Contrasting centromere structures in carnivorous sundews uncover a dynamic evolutionary trajectory shaped by polyploidy and chromosomal remodeling

Laura Ávila Robledillo[1,$], Steven J. Fleck[2,*], Jonathan Kirshner[2,*], Dirk Becker[3], Aaryan Bhatia[1], Gerhard Bringmann[4], Jordan R. Brock[5], Daniela Drautz-Moses[6], Matthias Freund[3], Rainer Hedrich[7], Luis Herrera-Estrella[8], Enrique Ibarra-Laclette[8], Ines Kreuzer[3], Tianying Lan[2], Sachiko Masuda[10], Martín Mata-Rosas[11], Todd P. Michael[12,13,14,15], Héctor Montero[16], Sitaram Rajaraman[17,18], Michaela Richter[2], David Sankoff[19], Stephan C. Schuster[6,17], Ken Shirasu[10,20,] Sonja Trebing[3], Yves Van de Peer[21,22,23,24], Gerd Vogg[25], Tan Qiao Wen[17], Yue Zhang[19], Chunfang Zheng[19], Kenji Fukushima[16,#], Jarkko Salojärvi[17,18,#], André Marques[1,27,#], Victor A. Albert[2,#]

[$]First author

[*]These authors contributed equally

[#]Corresponding authors. For correspondence: vaalbert@buffalo.edu (V. A. Albert), amarques@mpipz.mpg.de (A. Marques), Jarkko@ntu.edu.sg (J. Salojärvi), kenji.fukushima@nig.ac.jp (K. Fukushima)

1. Department of Chromosome Biology, Max Planck Institute for Plant Breeding Research, Cologne, Germany
2. Department of Biological Sciences, University at Buffalo, Buffalo, New York, USA
3. Julius-von-Sachs-Institute for Biosciences, Department of Molecular Plant Physiology and Biophysics, University of Würzburg, Julius-von-Sachs-Platz 2, 97082 Würzburg, Germany
4. Institute of Organic Chemistry, Am Hubland, 97074 Würzburg, Germany
5. Washington University, CB 1137, One Brookings Drive, St. Louis, MO 63130-4899, USA
6. Singapore Centre for Environmental Life Sciences Engineering, Nanyang Technological University, Singapore 637551
7. Faculty of Synthetic Biology, Shenzhen University of Advanced Technology (SUAT), China
8. Texas Tech University, Institute of Genomics for Crop Abiotic Stress Tolerance, 1006 Canton Ave, Lubbock, TX 79409
9. Red de Estúdios Moleculares Avanzados, Instituto de Ecología, A.C. (INECOL), Xalapa 91073, Veracruz, Mexico
10. RIKEN Center for Sustainable Resource Science (CSRS), Suehiro-cho 1-7-22 Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
11. Red Manejo Biotecnológico de Recursos, Instituto de Ecología, A.C., Carretera Antigua a Coatepec No.351, Col. El Haya, Xalapa, Veracruz, 91073, Mexico
12. The Plant Molecular and Cellular Biology Laboratory, The Salk Institute for Biological Studies, La Jolla, CA 92037, US
13. Department of Cell and Developmental Biology, School of Biological Sciences, University of California, San Diego, La Jolla, CA, 92093, USA
14. Center for Marine Biotechnology and Biomedicine, University of California, San Diego, La Jolla, CA, 92093, USA

15. Department of Science and Conservation, San Diego Botanical Garden, Encinitas, CA, 92024, USA
16. National Institute of Genetics, Center for Frontier Research, Plant Evolution Laboratory, 1111 Yata, Mishima, Shizuoka 411-8540, Japan
17. School of Biological Sciences, Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798
18. Organismal and Evolutionary Biology Research Program, Faculty of Biological and Environmental Sciences, and Viikki Plant Science Centre, University of Helsinki, Helsinki, Finland
19. Department of Mathematics and Statistics, University of Ottawa, Ottawa, Ontario, Canada
20. Graduate school of Science, The University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo, 113-8654, Japan
21. Department of Plant Biotechnology and Bioinformatics, Ghent University, 9052 Ghent, Belgium
22. VIB Center for Plant Systems Biology, VIB, 9052 Ghent, Belgium
23. College of Horticulture, Academy for Advanced Interdisciplinary Studies, Nanjing Agricultural University, 210095 Nanjing, China
24. Department of Biochemistry, Genetics and Microbiology, University of Pretoria, Pretoria, 0028 South Africa
25. Julius-von-Sachs-Institut für Biowissenschaften, Botanischer Garten der Universität Würzburg, Julius-von-Sachs-Platz 4, D - 97082 Würzburg
26. Cluster of Excellence on Plant Sciences (CEPLAS), Heinrich-Heine University, Düsseldorf, Germany

Author ORCIDs, where available:

Laura Ávila Robledillo - 0000-0002-7617-2613

Steven J. Fleck - 0000-0001-5370-2258

Jonathan Kirshner - 0009-0008-5521-0351

Aaryan Bhatia - 0000-0001-6973-8052

Dirk Becker - 0000-0003-0723-4160

Gerhard Bringmann - 0000-0002-3583-5935

Jordan Brock - 0000-0001-6231-1435

Chloe P. Drummond - 0000-0001-8086-9213

Matthias Freund - 0000-0002-5683-6275

Rainer Hedrich - 0000-0003-3224-1362

Luis Herrera-Estrella - 0000-0001-7936-3856

Enrique Ibarra-Laclette - 0000-0003-1441-1938

Ines Kreuzer - 0000-0002-8772-783X

Sachiko Masuda - 0000-0003-3733-0030

Martín Mata-Rosas - 0000-0001-7868-3305

Todd P. Michael - 0000-0001-6272-2875
Héctor Montero - 0000-0001-8590-6394
Sitaram Rajaraman - 0000-0001-5171-3578
Michaela Richter - 0000-0001-7537-4902
David Sankoff - 0000-0001-8415-5189
Ken Shirasu - 0000-0002-0349-3870
Sonja Trebing - 0000-0001-5190-4497
Yves Van de Peer - 0000-0003-4327-3730
Gerd Vogg - 0009-0000-9766-9189
Tan Qiao Wen - 0000-0001-5690-7478
Kenji Fukushima -0000-0002-2353-92747
Jarkko Salojärvi - 0000-0002-4096-6278
André Marques - 0000-0002-9567-2576
Victor A. Albert - 0000-0002-0262-826X

Abstract

Sundews (*Drosera*) are carnivorous plants that capture prey using mucilage-producing glands with both secretory and absorptive functions. Within a remarkably diverse clade that includes the Venus flytrap (*Dionaea*), East Asian pitcher plants (*Nepenthes*), the African carnivore *Triphyophyllum* (which bears three leaf types, one sundew-like) and the non-carnivorous vine *Ancistrocladus*, sundews show striking chromosomal variation. We present chromosome-scale genome assemblies for *Drosera regia* and *D. capensis*, comparing them with genomes of relatives. Complex allopolyploidies are revealed that prohibit depiction as bifurcating phylogenies. Characterization of centromere sequences demonstrated holocentric and monocentric chromosomes in *D. regia* and *D. capensis*, respectively, with corresponding structural divergence. Chromosomal fusions in *D. regia* likely trace to a shared allohexaploid ancestor shared with the dodecaploid and highly rearranged *D. capensis* genome. This comparative framework highlights genomic plasticity in a carnivorous plant lineage that offers a valuable model for studying morphological and ecophysiological diversity, polyploidy and centromere variation in angiosperms.

Introduction

Flowering plants have diversified into nearly every terrestrial habitat by evolving a wide array of physiological and morphological strategies to, among other adaptive features, acquire essential nutrients. Among the most striking deviations from typical angiosperm organization are carnivorous plants, which have repeatedly evolved the ability to trap and digest animal prey to supplement their nutrient intake, particularly nitrogen and phosphorus, in resource-poor environments[1,2]. The transition to carnivory involves dramatic shifts in leaf development[3,4], repurposing of defense responses[5,6], nutrient transport[7,8], and enzymatic secretion[8,9], including co-opting genetic pathways typically associated with plant immunity and senescence[10-13].

Carnivory has evolved independently at least ten times across flowering plants, and within five different angiosperm orders[2]. These convergent origins have given rise to an impressive diversity of trap morphologies, including snap-traps, suction bladders, and pitfall-trapping pitchers[2]. Particularly striking is the convergent evolution of pitcher traps in three distantly related lineages that diverged over 100 million years ago, being separated by vast numbers of non-carnivorous taxa[1]. Two major carnivorous clades within Lamiales (e.g., *Utricularia*, *Pinguicula*, *Genlisea*) and Caryophyllales (e.g., *Drosera*, *Dionaea*, *Nepenthes*), respectively, each encompass a striking diversity of trap morphologies[1,2], exemplifying some of the most profound evolutionary innovations among angiosperms.

Within Caryophyllales, the genus *Drosera* (sundews) comprises over 200 species that use a specialized adhesive flypaper-trapping strategy[14]. Their leaves are covered with mucilage-secreting glandular tentacles that immobilize and digest insect prey. Despite this basic structural uniformity in the genus, *Drosera* exhibits considerable ecological and morphological diversity. Moreover, closely allied genera include the snap-trapping *Dionaea* (Venus flytrap) and *Aldrovanda* (the waterwheel plant), the morphologically divergent flypaper trappers *Triphyophyllum*[15-17] and *Drosophyllum*[18,19], and a group of secondarily non-carnivorous lianas (*Ancistrocladus*, *Habropetalum*, *Dioncophyllum*)[15], which together form an extraordinary evolutionary radiation spanning multiple functional transitions.

Beyond their ecological specializations, sundews are remarkable for their reported diversity in chromosomal and genomic architecture[20-23]. Cytogenetic studies have proposed that the genus contains both monocentric chromosomes (with a single, localized centromere) and holocentric chromosomes (in which centromeric function is distributed in patches along entire chromatids)[20,22,23]. However, these claims have remained contentious due to the minute size of sundew chromosomes, which complicates cytological observations, and a historical lack of genomic sequence data to validate centromere structure and function.

Recent work across diverse eukaryotes has underscored the central role of centromere organization in shaping genome architecture[24-26]. Features such as the distribution of genes and transposable elements, patterns of tandem repeats, histone modifications, and crossover localization during meiosis all contribute to large-scale chromosomal behavior[24-27]. In plants with holocentric chromosomes, chromosomal breakage and fusion events can be tolerated without loss of centromere function, potentially accelerating genome restructuring and enabling novel evolutionary trajectories[24-26]. Despite this, the evolutionary and mechanistic consequences of holocentricity remain poorly characterized in flowering plants.

Here, we present high-quality, chromosome-scale genome assemblies for two *Drosera* species: *D. regia*, a morphologically and phylogenetically distinct taxon usually considered sister to all other sundews[28,29], and *D. capensis*, a more typical, phylogenetically derived and widely cultivated species[30,31]. These species differ not only in evolutionary placement but also in

cytogenetic properties, with *D. regia* previously proposed to possess monocentric chromosomes[20] and *D. capensis* reported to be holocentric[28,32,33]. They also differ in ploidy level, with *D. regia* inferred to be hexaploid (6*x*) and *D. capensis* tetraploid (4*x*)[34], although these estimates have not previously been confirmed with genome-scale data.

Our assemblies reveal that both *Drosera* genomes are relatively small (<300 Mbp) and shaped by ancient polyploidy. Contrary to prior reports, our cytogenomic analyses affirm that *D. regia* possesses holocentric, as opposed to monocentric chromosomes, suggesting that this feature may be more widespread in the genus than previously recognized, and that the species' post-hexaploid diploidization involved extensive chromosomal fusion events. In contrast, *D. capensis* unexpectedly displays monocentric chromosomes and a dodecaploid, ancestrally hexaploid genome characterized by substantial structural rearrangement and disrupted synteny relative to *D. regia* and other carnivorous Caryophyllales, including *Dionaea*, *Triphyophyllum*, *Ancistrocladus* and *Nepenthes*.

To further explore genomic architecture within the genus, we analyzed repetitive DNA content in both species and generated short-read data for 10 additional *Drosera* taxa. These data reveal lineage-specific variation in repeat dynamics, suggesting that centromere structure may influence patterns of genome evolution even among closely related species in a single genus. Together, our findings illuminate the complex evolutionary history of *Drosera* and its close relatives, providing new insight into the interplay between centromere biology, polyploidy, and genome architecture in one of the most morphologically and genomically dynamic plant lineages.

Chromosome-scale sundew genomes

We generated reference-quality genome assemblies for *D. regia* and *D. capensis* using a combination of PacBio SMRT and Illumina DNA sequencing, and Hi-C chromosome conformation capture (**Extended Data Table 1**). The number of large pseudochromosomal scaffolds in the *D. capensis* and *D. regia* assemblies matched their known haploid chromosome numbers, n=20 and n=17, respectively[34] (**Fig. 1**). Our chromosome-level assemblies were annotated using RNA-seq data from *D. capensis* and other species (see Methods). Furthermore, we annotated 32,230 and 32,090 high-confidence gene models along the 17 and 20 pseudomolecules of *D. regia* and *D. capensis*, respectively, with BUSCO scores of 96.7% for *D. regia* and 94.8% for *D. capensis* (**Extended Data Table 1**).

Genome architecture reveals contrasting centromere organization in sundews

Several species of *Drosera* have been reported to have holocentric chromosomes with centromeres decentralized over the length of the chromosome[15-18]. We combined Hi-C contact maps with genome-wide repeat annotations to infer the centromeric organization of the *D. capensis* and *D. regia* chromosomes (**Fig. 1a–h**). The *D. capensis* Hi-C map displayed a classic monocentric architecture, showing distinct A (euchromatin) and B (heterochromatin) compartments, including some degree of a telomere-to-centromere interaction axis (**Fig. 1a–b**). Correspondingly, repeat profiling of *D. capensis* revealed a single, size-restricted region of satellite DNA (satDNA) probably marking the centromere, flanked by a high density of transposable elements in the pericentromeric region, tapering toward the telomeres, while genes were concentrated on the distal chromosome arms (**Fig. 1c–d**). In contrast, the *D. regia* chromosomes lacked large-scale compartmentalization and showed no telomere-to-centromere

interaction axis in their Hi-C maps (**Fig. 1e–f**), suggesting a non-monocentric structure. The *D. regia* repeat landscape mirrored this difference: satellite DNA occurred as multiple short arrays dispersed along the chromosomes, and both transposable elements and genes were uniformly distributed, without the typical pericentromeric enrichment (**Fig. 1g–h**). These results are in close agreement with recent reports for other holocentric plants. As recently reported by Hofstatter, et al.[20], the concept of chromosome arms does not apply to holocentric species; however, as centromeric function is ubiquitous, holocentricity can be inferred at the chromatin contact pattern level observed in Hi-C matrices, and by the spatial organization of the different types of sequences along chromosomes. At present, our inference of holocentricity in *D. regia* is not matched by other previously-assembled genomes of carnivorous Caryophyllales under study; for example, the Hi-C contact map of *D. muscipula* reflected a chromosomal architecture comparable to that of *D. capensis*, in contrast to previous non-genomic research that suggested holocentricity for *Dionaea muscipula*[16] (**Extended data Fig. 1**).

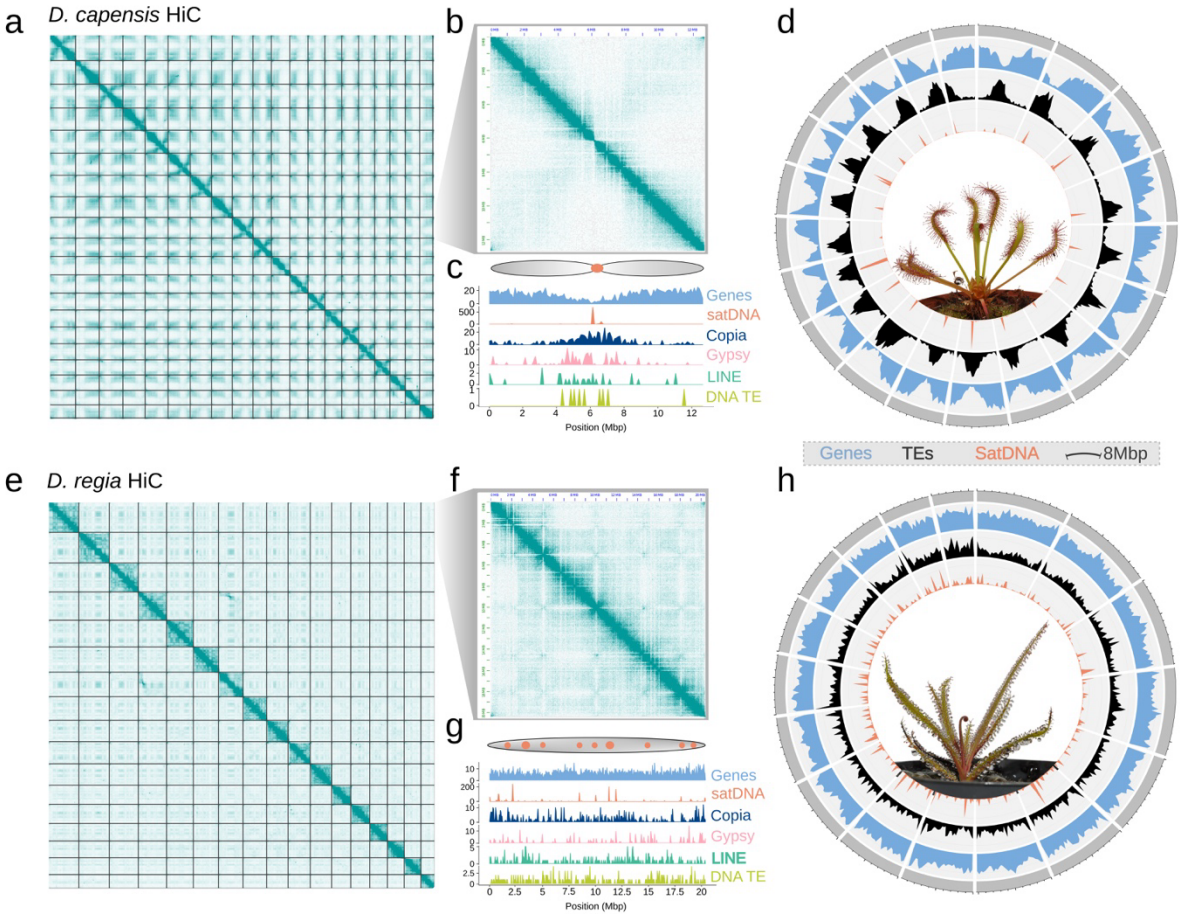


Fig. 1: Contrasting spatial genomic organization in monocentric *D. capensis* versus holocentric *D. regia* chromosomes. (**a, e**) Genome-wide Hi-C contact maps and (**b, f**) single chromosome views, showing the main repetitive sequence type distributions (**c, g**) –clustered in *D. capensis* and uniform in *D. regia*, which are typical patterns for monocentric and holocentric chromosomes, respectively. Window size for sequence type distribution density is 100 kb. (**d, h**) Circos plots showing gene, TE, and satellite DNA distributions across *D. capensis* (**d**) and *D. regia* (**h**) chromosomes again show patterns expected for mono- vs. holocentricity, respectively.

