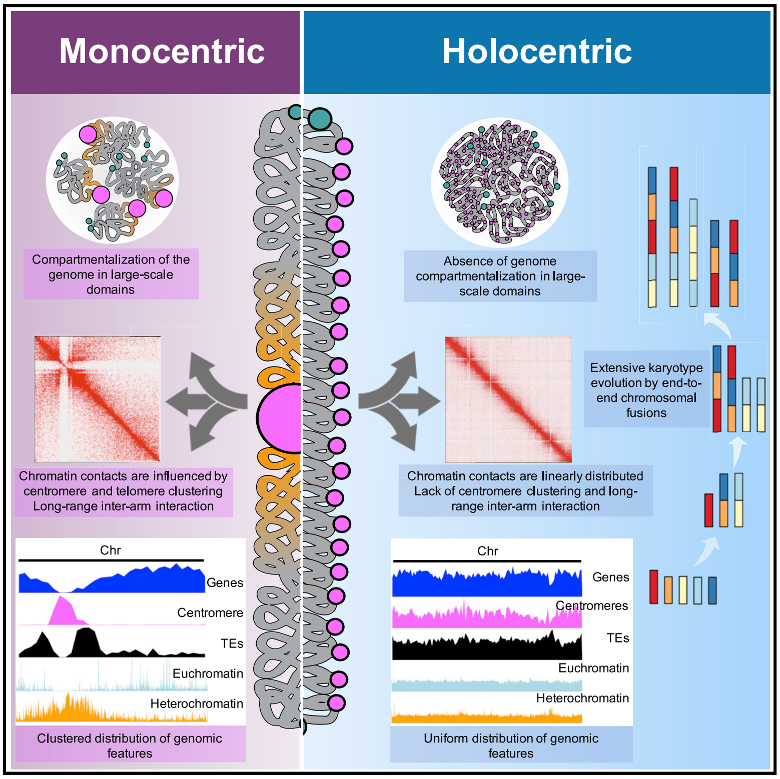
Cell

Article

Repeat-based holocentromeres influence genome architecture and karyotype evolution

Graphical abstract



Authors

Paulo G. Hofstatter,
Gokilavani Thangavel, Thomas Lux, ...,
Klaus F.X. Mayer, Andreas Houben,
André Marques

Correspondence

amarques@mpipz.mpg.de

In brief

While most eukaryotes contain single regional centromeres, several plant and animal lineages assemble holocentromeres along the entire chromosome length. The assembly of chromosome-scale holocentric genomes with repeat-based holocentromeres from beak-sedges and their closest monocentric relative sheds light on important aspects of genome architecture and evolution influenced by centromere organization.

Highlights

- Chromosome-scale genomes for holocentric plants with repeat-based holocentromeres
- Transition to holocentricity influenced the 3D (epi)genome architecture
- Regulation of repeat-based centromeres is conserved in mono- and holocentrics
- Chromosome fusions drive karyotype evolution and structural diploidization

Hofstatter et al., 2022, Cell *185*, 3153–3168
August 18, 2022 © 2022 The Author(s). Published by Elsevier Inc.
https://doi.org/10.1016/j.cell.2022.06.045

CellPress

Cell

CellPress
OPEN ACCESS

Article

Repeat-based holocentromeres influence genome architecture and karyotype evolution

Paulo G. Hofstatter,[1,12] Gokilavani Thangavel,[1,12] Thomas Lux,[2,12] Pavel Neumann,[3] Tihana Vondrak,[3,4] Petr Novak,[3] Meng Zhang,[1] Lucas Costa,[5] Marco Castellani,[1] Alison Scott,[1] Helena Toegelová,[6] Joerg Fuchs,[7] Yennifer Mata-Sucre,[5] Yhanndra Dias,[5] André L.L. Vanzela,[8] Bruno Huettel,[9] Cicero C.S. Almeida,[10] Hana Šimková,[6] Gustavo Souza,[5] Andrea Pedrosa-Harand,[5] Jiri Macas,[3] Klaus F.X. Mayer,[2,11] Andreas Houben,[7] and André Marques[1,13,*]

[1]Department of Chromosome Biology, Max Planck Institute for Plant Breeding Research, Cologne, NRW 50829, Germany
[2]Plant Genome and Systems Biology, German Research Center for Environmental Health, Helmholtz Zentrum München, Ingolstädter Landstraße 1, 85764 Neuherberg, Germany
[3]Biology Centre, Czech Academy of Sciences, Institute of Plant Molecular Biology, České Budějovice 37005, Czech Republic
[4]Faculty of Science, University of South Bohemia, České Budějovice 37005, Czech Republic
[5]Laboratory of Plant Cytogenetics and Evolution, Department of Botany, Centre of Biosciences, Federal University of Pernambuco, Recife, Pernambuco 50670-901, Brazil
[6]Institute of Experimental Botany of the Czech Academy of Sciences, Centre of Plant Structural and Functional Genomics, Olomouc 779 00, Czech Republic
[7]Breeding Research, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) Gatersleben, Seeland, Saxony-Anhalt 06466, Germany
[8]Laboratory of Cytogenetics and Plant Diversity, State University of Londrina, 86097-570 Paraná, Brazil
[9]Max Planck Genome-Centre Cologne, Max Planck Institute for Plant Breeding Research, Cologne, NRW 50829, Germany
[10]School of Agronomical Sciences, Campus Arapiraca, Federal University of Alagoas, Arapiraca 57309-005, Brazil
[11]School of Life Sciences Weihenstephan, Technical University of Munich, Alte Akademie 8, 85354 Freising, Germany
[12]These authors contributed equally
[13]Lead contact
*Correspondence: amarques@mpipz.mpg.de
https://doi.org/10.1016/j.cell.2022.06.045

