR method.

Supplementary Table 13. GO enrichment analysis of the syntelogs on the recessive subgenomes. *P* values were calculated using two-sided Fisher's exact tests. Corrections for multiple testing were made using the Benjamini-Hochberg FDR method.

Supplementary Table 14. GO enrichment analysis of the tandem duplicates on the dominant subgenomes. *P* values were calculated using two-sided Fisher's exact tests. Corrections for multiple testing were made using the Benjamini-Hochberg FDR method.

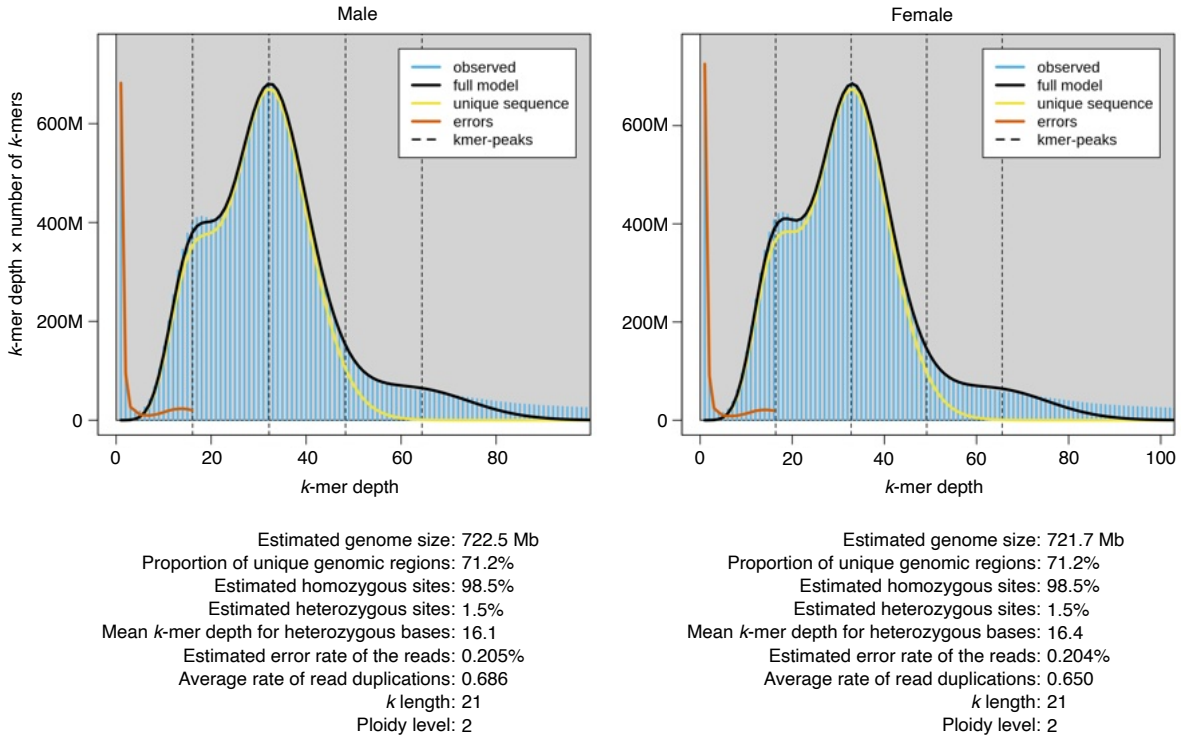
Supplementary Table 15. GO enrichment analysis of the tandem duplicates on the recessive subgenomes. *P* values were calculated using two-sided Fisher's exact tests. Corrections for multiple testing were made using the Benjamini-Hochberg FDR method.

Supplementary Table 16. The number of genes and codon alignment statistics in individual orthogroups.

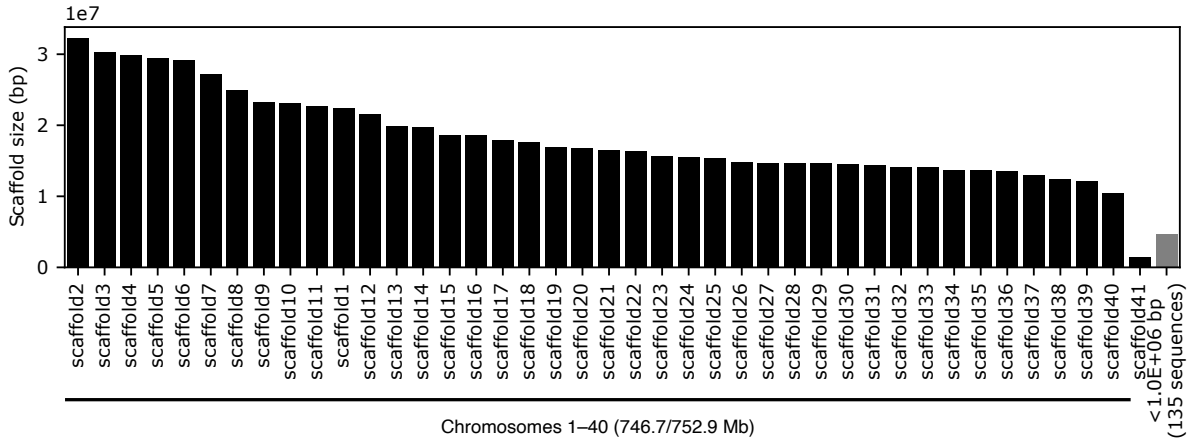
Supplementary Table 17. Detected gene duplication mapping to the best multi-labeled species tree.

Supplementary Table 18. Identification of BLAST-based orthologs encoding digestive fluid proteins.

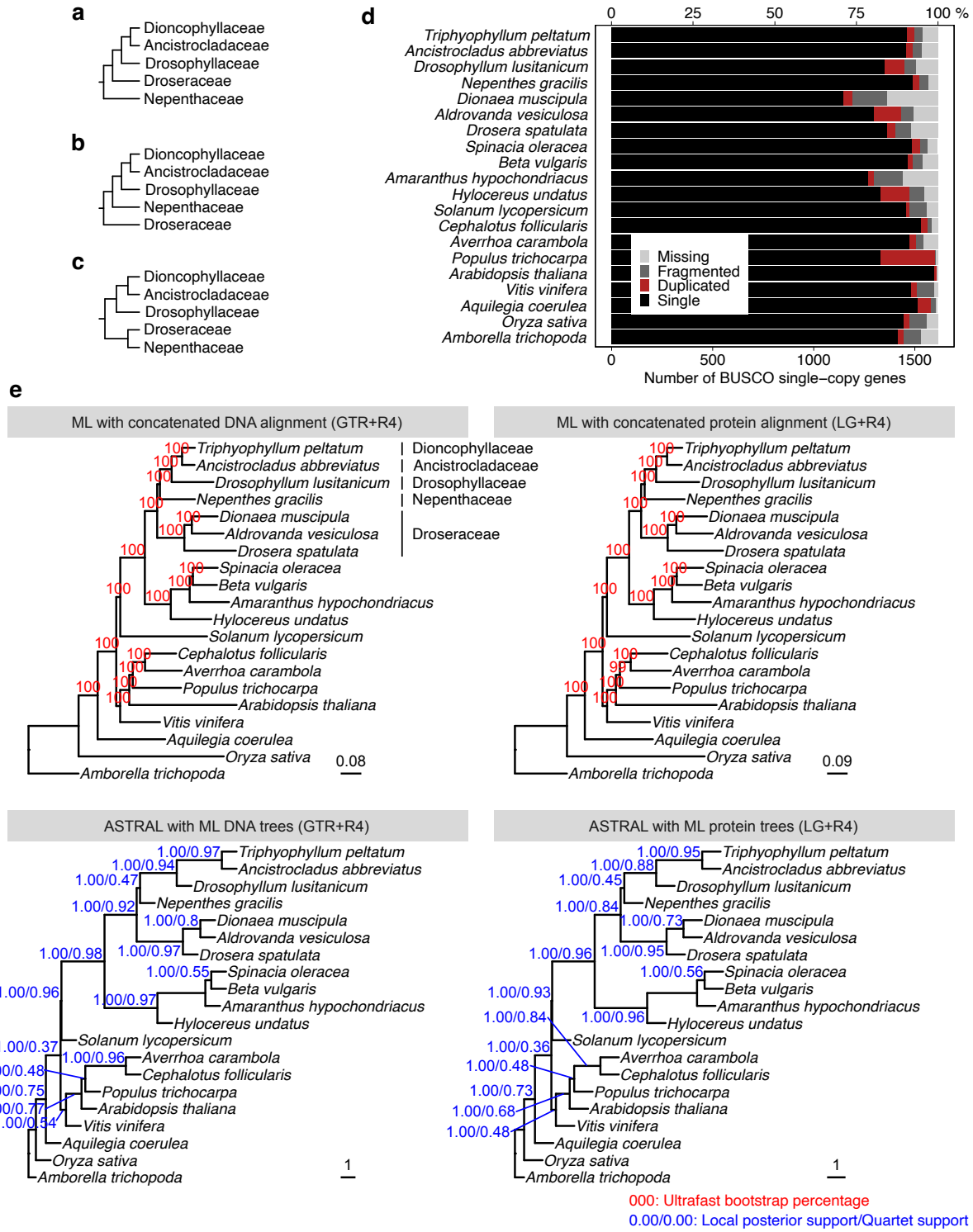
Supplementary Figures



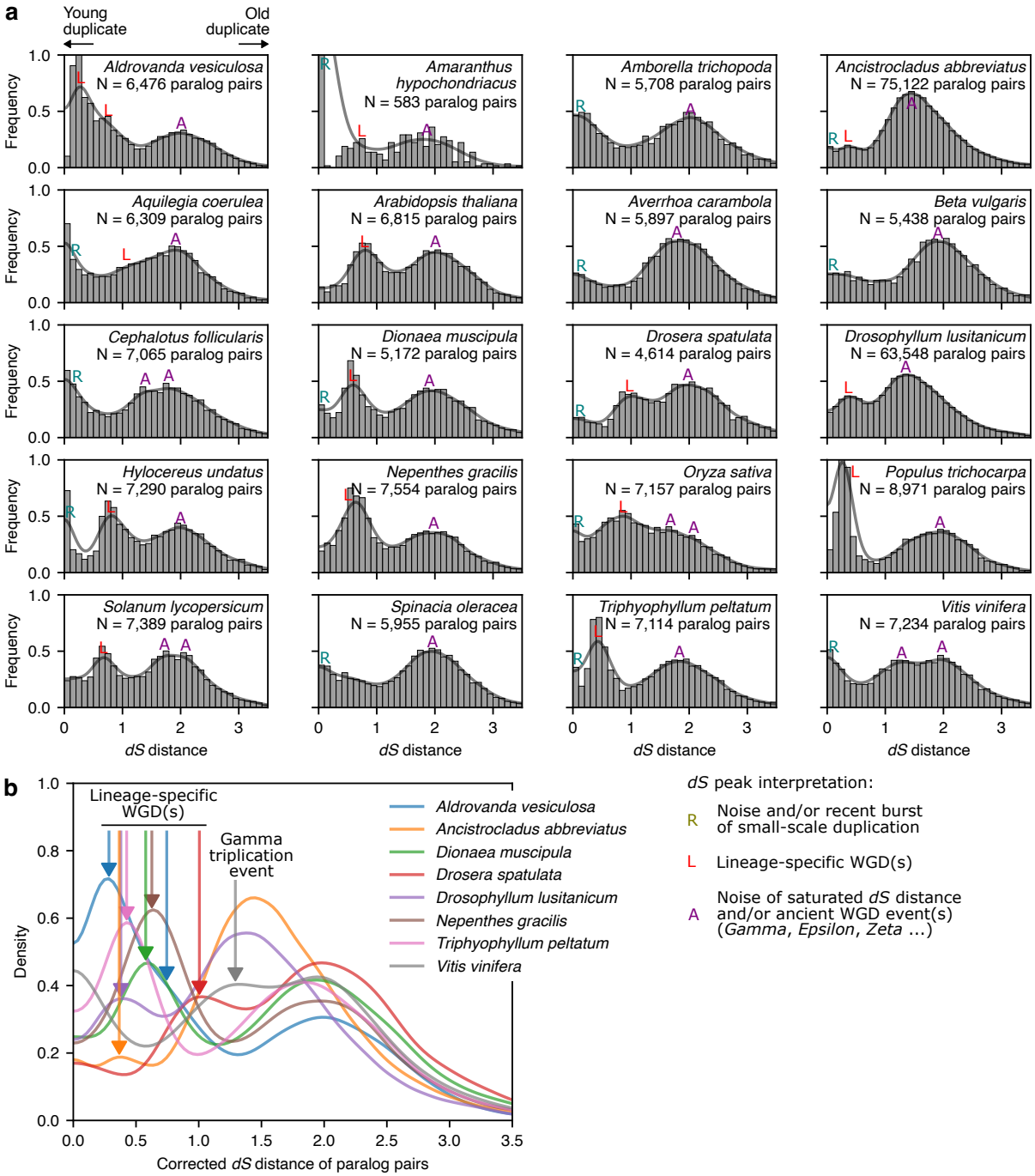
Supplementary Fig. 1. The *k*-mer frequency distribution of the *Nepenthes gracilis* male and female individuals. Illumina short reads (Supplementary Table 1) were analyzed using GenomeScope v2.0 (https://github.com/tbenavi1/genomescope2.0)[148].



