Centromere divergence and allopolyploidy reshape carnivorous sundew genomes

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[9], 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 -0000-0002-2353-9274
Jarkko Salojärvi - 0000-0002-4096-6278
André Marques - 0000-0002-9567-2576
Victor A. Albert - 0000-0002-0262-826X

Abstract

Centromeres are essential for chromosome function, yet their role in shaping genome evolution in polyploid plants remains poorly understood. Allopolyploidy, where post-hybridization genome doubling merges parental genomes that may differ markedly in chromosomal architecture, has the potential to increase centromeric complexity and influence genomic plasticity. We explore this possibility in carnivorous Caryophyllales, a morphologically and chromosomally diverse plant lineage encompassing sundews, Venus flytraps, and *Nepenthes* pitcher plants. Focusing on sundews (*Drosera*), we generated chromosome-scale assemblies of holocentric *D. regia* and monocentric *D. capensis*, which share an allohexaploid origin but have diverged dramatically in genome structure. *D. regia* retains ancestral chromosomal fusions, dispersed centromeric repeats, and conserved synteny, whereas *D. capensis* exhibits extensive chromosomal reorganization and regionally localized centromeres after a lineage-specific genome duplication. Phylogenomic evidence traces *D. regia* to an ancient hybridization between sundew- and Venus flytrap-like ancestors, setting it apart within its infrageneric context. Genus-wide satellite DNA repeat profiling reveals rapid turnover and species-level variation in centromere organization. Together, these results establish sundews as a natural system for investigating how centromere dynamics interact with recurrent polyploidization and episodes of ecological innovation to shape genomic resilience.

Introduction

Flowering plants have diversified into nearly every terrestrial habitat by evolving a wide array of physiological and morphological strategies to secure essential nutrients. Among the most prominent departures from typical angiosperm organization are carnivorous plants, which have independently evolved mechanisms to trap and digest animal prey, particularly to supplement nitrogen and phosphorus in nutrient-poor environments[1,2]. This transition to carnivory involves extensive shifts in leaf development[3,4] and repurposing of defense responses[5,6], nutrient transport[7,8] and enzymatic secretion[8,9]. These changes frequently co-opt genetic pathways associated with plant immunity and senescence[10-13].

Carnivory has evolved independently at least ten times across five angiosperm orders[2]. These convergent origins have produced a diversity of trap architectures, including snap traps, suction bladders, and pitfall pitchers[2]. Particularly notable is the convergent evolution of pitcher traps in three distantly related angiosperm lineages that diverged more than 100 million years ago and are separated by large numbers of non-carnivorous taxa[1,2]. Two especially diverse carnivorous groups are found within Lamiales (e.g., *Utricularia*, *Pinguicula*, *Genlisea*) and Caryophyllales (e.g., *Drosera*, *Dionaea*, *Nepenthes*)[1,2]. These clades exemplify some of the most conspicuous evolutionary innovations among angiosperms.

Within Caryophyllales, the genus *Drosera* (sundews) includes over 200 species that use adhesive flypaper traps to capture insect prey[14]. The trapping surface is formed by mucilage-secreting glandular tentacles on the leaf lamina. Although this architecture is broadly conserved across the genus, *Drosera* displays considerable ecological and morphological diversity. Closely related genera include the snap-trapping *Dionaea* (Venus flytrap) and *Aldrovanda*, the morphologically divergent flypaper trappers *Triphyophyllum*[15-17] and *Drosophyllum*[18,19], and a group of secondarily non-carnivorous lianas (*Ancistrocladus*, *Habropetalum*, *Dioncophyllum*)[15]. Together, these lineages form a unique evolutionary radiation involving multiple transitions in trap form and function.

Beyond their ecological specializations, sundews are notable for variation 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, where centromeric function is distributed along the length of the chromosome[20,22,23]. These claims have remained unresolved due to the small size of sundew chromosomes, which complicates cytological interpretation, and the absence of genome-scale data to confirm centromere structure and function.

Recent studies across eukaryotes have emphasized the central role of centromere organization in shaping genome architecture[24-26]. Factors such as repeat content, transposable element distribution, histone modification patterns, and recombination localization all influence large-scale chromosomal behavior[24-27]. In holocentric species, chromosomal fissions and fusions may be tolerated without loss of centromere function, potentially facilitating more rapid genome restructuring[24-26]. Despite its potential importance, the evolutionary and mechanistic significance of holocentricity in flowering plants remains poorly characterized.

Here we present chromosome-scale genome assemblies for two species of *Drosera*: *D. regia*, a morphologically and phylogenetically distinct lineage often considered sister to all other sundews[28,29], and *D. capensis*, a widely cultivated species positioned deeper within the genus[30,31]. These species differ in evolutionary placement, previously reported centromere structure (*D. regia* monocentric[20] and *D. capensis* holocentric[28,32,33]), and ploidy level, with *D.*

regia inferred to be hexaploid and *D. capensis* tetraploid[34]. Until now, however, these distinctions had not been supported by genome-scale data.

Our assemblies reveal that both species possess relatively small genomes (less than 300 megabases) shaped by ancient polyploidy. Contrary to earlier cytological reports, our cytogenomic analyses show that *D. regia* is holocentric, suggesting that holocentricity may be more widespread across the genus than previously recognized. Its post-polyploid history includes extensive chromosomal fusions and largely conserved synteny. In contrast, *D. capensis* possesses monocentric chromosomes and a structurally reorganized, dodecaploid genome characterized by disrupted synteny relative to *D. regia* and other carnivorous Caryophyllales, including *Dionaea*, *Triphyophyllum*, *Ancistrocladus*, and *Nepenthes*.

To extend our findings, we generated short-read sequencing data for ten additional *Drosera* taxa. This revealed lineage-specific variation in repeat dynamics, implying that centromere organization may strongly influence genome evolution even within a single genus. Finally, phylogenomic analysis supports an allopolyploid origin for *D. regia*, derived from hybridization between sundew-like and Venus flytrap-like ancestors. Together, these results shed light on the complex evolutionary history of *Drosera*, and demonstrate how centromere biology and hybridization have jointly shaped genome evolution in this dynamic carnivorous plant lineage.

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 centromeric foci distributed along the length of the chromosome[20-23]. 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**), indicating a non-monocentric structure. The *D. regia* repeat landscape mirrored this difference: satDNA 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.[24], 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 *Dionaea muscipula* reflected a chromosomal architecture comparable to that of *D. capensis*, in contrast to previous non-genomic research that suggested holocentricity for *Dionaea muscipula*[21] (**Extended data Fig. 1**).

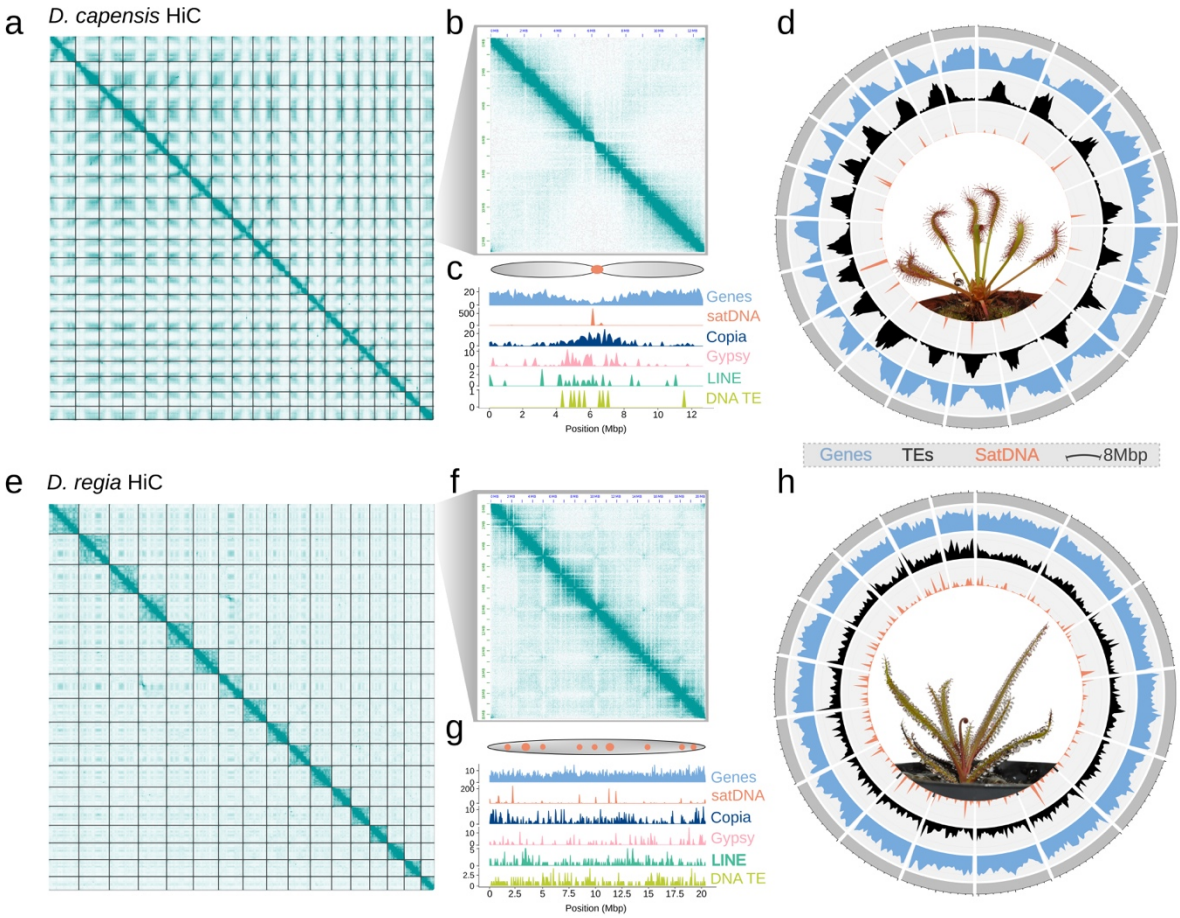


Fig. 1: Contrasting spatial genome organization in monocentric *D. capensis* and holocentric *D. regia* chromosomes. (**a, e**) Genome-wide Hi-C contact maps and (**b, f**) single-chromosome views show the distribution of major repetitive sequence types (**c, g**), clustered in *D. capensis* and uniform in *D. regia*, reflecting typical monocentric and holocentric patterns, respectively. Sequence-type density was calculated in 100 kb windows. (**d, h**) Circos plots of gene, TE, and satDNA distributions in *D. capensis* (**d**) and *D. regia* (**h**) further illustrate the expected mono- vs. holocentric organization.

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

We examined the genomes of *D. regia* and *D. capensis* to identify CENH3, a centromere-specific variant of histone H3 that plays an essential role in kinetochore assembly and is notably

divergent across species[24]. In both species, two copies of CENH3 were identified, and antibodies were generated against their variable N-terminal tails (**Extended Data Fig. 2a**). Immunostaining using these CENH3 antibodies produced distinct foci in the interphase nuclei of *D. capensis*, confirming monocentric chromosome organization (**Fig. 2a**). In contrast, no specific signal was observed in interphase nuclei of *D. regia* (**Fig. 2b**).