SUMMARY

The centromere represents a single region in most eukaryotic chromosomes. However, several plant and animal lineages assemble holocentromeres along the entire chromosome length. Here, we compare genome organization and evolution as a function of centromere type by assembling chromosome-scale holocentric genomes with repeat-based holocentromeres from three beak-sedge (*Rhynchospora pubera*, *R. breviuscula*, and *R. tenuis*) and their closest monocentric relative, *Juncus effusus*. We demonstrate that transition to holocentricity affected 3D genome architecture by redefining genomic compartments, while distributing centromere function to thousands of repeat-based centromere units genome-wide. We uncover a complex genome organization in *R. pubera* that hides its unexpected octoploidy and describe a marked reduction in chromosome number for *R. tenuis*, which has only two chromosomes. We show that chromosome fusions, facilitated by repeat-based holocentromeres, promoted karyotype evolution and diploidization. Our study thus sheds light on several important aspects of genome architecture and evolution influenced by centromere organization.

INTRODUCTION

Most eukaryotes are monocentric, meaning that their centromeres are restricted to single regions on each chromosome. These centromeric regions can range from kilobases (kbs) to megabases (Mbs) in length and comprise often specific repeats (Gohard et al., 2014). Holocentromeres, by contrast, consist of multiple centromeric units distributed along the poleward surface of metaphase chromosomes, extending from one telomere to the other, and are thus typically visible as a line on each chromatid (Heckmann et al., 2013; Senaratne et al., 2021; Steiner and Henikoff, 2014). Holocentromeres are hypothesized to stabilize chromosomal fragments and fusions that favor karyotype rearrangements and speciation (Mandrioli and Manicardi, 2020), directly influencing chromosome evolution (Schubert and Lysak, 2011). This hypothesis is supported by the fact that holocentromeres have evolved independently several times in different plant and animal lineages (Escudero et al., 2016; Melters et al., 2012).

Aside from their function in cell division, centromeres have an evolutionarily conserved role in determining large-scale genome architecture and chromatin composition (Muller et al., 2019). Centromeres in monocentric chromosomes influence the distribution of genes, euchromatin- and heterochromatin-specific

© 2022 The Author(s). Published by Elsevier Inc.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

post-translational histone modification domains, transposable elements (TEs), and meiotic crossovers (Fernandes et al., 2019; Fuchs et al., 2006; Muller et al., 2019; Naish et al., 2021). However, genome organization and chromatin composition of organisms with holocentric chromosomes is poorly understood, and it is likely that holocentric species differ markedly from the monocentric paradigm.

The beak-sedge *Rhynchospora pubera* (Cyperaceae, sedges) has repeat-based holocentromeres (Marques et al., 2015), as do other species from the same genus (Costa et al., 2021; Ribeiro et al., 2017). *R. pubera* holocentromeres are associated with a single tandem-repeat family (the centromeric 172-bp unit *Tyba* repeat) and the centromeric retrotransposon of *Rhynchospora* (*CRRh*), giving rise to thousands of small centromere units across the genome (Marques et al., 2015). The lack of a *Rhynchospora* reference genome has, however, hampered detailed studies about its intriguing centromere organization.

Here, we combined genomic and chromatin analyses to elucidate genomic adaptations related to different centromere organizations. We report the full characterization of a holocentric genome containing thousands of repeat-based centromere units. We show that this centromere organization influences the 3D genome architecture by redefining the extent of genomic compartments due to the lack of centromere clustering. Strikingly, despite substantial genome restructuring, the epigenetic regulation of centromere units in beak-sedges resembles that of monocentric centromeres, as in *Arabidopsis thaliana* (Naish et al., 2021). This observation suggests evolutionarily conserved epigenetic regulation of repeat-based centromeres in both monocentric and holocentric organisms. We further reveal that chromosome fusions facilitated by repeat-based holocentromeres reduce chromosome number and can act as an alternative to diploidization after genome doubling without the need for genome downsizing. Our work sheds light on the role of centromeres in overall genome organization and chromosome evolution.