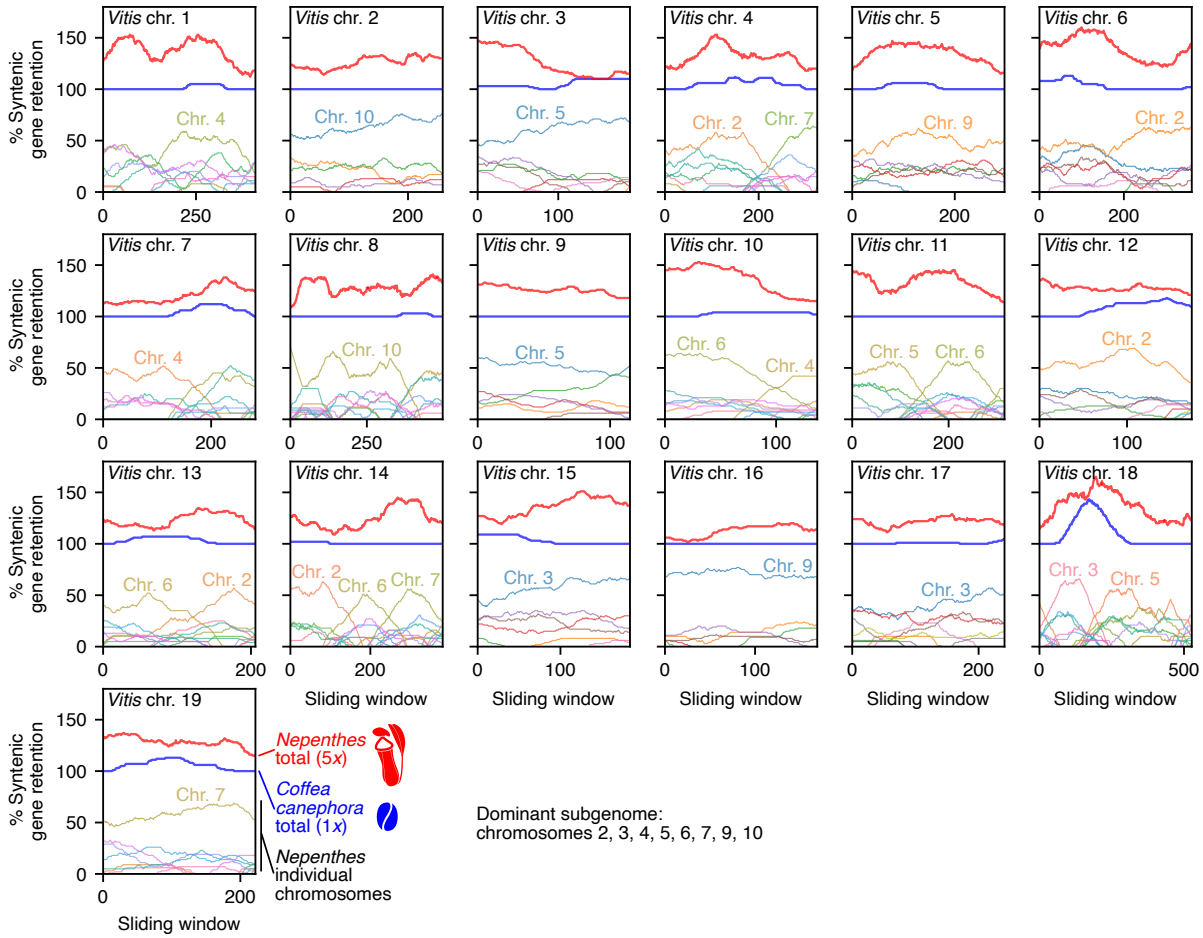
Supplementary Fig. 2. The scaffold size distribution of the *Nepenthes gracilis* male genome assembly. Scaffolds smaller than 1 Mb are merged to the right.



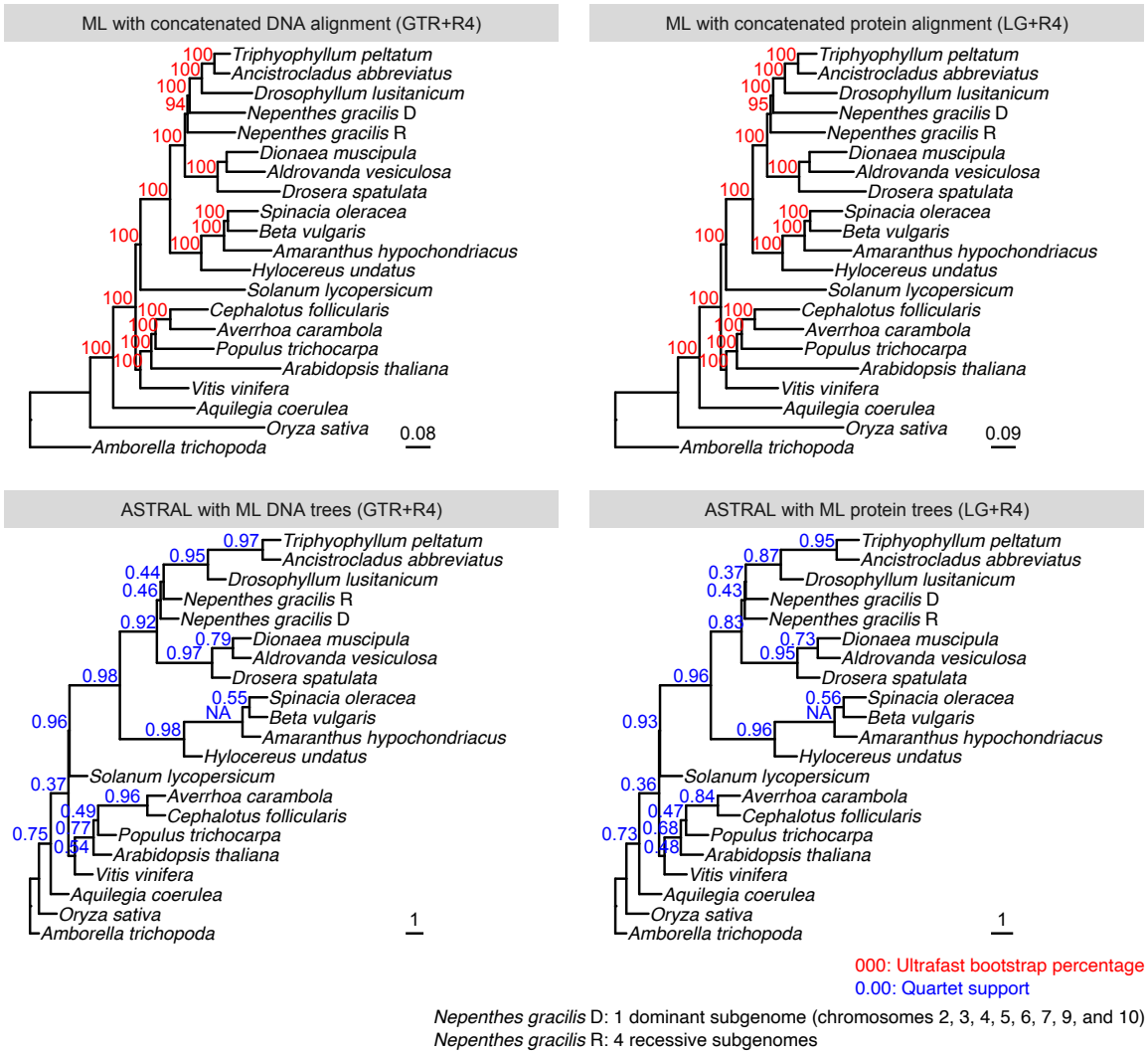
Supplementary Fig. 3. Phylogenetic hypotheses on the position of Nepenthaceae. (**a-c**) The positions in **a**[149–151], **b**[12,55,121–123], and **c**[13,118–120] are supported by previous research. (**d**) Phylogenetic relationships inferred by the maximum-likelihood (ML) method and a coalescence-based method (ASTRAL) with DNA and protein alignments of 1,614 single-copy genes conserved in Embryophyta (defined by BUSCO embryophyta_odb10; 2,575,635 nucleotide sites). All analyses support the tree topology shown in **b**. Branch lengths correspond to nucleotide and amino acid substitutions per site, respectively, in the upper left and the upper right panel. In the lower panels, lengths of internal branches correspond to coalescent units, while those of terminal branches are not estimated (see https://github.com/smirarab/ASTRAL/blob/master/astral-tutorial.md) and are arbitrarily shown only for visualization purposes.



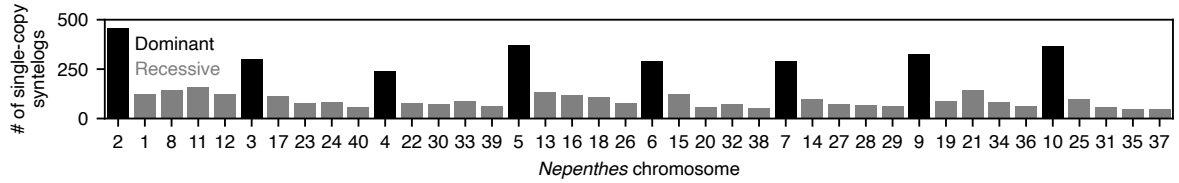
Supplementary Fig. 4. Distribution of synonymous divergence (*dS*) in paralogous gene pairs within the genome. (**a**) Analysis of the 20-genome dataset with WGDdetector[105]. Histograms are overlaid with a kernel density plot with the reflective lower boundary. Our interpretation of detected *dS* peaks is labeled in the plot. Recent bursts of small-scale gene duplications[152], as well as errors, such as incompletely purged haplotigs and splicing variants, can produce a peak near $dS = 0$ (R). Saturated *dS* distance and/or ancient WGD events, such as the *gamma* hexaploidization event shared by all extant eudicots, may produce a peak with large *dS* values (A). Lineage-specific WGDs (L) are expected to be located between the R and A peaks. (**b**) A comparison of lineage-specific WGDs in carnivorous Caryophyllales species. *V. vinifera* is included to anchor the location of the *gamma* hexaploidization event.



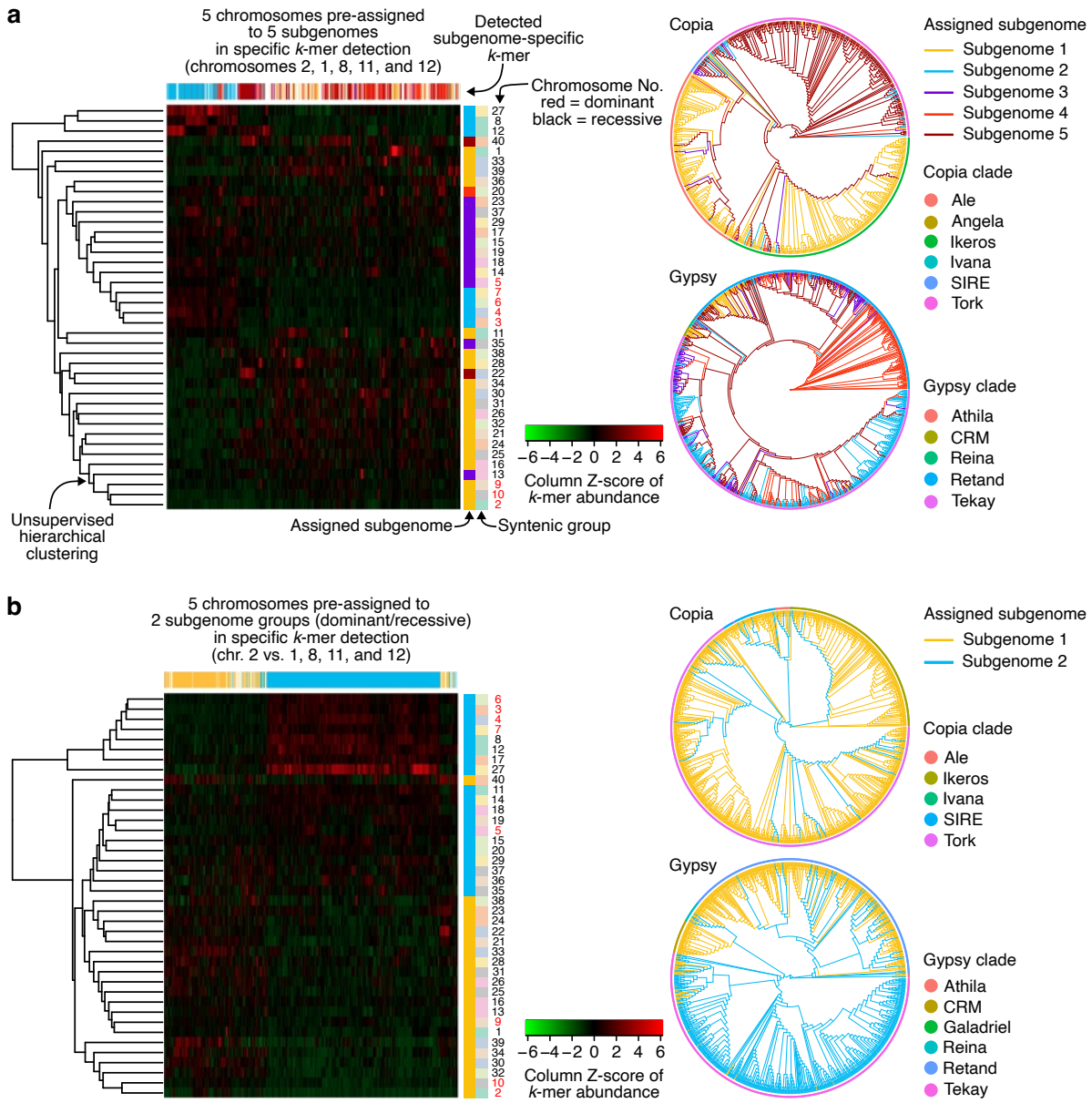
Supplementary Fig. 5. Analysis of fractionation bias in the *Nepenthes* genome. The rate of syntenic gene retention was evaluated by aligning the *Nepenthes* genome (5*x*) against the *V. vinifera* genome (1*x*), which is known not to have experienced a WGD after the *gamma* hexaploidization event[24]. Analysis of the *C. canephora* genome (1*x*) with the same FractBias parameters[26] is shown for comparison. Dominant chromosomes with clear signals are indicated in the plot. The results are reproducible on CoGe with the following links: *V. vinifera* – *N. gracilis* (https://genomevolution.org/r/1myic) and *V. vinifera* – *C. canephora* (https://genomevolution.org/r/1n0tj).