Genetic and epigenetic composition of centromeres in *D. capensis* and *D. regia*

We analyzed the genomes of *D. regia* and *D. capensis* to identify CENH3, a centromere-specific variant of histone H3 that is essential for kinetochore assembly and exhibits remarkable sequence divergence across species[24]. We identified two copies of CENH3 in each genome and generated specific antibodies against their variable N-terminal tails (**Extended Data Fig. 2a**). Immunostaining with the CENH3 antibodies showed distinct foci in *D. capensis* interphase nuclei, confirming monocentricity in this species (**Fig. 2a**). Surprisingly, no specific signal was detected in the interphase of *D. regia* **(Fig. 2b)**. Because it has been shown in some holocentric species that CENH3 is not essential to centromere function[25,26], we aimed to check for the presence of the kinetochore component protein KNL1, which has been reported to be conserved across different plant lineages[27]. Immunostaining with the KNL1 antibody in interphase nuclei of *D. capensis* showed distinct foci (**Fig. 2c**). However, once again, we failed to detect any specific KNL1 signal in *D. regia* **(Fig. 2d)**. Immunostaining of *D. capensis* metaphase chromosomes using KNL1 showed signals confined to a single chromosomal region colocalized with α-tubulin, reinforcing the conclusion of monocentricity in this species (**Fig. 2e–i**). Despite the absence of KNL1 signals in *D. regia*, α-tubulin was found to bind along the entire extent of its mitotic chromosomes (**Fig. 2j–m**). Thus, contrary to previous reports, our results provided solid evidence for monocentricity in *D. capensis*, whereas holocentricity instead for *D. regia*[17].

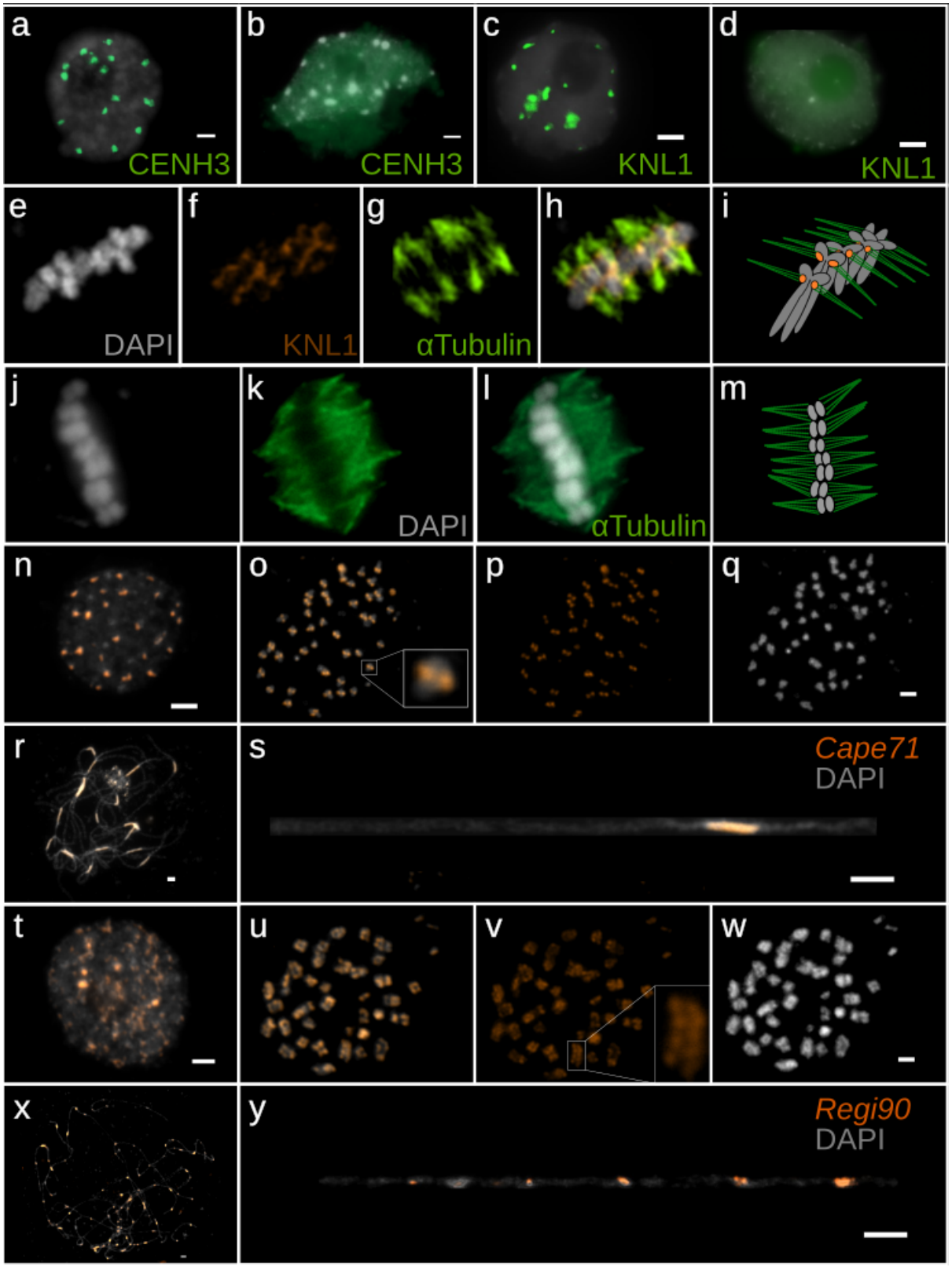


Fig. 2: Fluorescence i*n situ* hybridization and immunostaining using centromeric markers reveal contrasting centromere morphologies between *D. capensis* and D. *regia* chromosomes. (**a, d**) Immunostaining of *D. capensis* and *D. regia* interphase nuclei. CENH3 antibody showing localized signals in *D. capensis* **(a)** and nonspecific signals in *D. regia* **(b)**. The same pattern can be observed using the KNL1 antibody in *D. capensis* **(c)**

and *D. regia* **(d)**. **(e-i)** Immunostaining on the metaphase chromosomes of *D. capensis* using KNL1 and α-tubulin; note the α-tubulin attaching to single regions of the chromosomes. (**i**) Model summarizing the observations in *D. capensis*. (**j-m**) Metaphase chromosomes of *D. regia* showing α-tubulin attaching to the entire chromosome. (**m**) Model summarizing the observations in *D. regia.* (**n-s**) FISH with a *Cape71* probe in *D. capensis* showing the clustered centromeric distribution. (**t-y**) Centromeric distribution. Note the contrasting patterns on interphase nuclei (**n, t**), metaphase chromosomes (**o-q, u-w**) and pachytene chromosomes (**r, x**). Note that **s** and **y** show isolated and stretched chromosomes from the same pachytene cells.

We next analyzed the chromosomal distribution of satellite DNA, a key component of centromere architecture, to further dissect the centromeric landscape in both species[28,29]. Even though satDNA is not strictly required for centromeric function, most eukaryotic centromeres are predominantly composed of tandem satDNA repeats[30]. In our analysis, we identified one main satDNA family in each genome, *Cape71* (a tandem repeat with monomer length of 71 bp) in *D. capensis* and *Regi90* (a tandem repeat with monomer length of 90 bp) in *D. regia*, that do not share significant sequence identity with each other (**Extended Data Fig. 2b**).

Fluorescence *in situ* hybridization (FISH) using a probe for *D. capensis*-specific *Cape71* satellite DNA revealed discrete centromeric signals on interphase nuclei and metaphase chromosomes, while pachytene chromosomes clearly showed a single cluster of *Cape71* repeats in each chromosome (**Fig. 2n-s**), confirming the presence of monocentric chromosomes in *D. capensis*. In contrast, the *D. regia*-specific repeat *Regi90* is found as numerous dispersed foci in interphase nuclei (**Figure 2t**). Remarkably, *Regi90* signals were found to form line-like signals in condensed mitotic metaphase chromosomes in *D. regia* (**Fig. 2u-w**). Indeed, *Regi90* shows a similar behavior as recently reported for holocentromeric repeats in beaksedges[35,36]. Moreover, during meiosis we observed several foci of *Regi90* repeats along highly stretched pachytene chromosomes in *D. regia* (**Fig. 2x-y**). The *Regi90* repeat was also found irregularly distributed in short arrays (10-20 kb) across all *D. regia* chromosomes (**Extended Data Fig. 2c**). This pattern of *Regi90* repeats is intriguing as it resembles the pattern observed for repeat-based holocentromeres reported in *Rhynchospora*[20] despite the fact that we could not detect centromere-specific proteins on *D. regia* chromosomes.

Differential satellite DNA evolution reflects centromere architecture in *Drosera*

We generated phylogenetic trees based on *Regi90* and *Cape71* sequences throughout the *D. regia* and *D. capensis* genomes to characterize the evolutionary differences between centromeric arrays of these species. Separate alignments of individual repeat units from both species revealed considerable SNP diversity and indel differences, both of which we infer to be related to mutational time since element duplications and/or movements[37]. Phylogenies of the repeat sequences parsed by chromosome and location revealed mixtures of positionally localized repeat clades as well as highly structurally mixed clades, suggesting both (1) local duplications of repeats that seed a given region (i.e., younger repeat lineages) followed by (2) movement of duplicates throughout the genomes thereafter (therefore signifying older repeat lineages), which then seed new centromeric foci and repeat the process (**Fig. 3**). *Cape71* phylogeny (**Fig. 3a**) is characterized by multiple, same-chromosome clades on long branches, indicating that distinct subfamilies of monomers have arisen and undergone sustained, chromosome-specific amplification. In contrast, the *Regi90* tree (**Fig 3b**) exhibits a more star-burst topology (very short internal branches and greater intermixing of colored tips) reflecting a recent, genome-wide expansion, with only very few clades of same-chromosome tips marking nascent local amplifications (**Fig. 3b, right panel**). Compared to the genome-wide

mixing implied by the *D. regia* satDNA tree, the *D. capensis* pattern points to stronger, longer-lived local concerted evolution: repeat variants arose and expanded on particular chromosomes and have persisted as distinct subclades rather than being rapidly shuffled across the karyotype.

To assess chromosome-specific evolution of satellite DNA monomers, the tip labels were encoded with their chromosome of origin, and the resulting tree was midpoint-rooted to allow for relative branch-length comparisons. We then inspected terminal clades for non-random clustering of monomers by chromosome. In our analysis and graphic, each point represents the position of a monomer from the *Cape71* tree (**Fig. 3c**), plotted at each chromosome relative midpoint (y-axis) and colored by root-distance order. All points lie at distinct vertical positions (one per chromosome) and display a variety of colors from the rank, indicating that different-age subfamilies lie at *D. capensis* centromeres, with early-splitting lineages having since been amplified on their respective chromosomes. A magnification of the chr19 centromeric region (**Fig. 3d**) once again reveals a narrow gradient of root-distance values within that ~20 kb interval. Filtering by monomers from the more monochromatic *Cape71* tree (**Fig 3d, bottom panel**) does not change the gradient of root-distance ranks along the chromosomal landscape. Indeed, a mostly purple (recent) clade within that tight region reflects an early-splitting subfamily that has undergone sustained diversification (observable as variable terminal branch lengths). By contrast, the *Regi90* plot shows every monomer scattered continuously along the length of each chromosome (**Fig. 3e**), colored across the full gradient from recent (purple) to ancient (yellow). This even distribution of young and old repeats along *D. regia* chromosomes reflects the starburst tree morphology, possibly indicating recent, genome-wide satDNA amplification followed by pervasive gene conversion that homogenized satellite sequences across all chromosomes. In the plot of *D. regia* Chr13, monomers of all root-distance ranks, from recent (purple) to deeply diverged (yellow), are evenly scattered along the entire chromosome length, likely reflecting high homogenization (**Fig. 3f**). However, when we isolate and plot only the monomers from the **Fig. 3b** sub-tree on Chr13 (right panel), they form a cohesive light-green cluster (**Fig. 3f, bottom panel**). Taking into account the topology of the tree, we infer that this cluster corresponds to a sudden amplification of a *Regi90* monomer variant, perhaps expanding as a higher order repeat. It is worth noting that *Regi90* monomers exhibit higher absolute divergence (longer root-to-tip distances; **Fig. 3g**) while displaying less chromosome-specific clustering and stronger karyotype-wide mixing than *Cape71*; conversely, *Cape71* repeats, despite lower overall divergence, maintain more pronounced, array-specific subfamily structure and greater homogeneity within each chromosomal cluster.

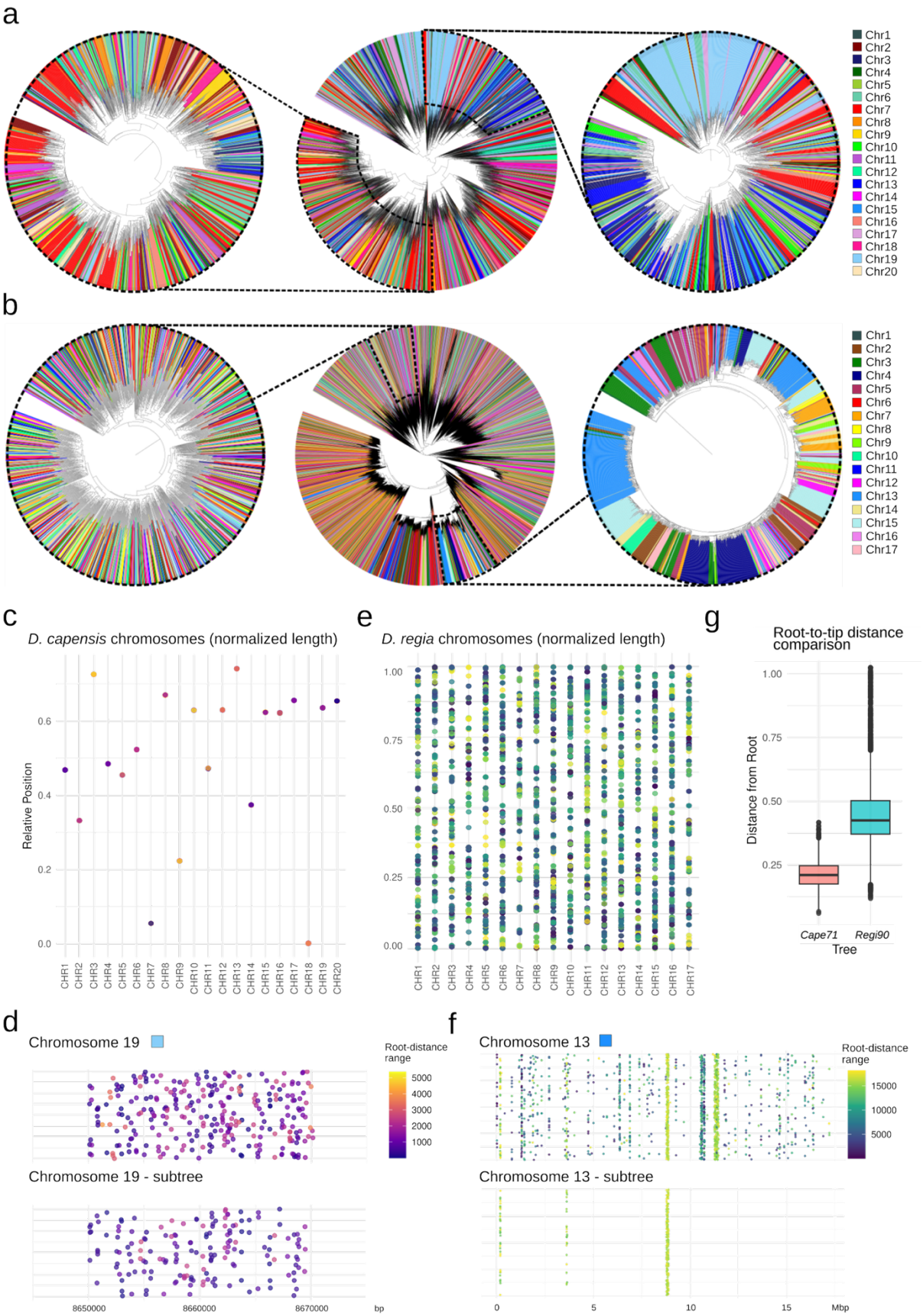

a
Chr1
Chr2
Chr3
Chr4
Chr5
Chr6
Chr7
Chr8
Chr9
Chr10
Chr11
Chr12
Chr13
Chr14
Chr15
Chr16
Chr17
Chr18
Chr19
Chr20
b
Chr1
Chr2
Chr3
Chr4
Chr5
Chr6
Chr7
Chr8
Chr9
Chr10
Chr11
Chr12
Chr13
Chr14
Chr15
Chr16
Chr17
c
D. capensis chromosomes (normalized length)
Relative Position
0.0
0.2
0.4
0.6
CHR1
CHR2
CHR3
CHR4
CHR5
CHR6
CHR7
CHR8
CHR9
CHR10
CHR11
CHR12
CHR13
CHR14
CHR15
CHR16
CHR17
CHR18
CHR19
CHR20
e
D. regia chromosomes (normalized length)
1.00
0.75
0.50
0.25
0.00
CHR1
CHR2
CHR3
CHR4
CHR5
CHR6
CHR7
CHR8
CHR9
CHR10
CHR11
CHR12
CHR13
CHR14
CHR15
CHR16
CHR17
g
Root-to-tip distance comparison
Distance from Root
1.00
0.75
0.50
0.25
Cape71
Regi90
Tree
d
Chromosome 19
Root-distance range
5000
4000
3000
2000
1000
Chromosome 19 - subtree
8650000
8660000
8670000
bp
f
Chromosome 13
Root-distance range
15000
10000
5000
Chromosome 13 - subtree
0
5
10
15
Mbp

Fig. 3: Centromeric satDNA repeat polymer phylogenies reveal both chromosome-specific and mixed clades in *D. capensis* (a) *D. regia* (b). Given the extremely large numbers of sequences analyzed, we zoomed in on subtrees to highlight large chromosome-specific clades, which are shaded blue for both species. **(c)** Chromosomal distribution of the sequences forming the *Cape71* tree in *D. capensis*. Each dot represents one repeat monomer, plotted at its relative position along the chromosome (all chromosomes are scaled to equal length for comparison). Dot color corresponds to the root-to-tip branch length in the tree (i.e., its relative evolutionary divergence). (**d**) Zoomed-in view of the chromosome 19 centromeric clade highlighted in blue in the *D. capensis Cape71* tree. In the upper panel, all branches from the *Cape71* tree that map to chromosome 19; in the bottom panel, only the branches from the subtree (**e**). Chromosomal distribution of the sequences forming the *Regi90* tree in *D. regia*, plotted exactly as in (c): each dot marks one repeat monomer at its normalized chromosomal position, colored by its root-to-tip branch length. (**f**), Zoomed-in view of the repeats from chromosome 13 in the *Regi90* tree, as in (**d**): the upper panel shows all branches from the *D. regia* tree that map to the target chromosome; the lower panel shows only the branches from the extracted subtree. (**g**), Comparison of root to tip distances for both tandem repeat trees.

