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*

Received: 14 June 2023

Accepted: 9 October 2023

Published online: 23 November 2023

Check for updates

Franziska Saul[1,14], Mathias Scharmann[2,14], Takanori Wakatake[1], Sitaram Rajaraman[3,4], André Marques[5], Matthias Freund[1], Gerhard Bringmann[6], Louisa Channon[1], Dirk Becker[1], Emily Carroll[7], Yee Wen Low[8], Charlotte Lindqvist[7], Kadeem J. Gilbert[9], Tanya Renner[10], Sachiko Masuda[11], Michaela Richter[7], Gerd Vogg[12], Ken Shirasu[11], Todd P. Michael[13], Rainer Hedrich[1], Victor A. Albert[7] ✉ & Kenji Fukushima[1] ✉

Subgenome dominance after whole-genome duplication generates distinction in gene number and expression at the level of chromosome sets, but it remains unclear how this process may be involved in evolutionary novelty. Here we generated a chromosome-scale genome assembly of the Asian pitcher plant *Nepenthes gracilis* to analyse how its novel traits (dioecy and carnivorous pitcher leaves) are linked to genomic evolution. We found a decaploid karyotype and a clear indication of subgenome dominance. A male-linked and pericentromerically located region on the putative sex chromosome was identified in a recessive subgenome and was found to harbour three transcription factors involved in flower and pollen development, including a likely neofunct ionalized LEAFY duplicate. Transcriptomic and syntenic analyses of carnivory-related genes suggested that the paleopolyploidization events seeded genes that subsequently formed tandem clusters in recessive subgenomes with specific expression in the digestive zone of the pitcher, where specialized cells digest prey and absorb derived nutrients. A genome-scale analysis suggested that subgenome dominance likely contributed to evolutionary innovation by permitting recessive subgenomes to diversify functions of novel tissue-specific duplicates. Our results provide insight into how polyploidy can give rise to novel traits in divergent and successful high-ploidy lineages.

Novel phenotypes often arise from the emergence of new genes that have undergone functional divergence after gene duplication. Polyploidy, which involves the duplication of entire genomes, has been argued to have great potential to drive adaptive evolution through the emergence of novel phenotypes[1]. Although polyploid lineages are sometimes perceived as evolutionary dead ends, characterized by short-term adaptive potential and high long-term extinction rates[2], there are notable exceptions to this trend. For instance, flowering plants (angiosperms)[3,4] and

A full list of affiliations appears at the end of the paper. ✉e-mail: vaalbert@buffalo.edu; kenji.fukushima@uni-wuerzburg.de

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

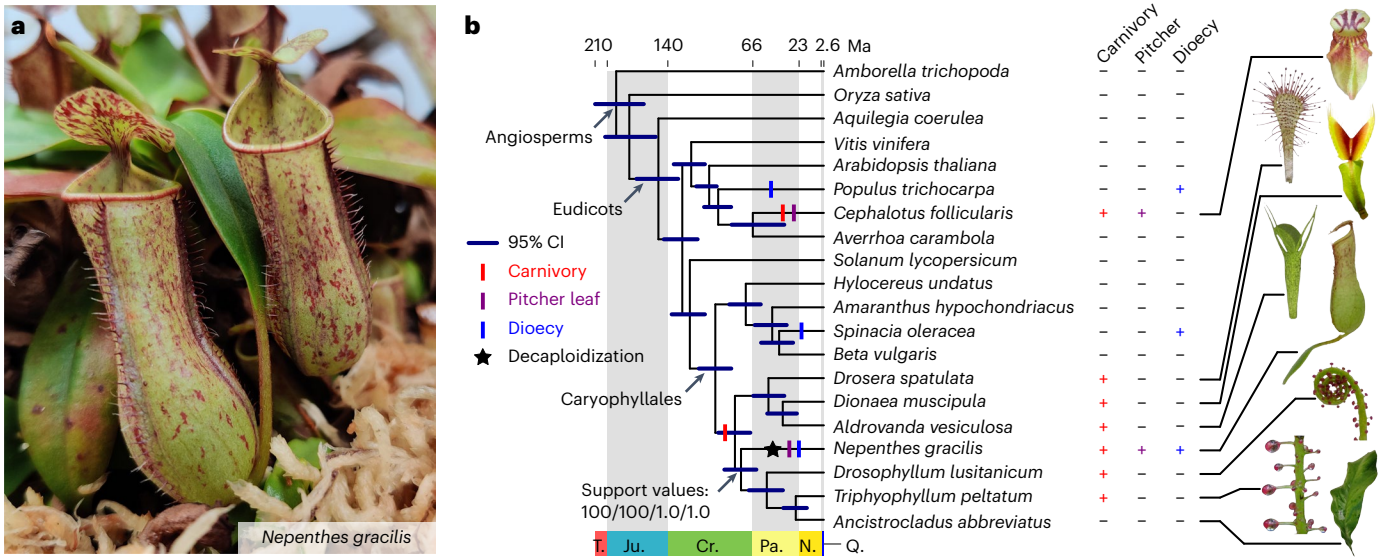


Fig. 1 | Evolution of novel traits in *Nepenthes*. **a**, *N. gracilis* plant with carnivorous pitcher leaves. **b**, The phylogenetic position of *Nepenthes*. Divergence time with a 95% CI is shown for the ML tree topology reconstructed using 1,614 single-copy protein sequences from 20 species (Supplementary Table 6). See Supplementary Fig. 3 for results with alternative phylogenetic methods, including a coalescence-based (CO) species tree approach. Bootstrap supports and posterior probability values for the position of *Nepenthes* are shown as follows: nucleotide ML/protein ML/nucleotide CO/protein CO. Character evolution was parsimoniously mapped to branches, while symbols do not indicate point estimates of evolutionary origin times. Note that carnivory was secondarily lost in *Ancistrocladus*[20]. Caricatures of leaves of plants belonging to carnivorous clades are shown to the right. CI, confidence interval; T., Triassic; Ju., Jurassic; Cr., Cretaceous; Pa., Paleogene; N., Neogene; Q., Quaternary.

vertebrates[5] have thrived and greatly diversified following multiple rounds of polyploidization early in their evolutionary history.

Following polyploidization, particularly when it occurs after a hybridization event[6,7], the coexistence of more than one haploid chromosome set often engenders subgenome dominance, which is initially characterized by the preferential expression of genes on one chromosome set over the other[8]. These early changes, as well as the subsequent differential retention of duplicated genes on dominant versus recessive subgenomes, can render versatile adaptive prospects among subgenomes over evolutionary time. Although such a genome-wide effect may constrain the long-term evolutionary paths of a large number of genes, it remains unclear how subgenome dominance may be involved in the evolution of novel gene functions. To study the long-term effects of subgenome dominance, it is necessary to examine genomes that include relatively ancient polyploidization events, yet also having well-preserved subgenome structures.

The Old World pitcher plant *Nepenthes* offers a unique opportunity to study the relationship between polyploidization-related genomic features and the evolution of novel genes. This genus of carnivorous plants includes at least 169 species of perennial vines[9] and is known for its highly modified leaves (pitchers) adapted for trapping and obtaining nutrients from arthropod prey[10]. In addition to carnivory, *Nepenthes* species possess other atypical traits, such as uniform dioecy across the genus. Less than 1% of all angiosperms are carnivorous[9], and approximately 6% of angiosperms are dioecious, possessing distinct male and female individuals[11]. No other carnivorous plant species outside of *Nepenthes* is known to be dioecious, rendering this lineage a rare and unique case. While it became evident from transcriptomic signatures that *Nepenthes* underwent ancient whole-genome duplications (WGDs)[12–14], the role that these genomic events played in the evolution of novel gene functions, many of which may underpin the aforementioned traits, remains unclear.

In this study, we present a chromosome-scale genome assembly of *Nepenthes gracilis* and demonstrate that the *Nepenthes* genome is decaploid, bearing five subgenomes and a clear signature of subgenome dominance. Our analyses suggest that recessive subgenomes played a crucial role in permitting the adaptive evolution of tissue-specific genes associated with the unique biology of *Nepenthes*, and we discuss how gene divergence occurs under the influence of WGDs and subgenome dominance as well as subsequent small-scale duplications (SSDs).

Results

Genome assembly and annotation

We used a combination of Oxford Nanopore Technology (ONT) long-read and Illumina short-read sequencing (Supplementary Table 1) to generate megabase-scale genome assemblies for male and female *N. gracilis* plants (Fig. 1a and Supplementary Table 2), whose *k*-mer-based genome size estimates were approximately 722 Mb (Supplementary Fig. 1). The male contigs were further improved through Hi-C scaffolding to generate a 752.9 Mb genome assembly containing 40 chromosome-scale scaffolds (N50 = 18.6 Mb), which accounted for 99.2% (746.7 Mb) of the total assembly size (Supplementary Fig. 2), consistent with the reported number of chromosomes in this species[15]. Unless otherwise noted, we report analyses of the male Hi-C assembly in this work. Repetitive elements accounted for 67.17% of the reference genome (Supplementary Table 3). Using RNA sequencing (RNA-seq) data (Supplementary Table 4), we predicted a total of 34,010 gene models (Supplementary Table 5) with a gene-space completeness score of 94.1% (1,519/1,614 genes in the embryophyta_odb10 dataset) as determined by Benchmarking Universal Single-Copy Orthologs (BUSCO)[16] (Supplementary Table 6).

Character evolution under a robust phylogeny

The phylogenetic position of *Nepenthes* and its monogeneric family, Nepenthaceae, has been a matter of debate, even with the use of improved gene-tree mining from transcriptomic data. Previous phylogenetic hypotheses have placed Nepenthaceae as sister (1) to Droseraceae, (2) to a clade of three families (Ancistrocladaceae, Dioncophyllaceae and Drosophyllaceae; ADD families) or (3) to a clade containing all four of these plant families (Supplementary Fig. 3a–c).

To more accurately determine the phylogenetic position of Nepenthaceae, we performed phylogenomic analyses using gene models mined from our newly sequenced *Nepenthes* genome. We also sequenced transcriptomes of the ADD families, for which genome sequences are currently unavailable (Supplementary Table 6). Our analysis, using 1,614 Embryophyta-wide single-copy BUSCO genes (Supplementary Fig. 3d), resulted in a single tree topology, regardless of the inference method (maximum likelihood (ML)[17] or coalescence based[18]) or substitution model used (GTR+R4 for nucleotides or LG+R4 for amino acids) (Supplementary Fig. 3e). This topology, limited of course to a bifurcating evolutionary history (see Supplementary Text 1 for the possibility of ancient hybridization events), indicated that plants in this lineage likely split in the sequential order of Droseraceae, Nepenthaceae and then the rest. Under parsimony, these phylogenetic relationships support the hypothesis that carnivory evolved only once in Caryophyllales and that the pitfall-type trap leaves of Nepenthaceae were ultimately derived from flypaper-type trap leaves[19–21], which are found in Droseraceae, Drosophyllaceae and Dioncophyllaceae (Fig. 1b). Dioecy is another fascinating character evolved uniquely in Nepenthaceae after its split from the other lineages. Despite the rapid evolution of the sexual system in many flowering plant lineages, even within single species[22], dioecy is consistently maintained across all species of *Nepenthes*[23].

Decaploidal origin of the *Nepenthes* genome

The large chromosome number in *Nepenthes* ($n = 40$)[15] could suggest a history of WGDs. Although ancient WGDs have been reported in *Nepenthes* and related lineages based on synonymous substitution plots of duplicate gene pairs[12–14], the exact ploidy level and its impact on *Nepenthes* genome structure have not yet been determined. To investigate polyploidization history in *Nepenthes*, we analysed patterns of internal synteny, revealing that the haploid-level *Nepenthes* genome has a clear structure of eight syntenic groups, each containing exactly five chromosomes (Fig. 2a), indicating a decaploidal (five-subgenome) origin with a basic chromosome number of eight (that is, $x = 8$). This finding rejects previous hypotheses of an octoploidal or hexadecaploidal (four- or eight-subgenome) origin with a basic chromosome number of five or ten (that is, $x = 5$ or $x = 10$)[15]. The distribution of synonymous distances between paralog pairs (Supplementary Fig. 4), coupled with clear 4:1 patterns of fractionation bias among homoeologous chromosomes (Supplementary Fig. 5), indicates that *Nepenthes* has a complex polyploid background (Supplementary Text 2), having undergone at least two sequential WGDs following the gamma hexaploidization event that occurred in the common ancestor of all extant core eudicots[24]. Given that the chromosome number is stable in *Nepenthes*, including in the earliest diverged species *Nepenthes pervillei*[15], decaploidy must have been established before the diversification of extant species (6.4–18.2 Ma (million years ago)[25]). Our phylogenomic analysis, which distinctly considered the dominant versus recessive subgenomes (see below for identification), indicated an ancient separation between the donor of the dominant subgenome and the donor(s) of the recessive subgenomes (Supplementary Fig. 6). This separation likely occurred around the time of their divergence from the ADD stem lineage and the Droseraceae stem lineage. Given these findings, it is conceivable that these ancient WGDs could have played a role in the evolution of novel traits shared by all known *Nepenthes* species, including their dioecy and carnivorous pitcher leaves.

Biased gene fractionation and subgenome dominance

Despite having a complete decaploid karyotype ($n = 5x = 40$ and $2n = 10x = 80$), the gene content in *Nepenthes* is highly fractionated, as was evident in syntenic comparisons with grape (*Vitis vinifera*), a species that has maintained its ploidy level since the gamma hexaploidization event (that is, $1x$ in *Vitis* versus $5x$ in *Nepenthes*)[24] (Fig. 2b). In total, 94.5% (5,002/5,293) of 100-gene genomic windows in the *Nepenthes* genome retained fewer than 1.5 syntelog copies on average (compared to the ploidy-based expectation of 5.0 copies). While fractionation was observed to be advanced overall, there was significant heterogeneity among syntenic chromosome groups (Fig. 2b). For example, the syntenic group corresponding to grape chromosome 6 contained an average of 1.4 syntelog copies, while the group corresponding to grape chromosome 16 had an average of only 1.1 copies. Homoeologous *Nepenthes* chromosome groups that are highly fractionated tended to show a clear distinction between one dominant, gene-rich chromosome and four recessive, gene-poor chromosomes (that is, fractionation bias[26]). This distinction was less obvious among chromosome groups that are less fractionated (Fig. 2c and Supplementary Fig. 5). In fact, the degree of fractionation bias significantly correlated with the degree of diploidization (Fig. 2d). The dominant subgenome carried 56% of the $1x$ equivalent syntelog set (average of 5,293 sliding 100-gene windows), and every chromosome in the recessive subgenomes had at least 46 single-copy syntelogs (that is, not detected in the other chromosomes) (Supplementary Fig. 7). Moreover, dominant chromosomes tended to be larger in assembled scaffold sizes (Fig. 2e), potentially reflecting differential rates of genomic DNA excision[27]. These results suggest that the establishment of dominant versus recessive subgenomes played a crucial role in efficient gene fractionation and in the differentiation of chromosomal sizes.