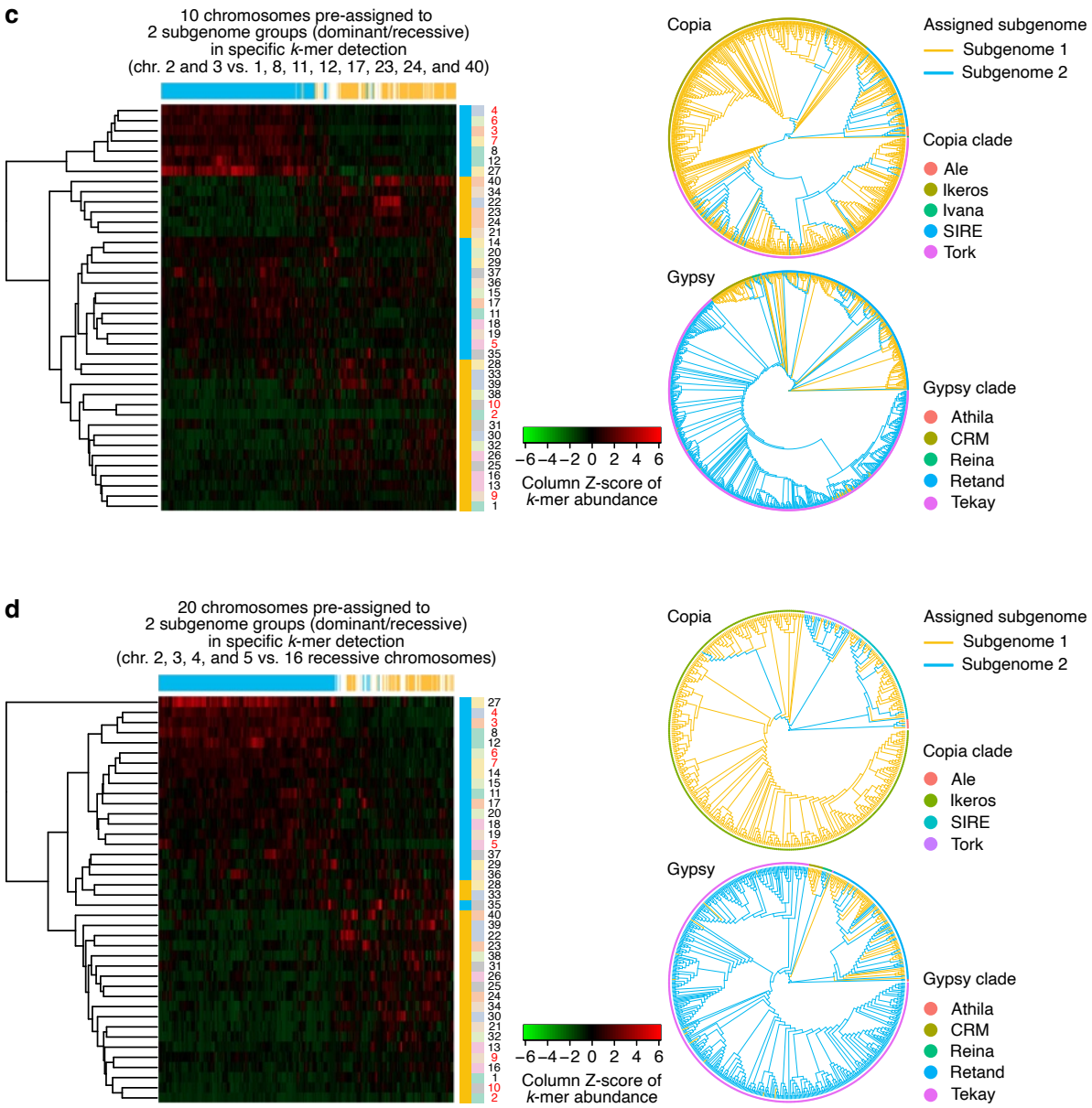
Supplementary Fig. 6. Phylogenetic analysis suggests that the separation of dominant and recessive subgenome donors is ancient. In this analysis, the *Nepenthes* genome is divided into two bins: one dominant subgenome, and the four recessive subgenome groups. This division permitted us to infer the timing of their separation. The branch lengths indicate that the separation of subgenome donors happened within a relatively short time frame, coinciding with the divergence of the ADD and Droseraceae stem lineages. This suggests that the WGDs in *Nepenthes* may be of ancient origin. Phylogenetic relationships were inferred by the maximum-likelihood (ML) method and a coalescence-based method (ASTRAL) with DNA and protein alignments of 1,614 single-copy genes conserved in Embryophyta (defined by BUSCO embryophyta_odb10). Branch lengths correspond to nucleotide and amino acid substitutions per site in the upper left and the upper right panel, respectively. In the lower panels, lengths of internal branches correspond to coalescent units, while those of terminal branches are not estimated (see https://github.com/smirarab/ASTRAL/blob/master/astral-tutorial.md) and are arbitrarily shown only for visualization purposes.



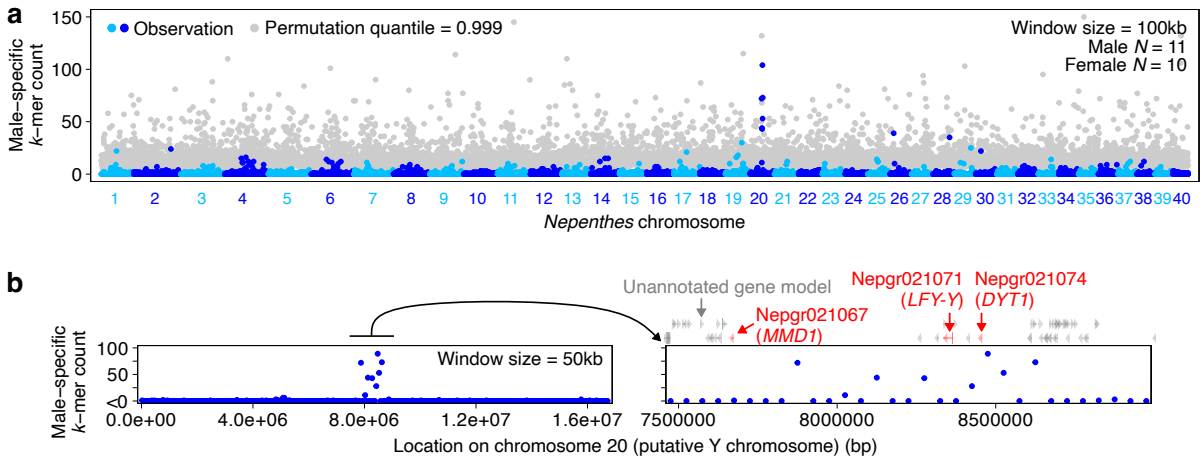
Supplementary Fig. 7. Chromosome-wise number of single-copy syntelogs in the *Nepenthes gracilis* genome as measured against the *Vitis* genome.



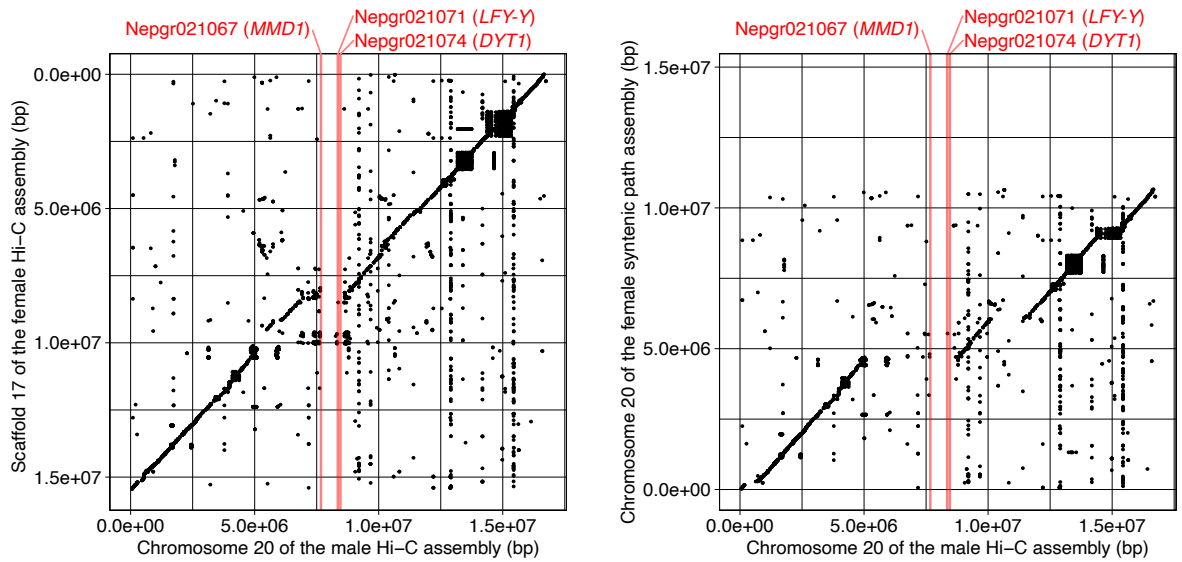
Supplementary Fig. 8. Subgenome phasing with specific *k*-mers. The *Nepenthes* male genome was analyzed using SubPhaser v1.2.5 (https://github.com/zhangrengang/SubPhaser)[29] with different sets of chromosomes in the subgenome specification for subgenome-specific *k*-mer detection. SubPhaser was run under four different conditions. In the first two conditions, five chromosomes were pre-assigned to either five (**a**) or two (**b**) subgenomes. In the third condition, ten chromosomes were pre-assigned to two subgenomes (**c**). In the fourth condition, 20 chromosomes were pre-assigned to two subgenomes (**d**). Heatmaps show the relative abundance of differential *k*-mers. Syntenic groups of homologous chromosomes are shown according to Fig. 2a. The relationships between the assigned subgenomes and Copia and Gypsy retrotransposons are depicted on the right. The subgenome phasing worked partially, with robust clustering of several chromosomes in the dominant subgenome (i.e., the clade of chromosomes 3, 4, 6, and 7 and the clade of chromosomes 2 and 10), but was not perfect even though different parameter settings were employed. No subgenome-specific *k*-mer was detected with all dominant and recessive chromosomes (8 and 32, respectively) specified for subgenome-specific *k*-mer detection. Depending on parameter settings, a tendency for specific members of Copia (e.g., Tork in **a**) and Gypsy (e.g., Tekay in **b-d**) elements to amplify in certain chromosome groups can be observed. A more detailed analysis of these transposable elements in the future could provide insights into the subgenomic evolution of *Nepenthes*.



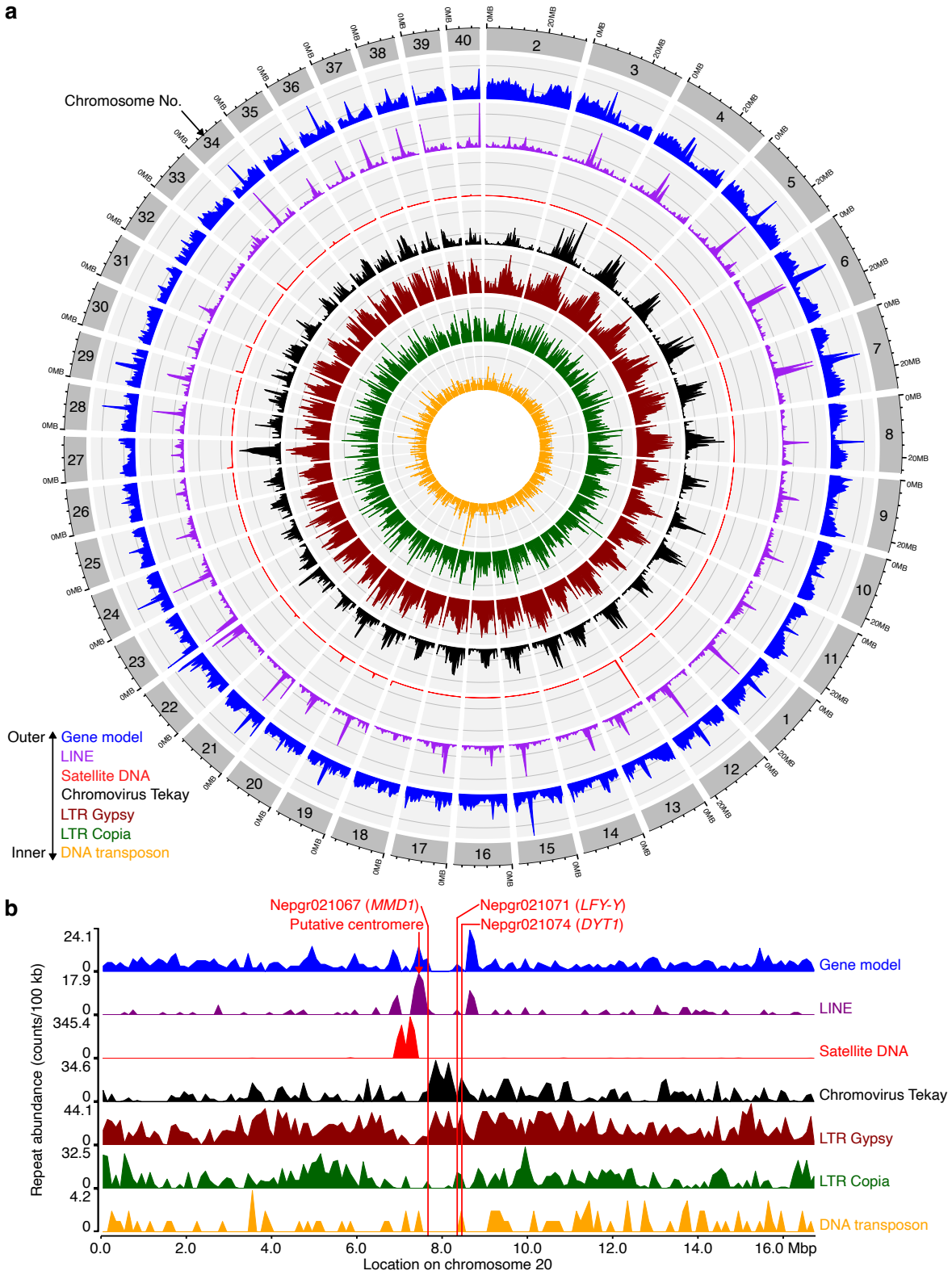
Supplementary Fig. 8 (continued)



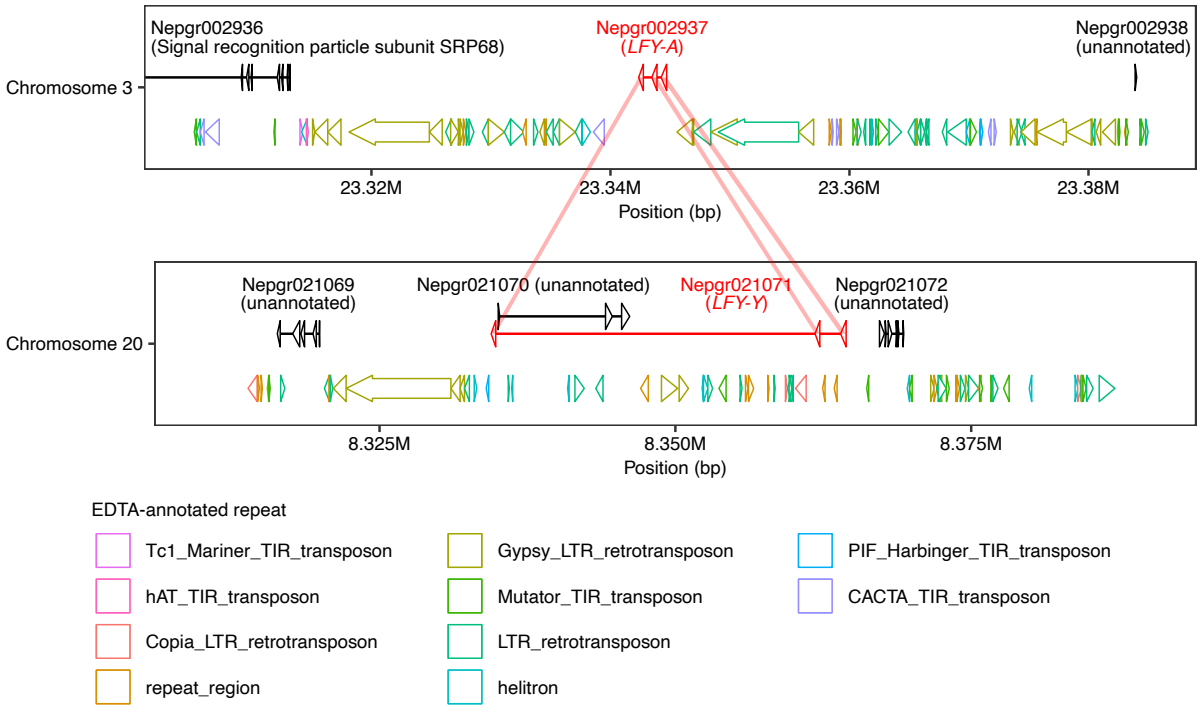
Supplementary Fig. 9. The frequency of male-specific *k*-mers in the *Nepenthes gracilis* male genome. (**a**) The genome-wide distribution of male-specific 16 mers. (**b**) Magnified views of the male-linked region on chromosome 20.