We further characterized satellite DNA dynamics by conducting a similarity-based assessment within each major satellite DNA family, revealing generally low levels of homogenization across the arrays (**Extended Data Fig. 3**). Interestingly, arrays of the centromeric *Cape71* repeat in monocentric *D. capensis* showed greater homogenization levels across different chromosomes and within chromosomes (**Extended Data Fig. 3a**). However, in *D. regia* the *Regi90* arrays exhibited lower levels of similarity (approx. 75%) among chromosomes, mirroring the similarity observed within individual chromosomes (**Extended Data Fig. 3b**). Notably, a large portion of the annotated satDNA in *D. capensis* was located in small unplaced scaffolds that were not linked to chromosomes, resulting in an underrepresented length of the arrays. From the total of 17 Mbp of *Cape71* found in the genome of *D. capensis* using TideCluster, only 1.12 Mbp were annotated to chromosomes, likely due to the high homogenization of *Cape71* hampering a contiguous assembly of these regions. In contrast, in *D. regia*, from a total of 3.8 Mb of annotated *Regi90*, 3.5 Mbp were placed into pseudochromosomes. These results are interesting as they suggest different rates of ongoing satDNA array size evolution and homogenization between monocentric *D. capensis* and holocentric *D. regia*.

High repeat turnover in sundews

Among the most abundant transposable elements, the genomes of *D. regia* and *D. capensis* were generally rich in *Copia* LTR retrotransposons[38], with a special dominance of *SIRE* in *D. capensis* and *TORK* in *D. regia* **(Extended Data Fig. 4a)**. Although the majority of TEs in *D. capensis* were found in pericentromeric regions, a fraction of *TORK* elements was present in proximity to genes at a greater frequency than other TEs **(Extended Data Fig. 4b)**, similar to the patterns observed in *D. regia* **(Extended Data Fig. 4b, right panels)**. However, comparison of the distribution of different TEs in relation to the putative centromere positions revealed a preference for LINEs in *D. capensis* **(Extended Data Fig. 4b, upper left)** and for the chromovirus *REINA* in *D. regia* **(Extended Data Fig. 4b, upper right)**. While chromoviruses preferentially target centromeric regions, LINE elements are often found pericentromerically[39,40]. In both cases, our results are in agreement with the assumption that *Cape71* and *Regi90* are the sequences associated with active centromeres.

Next, we aimed to characterize the dynamics and abundance of repetitive DNA across the *Drosera* genus. Although the specific composition of such DNAs is highly variable among angiosperm species, these sequences have proven to be major players in the evolution of eukaryotes[41]. For this purpose we included Illumina reads from an additional 10 *Drosera*

species to analyze their repetitive fractions. Comparative analysis of the repeatomes of the 12 related species revealed that most of the elements found were shared only among species within the same clade. These similarities apparently reflect their phylogenetic relationships, as the sister species *D. capensis* and *D. aliciae* share a great diversity of repetitive elements. Apart from some families of transposable elements, such as *ATHILA*, *SIRE* or LINE, the satellite repeat *Cape71* in *D. capensis* was found to be shared with *D. aliciae* (**Extended Data Fig. 4c**). In addition, a closer look into the similarities between *Cape71* and the remaining satellite repeats found among the 12 species revealed that *Cape71* is related to both of the satDNA clusters found in *D. spatulata* (**Extended Data Fig 4d**), thus reflecting the close evolutionary relationships of the three species. Another case of repeatome interrelationship that matches phylogeny was observed for *D. anglica*, *D. filiformis* and *D. intermedia,* where most of the different families of transposable elements were shared between these species but not with the other *Drosera* species studied. However, in this case, no shared satDNA repeats were found. Finally, great similarities among the elements of *D. erythoriza* and *D. peltata* were also found. Despite their different bursts of the *RETAND* and LINE families, these species shared most of their elements with their sister species *D. menziesii* (**Extended data Fig. 4c**). In addition, when the satDNA fractions of these three species were compared, we observed that the three taxa shared the same family of satDNA (**Extended data Fig. 4d**).

Interestingly, potentially holocentric species such as *D. regia* (**Fig. 1e-h**), *D. scorpioides*[42] and D. *broomensis* are particularly divergent in their repetitive element compositions, having unique families of transposable elements and satDNAs in the majority of their clusters (**Extended Data Fig. 4c**). In this regard, *Regi90* does not share significant similarity with any of the other satellites analyzed, except for short stretches of homopolymer/mononucleotide repeats (AAAs)n (**Extended data Fig. 4d, Extended data Fig. 2b,**). In summary, our comparative analysis revealed that repetitive fractions experienced rapid turnover across the different clades of *Drosera*, which could be potentially linked with the different types of centromere organization found.

Phylogeny and genome structural evolution of the Caryophyllales carnivore clade

Backbone phylogeny reveals unexpected relationships for *D. regia* and *N. gracilis*. The genome of *Nepenthes* revealed a complex allopolyploid history (five subgenomes of x=8)[13] that we suspected could confound inferences of relationships within *Drosera*. We therefore constructed a backbone phylogeny to model reticulate evolution among Caryophyllales carnivores by using conserved Benchmarking Universal Single-Copy Ortholog (BUSCO)[43] gene sets to generate phylogenetic trees for *Drosera*, its close relatives, and selected outgroup angiosperms. Two sets of data were analyzed: nucleotide and inferred amino acid sequences for 22 and 50 species, respectively (**Extended Data Fig. 5**). Distinct taxon samplings permitted evaluation of topological consistency under different possible outgroup effects. Phylogenetic reconstructions were made in two distinct ways: (1) maximum likelihood (ML) analyses[44] of concatenated BUSCO alignments, and (2) ASTRAL species tree reconstruction[45] summarizing all individual BUSCO gene trees. In all analyses, *D. regia* was resolved as sister to *Aldrovanda* plus *Dionaea* instead of to other species of *Drosera*. *Nepenthes* was usually resolved as sister to *Drosera* plus the snap-trapping carnivores, with the *Drosophyllum* plus *Triphyophyllum*/*Ancistrocladus* group forming that clade's sister lineage (**Extended Data Fig. 5**). These relationships for *N. gracilis* were incongruent with published results (that also employed nuclear-encoded BUSCO genes), which placed *Nepenthes* sister to the *Drosophyllum* plus *Triphyophyllum*/*Ancistrocladus* clade instead[13]. However, these

relationships, forced to a bifurcating topology, are clearly inadequate for representing phylogenetic relationships among alloploids, which have hybrid origins[46].

Single copy genes display substantial gene tree conflict. As a first approach to evaluating reticulate interrelationships, we analyzed taxonomically complete single-copy gene data for eight Caryophyllales carnivore species and their non-carnivorous relative, *Ancistrocladus*. Using such complete taxonomic representation, we visualized the results as a cloudogram[47] to qualitatively highlight evidence for ancient admixture or incomplete lineage sorting that may have given rise to conflicting gene and species trees (**Fig. 4**). Here, it became apparent that *D. regia* has conflicting gene-tree relationships, wherein some loci favor a close relationship to *Aldrovanda* plus *Dionaea*, whereas others place *D. regia* as sister to the other *Drosera* species pair. Moreover, the gene tree conflict within the snap-trapping *Aldrovanda* plus *Dionaea* group suggests complex histories for these genera.

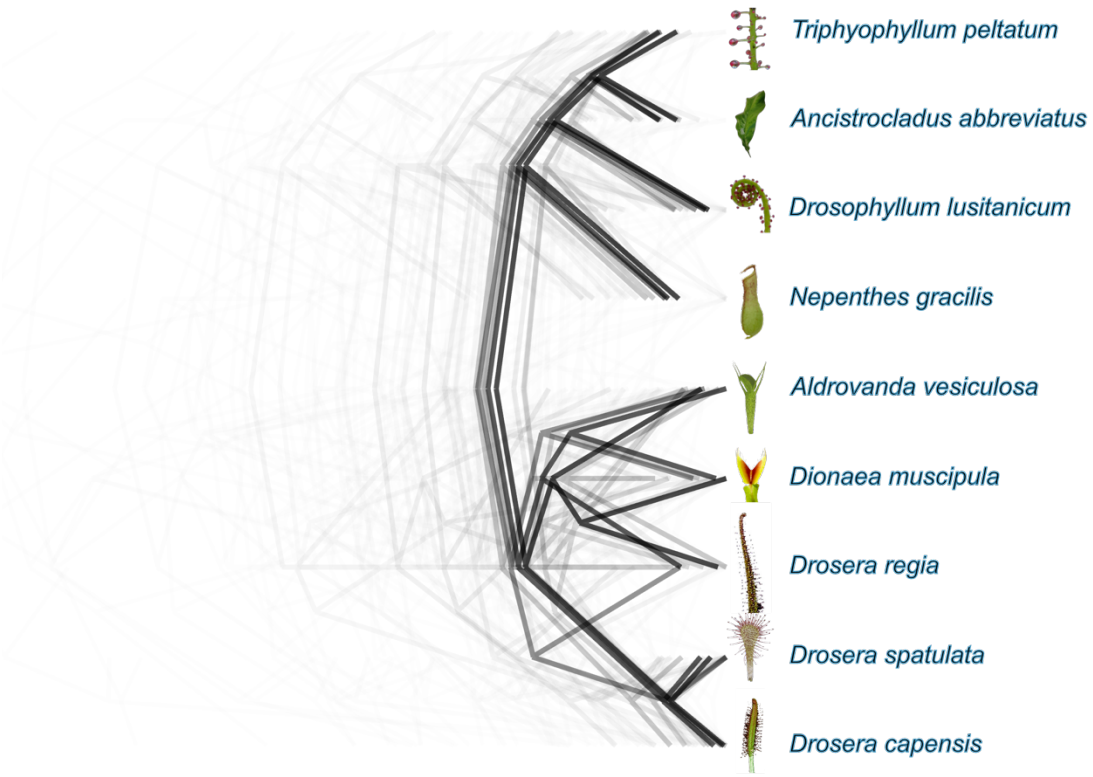


Fig. 4. A Densitree cloudogram depiction of gene tree conflict among 563 phylogenies of single-copy orthologs. Consensus trees representing distinct tree topologies found among the set are shown with their intensities proportional to their frequencies among the tree set[47]. *Dionaea*, *Aldrovanda* and *D. regia* reveal highly reticulate cloudogram structure consistent with allopolyploid origins for each species.

Genome structure suggests complex polyploidy. We further explored reticulate relationships within this Caryophyllales lineage by tracking genome structural evolutionary relationships. We used GENESPACE[48] to perform gene-based synteny analyses between *D. capensis* and *D. regia* against available genomes from closely related outgroups: the pitcher plant *N. gracilis*, the Venus flytrap *Dionaea muscipula*, *Ancistrocladus abbreviatus* and *Triphyophyllum peltatum*. Synteny of the *N. gracilis* genome was highly conserved with the other

Caryophyllales genomes studied. *Ancistrocladus* and *Triphyophyllum* shared a chromosomal fission followed by a tetraploidy event, which explains higher basic chromosome number of these species (x=9, **Fig. 5a**). *Dionaea* was also resolved as a tetraploid, but with two clearly different subgenomes (**Fig. 5b, Extended Data Fig. 6**), indicating an allotetraploidy following lineage hybridization. Altogether our results suggest the basic chromosome number of the lineage to be x=8 (see also Mohn et al., 2023[28]), with an ancestral karyotype similar to the dominant subgenome of *N. gracilis*.

The genome of *N. gracilis* bears five highly syntenic subgenomes, four of which are recessive in terms of genome expression and fractionation biases, while one is dominant[13]. Thus, for the sake of simplicity we compare only the dominant subgenome of *N. gracilis* with the other genomes. Remarkably, a high level of synteny was observed across most of the species' genomes with a relative decrease in synteny found among sundews, especially *D. capensis*. Indeed, most *D. regia* chromosomes match *N. gracilis* dominant chromosomes in a 3:1 manner (**Fig. 5a**), consistent with the former species' proposed hexaploid subgenome structure, which can in fact be phased 2 parts to 1 using Ks values (**Fig. 5c**). In contrast, the *D. capensis* genome was much more rearranged, showing much less syntenic block collinearity with the *N. gracilis* genome. Nonetheless, an unprecedented 6:1 match with the *N. gracilis* dominant subgenome supports a dodecaploid origin of its genome via tetraploidization of a hexaploid ancestor (**Fig. 5a**), in contrast to previous assumptions.

We were further able to underscore the paralogous/orthologous syntenic relationships within and between *D. regia* and *N. gracilis* using a concatenated *D. regia* and *N. gracilis* hybrid assembly and the SynMap application in CoGe[49] (**Fig. 6a,b**). Three Ks coloration ranges for syntenic blocks in the dot plot matched WGM or speciation events. Within *D. regia*, the 2:1 subgenome phasing was apparent as the purple:cyan blocks, with the purple block – plus the diagonal – representing the two youngest of the three subgenomes, and the single cyan block the oldest (see also **Fig. 5c**). These blockwise relationships can be discerned counting either horizontally or vertically. Within *N. gracilis*, its 4:1 subgenome structure can be recognized as four blocks on the dot plot plus the diagonal, and its pentuplicate structure is even clearer in its 5:3 blockwise to *D. regia*. From colorations in the Ks histogram (**Fig. 6b**), it can be seen that both of the *D. regia* events (purple younger that cyan) occurred more recently than those of *N. gracilis* (green and orange), with the species split between *D. regia* and *N. gracilis* (also orange) as old as some of the *Nepenthes*-specific subgenomes resulting from the latter taxon's two lineage-specific tetraploidies. Remarkably, window-wise fractionation bias calculations[49,50] (Fig. 6c-e) demonstrated that all three *D. regia* subgenomes share strong and highly similar post-polyploidy gene retention profiles with the dominant subgenome of *N. gracilis*, the recessive subgenomes of which are much more heavily fractionated. As such, it is plausible that the *N. gracilis* dominant subgenome shares close ancestry with a near relative of *D. regia*.

We investigated whether the hexaploid ancestry of *D. capensis* is shared with *D. regia* by conducting additional synteny analyses using *ksrates*[51] and targeted GENESPACE analyses. Despite extensive post-polyploid rearrangements in *D. capensis*, some chromosomes shared between *D. regia* and *D. capensis* (denoted by pairs of different colors detailed in **Extended Data Fig. 8**) displayed plausibly shared fusion events, suggesting common ancestral genomic restructuring that occurred prior to their species split, i.e., in their common ancestor. *ksrates* analysis further confirmed that the allotetraploidy in *Dionaea* and the hexaploidy in both *D. regia* and *D. capensis* predated their respective species splits, which points toward shared triplicate structure for *D. regia* and *D. capensis* followed by *D. capensis* subsequently undergoing an additional lineage-specific tetraploidy event (**Extended Data Fig. 9**).

Furthermore, congruent with the SynMap Ks coloration in **Fig. 6a,b**, *ksrates* revealed a species split between *N. gracilis* and the *Dionaea-Drosera* group far earlier than the phylogenetic splits among the latter.

Nonetheless, definitively asserting a single, shared hexaploidy event for both *D. regia* and *D. capensis* based solely on shared rearrangements is complex. While shared rearrangements are a useful, parsimonious proxy for inferring a common polyploid origin[52,53], this association is not unassailable. Given triplicate structure in both *Drosera* species, the probability of observing seemingly shared rearrangements by chance, or at sites prone to recurrent breakage, is low but cannot be entirely excluded. Interpretation of a shared hexaploidy is further nuanced by phylogenetic evidence (e.g., our BUSCO topologies; **Extended Data Fig. 5**) suggesting *D. regia*'s sister-group relationship to *Dionaea*, rather than to the *Drosera* clade that contains *D. capensis*. Future work should include the analysis of additional *Drosera* genomes to robustly resolve the precise timing and shared ancestry of one or more hexaploidy events. This will allow a clearer understanding of whether pervasive fusions in *D. regia* occurred before or after the evolution of holocentricity, and if reversal from holocentricity to monocentricity in *D. capensis* is plausible.

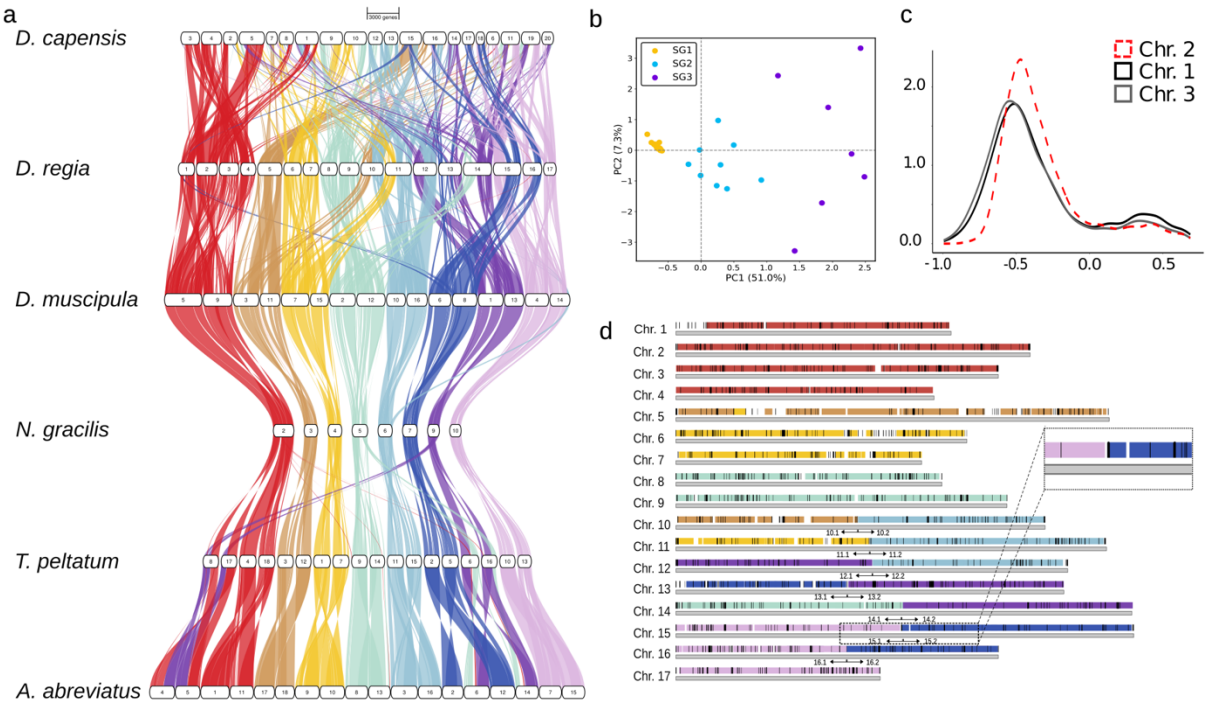


Fig 5: Differential structural rearrangements shaped the evolution of mono- and holocentric *Drosera* genomes. (a) Synteny comparison of the dominant subgenome of decaploid *N. gracilis* (8 chromosomes) against tetraploid *Ancistrocladus abbreviatus* and *Triphyophyllum peltatum* (below *Nepenthes*), allotetraploid *Dionaea muscipula*, hexaploid and holocentric *D. regia* and the dodecaploid and monocentric *D. capensis* (above *Nepenthes*). *Dionaea* is doubled compared to *Nepenthes*, and very little rearranged; *D. regia* and *D. capensis* share triplicate structure and some chromosomal fusions compared to *Nepenthes*. *D. capensis* shows a further duplication and complex rearrangements. (**b**) Subphaser PCA plot based on a concatenation of the *Dionaea* and *D. regia* assemblies demonstrates that *Dionaea* is an allopolyploid of subgenomes 2 (cyan) and 3 (purple), and that *D. regia* (yellow) has evolved distinct *k*-mers since its divergence from the former species (see also **Extended Data Fig. 6**). (**c**) Density Ks plot of *D. regia* homeologous gene pairs clearly phases *D. regia*'s triplicate subgenome structure into a 2:1 pattern. Chromosomes 1 and 3 match, whereas chromosome 2 represents an older subgenome at higher Ks (x-axis = log10 Ks; y-axis = density; see also **Extended Data Fig. 7**). (**d**) Analysis of

Regi90 repeat localization (black bars) at chromosomal fusion points found in *D. regia*. Note the high frequency of *Regi90* found at fusion regions. The colored rectangles represent syntenic blocks to ancestral *N. gracilis* chromosomes in panel **a**. Note the black bar at the break point of chromosome 15.