Given that some holocentric species do not require CENH3 for centromere function[25,26], we investigated the presence of the kinetochore component KNL1, a protein conserved across many plant lineages[35]. In *D. capensis*, KNL1 antibodies revealed distinct nuclear foci at interphase (**Fig. 2c**), consistent with localized centromeric activity. However, no specific KNL1 signal could be detected in *D. regia* interphase nuclei **(Fig. 2d)**. Metaphase chromosomes of *D. capensis* stained with KNL1 displayed signals confined to discrete chromosomal regions, colocalizing with α-tubulin, further supporting monocentricity (**Fig. 2e-i**). Meanwhile, in *D. regia*, although KNL1 signals were absent, α-tubulin was observed to associate along the entire length of the mitotic chromosomes (**Fig. 2j-m**). These results challenge earlier assumptions and provide strong evidence for monocentricity in *D. capensis* and holocentricity in *D. regia*.

To further investigate centromere architecture, we analyzed the chromosomal distribution of satDNA, which, while not universally required for centromere function, is a prominent feature of most eukaryotic centromeres[36,37]. Each species exhibited a dominant satDNA family: *Cape71* in *D. capensis*, consisting of a 71 base pair monomer tandem repeat, and *Regi90* in *D. regia*, a 90 base pair tandem repeat. These sequences showed no significant similarity to each other (**Extended Data Fig. 2b**).

Fluorescence in situ hybridization (FISH) using a probe for *Cape71* revealed concentrated centromeric signals in *D. capensis* during interphase and metaphase. Pachytene chromosomes showed a single, clearly defined cluster of *Cape71* repeats per chromosome (**Fig. 2n-s**), confirming a monocentric organization. By contrast, *Regi90* in *D. regia* appeared as numerous dispersed foci in interphase nuclei (**Figure 2t**). During metaphase, *Regi90* signals formed continuous, line-like structures along chromosomes (**Fig. 2u-w**), a pattern that resembles holocentromeric satellite organization previously observed in other plant lineages[38,39]. During meiosis, *Regi90* repeats appeared as multiple foci along stretched pachytene chromosomes (**Fig. 2x-y**), consistent with a diffuse centromere model. Genomic analyses further revealed that *Regi90* was distributed in short arrays of approximately 10 to 20 kilobases across all *D. regia* chromosomes (**Extended Data Fig. 2c**). This pattern closely resembles the repeat-based centromere distribution reported in holocentric species like *Rhynchospora*[39], although we did not detect CENH3 or KNL1 localization in *D. regia*.

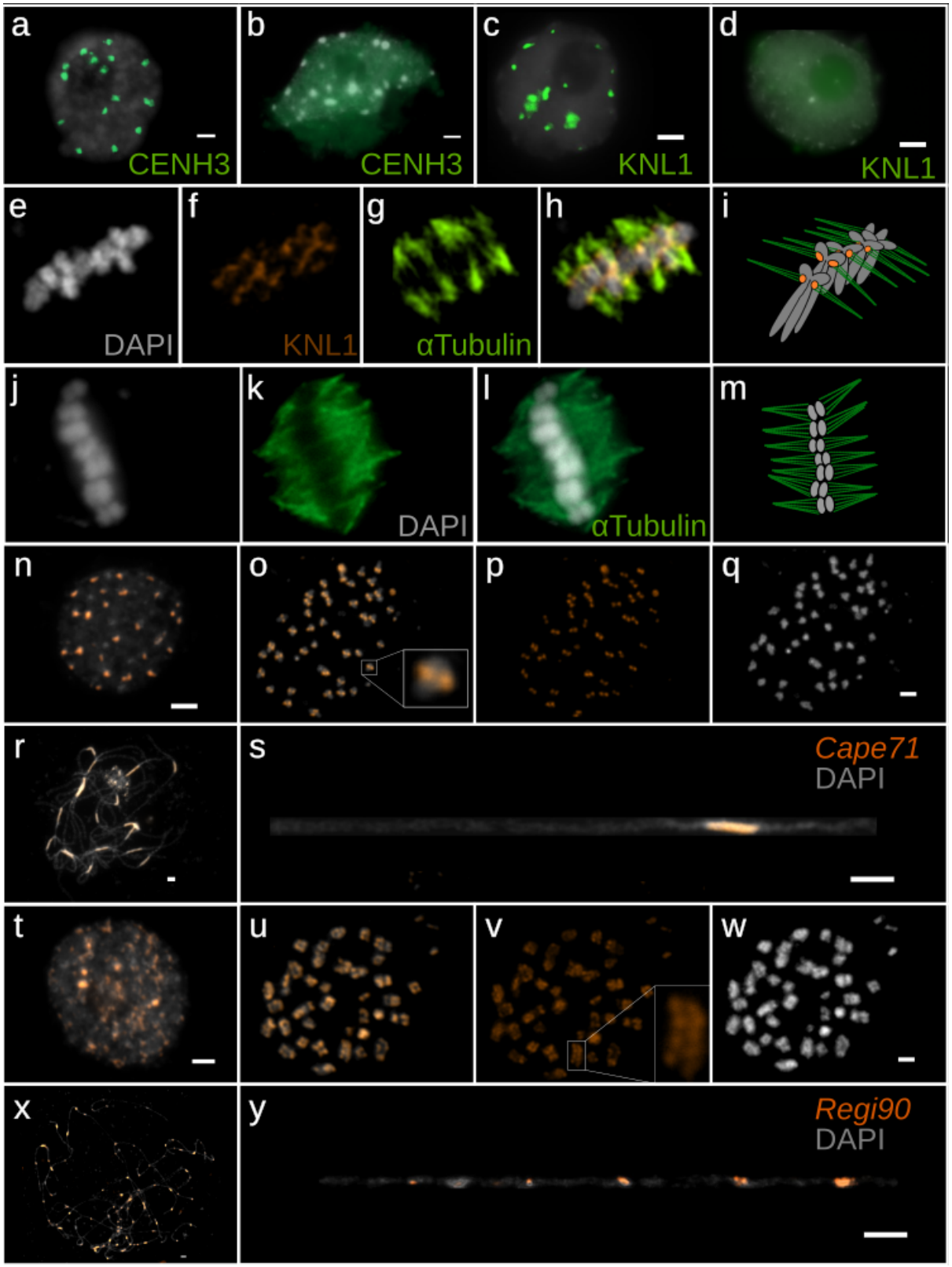


Fig. 2: Centromere morphologies in *D. capensis* and *D. regia* revealed by immunostaining and FISH. Immunostaining of interphase nuclei with CENH3 antibodies shows localized signals in *D. capensis* (**a**) and diffuse signals in *D. regia* (**b**). Similar patterns are observed with KNL1 antibodies in *D. capensis* (**c**) and *D. regia* (**d**). (**e-h**) Immunostaining of *D. capensis* metaphase chromosomes with KNL1 and α-tubulin reveals α-

tubulin attachment at single, localized regions. (**i**) Model summarizing *D. capensis* observations. (**j-l**) In *D. regia*, α-tubulin attaches along the entire metaphase chromosome length. (**m**) Model summarizing *D. regia* observations. (**n-s**) FISH with the *Cape71* probe in *D. capensis* shows clustered centromeric signals. (**t-y**) In *D. regia*, *Regi90* centromeric signals are distributed along the chromosomes. Panels illustrate contrasting patterns in interphase nuclei (**n, t**), metaphase chromosomes (**o-q**, **u-w**), and pachytene chromosomes (**r, x**). Panels **s** and **y** show isolated and stretched pachytene chromosomes from the same cells.

Differential satellite DNA evolution reflects centromere architecture in *Drosera*

To examine how centromeric satellite repeats evolved in *D. regia* and *D. capensis*, we constructed phylogenetic trees based on sequences from the *Regi90* and *Cape71* families. Repeat units were aligned separately for each species, showing substantial SNP variation and indel diversity. We interpret these patterns to reflect the time elapsed since element duplication or genomic repositioning events. Phylogenies grouped by chromosomal origin and location showed region-specific clades and mixed-repeat clusters (**Fig. 3**). This pattern suggests two evolutionary phases: initial local duplication/amplification, followed by dispersal of repeats to new genomic sites, establishing additional centromeric foci.

The *Cape71* phylogeny (**Fig. 3a**) includes multiple chromosome-specific clades with long branches, indicating that monomer subfamilies have arisen independently and undergone sustained amplification in a chromosome-restricted manner. In contrast, the *Regi90* tree (**Fig. 3b**) shows a starburst-like topology with short internal branches and widespread intermixing of terminal tips from different chromosomes. This pattern reflects a recent, genome-wide amplification of *Regi90*, followed by minimal local structuring. Few same-chromosome clades are present in *D. regia*, suggesting regionally distributed, nascent amplifications.

These differences imply that *D. capensis* experiences stronger and more sustained local concerted evolution. In this species, repeat variants that arise on a chromosome tend to remain and amplify there, maintaining their identity over time. *D. regia*, by contrast, exhibits faster turnover and inter-chromosomal mixing of repeats, limiting the persistence of array-specific subfamilies.

To more directly assess chromosome-specific structure, we labeled tree tips by chromosome of origin and applied midpoint-rooting to visualize relative branch lengths. In *D. capensis* (**Fig. 3c**), each point represents a *Cape71* monomer, plotted at the midpoint of its chromosome on the vertical axis and colored by relative evolutionary age (distance from root). The color spread within each vertical band indicates that centromeric regions on individual chromosomes house repeats of different ages. These distinct-age subfamilies reflect early-splitting lineages that were later amplified on specific chromosomes.

Zooming in on chromosome 19 (**Fig. 3d**), a narrow gradient in root-distance values is visible across the centromeric region. This gradient persists even when the analysis is restricted to monomers from a particular *Cape71* subtree. A nearly monochromatic cluster within this ~20 kb region marks an early-diverging subfamily that underwent sustained diversification, evident in variable terminal branch lengths.

In *D. regia*, the *Regi90* distribution plot (**Fig. 3e**) shows a continuous scatter of monomers along chromosomes, with color variation spanning the full spectrum of divergence. This supports genome-wide amplification followed by gene conversion, homogenizing repeats across chromosomes. On chromosome 13 (**Fig. 3f**), monomers span the full color gradient and are evenly distributed. However, when a subtree specific to this chromosome is isolated from

Fig. 3b, the monomers cluster as a light-green group, indicating a localized burst of *Regi90* amplification, perhaps involving a higher-order repeat structure.

Comparing both trees, *Regi90* monomers show higher absolute divergence (longer root-to-tip paths; **Fig. 3g**) but weaker chromosome-specific clustering and more extensive inter-chromosomal mixing. *Cape71* repeats, by contrast, exhibit lower overall divergence but stronger subfamily structure, with more homogeneous clades confined to specific chromosomal arrays.

We further analyzed satDNA dynamics by assessing sequence similarity within each major repeat family. Results showed generally low homogenization across arrays (**Extended Data Fig. 3**). Notably, *Cape71* repeats in the monocentric *D. capensis* genome exhibited higher homogenization within and between chromosomes (**Extended Data Fig. 3a**). In contrast, *Regi90* arrays in *D. regia* were less similar, with approximately 75 percent identity observed in both intra- and inter-chromosomal comparisons (**Extended Data Fig. 3b**).