RESULTS

Holocentricity affects spatial genome organization

To identify the genomic adaptations related to the transition to holocentricity, we constructed chromosome-scale reference genomes using PacBio HiFi sequencing and Dovetail Omni-C (DNase-based Hi-C) for three holocentric *Rhynchospora* species, *R. pubera* (n = 5; haploid nuclear genome size [1C] = 1.61 Gb), *R. breviuscula* (n = 5; 1C = 415 Mb), and *R. tenuis* (a plant with the fewest known chromosomes; n = 2; 1C = 394 Mb) (Castiglione and Cremonini, 2012; Vanzela et al., 1996), as well as their closest monocentric relative, the rush *J. effusus* (n = 21; 1C = 271 Mb) (Guerra et al., 2019; (Figures 1, 2, S1A, and S1B; Table S1; STAR Methods).

J. effusus showed a typical monocentric configuration of chromatin interaction within A (euchromatin) and B (heterochromatin) compartments, including some degree of a telomere-to-centromere axis (Figures 2A and 2B; see Hoencamp et al., 2021).

The concept of chromosome arms does not apply to holocentric species, as centromeres are ubiquitous. Consequently, we observed no large-scale compartmentalization or telomere-to-centromere axes, as evidenced by the chromatin configuration capture (Hi-C) contact matrices of our three *Rhynchospora* species (Figures 2C, 2D, S1A, and S1B). Further quantification of intrachromosomal (*cis*) and interchromosomal (*trans*) chromatin contacts revealed a significantly higher ratio ($p < 4.04e-05$) of *cis* versus *trans* interactions in all *Rhynchospora* species compared with the monocentric *J. effusus* (Figure S1C). Thus, holocentric beak-sedges are characterized by higher intrachromosomal spatial genome organization and lack of centromere clustering.

The distribution of genomic features differed markedly between holocentric *Rhynchospora* and monocentric *J. effusus* (Figures 2E and 2F). *Rhynchospora* had a uniform distribution of genes, transcriptional activity, *Tyba* centromeric repeats, TEs, and DNA methylation (Figures 2F, S1D, and S1E). By contrast, *J. effusus* genes were concentrated toward telomeric regions, while TEs and tandem repeats clustered toward centromeric regions (Figure 2E). Genome-wide gene distribution and transcriptional activity were positively correlated, while repeat distribution was positively correlated with overall DNA methylation levels (Figure 2E). Genome-wide CpG methylation (mCpG) was lower in *R. pubera* than in *J. effusus*, whereas CHG methylation was higher and CHH methylation was the same in both species (Figures S1F and S1G). Thus, transition to holocentricity likely affects 3D genome architecture by redefining the extents of genomic compartments and their relationships to each other.

Genetic and epigenetic composition of repeat-based holocentromeres

We analyzed the sequence organization and chromatin structure of the *Rhynchospora* repeat-based holocentromeres. The contiguity of our assemblies, coupled with the short array size of centromeric *Tyba* repeats, allowed us to resolve mostly complete *Tyba* arrays in the three *Rhynchospora* genomes. While total number and amount of *Tyba* arrays increased with chromosome size (Figures 3A and 3B), the density of arrays decreased (Figure 3C). Average array sizes of 20.3, 20.5, and 19.8 kb, and average spacing between two consecutive arrays of 368, 492, and 424 kb were found in *R. breviuscula*, *R. pubera*, and *R. tenuis*, respectively (Figures 3D and 3E). These results confirm a similar overall organization of centromeric *Tyba* repeats among the three *Rhynchospora* species. In common with monocentric centromeric repeats (Kasinathan and Henikoff, 2018), we also found a high frequency of dyad symmetries in the *Tyba* consensus sequences of all three *Rhynchospora* species (Figure 3F).

Chromatin immunoprecipitation followed by sequencing (ChIP-seq) confirmed the highest enrichment of centromeric histone H3 (CENH3) for the *Tyba* repeats and lower enrichment for *CRRh* throughout the entire *R. pubera* and *R. breviuscula* genomes (Figures 4A–4C and S1H; Table S2). We detected 2,753 and 995 CENH3-binding regions (hereafter CENH3 domains) evenly distributed across the five chromosomes of *R. pubera* and *R. breviuscula*, respectively. In both species, length, density, and spacing of CENH3 domains followed a similar pattern to the number of *Tyba* arrays detected (Figures 3A–3E). Considering that one CENH3 domain is equivalent to one centromere unit,