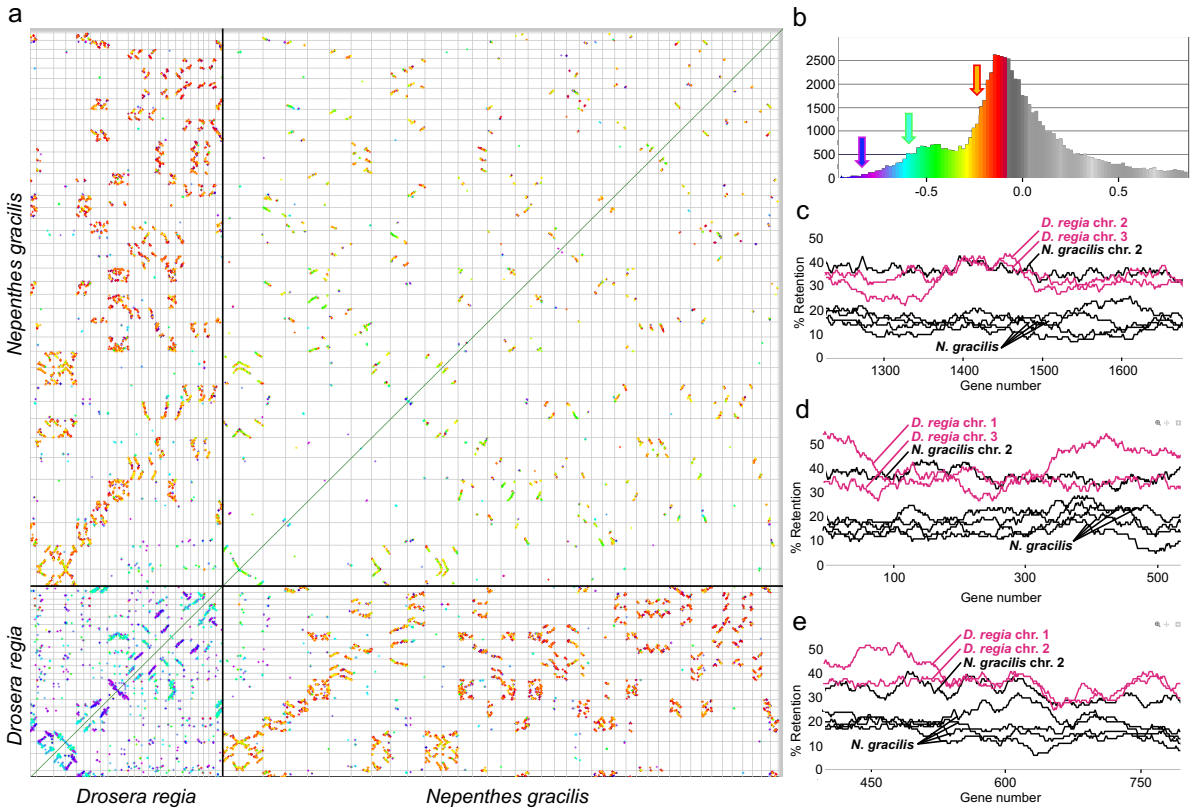


Fig. 6. **Ancestral subgenome sharing between *Nepenthes* and *D. regia*.** (**a**) Self:self syntenic dot plot of a concatenated assembly of *D. regia* and *N. gracilis*, (**b**) colored by with Ks. (**c**) FractBias plots showing a close relationship, through very similar gene retention profiles, among all three *D. regia* subgenomes and the *Nepenthes* dominant subgenome. The four recessive *Nepenthes* chromosomes (generated by two tetraploidy events) have much lower gene retention profiles and are shown in gray. Views (**c-e**) show the three different possible chromosome pairs mapped to *D. regia* chromosomes 1, 2 and 3, respectively; *Nepenthes* dominant chromosome 2 is their syntenic match.

In summary, our inferred hybrid origin of *D. regia* and *D. capensis* likely involves other *Drosera* species in addition to the *Dionaea* lineage, as is apparent from the extreme conflict among independent phylogenies based on a large collection of single-copy genes (**Fig. 4**). Originating from three x=8 ancestral subgenomes, the immediate progenitor of *D. regia* was likely n=24 prior to chromosomal rearrangements that led to its modern n=17 karyotype. The *D. capensis* genome also contains an underlying triplicate structure, many species-specific chromosome fusion events, and a recent, lineage-specific WGM that led to its dodecaploid karyotype (**Fig. 5a**). *Triphyophyllum peltatum* includes, through a lineage-specific WGM, n=18 chromosomes descendant from a x=9 progenitor derived from a chromosomal fission event within the *Caryophyllales* x=8 ancestor. *Ancistrocladus* is sister to *Triphyophyllum* and shares its karyotype (**Fig. 5a**).

Based on our backbone phylogenetic tree analyses, syntenic relationships, subgenome fractionation patterns, Ks distributions, and chromosome numbers, we offer a tentative scenario (**Fig. 7**) for reticulate allopolyploid interrelationships among the Caryophyllales carnivore genomes analyzed here that will require additional high-quality genome assemblies for validation.

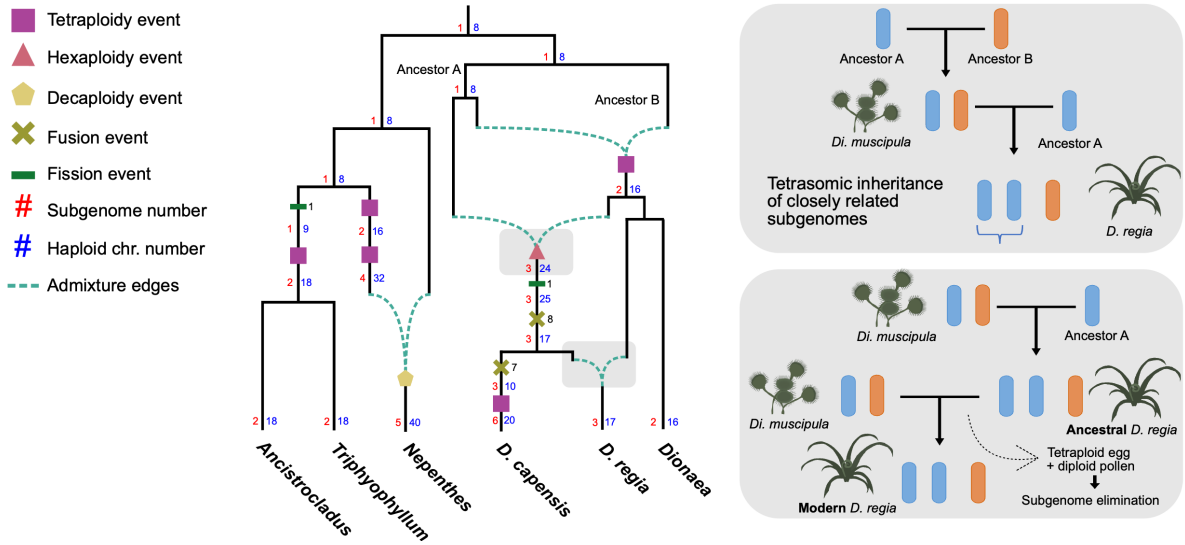


Fig. 7. Phylogenetic network hypothesis for interrelationships among the six Caryophyllales genomes under study. An *x*=8 ancestor is proposed. All other Caryophyllales carnivores share *Nepenthes*' *x*=8 ancestral subgenome, or a modification of it. *Dionaea* is an *x*=16 allotetraploid hybrid of two different 8-chromosome ancestors. *D. regia* appears to share the two subgenomes of that species, along with a third subgenome joining in to yield another hybrid lineage, an allohexaploid one that precludes phylogenetic depiction as deriving from a single phylogenetic split. Polyploidy events, chromosome fusions and fissions, numbers of subgenomes, haploid chromosome numbers, and admixture (hybridization) edges are denoted as in the legend. The first and second grey panels detail the possible events that gave rise to the modern *D. regia* genome. In the upper panels, two successive hybridization events produce an ancestral *D. regia* that possesses three distinct subgenomes: two closely related subgenomes (blue rectangles) and one divergent subgenome (orange rectangle). In the lower panel, a single cross between an ancestral *D. regia* lineage and an ancestral *Dionaea* lineage gives rise to modern *D. regia*.

Discussion

We present chromosome-scale assemblies for *Drosera capensis* and *D. regia*, two carnivorous sundews that exemplify the rare coexistence of divergent centromere architectures, monocentric and holocentric, within a single genus. This unusual pairing provides a powerful system for investigating how centromere type influences genome architecture, repeat evolution, and chromosomal behavior, or conversely, how genome dynamics may shape centromere evolutionary transitions.

Integrating cytogenetic, genome sequence, and 3D chromatin conformation data, we confirm monocentricity in *D. capensis* and holocentricity in *D. regia*, an infrageneric contrast otherwise reported in plants only in the parasitic genus *Cuscuta*. While holocentricity has evolved independently across angiosperms[54-56], the trait is otherwise infrafamilial only. These findings underscore the need for high-resolution chromosomal and cytological assays when inferring

centromere types, particularly in groups with complex karyotypes. Indeed, the recent discovery of extended regional centromeres in the related *Dionaea muscipula*[57] suggests a continuum in centromere organization across Droseraceae, from monocentric to extended regional to holocentric forms, potentially shaped by lineage-specific responses to whole-genome duplication and the "genomic shock" caused by TE activation[58].

Structurally, *D. capensis* and *D. regia* exhibit striking divergence in centromeric repeat profiles. The satDNA *Cape71* is localized and structured in *D. capensis*, consistent with point centromere function, while *Regi90* in *D. regia* is dispersed along chromosomes, resembling other repeat-based holocentromeres[24,26]. The absence of CENH3 and KNL1 immunosignals in *D. regia* may also support a non-canonical centromeric structure defined by unusual epigenetic mechanisms. Species-specific TE enrichments, SIRE in *D. capensis* and Tork in *D. regia*, may contribute to centromere remodeling, influencing chromatin context and kinetochore plasticity.

Comparative repeat profiling across the *Drosera* genus reveals high variability in satDNA and TE composition, consistent with rapid repeat turnover and centromere repositioning, possibly in response to polyploidy. These patterns reinforce emerging models linking transposon dynamics to the restructuring of plant centromeres and overall genome fluidity.

Synteny analyses, particularly between the shared chromosomal rearrangements in both *D. capensis* (dodecaploid) and *D. regia* (hexaploid) suggest a possible common hexaploid ancestor that had already undergone extensive chromosomal fusions. However, the two species exhibit sharply divergent structural outcomes: holocentric *D. regia* retained this ancestral karyotype with minimal rearrangements, whereas monocentric *D. capensis* experienced substantial chromosomal reshuffling. The observed divergence following whole genome duplication suggests that centromere type may influence genome stability, with holocentricity potentially buffering against large-scale rearrangements.

It is remarkable that such divergent chromosomal architectures evolved within *Drosera*, a genus embedded within the morphologically and physiologically complex Caryophyllales carnivorous plant clade. This monophyletic lineage encompasses species with diverse trapping strategies: flypaper traps in *Drosera*, *Drosophyllum*, and *Triphyophyllum*; snap-traps in *Dionaea* and *Aldrovanda*; and pitfall traps in *Nepenthes*[14]. Within this context, our findings could suggest that centromere and polyploid mechanistic diversity may correlate with ecological diversification or even promote it, which would raise questions about the interplay between chromosomal mechanics, genome architecture, and adaptation.

Together, our results demonstrate that centromere organization, polyploid history, and repeat dynamics jointly shape genome evolution in *Drosera*. The presence of both holocentric and monocentric species, each likely derived from the same, distinctly chromosomally fused hexaploid ancestor yet diverging markedly in chromosome structural outcomes, points to centromere type as a possible determinant of post-polyploid genome trajectories. Future studies should investigate how these architectural differences influence chromosome segregation, recombination, and gene expression, ultimately shaping adaptation in diverse plant lineages.

Materials and methods

Plant materials and genome sequencing. In vitro cultures of *D. regia* and *D. capensis* were initiated from seeds derived from plants maintained in the carnivorous plant collection at the Red Manejo Biotecnológico de Recursos, Instituto de Ecología, A.C., Xalapa, Veracruz, Mexico. For superficial disinfection, seeds were placed in filter paper envelopes (Whatman

No. 1, 110 mm diameter) and submerged in sterile distilled water for 30 minutes. They were then soaked in a 10% (v/v) commercial bleach solution (1.8% NaOCl) containing two drops of Tween-80 per 100 mL (Sigma, St. Louis, MO) for 10 minutes. This was followed by four rinses with sterile distilled water under aseptic conditions. The disinfected seeds were sown in 125-mL baby food jars containing 25 mL of half-strength MS medium (Murashige and Skoog, 1962), supplemented with 30 $g \cdot L^{-1}$ sucrose. The pH was adjusted to 5.0 ± 0.1 using 0.5 N NaOH and 0.5 N HCl prior to the addition of 7.5 $g \cdot L^{-1}$ Agar, Plant TC (Caisson A111), followed by autoclaving at 1.2 $kg \cdot cm^{-2}$ and 120 °C for 15 minutes. Cultures were incubated in a growth chamber at 25 ± 1 °C under a 16-hour photoperiod provided by LED lamps (50 $\mu mol \cdot m^{-2} \cdot s^{-1}$). Plants obtained from germinated seeds were subcultured every 3–4 months on the same medium to promote growth and multiplication.

High-molecular-weight DNA was extracted from nuclei isolated from the young leaves of *Drosera* species, following the protocol by Steinmüller and Apel, 1986[59]. To reduce contamination from chloroplast and mitochondrial DNA, nuclei were first collected from a 60% Percoll (Invitrogen) density gradient after centrifugation at 4000g for 10 minutes at 4°C. Next, high-quality megabase-sized DNA was obtained using the MagAttract HMW DNA Kit (Qiagen). Before library preparation for sequencing, the integrity of the high molecular weight (HMW) DNA was confirmed using pulsed-field gel electrophoresis (CHEF-DRIII system, Bio-Rad), as described elsewhere[60]. For library preparation, 10 µg of DNA were sheared to a fragment size range of 10–40 kb using a Covaris g-TUBE. The resulting fragment distribution was verified by pulsed-field gel electrophoresis. The sheared DNA was purified using 0.45× AMPure PB beads (Pacific Biosciences) following the manufacturer's protocol.

Library preparation For the PacBio RS II instrument was carried out according to the PacBio 20 kb SMRTbell Template Preparation Protocol, using 5 µg of the sheared DNA as input material. After preparation, the library size distribution was analyzed on an Agilent DNA 12000 Bioanalyzer chip to determine the appropriate size-selection cut-off. Libraries were size-selected with a Sage Science BluePippin system, employing a dye-free 0.75% agarose cassette and a 15 kb cut-off. The selected libraries were reanalyzed on the Bioanalyzer to confirm size distribution. Two libraries per species were prepared for *D. capensis* and *D. regia* The *D. capensis* library was sequenced on two SMRT cells of the PacBio RSII single-molecule sequencing platform at loading concentrations of 0.15 nM and 0.2 nM, respectively. The *D. regia* library was sequenced on eight SMRT cells at a loading concentration of 0.2 nM. For Illumina sequencing of *D. capensis*, *D. regia* and the ten additional *Drosera* species, the Illumina HiSeq 2500 (rapid run, 2x250bp; https://www.illumina.com/documents/products/datasheets/datasheet_hiseq2500.pdf) was employed.

Genome assembly and Hi-C scaffolding. The 2x250bp Illumina reads were first filtered for adapter sequences using Trimmomatic v0.36[61]. A hybrid assembly including the filtered reads and PacBio RS II reads was then carried out using MaSuRCA v3.2.1[62]. The contig-level assemblies were evaluated for completeness with BUSCO v3.0.2[43] using the embryophyta odb9 database, and for contiguity, using QUAST v4.3[63].

Dovetail Hi-C reads were first mapped to the contig file obtained from the MaSuRCA assembler using BWA[64] following the hic-pipeline (https://github.com/esrice/hic-pipeline). Hi-C scaffolding was performed using 3D-DNA pipeline[65,66] with default parameters using '*GATC, GAATC, GATTC, GAGTC, GACTC*' as restriction sites. After testing several minimum mapping quality values of bam alignments, the final scaffolding was performed with MAPQ10. Following the automated scaffolding by 3D-DNA, several rounds of visual assembly correction

guided by Hi-C heatmaps were performed. When regions showed multiple contact patterns, manual re-organization of the scaffolds was performed with Juicebox and 3D-DNA assembly pipeline[65] to correct position/orientation and to obtain the pseudomolecules.

Transcriptome sequencing and genome annotation. Total RNA was extracted from *D. capensis* leaf tissues and petioles. Sample preparation employed a single stranded mRNA library kit, and libraries were subsequently sequenced using an Illumina HiSeq instrument. We obtained a total data of 105,784,845 read pairs. To generate our transcriptome assembly, we first merged the RNA-Seq data from the two tissues. We then assembled one de novo transcriptome using transAbyss v2.0.1[67]. In transAbyss, we used multiple *k*-mers (33-75) in steps of 2 to generate multiple assemblies before merging them all into a single large set. We then assembled a second de novo transcriptome using Trinity v2.6.6[68] using the default *k*-mer size of 25. We generated one reference-guided transcriptome by first mapping the RNA-Seq reads against the reference *D. capensis* genome using HISAT2 v.2.1.0[69] and then assembling the transcriptome using StringTie v.1.3.4c[70]. We then passed the three transcriptome assemblies to EvidentialGene v2017.12.21[71] which produced a final high confidence transcriptome assembly.