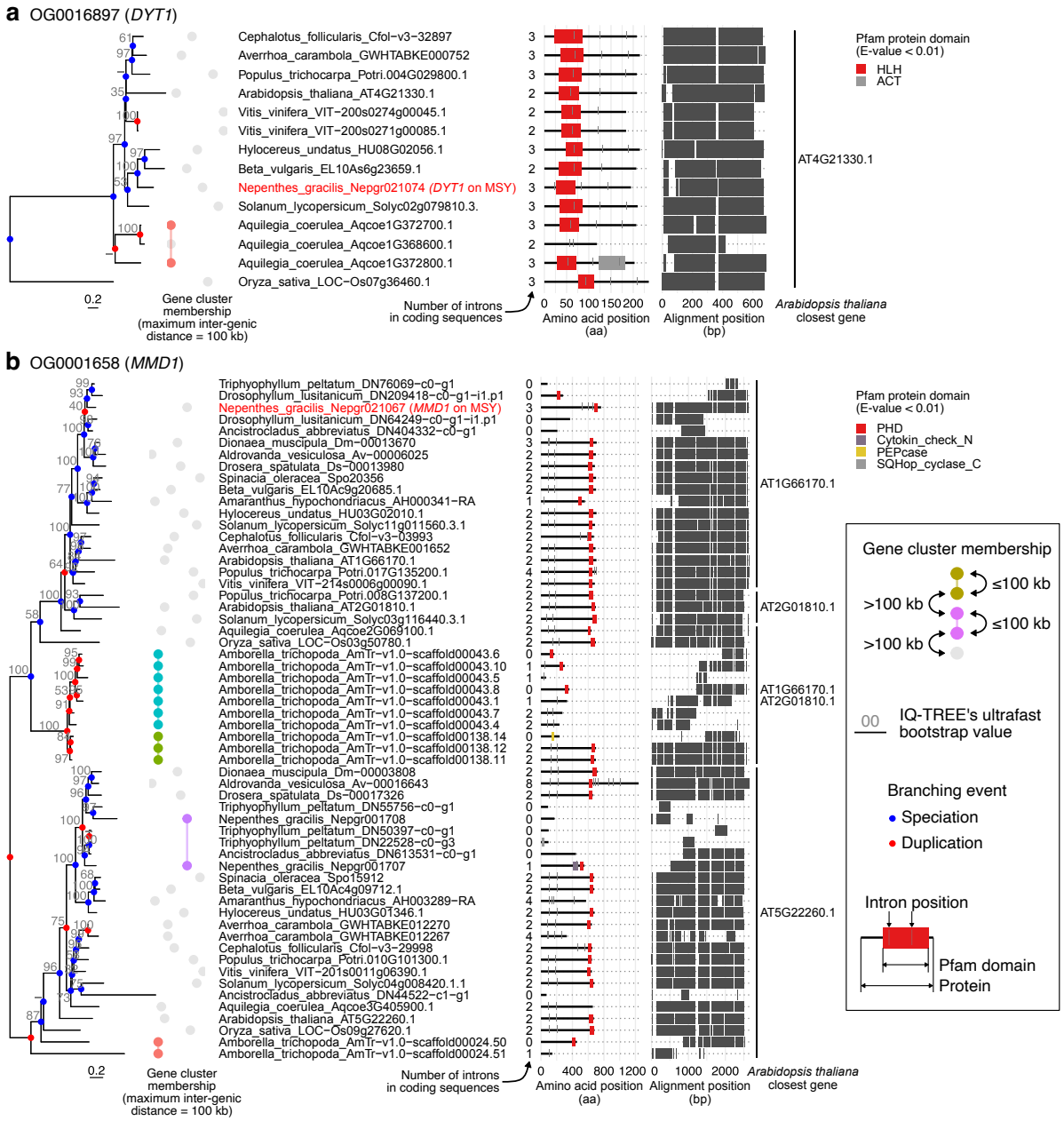
Supplementary Fig. 10. Syntenic comparison of the putative X and Y chromosomes. The female Nanopore assembly was scaffolded with the male Hi-C reads (left) or via syntenic path assembly against the male Hi-C assembly (right), and the resulting chromosome-scale female genome assemblies were then used for the dot plot comparison. The male coding sequences were searched against the female genomic sequences using LAST, and all pairwise hits are shown in the plot. The positions of the three male-specific transcription factor genes are indicated in red. The results are reproducible using the LAST outputs that are available for download from the following CoGe links: male Hi-C assembly versus female Hi-C assembly (https://genomevolution.org/r/1ir4m) and male Hi-C assembly versus female syntenic path assembly (https://genomevolution.org/r/1nqd6).



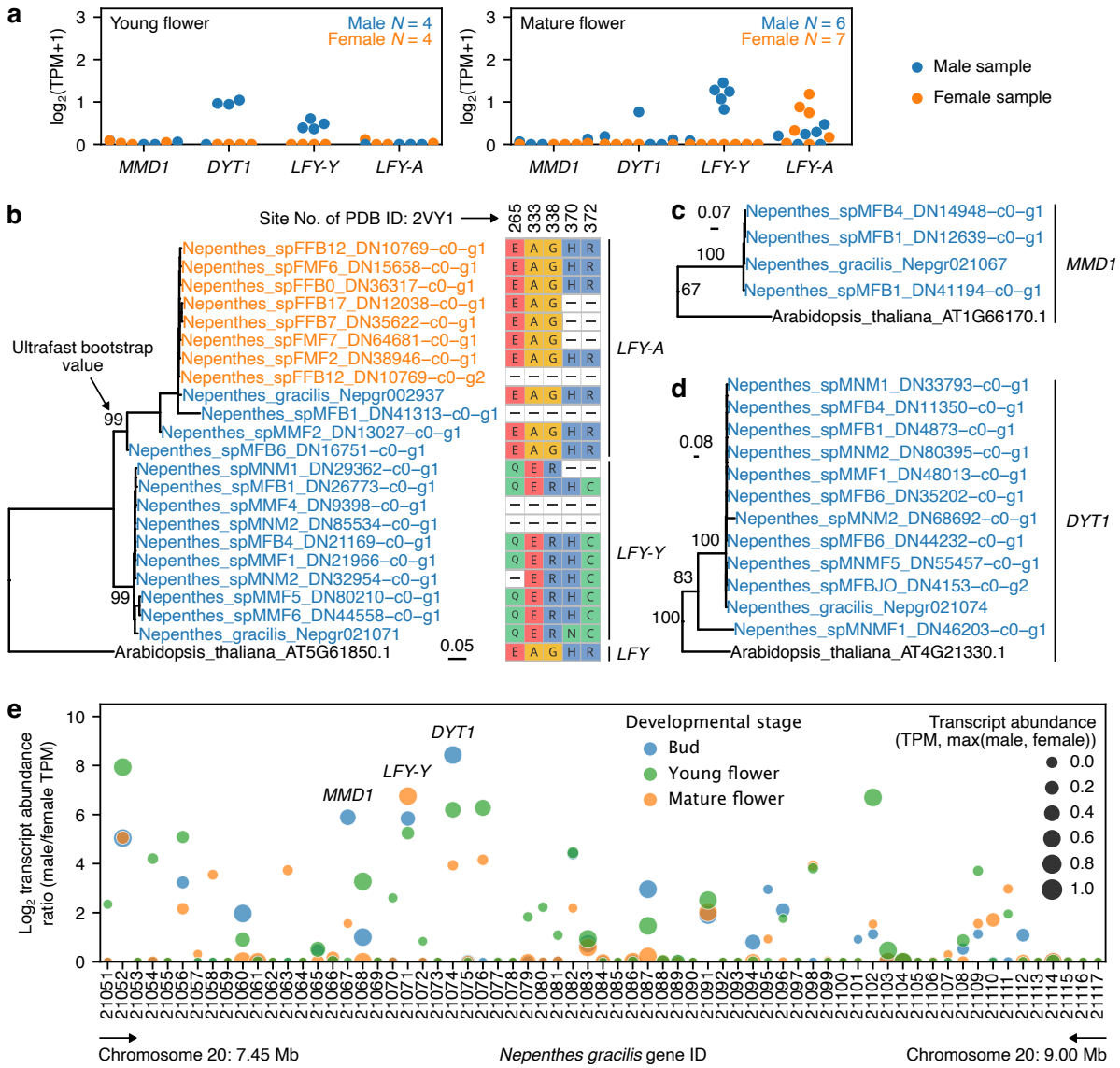
Supplementary Fig. 11. The repeat sequence profiles of the *N. gracilis* male genome. (**a**) Chromosome-wise distributions of representative repeat sequences. Window size: 500 kb. (**b**) A close-up view of chromosome 20. The locations of the putative centromere and male-specific transcription factor genes are indicated. Window size: 100 kb. LINE, long interspersed nuclear elements; LTR, long terminal repeat.



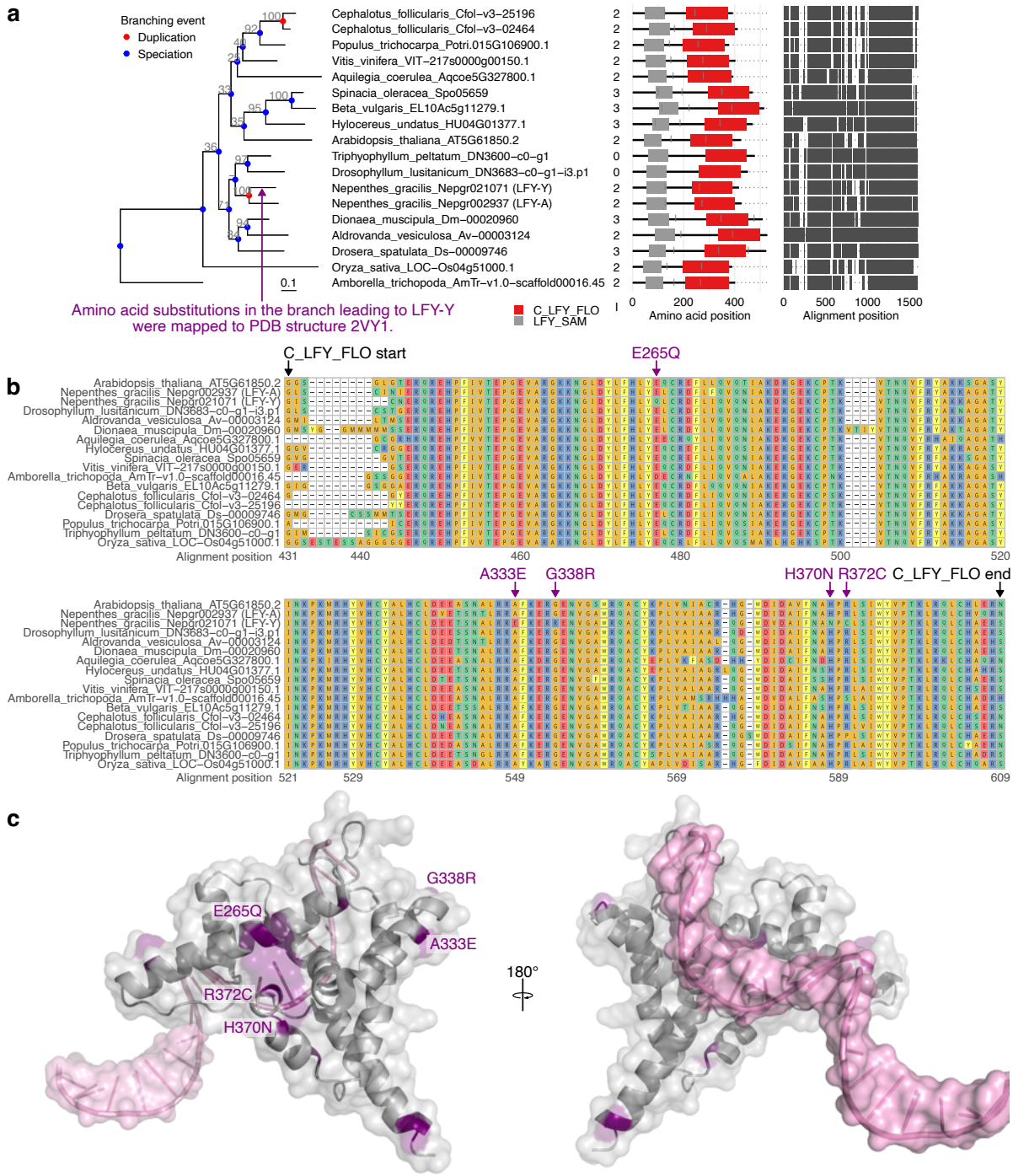
Supplementary Fig. 12. Gene structures of *LFY-A* and *LFY-Y*. Repeat sequences identified by EDTA[153] are shown below gene models. Corresponding coding sequences of *LFY-A* and *LFY-Y* are connected. Trinotate-based gene product names are provided in parentheses. The coordinates on chromosomes 3 and 20 have the same scale.