In line with subgenome fractionation dominance, genes in the dominant subgenome tended to show higher expression levels compared with corresponding syntelogs in recessive subgenomes (Fig. 2f). Thus, the dominant or recessive distinction was clear in both gene retention and gene expression differences. Although the WGDs in *Nepenthes* are too ancient to be analysed for any potential initial genomic shock, the differential transposable element density between homoeologous chromosomes, coupled with corresponding DNA methylation levels, might have contributed to establishing subgenomic expression dominance immediately after the WGDs, as is often suggested to occur in recent polyploids[28]. It is noteworthy that, despite the clear chromosome-level expression dominance, 47% (972/2,055) of syntelog pairs showed higher expression levels in gene copies on the recessive subgenomes (Fig. 2g), showing a significant functional contribution from these subgenomes as well (Supplementary Text 3).

The odd-numbered subgenomic structure of *Nepenthes*, along with its clear 1:4 subgenome dominance, strongly suggests an allopolyploid evolutionary history[28]. Our attempts to use subgenome-specific *k*-mers identified by SubPhaser, which has successfully phased the subgenomes of many well-known allopolyploids[29], did not result in complete subgenome phasing in *Nepenthes* (Supplementary Fig. 8). Disentangling exact polyploidization history can be difficult, especially in high-ploidy lineages (for example, in decaploid cotton[30]). Chromosome-scale syntenic comparisons between *Nepenthes* and related lineages, once they are available, will be necessary to discern among alternative allopolyploidization scenarios.

Sex chromosome evolution

Flowering plants in general produce bisexual, hermaphroditic flowers with both male and female reproductive functions. In contrast to their closest relatives and most other angiosperms, *Nepenthes* plants develop unisexual male or female flowers on separate individuals (Fig. 1b). This sexual system is termed dioecy (in zoology: gonochorism) and typically has a genetic basis, often involving sex chromosomes. The many evolutionary origins of dioecy and sex chromosomes in diverse organisms[31] hint towards fundamentally shared principles, yet general understanding is hindered by the lack of details from independent lineages. Scharmann et al.[23] showed that an XY-type sex determination system operates in *Nepenthes*, but the sex chromosome has so far remained unidentified. To locate the male-specific region of the Y chromosome (MSY) in our male *N. gracilis* genome assembly, we mapped sequencing reads of double-digest restriction-site-associated

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

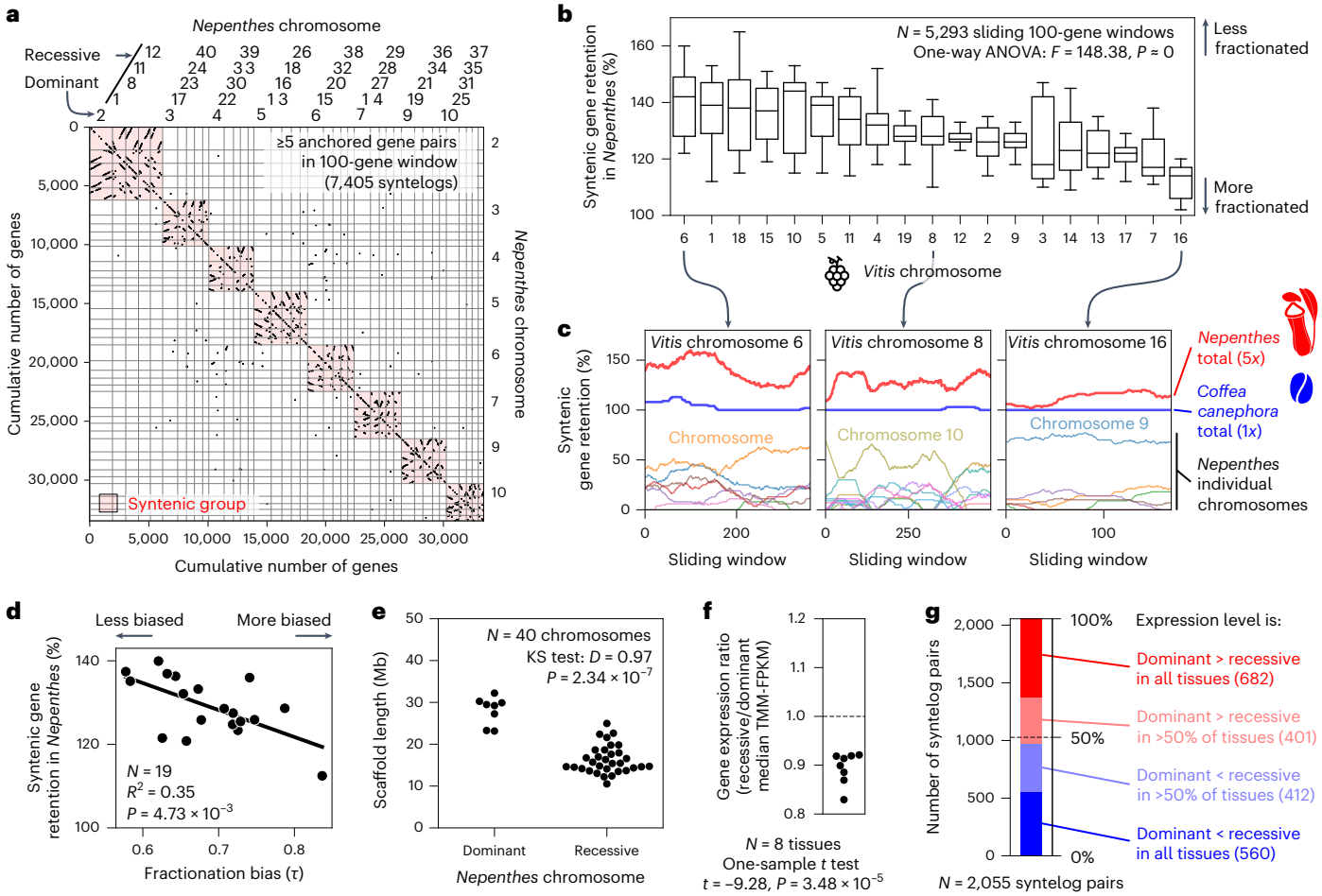


Fig. 2 | Subgenome dominance in the decaploid *Nepenthes* genome. **a**, Self–self syntenic dotplot of protein-coding sequences shows clear evidence for syntenic groups. Chromosomes are ordered by identified homoeologous groups (red). For the identification of dominant/recessive chromosomes, see **c** and Supplementary Fig. 5. **b**, The unequal rate of fractionation in the *Nepenthes* genome. Distribution of the syntenic gene retention rate is shown for syntenic blocks, which are grouped by corresponding *Vitis* chromosomes. Statistical significance of unequal rates was tested with a one-way analysis of variance (ANOVA). Box plot elements are defined as follows: centre line, median; box limits, upper and lower quartiles; whiskers, 1.5 × interquartile range. **c**, Different levels of fractionation in the *Nepenthes* syntenic blocks (5*x*) mapped to representative *Vitis* chromosomes (1*x*): lowest in chromosome 6, middle in chromosome 8 and highest in chromosome 16. Analysis of the *Coffea canephora* genome (also 1*x*) with the same FractBias parameters[26] is shown for comparison. For a full comparison, see Supplementary Fig. 5. For the count of retained syntelogs for each chromosome, see Supplementary Fig. 7. **d**, A negative correlation exists between fractionation bias and syntenic gene retention. Yanai's τ[69] was utilized as a proxy for fractionation bias. Linear regression and associated statistics are provided in the plot. **e**, Dominant chromosomes tend to be larger than recessive ones. Statistical significance was examined with a two-sided Kolmogorov–Smirnov (KS) test. **f**, Syntelogs in the dominant chromosomes tend to express at higher levels than those in the recessive chromosomes. Statistical significance was tested with a two-sided one-sample *t*-test. Eight tissues in *Nepenthes* were analysed (Supplementary Table 4). **g**, Syntelog-wise comparison of expression levels in genes on a dominant chromosome versus on a recessive chromosome.

DNA (ddRAD-seq) data from wild, sex-identified *N. gracilis* individuals[23]. The result showed a clear, sharply delimited signal of male-specific sequences in a 1 Mb region on chromosome 20 (Fig. 3a, Supplementary Fig. 9 and Supplementary Text 4). Chromosome 20 is a member of one recessive subgenome (Fig. 2a) and shows a substantial reduction in syntenic gene numbers compared with its dominant homoeolog (chromosome 6) (Supplementary Fig. 7). We interpret this reduction as evidence of chromosome-wide relaxed purifying selection, a feature we specifically attribute to recessive subgenomes and concomitant biased fractionation, rather than to sex linkage. Nonetheless, the presence of ancestrally relaxed purifying selection for gene content could have facilitated the emergence of novel mutations at the origin of the sex-linked region in *Nepenthes*. Structural comparison of the full Y chromosome from the male assembly against its corresponding X chromosome from the female assembly (Supplementary Fig. 10) revealed that the MSY cannot be aligned, incurring a large gap in the female instead. For the genes located in the MSY, we did not detect syntenic counterparts on the X chromosome, on the four homoeologs of the sex chromosome (chromosomes 6, 15, 32 and 38) nor anywhere else in the male or female assemblies. These findings suggest that the MSY is fully hemizygous and that no gametologs (regions with copies on both X and Y chromosomes) exist, similar to the independently evolved MSY of *Asparagus*[32].

A major question in sex-chromosome research is about the causes and consequences of suppressed recombination. A lack of X–Y recombination throughout the *Nepenthes* MSY is implicit in the population-resequencing approach that we used for the delimitation of the MSY and is furthermore implied by the observed hemizygosity. However, in earlier stages before and during the establishment of the MSY, adjacency to the satellite-rich putative centromere (Fig. 3b and Supplementary Fig. 11) may have facilitated X–Y divergence by providing a recombination-depauperate substrate within which novel

haplotypes could more easily have become fixed (for example, translocations into the proto-Y and deletions from the proto-X). Pericentromeric regions, due to their heterochromatin and low recombination rates, can favour the formation of MSYs and other supergenes, as shown in other plant and animal systems[33–39]. In the classical population genetic model for the transition from hermaphroditism towards dioecy[40,41], two independent sex-determining mutations arise (one factor controlling male reproduction, another controlling female reproduction), whereafter additional mutations that modify recombination rates between them become favoured by selection. Naturally, pre-existing non-recombining regions, such as pericentromeric regions, can facilitate the establishment of dioecy and sex-chromosome evolution by a two-factor sex determination system. In the *Nepenthes* MSY, three male-specific transcription factor genes (described below) are embedded in a tandem cluster of long interspersed nuclear elements, a group of retrotransposons (Fig. 3b and Supplementary Fig. 11). Transposable elements have accumulated not only in these intergenic regions but also in the introns of one of these genes itself (Supplementary Fig. 12). It seems likely that this transposable element accumulation is associated with suppressed recombination.

Genes in the MSY

Notably, the MSY, which spans approximately 1 Mb in the 16.7 Mb chromosome, contains DYSFUNCTIONAL TAPETUM 1 (*DYT1*), which is the only fully male-linked gene known to date in *Nepenthes*[23] (Fig. 3b). *DYT1* encodes a basic helix–loop–helix transcription factor that, in *Arabidopsis*, functions in the cell maturation of tapetum, the nutritive cell layer that aids microsporogenesis in the developing anther[42]. The potential function of *DYT1* may therefore be essential for male reproduction, and its absence in females would constitute a recessive male-sterility mutation, corresponding to the two-factor model for the evolution of dioecy[41]. Interestingly, *DYT1* in *Arabidopsis* directly regulates the expression of another tapetum development gene, DEFECTIVE IN TAPETAL DEVELOPMENT AND FUNCTION 1 (*TDF1*)[43], an ortholog of which acts as the male-promoting factor in *Asparagus*[44]. An ortholog of MALE MEIOCYTE DEATH 1 (*MMD1*), which encodes a PHD-finger transcription factor whose loss causes male meiotic defects[45], is another transcription factor gene located in the male-specific region. While the characterized functions of *DYT1* and *MMD1* suggest their involvement in microspore development, which begins later in anther development, female flowers in *Nepenthes* lack not only microspores but also, almost entirely, staminal structures[46]. We, therefore, hypothesized that another gene upstream of *DYT1* and *MMD1* locates to the male-specific region and determines floral sex. Consistent with this idea, we found a male-specific copy (*LFY-Y*) of the LEAFY (*LFY*) gene, which in hermaphroditic angiosperms encodes a plant-specific transcription factor that assigns the floral fate of meristems[47]. *LFY* is one of the earliest expressed genes in flower development, where it acts as a master regulator of the reproductive state (Supplementary Text 5). Loss of *LFY* gene function converts lateral floral organs, including stamens and carpels, into leaf-like structures[47].

LFY is well known to be a highly conserved gene in angiosperms, where it is maintained as a single copy in most species[48]. The *Nepenthes* genome harbours what is likely to be the principal *LFY* copy on an autosome (*LFY-A* on chromosome 3) that is not homoeologous to the putative sex chromosome. Gene phylogeny (Fig. 3c), chromosomal syntenic groups (Fig. 2a) and the presence of introns in both copies (Supplementary Fig. 12) suggest that the two *LFY* genes emerged by a lineage-specific SSD rather than via retrotransposition or WGD in *Nepenthes* after its split from the other carnivorous lineages in Caryophyllales, consistent with the emergence of dioecy in Nepenthaceae alone. This situation contrasts with *DYT1* and *MMD1*, both of which are maintained as single-copy genes in the decaploid genome of *Nepenthes* (Supplementary Fig. 13). As expected, *DYT1*, *MMD1* and *LFY-Y* were expressed in the developing buds of male flowers, but not in female buds (Fig. 3d,e). Furthermore, the male-specific expression of *LFY-Y* was maintained until flower maturation (Supplementary Fig. 14a). By contrast, we found no significant difference in *LFY-A* expression between male and female buds. Although this expression analysis has been performed by heterologous mapping to the *N. gracilis* genome using RNA-seq reads from related species (Methods), the read mapping rates were comfortably high (75.9–83.9% in 31 samples), and we further confirmed by transcriptome assembly that the transcripts of *DYT1*, *MMD1* and *LFY-Y* were detected only in male samples (Supplementary Fig. 14b–d). In the MSY, only these three genes could be functionally annotated with sequence similarity against the UniProt database (Supplementary Table 5), and they showed the highest male/female transcript abundance ratios in developing buds among all physically linked gene models in the region (Supplementary Fig. 14e). In summary, the three transcription factor genes in the MSY are good candidates for involvement in the establishment of dioecy in *Nepenthes*.