For repeat masking, we first generated a de novo repeat library for the *D. capensis* genome using RepeatModeler v1.0.9[72] and then masked the genome using RepeatMasker v4.0.7[73]. For gene model prediction, the transcriptome assembly was first splice-aligned against the unmasked reference *D. capensis* genome using PASA v2.2.0[74] to generate ORFs. Secondly, ab initio gene model prediction was carried out using BRAKER v.2.0.3[75], which internally used the reference aligned RNA-Seq data and Genemark-ET to train AUGUSTUS v3.3[75,76] for its final prediction. Additionally, the self-training Genemark-ES v.4.33[77] was run independently to generate a second set of predictions. Finally the 2 prediction tracks and two spliced-alignment tracks were passed to the combiner tool Evidence Modeler v1.1.1[74] with highest weights to transcriptome evidence and lowest weights to Genemark-ES to generate a final high confidence gene prediction set. This final prediction set was re-run through PASA to update the gene models, add UTRs and identify and generate alternate spliced models.

For *D. regia*, the *D. capensis* transcriptome assembly was used as evidence for training purposes. First, the assembly was splice-aligned against the *D. regia* genome using PASA v.2.2.0[74] to generate ORFs. Additionally, *Arabidopsis thaliana* gene models were aligned against the genome using exonerate v2.2.0[78]. Genemark-ES v4.33[77] was then used to generate one set of predictions. BRAKER v2.0.3[75] was used in the protein mode where the *D. capensis* gene models were aligned against the *D. regia* genome using GenomeThreader v1.7.1 (https://genomethreader.org/), and that set was used to train BRAKER v2.0.3 to generate a second set of gene models. AUGUSTUS v3.3[75,76] was run independently using *D. capensis* parameters to generate a third set of predictions for *D. regia*. All of these gene predictions were then passed to Evidence Modeler v1.1.1[74] to generate a single high confidence gene prediction set. This set was once again passed through PASA to update the gene models with UTR regions and also to generate the splice variants.

Synteny analysis. Synteny analyses were performed with GENESPACE (https://github.com/jtlovell/GENESPACE)[48]. To identify shared chromosomal rearrangements, we first identified all end-to-end fusions of non-homoeologous chromosomes within the *D. regia* genome. For each such fusion event, we then extracted corresponding genomic regions from *D. capensis* that involved the same ancestral *Nepenthes*-like chromosomes. Shared rearrangements were inferred when the breakpoint boundary regions in

D. regia precisely matched the syntenic block locations in *D. capensis* that represented these fusions.

Further synteny analyses were performed using the CoGe SynMap platform (https://genomevolution.org/coge/SynMap.pl)[49]. Synteny plots were obtained using the following steps: (1) using the Last tool, (2) synteny analysis was performed using DAGChainer, using 20 genes as the maximum distance between two matches (-D) and 5 genes as the minimum number of aligned pairs (-A). Then (3) either default (no syntenic depth) or Quota Align (with syntenic depth) was used with overlap distance of 40 genes, and (4) orthologous and paralogous blocks were differentiated according to the synonymous substitution rate (Ks) using CoGe-integrated CodeML, and represented with different colors in the dot plot. FractBias[50] was run using Quota Align window size of 100 for all genes in the target genome, with a syntenic depth ratio of 12:12 with maximum query and target chromosome numbers of 64 each for the concatenated *D. regia-N. gracilis* genome assembly.

For the characterization of the regions involved in fusions, we followed Hofstatter et al, 2022[24]. The synteny alignment between *D. regia* genome and *N. gracilis* dominant subgenome obtained in GENESPACE allowed us to pin the putative regions around the borders of the fusion events. In order, to identify the underlying sequences at the fusion regions, we loaded annotation features for genes, TEs, and tandem repeats on synteny alignments. This allowed us to detect the sequence types in the putative fused regions. Further inspection and characterization of such regions were done by checking the genome coordinates and annotation features with Geneious (https://www.geneious.com).

Subgenome-aware phasing of *D. regia* and *Dionaea muscipula* genomes. We used SubPhaser[79] (default parameters) to phase and partition the subgenomes of *D. regia* and *D. muscipula* genomes simultaneously by assigning chromosomes to subgenomes based on differential repetitive *k*-mers. Additionally, Ks distributions of homeologous duplicate gene pairs were extracted from CoGe SynMap calculations (above) to generate density plots for each triplet of *D. regia* ancestral chromosomes. Density plots were generated in R using the tidyverse[80], ggplot2[81], RColorBrewer[82], ggridges[83], and ggpmisc[84] packages.

***ksrates* analysis.** *ksrates* version 1.1.4[85] was used to position species splits relative to polyploidy events. Coding sequence (CDS) fasta files were extracted using AGAT version 1.4.0[86]. Paralogous Ks peaks were generated for the following focal species: *Ancistrocladus abbreviatus* (this study), *Dionaea muscipula* (this study), *D. capensis* (this study), *D. regia* (this study), *Nepenthes gracilis*[13], and *Triphyophyllum peltatum* (this study). *Beta vulgaris* (GCA_026745355.1), *Coffea canephora*[87], *Gelsemium elegans* (CoGe id64491), and *Spinacia oleracea* (GCA_020520425.1) were also included in the analysis, but not as focal species. Orthologous Ks peaks were generated for each required species pair.

For the tree topology used in *ksrates*, OrthoFinder v2.5.5[88] was run with default settings to generate a tree for all species listed above. Each annotation was reduced to the longest isoform and proteins were extracted using AGAT version 1.4.0[86]. The species tree inference was performed with STAG[89], which uses the proportion of species trees derived from single-locus gene trees supporting each bipartition as its measure of support, and rooted using STRIDE[90].

Repeat characterization. DANTE and DANTE-LTR retrotransposon identification (Galaxy Version 3.5.1.1) pipelines[91] were used to identify full-length LTR retrotransposons in the assembled genomes of *D. capensis* and *D. regia*, using a set of protein domains from REXdb[92]. All complete LTR-RTs contain GAG, PROT, RT, RH and INT domains, including some lineages encoding additional domains, such as chromodomains (CHD and CHDCR) from

chromoviruses or ancestral RNase H (aRH) from Tat elements. DANTE_LTR retrotransposon filtering (Galaxy Version 3.5.1.1) was used to search for high quality retrotransposons, those with no cross-similarity between distinct lineages. This tool produced a GFF3 output file with detailed annotations of the LTR-RTs identified in the genome and a summary table with the numbers of the identified elements. Overall repeat composition was calculated excluding clusters of organelle DNA (chloroplast and mitochondrial DNA).

Tandem repeat sequences were identified using RepeatExplorer2 (https://repeatexplorer-elixir.cerit-sc.cz/) and further verified using the TideCluster pipeline (https://github.com/kavonrtep/TideCluster)[93]. All putative tandem sequences were compared for homology using DOTTER. These tandem sequences were individually mapped to the genome by BLAST[94], with 95% similarity in Geneious (https://www.geneious.com). The mapped sequence files were converted to BED and used as an input track for a genome-wide overview with ShinyCircos[95] using a 100kb window**.**

Comparative repeatome analysis of *Drosera* genomes was made using Illumina reads from the species listed in **Extended Data Table 1**. First, reads were filtered by quality with 95% of bases equal to or above the quality cut-off value of 10 using the RepeatExplorer2 pipeline (https://repeatexplorer-elixir.cerit-sc.cz/)[93]. The clustering was performed using the default settings of 90% similarity over 55% of the read length. For the comparative analyses, we performed an all-to-all similarity comparison across all species following the same approach. Because genome sizes are unknown for some analyzed species, each set of reads was down-sampled to 1,000,000 for each species. The automated annotations of repeat clusters obtained by RepeatExplorer2 were manually inspected and reviewed, followed by recalculation of the genomic proportion of each repeat type when appropriate.

Immunostaining and FISH. Flower buds of *D. capensis and D. regia* at various developmental stages were harvested and immediately fixed in freshly prepared 4% formaldehyde in Tris buffer (10 mM Tris-HCl (pH 7.5), 10 mM EDTA, and 0.5% Triton X-100), when immunostaining was intended, or ethanol-acetic acid (3:1 v/v) when only FISH was performed.

For immunostaining, fixation was carried out under vacuum infiltration for at least 15 minutes at room temperature, followed by an additional 45 minutes of incubation without vacuum. Fixed tissues were washed twice in 1× phosphate-buffered saline (1x PBS) at 4 °C until further processing. For enzymatic digestion, individual fixed buds were incubated in a solution containing 2% (w/v) cellulase (Onozuka R-10) and 2% (w/v) pectinase (Sigma) prepared in 1× PBS. Digestion was performed at 37 °C for 1 hour to facilitate cell wall degradation and release of nuclei. Following digestion, the softened tissue was gently macerated on a clean slide and squashed under the coverslip. Chromosome spreads were washed in 1× PBS for 5 minutes, followed by incubation in PBS-T1 buffer (1× PBS, 0.5% Triton X-100, pH 7.4) for 25 minutes. After two additional 5-minute washes in PBS, slides were incubated in PBS-T2 buffer (1× PBS, 0.1% Tween 20, pH 7.4) for 30 minutes. Primary antibody incubation was performed overnight at 4 °C using rabbit anti-CENH3 (specific for each species; **Extended data Fig.2**), rabbit anti-KNL1[96] (GenScript, NJ, USA) and mouse anti α-tubulin (Sigma-Aldrich, St. Louis, MO; catalog number T6199) diluted 1:1,000 in blocking buffer (3% BSA, 1x PBS, 0.1% Tween 20, pH 7.4). Slides were then washed twice in 1x PBS (5 minutes each) and once in PBS-T2 (5 minutes), followed by incubation with secondary antibodies for at least 1 hour at room temperature. As the secondary antibody, goat anti-rabbit IgG antibody conjugated with Alexa Fluor 488 (Invitrogen; catalog number A27034), goat anti-rabbit conjugated with Rhodamine Red X (Jackson ImmunoResearch, catalog number: 111-295-144)

or goat anti-mouse conjugated with Alexa Fluor 488 (Jackson ImmunoResearch; catalog number 115-545-166) were used in a 1:500 dilution. Final washes included two rounds in PBS and one in PBS-T2, each for 5 minutes. Slides were then mounted for fluorescence microscopy. Microscopic images were recorded using a Zeiss Axiovert 200M microscope equipped with a Zeiss AxioCam CCD. Images of at least 5 cells were analyzed using the ZEN software (Carl Zeiss GmbH).

For FISH experiments, material fixation was performed for 2 days at 4 °C, washed in ice-cold water, and digested in a solution of 4% cellulase (Onozuka R10, Serva Electrophoresis, Heidelberg, Germany), 2% pectinase, and 0.4% pectolyase Y23 (both MP Biomedicals, Santa Ana, CA) in 0.01 M citrate buffer (pH 4.5) for 60 min at 37 °C. The digested material was transferred to a drop of 45% acetic acid, macerated and squashed under a coverslip. FISH was performed using either oligonucleotide probes that were 5′-labeled with Cy3 during their synthesis (*D. regia*__Regi90 (Cy3)-CAAGTATTTCAATGGAAATGGTGAAATAACATGTTTTTACACCTATTTCC;

*D. capensis*__Cape71 (Cy3)-CCCTTTAAATGAGCTTAAAACACTCAAAACCCCTTGAAAAGGCTAAAAAC). FISH was performed as described in Macas, et al. 2007[97], with hybridization and washing temperatures adjusted to account for AT/GC content and hybridization stringency allowing for 10–20% mismatches. The slides were counterstained with 2 µg/mL DAPI in Vectashield (Vector) mounting medium. The images of at least 10 cells were captured as described above.

Satellite DNA phylogeny. Centromere tracks were generated using the consensus sequences of the main satellite repeats identified using TideCluster, and the coordinates of *Cape71* and *Regi90* satDNA polymers were used to guide sequence extraction in Geneious Prime (https://www.geneious.com). In total, 5,371 *Cape71* and 18,178 *Regi90* were extracted from *D. capensis* and *D. regia*, respectively. The collected centromeric repeat sequences were aligned using MAFFT v7.490[98]. Phylogenetic trees were inferred using FastTree v2.1.11[99] under the generalized time-reversible (GTR) model. Resulting trees were visualized and annotated using the Interactive Tree of Life (iTOL) web server[100], with colors corresponding to their chromosome of origin.

The phylogenetic trees were imported into R using the ape package[101]. Tip labels included chromosome ID and genomic coordinates. Genomic positions of sampled tips were plotted using ggplot2[81] as normalized positions (relative to chromosome length) in order to show the spatial distribution of tips along individual chromosomes. Tips were plotted using ggplot2, with color representing the order of appearance in the tree (NodeOrder), allowing assessment of how phylogenetic relationships correlate with genomic location. All analyses and plotting were performed in R (version 4.1.2) using the packages ape[101], dplyr[102], ggplot2[81], and viridis (https://sjmgarnier.github.io/viridis/).

Species tree inference with BUSCO genes. To extract BUSCO genes, we collected coding sequence sets from sequenced genomes as well as previously assembled transcriptomes[103] (**Extended Data Table 2**). We identified eudicot-conserved single-copy genes with BUSCO v5.3.2 (https://gitlab.com/ezlab/busco). Genes classified as single-copy (S) or fragmented (F) were retained, whereas those classified as duplicated (D) or missing (M) were considered absent, as described previously. Protein sequences were aligned with MAFFT v7.508 (https://mafft.cbrc.jp/alignment/server/index.html), trimmed with ClipKIT v2.1.1 (https://github.com/JLSteenwyk/ClipKIT), and back-translated to codons with CDSKIT

v0.10.10 (https://github.com/kfuku52/cdskit) to produce in-frame nucleotide alignments. For each gene, nucleotide and protein maximum-likelihood trees were inferred with IQ-TREE v2.2.5 using GTR+R4 and LG+R4 models, respectively. Individual gene trees were then used for coalescent species-tree inference with ASTRAL v5.7.8[45]. Concatenated nucleotide and protein alignments, generated by catfasta2phyml v2018-09-28 (https://github.com/nylander/catfasta2phyml), served as additional input to IQ-TREE[44] with the same substitution models. *Amborella trichopoda* was specified as the outgroup.

Data and code availability

The processed reference genomes, sequencing data, annotations and all data tracks generated in this study will be available on journal publication. The REXdb database Viridiplantae v.3.0 is publicly available at Github [https://github.com/repeatexplorer/rexdb]. Any original code used in this study will be made available at GitHub with journal publication

Acknowledgments

We acknowledge the following sources for funding: United States National Science Foundation grants (2030871 to V.A.A. and PRFB-2109178 to J.R.B.),), Human Frontier Science Program Young Investigators grant (RGY0082/2021 to K.F. and R.T.), JSPS KAKENHI (23K20050 to K.F.), Max Planck Society (core funding to A.M.), European Research Council (ERC StG HoloRECOMB, grant no. 101114879 to A.M.), the German Science Foundation (DFG MA 9363/6-1 to A.M.), the Research Council of Finland (decision 318288 to J.S). L.A.R. is financially supported by DFG (German Research Foundation) grant MA 9363/6-1. K.S. is supported by RIKEN TRIP initiative. The DFG also funded this work under Germany's Excellence Strategy—EXC 493 2048/1–390686111 (to A.M.).

Competing Interests

The authors declare no competing interests.

Extended Data Table 1. Quality assessment of the genomes.

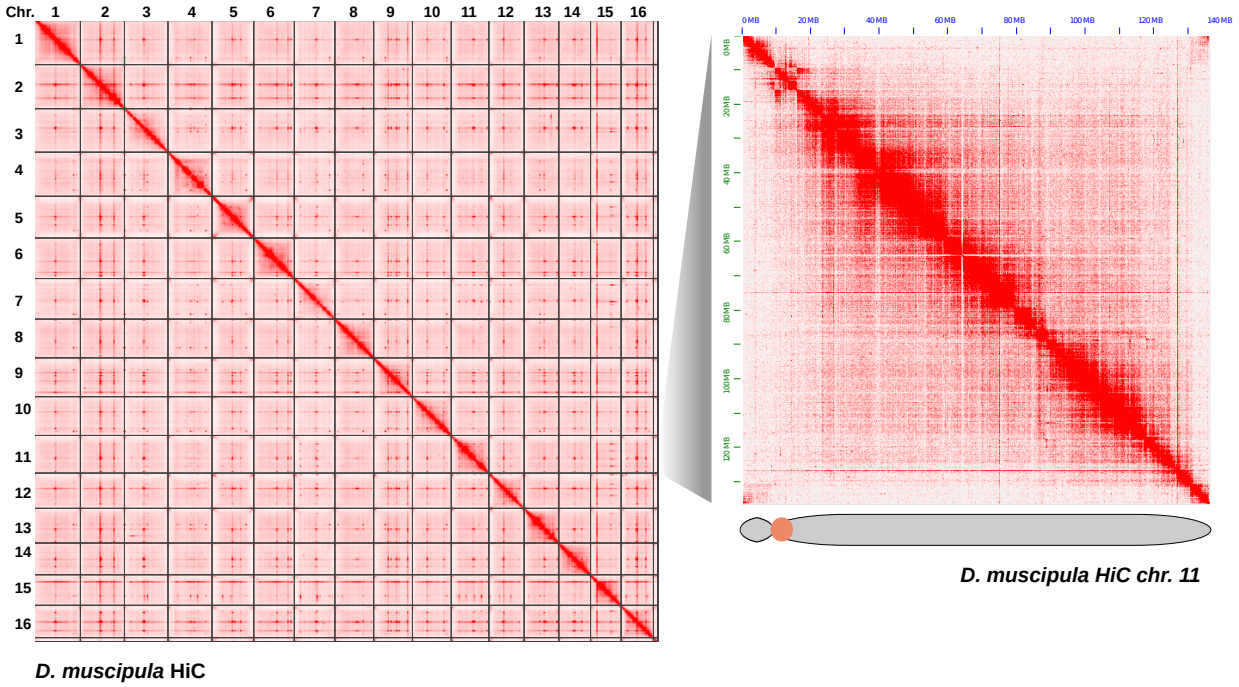
	BUSCO Library: eudicot	*Drosera_capensis*	*Drosera_regia*	*Triphyophyllum_peltatum*	*Ancistrocladus_abbreviatus*	*Dionaea_muscipula*
BUSCO	Complete BUSCOs (C)	2,086 (89.7%)	2,179 (93.7%)	2,255 (96.9%)	2,240 (96.3%)	2,128 (91.5%)
	Complete and single-copy BUSCOs (S	1,471 (63.2%)	1,859 (79.9%)	2,089 (89.8%)	2,033 (87.4%)	1,995 (85.8%)
	Complete and duplicated BUSCOs (D)	615 (26.4%)	320 (13.8%)	166 (7.1%)	207 (8.9%)	133 (5.7%)
	Fragmented BUSCOs (F)	37 (1.6%)	27 (1.2%)	18 (0.8%)	29 (1.2%)	37 (1.6%)
	Missing BUSCOs (M)	203 (8.7%)	120 (5.2%)	53 (2.3%)	57 (2.5%)	161 (6.9%)
	Total BUSCO groups searched	2,326 (100.0%)	2,326 (100.0%)	2,326 (100.0%)	2,326 (100.0%)	2,326 (100.0%)
	Number of scaffolds	1,089	1,338	429	131	3,706
	Number of contigs	1,431	3,275	468	653	11,425
	Total length	284,038,097	282,006,236	552,408,181	1,187,329,241	2,551,530,229
	Percent gaps	0	0	0	0	0
	Scaffold N50	12 MB	15 MB	26 MB	64 MB	144 MB
	Contigs N50	1 MB	287 KB	14 MB	4 MB	432 KB
	BUSCO Library: embryophyta	*Drosera_capensis*	*Drosera_regia*	*Triphyophyllum_peltatum*	*Ancistrocladus_abbreviatus*	*Dionaea_muscipula*
BUSCO	Complete BUSCOs (C)	1,530 (94.8%)	1,561 (96.7%)	1,578 (97.8%)	1,572 (97.4%)	1,540 (95.4%)
	Complete and single-copy BUSCOs (S	1,115 (69.1%)	1,390 (86.1%)	1,507 (93.4%)	1,459 (90.4%)	1,466 (90.8%)
	Complete and duplicated BUSCOs (D)	415 (25.7%)	171 (10.6%)	71 (4.4%)	113 (7.0%)	74 (4.6%)
	Fragmented BUSCOs (F)	23 (1.4%)	11 (0.7%)	15 (0.9%)	18 (1.1%)	25 (1.5%)
	Missing BUSCOs (M)	61 (3.8%)	42 (2.6%)	21 (1.3%)	24 (1.5%)	49 (3.0%)
	Total BUSCO groups searched	1,614 (100.0%)	1,614 (100.0%)	1,614 (100.0%)	1,614 (100.0%)	1,614 (100.0%)