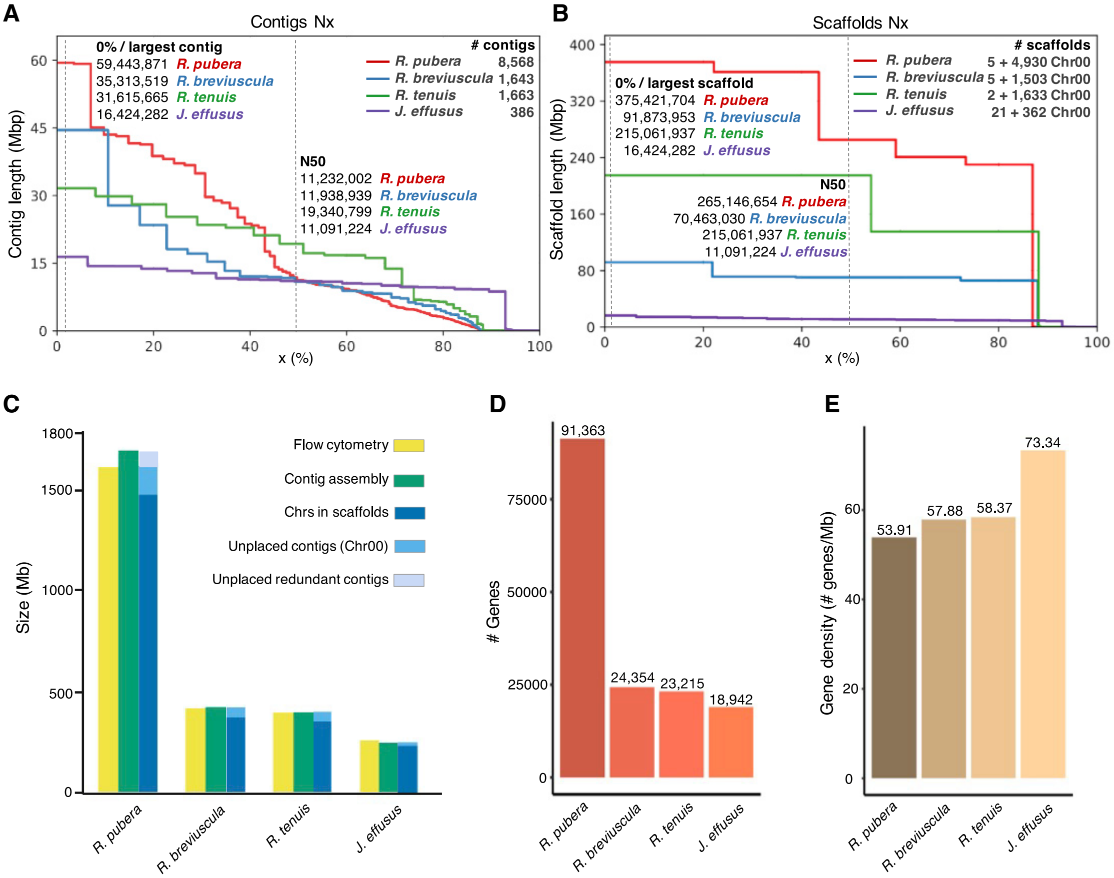


Figure 1. Summary of genome sizes, assemblies, scaffolding, and annotations

(A and B) Assembly (A) and final scaffolding (B) statistics.

(C) Comparison of estimated genome size and assembly and scaffolding sizes.

(D) Total number of high-confidence annotated genes.

(E) Gene density per Mb.

See also Table S1.

on average, each *R. pubera* chromosome carried 600 centromere units (1.88 domains/Mb), while the smaller chromosomes of *R. breviuscula* carried 200 centromere units each on average (2.69 domains/Mb) (Figures 3A–3C). Thus, genome/chromosome size may be negatively correlated with centromere unit density in beak-sedges. Genome-wide there was a significant association between CENH3 domains and *Tyba* repeats for both species ($p < 0.05$), confirming that *Tyba* repeats are the main CENH3-binding sites. Therefore, repeat-based holocentromeres are likely to be conserved and associated with *Tyba* repeats in beak-sedges.

In the monocentric *J. effusus*, the histone mark H3K4me3, which is euchromatin specific, showed dispersed labeling along chromosome arms, while H3K9me2 (heterochromatin specific) was concentrated at pericentromeric regions and co-localized with chromocenters in interphase (Figure S1I). By contrast, in the holocentric *R. pubera*, both euchromatin- and heterochromatin-specific histone marks were intermingled all along the chromosomes with a constant density even toward the subtelomeric and central chromosomal regions (Figures 4A–4C). Locally, H3K4me3 was mostly highly enriched at the promoter regions of protein-coding genes, whereas H3K9me2 was enriched on small heterochromatic islands, typically resembling TEs (Figure 4C). H3K4me3 was depleted at CENH3 domains, while H3K9me2 showed residual enrichment. We noticed a slight increase in H3K9me2 enrichment flanking CENH3 domains relative to the core region, mimicking the pericentromeric chromatin composition in monocentromeres (Figure 4C).

Irrespective of centromere type, gene bodies were highly enriched for mCpG in both *R. pubera* and *J. effusus*, with a sharp decrease at promoters and terminal regions. Methylation in the CHH and CHG contexts was much lower for the gene bodies than for intergenic regions (Figures 4D and 4E), as previously reported for other plants (Feng et al., 2020). Remarkably, despite the differences in chromosome organization, both the *Tyba* repeats in *R. pubera* and tandem repeats in centromeric regions of *J. effusus* chromosomes were highly enriched for mCpG at levels similar to those for TEs (Figures 4D and 4E). mCHG was sharply enriched flanking CENH3-binding regions in *R. pubera*, resembling the H3K9me2 pattern (Figure 4C). We obtained a