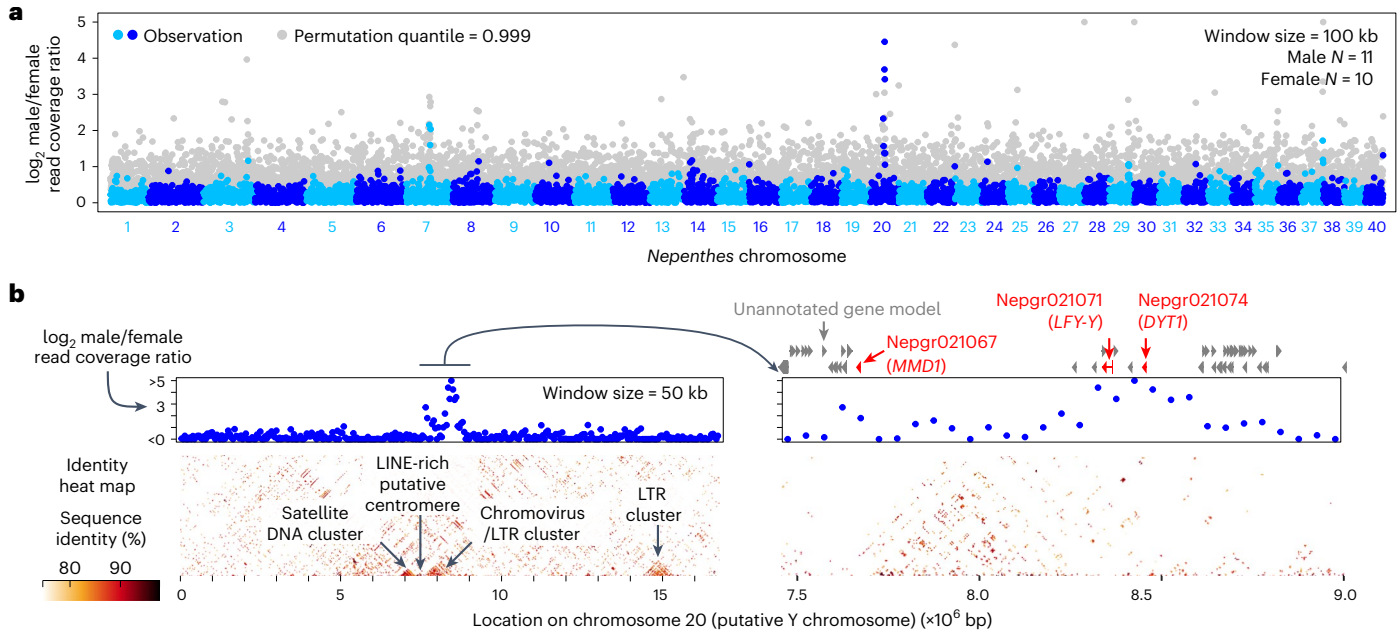
Neofunctionalization of the male-specific *LFY-Y* gene copy

To examine how two *LFY* copies acquired distinct expression patterns, we characterized *cis*-regulatory motifs in their promoters by mapping promoter motif sequences using the JASPAR database[49]. Under a false discovery rate (FDR) cut-off of 0.05, we detected 23 types of putative transcription factor-binding motif in the *LFY-Y* promoter (Supplementary Table 7), of which three motifs were found specifically in *LFY-Y* among all *LFY* promoter sequences analysed in the 20 species shown in Fig. 1b, including the *LFY-A* promoter: these were motifs bound by AUXIN RESPONSE FACTOR 18 (ARF18), TESMIN/TSO1-LIKE CXC 3 (TCX3) or ZINC-FINGER HOMEODOMAIN PROTEIN 6 (ZHD6) (Fig. 3f). In *Arabidopsis*, *ARF18* is highly expressed in developing carpels (Fig. 3g) and encodes a transcriptional repressor[50], whereas *TCX3* and *ZHD6* are expressed mainly in anthers (Fig. 3g), suggesting roles for the encoded regulators in modulating *LFY-Y* expression in both male and female organs.

Fig. 3 | The male-specific chromosomal region harbours transcriptional regulators of flower development. a,**b**, Analysis of genomic regions specific to male individuals in *N. gracilis*. ddRAD-seq reads were mapped to the reference genome to delimit the MSY. A genome-wide overview (**a**), alongside a magnified view of chromosome 20 (**b**), is shown. In the bottom panels of **b**, tandem repeats are visualized with identity heat maps using StainedGlass[117]. The range of the right panel corresponds to the range of the black line on the left panel. Positions of Trinotate-annotated and unannotated gene models are indicated in red and grey, respectively. For a male-versus-female syntenic comparison of chromosome 20, see Supplementary Fig. 10. The position of the putative centromere, which is typically LINE-rich in *N. gracilis*, is indicated in **b** (for details, see Supplementary Fig. 11). LINE, long interspersed nuclear elements; LTR, long terminal repeat. **c**, Phylogenetic relationships of *LFY* genes in angiosperms. Duplicates in *Nepenthes* are indicated with colours. The bar indicates 0.1 substitutions per nucleotide site. Supplementary Fig. 15a provides a complete phylogeny. **d**, Male and female flowers of *N. gracilis*. The male flower picture is licensed under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/) by H. P. Lim. **e**, Expression of genes in the MSY. Cultivated individuals of *Nepenthes* spp. (Supplementary Table 4) were used for RNA-seq, and reads were mapped to the *N. gracilis* genome. *P* values and *t* statistics of Welch's two-sided *t*-tests are provided above the plot. TPM, transcripts per million. **f**, Promoter differentiation of duplicated *LFY* genes. The transcription factor-binding motifs in JASPAR (FDR < 0.05) were detected in the 2 kb promoter sequences of all genes in **c**, and their overlap is shown in the Venn diagram. A comparison of the duplicated *LFY* promoters in *N. gracilis* is shown to the right. Commonly found motifs are connected with lines in order of proximity to the transcription start site. **g**, Tissue-specific expression patterns of transcription factor genes in *A. thaliana* flowers[59]. The expression heat map shows genes encoding transcription factors that bind to motifs detected in the *LFY-Y* promoter. The flower illustration is licensed under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/) by F. Bouché.

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

These *LFY-Y*-specific motifs (ARF18, TCX3 and ZHD6) were detected 1.0–1.5 kb upstream from the start codon. This region may have been newly acquired by *LFY-Y* during sex chromosome evolution. Within this promoter region, we also detected paired PISTILLATA (PI)-binding motifs only in *Nepenthes LFY-Y*, not in *LFY-A*, and interestingly, in *LFY* promoters of the comparator dioecious species spinach (*Spinacia oleracea*). LFY in fact also up-regulates *PI* expression in *Arabidopsis*[51]. PI is a B-class MADS-box protein that specifies petal and stamen identity in *Arabidopsis* and other plants[52]. Spinach evolved dioecy independently from *Nepenthes* (Fig. 1b), and the suppression of spinach *PI* converts its male flowers into phenotypically normal female flowers[53], suggesting a potentially convergent mechanism underlying the evolution of dioecy between the two Caryophyllales lineages through regulatory interactions among LFY, PI and other transcription factors (for known sex determination systems in other Caryophyllales lineages, see Supplementary Text 7). Taken together, these results suggest that *Nepenthes LFY-Y* was neofunctionalized in terms of expression pattern, likely as a consequence of the change in *cis*-regulatory elements. We detected five amino acid substitutions specifically found in the C-terminal DNA-binding domain of LFY-Y (Supplementary Fig. 15), suggesting potential changes in protein properties in addition to the expression patterns.



Pitcher-tissue-specific prey capture responses

Besides dioecy, trapping leaf organization also serves as an evolutionary innovation in *Nepenthes*. A single *Nepenthes* leaf consists of segmented parts, including a carnivorous pitcher trap that is further elaborated by tissue differentiations along the proximodistal axis, representing one of the most complex leaf shapes known among angiosperms[54]. In Caryophyllales, plant carnivory evolved before the origin of Nepenthaceae (Fig. 1b), presumably with the flypaper-type trapping mechanism and within a conventional leaf organization[19,20,55], with pitcher organization representing an evolutionary novelty that emerged in *Nepenthes* only.

To examine whether gene expression patterns parallel unique pitcher tissue differentiation, we performed RNA-seq experiments in six dissected leaf parts: from proximal to distal, flat part, tendril, digestive zone, waxy zone, peristome and lid (Fig. 4a,b). For a cross-species comparison, we also generated the pitcher tissue-specific transcriptomes of *Cephalotus*, which evolved pitcher leaves independently from *Nepenthes* in an altogether different angiosperm order (Fig. 1b), thus representing a product of convergent evolution[19,21]. We further integrated additional RNA-seq data from *Dionaea*[56–58] and *Arabidopsis*[59] to form a four-species expression level dataset normalized with trimmed mean of *M* values (TMM)[60] for a total of 3,572 single-copy genes. Dimensionality reduction analysis with different methods consistently showed that, in gene expression profiles, the pitcher tissues of *Nepenthes* were distinct from conventional photosynthetic organs (Supplementary Fig. 16a). This contrasts with the pitcher leaves of *Cephalotus*, whose tissues showed expression profiles largely overlapping with the photosynthetic organs, suggesting that pitcher tissues in *Nepenthes* traps are much more strongly differentiated from ancestral photosynthetic leaf structures than those of *Cephalotus*. This view is consistent with the contrasting patterns of tissue-specific photosynthetic activity in the two pitcher plant lineages, with *Nepenthes* pitchers showing almost no photosynthetic assimilation[61,62].

Next, we characterized tissue-specific prey capture responses in *Nepenthes*. Pitchers were treated with an insect homogenate to mimic prey capture, and tissues were collected after 24 h, coincident with the expected onset of jasmonic-acid-mediated prey responses[63]. Jasmonic acid is a known regulator of gene expression in plants and can mediate plant–herbivore and plant–pathogen interactions[64,65], as well as prey capture responses in the Caryophyllales lineage of carnivorous plants including *Nepenthes*[66]. As anticipated, the number of significantly differentially expressed genes (DEGs; FDR < 0.05) was highest in the digestive zone, where digestive and absorptive glands[21] were directly exposed to the insect homogenate (Fig. 4b). A clear response of the digestive zone was further confirmed by dimensionality reduction analysis (Supplementary Fig. 16b). Significantly enriched Gene Ontology (GO) terms (Supplementary Tables 8–11) among the up-regulated genes clearly indicated the activation of translation machinery upon prey capture (Supplementary Table 8), which would be anticipated for rapid response by the palette of necessary digestive enzymes. Other enriched GO terms included those presumably associated with prey digestion, nutrient metabolism and biological interactions (Fig. 4b). The second-highest number of DEGs was found in the tendril, which connects the pitcher to the rest of the plant body. Interestingly, with the above threshold, we did not detect DEGs in other trapping leaf tissues, including the flat part, where absorbed nutrients flow through. While long-term prey feeding over weeks can alter the abundance of photosynthetic proteins in the photosynthetic lamina[67], our findings suggest that daylong prey response is well compartmentalized within a *Nepenthes* trapping leaf.

Tandem gene clusters with pitcher-tissue-specific expression

The complexity of *Nepenthes* leaves may be reflected by considerable subfunctionalized or neofunctionalized duplicate gene expression in novel tissues, where amplification of gene copies can be mediated by tandem gene duplications[68]. We therefore searched for tandem gene clusters with high tissue specificity in gene expression, measured by the τ metric[69], which outperformed alternative measures in a benchmark study[70]. A conspicuous example we detected was gene clusters of SENESCENCE-RELATED GENE 1 (*SRG1*) orthologs[71]. This senescence marker gene encodes cytoplasm-localized melatonin 3-hydroxylase (M3H) and produces cyclic 3-hydroxymelatonin, whose antioxidant activity is 15-fold higher than that of its precursor melatonin, thereby promoting growth in *Arabidopsis* and rice[72–74]. This family of genes forms tandem duplicates on separate *Nepenthes* chromosomes within the same syntenic group (Fig. 4c), with the largest cluster (19 gene models with 15 cases with complete protein domain structure (Supplementary Fig. 17)) on chromosome 20 (the sex chromosome), which belongs to a recessive subgenome. Phylogenetic analyses suggested that the first copy of this cluster originated as a WGD duplicate whose counterpart in the dominant subgenome (chromosome 6) remains a single-copy gene (Nepgr007022). The tandem genes in the dominant subgenome showed well-conserved microsynteny among eudicots (Supplementary Fig. 17), suggesting ancient origins (>131 Ma; Fig. 1b). Interestingly, the tandem array on chromosome 20 contains many genes with expression specific to the digestive zone in the pitcher, which forms the interface to the digestive fluid, where peroxidases likely produce reactive oxygen species (ROS) to facilitate proteolysis for prey degradation[75–78]. Given this characteristic expression, SRG1 proteins may participate in scavenging cytotoxic ROS produced during prey digestion and nutrient absorption.

To further examine whether other carnivory-related genes similarly formed tandem clusters, we analysed genes encoding digestive enzymes. Using a list of experimentally confirmed digestive fluid proteins, many of which are digestive enzymes[77], we characterized their phylogeny, synteny and expression in *Nepenthes* (Supplementary Fig. 18). Among the 11 families of genes we analysed, we detected 8 tandem gene clusters (6 on recessive subgenomes) that were formed in *Nepenthes* after its split from close relatives in Caryophyllales. Tandem clusters encoding typical digestive enzymes (aspartic protease, class III peroxidase, glycoside hydrolase family 19 chitinase, purple acid phosphatase, RNase T2 and β-1,3-glucanase) included genes whose transcript abundance was highest in the digestive zone, with increased expression following the feeding treatment. These results suggest that specific tandem cluster evolution is not restricted to *SRG1* and may be found in many other genes involved in *Nepenthes* carnivory.