	Number of scaffolds	1,089	1,338	429	131	3,706
	Number of contigs	1,431	3,275	468	653	11,425
	Total length	284,038,097	282,006,236	552,408,181	1,187,329,241	255,153,029
	Percent gaps	0	0	0	0	0
	Scaffold N50	12 MB	15 MB	26 MB	64 MB	144 MB
	Contigs N50	1 MB	287 KB	14 MB	4 MB	432 KB
QUAST	# contigs	1,089	1,338	429	131	3,706
	Largest contig	17,145,115	20,363,681	31,991,434	85,057,085	169,349,142
	Total length	284,038,097	282,006,236	552,408,181	1,187,329,241	2,551,530,229
	GC (%)	37	37	40	36	44
	N50	12,650,043	15,761,547	26,241,775	64,216,489	144,508,999
	N75	10,814,855	12,243,869	24,909,253	58,859,448	126,729,089
	L50	10	8	10	9	9
	L75	16	13	16	14	13
	# N's per 100 kbp	34	120	1	4	30

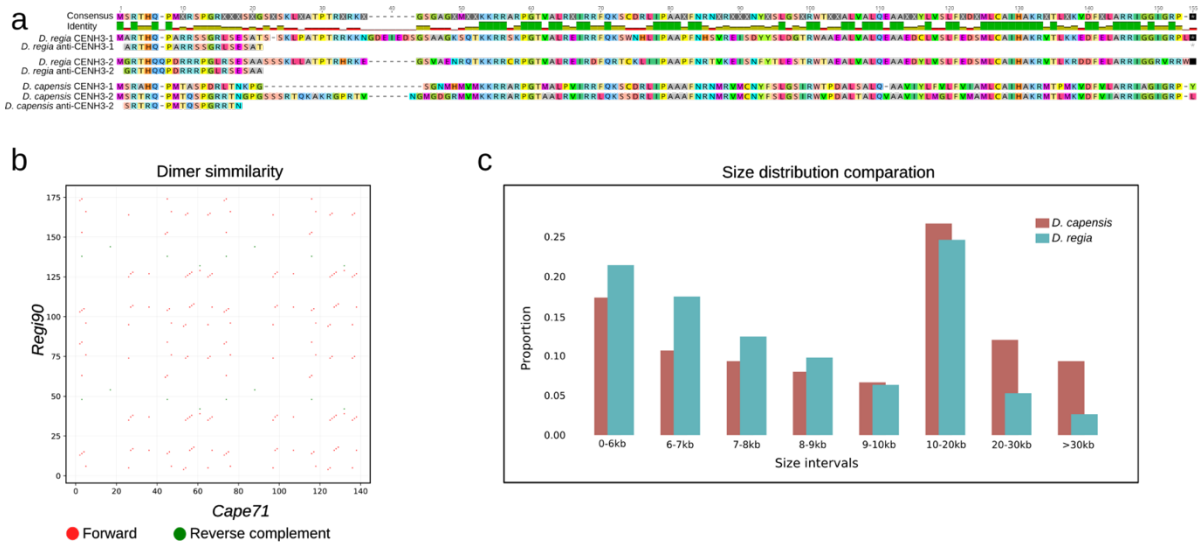
Extended data Table 2. Short reads analyzed in the repeatome comparative analysis.

Species	Total reads analyzed	Read length	Reads: repetitive Elements	% reads repetitive elements
N. gracilis	236096	150	78331	33.18
D. regia	235232	251	161825	68.79
D. aliciae	236124	251	156822	66.42
D. capensis	235410	251	178862	75.98
D. tokaiensis	235912	145	172730	73.22
D. anglica	235654	101	136961	58.12
D. intermedia	235694	251	203669	86.41
D. filliformis	234720	251	203766	86.81
D. broomensis	235806	251	158050	67.03
D. erythorhiza	235186	151	113059	48.07
D. peltata	235214	251	195065	82.93
D. menziesii	235850	251	184643	78.29
D. scorpioides	234890	251	172536	73.45

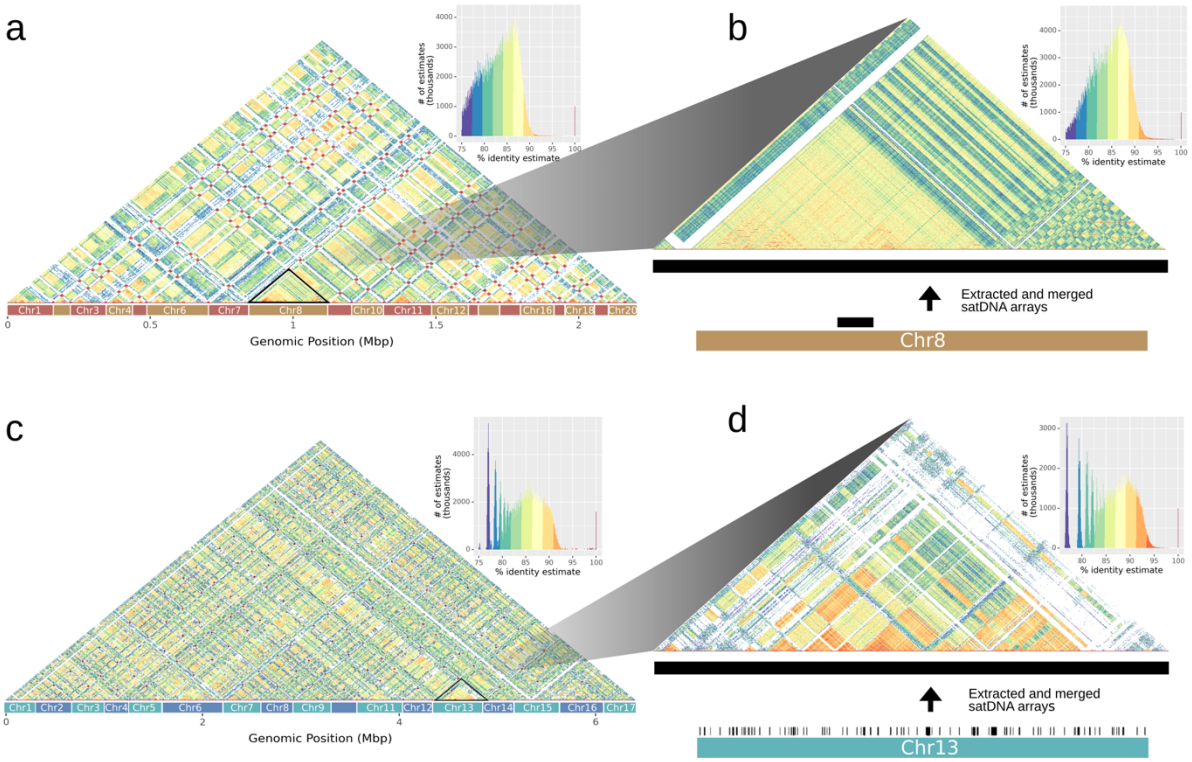
Extended Data Figures



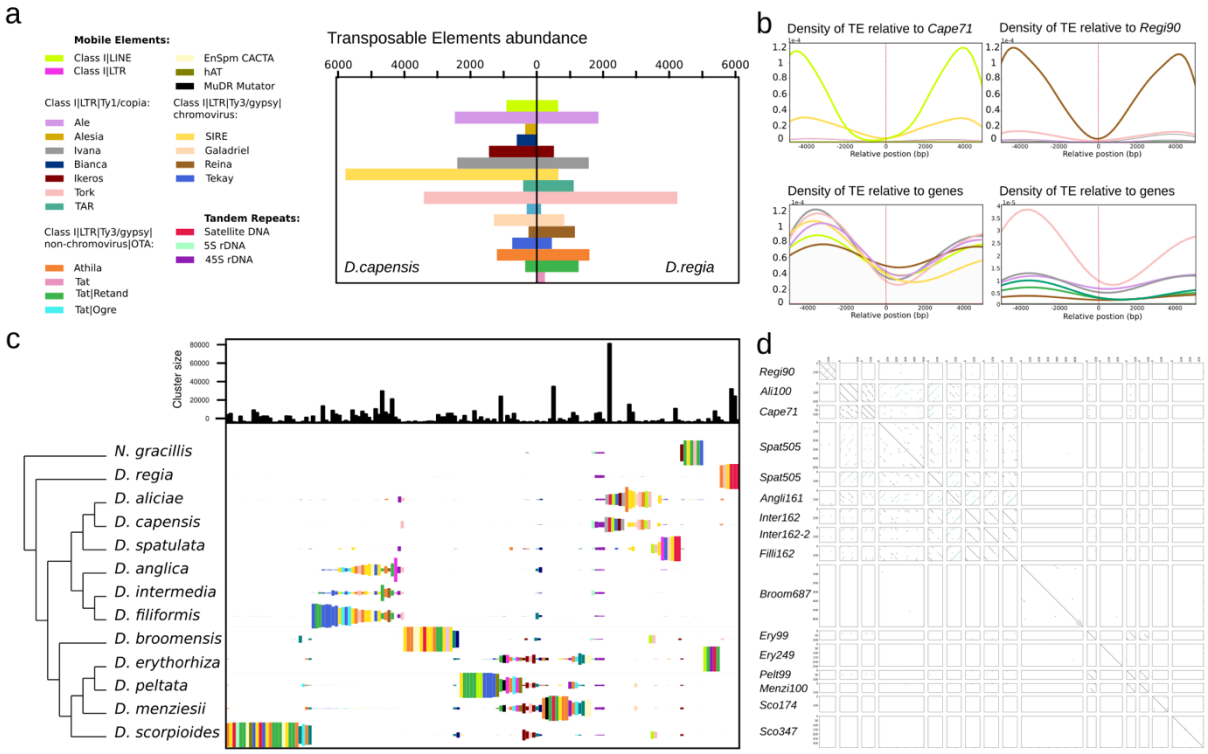
Extended data Fig. 1. *Dionaea muscipula* Hi-C contact map. Hi-C contact map showing every *D. muscipula* chromosome., plus a zoom-in to chromosome 11.



Extended Data Fig. 2: Characterization of CENH3 sequence and centromeric repeats in *D. regia* and *D. capensis*. (a) Sequence alignment of inferred CENH3 proteins found in *D. regia* and *D. capensis*. The antibody against CENH3 in each species is shown below each CENH3 sequence. **(b)** Dot plot highlighting short stretches of identity between the two satDNA families *Cape71* in *D. capensis* and *Regi90* in *D. regia.* **(c)** Size distribution comparison of *Cape71* and *Regi90*. Notice the prevalent array size of 10-20kbp for both species.

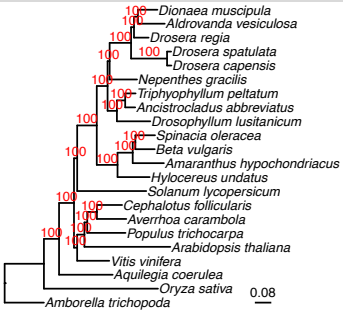


Extended Data Fig. 3: Comparison of centromeric sequence homogenization and the abundance of different transposable elements present in the *D. capensis* and *D. regia* genomes. **(a)** Genome-wide ModDot plot histograms of concatenated isolated centromeric arrays of all *D. capensis* chromosomes. **(b)** Close-up view of satellite array homogenization in chromosome 8 centromeric arrays. **(c)** Genome-wide ModDot plot histograms of concatenated isolated centromeric arrays of all *D. regia* chromosomes. **(d)** Close-up view of satellite array homogenization in chromosome 13 centromeric arrays.

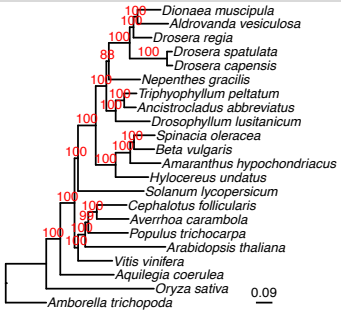


Extended Data Fig. 4: Comparison of the most abundant transposable-element (TE) families across *Drosera capensis* and *D. regia* genomes, and comparison of total repeats among 12 *Drosera* species plus *Nepenthes gracilis*. **(a)** Abundance differences of major TE families between *D. capensis* and *D. regia*. **(b)** Distribution of densities of TEs in relation to their position (+- 5000 bp) to nearby satDNA (upper panels), and genes (lower panels) **(c)** Relative contribution of each repetitive family to the total repetitive fraction in 12 *Drosera* species and *N. gracilis*. Black bars above the panels denote the total cluster size for each TE family (columns), while colored bars within each species row indicate the relative abundance of that family in that genome. **(d)** Dot plot comparison of the main satDNA families found in the genome of the 12 *Drosera* species analyzed.

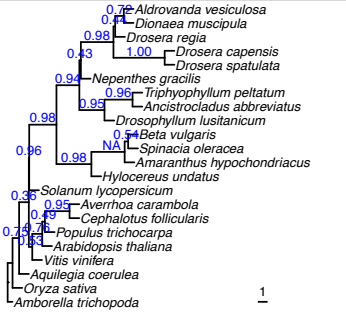
IQ–TREE v2.2.5 with concatenated DNA alignment
(3,548,928 sites, GTR+R model, loglik = –42,388,961.8)



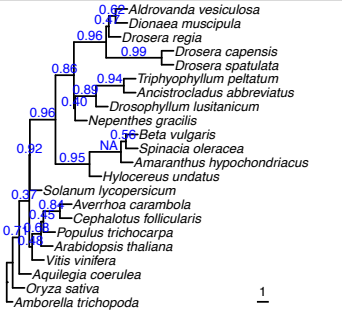
IQ–TREE v2.2.5 with concatenated protein alignment
(1,182,976 sites, LG+R model, loglik = –22,070,167.6)



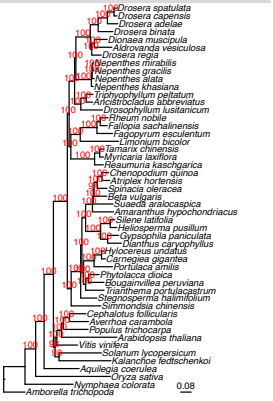
ASTRAL v5.7.8 with 2,325 DNA ML gene trees (GTR+R)



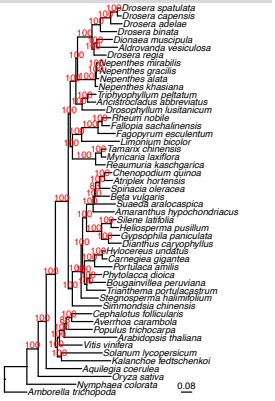
ASTRAL v5.7.8 with 2,325 protein ML gene trees (LG+R)



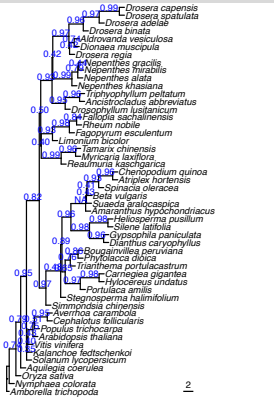
IQ–TREE v2.2.5 with concatenated DNA alignment
(3,349,866 sites, GTR+R model, loglik = –68,503,500.7)



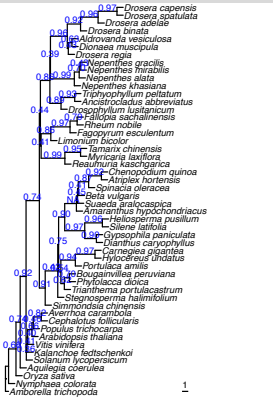
IQ–TREE v2.2.5 with concatenated protein alignment
(1,116,622 sites, LG+R model, loglik = –33,410,847.7)