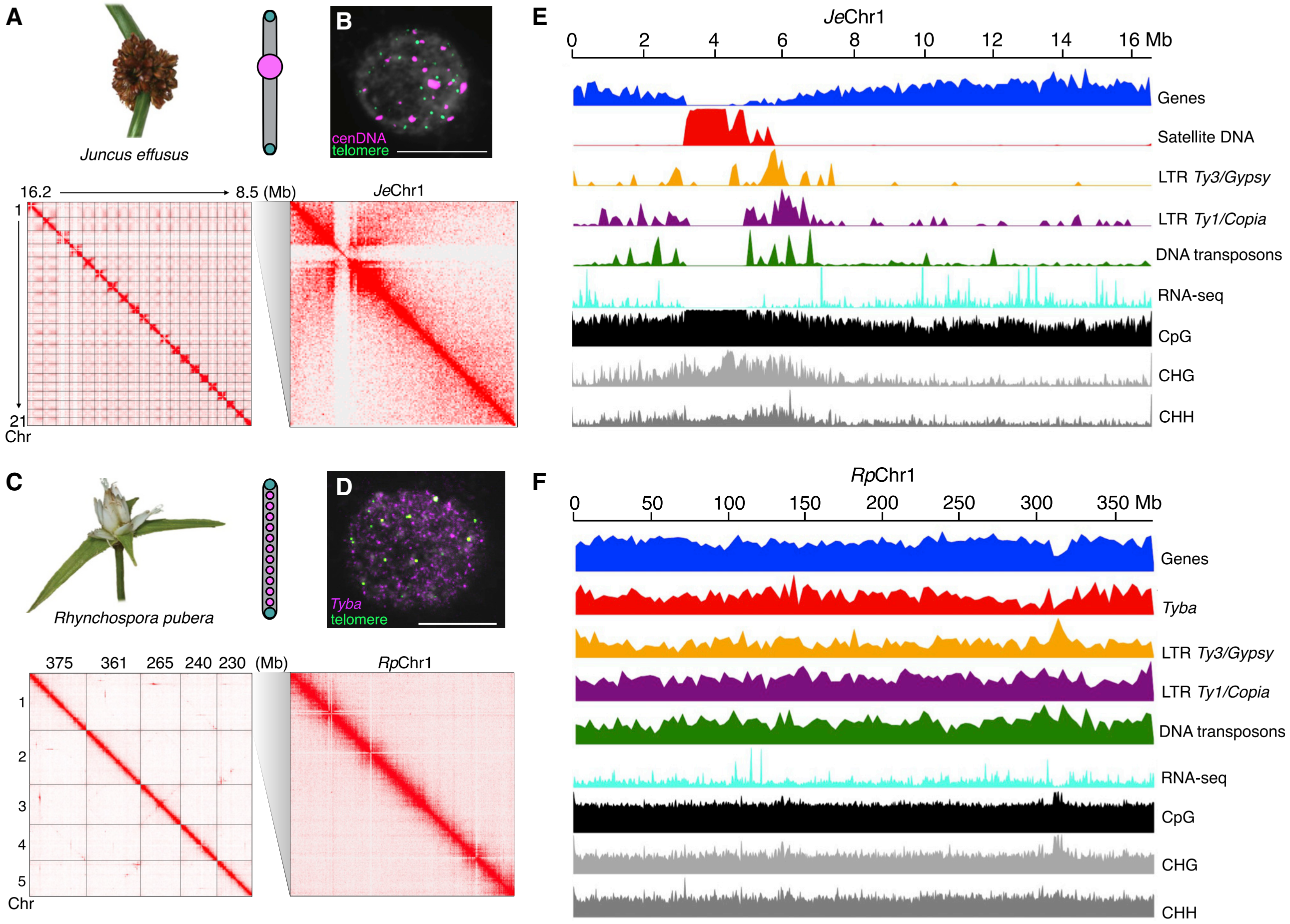


Figure 2. Spatial genome organization: monocentric versus holocentric chromosomes

(A) *J. effusus* (top left) genome contact map (bottom left) and chromosome 1 (*Je*Chr1) detailed view (bottom right). Centromere organization in monocentric chromosomes (top right).

(B) Interphase nucleus hybridized with DNA probes for the centromeric DNA (cenDNA) and telomeric sequence in *J. effusus*.

(C) *R. pubera* (top left) genome contact map (bottom left) and *Rp*Chr1 detailed view (bottom right). Centromere organization in holocentric chromosomes (top right).

(D) Interphase nucleus hybridized with DNA probes for the centromeric repeat *Tyba* and telomeric sequence in *Rhynchospora*.

(E) *Je*Chr1 and (F) *Rp*Chr1 detailed view showing the clustered (*Je*Chr1) and uniform (*Rp*Chr1) distribution of main genomic features, which are typical for monocentric chromosomes and holocentric chromosomes, respectively. Window sizes for sequence-type distribution density, 100 kb (*J. effusus*) and 3 Mb (*R. pubera*). Centromeres and telomeres in chromosome models are represented by magenta and green circles, respectively. Scale bars, 10 μm.

See also Figure S1.

similar pattern for mCHG at centromeric repeats in *J. effusus* (Figures 2D, 4D, and 4E). TEs showed the highest enrichment for mCHG and mCHH, while *Tyba* repeats displayed lower levels of mCHH, similar to genes (Figures 4D and 4E). Our results argue for the presence of a pericentromere-like chromatin state around the ends of centromere units in *Rhynchospora* that may mark the borders for CENH3 loading.