One confounding factor in *D. capensis* is the large amount of satDNA located on small, unplaced scaffolds. Of the 17 megabases of *Cape71* identified by TideCluster, only 1.12 megabases were assigned to chromosomes, possibly reflecting assembly challenges due to high sequence similarity. In *D. regia*, 3.8 megabases of *Regi90* were annotated, with 3.5 megabases placed into pseudochromosomes.

Collectively, these data reveal distinct evolutionary trajectories for centromeric satDNAs in sundews. In monocentric *D. capensis*, satellite arrays show stronger local identity and homogenization. In holocentric *D. regia*, repeat families exhibit faster turnover, lower sequence similarity, and broader genomic dispersion. These patterns support a link between centromere architecture and satDNA evolutionary dynamics.

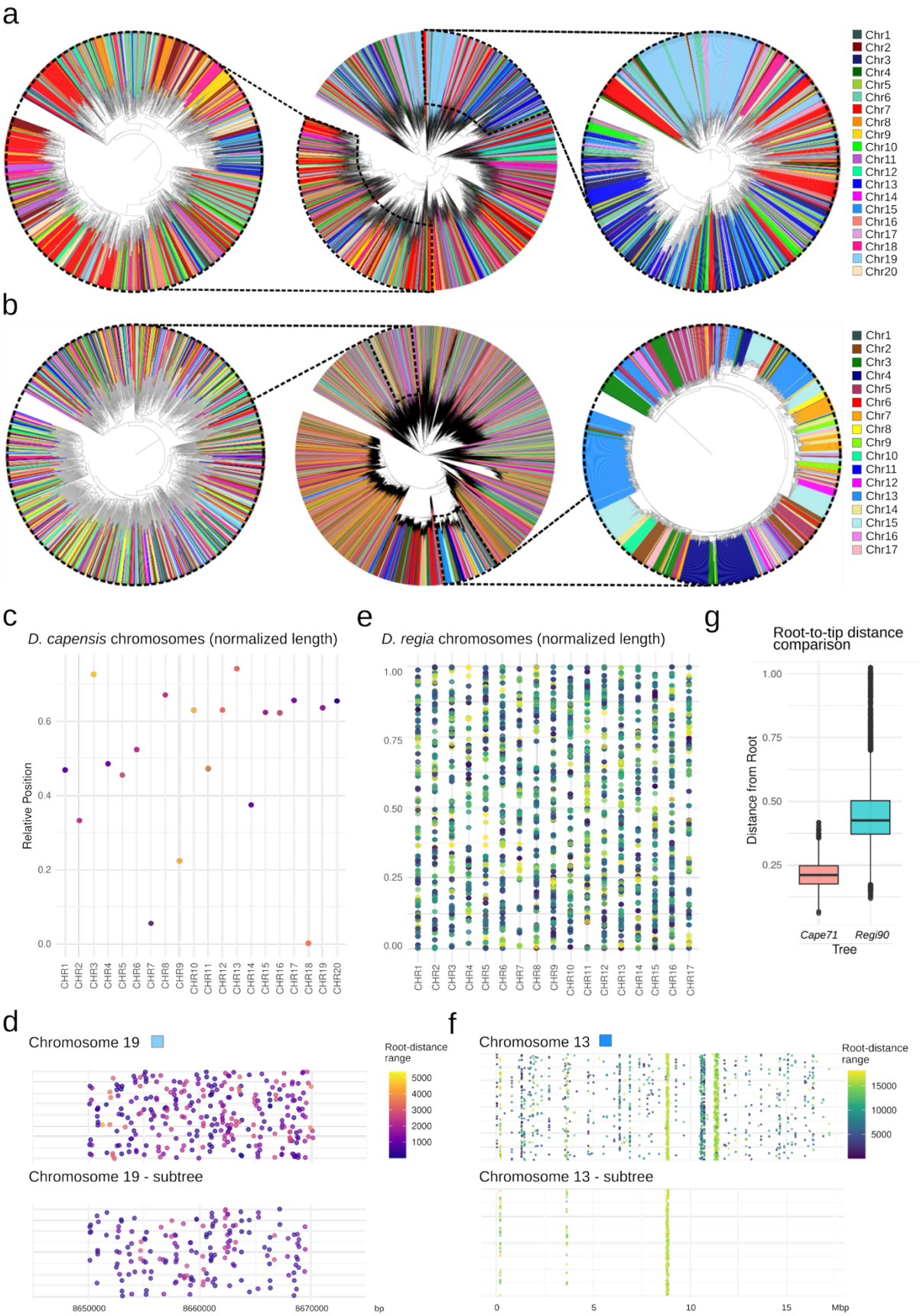

a
Chr1
Chr2
Chr3
Chr4
Chr5
Chr6
Chr7
Chr8
Chr9
Chr10
Chr11
Chr12
Chr13
Chr14
Chr15
Chr16
Chr17
Chr18
Chr19
Chr20
b
Chr1
Chr2
Chr3
Chr4
Chr5
Chr6
Chr7
Chr8
Chr9
Chr10
Chr11
Chr12
Chr13
Chr14
Chr15
Chr16
Chr17
c
D. capensis chromosomes (normalized length)
Relative Position
0.0
0.2
0.4
0.6
CHR1
CHR2
CHR3
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 polymer phylogenies in *D. capensis* (a) and *D. regia* (b) reveal both chromosome-specific and mixed clades. Large chromosome-specific clades are shaded blue. (**c**) Chromosomal distribution of *Cape71* monomers in *D. capensis*; each dot marks a repeat at its normalized position on given chromosomes (which are scaled to equal lengths for comparison), colored by root-to-tip branch length (i.e., relative evolutionary divergence). (**d**) Zoom-in on the chromosome 19 clade from (**c**) with all mapped branches (upper) and extracted subtree (lower). (**e**) Chromosomal distribution of *Regi90* monomers in *D. regia*, plotted as in (**c**). (**f**) Zoom-in on chromosome 13 from (**e**), with full set (upper) and subtree (lower). (g) Root-to-tip distances for both trees.

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[40], 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[41,42]. In both cases, our results agree 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[43]. 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. erythroriza* 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*[33] 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 with five subgenomes of x=8[13], which was suspected to complicate inferences of relationships within *Drosera*. To address this, a backbone phylogeny was constructed to model reticulate evolution among Caryophyllales carnivores using conserved Benchmarking Universal Single-Copy Ortholog (BUSCO)[44] gene sets to generate phylogenetic trees for *Drosera*, its close relatives, and selected outgroup angiosperms. Two datasets were analyzed: nucleotide and inferred amino acid sequences for 22 and 50 species, respectively (**Extended Data Fig. 5**). Distinct taxon samplings allowed assessment of topological consistency under varied outgroup effects. Phylogenetic reconstructions were performed using two methods: maximum likelihood (ML) analyses[45] of concatenated BUSCO alignments, and ASTRAL species tree reconstruction[46] summarizing individual BUSCO gene trees. In all analyses, *D. regia* consistently appeared as sister to *Aldrovanda* plus *Dionaea*, rather than clustering with other *Drosera* species. *Nepenthes* was generally resolved as sister to *Drosera* plus the snap-trapping carnivores, while the *Drosophyllum* plus *Triphyophyllum*/*Ancistrocladus* group formed the sister lineage to that clade (**Extended Data Fig. 5**). These relationships for *N. gracilis* differed from earlier published results, which also used nuclear-encoded BUSCO genes but placed *Nepenthes* as sister to the *Drosophyllum* plus *Triphyophyllum*/*Ancistrocladus* clade[13]. However, when constrained to a bifurcating topology, such relationships fail to adequately represent phylogenetic patterns among allopolyploids, which arise through hybrid origins[47].

Single copy genes display substantial gene tree conflict. To explore reticulate interrelationships (as anticipated from earlier results on gene-tree conflict within the group[48]), single-copy gene data were analyzed for eight Caryophyllales carnivore species and their non-carnivorous relative *Ancistrocladus*. With complete taxonomic representation, a cloudogram visualization[49] was used to highlight gene-tree conflict (**Fig. 4**), which can result from ancient admixture or incomplete lineage sorting. The analysis revealed that *D. regia* shows conflicting gene-tree relationships, with some loci grouping it with *Aldrovanda* and *Dionaea*, while others aligned it with the remaining *Drosera* species. Furthermore, substantial conflict within the *Aldrovanda* and *Dionaea* grouping points to complex evolutionary histories for these snap-trapping genera.

Fig. 4: DensiTree cloudogram of 563 single-copy ortholog phylogenies. Consensus topologies are overlaid, with line intensity proportional to frequency across the tree set[49]. *Dionaea*, *Aldrovanda*, and *D. regia* show highly reticulate patterns, consistent with allopolyploid origins.

Genome structure suggests complex polyploidy. To delve further into these relationships, gene-based synteny analyses were performed using GENESPACE[50]. Genomes of *D. capensis* and *D. regia* were compared to those of closely related outgroups, including *N. gracilis*, *Dionaea muscipula*[51], *Ancistrocladus abbreviatus*, and *Triphyophyllum peltatum*. The *N. gracilis* genome, which contains five x=8 subgenomes[13], exhibited strongly conserved synteny across the Caryophyllales species evaluated. Accordingly, x=8 is supported as the ancestral karyotype for the lineage[28]. *Ancistrocladus* and *Triphyophyllum* shared a chromosomal fission followed by a tetraploidy event, accounting for their higher basic chromosome number (x=9, **Fig. 5a**). *Dionaea* was resolved as an allotetraploid, with two distinct subgenomes (**Fig. 5a,b, Extended Data Fig. 6**), indicating a hybrid origin. Among sundews, most *D. regia* chromosomes matched *N. gracilis* chromosomes in a 3:1 pattern (**Fig. 5a**), consistent with *D. regia's* proposed hexaploid structure. This could be further resolved into 2:1 phasing using Ks values (**Fig. 5c**). In contrast, *D. capensis* showed substantial chromosomal rearrangement and reduced syntenic collinearity. Nonetheless, a 6:1 match with the eight-chromosome dominant subgenome of *N. gracilis* supports a dodecaploid origin for *D. capensis*, involving tetraploidization of a hexaploid ancestor (**Fig. 5a**).

Using a concatenated assembly of *D. regia* and *N. gracilis* and SynMap in CoGe[52], paralogous and orthologous syntenic relationships were clarified. Ks-based coloring of syntenic blocks in dot plots corresponded with whole genome multiplications or speciation events. Within *D. regia*, subgenomes displayed a 2:1 phasing pattern (**Fig. 6a,b**): two recently diverged

chromosomes per group (low Ks) and one older (high Ks), indicating that a tetraploid genome was joined by a third subgenome to form an allohexaploid.

In *N. gracilis*, its 4:1 subgenome structure was visible as four blocks plus the diagonal in the SynMap dot plot, and the 5:3 syntenic ratio against *D. regia* further emphasized this structure (**Fig. 6a**). Ks histograms showed the duplication events in *D. regia* occurred more recently than those in *N. gracilis*, with the divergence between the species being as old as some of the *Nepenthes*-specific subgenomes (**Fig. 6b**). Fractionation bias calculations[52,53] revealed that all three *D. regia* subgenomes shared similar gene retention patterns with the dominant *N. gracilis* subgenome (**Fig. 6c-e**), suggesting close common ancestry.