Unequal subgenomic contributions for novel genes

As the male-specific region harbouring *LFY-Y* and many tandem gene clusters expressed in the digestive zone (*SRG1* and digestive enzyme genes) were found in recessive subgenomes rather than in the dominant subgenome, subgenomes may differentially serve as hosts of novel duplicated genes. However, these observations may also be explained by chance, as there is a 70% probability that a novel gene will be located in a recessive subgenome if its occurrence is proportional to the chromosome assembly size (528/753 Mb) and a 67.7% probability if it occurs proportionally to the number of genes (23,039/34,010 genes). Therefore, statistical analysis was necessary. To this end, we analysed lineage-specific genes in the *Nepenthes* genome. First, DIAMOND BLASTP v2.0.15 searches[79] were conducted against 20 other plant genomes to identify *Nepenthes* genes that are most similar to another *Nepenthes* gene rather than to genes from other species and thus likely to have emerged by gene duplication after the split from those other lineages. Identified lineage-specific genes were significantly enriched by those having specific expression in all analysed tissues (Fig. 5b). Next, the types of duplication (that is, WGD or SSD) and involved subgenomes (dominant or recessive) were estimated by whether the best-hit gene was on the same chromosome, on homoeologous chromosomes or on other chromosomes (Fig. 5a). Finally, the frequency of each category of tissue-specific genes was compared with the overall average

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

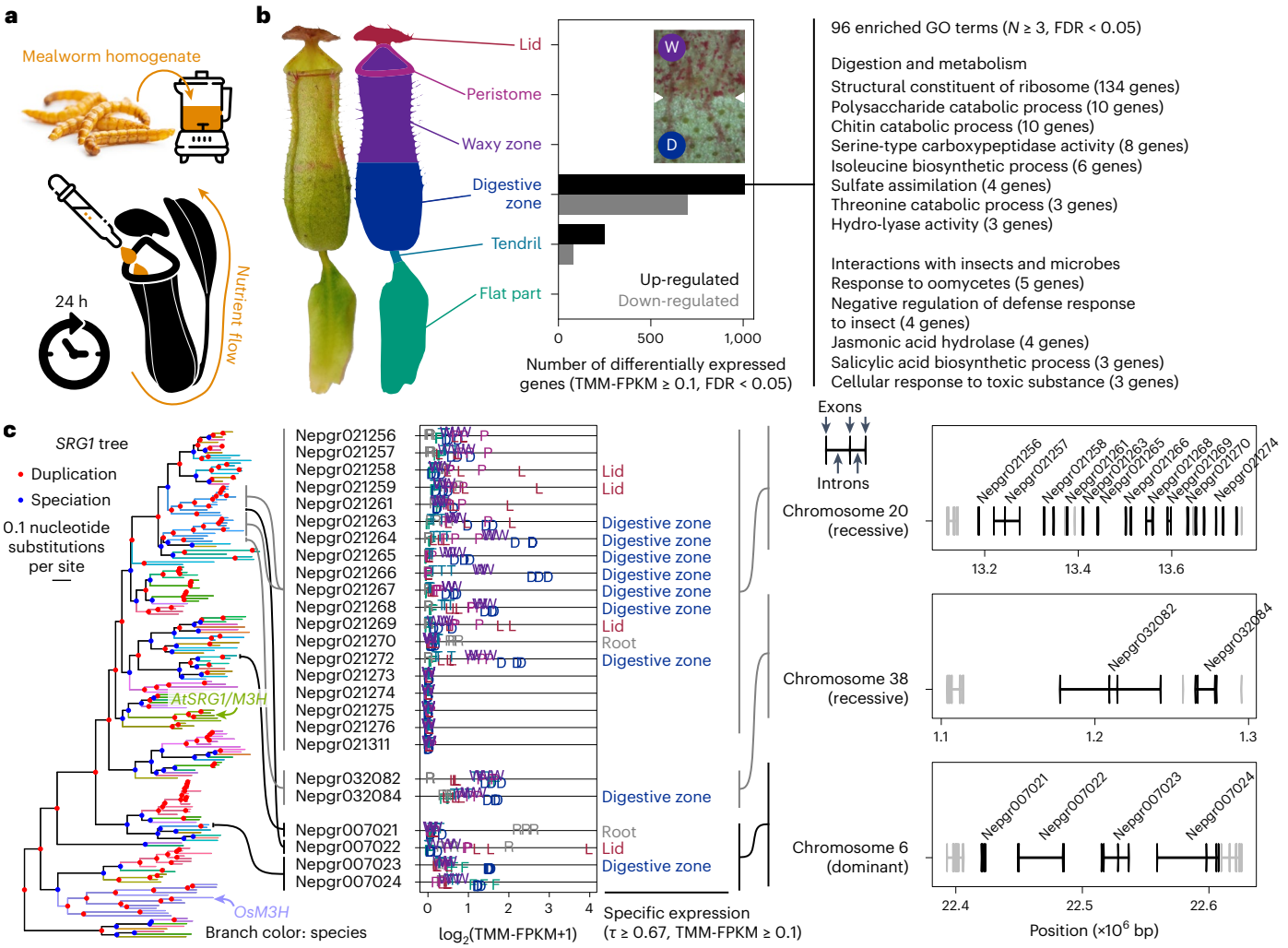


Fig. 4 | Tissue-specific gene expression in trapping pitcher leaves. a, The feeding experiment (see Methods for details). **b**, Genes responsive to prey in pitcher tissues. The functionally differentiated proximal–distal segments of pitcher tissues are indicated with colours. The boundary between the waxy zone (W) and the digestive zone (D) is highlighted as an inset in the bar plot. On the right, selected GO terms enriched in the up-regulated genes in the digestive zone are displayed. For comprehensive lists, see Supplementary Tables 8–11. **c**, The massive amplification of *SRG1* paralogs in *Nepenthes*. Left: the ML phylogeny; middle: expression level and tissue specificity; right: chromosomal location. The positions of *A. thaliana SRG1*/*M3H* and *Oryza sativa M3H* are marked in the tree, which was generated using the 20-genome dataset (Supplementary Table 6). Branch colours indicate species identifications. For the complete phylogeny, see Supplementary Fig. 17. The expression levels for each tissue, from three replicates, are denoted by the initial letter: L, lid; P, peristome; W, waxy zone; D, digestive zone; T, tendril; F, flat part; R, root. The tissue primarily expressing the gene is indicated to the right of the expression level plot when the gene expression specificity and abundance exceed the thresholds ($\tau \geq 0.67$ and TMM-FPKM ≥ 0.1). The chromosomal locations of non-*SRG1* gene models are marked in grey. Image credits: the mealworm photo and icons from freepik.com.

obtained from all expressed genes. Interestingly, the recessive subgenomes tended to host more tissue-specific genes than the dominant subgenome. Such distribution shifts from the null expectation with all expressed genes were statistically significant in genes specifically expressed in six out of the seven analysed tissues (Fig. 5b), including pitcher tissues, whose differentiation evolved in *Nepenthes* to form pitfall traps. The difference is likely to have emerged during functional divergence after gene duplications because the frequency of tandem duplications themselves was comparable between dominant and recessive subgenomes (Supplementary Fig. 19).

The increased contribution to the tissue-specific genes by recessive subgenomes was most pronounced among within-chromosome SSD duplicates, which are enriched in tandemly duplicated genes. The rate of protein evolution (*dN*/*dS*) was higher in SSD duplicates than in WGD duplicates, with the within-chromosome SSD duplicates showing the highest evolutionary rates (Supplementary Fig. 20). Because there was no overall difference in *dN*/*dS* between dominant and recessive subgenomes, it is possible that subgenome dominance may have only transiently relaxed purifying selection among the recessive subgenomes and that its signature has been difficult to capture in extant gene sequences, other than via the substantial gene losses with subgenome fractionation discussed above (Fig. 2).

Different modes of gene duplication have been shown to exert different long-term impacts on the complements of genes retained in plant genomes. Generally, as similarly noted for many other plant systems[80], WGD duplicates tend to be enriched for regulatory functions, whereas small-scale (for example, tandem) duplicates are more likely to be enriched in functions related to plant defence (and possibly by extension, to plant carnivory). To examine how functional gene categories are differentially enriched in the dominant and recessive subgenomes of *Nepenthes*, we analysed overrepresented GO categories (Supplementary Fig. 21 and Supplementary Tables 12–15). Among other functions, syntelogs on the dominant subgenome were significantly enriched for genes annotated with 'flower development' and

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

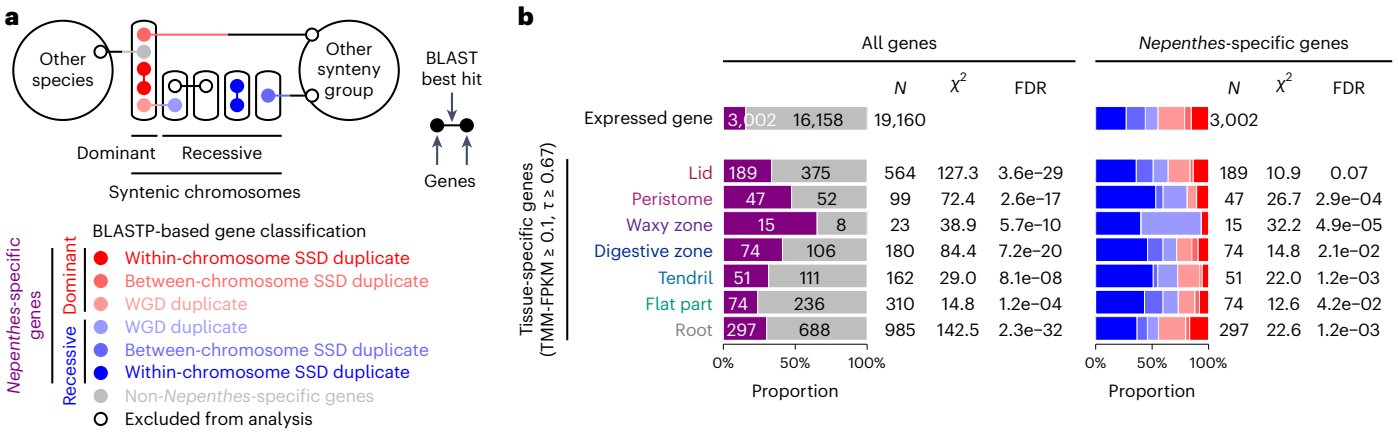


Fig. 5 | Differential contributions of subgenomes to the evolution of novel genes. a, Classification of duplication events in *Nepenthes*-specific genes. The categories of duplication, whether from SSDs or WGDs, and the timing of duplication (before or after the divergence of *Nepenthes* from its sister lineage) were determined by the best DIAMOND BLASTP hit within the 20-genome dataset (Supplementary Table 6). Circles represent genes, and edges connecting two circles indicate the best hit relationships. **b**, Differential contributions of duplication categories to tissue-specific novel *Nepenthes* genes. All tissue-specific genes (left) and *Nepenthes*-specific subsets (right) were analysed for individual tissues. The colours of the bar plot elements correspond to the subgenomic category (dominant/recessive) and the duplication origin of the analysed genes, matching those in **a**. The colours of the *y*-axis labels match those used in Fig. 4. The proportions of *Nepenthes*-specific genes (right) and duplication categories (right) in tissue-specific genes were compared with that of expressed genes (TMM-FPKM ≥ 0.1) using χ^2 tests controlled for FDRs. The numbers of genes (*N*), χ^2 statistics and FDR are provided in the plots.

'ethylene-activated signalling pathway' GOs (Bonferroni-corrected $P < 0.05$). Syntelogs on the recessive subgenomes had potentially related GO enrichments such as 'abaxial cell fate specification', 'floral meristem determinacy' and 'leaf morphogenesis' and were also marked by various plant hormone GOs, including 'response to jasmonic acid'. As such, the evolution of dominant and recessive subgenomes following polyploidization may have included both specialization and partitioning of ancestral regulatory networks, in a manner analogous to neofunctionalization and subfunctionalization, respectively, at the individual gene level[81]. Regarding small-scale, local duplication events, dominant tandems were enriched in genes annotated with 'cell surface receptor signalling pathway' and 'response to oomycetes', whereas recessive tandems were enriched, for example, with GOs such as 'methyl jasmonate methylesterase activity' and 'salicylic acid glucosyltransferase (glucoside-forming) activity', which could conceivably be associated with prey recognition pathways[82].

Taken together, these results suggest that recessive subgenomes play an important role in hosting novel tissue-specific genes that evolved through SSDs, possibly in an environment of relaxed purifying selection compared to that in the dominant subgenome, thereby contributing to the unique biology of *Nepenthes*, including dioecy (Fig. 3) and carnivory (Fig. 4).

Discussion

In this study, we elucidated the decaploid structure of the *Nepenthes* genome and identified a clear signature of 1:4 subgenome dominance (Fig. 2). We also highlighted how the four recessive subgenomes have contributed to evolutionary novelties in *Nepenthes* (Fig. 1), especially in relation to dioecy and carnivorous trapping leaves.

Our analyses of the organization of the MSY (Fig. 3) suggested that a series of mutational events may have led to the evolution of dioecy in *Nepenthes* (Supplementary Fig. 22a). A likely scenario is an evolutionary path from hermaphroditism via gynodioecy towards full dioecy[41,83]. First, an ancestral population of hermaphrodites could have given rise to a gynodioecious population through loss of function of *MMD1* or *DYT1* in the ancestral X chromosome, which belongs to a recessive subgenome. This event would have resulted in the emergence of homozygous, recessive, female-only plants without functional pollen. The linkage of *MMD1* and *DYT1* and their double loss of function on the X chromosome might have enhanced the female-only trait. Frequency-dependent selection may then have favoured male function in hermaphroditic plants, until the gain of the masculinization gene on the ancestral Y chromosome, in accordance with the two-factor model for the transition towards dioecy and the origin of sex chromosomes[41]. Such a masculinization gene should dominantly suppress the production of carpels or at least their function. From our association and expression analysis, *LFY-Y* was a plausible candidate for the masculinization gene, which, in the above scenario, would have resulted in the evolution of a completely dioecious proto-*Nepenthes* population. However, male reproductive organs (anthers and filaments) are almost completely suppressed in the flowers of female *Nepenthes*, such that male meiosis and pollen formation (that is, the later stages where loss of *MMD1* and/or *DYT1* would manifest as male sterility) are precluded. We hence hypothesize that additional mutations suppressing anthers and filaments (recessive to the MSY) occurred in genomic regions present in females (autosomes and X chromosomes). This important argument hints at the possibility that the current sex determination system of *Nepenthes* is no longer based on the joint action of the two MSY factors but instead on a masculinizing effect of the *LFY-Y* gene acting in a genomic background of 'default' unisexual female development[84]. Such single-factor sex determination is known from a range of plants[85] and is the norm in vertebrates. Interestingly, *Nepenthes* may represent a case where a 'single-factor system with default sex' was derived from a previous stage having two-factor sex determination.