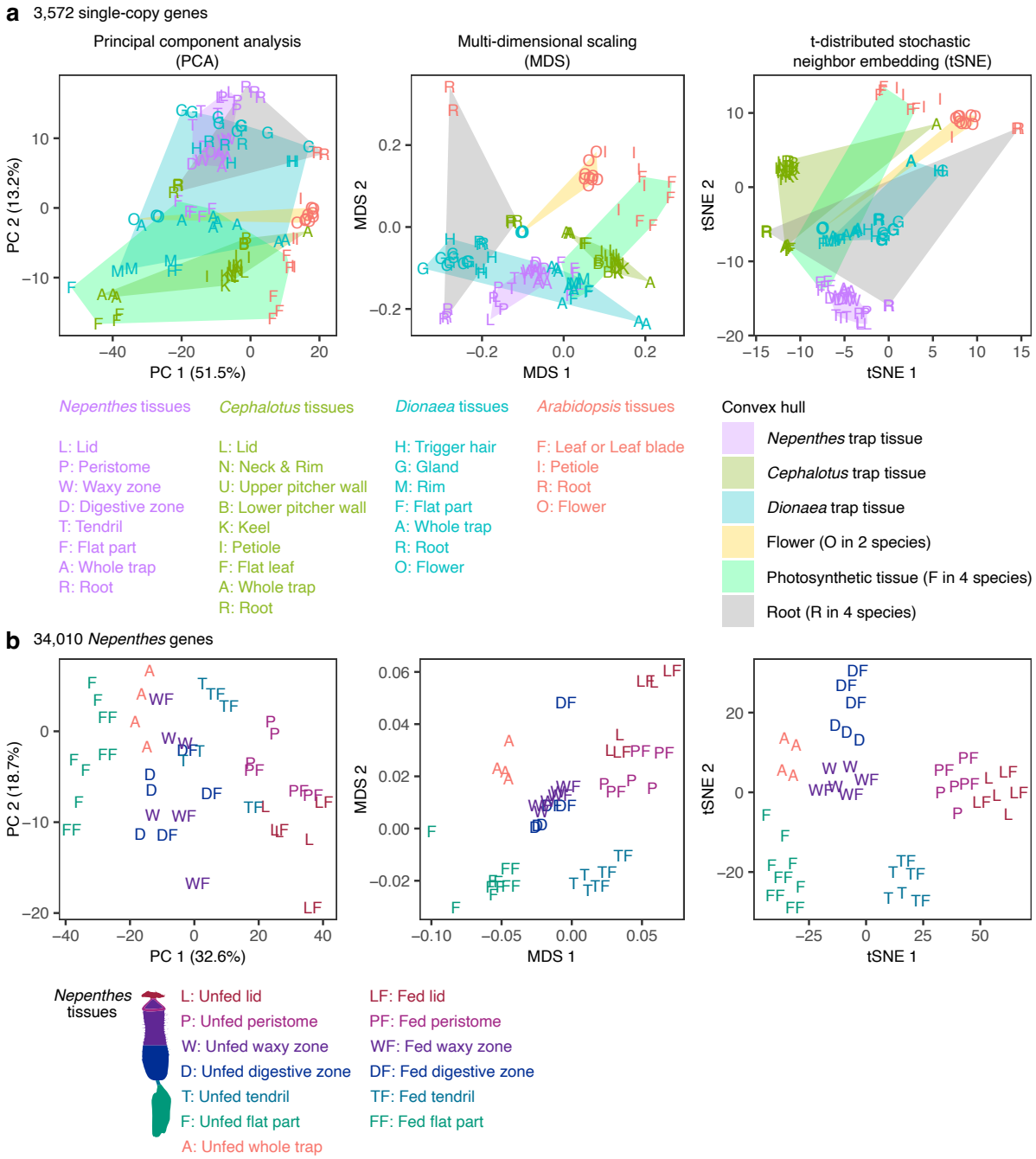
Supplementary Fig. 13. The phylogenetic relationships of *DYT1* and *MMD1*. The orthogroup phylogeny for *DYT1* (**a**) and *MMD1* (**b**) was reconstructed using the maximum-likelihood method in conjunction with phylogeny reconciliation with the species tree (see Methods). We could not obtain gene cluster memberships for species for which only transcriptomes are available: i.e., *Ancistrocladus*, *Drosophyllum*, and *Triphyophyllum*. There may be false negatives in gene cluster detection for species with highly fragmented genome assemblies. Trimmed codon sequence alignments are visualized (black denotes site presence) to indicate fragmented gene models. The bars in the phylogenetic trees represent 0.2 nucleotide substitutions per site.



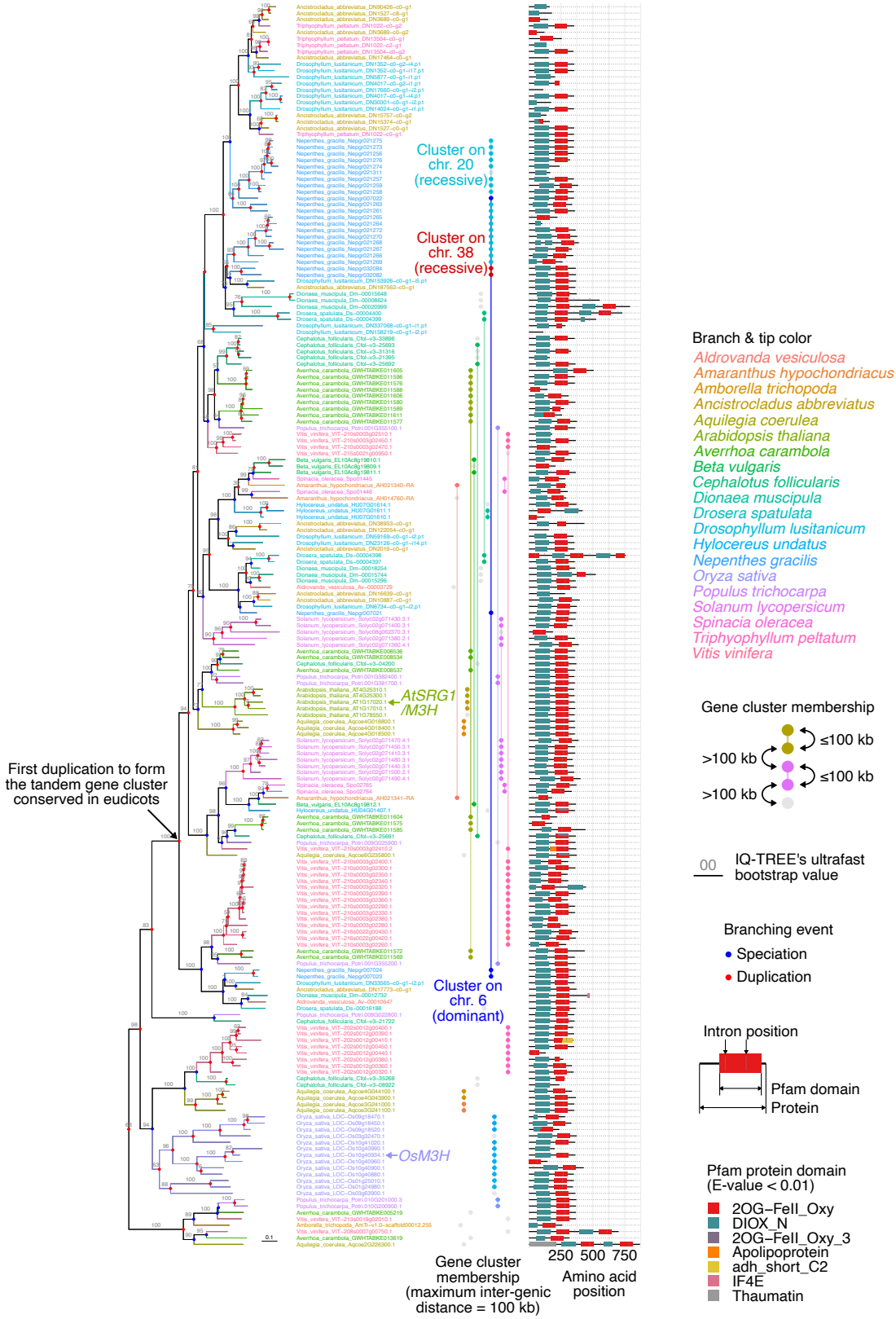
Supplementary Fig. 14. Expression and segregation of male-specific genes. (**a**) The expression levels of male-specific genes in the late developmental stages. RNA-seq reads from flowers in *Nepenthes* cultivars and other species were mapped to the *N. gracilis* male gene models to estimate the transcript abundance as in Fig. 3e. For sample dissection, see Supplementary Fig. 23. (**b**) Phylogenetic identification of *LFY-A* and *LFY-Y* in individually assembled flower transcriptomes. *LFY* transcripts were identified from 17 out of 31 transcriptome assemblies using TBLASTX search with the *N. gracilis* genes as query sequences and an E-value cutoff of 0.01 with a minimum query sequence coverage of BLAST hits of 25%. The amino acid sites where a substitution occurred in the C_LFY_FLO domain of the *N. gracilis* LFY-Y (Supplementary Fig. 15) are shown to the right. Note that the missing amino acid positions represent incomplete assembly of the transcript. The *LFY-A* clade contains transcripts from both male and female individuals (blue and orange, respectively), while that of *LFY-Y* is composed strictly of male transcripts. The bar indicates nucleotide substitutions per nucleotide site. (**c**) Phylogenetic identification of *MMD1* in individually assembled flower transcriptomes. (**d**) Phylogenetic identification of *DYT1* in individually assembled flower transcriptomes. In **c**–**d**, orthologous clades were extracted from the full trees containing paralogs after midpoint rooting of the ML tree constructed with the GTR+G4 substitution model. (**e**) Transcript abundance ratios of gene models in the MSY.



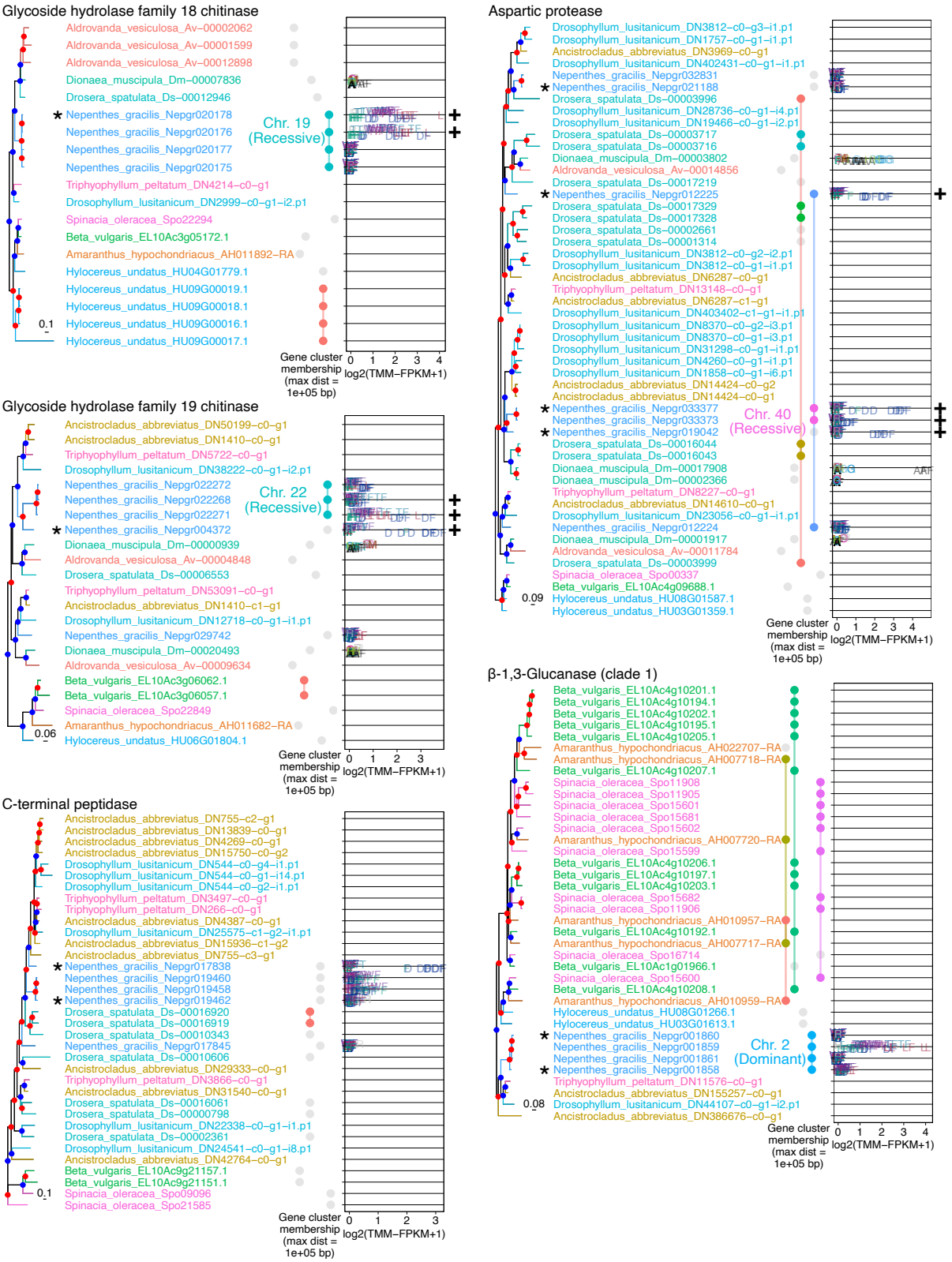
Supplementary Fig. 15. Amino acid substitutions in the conserved C_LFY_FLO domain (pfam17538) of LFY-Y. (**a**) A maximum-likelihood phylogenetic tree of *LFY* homologs. IQ-TREE's ultrafast bootstrap values are shown above branches. A hyphen (-) marks a branch reconciled by GeneRax. Node colors in the trees indicate inferred branching events of speciation (blue) versus gene duplication (red). To the right of the tip labels, numbers of introns in protein-coding sequences (I), Pfam domain structures (E-value < 0.01), and codon alignment structures are shown. Note that intron numbers are not available in *Ancistrocladus*, *Drosophyllum*, and *Triphyophyllum* as their sequences were obtained from transcriptome assemblies. Both *Nepenthes LFY-A* and *LFY-Y* retain two introns, indicating that the gene duplication was DNA-based. (**b**) Protein alignment of the C_LFY_FLO domain. The positions of amino acid substitutions in the branch leading to LFY-Y are indicated. Note that residue H370N is not shared by other *Nepenthes* species analyzed in Supplementary Fig. 14, but the others are. (**c**) Amino acid substitutions in *N. gracilis* LFY-Y (purple) mapped to the structure of *A. thaliana* LFY in complex with DNA (pink) from the APETALA 1 (*AP1*) promoter (PDB ID: 2VY1)[154]. The substitution mapping was performed using CSUBST[116].



Supplementary Fig. 16. Gene expression profiles of carnivorous trap leaves. (**a**) Comparison among carnivorous trap leaves (*Nepenthes*, *Cephalotus*, and *Dionaea*) and *Arabidopsis* leaves. The cross-species TMM-normalized FPKM values of OrthoFinder-based single-copy genes were analyzed. To draw a robust conclusion, three methods (PCA, MDS, and tSNE) were compared. The distribution of *Nepenthes* trap tissues overlaps with that of roots (PCA) and *Dionaea* trap tissues (PCA, and MDS), but not with others, including photosynthetic tissues. (**b**) The effect of feeding treatments in the pitcher tissues of *N. gracilis*.