To determine whether *D. capensis* shares its hexaploid ancestry with *D. regia*, additional synteny analyses were conducted using *ksrates*[54] and GENESPACE. Despite post-polyploid rearrangements, some chromosomes in both *Drosera* species displayed plausibly shared fusion events (denoted by pairs of different colors detailed in **Extended Data Fig. 8**), implying common ancestral restructuring prior to their divergence. *Ksrates* analysis confirmed that the allotetraploidy in *Dionaea* and hexaploidy in *D. regia* and *D. capensis* occurred before their respective species splits (**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. However, although shared rearrangements support a shared hexaploidy origin, this cannot be concluded definitively. While shared rearrangements are a useful, parsimonious proxy for inferring a common polyploid origin[55,56], the probability of convergent rearrangements, while low, cannot be ruled out. Interpretation of a shared hexaploidy is further nuanced by phylogenetic evidence (e.g., our BUSCO topologies; **Extended Data Fig. 5**) that shows *D. regia* as sister to *Dionaea*, rather than grouping with the *Drosera* clade containing *D. capensis*. Additional genome analyses within *Drosera* will be necessary to resolve the precise timing and origin of one or more hexaploidy events. This is crucial for determining whether chromosomal fusions in *D. regia* occurred before or after holocentric chromosome evolution, and whether *D. capensis* could represent a reversion to monocentricity.

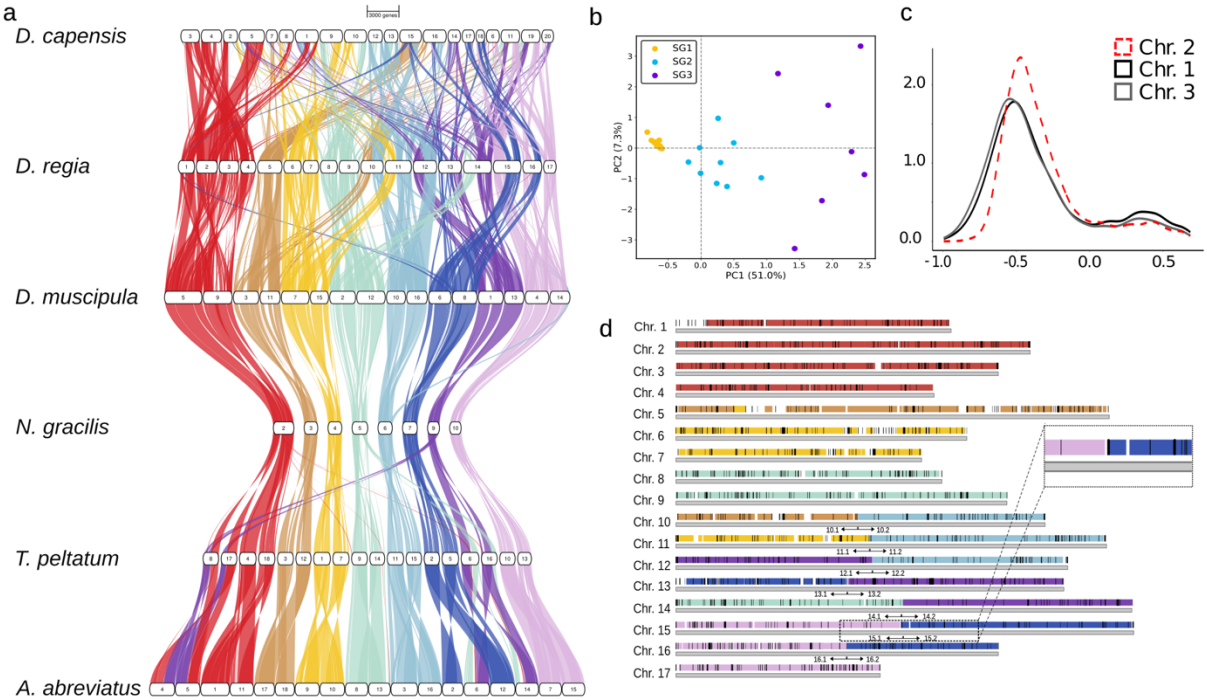


Fig. 5: Structural rearrangements in mono- and holocentric *Drosera* genomes. (**a**) Synteny of the dominant *N. gracilis* subgenome (8 chromosomes) with *Ancistrocladus abbreviatus*, *Triphyophyllum peltatum*, *Dionaea muscipula*, hexaploid *D. regia*, and dodecaploid *D. capensis*. *Dionaea* is doubled but minimally rearranged; *D. regia* and *D. capensis* share triplicate structure and several fusions, while *D. capensis* shows further duplication and complex rearrangements. (**b**) SubPhaser PCA of *Dionaea* and D. regia assemblies indicates *Dionaea* is an allopolyploid of subgenomes 2 (cyan) and 3 (purple), with *D. regia* (yellow) diverging in k-mer composition (see **Extended Data Fig. 6**). (**c**) Ks density plot phases *D. regia*'s triplicate subgenomes into a 2:1 pattern: chromosomes 1 and 3 match; chromosome 2 represents an older subgenome at higher Ks (x = $\log_{10}$ Ks; y = density; see **Extended Data Fig. 7**). (**d**) *Regi90* repeats (black bars) localize at *D. regia* fusion points. Colored blocks indicate synteny with ancestral *N. gracilis* chromosomes (panel **a**). Note black bar at the chromosome 15 breakpoint.

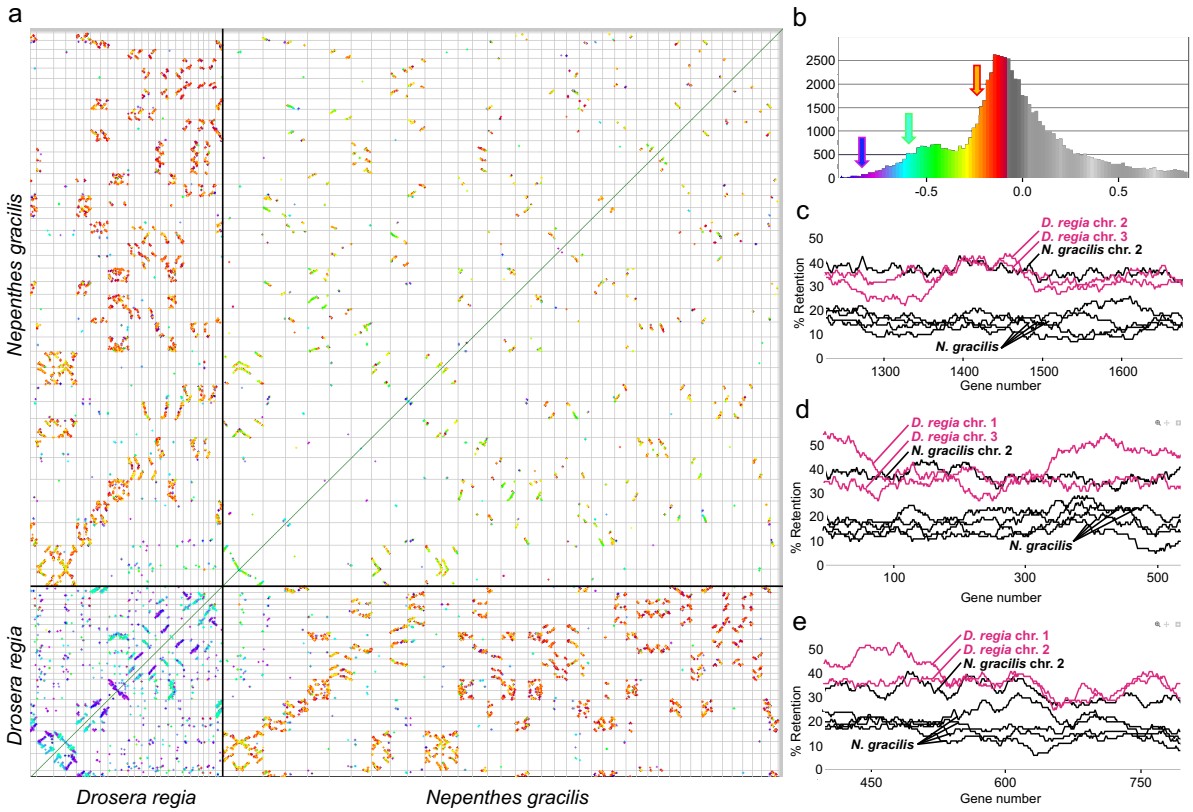


Fig. 6. Ancestral subgenome sharing between *Nepenthes* and *D. regia*. (**a**) Self:self synteny of concatenated *D. regia* and *N. gracilis* assemblies, colored by $\log_{10}$ Ks values as in (**b**). (**b**) Ks histogram, with x-axis showing number of gene pairs; arrows indicate three relative Ks time ranges (purple = youngest). (**c-e**) FractBias profiles (with chromosomal queries and targets in all three possible combinations) reveal close similarity between all three *D. regia* subgenomes and the *Nepenthes* dominant subgenome. Four recessive *Nepenthes* chromosomes, from two tetraploidy events, show much lower gene retention profiles (gray).

Drawing on our backbone phylogenetic analyses, syntenic relationships, subgenome fractionation patterns, Ks distributions, and chromosome counts, we propose a tentative scenario (**Fig. 7**) illustrating the reticulate allopolyploid relationships among the Caryophyllales carnivorous plant genomes analyzed. Notably, *D. regia* is a highly complex allopolyploid with likely subgenome contributions from the *Dionaea* lineage. We hypothesize that the modern *D. regia* genome evolved in a two-stage process: the first step forming its allohexaploid structure with one input from the *Dionaea* lineage, and a second involving the *Dionaea* clade again via introgression and consequent chromosome elimination (**Fig. 7**). This scenario may explain *D. regia*'s consistent sister-group relationship to snap-trappers, rather than to other *Drosera* species, in our phylogenetic reconstructions (**Extended Data Fig. 5**).

Considering these findings, *D. regia* may warrant recognition as a separate, monotypic genus outside of *Drosera*. The name *Freatulina regia* is already available for this purpose[57]. Morphologically, the species is distinct from most sundews, bearing woody rhizomes, operculate pollen released in tetrads, undivided floral styles, and a non-circinate flowering scape: traits that are either considered ancestral or absent from other members of the genus[57-59]. The species also exhibits unique wrapping leaf movements, capable of repeatedly coiling around prey[59].