An alternative scenario is that *LFY-Y* as a dominant masculinization gene was gained in the MSY before the localization and activity of *DYT1* and *MMD1*, implying an intermediate androdioecious step towards full dioecy. The sequence of these mutational events cannot be discerned by molecular clocks because *Nepenthes DYT1* and *MMD1* lack gametologs or other paralogs. However, population genetic theory asserts that androdioecy is generally less plausible than gynodioecy[83], so it is difficult to explain the linkage of *MMD1* and *DYT1* in the MSY if the masculinization gene evolved first to form androdioecious intermediates. It is important to note that our interpretations are based on the well-understood functions of *MMD1*, *DYT1* and *LFY*, which have been characterized in other plants. This leaves room for the possibility that Y-linked *Nepenthes* orthologs could possess other, lineage-specific functions, thereby providing alternative explanations for the evolution of dioecy, distinct from the scenarios discussed above.

Furthermore, *LFY-Y* expression in male flowers and *LFY-A* expression in both flower sexes are consistent with neofunctionalization of *LFY-Y* rather than *LFY* duplicate subfunctionalization[81] as a mode of functional change after gene duplication (Supplementary Fig. 22b). Indeed, the functional partitioning we describe here resembles the opposite of a sex-aggregating partial *LFY* redundancy and deletion scenario developed to explain the evolution of the bisexual angiosperm flower from separate male and female reproductive axes controlled by two distinct *LFY* copies in gymnosperms[86]. Eventual heterologous experimental characterization of the *LFY-Y* gene in a genetically tractable species is expected to yield further insight into the evolution of dioecy in *Nepenthes* (Supplementary Text 6).

In addition to our inferences on the evolution of *Nepenthes* dioecy, our analyses of tissue-specific genes in *N. gracilis* pitcher leaves (Fig. 4) showed how molecular evolution likely paralleled the increased complexity of tissue organization in this carnivorous plant. As one of the most prominent examples, *SRG1* genes, which may participate in scavenging ROS during prey digestion and nutrient absorption, formed a massive tandem cluster in a recessive subgenome, many of the members of which acquired tissue-specific expression in *Nepenthes*-specific tissues, such as the digestive zone. In contrast to the sex-linked *LFY* duplication, the mode of functional divergence (that is, neofunctionalization or subfunctionalization) in the *SRG1* cluster is not clear. However, it seems evident from gene expression and phylogenetic relationship data (Fig. 4c) that successive functional divergence has taken place during tandem cluster formation. The novel gene acquisitions discussed here (that is, *LFY-Y* and *SRG1*) occurred in recessive subgenomes, as did many other *Nepenthes*-specific gene duplications that acquired tissue-specific expression (Fig. 5), including digestive enzyme genes (Supplementary Fig. 18). This genome-wide trend (Fig. 5) suggests that WGDs influenced the fates of subsequently produced SSD duplicates through recessivity and concomitant relaxed purifying selection, in a system with strongly divergent subgenome dominance patterns.

Our findings of novel gene duplicates suggest that the greater prevalence of tissue-specific genes after SSDs in recessive subgenomes may have contributed to adaptive potential in *Nepenthes* and, thus, to the maintenance of these recessive subgenomes. Thus, the recessive subgenomes have not degenerated completely but have instead persisted for long periods of time (58.3–93.8 million years in *Nepenthes*[25]), contributing to the emergence of evolutionary novelties. Although the myriad of polyploids that have arisen in plant evolution may frequently represent evolutionary dead ends[2], as we have shown in *Nepenthes*, recessive subgenomes may have served as sources of adaptive potential in some highly radiated polyploid lineages. Our findings may therefore contribute to revised models of karyotype stability and gene divergence during polyploid evolution.

Methods

Plant materials

For genome sequencing, male and female *N. gracilis* individuals were collected from the field. The collection date was 26 March 2019, and the collection locality was Tampines Avenue 10, 1° 21′ 27.1″ N, 103° 55′ 49.9″ E (1.35753° N, 103.93053° E), Tampines, Singapore. The habitat was an open secondary forest on seasonally waterlogged flat ground with *Dillenia suffruticosa*, *Acrostichum aureum*, *Dicranopteris linearis*, sedges and grasses. For transcriptome analysis of vegetative tissues, we purchased *N. gracilis* from CZ Plants. The plants were grown on a mixture of sphagnum moss and perlite in the greenhouse. As it was difficult to obtain fresh flower samples from *N. gracilis*, we obtained developing inflorescences from other *Nepenthes* cultivars and species (Supplementary Table 4). Inflorescences were dissected into early flower buds, late flower buds, young flowers and mature flowers (Supplementary Fig. 23). Plants of *Ancistrocladus abbreviatus* (Ancistrocladaceae)[87] and *Drosophyllum lusitanicum* (Drosophyllaceae) were grown on soil in a greenhouse. Axenically grown strains of *Cephalotus follicularis* (Cephalotaceae)[77,88] and *Triphyophyllum peltatum* (Dioncophyllaceae)[89,90] were maintained in half-strength Murashige and Skoog solid medium[91] supplemented with 3% sucrose, 1× Gamborg's vitamins, 0.1% 2-(*N*-morpholino)ethanesulfonic acid, 0.05% Plant Preservative Mixture (Plant Cell Technology) and 0.3% Phytagel, at 25 °C (*C. follicularis*) or 20 °C (*T. peltatum*) in continuous light.

High-molecular-weight genomic DNA isolation

Young leaf tissues of wild-collected male and female *N. gracilis* individuals were gathered, cleaned, flash frozen in liquid nitrogen and then stored at −80 °C before extraction. About 10 g of flash-frozen tissue was used for high-molecular-weight (HMW) genomic DNA isolation. The first step followed the BioNano NIBuffer nuclei isolation protocol, in which frozen leaf tissue was homogenized in liquid nitrogen, with a subsequent nuclei lysis step using the isolation buffer with Triton X-100 and β-mercaptoethanol (IBTB) with spermine and spermidine added and filtered just before use. IBTB consists of isolation buffer (15 mM Tris, 10 mM EDTA, 130 mM KCl, 20 mM NaCl, 8% (*w*/*v*) PVP-10, pH 9.4) with 0.1% Triton X-100 and 7.5% (*V*/*V*) β-mercaptoethanol mixed in and chilled on ice. The mixture of homogenized leaf tissue and IBTB was strained to remove undissolved plant tissue. Then, 1% Triton X-100 was added to lyse the nuclei before centrifugation at 2,000 × *g* for 10 min to pellet the nuclei. Once the nuclei pellet was obtained, we proceeded with CTAB (cetyltrimethylammonium bromide) DNA extraction with modifications for the Oxford Nanopore sequencer as described previously[92]. The quality and concentration of HMW genomic DNA was checked using a Thermo Scientific NanoDrop Spectrophotometer, as well as with agarose gel electrophoresis following standard protocols. The genomic DNA was further purified with a Qiagen Genomic-tip 500/G according to the protocol provided by the developer.

Genome sequencing

HMW DNA from male and female plants were used to generate sequencing libraries for use with ONT flow cells (SQK-LSK109). The resulting libraries were run on a PromethION sequencer running for 48 h. All bases were called on the PromethION using Guppy v2.0, and the resulting fastq files were used for genome assembly.

Genome assembly and polishing

The fastq sequencing data were filtered for >10 kb reads using seqtk v1.2 (https://github.com/lh3/seqtk). Filtered reads resulting from male and female plants were separately assembled using WTDBG2 v2.2 (wtdbg2 -t64 -p19 -AS2 -e2 -L10000) (https://github.com/ruanjue/wtdbg2). Consensus genome assemblies were generated by mapping reads >10 kb to the assembly with minimap2 v0.2 (https://github.com/lh3/minimap2) and then running Racon v1.3.1 (https://github.com/isovic/racon); the consensus process was repeated three times. The contig assemblies were further polished using paired-end 2 × 150 Illumina sequence by first aligning the reads to the consensus assemblies using minimap2 followed by running the assembly tool Pilon v1.18 (https://github.com/broadinstitute/pilon) three times using the paired-end 2 × 150 Illumina sequence data. Purge Haplotigs v1.0.0 (https://bitbucket.org/mroachawri/purge_haplotigs/src/master/) was applied to both the female and male Nanopore assemblies separately. The raw Illumina reads were aligned to the genome assemblies using bwa mem v0.7.17 (https://github.com/lh3/bwa), and input files were prepared using SAMtools v1.3 (https://github.com/samtools/samtools). Purge Haplotigs was then run using the prepared bam files and genome assemblies.

Hi-C scaffolding and syntenic path assembly

While Illumina-polished ONT-based wtdbg2 assemblies were generated independently for male and female specimens, we further improved the contiguity of the male assembly using HiRise scaffolding of Chicago and Dovetail Hi-C libraries by Dovetail Genomics[93] (Supplementary Fig. 24 and Supplementary Table 1). Heterologous Hi-C scaffolding of

the female genome assembly was performed using the female wtdbg2 assembly and the male Hi-C sequencing data. A list of Hi-C contact positions for the female was generated with Juicer v1.6 (https://github.com/aidenlab/juicer). This file was then used as input for 3D-DNA v170123 (https://github.com/aidenlab/3d-dna) to order and orient fragments of the genome assembly. Because of its lack of detectable synteny to any regions in the male assembly, an unnaturally large scaffold of the female Hi-C assembly was collapsed to its original contigs. We also used another approach to scaffolding the female ONT assembly. With the male Hi-C assembly as a reference, we applied the syntenic path assembly to the female genome using the SynMap function of CoGe[94]. Assembly statistics are available in Supplementary Table 2. The scaffold numbering in the final assembly corresponds to the chromosome numbers discussed in this paper.

Feeding experiment

Dried mealworms (batch number L400518, MultiFit Tiernahrungs GmbH) were ground into a fine powder with a mortar and a pestle to prepare a 100 mg ml^{-1} homogenate in water. The homogenate was subsequently centrifuged at 2,000 × *g* for 30 s, and the supernatant (mealworm extract) was obtained. Upon feeding, we measured the total volume of digestive fluid and added the mealworm extract adjusted to 10% of the total volume to ensure a comparable concentration among pitchers with different sizes. The pitchers were then sealed using parafilm and left in the greenhouse for 24 h. The same procedure was applied to control plants fed with water alone.

RNA extraction and transcriptome sequencing

Leaves of *N. gracilis* were dissected into the lid, peristome, waxy zone, digestive zone, tendril and flat part (Supplementary Fig. 23). Digestive fluid was discarded. After washing with water and drying with a paper towel, the tissues were immediately frozen in liquid nitrogen. Tissues from five leaves (from multiple individuals) were pooled for one replicate, and a total of three biological replicates were prepared. Root samples were collected from untreated plants, and each replicate was derived from a single individual. Frozen samples were homogenized in liquid nitrogen using mortar and pestle. RNA extraction was performed with PureLink Plant RNA Reagent solution (Thermo Fisher) according to the manufacturer's instructions. The RNA pellet was resuspended in RNase-free water at 4 °C. After centrifuging at 18,850 × *g* for 10 min at 4 °C, the solution was transferred to a new tube and purified using an RNeasy Mini kit (Qiagen). While the above extraction method was used for *Nepenthes*, *Ancistrocladus* (flowers, leaves and roots (Supplementary Fig. 25)), *Triphyophyllum* (juvenile non-trap leaves (Supplementary Fig. 26)) and *Cephalotus* (Supplementary Fig. 28), as well as trap leaves and roots of *Drosophyllum*, a modified CTAB protocol was used for *Drosophyllum* flowers (Supplementary Fig. 27). The frozen *Drosophyllum* flower samples were homogenized with mortar and pestle. The ground sample was transferred to a pre-cooled 2 ml tube, and 0.75 ml of preheated 2× CTAB buffer (65 °C) was added. The tube was shaken vigorously by using a vortex mixer and incubated in a thermoshaker at 1,400 r.p.m. for 20 min at 60 °C. An aliquot of 0.75 ml chloroform:isoamyl alcohol (25:1) was added and mixed vigorously. The homogenate was centrifuged at 21,630 × *g* for 15 min at room temperature. The supernatant was then transferred to a new 2 ml tube. For RNA precipitation, 1.5 ml of 100% ethanol was mixed with 60 µl of 3 M aqueous sodium acetate and added to the supernatant, which was then shaken at 500 r.p.m. for 60 min at room temperature. The homogenate was centrifuged at 21,630 × *g* for 20 min at room temperature. The supernatant was discarded carefully to not disturb the pellet. The pellet was then washed with 1 ml of 75% ethanol. The supernatant was discarded, and the pellet was air dried for 15 min in a thermoshaker at 37 °C. The pellet was resuspended in 100 µl of RNase-free water at 4 °C by shaking at 500 r.p.m. for 15 min. The samples were again centrifuged at 21,630 × *g* for 5 min at 4 °C. The solution was transferred to a new tube, and the RNA was purified using the RNeasy Mini kit (Qiagen). The quality of RNA was examined using Nanodrop (Thermo Fisher) and the Qubit IQ assay kit (Invitrogen). Total RNA was sent to Novogene, where paired-end mRNA sequencing libraries were prepared using the Novogene NGS RNA Library Prep Set (PT042). Briefly, after the poly-A enrichment and fragmentation, the first-strand complementary DNA was synthesized using random hexamer primers followed by the second-strand cDNA synthesis, end repair, A-tailing, adapter ligation, size selection for 250–300 bp insert size, amplification and purification. Libraries were paired-end sequenced for 150 bps with an Illumina Novaseq 6000 platform.

Transcriptome assembly

Transcriptome assemblies for *A. abbreviatus*, *D. lusitanicum* and *T. peltatum* were generated with the RNA-seq reads (Supplementary Table 4). The reads were preprocessed with fastp v0.20.0 (https://github.com/OpenGene/fastp) and assembled using Trinity (see Supplementary Table 6 for version) (https://github.com/trinityrnaseq/trinityrnaseq). Open reading frames (ORFs) were obtained with TransDecoder v5.5.0 with a minimum length of 50 bp (-m 50) (https://github.com/TransDecoder/TransDecoder). The longest ORFs among isoforms were extracted with the 'aggregate' function of CDSKIT v0.9.1 (https://github.com/kfuku52/cdskit).