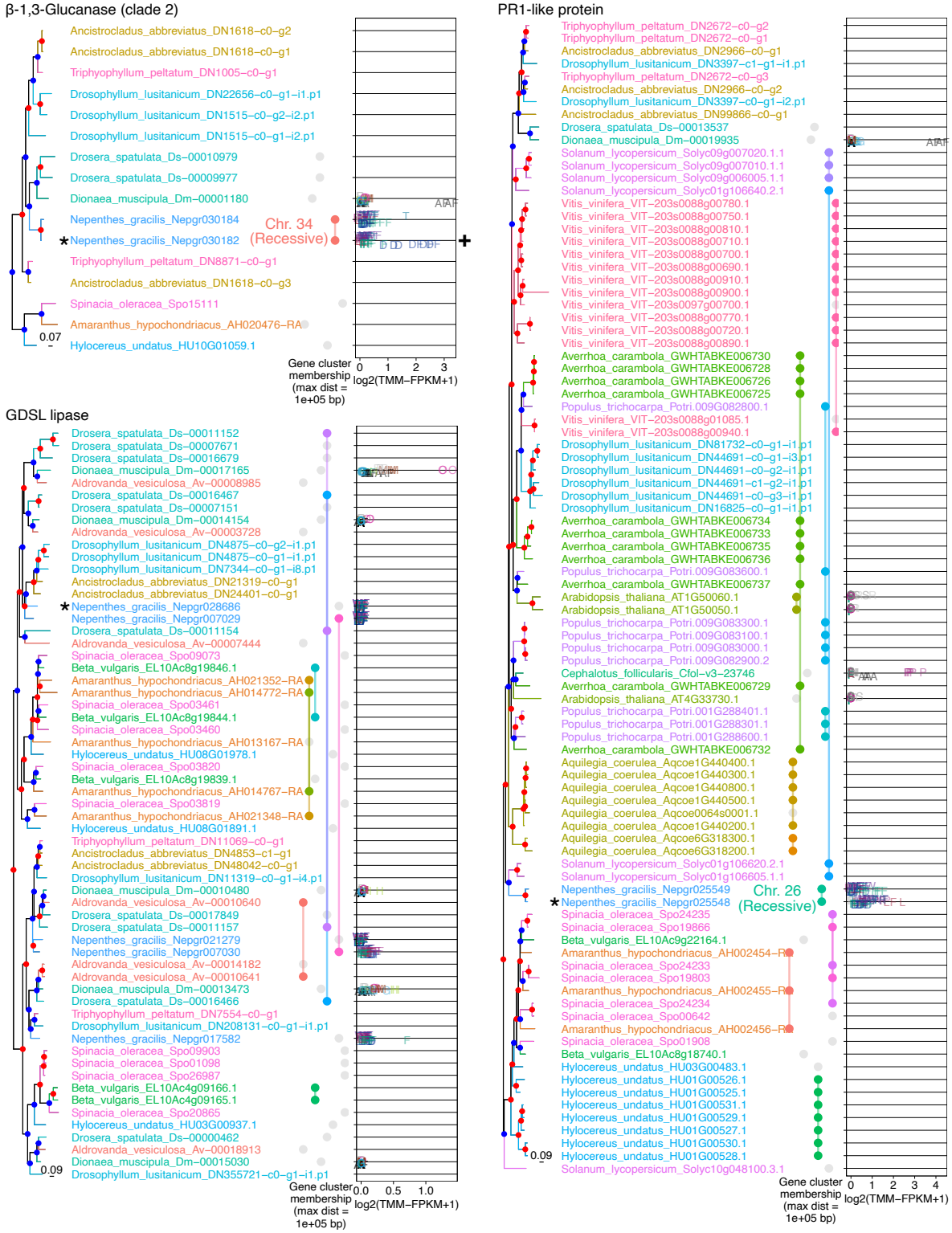


Supplementary Fig. 17. A complete phylogeny of *SRG1*-like genes. Homologous genes were identified by a TBLASTX search with *Arabidopsis* and *Nepenthes* genes as query sequences using an E-value cutoff of 0.01 and a minimum query sequence coverage of BLAST hits of 50%. Phylogenetic analysis was performed as described in Methods. Note that gene cluster memberships could not be obtained for species for which only transcriptomes were available (*Ancistrocladus*, *Drosophyllum*, and *Triphyophyllum*). False negatives of gene cluster detection may occur in species whose genome assemblies are highly fragmented. The bar indicates 0.1 nucleotide substitutions per site.

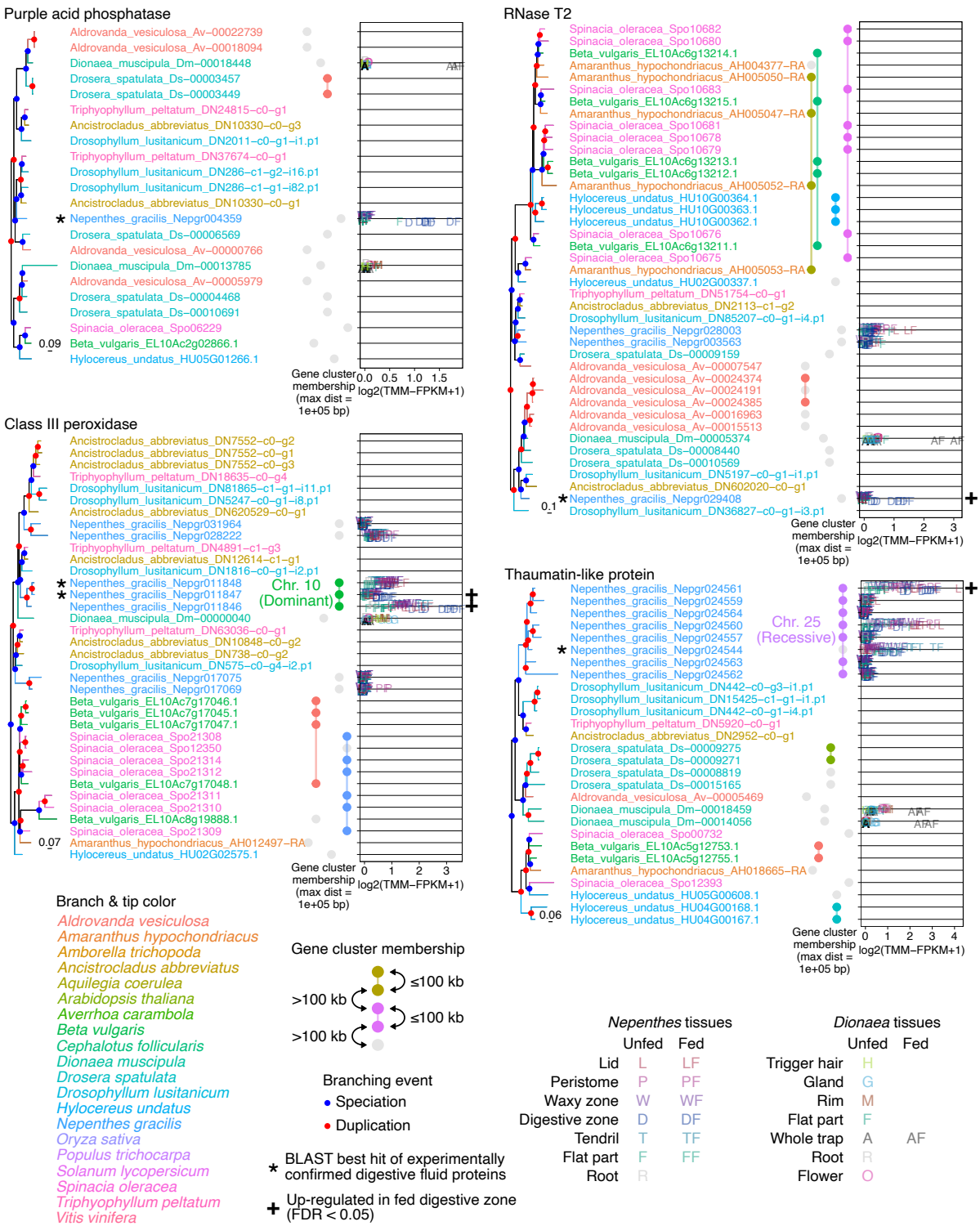


Supplementary Fig. 18. Tandem arrays associated with the genes encoding digestive fluid proteins. The list of experimentally confirmed cases of digestive fluid proteins in *Nepenthes* was obtained from previous work[77], and their orthologs in the *N. gracilis* genome were identified by a TBLASTX search (asterisks; Supplementary Table 18). Orthologous groups of genes were extracted from the full nucleotide ML trees (Supplementary Dataset). The figure displays, from left to right, the phylogenetic tree, gene IDs, gene cluster memberships, and tissue-specific transcript abundance for each protein family. Transcript abundance is shown for *N. gracilis* and *D. muscipula*. The colors used

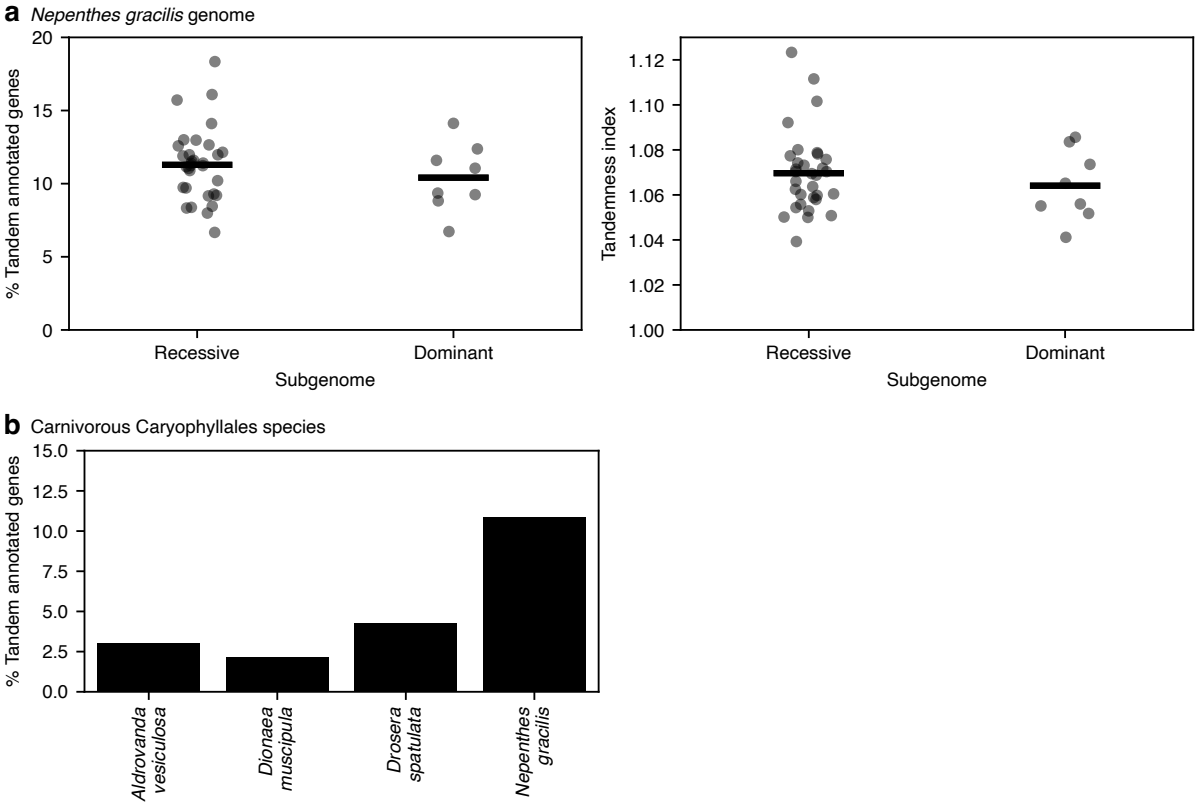
to represent gene cluster memberships are chosen to differentiate the gene clusters and do not correspond to the colors of the species. Gene clusters associated with digestive fluid proteins are labeled with chromosome IDs and subgenome categories (dominant/recessive).



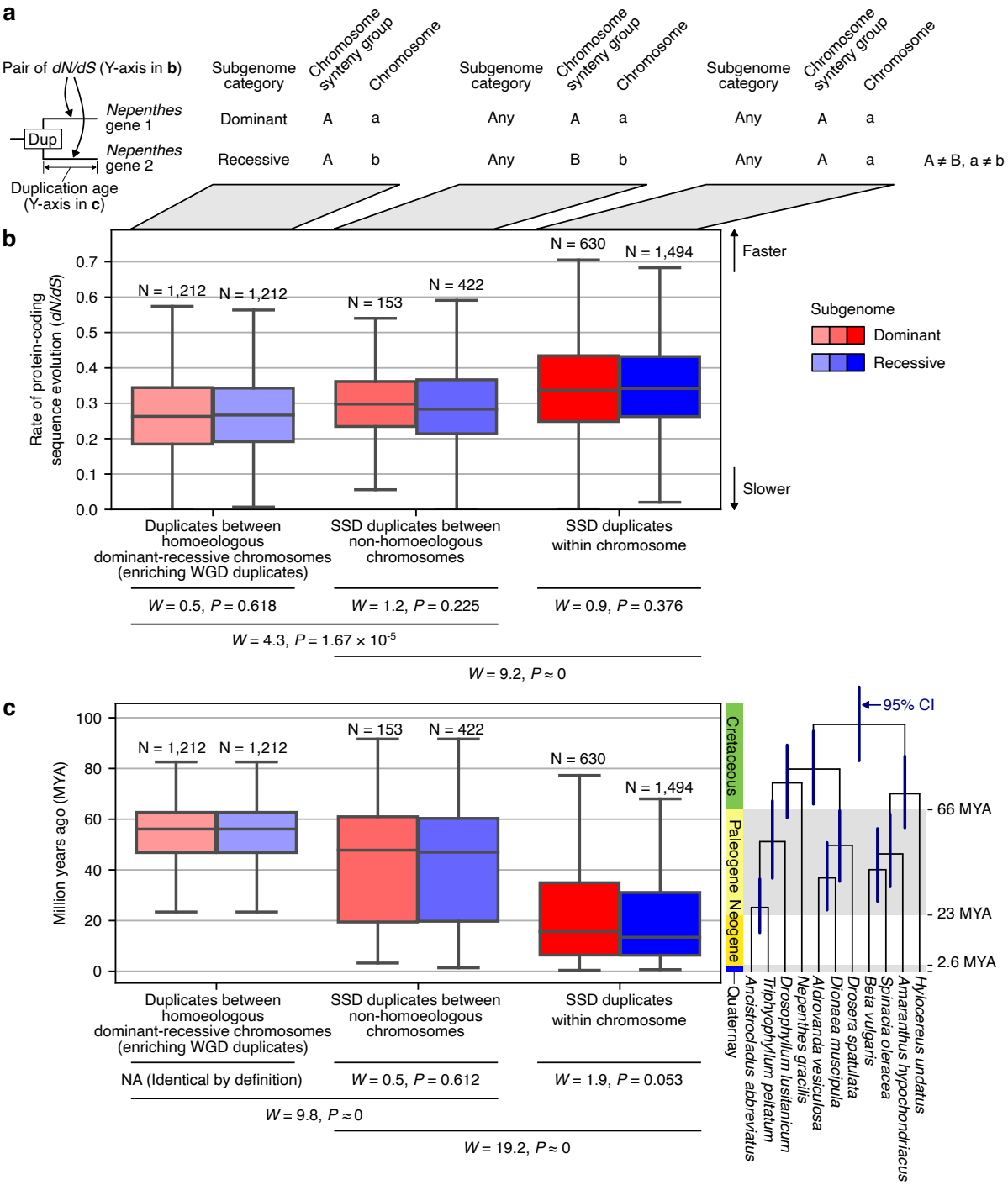
Supplementary Fig. 18 (continued)



Supplementary Fig. 18 (continued)



Supplementary Fig. 19. Genome-wide analysis of gene tandemness. (**a**) Analysis of chromosome-wise gene tandemness in *Nepenthes gracilis* using two different measures. A total of 18,425 *Nepenthes* genes with hits in the UniProt database were analyzed (E-value < 0.01). Tandem arrangements were identified as adjacent genes (omitting gene models with no-hit in UniProt) with the same best-hit UniProt entry. The tandemness index was calculated as the mean number of tandem gene groups (including singletons without a tandem duplicate). Points correspond to chromosomes. The numbers of samples are 32 and 8 for recessive and dominant subgenomes, respectively. Bars indicate mean values. (**b**) Cross-species comparison in the carnivorous lineage of Caryophyllales. Average values across the entire genome are presented. Values vary among species, with *Nepenthes* displaying a relatively higher value compared to the others. However, caution is advised when comparing different genomes. The methodology employed is anticipated to be sensitive to several factors, including genome assembly contiguity, the quality of gene models, and the degree of relatedness to species that are well-represented in the UniProt database.