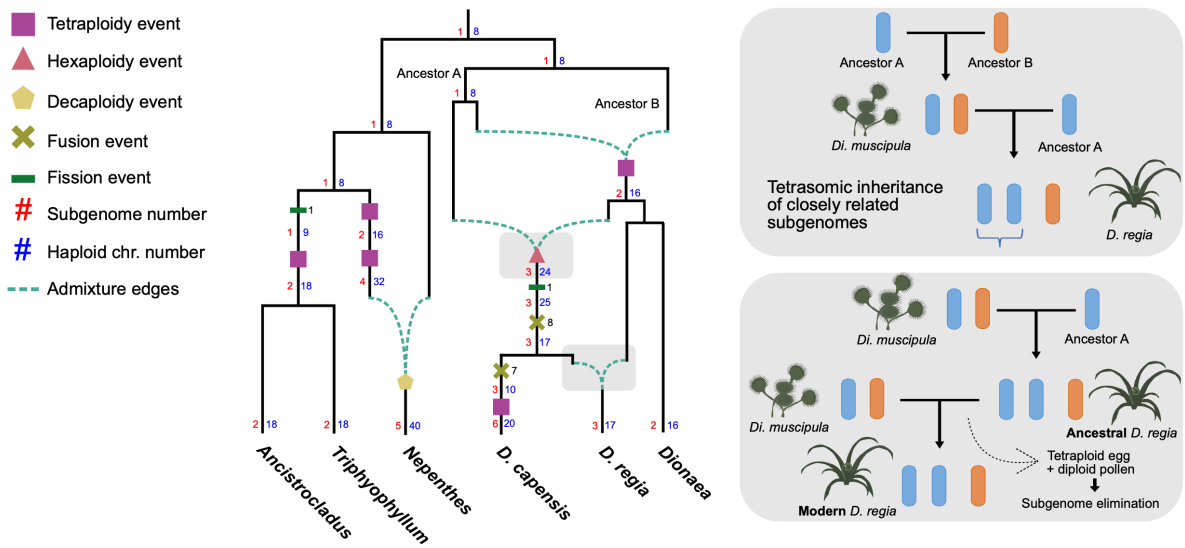


Fig. 7. Phylogenetic network hypothesis for interrelationships among the six genomes under study. An *x*=8 ancestor is proposed. All other Caryophyllales carnivores and the revertant *Ancistrocladus* share *Nepenthes*' *x*=8 ancestral subgenome, or a modification of it. *Dionaea* is an *x*=16 allotetraploid 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 an allohexaploid hybrid lineage. 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 gray panels detail hypothetical events that gave rise to the modern *D. regia* genome. In the upper panel, two successive hybridizations produce an ancestral *D. regia* that possesses three distinct subgenomes: two closely related (blue rectangles), enabling tetrasomic inheritance, and one divergent (orange rectangle). In the lower panel, introgression involving the ancestral *D. regia* lineage and a close relative of *Dionaea* gives rise to modern *D. regia* via subgenome elimination. 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.

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 an integrative 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 demonstrate the occurrence of monocentric versus holocentric chromosomes in *D. capensis* and *D. regia*, respectively, revealing a highly unusual architectural contrast within a single genus of angiosperms[60-62]. 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 Venus flytrap, *Dionaea*[35], 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[63].

Structurally, *D. capensis* and *D. regia* exhibit pronounced 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 support recent models connecting transposon dynamics to centromere restructuring and genomic plasticity.

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 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 in *D. capensis* suggests that centromere type may influence genome stability, with holocentricity potentially buffering against large-scale rearrangements.

It is notable that such divergent chromosomal architectures evolved within *Drosera*, a genus embedded within an equivalently morphologically and physiologically complex carnivorous plant clade. The monophyletic Caryophyllales carnivore 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*[2]. In this context, our findings suggest that variation in centromere structure and polyploid mechanisms may not only correlate with ecological diversification but could also actively promote it, highlighting a potential interplay between chromosomal dynamics, genome architecture, and adaptive evolution.

Together, our results demonstrate that centromere organization, polyploid history, chromosomal fusions, and repeat dynamics collectively shape genome evolution in *Drosera*. The presence of both holocentric and monocentric species interpreted to derive from the same hexaploid ancestor - while diverging markedly in genome structural outcomes - points to centromere type as a possible determinant of post-WGM evolutionary 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[64]. 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[65]. 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[66]. A hybrid assembly including the filtered reads and PacBio RS II reads was then carried out using MaSuRCA v3.2.1[67]. The contig-level assemblies were evaluated for completeness with BUSCO v3.0.2[44] using the embryophyta odb9 database, and for contiguity, using QUAST v4.3[68].

Dovetail Hi-C reads were first mapped to the contig file obtained from the MaSuRCA assembler using BWA[69] following the hic-pipeline (https://github.com/esrice/hic-pipeline). Hi-C scaffolding was performed using 3D-DNA pipeline[70,71] 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[70] 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 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[72]. 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[73] 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[74] and then assembling the transcriptome using StringTie v.1.3.4c[75]. We then passed the three transcriptome assemblies to EvidentialGene v2017.12.21[76] 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[77] and then masked the genome using RepeatMasker v4.0.7[78]. For gene model prediction, the transcriptome assembly was first splice-aligned against the unmasked reference *D. capensis* genome using PASA v2.2.0[79] to generate ORFs. Secondly, ab initio gene model prediction was carried out using BRAKER v.2.0.3[80], which internally used the reference aligned RNA-Seq data and Genemark-ET to train AUGUSTUS v3.3[80,81] for its final prediction. Additionally, the self-training Genemark-ES v.4.33[82] 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[79] 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[79] to generate ORFs. Additionally, *Arabidopsis thaliana* gene models were aligned against the genome using exonerate v2.2.0[83]. Genemark-ES v4.33[82] was then used to generate one set of predictions. BRAKER v2.0.3[80] 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[80,81] 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[79] 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)[50]. 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)[52]. 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[53] 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 regions involved in fusions, we followed Hofstatter et al, 2022[24]. The syntenic alignment obtained in GENESPACE between the *D. regia* genome and the *N. gracilis* dominant subgenome allowed us to pinpoint regions around the borders of proposed homologous fusion events. To evaluate homology, we loaded and compared annotation features for genes, TEs, and tandem repeats along the syntenic alignments using Geneious (https://www.geneious.com).

Subgenome-aware phasing of *D. regia* and *Dionaea muscipula* genomes. We used SubPhaser[84] (default parameters) to phase and partition the subgenomes of *D. regia* and *Dionaea 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[85], ggplot2[86], RColorBrewer[87], ggridges[88], and ggpmisc[89] packages.

***ksrates* analysis.** *ksrates* version 1.1.4[90] was used to position species splits relative to polyploidy events. Coding sequence (CDS) fasta files were extracted using AGAT version 1.4.0[91]. 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*[92], *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[93] 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[91]. The species tree inference was performed with STAG[94], 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[95].

Repeat characterization. DANTE and DANTE-LTR retrotransposon identification (Galaxy Version 3.5.1.1) pipelines[96] 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[97]. 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)[98]. All putative tandem sequences were compared for homology using DOTTER. These tandem sequences were individually mapped to the genome by BLAST[99], 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[100] 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/)[98]. 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[35] (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[101], 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[102]. Phylogenetic trees were inferred using FastTree v2.1.11[103] under the generalized time-reversible (GTR) model. Resulting trees were visualized and annotated using the Interactive Tree of Life (iTOL) web server[104], with colors corresponding to their chromosome of origin.

The phylogenetic trees were imported into R using the ape package[105]. Tip labels included chromosome ID and genomic coordinates. Genomic positions of sampled tips were plotted using ggplot2[86] 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[105], dplyr[106], ggplot2[86], 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[107] (**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[46]. Concatenated nucleotide and protein alignments, generated by catfasta2phyml v2018-09-28 (https://github.com/nylander/catfasta2phyml), served as additional input to IQ-TREE[45] with the same substitution models. *Amborella trichopoda* was specified as the outgroup.

Data and code availability

The *Triphyophyllum peltatum*, *Ancistrocladus abbreviatus*, and *Dionaea muscipula* genomes are available at NCBI via accessions BAAHMP010000001-BAAHMP010000429, BAAHMO010000001-BAAHMO010000131 and BAAGIJ010000001-BAAGIJ010003706, respectively. The processed *Drosera* 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.).

Author contributions

V.A.A. conceived the research. V.A.A. and A.M. led the study, with contributions from J.S. and K.F. M.M.-R. provided *Drosera* plant material, and E.I.-L. extracted DNAs. S.C.S. and D.D.-M. sequenced *Drosera* genomes, J.S. and A.M. assembled them and S.R. annotated them. K.F. led the sequencing, assembly and annotation of the *Dionaea*, *Triphyophyllum* and *Ancistrocladus* genomes with sequencing support from K.S. and S.M. S.J.F. and A.M. phased the *Drosera* and *Dionaea* genomes, with contributions from J.R.B. L.A.R. performed all cytological analyses and led research on repetitive DNAs, the latter with contributions from J.K. (phylogenetic analyses) and T.P.M. L.A.R., S.J.F., A.B., A.M. and V.A.A. performed syntenic comparisons, with contributions from D.S., Y.Z. and C.Z. S.J.F. performed *ksrates* analyses. H.M. performed gene family and phylogenetic analyses, with contributions from S.J.F. and V.A.A. V.A.A., J.K., A.M. and L.A.R. developed phylogenetic network and allopolyploidy reconstructions. Additional bioinformatic analyses were done by T.L., M.R., T.Q.W. and J.S. Additional materials and research support were provided by D.B., G.B., M.F., R.H., L.H.-E., I.K., S.T., Y.V.d.P. and G.V. V.A.A., L.A.R. and A.M. wrote the manuscript, and all authors approved the final version.

Competing Interests

The authors declare no competing interests.

Extended Data Table 1. Quality assessment of the genomes.

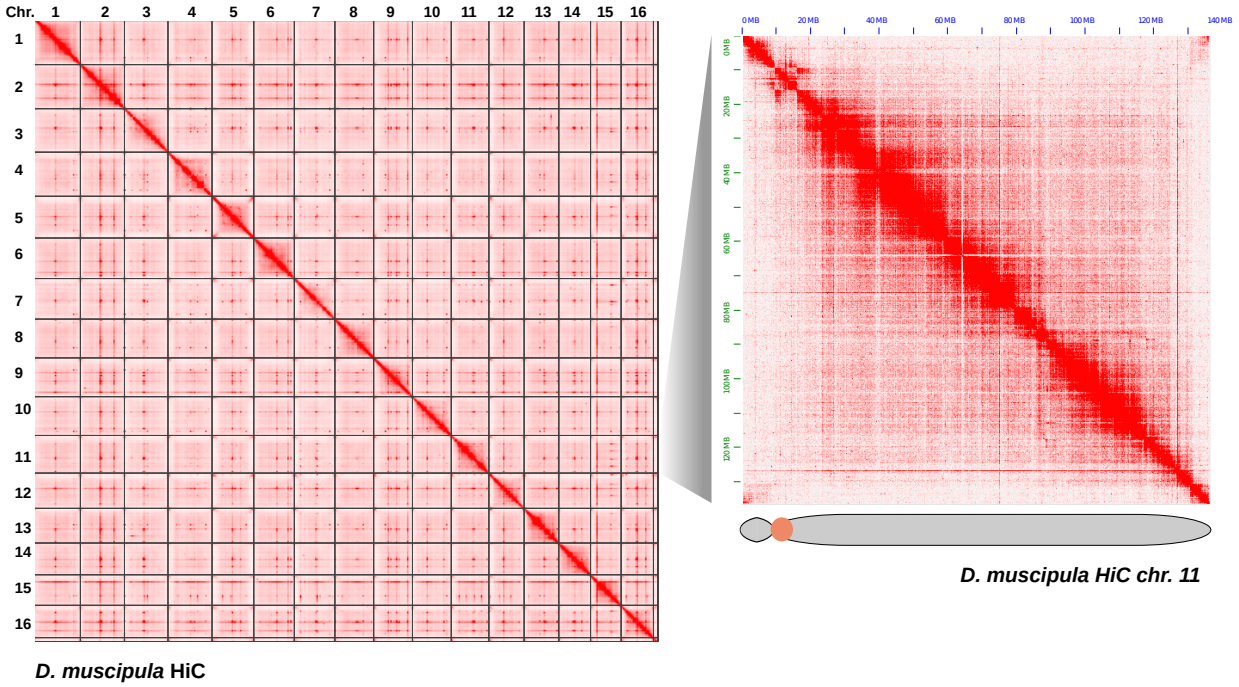
	BUSCO Library: eudicot	*Drosera_capensis*	*Drosera_regia*	*Triphyophyllum_peltatum*	*Ancistrocladus_abbreviatus*	*Dionaea_muscipula*
BUSCO	Complete BUSCOs (C)	2,086 (89.7%)	2,179 (93.7%)	2,255 (96.9%)	2,240 (96.3%)	2,128 (91.5%)
	Complete and single-copy BUSCOs (S	1,471 (63.2%)	1,859 (79.9%)	2,089 (89.8%)	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