Repeat identification and annotation

For subsequent gene model prediction, repetitive elements on the reference genome were masked using RepeatMasker v4.0.9 (https://github.com/rmhubley/RepeatMasker) with a species-specific repetitive sequence library generated by RepeatModeler v2.0 (https://github.com/Dfam-consortium/RepeatModeler) (Supplementary Table 3). For further identification and annotation of repetitive elements, we used EDTA v2.1.0 (https://github.com/oushujun/EDTA). To identify the overall repetitiveness of genomes, we performed de novo repeat discovery with RepeatExplorer2[95]. We used a repeat library obtained from the RepeatExplorer2 analysis of Illumina paired-end reads. All clusters representing at least 0.005% of the genomes were manually checked, and the automated annotation was corrected if needed. Contigs from the annotated clusters were used to build a repeat library. Transposable element protein domains[96] found in the assembled genomes were annotated using the DANTE tool available from the RepeatExplorer2 Galaxy portal (https://galaxy-elixir.cerit-sc.cz/). Further repeat density distribution plots shown in Supplementary Fig. 11 were made using shinyCircos (https://venyao.xyz/shinycircos/) and pyGenomeTracks v3.6 (ref. 97).

Analysis of tandem repeat clusters

We used StainedGlass v0.4 (https://github.com/mrvollger/StainedGlass) to identify repeat sequence clusters. The genomic distribution of repeat sequences was visualized using HiGlass v0.10.1 (https://github.com/higlass/higlass) and HiGlass Manage (https://github.com/higlass/higlass-manage) with the gene annotation track (https://github.com/higlass/gene_annotations).

Transcriptome assembly for gene modelling

The transcriptome assembly was performed using a combination of genome-guided and de novo approaches. The genome-guided approach used StringTie v2.1.4 (https://github.com/gpertea/stringtie) with aligned reads from HiSat2 v2.1.0 (https://github.com/DaehwanKimLab/hisat2). For the de novo approach, we first ran Trinity v2.8.5 with default parameters, followed by running TransAbyss v2.0.1 (https://github.com/bcgsc/transabyss) for multiple *k*-mers (51, 61, 71, 81, 91 and 101). The resulting files from both approaches were merged to generate a single high-confidence transcriptome assembly using EvidentialGene v2022.01.20 (https://sourceforge.net/projects/evidentialgene/). This approach was repeated for six RNA-seq libraries (DRR461683–DRR461688).

Gene model prediction

Gene model prediction was performed using a combination of ab initio and homology-based approaches. First, six transcriptome assemblies were splice-aligned against the genome using PASA v2.3.3 (https://github.com/PASApipeline/PASApipeline). The longest ORFs from these PASA alignments were also extracted using TransDecoder v.5.5.0. Next, we used the ab initio gene prediction tool GeneMark-ES v4.65 (ref. 98) and Braker v2.1.2 (https://github.com/Gaius-Augustus/BRAKER) to produce two separate sets of candidate gene models on the reference genome soft-masked by RepeatMasker as described above. The initial RNA-seq alignments for Braker were produced using STAR aligner v2.7.2b (https://github.com/alexdobin/STAR). The final prediction step in Braker was carried out using Augustus v3.3.2 (https://github.com/Gaius-Augustus/Augustus). Braker was run for six RNA-seq libraries. The homology-based predictor GeMoMa v1.6.1 (ref. 99) was used to produce two additional sets of candidate gene models using gene models from *Arabidopsis thaliana* (Athaliana_167_TAIR10 from Phytozome) and *Populus trichocarpa* (Ptrichocarpa_444_v3.1 from Phytozome). All candidate gene models were then combined to form a single high-quality set of 34,010 gene models using EVidenceModeler v1.1.1 (https://github.com/EVidenceModeler/EVidenceModeler). The completeness of gene models was evaluated with BUSCO v5.3.2 (https://gitlab.com/ezlab/busco). The male gene models were transferred to the female genome assembly using GeMoMa.

Gene annotation

Functional annotation of predicted coding sequences was performed with Trinotate v3.2.1 (https://github.com/Trinotate/Trinotate) with the DIAMOND BLASTP search (v2.0.15, *E* value cut-off = 0.01, https://github.com/bbuchfink/diamond) against the UniProt database downloaded on 21 June 2022, resulting in the GO annotation in 53% (18,037/34,010) of gene models. SignalP v4.1 (ref. 100) and TMHMM v2.0c (ref. 101) were used to predict signal peptides and transmembrane domains, respectively. Coding sequences were used for RPS-BLAST v2.9.0 searches[102] against Pfam-A families[103] (released on 25 November 2020) with an *E* value cut-off of 0.01 to obtain protein domain architectures.

Analysis of synteny and WGD

Syntelogs between homologous *Nepenthes* chromosomes were identified using JCVI v1.2.7 (https://github.com/tanghaibao/jcvi) with the MCscan pipeline. The identification of syntelogs between species was performed using SynMap2 (https://genomevolution.org/wiki/index.php/SynMap2), which internally uses LAST for sequence alignments[104], and then fractionation bias was analysed with FractBias[26]. The reproducible links are as follows: *Vitis* versus *Nepenthes* (https://genomevolution.org/r/1myic) and *Vitis* versus *Coffea* (https://genomevolution.org/r/1myu9). Synonymous divergence of paralogous pairs was obtained using WGDdetector v1.1 (ref. 105).

Analysis of the male-specific region

The male-specific region was delimited on the male genome assembly using re-sequencing data (ddRAD-seq[106]) of 11 male and 10 female individuals sampled from wild populations, including data from previous work[23]. Male versus female read coverage was evaluated by mapping the read data to the genome with bwa mem v0.7.17-r1188, and alignments were filtered with SAMtools v1.12 against non-primary and supplementary alignments, and in the case of paired-end reads, by enforcing the 'properly paired' status. Read depth per sample was counted in genomic windows using bedtools v2.30.0 (https://github.com/arq5x/bedtools2) and compared between the sexes by taking the $\log_2$ of the ratio of male and female normalized depth sums given by

$$r_{\mathrm{window}} = \log_2\left(\left(m_{\mathrm{window}}/m_{\mathrm{total}} + 10^{-6}\right)/\left(f_{\mathrm{window}}/f_{\mathrm{total}} + 10^{-6}\right)\right),$$

where m is the sum of male read counts and f is the sum of female read counts. Window-specific null distributions for this statistic were obtained by 1,000 permutations of the sex assignment of the 21 individuals. Male-specific sequences were further evaluated by counting *k*-mers ($k = 16$) in the ddRAD-seq data of males and females using KMC v3.1.0 (https://github.com/refresh-bio/KMC), requiring *k*-mers to occur at least two times. Observed *k*-mers were classified as male specific if they were present in at least 9 out of 11 male samples and absent in all 10 female samples, using kmc_tools. All possible alignments of such male-specific *k*-mers in the reference genome were found by bwa mem, allowing at most one mismatch. The abundance of such *k*-mer alignments was counted in genomic windows using bedtools. Window-specific null distributions for the abundance of male-specific *k*-mers were generated by 1,000 permutations of the sexes and repetition of the above kmc_tools and alignment procedure.

Analysis of differential expression and GO enrichment

RNA-seq reads were preprocessed, pseudo-aligned to reference and TMM corrected using AMALGKIT v0.6.8.0 (https://github.com/kfuku52/amalgkit), which internally uses fastp, kallisto v0.48.0 (https://github.com/pachterlab/kallisto) and edgeR v3.36.0 (ref. 107). OrthoFinder-based single-copy genes were used for the cross-species TMM correction of expression levels in *A. thaliana* (Brassicaceae), *C. follicularis*, *Dionaea muscipula* (Droseraceae) and *N. gracilis*. In *N. gracilis*, the gene models of the male assembly were used for all samples, including flower samples from other species and cultivars. One flower sample (DRR461757) was removed from the analysis due to a low mapping rate (4.7%; Supplementary Table 4). DEGs between fed and unfed samples were detected using edgeR. While we utilized transcripts per million to analyse gene expression levels in *Nepenthes* flowers (Fig. 3e and Supplementary Fig. 14), we used fragments per kilobase million normalized by the trimmed mean of *M* values (TMM-FPKM) to account for the increased heterogeneity among samples from different species in the cross-species analyses presented in Supplementary Figs. 16 and 18, as well as the subset of this dataset shown in Fig. 3g.

Analysis of *cis*-regulatory elements

The 2 kb sequences upstream of the start codons of *LFY* genes were obtained with SeqKit v2.3.1 (https://github.com/shenwei356/seqkit). The JASPAR CORE v2022 non-redundant set of experimentally defined transcription factor binding sites for plants[49] was used to search *cis*-regulatory elements using FIMO in the MEME Suite v5.4.1 (ref. 108) with the FDR cut-off value of 0.05.

Species tree inference

A total of 1,614 single-copy genes conserved in land plants were searched in the genomes and transcriptomes from the 20 species using BUSCO with the Embryophyta dataset in OrthoDB v10 (embryophyta_odb10)[109]. All 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 Fig. 29). In-frame codon alignments were created by aligning translated protein sequences with MAFFT v7.475 (ref. 110), trimming by ClipKIT v1.3.0 (https://github.com/JLSteenwyk/ClipKIT) and back-translation by CDSKIT v0.10.2 (https://github.com/kfuku52/cdskit). For each single-copy gene, nucleotide and protein ML trees were generated using IQ-TREE v2.2.0.3 (https://github.com/iqtree/iqtree2) with the GTR+R4 model and the LG+R4 model, respectively. The collection of 1,614 single-copy gene trees was subjected to the coalescence-based species tree inference with ASTRAL v.5.7.3 (https://github.com/smirarab/ASTRAL). In addition, concatenated alignments were generated with catfasta2phyml v2018-09-28 (https://github.com/nylander/catfasta2phyml) and used as input to IQ-TREE for nucleotide and protein ML tree inference with the above substitution models. *Amborella trichopoda* was used as the outgroup for rooting.

Divergence time estimation
Divergence time estimation was performed with, as input, the ML species tree and the concatenated codon alignment of single-copy genes. Fossil constraints used in previous studies[111,112] were introduced with NWKIT v0.11.2 (https://github.com/kfuku52/nwkit). Species divergence was estimated with MCMCTREE in the PAML package v4.9 (https://github.com/abacus-gene/paml). Priors and parameters were chosen as described in a tutorial (http://abacus.gene.ucl.ac.uk/software/paml.html). Branch lengths and substitution model parameters were pre-estimated using BASEML with a global clock using the GTR+G model.

Orthogroup classification
Gene sets from the 20 species (Supplementary Table 6) were grouped into 126,597 orthogroups by OrthoFinder v2.5.4 (https://github.com/davidemms/OrthoFinder) with the inferred species tree as a guide tree. For downstream analysis, we extracted 12,123 orthogroups with genes from at least 50% of species and no more than 1,000 genes (Supplementary Table 16).

Gene tree inference
For each orthogroup, the nucleotide ML tree was generated with the GTR+G4 model as described above. The confidence of tree topology was evaluated with ultrafast bootstrapping (–ufboot 1000) with the optimization by hill-climbing nearest neighbour interchange (-bnni). The ML tree was used as the starting gene tree for GeneRax v2.0.4 (https://github.com/BenoitMorel/GeneRax) to generate a rooted, species-tree-aware gene tree. The divergence times of gene trees were inferred by the reconciliation-assisted divergence time estimation using RADTE v0.2.0 (https://github.com/kfuku52/RADTE), which uses a dated species tree as a reference to anchor the divergence time of gene tree nodes[113].

Analysis of gene trees
Branching events in gene trees were categorized into speciation or gene duplication by a species-overlap method[114]. The *dN*/*dS* values were obtained for all branches in the 12,123 orthogroup trees using the *mapdNdS* approach with mapnh v2 (ref. 115) with parameter estimation using IQ-TREE according to a previous report[113]. Nonsynonymous codon substitutions in *LFY-Y* were estimated with IQ-TREE and were mapped to the protein structure with CSUBST v.1.1.0 (https://github.com/kfuku52/csubst)[116].

Data visualization
Phylogenetic trees were visualized using the R package ggtree v3.2.0 (https://github.com/YuLab-SMU/ggtree). General data visualization was performed with Python packages matplotlib v3.6.1 (https://github.com/matplotlib/matplotlib) and seaborn v0.12.0 (https://github.com/mwaskom/seaborn) as well as the R package ggplot2 v3.3.5 (https://github.com/tidyverse/ggplot2). Protein structures were visualized using Open-Source PyMOL v2.4.0 (https://github.com/schrodinger/pymol-open-source).

Reporting summary
Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability
Raw data and results are available at Dryad (https://doi.org/10.5061/dryad.xsj3tx9mj). The *N. gracilis* genome assembly and gene models are available from the DNA Data Bank of Japan (DDBJ) with the accession numbers BSYO01000001 to BSYO01000176. The *N. gracilis* genome assemblies are also available on CoGe (https://genomevolution.org/coge/) (genome ID: male assembly, 61566; female Hi-C assembly, 61892; female syntenic path assembly, 61931) and Dryad (https://doi.org/10.5061/dryad.xsj3tx9mj). DNA and mRNA sequencing reads were deposited to DDBJ (PRJDB15224, PRJDB15738, PRJDB15742 and PRJDB15737) and EBI (PRJEB20488), and the accession numbers are shown in Supplementary Tables 1 and 4. In this study, data were sourced from the following publicly accessible databases: DDBJ (https://www.ddbj.nig.ac.jp/index-e.html), JASPAR (https://jaspar.genereg.net/), NCBI (https://www.ncbi.nlm.nih.gov/), OrthoDB (https://www.orthodb.org/), Pfam (https://www.ebi.ac.uk/interpro/) and UniProt (https://www.uniprot.org/).

Code availability
Scripts used in this study are available at Dryad (https://doi.org/10.5061/dryad.xsj3tx9mj).