Supplementary Fig. 20. Evolutionary rates and ages of *Nepenthes*-specific duplicates. (**a**) Design of the analysis. *dN/dS* was calculated for all branches of reconciled nucleotide ML trees for all orthogroups among the 20 genomes using the *mapdNdS* approach[115], and pairs of *N. gracilis* genes forming terminal sister clades in gene trees were extracted and analyzed. Only expressed genes were included in the analysis (maximum TMM-FPKM > 0.1). (**b**) The effect of duplication type and subgenome dominance status on the rate of protein evolution. Colors match those in Fig. 5. (**c**) Age distributions of analyzed duplicates. The result showed that the majority of the analyzed SSDs were younger than the WGDs in *Nepenthes*. The divergence time on gene trees was estimated using RADTE[113]. The species divergence in Caryophyllales, as shown in Fig. 1b, is aligned to the right. The error bars on the internal nodes of the phylogenetic tree represent a 95% confidence interval for the divergence time. In **b** and **c**, box plot elements are defined as follows: center line, median; box limits, upper and lower quartiles; whiskers, 1.5 × interquartile range. The stochastic equality of the data in **b** and **c** was tested by a two-sided Brunner–Munzel test with *W* as the test statistic[155].

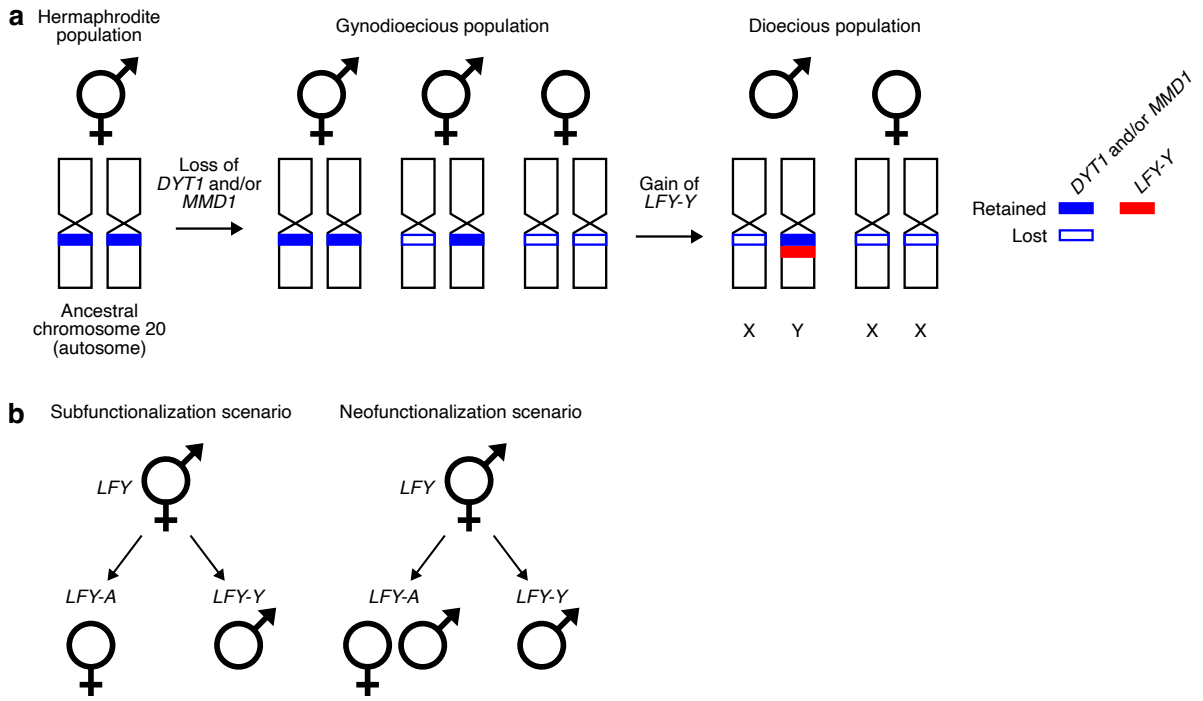
Biological Process	Dominant subgenome	Recessive subgenomes
WGD duplicates	cell differentiation negative regulation of DNA-templated transcription regulation of growth positive regulation of DNA-templated transcription response to cold regulation of DNA-templated transcription regulation of transcription by RNA polymerase II flower development ethylene-activated signaling pathway protein ubiquitination	negative regulation of gibberellic acid mediated signaling pathway mitotic cell cycle phase transition abaxial cell fate specification ethylene-activated signaling pathway response to gibberellin proteasome-mediated ubiquitin-dependent protein catabolic process response to jasmonic acid cell differentiation response to cold protein autophosphorylation negative regulation of DNA-templated transcription intracellular signal transduction root development response to salt stress regulation of transcription by RNA polymerase II response to water deprivation auxin-activated signaling pathway response to auxin response to osmotic stress gibberellic acid mediated signaling pathway positive regulation of DNA-templated transcription abscisic acid-activated signaling pathway protein ubiquitination response to abscisic acid floral meristem determinacy polarity specification of adaxial/abaxial axis regulation of DNA-templated transcription anther development regulation of flower development leaf morphogenesis plant organ development regulation of cyclin-dependent protein serine/threonine kinase activity leaf development response to light stimulus
Tandem duplicates	cell surface receptor signaling pathway response to oomycetes	positive regulation of anaphase-promoting complex-dependent catabolic process monoterpenoid biosynthetic process monoatomic anion transport DNA integration diterpenoid biosynthetic process purine nucleotide transport

Supplementary Fig. 21. Differential GO enrichments of WGD and tandem duplicates in dominant and recessive subgenomes. For complete information, see Supplementary Table 12, Supplementary Table 13, Supplementary Table 14, and Supplementary Table 15. Following gene annotation, we performed gene ontology (GO) enrichment analysis using GOATOOLS v0.9.9[156] to test the overrepresentation of gene ontology terms among duplicated genes in dominant chromosomes and recessive chromosomes. The background consisted of all annotated genes in the genome and the foreground subset consisted of syntenic (polyploid) pairs or tandem duplicates, respectively, for both dominant and recessive chromosomes, using Bonferroni-adjusted $P < 0.05$ as the threshold for significance. Syntenic vs. tandem duplicates were downloaded from a self-vs-self analysis of the *Nepenthes gracilis* genome using CoGe's SynMap tool[94].

	Cellular Component	Dominant subgenome	Recessive subgenomes
WGD duplicates		cytoplasm plasma membrane membrane nucleus cytosol Golgi apparatus	nucleosome Golgi apparatus plant-type vacuole cytoplasm plasma membrane cytosol plasmodesma membrane nucleus protein phosphatase type 2A complex cyclin-dependent protein kinase holoenzyme complex
Tandem duplicates		extracellular space	mitotic checkpoint complex

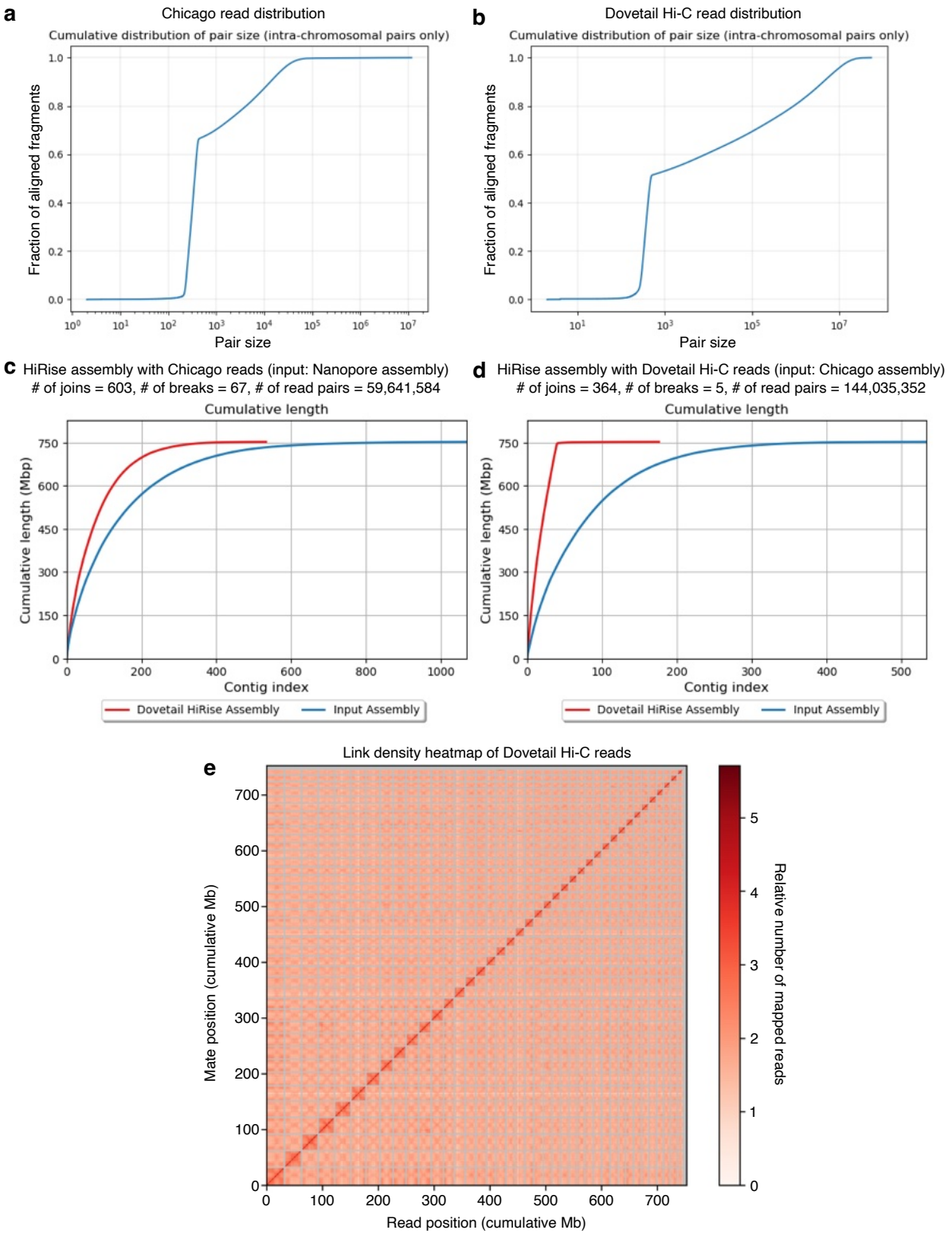
	Molecular Function	Dominant subgenome	Recessive subgenomes
WGD duplicates			minor groove of adenine-thymine-rich DNA binding cyclin-dependent protein serine/threonine kinase regulator activity protein heterodimerization activity ubiquitin-protein transferase activity protein dimerization activity
	RNA polymerase II cis-regulatory region sequence-specific DNA binding		DNA-binding transcription factor activity, RNA polymerase II-specific
	sequence-specific DNA binding		RNA polymerase II cis-regulatory region sequence-specific DNA binding
	DNA-binding transcription factor activity, RNA polymerase II-specific		sequence-specific DNA binding
	protein dimerization activity		protein homodimerization activity
	transcription cis-regulatory region binding		mRNA binding
	DNA binding		calmodulin binding
	DNA-binding transcription factor activity		transcription cis-regulatory region binding
	metal ion binding		DNA binding
	protein serine kinase activity		protein serine kinase activity
	protein serine/threonine/tyrosine kinase activity		protein serine/threonine/tyrosine kinase activity
	protein serine/threonine kinase activity		DNA-binding transcription factor activity
			protein serine/threonine kinase activity
			metal ion binding
			protein serine/threonine phosphatase activity
			identical protein binding
			ATP binding
Tandem duplicates			ubiquitin ligase activator activity geraniol 10-hydroxylase activity mechanosensitive monoatomic ion channel activity ADP binding anaphase-promoting complex binding S-adenosylmethionine-dependent methyltransferase activity manganese ion binding oxidoreductase activity, acting on paired donors, with ...
		raffinose alpha-galactosidase activity	ABC-type transporter activity
		monooxygenase activity	RNA-directed DNA polymerase activity
		polysaccharide binding	quercetin 7-O-glucosyltransferase activity
		iron ion binding	quercetin 3-O-glucosyltransferase activity
		heme binding	monooxygenase activity
		scopoletin glucosyltransferase activity	methyl jasmonate methylesterase activity iron ion binding alcohol dehydrogenase (NAD+) activity oxidoreductase activity, acting on paired donors, with ... endonuclease activity heme binding O-methyltransferase activity serine-type endopeptidase inhibitor activity salicylic acid glucosyltransferase (glucoside-forming) activity

Supplementary Fig. 21 (continued)



Supplementary Fig. 22. A model for the evolution of dioecy in *Nepenthes*. (**a**) A plausible scenario for the transition from hermaphroditic to dioecious populations. (**b**) Functional divergence after gene duplication of *LFY*.