A typical centromere unit in *R. pubera* comprised a single *Tyba* array surrounded by genes and TEs (Figure 4F). We detected CENH3 domains all along the chromosomes, even in *Tyba* arrays located near telomeres like those at both ends of *R. pubera* chromosome 2 (*Rp*Chr2) (Figures 4F and 4G), confirming the telomere-to-telomere centromere activity of holocentric chromosomes. Notably, we observed an enrichment for H3K4me3 and actively transcribed genes close to centromere units, with an average distance of 6.3 kb (Figures 4H and 4I). We identified 313 genes that showed at least a 1-base-pair (bp) overlap with CENH3 domains. We even detected actively transcribed genes with typical H3K4me3 enrichment inside CENH3 domains (Figure 4H), a characteristic only rarely observed in monocentric organisms (Mizuno et al., 2011; Schotanus et al., 2021). Both CENH3 association and transcription were frequently reduced in genic regions inserted into centromere units, compared with genic regions residing outside the core centromere unit (Figure 4H), reflecting the precise regulation of chromatin composition of the *R. pubera* genome. *CRRh* was frequently inserted into *Tyba* arrays enriched for CENH3, but also H3K9me2 and some level of H3K4me3, suggesting a different epigenetic regulation

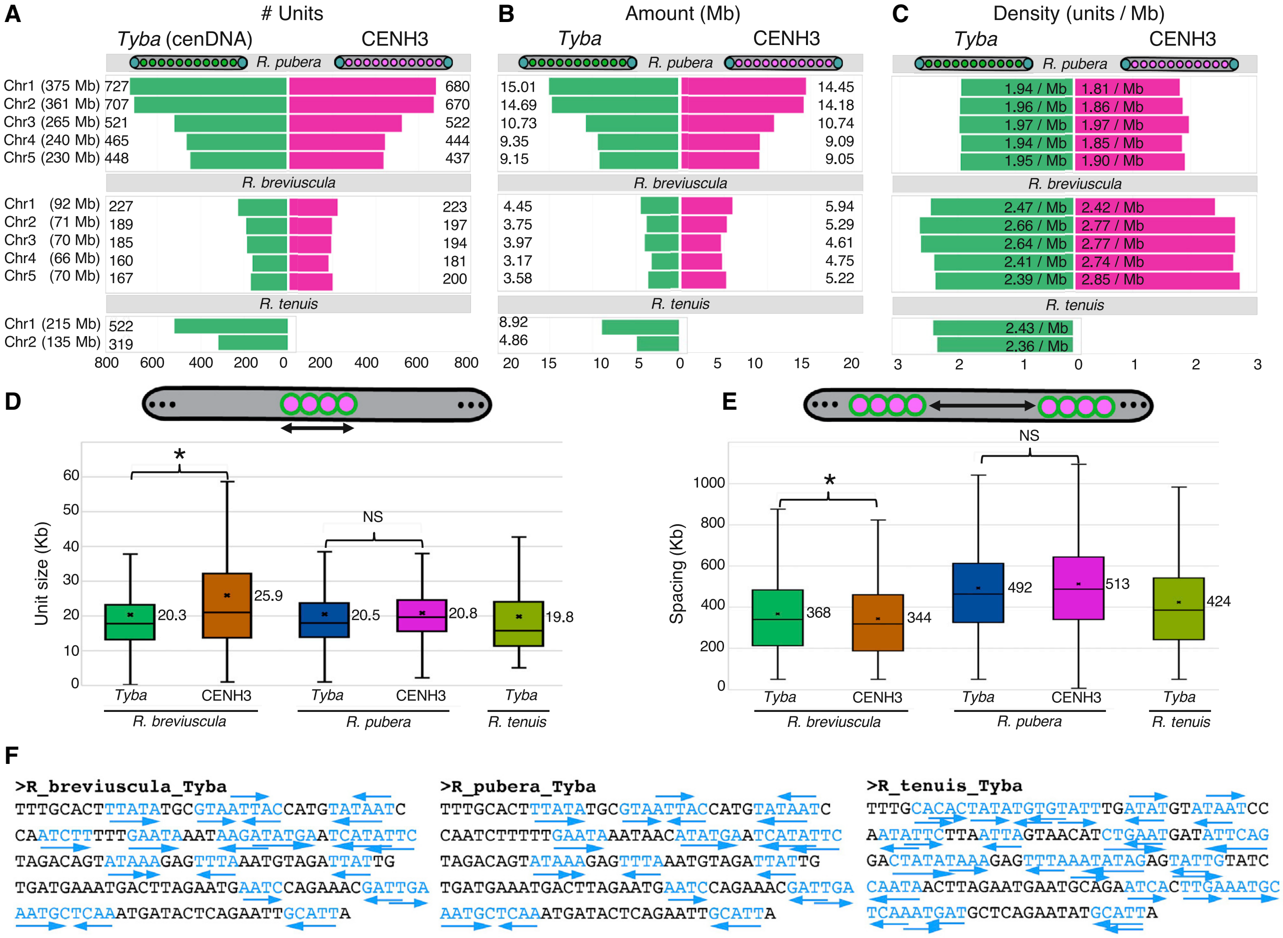


Figure 3. Features of *Tyba* centromeric DNA and CENH3 domains among *Rhynchospora* species

(A) Total number per chromosome of annotated *Tyba* arrays and CENH3 domains.