References

1. Soltis, D. E. et al. Polyploidy and angiosperm diversification. *Am. J. Bot.* **96**, 336–348 (2009).
2. Van de Peer, Y., Mizrachi, E. & Marchal, K. The evolutionary significance of polyploidy. *Nat. Rev. Genet.* **18**, 411–424 (2017).
3. Amborella Genome Project et al. The *Amborella* genome and the evolution of flowering plants. *Science* **342**, 1241089 (2013).
4. Chanderbali, A. S. et al. *Buxus* and *Tetracentron* genomes help resolve eudicot genome history. *Nat. Commun.* **13**, 643 (2022).
5. Dehal, P. & Boore, J. L. Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biol.* **3**, e314 (2005).
6. Edger, P. P. et al. Subgenome dominance in an interspecific hybrid, synthetic allopolyploid, and a 140-year-old naturally established neo-allopolyploid monkeyflower. *Plant Cell* **29**, 2150–2167 (2017).
7. Li, Z. et al. Patterns and processes of diploidization in land plants. *Annu. Rev. Plant Biol.* **72**, 387–410 (2021).
8. Cheng, F. et al. Gene retention, fractionation and subgenome differences in polyploid plants. *Nat. Plants* **4**, 258–268 (2018).
9. Cross, A. T., Krueger, T. A., Gonella, P. M., Robinson, A. S. & Fleischmann, A. S. Conservation of carnivorous plants in the age of extinction. *Glob. Ecol. Conserv.* **24**, e01272 (2020).
10. Ellison, A. M. & Adamec, L. R. (eds) *Carnivorous Plants: Physiology, Ecology, and Evolution* (Oxford Univ. Press, 2018).
11. Renner, S. S. & Ricklefs, R. E. Dioecy and its correlates in the flowering plants. *Am. J. Bot.* **82**, 596–606 (1995).
12. Walker, J. F. et al. Widespread paleopolyploidy, gene tree conflict, and recalcitrant relationships among the carnivorous Caryophyllales. *Am. J. Bot.* **104**, 858–867 (2017).
13. Yang, Y. et al. Improved transcriptome sampling pinpoints 26 ancient and more recent polyploidy events in Caryophyllales, including two allopolyploidy events. *New Phytol.* **217**, 855–870 (2018).
14. Palfalvi, G. et al. Genomes of the Venus flytrap and close relatives unveil the roots of plant carnivory. *Curr. Biol.* **30**, 2312–2320.e5 (2020).
15. Heubl, G. & Wistuba, A. A cytological study of the genus *Nepenthes* L. (Nepenthaceae). *Sendtnera* **4**, 169–174 (1997).
16. Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A. & Zdobnov, E. M. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol. Biol. Evol.* **38**, 4647–4654 (2021).
17. Minh, B. Q. et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
18. Zhang, C., Rabiee, M., Sayyari, E. & Mirarab, S. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinform.* **19**, 153 (2018).

19. Albert, V. A., Williams, S. E. & Chase, M. W. Carnivorous plants: phylogeny and structural evolution. *Science* **257**, 1491–1495 (1992).
20. Heubl, G., Bringmann, G. & Meimberg, H. Molecular phylogeny and character evolution of carnivorous plant families in Caryophyllales—revisited. *Plant Biol.* **8**, 821–830 (2006).
21. Freund, M. et al. The digestive systems of carnivorous plants. *Plant Physiol.* **190**, 44–59 (2022).
22. Barrett, S. C. H., Yakimowski, S. B., Field, D. L. & Pickup, M. Ecological genetics of sex ratios in plant populations. *Philos. Trans. R. Soc. B* **365**, 2549–2557 (2010).
23. Scharmann, M., Grafe, T. U., Metali, F. & Widmer, A. Sex is determined by XY chromosomes across the radiation of dioecious *Nepenthes* pitcher plants. *Evol. Lett.* **3**, 586–597 (2019).
24. Jaillon, O. et al. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* **449**, 463–467 (2007).
25. Scharmann, M., Wistuba, A. & Widmer, A. Introgression is widespread in the radiation of carnivorous *Nepenthes* pitcher plants. *Mol. Phylogenet. Evol.* **163**, 107214 (2021).
26. Joyce, B. L. et al. FractBias: a graphical tool for assessing fractionation bias following polyploidy. *Bioinformatics* **33**, 552–554 (2017).
27. Yu, Z., Zheng, C., Albert, V. A. & Sankoff, D. Excision dominates pseudogenization during fractionation after whole genome duplication and in gene loss after speciation in plants. *Front. Genet.* **11**, 603056 (2020).
28. Alger, E. I. & Edger, P. P. One subgenome to rule them all: underlying mechanisms of subgenome dominance. *Curr. Opin. Plant Biol.* **54**, 108–113 (2020).
29. Jia, K.-H. et al. SubPhaser: a robust allopolyploid subgenome phasing method based on subgenome-specific *k*-mers. *New Phytol.* **235**, 801–809 (2022).
30. Wang, X. et al. Comparative genomic de-convolution of the cotton genome revealed a decaploid ancestor and widespread chromosomal fractionation. *New Phytol.* **209**, 1252–1263 (2016).
31. Ashman, T.-L. et al. Tree of sex: a database of sexual systems. *Sci. Data* **1**, 140015 (2014).
32. Harkess, A. et al. Sex determination by two Y-linked genes in garden asparagus. *Plant Cell* **32**, 1790–1796 (2020).
33. Schwander, T., Libbrecht, R. & Keller, L. Supergenes and complex phenotypes. *Curr. Biol.* **24**, R288–R294 (2014).
34. Li, J. et al. Integration of genetic and physical maps of the *Primula vulgaris S* locus and localization by chromosome in situ hybridization. *New Phytol.* **208**, 137–148 (2015).
35. Rifkin, J. L. et al. Widespread recombination suppression facilitates plant sex chromosome evolution. *Mol. Biol. Evol.* **38**, 1018–1030 (2021).
36. Potente, G. et al. Comparative genomics elucidates the origin of a supergene controlling floral heteromorphism. *Mol. Biol. Evol.* **39**, msac035 (2022).
37. Akagi, T. et al. Recurrent neo-sex chromosome evolution in kiwifruit. *Nat. Plants* **9**, 393–402 (2023).
38. Horiuchi, A. et al. Ongoing rapid evolution of a post-Y region revealed by chromosome-scale genome assembly of a hexaploid monoecious persimmon (*Diospyros kaki*). *Mol. Biol. Evol.* **40**, msad151 (2023).
39. Yue, J. et al. The origin and evolution of sex chromosomes, revealed by sequencing of the *Silene latifolia* female genome. *Curr. Biol.* **33**, 2504–2514.e3 (2023).
40. Westergaard, M. The mechanism of sex determination in dioecious flowering plants. in *Advances in Genetics* Vol. 9 (ed. Demerec, M.) 217–281 (Academic, 1958).
41. Charlesworth, B. & Charlesworth, D. A model for the evolution of dioecy and gynodioecy. *Am. Nat.* **112**, 975–997 (1978).
42. Zhang, W. et al. Regulation of *Arabidopsis tapetum* development and function by DYSFUNCTIONAL TAPETUM1 (DYT1) encoding a putative bHLH transcription factor. *Development* **133**, 3085–3095 (2006).
43. Zhu, J. et al. *Defective in Tapetal Development and Function 1* is essential for anther development and tapetal function for microspore maturation in *Arabidopsis*. *Plant J.* **55**, 266–277 (2008).
44. Murase, K. et al. MYB transcription factor gene involved in sex determination in *Asparagus officinalis*. *Genes Cells* **22**, 115–123 (2017).
45. Yang, X., Makaroff, C. A. & Ma, H. The *Arabidopsis MALE MEIOCYTE DEATH1* gene encodes a PHD-finger protein that is required for male meiosis. *Plant Cell* **15**, 1281–1295 (2003).
46. Subramanyam, K. & Narayana, L. L. A contribution to the floral anatomy of *Nepenthes khasiana* Hook F. *Proc. Indian Acad. Sci.* **73**, 124–131 (1971).
47. Moyroud, E., Kusters, E., Monniaux, M., Koes, R. & Parcy, F. LEAFY blossoms. *Trends Plant Sci.* **15**, 346–352 (2010).
48. Moyroud, E., Tichtinsky, G. & Parcy, F. The LEAFY floral regulators in angiosperms: conserved proteins with diverse roles. *J. Plant Biol.* **52**, 177–185 (2009).
49. Castro-Mondragon, J. A. et al. JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **50**, D165–D173 (2022).
50. Liu, J. et al. Natural variation in *ARF18* gene simultaneously affects seed weight and silique length in polyploid rapeseed. *Proc. Natl Acad. Sci. USA* **112**, E5123–E5132 (2015).
51. Honma, T. & Goto, K. The *Arabidopsis* floral homeotic gene *PISTILLATA* is regulated by discrete *cis*-elements responsive to induction and maintenance signals. *Development* **127**, 2021–2030 (2000).
52. Theißen, G., Melzer, R. & Rümpler, F. MADS-domain transcription factors and the floral quartet model of flower development: linking plant development and evolution. *Development* **143**, 3259–3271 (2016).
53. Sather, D. N., Jovanovic, M. & Golenberg, E. M. Functional analysis of B and C class floral organ genes in spinach demonstrates their role in sexual dimorphism. *BMC Plant Biol.* **10**, 46 (2010).
54. Tsukaya, H. Comparative leaf development in angiosperms. *Curr. Opin. Plant Biol.* **17**, 103–109 (2014).
55. Renner, T. & Specht, C. D. A sticky situation: assessing adaptations for plant carnivory in the Caryophyllales by means of stochastic character mapping. *Int. J. Plant Sci.* **172**, 889–901 (2011).
56. Bemm, F. et al. Venus flytrap carnivorous lifestyle builds on herbivore defense strategies. *Genome Res.* **26**, 1–14 (2016).
57. Iosip, A. L. et al. The Venus flytrap trigger hair-specific potassium channel KDM1 can reestablish the K^+ gradient required for hapto-electric signaling. *PLoS Biol.* **18**, e3000964 (2020).
58. Procko, C. et al. Stretch-activated ion channels identified in the touch-sensitive structures of carnivorous Droseraceae plants. *eLife* **10**, e64250 (2021).
59. Klepikova, A. V., Kasianov, A. S., Gerasimov, E. S., Logacheva, M. D. & Penin, A. A. A high resolution map of the *Arabidopsis thaliana* developmental transcriptome based on RNA-seq profiling. *Plant J.* **88**, 1058–1070 (2016).
60. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11**, 2010–2011 (2010).
61. Pavlovič, A., Masarovičová, E. & Hudák, J. Carnivorous syndrome in Asian pitcher plants of the genus *Nepenthes*. *Ann. Bot.* **100**, 527–536 (2007).
62. Pavlovič, A. Photosynthetic characterization of Australian pitcher plant *Cephalotus follicularis*. *Photosynthetica* **49**, 253–258 (2011).

63. Yilamujiang, A., Reichelt, M. & Mithöfer, A. Slow food: insect prey and chitin induce phytohormone accumulation and gene expression in carnivorous *Nepenthes* plants. *Ann. Bot.* **118**, 369–375 (2016).
64. Glazebrook, J. Contrasting mechanisms of defense against biotrophic and necrotrophic pathogens. *Annu. Rev. Phytopathol.* **43**, 205–227 (2005).
65. Erb, M., Meldau, S. & Howe, G. A. Role of phytohormones in insect-specific plant reactions. *Trends Plant Sci.* **17**, 250–259 (2012).
66. Pavlovič, A. & Mithöfer, A. Jasmonate signalling in carnivorous plants: copycat of plant defence mechanisms. *J. Exp. Bot.* **70**, 3379–3389 (2019).
67. Capó-Bauçà, S., Font-Carrascosa, M., Ribas-Carbó, M., Pavlovič, A. & Galmés, J. Biochemical and mesophyll diffusional limits to photosynthesis are determined by prey and root nutrient uptake in the carnivorous pitcher plant *Nepenthes × ventrata*. *Ann. Bot.* **126**, 25–37 (2020).
68. Durand, D. & Hoberman, R. Diagnosing duplications—can it be done? *Trends Genet.* **22**, 156–164 (2006).
69. Yanai, I. et al. Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics* **21**, 650–659 (2005).
70. Kryuchkova-Mostacci, N. & Robinson-Rechavi, M. A benchmark of gene expression tissue-specificity metrics. *Brief. Bioinform.* **44**, bbw008 (2016).
71. Callard, D., Axelos, M. & Mazzolini, L. Novel molecular markers for late phases of the growth cycle of *Arabidopsis thaliana* cell-suspension cultures are expressed during organ senescence. *Plant Physiol.* **112**, 705–715 (1996).
72. Lee, K., Zawadzka, A., Czarnocki, Z., Reiter, R. J. & Back, K. Molecular cloning of melatonin 3-hydroxylase and its production of cyclic 3-hydroxymelatonin in rice (*Oryza sativa*). *J. Pineal Res.* **61**, 470–478 (2016).
73. Choi, G.-H. & Back, K. Cyclic 3-hydroxymelatonin exhibits diurnal rhythm and cyclic 3-hydroxymelatonin overproduction increases secondary tillers in rice by upregulating *MOC1* expression. *Melatonin Res.* **2**, 120–138 (2019).
74. Lee, H. Y. & Back, K. The antioxidant cyclic 3-hydroxymelatonin promotes the growth and flowering of *Arabidopsis thaliana*. *Antioxidants* **11**, 1157 (2022).
75. Chia, T. F., Aung, H. H., Osipov, A. N., Goh, N. K. & Chia, L. S. Carnivorous pitcher plant uses free radicals in the digestion of prey. *Redox Rep.* **9**, 255–261 (2004).
76. Hatano, N. & Hamada, T. Proteomic analysis of secreted protein induced by a component of prey in pitcher fluid of the carnivorous plant *Nepenthes alata*. *J. Proteomics* **75**, 4844–4852 (2012).
77. Fukushima, K. et al. Genome of the pitcher plant *Cephalotus* reveals genetic changes associated with carnivory. *Nat. Ecol. Evol.* **1**, 0059 (2017).
78. Wal, A., Staszek, P., Pakula, B., Paradowska, M. & Krasuska, U. ROS and RNS alterations in the digestive fluid of *Nepenthes × ventrata* trap at different developmental stages. *Plants* **11**, 3304 (2022).
79. Buchfink, B., Reuter, K. & Drost, H.-G. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat. Methods* **18**, 366–368 (2021).
80. Freeling, M. Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition. *Annu. Rev. Plant Biol.* **60**, 433–453 (2009).
81. Conant, G. C. & Wolfe, K. H. Turning a hobby into a job: how duplicated genes find new functions. *Nat. Rev. Genet.* **9**, 938–950 (2008).
82. Hedrich, R. & Fukushima, K. On the origin of carnivory: molecular physiology and evolution of plants on an animal diet. *Annu. Rev. Plant Biol.* **72**, 133–153 (2021).
83. Pannell, J. R. & Jordan, C. Y. Evolutionary transitions between hermaphroditism and dioecy in animals and plants. *Annu. Rev. Ecol. Evol. Syst.* **53**, 183–201 (2022).
84. Cronk, Q. & Müller, N. A. Default sex and single gene sex determination in dioecious plants. *Front. Plant Sci.* **11**, 1162 (2020).
85. Renner, S. S. & Müller, N. A. Plant sex chromosomes defy evolutionary models of expanding recombination suppression and genetic degeneration. *Nat. Plants* **7**, 392–402 (2021).
86. Albert, V. A., Oppenheimer, D. G. & Lindqvist, C. Pleiotropy, redundancy and the evolution of flowers. *Trends Plant Sci.* **7**, 297–301 (2002).
87. Bringmann, G., Rischer, H., Schlauer, J. & Aké Assi, L. *In vitro* propagation of *Ancistrocladus abbreviatus* Airy Shaw (Ancistrocladaceae). *Plant Cell Tissue Organ Cult.* **57**, 71–73 (1999).
88. Fukushima, K., Narukawa, H., Palfalvi, G. & Hasebe, M. A discordance of seasonally covarying cues uncovers misregulated phenotypes in the heterophyllous pitcher plant *Cephalotus follicularis*. *Proc. R. Soc. B* **288**, 20202568 (2021).
89. Bringmann, G., Rübenacker, M., Jansen, J. R., Scheutzow, D. & Aké Assi, L. On the structure of the Dioncophyllaceae alkaloids dioncophylline A ('triphyophylline') and '*O*-methyl-triphyophylline'. *Tetrahedron Lett.* **31**, 639–642 (1990).
90. Bringmann, G. & Rischer, H. In vitro propagation of the alkaloid-producing rare African liana, *Triphyophyllum peltatum* (Dioncophyllaceae). *Plant Cell Rep.* **20**, 591–595 (2001).
91. Murashige, T. & Skoog, F. A revised medium for rapid growth and bio assays with tobacco tissue cultures. *Physiol. Plant.* **15**, 473–497 (1962).
92. Michael, T. P. et al. High contiguity *Arabidopsis thaliana* genome assembly with a single nanopore flow cell. *Nat. Commun.* **9**, 541 (2018).
93. Putnam, N. H. et al. Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. *Genome Res.* **26**, 342–350 (2016).
94. Haug-Baltzell, A., Stephens, S. A., Davey, S., Scheidegger, C. E. & Lyons, E. SynMap2 and SynMap3D: web-based whole-genome synteny browsers. *Bioinformatics* **33**, 2197–2198 (2017).
95. Novák, P., Neumann, P. & Macas, J. Global analysis of repetitive DNA from unassembled sequence reads using RepeatExplorer2. *Nat. Protoc.* **15**, 3745–3776 (2020).
96. Neumann, P., Novák, P., Hoštáková, N. & Macas, J. Systematic survey of plant LTR-retrotransposons elucidates phylogenetic relationships of their polyprotein domains and provides a reference for element classification. *Mob. DNA* **10**, 1 (2019).
97. Lopez-Delisle, L. et al. pyGenomeTracks: reproducible plots for multivariate genomic datasets. *Bioinformatics* **37**, 422–423 (2021).
98. Brůna, T., Lomsadze, A. & Borodovsky, M. GeneMark-EP+: eukaryotic gene prediction with self-training in the space of genes and proteins. *NAR Genom. Bioinform.* **2**, lqaa026 (2020).
99. Keilwagen, J., Hartung, F., Paulini, M., Twardziok, S. O. & Grau, J. Combining RNA-seq data and homology-based gene prediction for plants, animals and fungi. *BMC Bioinform.* **19**, 189 (2018).
100. Petersen, T. N., Brunak, S., von Heijne, G. & Nielsen, H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* **8**, 785–786 (2011).
101. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305**, 567–580 (2001).
102. Camacho, C. et al. BLAST+: architecture and applications. *BMC Bioinform.* **10**, 421 (2009).
103. El-Gebali, S. et al. The Pfam protein families database in 2019. *Nucleic Acids Res.* **47**, D427–D432 (2019).