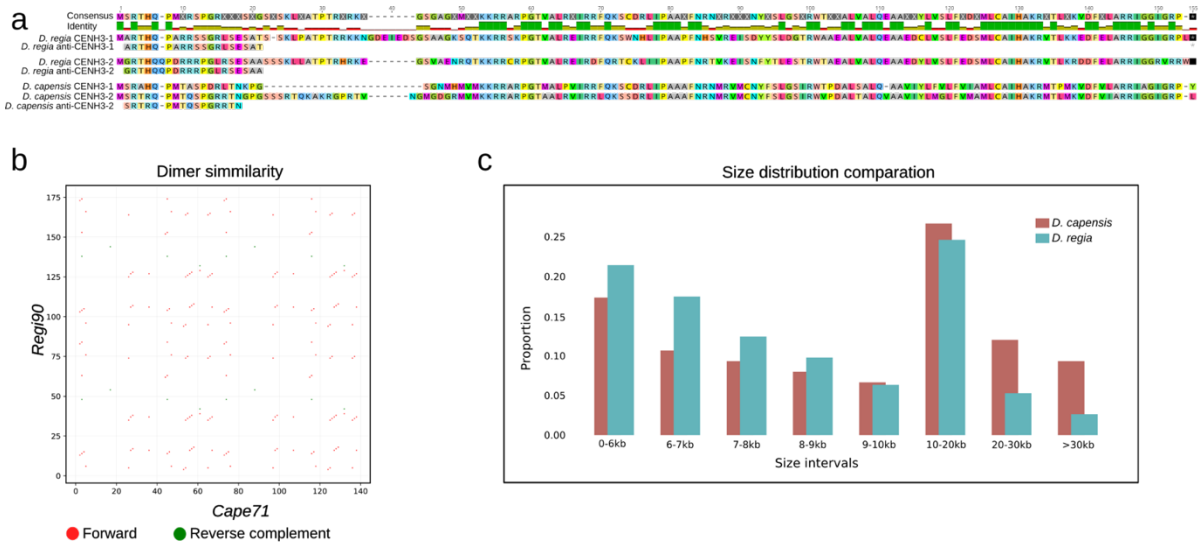
Extended data Table 2. Short reads analyzed in the repeatome comparative analysis.

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. erythrorhiza	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

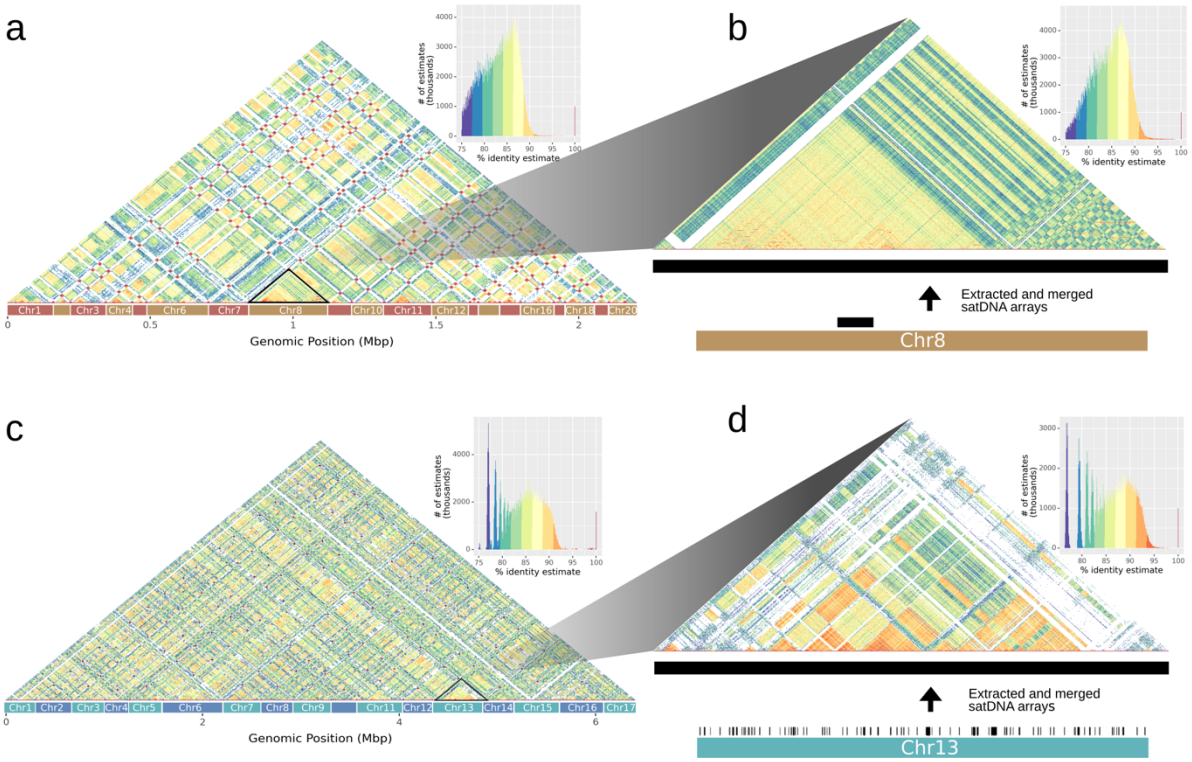
Extended Data Figures



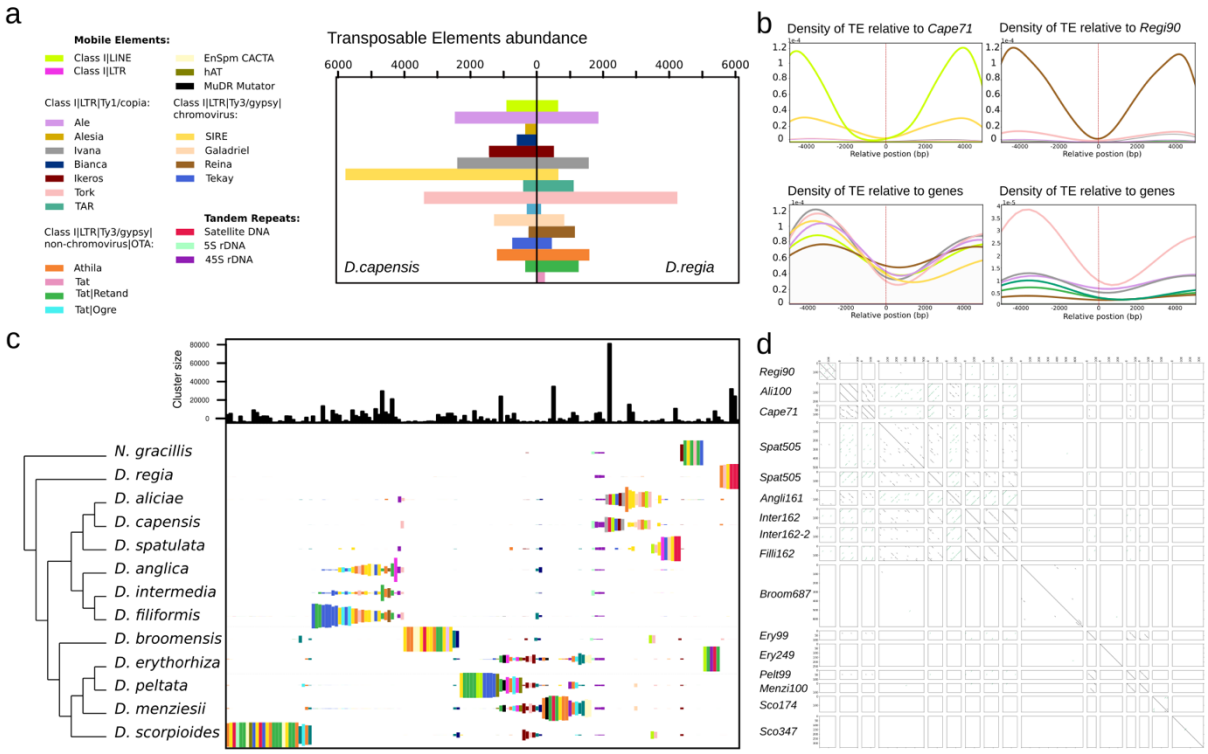
Extended data Fig. 1. *Dionaea muscipula* Hi-C contact map. Hi-C contact map showing every *Dionaea muscipula* chromosome, plus a zoom-in to chromosome 11.



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.

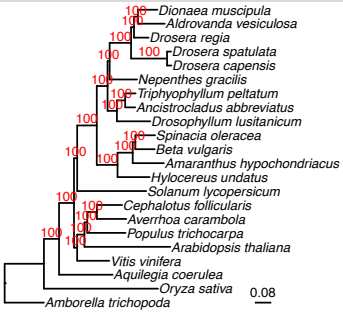


Extended Data Fig. 3: Comparison of centromeric sequence homogenization and the abundance of different transposable elements present in the *D. capensis* and *D. regia* genomes. **(a)** Genome-wide ModDot plot histograms of concatenated isolated centromeric arrays of all *D. capensis* chromosomes. **(b)** 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.

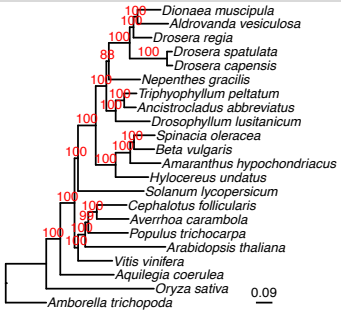


Extended Data Fig. 4: Comparison of the most abundant transposable-element (TE) families across *Drosera capensis* and *D. regia* genomes, and comparison of total repeats among 12 *Drosera* species plus *Nepenthes gracilis*. **(a)** Abundance differences of major TE families between *D. capensis* and *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.