(B) Total amount of bases associated with *Tyba* arrays and CENH3 domains.

(C) Density of *Tyba* arrays and CENH3 domains per chromosome.

(D) Size distribution of *Tyba* arrays and CENH3 domains. NS = not significant.

(E) Spacing between two consecutive centromere arrays/domains among *Rhynchospora* species. Asterisks indicate Dunn's test, $p < 0.05$.

(F) Patterns of DNA dyad symmetry in the *Tyba* consensus sequences of the three *Rhynchospora* species.

of this retroelement compared with *Tyba* repeats (Figure 4I). Our results thus point to fine-scale epigenetic regulation of genomes with repeat-based holocentromeres.

Transposition partially explains genome-wide *Tyba* dispersal and expansion

Tyba repeats in *R. pubera* can be flanked by *TCR1* and *TCR2* repeats, suggesting that some *Tyba* arrays are part of larger repetitive elements (Marques et al., 2015). The consensus full-length *TCR1* element contained a *Tyba* array with a 5′ sequence of approximately 4.8-kb and a 136-bp 3′ sequence. The element possessed no open reading frame (ORF) and lacked terminal repeats, and its 5′ and 3′ ends harbored the ATC and CTAGT sequence motifs, respectively, suggesting that *TCR1* is a nonautonomous *Helitron* TE (Thomas and Pritham, 2015), from the same family as a fully autonomous *Helitron* element (*Helitron-27*) in the *R. pubera* genome. Despite sharing conserved terminal sequences, *TCR1* and *Helitron-27* exhibited no similarities in their internal regions. We identified three intact copies of the autonomous *Helitron-27* in the genome with high mutual similarity (Table S3), each encoding a full *Helitron* helicase (1,340 amino acids), indicating that *TCR1* and *Helitron-27* elements are still capable of transposition. We further identified an additional 322 full-length elements (Table S3) with both *TCR1* termini as well as *Tyba* and another 146 partial elements with the 3′-terminal sequence and containing *Tyba* within the upstream 500-bp region. We conclude that at least 468 *Tyba*-containing loci in the genome resulted from the transposition activity of *TCR1* elements. The full-length *TCR1* sequences were 6.9–49.6 kb (24.8 kb average), containing 1.2–31.3 kb from *Tyba* (15.7 kb average). In many *TCR1* elements, *Tyba* arrays were split into multiple segments due to insertions of other sequences, showing that multiple *Tyba* loci can originate from a single *TCR1* insertion (Figures 4J and 4K). Importantly, a comparison of *TCR1* and