104. Kiełbasa, S. M., Wan, R., Sato, K., Horton, P. & Frith, M. C. Adaptive seeds tame genomic sequence comparison. *Genome Res.* **21**, 487–493 (2011).
105. Yang, Y., Li, Y., Chen, Q., Sun, Y. & Lu, Z. WGDdetector: a pipeline for detecting whole genome duplication events using the genome or transcriptome annotations. *BMC Bioinform.* **20**, 75 (2019).
106. Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S. & Hoekstra, H. E. Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE* **7**, e37135 (2012).
107. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
108. Bailey, T. L., Johnson, J., Grant, C. E. & Noble, W. S. The MEME Suite. *Nucleic Acids Res.* **43**, W39–W49 (2015).
109. Kriventseva, E. V. et al. OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic Acids Res.* **47**, D807–D811 (2019).
110. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
111. Zhang, G.-Q. et al. The *Apostasia* genome and the evolution of orchids. *Nature* **549**, 379–383 (2017).
112. Suetsugu, K. et al. Transcriptomic heterochrony and completely cleistogamous flower development in the mycoheterotrophic orchid *Gastrodia*. *New Phytol.* **237**, 323–338 (2023).
113. Fukushima, K. & Pollock, D. D. Amalgamated cross-species transcriptomes reveal organ-specific propensity in gene expression evolution. *Nat. Commun.* **11**, 4459 (2020).
114. Huerta-Cepas, J. et al. The human phylome. *Genome Biol.* **8**, 934–941 (2007).
115. Guéguen, L. & Duret, L. Unbiased estimate of synonymous and nonsynonymous substitution rates with nonstationary base composition. *Mol. Biol. Evol.* **35**, 734–742 (2018).
116. Fukushima, K. & Pollock, D. D. Detecting macroevolutionary genotype–phenotype associations using error-corrected rates of protein convergence. *Nat. Ecol. Evol.* **7**, 155–170 (2023).
117. Vollger, M. R., Kerpedjiev, P., Phillippy, A. M. & Eichler, E. E. StainedGlass: interactive visualization of massive tandem repeat structures with identity heatmaps. *Bioinformatics* **38**, 2049–2051 (2022).

Acknowledgements

We acknowledge the following sources for funding: the Sofja Kovalevskaja programme of the Alexander von Humboldt Foundation (to K.F.), a Human Frontier Science Program Young Investigators grant RGY0082/2021 (to K.F. and T.R.), Deutsche Forschungsgemeinschaft research grants (454506241 to K.F., 415282803 to R.H. and 699/14-2 to G.B.), JSPS KAKENHI JP20H05909 (to K.S.), United States National Science Foundation grants (1442190 to V.A.A. and 2030871 to T.R. and V.A.A.), United States Department of Agriculture-National Institutes of Food and Agriculture grant 2019-67012-37587 (to K.J.G.) and research-leave funding from the School of Biological Sciences, Nanyang Technological University, to V.A.A. and C.L. that supported *N. gracilis* collecting and sequencing. M. Niissalo, J. H. Ang, Q. Y. Tan and G. Khew (Singapore Botanic Gardens, National Parks Board, Singapore) are thanked for their assistance with collecting and processing the *N. gracilis* material for long-read DNA sequencing. We thank J. Danz, H. Doi and L. Steffen for providing additional plant materials of *Nepenthes*, T. Winkelmann for propagating in vitro cultures of *T. peltatum* and J. Rothenhöfer for plant cultivation. Computations were partially performed on the National Institute of Genetics supercomputer, the Data Integration and Analysis Facility at the National Institute for Basic Biology, the Erlangen National High Performance Computing Center, the University at Buffalo Center for Computational Research and the High-Performance Computing Clusters at the University of Würzburg.

Author contributions

V.A.A. and K.F. conceptualized the project. F.S., G.V., G.B., Y.W.L., C.L., V.A.A. and K.F. acquired materials. Y.W.L., C.L., T.P.M., S.M. and K.S. conducted and/or oversaw DNA extraction and sequencing. F.S., E.C., T.P.M., V.A.A. and K.F. conducted genome assembly. F.S., L.C., M.F., T.W. and K.F. conducted RNA extraction. S.R. conducted gene model prediction. F.S., A.M., M.R., M.S., V.A.A. and K.F. analysed the results. F.S., M.S., V.A.A. and K.F. wrote the manuscript with input from all authors. K.J.G. and T.R. contributed to improving the manuscript. D.B., R.H., V.A.A. and K.F. supervised the project and coordinated collaborations. All authors reviewed the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41477-023-01562-2.

Correspondence and requests for materials should be addressed to Victor A. Albert or Kenji Fukushima.

Peer review information *Nature Plants* thanks Takashi Akagi, Andrej Pavlovic and Feng Cheng for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

[1]Institute for Molecular Plant Physiology and Biophysics, University of Würzburg, Würzburg, Germany. [2]Institute for Biochemistry and Biology (IBB), University of Potsdam, Potsdam, Germany. [3]School of Biological Sciences, Nanyang Technological University, Singapore, Singapore. [4]Organismal and Evolutionary Biology Research Programme, Faculty of Biological and Environmental Sciences, University of Helsinki, Helsinki, Finland. [5]Department of Chromosome Biology, Max Planck Institute for Plant Breeding Research, Cologne, Germany. [6]Institute of Organic Chemistry, University of Würzburg, Am Hubland, Würzburg, Germany. [7]Department of Biological Sciences, University at Buffalo, Buffalo, NY, USA. [8]Singapore Botanic Gardens, National Parks Board, Singapore, Singapore. [9]Department of Plant Biology & W.K. Kellogg Biological Station & Program in Ecology, Evolution, and Behavior, Michigan State University, Hickory Corners, MI, USA. [10]Department of Entomology, The Pennsylvania State University, University Park, PA, USA. [11]Riken Center for Sustainable Resource Science, Yokohama, Japan. [12]Botanical Garden, University of Würzburg, Würzburg, Germany. [13]Plant Molecular and Cellular Biology Laboratory, The Salk Institute for Biological Studies, La Jolla, CA, USA. [14]These authors contributed equally: Franziska Saul, Mathias Scharmann. ✉e-mail: vaalbert@buffalo.edu; kenji.fukushima@uni-wuerzburg.de

nature portfolio

Corresponding author(s): Kenji Fukushima

Last updated by author(s): Sep 12, 2023

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	no software was used.
Data analysis	Guppy v2.0, seqtk v1.2, WTDBG2 v2.2, minimap2 v0.2, Racon v1.3.1, Pilon v1.18, Purge Haplotigs v1.0.0, bwa mem v0.7.17, SAMtools v1.3, Juicer v1.6, 3D-DNA v170123, fastp v0.20.0, TransDecoder v5.5.0, CDSKIT v0.9.1, RepeatMasker v4.0.9, RepeatModeler v2.0, EDTA v2.1.0, pyGenomeTracks v3.6, StainedGlass v0.4, HiGlass v0.10.1, StringTie v2.1.4, HiSat2 v2.1.0, Trinity v2.8.5, TransAbyss v2.0.1, EvidentialGene v2022.01.20, PASA v2.3.3, GeneMark-ES v4.65, Braker v2.1.2, STAR v2.7.2b, Augustus v3.3.2, GeMoMa v1.6.1, EVidenceModeler v1.1.1, BUSCO v5.3.2, Trinotate v3.2.1, SignalP v4.1, TMHMM v2.0c, RPS-BLAST v2.9.0, JCVI v1.2.7, WGDdetector v1.1, bwa mem v0.7.17-r1188, SAMtools v1.12, bedtools v2.30.0, KMC v3.1.0, AMALGKIT v0.6.8.0, kallisto v0.48.0, edgeR v3.36.0, SeqKit v2.3.1, MEME Suite v5.4.1, MAFFT v7.475, ClipKIT v1.3.0, CDSKIT v0.10.2, IQ-TREE v2.2.0.3, ASTRAL v.5.7.3, catfasta2phyml v2018-09-28, NWKIT v0.11.2, MCMCTREE in PAML v4.9, OrthoFinder v2.5.4, GeneRax v2.0.4, RADTE v0.2.0, mapnh v2, CSUBST v.1.1.0, ggtree v3.2.0, matplotlib v3.6.1, seaborn v0.12.0, ggplot2 v3.3.5, Open-Source PyMOL v2.4.0, SubPhaser v1.2.5, Custom codes available at https://doi.org/10.5061/dryad.xsj3tx9mj

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Raw data and results are available at Dryad (https://doi.org/10.5061/dryad.xsj3tx9mj). The N. gracilis genome assembly and gene models are available from the DNA Data Bank of Japan (DDBJ) with the accession numbers BSYO01000001 to BSYO01000176 The N. gracilis genome assemblies are also available on CoGe (https://genomevolution.org/coge/) (genome ID: male assembly, 61566; female Hi-C assembly, 61892; female syntenic path assembly, 61931) and Dryad (https://doi.org/10.5061/dryad.xsj3tx9mj). DNA and mRNA sequencing reads were deposited to DDBJ (PRJDB15224, PRJDB15738, PRJDB15742, and PRJDB15737) and EBI (PRJEB20488), and the accession numbers are shown in Supplementary Table 1 and Supplementary Table 4. In this study, data were sourced from the following publicly accessible databases: DDBJ (https://www.ddbj.nig.ac.jp/index-e.html), JASPAR (https://jaspar.genereg.net/), NCBI (https://www.ncbi.nlm.nih.gov/), OrthoDB (https://www.orthodb.org/), Pfam (https://www.ebi.ac.uk/interpro/), and UniProt (https://www.uniprot.org/).

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	N/A
Reporting on race, ethnicity, or other socially relevant groupings	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Differential expression analysis of RNA-seq data was performed with triplicate samples (or more) to meet the accepted standards in the field (Conesa et al., 2016, https://doi.org/10.1186/s13059-016-0881-8).
Data exclusions	In the heterologous RNA-seq mapping analysis of flower samples, one sample (DRR461757) was excluded from the analysis due to its low mapping rate of 4.7% (see Supplementary Table 4).
Replication	A single RNA-seq experiment was performed for each analysis, and it included a minimum of three biological replicates.
Randomization	In the feeding experiments, individuals of Nepenthes gracilis were randomly assigned to either the control group or the feeding treatment group. In the flower RNA-seq experiments, no randomization was performed because we collected all available flowers that were spontaneously produced by cultivated plants.
Blinding	Blinding was not implemented in this study due to the nature of the interventions and the logistical constraints involved in executing them.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☐	☒ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Dual use research of concern

Policy information about dual use research of concern

Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

No	Yes
☒	☐ Public health
☒	☐ National security
☒	☐ Crops and/or livestock
☒	☐ Ecosystems
☒	☐ Any other significant area

Experiments of concern

Does the work involve any of these experiments of concern:

No	Yes
☒	☐ Demonstrate how to render a vaccine ineffective
☒	☐ Confer resistance to therapeutically useful antibiotics or antiviral agents
☒	☐ Enhance the virulence of a pathogen or render a nonpathogen virulent
☒	☐ Increase transmissibility of a pathogen
☒	☐ Alter the host range of a pathogen
☒	☐ Enable evasion of diagnostic/detection modalities
☒	☐ Enable the weaponization of a biological agent or toxin
☒	☐ Any other potentially harmful combination of experiments and agents