Supplementary Fig. 23. Dissection of *Nepenthes* samples for RNA-seq analysis. (**a**) Inflorescence dissection. Early flower buds included the six youngest flower buds at the tip of developing inflorescences. For the expression analysis (Fig. 3 and Supplementary Fig. 14), early and late flower buds were not distinguished and were instead analyzed together as flower buds. Inflorescences were dissected in the same manner regardless of species and sex. Flowers from several *Nepenthes* species were utilized in this study (Supplementary Table 4), with *N. robcantleyi* presented here as a representative example. (**b**) Pitcher dissection. The transitional region between the digestive zone and the tendril was carefully removed to prevent cross-contamination of tissues.

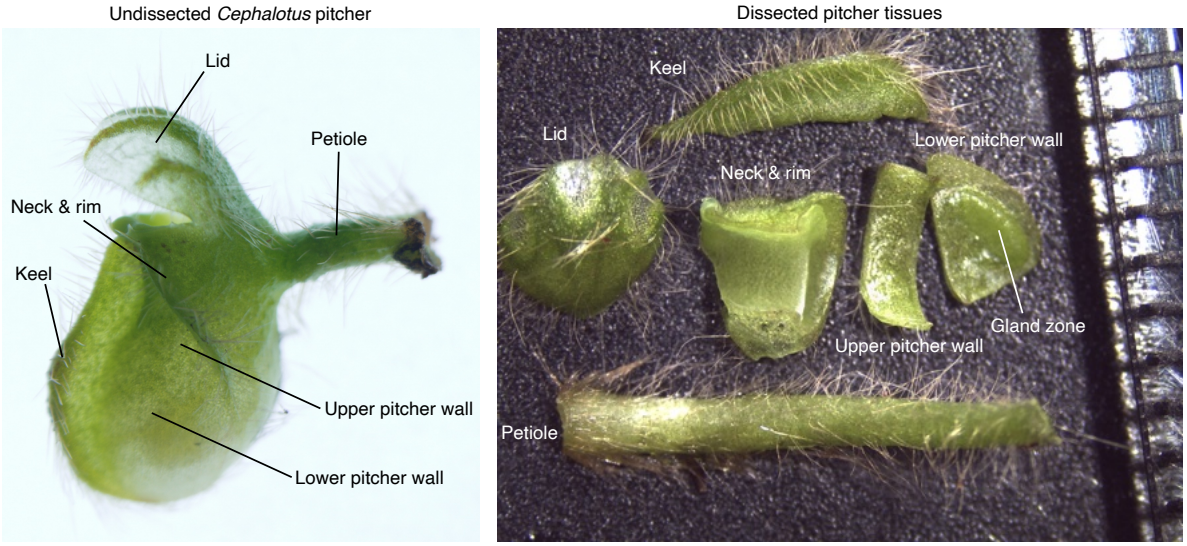


Supplementary Fig. 24. Hi-C assembly of the *N. gracilis* male genome. (**a-b**) The distributions of read pair sizes in Chicago (**a**) and Dovetail Hi-C (**b**) libraries. (**c-d**) Improved contiguity of the genome assembly using Chicago (**c**) and Dovetail Hi-C (**d**) reads. (**e**) The link density heatmap of the Dovetail Hi-C reads. The X-axis and Y-axis represent the mapping positions of the first and second reads in read pairs, respectively. Gray lines indicate the boundaries between scaffolds. The value range of heatmap colors for the number of mapped reads was set to 0–351 in Juicebox (https://github.com/aidenlab/Juicebox).

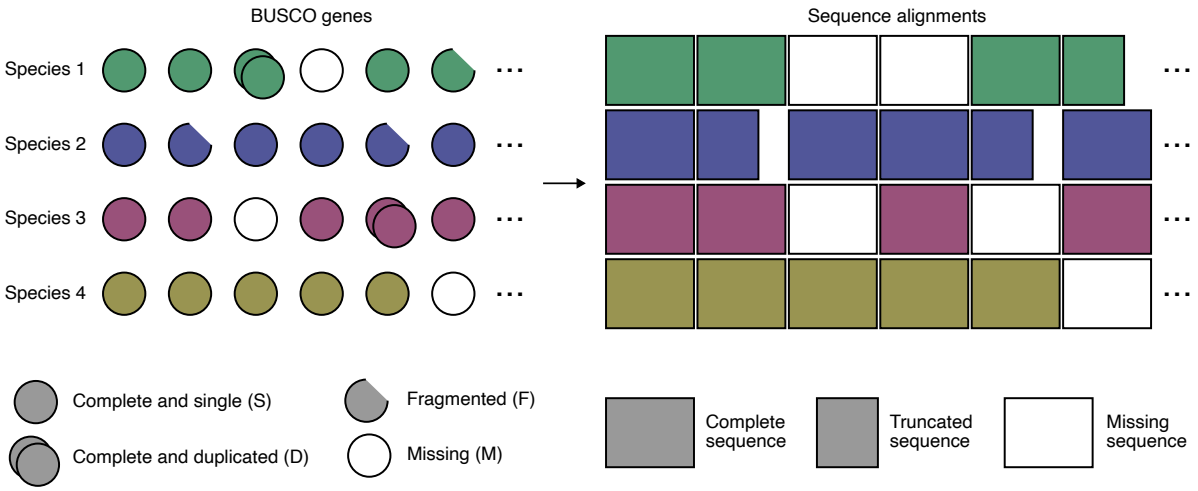
Supplementary Fig. 25. ***Ancistrocladus abbreviatus*** **samples used for RNA-seq analysis.** The smallest increment of the ruler is 1 mm.

Supplementary Fig. 26. *Triphyophyllum peltatum* samples used for RNA-seq analysis. Juvenile non-trapping leaves were harvested from axenic cultures.

Supplementary Fig. 27. *Drosophyllum lusitanicum* samples used for RNA-seq analysis. The smallest increment of the ruler is 1 mm.



Supplementary Fig. 28. Dissection of pitcher leaves in *Cephalotus follicularis* for RNA-seq analysis. The smallest increment of the ruler is 1 mm.



Supplementary Fig. 29. Procedure for generating single-copy gene alignments. All detected BUSCO genes marked as single-copy (S) or fragmented (F) were extracted, while those marked as duplicated (D) or missing (M) were treated as missing data.

Supplementary References

118. Brockington, S. F. *et al.* Phylogeny of the Caryophyllales *sensu lato*: Revisiting hypotheses on pollination biology and perianth differentiation in the core Caryophyllales. *Int. J. Plant Sci.* **170**, 627–643 (2009).
119. Schäferhoff, B., Müller, K. F. & Borsch, T. *Caryophyllales* phylogenetics: disentangling *Phytolaccaceae* and *Molluginaceae* and description of *Microteaceae* as a new isolated family. *Willdenowia* **39**, 209–228 (2010).
120. Yao, G. *et al.* Plastid phylogenomic insights into the evolution of Caryophyllales. *Mol. Phylogenet. Evol.* **134**, 74–86 (2019).
121. Meimberg, H., Dittrich, P., Bringmann, G., Schlauer, J. & Heubl, G. Molecular phylogeny of Caryophyllidae s.l. based on matK sequences with special emphasis on carnivorous taxa. *Plant Biol.* **2**, 218–228 (2000).
122. Cameron, K. M., Wurdack, K. J. & Jobson, R. W. Molecular evidence for the common origin of snap-traps among carnivorous plants. *Am. J. Bot.* **89**, 1503–1509 (2002).
123. Walker, J. F. *et al.* From cacti to carnivores: Improved phylotranscriptomic sampling and hierarchical homology inference provide further insight into the evolution of Caryophyllales. *Am. J. Bot.* **105**, 446–462 (2018).
124. Thomas, G. W. C., Ather, S. H. & Hahn, M. W. Gene-tree reconciliation with MUL-trees to resolve polyploidy events. *Syst. Biol.* **66**, 1007–1018 (2017).
125. Chen, K., Durand, D. & Farach-Colton, M. NOTUNG: a program for dating gene duplications and optimizing gene family trees. *J. Comput. Biol.* **7**, 429–447 (2000).
126. Solís-Lemus, C. & Ané, C. Inferring phylogenetic networks with maximum pseudolikelihood under incomplete lineage sorting. *PLOS Genet.* **12**, e1005896 (2016).
127. Cai, R. & Ané, C. Assessing the fit of the multi-species network coalescent to multi-locus data. *Bioinformatics* **37**, 634–641 (2021).
128. Bird, K. A., VanBuren, R., Puzey, J. R. & Edger, P. P. The causes and consequences of subgenome dominance in hybrids and recent polyploids. *New Phytol.* **220**, 87–93 (2018).
129. Blázquez, M. A., Soowal, L. N., Lee, I. & Weigel, D. *LEAFY* expression and flower initiation in *Arabidopsis*. *Development* **124**, 3835–3844 (1997).
130. Shu, G., Amaral, W., Hileman, L. C. & Baum, D. A. *LEAFY* and the evolution of rosette flowering in violet cress (*Jonopsidium acaule*, Brassicaceae). *Am. J. Bot.* **87**, 634–641 (2000).
131. Hofer, J. *et al. UNIFOLIATA* regulates leaf and flower morphogenesis in pea. *Curr. Biol.* **7**, 581–587 (1997).
132. Wang, Z. J., Huang, J. Q., Huang, Y. J., Chen, F. F. & Zheng, B. S. Cloning and characterization of a homologue of the *FLORICAULA/LEAFY* gene in hickory (*Carya cathayensis* Sarg). *Plant Mol. Biol. Report.* **30**, 794–805 (2012).
133. Carmona, M. J., Cubas, P. & Martínez-Zapater, J. M. *VFL*, the grapevine *FLORICAULA/LEAFY* ortholog, is expressed in meristematic regions independently of their fate. *Plant Physiol.* **130**, 68–77 (2002).
134. Coen, E. S. *et al. floricaula*: A homeotic gene required for flower development in *Antirrhinum majus*. *Cell* **63**, 1311–1322 (1990).
135. Kelly, A. J., Bonnlander, M. B. & Meeks-Wagner, D. R. *NFL*, the tobacco homolog of *FLORICAULA* and *LEAFY*, is transcriptionally expressed in both vegetative and floral meristems. *Plant Cell* **7**, 225–234 (1995).
136. Molinero-Rosales, N. *et al. FALSIFLORA*, the tomato orthologue of *FLORICAULA* and *LEAFY*, controls flowering time and floral meristem identity. *Plant J.* **20**, 685–693 (1999).
137. Ma, Y.-P., Fang, X.-H., Chen, F. & Dai, S.-L. *DFL*, a *FLORICAULA/LEAFY* homologue gene from *Dendranthema lavandulifolium* is expressed both in the vegetative and reproductive tissues. *Plant Cell Rep.* **27**, 647–654 (2008).
138. Busch, A. & Gleissberg, S. *EcFLO*, a *FLORICAULA*-like gene from *Eschscholzia californica* is expressed during organogenesis at the vegetative shoot apex. *Planta* **217**, 841–848 (2003).
139. Zhang, J.-X., Wu, K.-L., Zeng, S.-J., Duan, J. & Tian, L.-N. Characterization and expression analysis of *PhalLFY*, a homologue in Phalaenopsis of *FLORICAULA/LEAFY* genes. *Sci. Hortic.* **124**, 482–489 (2010).

140. Bomblies, K. *et al.* Duplicate *FLORICAULA/LEAFY* homologs *zfl1* and *zfl2* control inflorescence architecture and flower patterning in maize. *Development* **130**, 2385–2395 (2003).
141. Wang, C.-N., Möller, M. & Cronk, Q. C. B. Altered expression of *GFLO*, the Gesneriaceae homologue of *FLORICAULA/LEAFY*, is associated with the transition to bulbil formation in *Titanotrichum oldhamii*. *Dev. Genes Evol.* **214**, 122–127 (2004).
142. Yu, Q., Moore, P. H., Albert, H. H., Roader, A. H. K. & Ming, R. Cloning and characterization of a *FLORICAULA/LEAFY* ortholog, *PFL*, in polygamous papaya. *Cell Res.* **15**, 576–584 (2005).
143. Kazama, Y. *et al.* A *CLAVATA3*-like gene acts as a gynoecium suppression function in white campion. *Mol. Biol. Evol.* **39**, msac195 (2022).
144. Akagi, T. *et al.* A Y-encoded suppressor of feminization arose via lineage-specific duplication of a cytokinin response regulator in kiwifruit. *Plant Cell* **30**, 780–795 (2018).
145. Akagi, T. *et al.* Two Y-chromosome-encoded genes determine sex in kiwifruit. *Nat. Plants* **5**, 801–809 (2019).
146. Käfer, J. *et al.* Dioecy is associated with higher diversification rates in flowering plants. *J. Evol. Biol.* **27**, 1478–1490 (2014).
147. Ma, X. *et al.* The spinach YY genome reveals sex chromosome evolution, domestication, and introgression history of the species. *Genome Biol.* **23**, 75 (2022).
148. Ranallo-Benavidez, T. R., Jaron, K. S. & Schatz, M. C. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat. Commun.* **11**, 1432 (2020).
149. Lledo, M. D., Crespo, M. B., Cameron, K. M., Fay, M. F. & Chase, M. W. Systematics of Plumbaginaceae based upon cladistic analysis of *rbcL* sequence data. *Syst. Bot.* **23**, 21–29 (1998).
150. Crawley, S. S. & Hilu, K. W. Impact of missing data, gene choice, and taxon sampling on phylogenetic reconstruction: the Caryophyllales (angiosperms). *Plant Syst. Evol.* **298**, 297–312 (2012).
151. Crawley, S. S. & Hilu, K. W. Caryophyllales: Evaluating phylogenetic signal in *trnK* intron versus *matK*. *J. Syst. Evol.* **50**, 387–410 (2012).
152. Zwaenepoel, A., Li, Z., Lohaus, R. & Peer, Y. V. de. Finding evidence for whole genome duplications: A reappraisal. *Mol. Plant* **12**, 133–136 (2019).
153. Ou, S. *et al.* Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biol.* **20**, 275 (2019).
154. Hamès, C. *et al.* Structural basis for LEAFY floral switch function and similarity with helix-turn-helix proteins. *EMBO J.* **27**, 2628–2637 (2008).
155. Brunner, E. & Munzel, U. The nonparametric Behrens-Fisher problem: asymptotic theory and a small-sample approximation. *Biom. J.* **42**, 17–25 (2000).
156. Klopfenstein, D. V. *et al.* GOATOOLS: A Python library for Gene Ontology analyses. *Sci. Rep.* **8**, 10872 (2018).