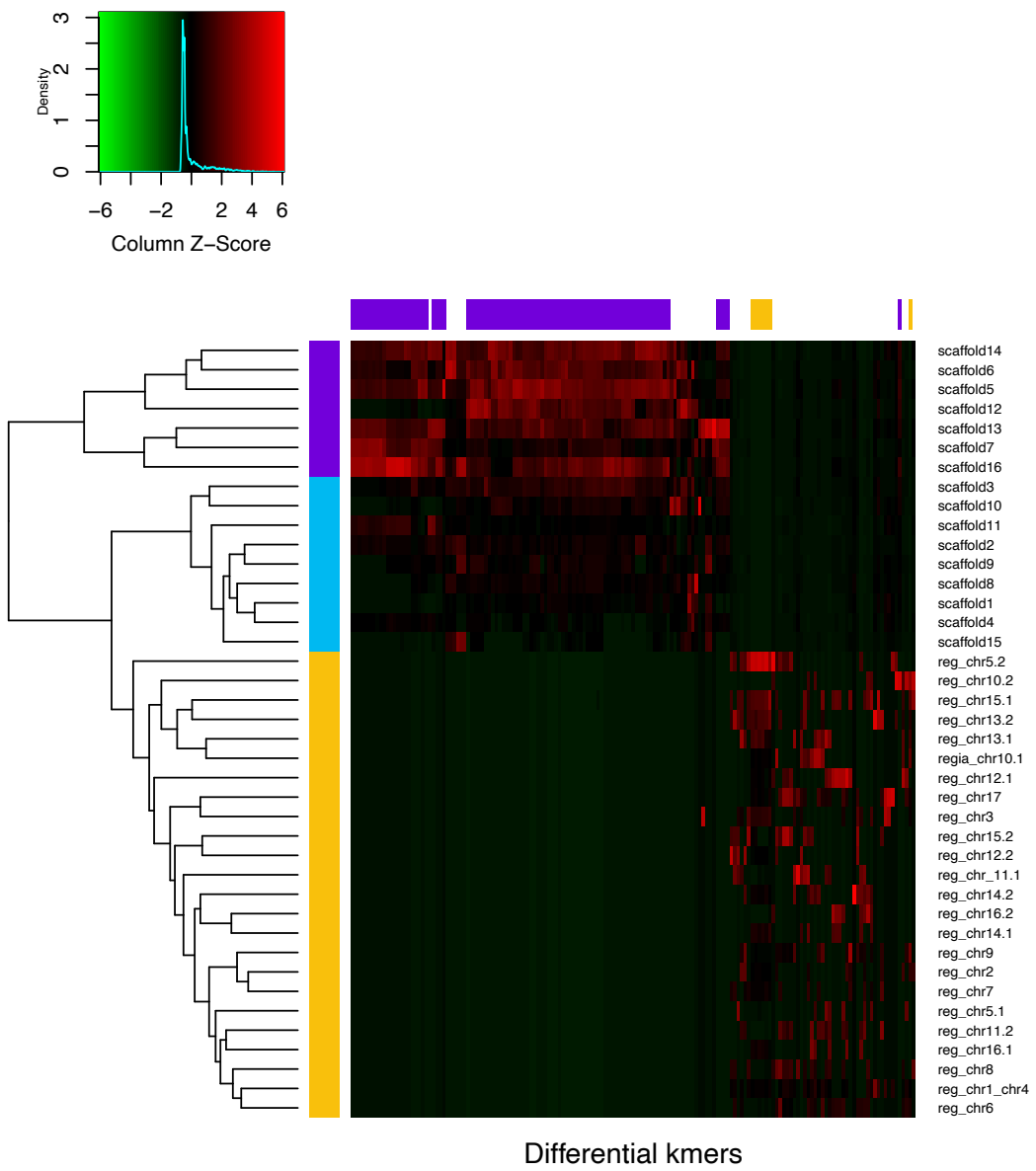
ASTRAL v5.7.8 with 2,325 DNA ML gene trees (GTR+R)



ASTRAL v5.7.8 with 2,325 protein ML gene trees (LG+R)

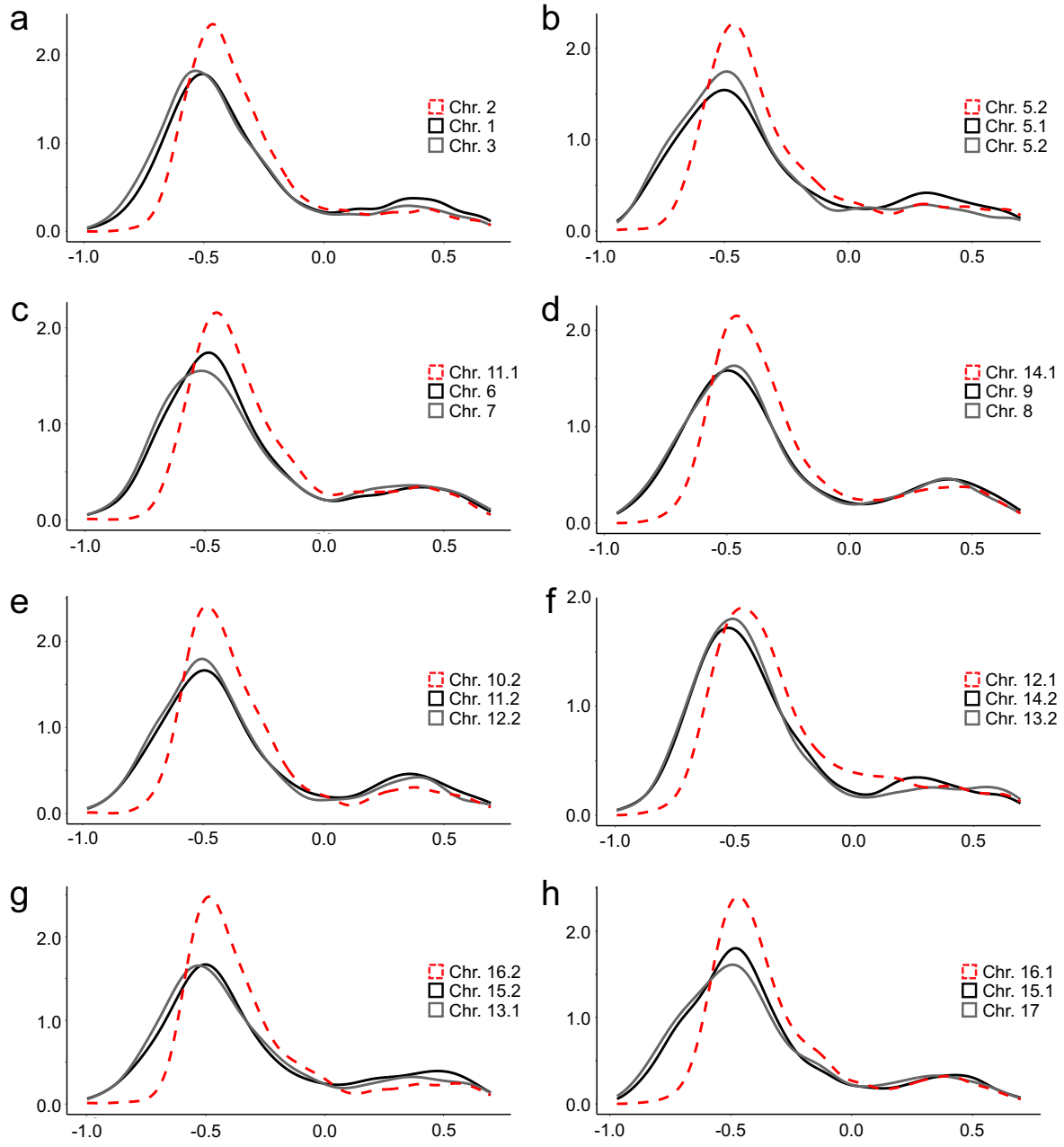


Extended Data Fig. 5. BUSCO gene and species trees. Two datasets were examined: nucleotide coding sequences or inferred amino acids from 22 species or 50 species. Phylogenetic relationships were reconstructed using two different approaches: (1) maximum likelihood (ML) analysis based on concatenated BUSCO gene alignments (**top**), and (2) ASTRAL species tree inference summarizing gene trees from individual BUSCOs (**bottom**). Both methods consistently placed *D. regia* as the sister group to *Aldrovanda* plus *Dionaea*, rather than to other *Drosera* species. *Nepenthes* was generally recovered as sister to *Drosera* and the snap-trapping lineages, while the clade containing *Drosophyllum* and *Triphyophyllum/Ancistrocladus* formed their sister group.

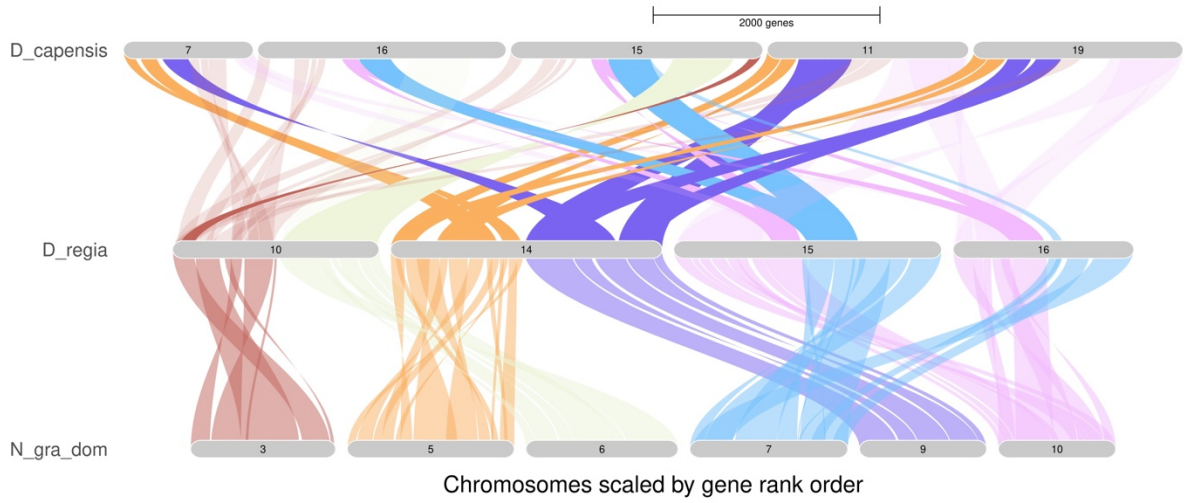


Extended Data Fig. 6: Unsupervised hierarchical clustering of ancestral chromosomes of *D. regia* and modern chromosomes of Dionaea. The horizontal color bar at the top (x-axis) indicates to which subgenome the *k*-mer is specific; the vertical color bar on the left (y-axis) indicates the subgenome to which the chromosome is assigned. The heatmap indicates the Z-scale relative abundance of *k*-mers. The larger the Z-score is, the greater the relative

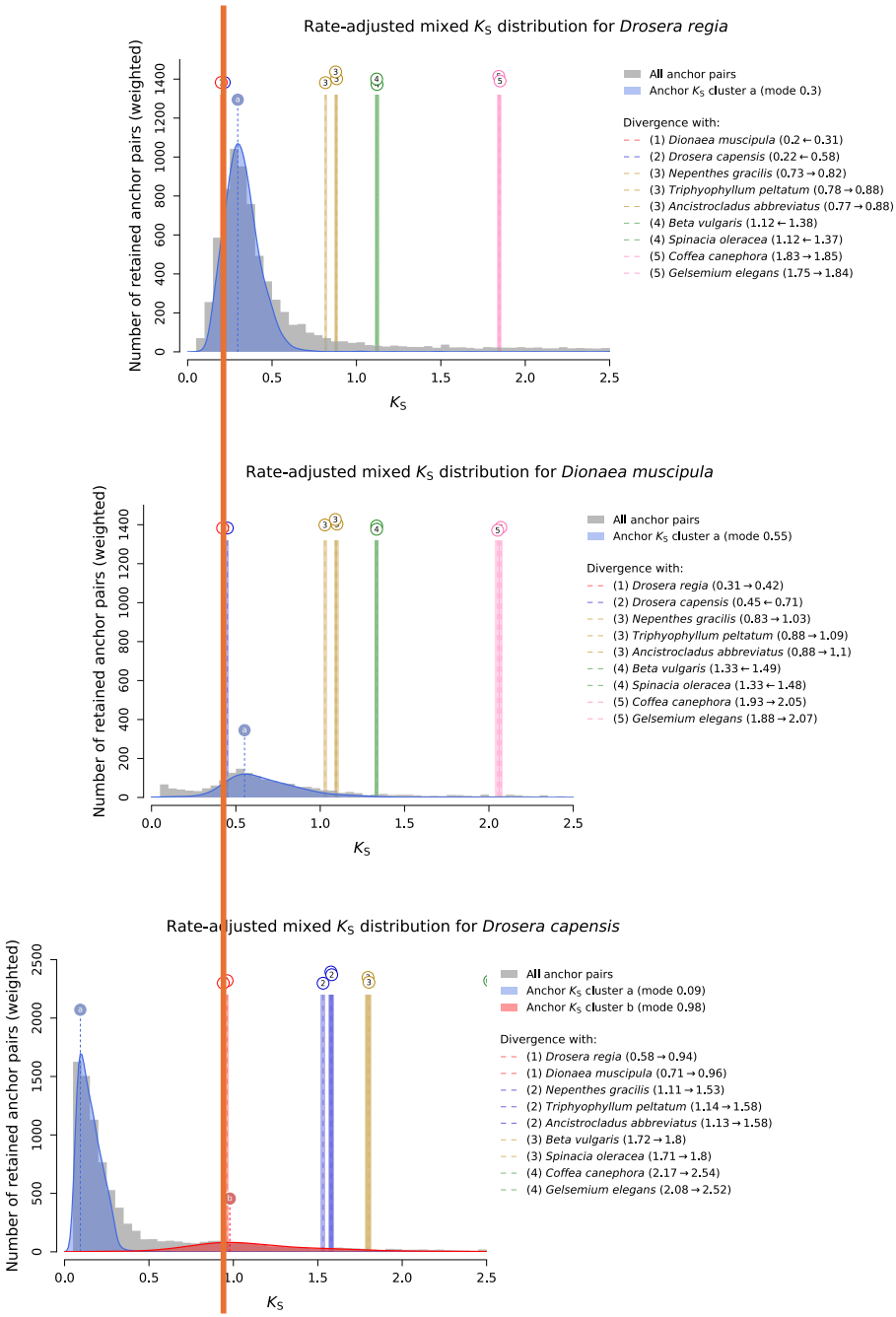
abundance of a *k*-mer. *Dionaea* subgenomes are represented in purple and blue colors, while *D. regia* (yellow) did not show any particular subgenome differentiation based on this approach.



Extended data Fig. 7: Ks density plots of homeologous gene pairs in *D. regia* reveals a clear triplicate subgenomic structure that exhibits a 2:1 configuration. Two of three chromosomes show similar distributions (gray), while a third (dashed red lines) stands out as an older subgenome, characterized by a higher Ks value (x-axis = $\log_{10}$ Ks; y-axis = density).



Extended Data Fig. 8: Shared chromosomal rearrangements and subgenome relationships across *Drosera* species, as compared to *Nepenthes*. GENESPACE riparian plot illustrating shared chromosomal rearrangements among selected chromosomes from the *N. gracilis* dominant subgenome, *D. regia* and *D. capensis*. This subset of chromosomes highlights key patterns supporting shared ancestral genomic restructuring. Note the orange:purple fusion in *D. regia* chromosome 14, and the three *D. capensis* chromosomes in this view (7, 11 and 19) that appear to show the same fusion. Likewise, note the pink:sky-blue fusion shared by *D. regia* chromosomes 15 and 16, which appears similarly reflected in *D. capensis* chromosomes 15 and 16.



Extended Fig. 9. *ksrates* synonymous substitutions calibration clarifies split times and polyploidy events for *D. regia*, *D. capensis*, and *Dionaea muscipula*. Here, focal species are (top to bottom) *D. regia*, *Dionaea*, and *D. capensis*. Colorations for events are the same for *D. regia* and *Dionaea*; they share the blue tetraploidy, which is the allotetraploidy still visible in *Dionaea*. For *D. regia*, one third of its hexaploid genome matches *Dionaea*; the hexaploidy was likely time-coincident with *Dionaea*'s allotetraploidy, with *D. regia*'s third subgenome possibly being one of the two original *Dionaea* parents. The species splits for *D. regia*, *Dionaea*, and *D. capensis* (aligned at the

orange vertical line) are also time-coincident, lying after the allotetraploidy and allohexaploidy events. While the *D. capensis*-focused analysis at the bottom appears different at a glance, it is not. Due to how the *ksrates* software handles plotting, the most recent tetraploidy is blue, and it occurred after *D. capensis* split from *D. regia* and *Dionaea* (orange line). With color coding changed, the older event shared with *D. regia* and *Dionaea* is red and is barely visible, with its remains in the modern *D. capensis* genome having been overlaid by an extremely recent tetraploidy event.

References cited

1. Albert, V.A., Williams, S.E., and Chase, M.W. (1992). Carnivorous plants: phylogeny and structural evolution. Science *257*, 1491-1495.
2. Fleischmann, A., Schlauer, J., Smith, S.A., Givnish, T.J., Ellison, A., and Adamec, L. (2018). Evolution of carnivory in angiosperms. Carnivorous plants: physiology, ecology, and evolution *1*, 22-42.
3. Fukushima, K., Fujita, H., Yamaguchi, T., Kawaguchi, M., Tsukaya, H., and Hasebe, M. (2015). Oriented cell division shapes carnivorous pitcher leaves of Sarracenia purpurea. Nature Communications *6*, 6450.
4. Whitewoods, C.D., Gonçalves, B., Cheng, J., Cui, M., Kennaway, R., Lee, K., Bushell, C., Yu, M., Piao, C., and Coen, E. (2020). Evolution of carnivorous traps from planar leaves through simple shifts in gene expression. Science *367*, 91-96.
5. Bemm, F., Becker, D., Larisch, C., Kreuzer, I., Escalante-Perez, M., Schulze, W.X., Ankenbrand, M., Van de Weyer, A.-L., Krol, E., and Al-Rasheid, K.A. (2016). Venus flytrap carnivorous lifestyle builds on herbivore defense strategies. Genome Research *26*, 812-825.
6. Pavlovič, A., and Mithöfer, A. (2019). Jasmonate signalling in carnivorous plants: copycat of plant defence mechanisms. Journal of experimental botany *70*, 3379-3389.
7. Adlassnig, W., Koller-Peroutka, M., Bauer, S., Koshkin, E., Lendl, T., and Lichtscheidl, I.K. (2012). Endocytotic uptake of nutrients in carnivorous plants. The Plant Journal *71*, 303-313.
8. Freund, M., Graus, D., Fleischmann, A., Gilbert, K.J., Lin, Q., Renner, T., Stigloher, C., Albert, V.A., Hedrich, R., and Fukushima, K. (2022). The digestive systems of carnivorous plants. Plant physiology *190*, 44-59.
9. Płachno, B., Adamec, L., Lichtscheidl, I., Peroutka, M., Adlassnig, W., and Vrba, J. (2006). Fluorescence labelling of phosphatase activity in digestive glands of carnivorous plants. Plant Biology *8*, 813-820.
10. Fukushima, K., Fang, X., Alvarez-Ponce, D., Cai, H., Carretero-Paulet, L., Chen, C., Chang, T.-H., Farr, K.M., Fujita, T., and Hiwatashi, Y. (2017). Genome of the pitcher plant Cephalotus reveals genetic changes associated with carnivory. Nature Ecology & Evolution *1*, 0059.
11. Lan, T., Renner, T., Ibarra-Laclette, E., Farr, K.M., Chang, T.-H., Cervantes-Pérez, S.A., Zheng, C., Sankoff, D., Tang, H., and Purbojati, R.W. (2017). Long-read sequencing uncovers the adaptive topography of a carnivorous plant genome. Proceedings of the National Academy of Sciences *114*, E4435-E4441.
12. Pavlovič, A. (2025). How the diversity in digestion in carnivorous plants may have evolved. New Phytologist.
13. Saul, F., Scharmann, M., Wakatake, T., Rajaraman, S., Marques, A., Freund, M., Bringmann, G., Channon, L., Becker, D., and Carroll, E. (2023). Subgenome dominance shapes novel gene evolution in the decaploid pitcher plant Nepenthes gracilis. Nature plants *9*, 2000-2015.
14. Fleischmann, A., Cross, A.T., Gibson, R., Gonella, P.M., and Dixon, K.W. (2018). Systematics and evolution of Droseraceae. Oxford Scholarship Online.
15. Metcalfe, C.R. (1951). The Anatomical Structure of the Dioncophyllaceae in Relation to the Taxonomic Affinities of the Family. Kew Bulletin *6*, 351-368. 10.2307/4118003.