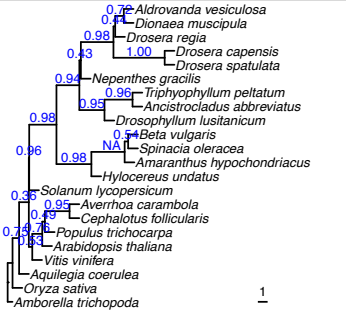
IQ–TREE v2.2.5 with concatenated DNA alignment
(3,548,928 sites, GTR+R model, loglik = –42,388,961.8)



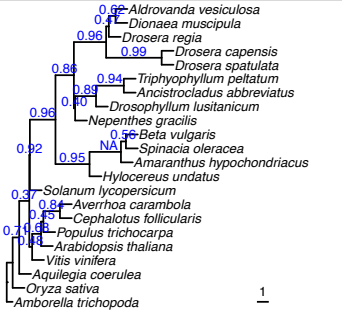
IQ–TREE v2.2.5 with concatenated protein alignment
(1,182,976 sites, LG+R model, loglik = –22,070,167.6)



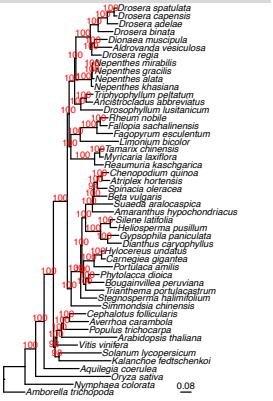
ASTRAL v5.7.8 with 2,325 DNA ML gene trees (GTR+R)



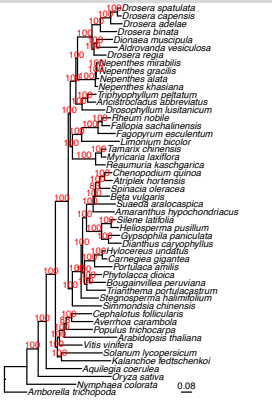
ASTRAL v5.7.8 with 2,325 protein ML gene trees (LG+R)



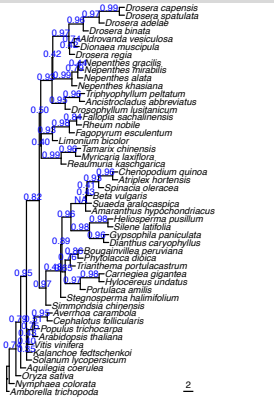
IQ–TREE v2.2.5 with concatenated DNA alignment
(3,349,866 sites, GTR+R model, loglik = –68,503,500.7)



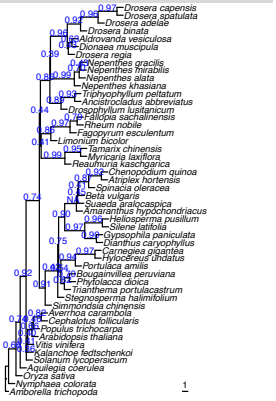
IQ–TREE v2.2.5 with concatenated protein alignment
(1,116,622 sites, LG+R model, loglik = –33,410,847.7)



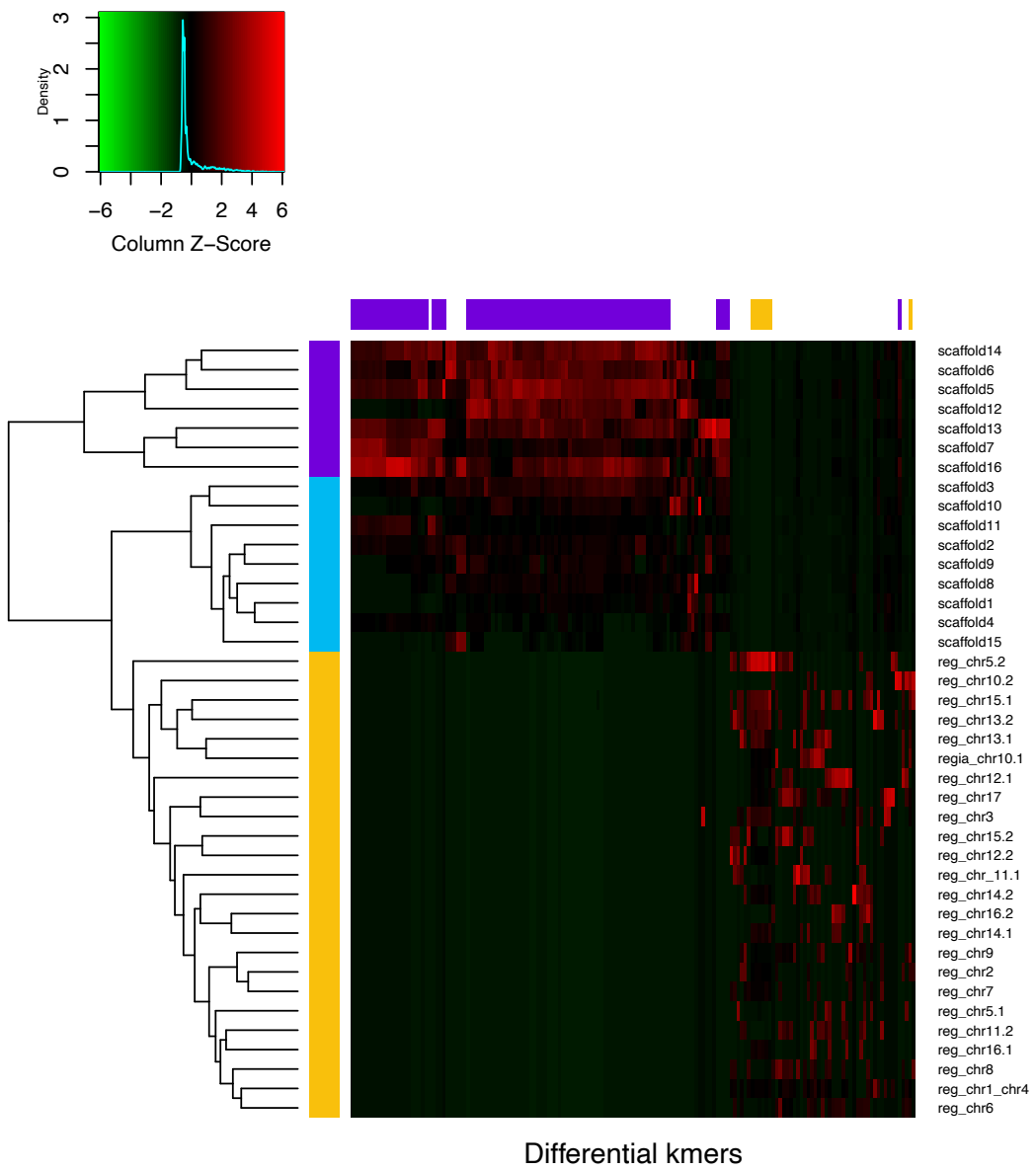
ASTRAL v5.7.8 with 2,325 DNA ML gene trees (GTR+R)



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

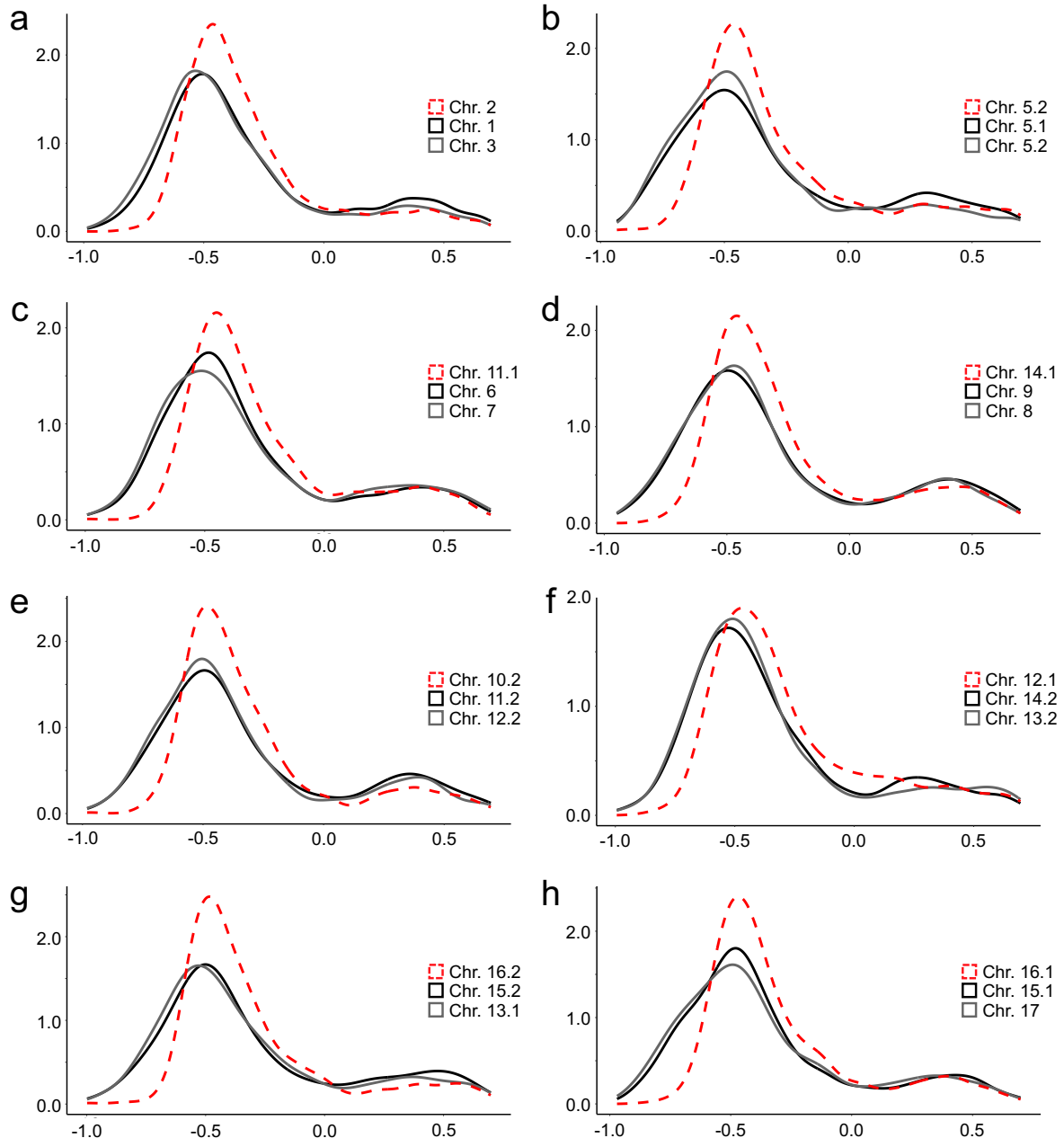


Extended Data Fig. 5. BUSCO gene and species trees. Two datasets were examined: nucleotide coding sequences or inferred amino acids from 22 species or 50 species. Phylogenetic relationships were reconstructed using two different approaches: (1) maximum likelihood (ML) analysis based on concatenated BUSCO 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.