16. GREEN, S., GREEN, T.L., and HESLOP-HARRISON, Y. (2008). Seasonal heterophylly and leaf gland features in Triphyophyllum (Dioncophyllaceae), a new carnivorous plant genus. Botanical Journal of the Linnean Society *78*, 99-116. 10.1111/j.1095-8339.1979.tb02188.x.
17. Winkelmann, T., Bringmann, G., Herwig, A., and Hedrich, R. (2023). Carnivory on demand: phosphorus deficiency induces glandular leaves in the African liana Triphyophyllum peltatum. New Phytologist *239*, 1140-1152. https://doi.org/10.1111/nph.18960.
18. Martín-Rodríguez, I., Vargas, P., Ojeda, F., and Fernández-Mazuecos, M. (2020). An enigmatic carnivorous plant: ancient divergence of Drosophyllaceae but recent differentiation of Drosophyllum lusitanicum across the Strait of Gibraltar. Systematics and Biodiversity *18*, 525-537. 10.1080/14772000.2020.1771467.
19. Płachno, B.J., Kapusta, M., Stolarczyk, P., Świątek, P., and Lichtscheidl, I. (2023). Differences in the Occurrence of Cell Wall Components between Distinct Cell Types in Glands of Drosophyllum lusitanicum. International Journal of Molecular Sciences *24*, 15045.
20. Junichi, S., Nagano, K., and Hoshi, Y. (2011). A chromosome study of two centromere differentiating Drosera species, D. arcturi and D. regia. Caryologia *64*, 453-463.
21. Kolodin, P., Cempírková, H., Bures, P., Horová, L., Veleba, A., Francová, J., Adamec, L., and Zedek, F. (2018). Holocentric chromosomes may be an apomorphy of Droseraceae. Plant Syst Evol *304*, 1289-1296. 10.1007/s00606-018-1546-8.
22. Sheikh, S.A., Kondo, K., and Hoshi, Y. (1995). Study of diffused centromeric nature of *Drosera* chromosomes. Cytologia *60*, 43-47.
23. Shirakawa, J., Hoshi, Y., and Kondo, K. (2011). Chromosome differentiation and genome organization in carnivorous plant family Droseraceae. Chromosome Botany *6*, 111-119. 10.3199/iscb.6.111.
24. Hofstatter, P.G., Thangavel, G., Lux, T., Neumann, P., Vondrak, T., Novak, P., Zhang, M., Costa, L., Castellani, M., Scott, A., et al. (2022). Repeat-based holocentromeres influence genome architecture and karyotype evolution. Cell. 10.1016/j.cell.2022.06.045.
25. Kuo, Y.T., Câmara, A.S., Schubert, V., Neumann, P., Macas, J., Melzer, M., Chen, J.Y., Fuchs, J., Abel, S., Klocke, E., et al. (2023). Holocentromeres can consist of merely a few megabase-sized satellite arrays. Nature Communications *14*. ARTN 3502 10.1038/s41467-023-38922-7.
26. Mata-Sucre, Y., Kratka, M., Oliveira, L., Neumann, P., Macas, J., Schubert, V., Huettel, B., Kejnovsky, E., Houben, A., Pedrosa-Harand, A., et al. (2024). Repeat-based holocentromeres of the woodrush Luzula sylvatica reveal insights into the evolutionary transition to holocentricity. Nat Commun *15*, 9565. 10.1038/s41467-024-53944-5.
27. Castellani, M., Zhang, M., Thangavel, G., Mata-Sucre, Y., Lux, T., Campoy, J.A., Marek, M., Huettel, B., Sun, H., Mayer, K.F.X., et al. (2024). Meiotic recombination dynamics in plants with repeat-based holocentromeres shed light on the primary drivers of crossover patterning. Nat Plants *10*, 423-438. 10.1038/s41477-024-01625-y.
28. Mohn, R.A., Zenil-Ferguson, R., Krueger, T.A., Fleischmann, A.S., Cross, A.T., and Yang, Y. (2023). Dramatic difference in rate of chromosome number evolution among sundew (Drosera L., Droseraceae) lineages. Evolution *77*, 2314-2325. 10.1093/evolut/qpad153.

29. Stephens, E.L. (1925). A NEW SUNDEW, DROSERA REGIA (STEPHENS), FROM THE CAPE PROVINCE. Transactions of the Royal Society of South Africa *13*, 309-312. 10.1080/00359192509519615.
30. Pavlovič, A., Vrobel, O., and Tarkowski, P. (2023). Water cannot activate traps of the carnivorous sundew plant Drosera capensis: On the trail of Darwin's 150-years-old mystery. Plants *12*, 1820.
31. Krausko, M., Perutka, Z., Šebela, M., Šamajová, O., Šamaj, J., Novák, O., and Pavlovič, A. (2017). The role of electrical and jasmonate signalling in the recognition of captured prey in the carnivorous sundew plant Drosera capensis. New Phytologist *213*, 1818-1835. https://doi.org/10.1111/nph.14352.
32. Zedek, F., and Bureš, P. (2017). Holocentric chromosomes: from tolerance to fragmentation to colonization of the land. Annals of Botany *121*, 9-16. 10.1093/aob/mcx118.
33. Veleba, A., Šmarda, P., Zedek, F., Horová, L., Šmerda, J., and Bureš, P. (2016). Evolution of genome size and genomic GC content in carnivorous holokinetics (Droseraceae). Annals of Botany *119*, 409-416. 10.1093/aob/mcw229.
34. Behre, K. (1929). Physiologische und zytologische Untersuchungen über Drosera (Springer).
35. Costa, L., Castro, N., Buddenhagen, C.E., Marques, A., Pedrosa-Harand, A., and Souza, G. (2024). Repeat competition and ecological shifts drive the evolution of the mobilome in Rhynchospora Vahl (Cyperaceae), the holocentric beaksedges. Annals of Botany *135*, 909-924. 10.1093/aob/mcae220.
36. Marques, A., Ribeiro, T., Neumann, P., Macas, J., Novák, P., Schubert, V., Pellino, M., Fuchs, J., Ma, W., and Kuhlmann, M. (2015). Holocentromeres in Rhynchospora are associated with genome-wide centromere-specific repeat arrays interspersed among euchromatin. Proceedings of the National Academy of Sciences *112*, 13633-13638.
37. Logsdon, G.A., Rozanski, A.N., Ryabov, F., Potapova, T., Shepelev, V.A., Catacchio, C.R., Porubsky, D., Mao, Y., Yoo, D., and Rautiainen, M. (2024). The variation and evolution of complete human centromeres. Nature *629*, 136-145.
38. Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J.L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., and Panaud, O. (2007). A unified classification system for eukaryotic transposable elements. Nature reviews genetics *8*, 973-982.
39. Neumann, P., Navrátilová, A., Koblížková, A., Kejnovský, E., Hřibová, E., Hobza, R., Widmer, A., Doležel, J., and Macas, J. (2011). Plant centromeric retrotransposons: a structural and cytogenetic perspective. Mobile DNA *2*, 4.
40. Bousios, A., Kakutani, T., and Henderson, I.R. (2025). Centrophilic retrotransposons of plant genomes. Annual Review of Plant Biology *76*.
41. Wells, J.N., and Feschotte, C. (2020). A field guide to eukaryotic transposable elements. Annual review of genetics *54*, 539-561.
42. Veleba, A., Smarda, P., Zedek, F., Horova, L., Smerda, J., and Bures, P. (2017). Evolution of genome size and genomic GC content in carnivorous holokinetics (Droseraceae). Annals of Botany *119*, 409-416. 10.1093/aob/mcw229.
43. Manni, M., Berkeley, M.R., Seppey, M., and Zdobnov, E.M. (2021). BUSCO: assessing genomic data quality and beyond. Current Protocols *1*, e323.

44. Wong, T.K., Ly-Trong, N., Ren, H., Baños, H., Roger, A.J., Susko, E., Bielow, C., De Maio, N., Goldman, N., and Hahn, M.W. (2025). IQ-TREE 3: Phylogenomic Inference Software using Complex Evolutionary Models.
45. Mirarab, S., Reaz, R., Bayzid, M.S., Zimmermann, T., Swenson, M.S., and Warnow, T. (2014). ASTRAL: genome-scale coalescent-based species tree estimation. Bioinformatics *30*, i541-i548.
46. Stull, G.W., Pham, K.K., Soltis, P.S., and Soltis, D.E. (2023). Deep reticulation: the long legacy of hybridization in vascular plant evolution. The Plant Journal *114*, 743-766.
47. Bouckaert, R.R. (2010). DensiTree: making sense of sets of phylogenetic trees. Bioinformatics *26*, 1372-1373. 10.1093/bioinformatics/btq110.
48. Lovell, J.T., Sreedasyam, A., Schranz, M.E., Wilson, M., Carlson, J.W., Harkess, A., Emms, D., Goodstein, D.M., and Schmutz, J. (2022). GENESPACE tracks regions of interest and gene copy number variation across multiple genomes. elife *11*, e78526.
49. Albert, V.A., and Krabbenhoft, T.J. (2023). Navigating the CoGe Online Software Suite for Polyploidy Research. In Polyploidy: Methods and Protocols, Y. Van de Peer, ed. (Springer US), pp. 19-45. 10.1007/978-1-0716-2561-3_2.
50. Joyce, B.L., Haug-Baltzell, A., Davey, S., Bomhoff, M., Schnable, J.C., and Lyons, E. (2016). FractBias: a graphical tool for assessing fractionation bias following polyploidy. Bioinformatics *33*, 552-554. 10.1093/bioinformatics/btw666.
51. Sensalari, C., Maere, S., and Lohaus, R. (2021). ksrates: positioning whole-genome duplications relative to speciation events in KS distributions. Bioinformatics *38*, 530-532. 10.1093/bioinformatics/btab602.
52. Sankoff, D., and Blanchette, M. (1998). Multiple genome rearrangement and breakpoint phylogeny. Journal of computational biology *5*, 555-570.
53. Sankoff, D., Zheng, C., and Zhu, Q. (2007). Polyploids, genome halving and phylogeny. Bioinformatics *23*, i433-i439. 10.1093/bioinformatics/btm169.
54. Marques, A., and Drinnenberg, I.A. (2025). Same but different: Centromere regulations in holocentric insects and plants. Current Opinion in Cell Biology *93*, 102484.
55. Melters, D.P., Paliulis, L.V., Korf, I.F., and Chan, S.W. (2012). Holocentric chromosomes: convergent evolution, meiotic adaptations, and genomic analysis. Chromosome Research *20*, 579-593.
56. Escudero, M., Márquez-Corro, J.I., and Hipp, A.L. (2016). The phylogenetic origins and evolutionary history of holocentric chromosomes. Systematic Botany *41*, 580-585.
57. Oliveira, L., Neumann, P., Mata-Sucre, Y., Kuo, Y.T., Marques, A., Schubert, V., and Macas, J. (2024). KNL1 and NDC80 represent new universal markers for the detection of functional centromeres in plants. Chromosome Research *32*. ARTN 3
10.1007/s10577-024-09747-x.
58. de Tomás, C., and Vicient, C.M. (2023). The genomic shock hypothesis: genetic and epigenetic alterations of transposable elements after interspecific hybridization in plants. Epigenomes *8*, 2.
59. Steinmüller, K., and Apel, K. (1986). A simple and efficient procedure for isolating plant chromatin which is suitable for studies of DNase I-sensitive domains and hypersensitive sites. Plant molecular biology *7*, 87-94.
60. Hatano, S., Yamaguchi, J., and Hirai, A. (1992). The preparation of high-molecular-weight DNA from rice and its analysis by pulsed-field gel electrophoresis. Plant Science *83*, 55-64.

61. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114-2120.
62. Zimin, A.V., Puiu, D., Luo, M.-C., Zhu, T., Koren, S., Marçais, G., Yorke, J.A., Dvořák, J., and Salzberg, S.L. (2017). Hybrid assembly of the large and highly repetitive genome of Aegilops tauschii, a progenitor of bread wheat, with the MaSuRCA mega-reads algorithm. Genome research *27*, 787-792.
63. Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. Bioinformatics *29*, 1072-1075.
64. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. bioinformatics *25*, 1754-1760.
65. Durand, N.C., Shamim, M.S., Machol, I., Rao, S.S., Huntley, M.H., Lander, E.S., and Aiden, E.L. (2016). Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell systems *3*, 95-98.
66. Rao, S.S., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., and Lander, E.S. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell *159*, 1665-1680.
67. Robertson, G., Schein, J., Chiu, R., Corbett, R., Field, M., Jackman, S.D., Mungall, K., Lee, S., Okada, H.M., and Qian, J.Q. (2010). De novo assembly and analysis of RNA-seq data. Nature methods *7*, 909-912.
68. Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., and Lieber, M. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nature protocols *8*, 1494-1512.
69. Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nature biotechnology *37*, 907-915.
70. Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.-C., Mendell, J.T., and Salzberg, S.L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nature biotechnology *33*, 290-295.
71. Gilbert, D. (2013). Gene-omes built from mRNA-seq not genome DNA.
72. Flynn, J.M., Hubley, R., Goubert, C., Rosen, J., Clark, A.G., Feschotte, C., and Smit, A.F. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. Proceedings of the National Academy of Sciences *117*, 9451-9457.
73. Chen, N. (2004). Using Repeat Masker to identify repetitive elements in genomic sequences. Current protocols in bioinformatics *5*, 4.10. 11-14.10. 14.
74. Haas, B.J., Salzberg, S.L., Zhu, W., Pertea, M., Allen, J.E., Orvis, J., White, O., Buell, C.R., and Wortman, J.R. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome biology *9*, R7.
75. Hoff, K.J., Lange, S., Lomsadze, A., Borodovsky, M., and Stanke, M. (2016). BRAKER1: unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. Bioinformatics *32*, 767-769.
76. Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., and Morgenstern, B. (2006). AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic acids research *34*, W435-W439.

77. Lomsadze, A., Ter-Hovhannisyan, V., Chernoff, Y.O., and Borodovsky, M. (2005). Gene identification in novel eukaryotic genomes by self-training algorithm. Nucleic Acids Research *33*, 6494-6506. 10.1093/nar/gki937.
78. Slater, G.S.C., and Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. BMC bioinformatics *6*, 31.
79. Jia, K.H., Wang, Z.X., Wang, L., Li, G.Y., Zhang, W., Wang, X.L., Xu, F.J., Jiao, S.Q., Zhou, S.S., and Liu, H. (2022). SubPhaser: a robust allopolyploid subgenome phasing method based on subgenome-specific k-mers. New Phytologist *235*, 801-809.
80. Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L.D.A., François, R., Grolemund, G., Hayes, A., Henry, L., and Hester, J. (2019). Welcome to the Tidyverse. Journal of open source software *4*, 1686.
81. Hadley, W. (2016). Ggplot2: Elegrant graphics for data analysis (Springer).
82. Neuwirth, E., and Neuwirth, M.E. (2014). Package 'RColorBrewer'. ColorBrewer Palettes.
83. Wilke, C.O. (2021). Ridgeline Plots in 'ggplot2'[R Package Ggridges Version 0.5. 3]. January. https://cran. r-project. org/web/packages/ggridges/index. html.
84. Aphalo, P. (2020). ggpmisc: Miscellaneous Extensions to "ggplot2"(R package version 0.3. 6).
85. Sensalari, C., Maere, S., and Lohaus, R. (2022). ksrates: positioning whole-genome duplications relative to speciation events in KS distributions. Bioinformatics *38*, 530-532.
86. Dainat, J. (2022). AGAT: Another Gff Analysis Toolkit to handle annotations in any GTF/GFF format (version 1.4.0). https://zenodo.org/records/6488306.
87. Denoeud, F., Carretero-Paulet, L., Dereeper, A., Droc, G., Guyot, R., Pietrella, M., Zheng, C., Alberti, A., Anthony, F., and Aprea, G. (2014). The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. science *345*, 1181-1184.
88. Emms, D.M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome biology *20*, 1-14.
89. Emms, D.M., and Kelly, S. (2018). STAG: Species tree inference from all genes. Preprint at https://doi.org/10.1101/267914.
90. Emms, D.M., and Kelly, S. (2017). STRIDE: species tree root inference from gene duplication events. Molecular biology and evolution *34*, 3267-3278.
91. Novák, P., Hoštáková, N., Neumann, P., and Macas, J. (2024). DANTE and DANTE_LTR: lineage-centric annotation pipelines for long terminal repeat retrotransposons in plant genomes. NAR Genomics and Bioinformatics *6*. 10.1093/nargab/lqae113.
92. Neumann, P., Novák, P., Hoštáková, N., and Macas, J. (2019). Systematic survey of plant LTR-retrotransposons elucidates phylogenetic relationships of their polyprotein domains and provides a reference for element classification. Mobile DNA *10*, 1. 10.1186/s13100-018-0144-1.
93. Novák, P., Neumann, P., and Macas, J. (2020). Global analysis of repetitive DNA from unassembled sequence reads using RepeatExplorer2. Nature Protocols *15*, 3745-3776. 10.1038/s41596-020-0400-y.
94. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. Journal of molecular biology *215*, 403-410.
95. Yu, Y., Ouyang, Y., and Yao, W. (2017). shinyCircos: an R/Shiny application for interactive creation of Circos plot. Bioinformatics *34*, 1229-1231. 10.1093/bioinformatics/btx763.

96. Oliveira, L., Neumann, P., Mata-Sucre, Y., Kuo, Y.-T., Marques, A., Schubert, V., and Macas, J. (2024). KNL1 and NDC80 represent new universal markers for the detection of functional centromeres in plants. Chromosome Research *32*, 3. 10.1007/s10577-024-09747-x.
97. Macas, J., Neumann, P., and Navrátilová, A. (2007). Repetitive DNA in the pea (Pisum sativum L.) genome: comprehensive characterization using 454 sequencing and comparison to soybean and Medicago truncatula. BMC Genomics *8*, 427. 10.1186/1471-2164-8-427.
98. Katoh, K., and Standley, D.M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Molecular Biology and Evolution *30*, 772-780. 10.1093/molbev/mst010.
99. Price, M.N., Dehal, P.S., and Arkin, A.P. (2010). FastTree 2–approximately maximum-likelihood trees for large alignments. PloS one *5*, e9490.
100. Letunic, I., and Bork, P. (2024). Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. Nucleic acids research *52*, W78-W82.
101. Paradis, E., and Schliep, K. (2018). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics *35*, 526-528. 10.1093/bioinformatics/bty633.
102. Wickham, H. (2015). dplyr: A grammar of data manipulation. R package version 04. *3*, p156.
103. Montero, H., Freund, M., and Fukushima, K. (2025). Convergent losses of arbuscular mycorrhizal symbiosis in carnivorous plants. bioRxiv, 2025.2004. 2003.646726.