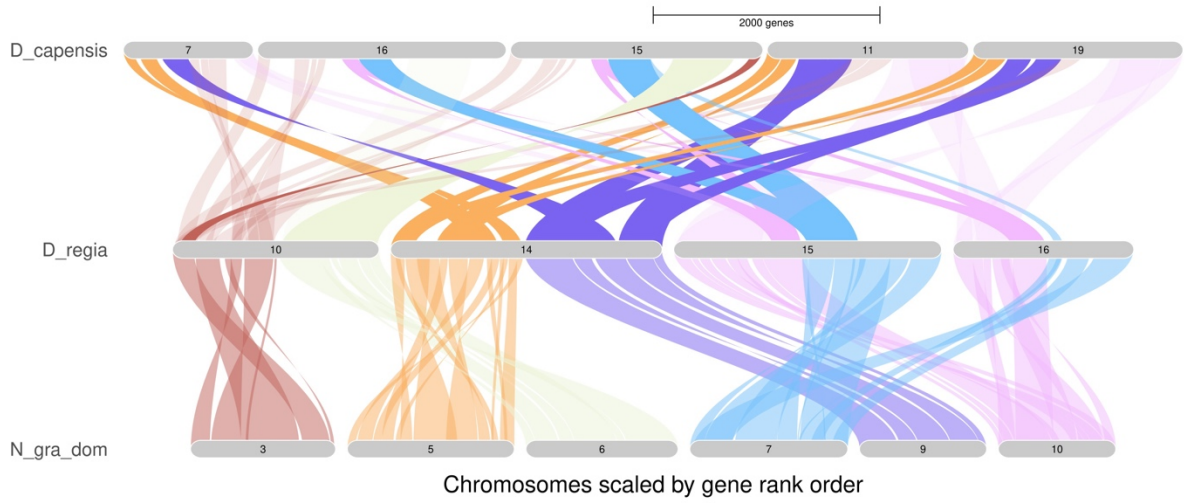


Extended Data Fig. 6: Unsupervised hierarchical clustering of ancestral chromosomes of *D. regia* and modern chromosomes of Dionaea. The horizontal color bar at the top (x-axis) indicates to which subgenome the *k*-mer is specific; the vertical color bar on the left (y-axis) indicates the subgenome to 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

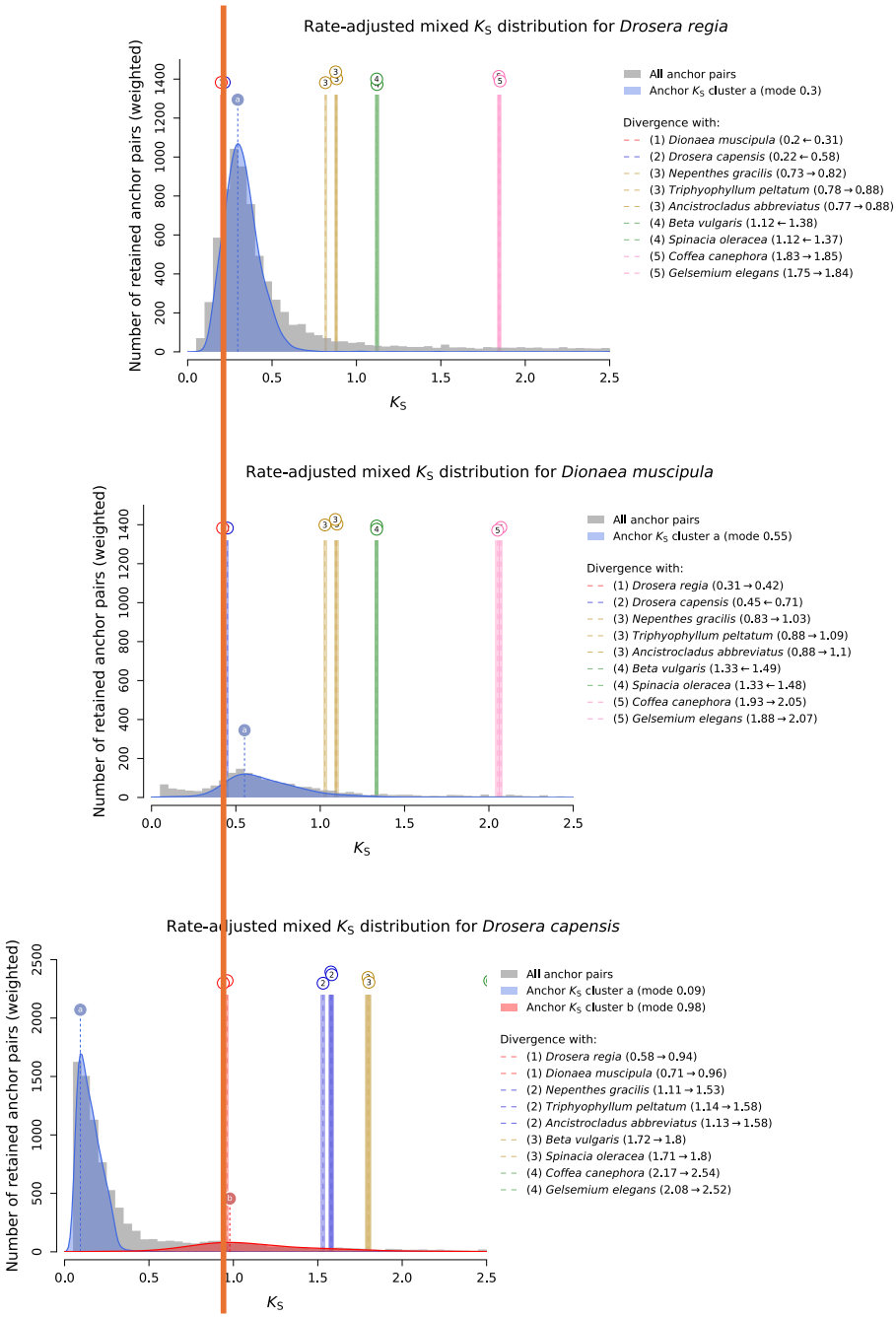
abundance of a *k*-mer. *Dionaea* subgenomes are represented in purple and blue colors, while *D. regia* (yellow) did not show any particular subgenome differentiation based on this approach.



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. 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.
36. Thakur, J., Packiaraj, J., and Henikoff, S. (2021). Sequence, chromatin and evolution of satellite DNA. International Journal of Molecular Sciences *22*, 4309.
37. Comai, L., Maheshwari, S., and Marimuthu, M.P. (2017). Plant centromeres. Current Opinion in Plant Biology *36*, 158-167.
38. 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.
39. 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.
40. 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.
41. 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.
42. Bousios, A., Kakutani, T., and Henderson, I.R. (2025). Centrophilic retrotransposons of plant genomes. Annual Review of Plant Biology *76*.
43. Wells, J.N., and Feschotte, C. (2020). A field guide to eukaryotic transposable elements. Annual review of genetics *54*, 539-561.
44. Manni, M., Berkeley, M.R., Seppey, M., and Zdobnov, E.M. (2021). BUSCO: assessing genomic data quality and beyond. Current Protocols *1*, e323.

45. 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.
46. 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.
47. 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.
48. Walker, J.F., Yang, Y., Moore, M.J., Mikenas, J., Timoneda, A., Brockington, S.F., and Smith, S.A. (2017). Widespread paleopolyploidy, gene tree conflict, and recalcitrant relationships among the carnivorous Caryophyllales. American Journal of Botany *104*, 858-867. https://doi.org/10.3732/ajb.1700083.
49. Bouckaert, R.R. (2010). DensiTree: making sense of sets of phylogenetic trees. Bioinformatics *26*, 1372-1373. 10.1093/bioinformatics/btq110.
50. 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.
51. Trebing, S., Freund, M., Iosip, A.L., Diblasi, C., Krennerich, V., Kirshner, J., Sato, M.P., Marques, A., Saitou, M., and Becker, D. (2025). Impaired trap closure in the counting-deficient Venus flytrap mutant DYSCALCULIA is caused by cell wall biomechanics. bioRxiv, 2025.2006. 2026.661685.
52. 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.
53. 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.
54. 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.
55. Sankoff, D., and Blanchette, M. (1998). Multiple genome rearrangement and breakpoint phylogeny. Journal of computational biology *5*, 555-570.
56. Sankoff, D., Zheng, C., and Zhu, Q. (2007). Polyploids, genome halving and phylogeny. Bioinformatics *23*, i433-i439. 10.1093/bioinformatics/btm169.
57. Chrtek, J., and Slavíková, Z. (1996). Comments on the families Drosophyllaceae and Droseraceae. Journal of the National Museum (Prague), Natural History Series *165*, 139-141.
58. Chrtek, J., and Slavíková, Z. (1999). Genera and families of the Droserales. Novitates Botanicae Universitatis Carolinae *13*, 39-46.
59. Poppinga, S., Masselter, T., Hartmeyer, I., Hartmeyer, S., and Speck, T. (2013). Trap diversity and evolution in the family Droseraceae. Plant Signaling & Behavior *8*, e24685. 10.4161/psb.24685.
60. Marques, A., and Drinnenberg, I.A. (2025). Same but different: Centromere regulations in holocentric insects and plants. Current Opinion in Cell Biology *93*, 102484.

61. 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.
62. 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.
63. 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.
64. 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.
65. 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.
66. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114-2120.
67. 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.
68. Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. Bioinformatics *29*, 1072-1075.
69. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. bioinformatics *25*, 1754-1760.
70. 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.
71. 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.
72. 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.
73. 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.
74. 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.
75. 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.
76. Gilbert, D. (2013). Gene-omes built from mRNA-seq not genome DNA.

77. 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.
78. Chen, N. (2004). Using Repeat Masker to identify repetitive elements in genomic sequences. Current protocols in bioinformatics *5*, 4.10. 11-14.10. 14.
79. 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.
80. 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.
81. 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.
82. 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.
83. Slater, G.S.C., and Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. BMC bioinformatics *6*, 31.
84. 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.
85. 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.
86. Hadley, W. (2016). Ggplot2: Elegrant graphics for data analysis (Springer).
87. Neuwirth, E., and Neuwirth, M.E. (2014). Package 'RColorBrewer'. ColorBrewer Palettes.
88. 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.
89. Aphalo, P. (2020). ggpmisc: Miscellaneous Extensions to "ggplot2"(R package version 0.3. 6).
90. Sensalari, C., Maere, S., and Lohaus, R. (2022). ksrates: positioning whole-genome duplications relative to speciation events in KS distributions. Bioinformatics *38*, 530-532.
91. 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.
92. 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.
93. Emms, D.M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome biology *20*, 1-14.
94. Emms, D.M., and Kelly, S. (2018). STAG: Species tree inference from all genes. Preprint at https://doi.org/10.1101/267914.
95. Emms, D.M., and Kelly, S. (2017). STRIDE: species tree root inference from gene duplication events. Molecular biology and evolution *34*, 3267-3278.

96. 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.
97. 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.
98. 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.
99. 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.
100. 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.
101. 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.
102. 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.
103. Price, M.N., Dehal, P.S., and Arkin, A.P. (2010). FastTree 2–approximately maximum-likelihood trees for large alignments. PloS one *5*, e9490.
104. 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.
105. 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.
106. Wickham, H. (2015). dplyr: A grammar of data manipulation. R package version 04. *3*, p156.
107. Montero, H., Freund, M., and Fukushima, K. (2025). Convergent losses of arbuscular mycorrhizal symbiosis in carnivorous plants. bioRxiv, 2025.2004. 2003.646726.