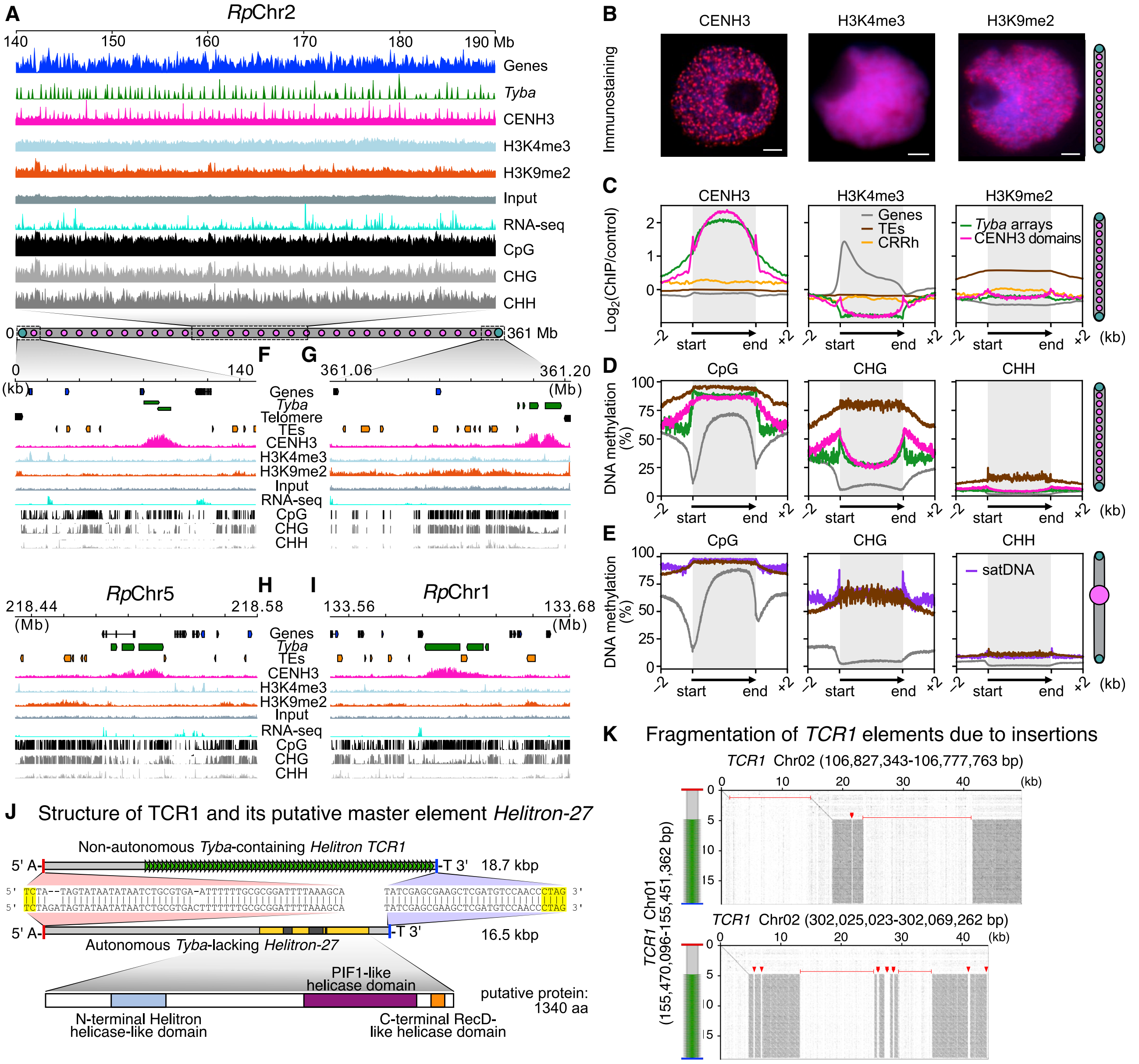


Figure 4. Genetic and epigenetic composition of repeat-based holocentromeres in *R. pubera*

(A) Zoomed-in view of *Rp*Chr2 showing a 50-Mb region with multiple CENH3 domains that are closely correlated with *Tyba* repeat distribution. Gene and *Tyba* densities were calculated over 100-kb windows.

(B) Immunostaining of *R. pubera* interphase nuclei for CENH3, H3K4me3, and H3K9me2. Scale bars, 2 μm.

(C) Enrichment of CENH3, H3K4me3, and H3K9me2 from the start and end of different types of sequences: genes (gray line), TEs (brown), *CRRh* (yellow), *Tyba* repeats (green), and CENH3 domains (magenta). ChIP-seq signals are shown as $\log_2$ (normalized RPKM ChIP/input).

(D and E) Enrichment of DNA methylation in the CpG, CHG, and CHH contexts for the same sequence types as shown I(C) for *R. pubera* (D) and *J. effusus* (E), genes (gray line), satDNA (purple) and TEs (brown). Gray boxes in (C)–(E) highlight the modification enrichment over the body of each sequence type.

(F and G) Close-up view of the first (F) and last (G) centromere units of *Rp*Chr2, which are composed of a *Tyba* repeat array very close to the telomere and showing the typical CENH3 enrichment.

(H) A centromere unit where an active gene is intermingled with the *Tyba* repeat.

(I) A *Tyba* array showing an insertion of the centromeric retrotransposon *CRRh* and CENH3-binding activity.

(J) Structures of the typical nonautonomous *TCR1* element (Chr01:155470096–155451362) and its likely master element *Helitron-27* (Chr05:40901972–40918485). Similarities between *TCR1* and *Helitron-27* are mostly restricted to the terminal sequences. The 5′- and 3′-terminal sequences are in red and blue,

(legend continued on next page)

CENH3 domains revealed that the vast majority (98.7%) of full-length *TCR1* elements are embedded within or overlap with the centromere units (Table S2).

*Helitron*s with boundaries similar to *TCR1*/*Helitron-27* were present in *R. tenuis* and *R. breviuscula*; however, all but one of the full-length elements in these two genomes lacked *Tyba*. The sole exception was a single element from *R. breviuscula* (Chr1:69162288–69195619) with 5′ and 3′ boundary sequences characteristic of this *Helitron* family as well as a *Tyba* array; however, the remaining sections lacked any similarity to the *TCR1* of *R. pubera*. These results suggest that *Tyba* was amplified as a part of a *TCR1 Helitron* only in the genome of *R. pubera*.

The *TCR2* element was found to be a miniature inverted-repeat TE (MITE) and ranged from 672 to 1,235 bp, likely originated from the DNA transposon MuDR with shared similarity (up to 97%) in the terminal inverted repeats. All 158 full-length *TCR2* elements identified in the *R. pubera* genome were in *Tyba* arrays, but none were characterized by *Tyba* insertions. Thus, *TCR2* elements did not contribute to the dispersal of *Tyba* in the *R. pubera* genome.

R. pubera is a cryptic auto-octoploid with n = 5 chromosomes

The *R. pubera* genome is 4 times larger than that of its closely related species, despite sharing the same ancestral chromosome number (ACN) (*x* = 5) (Burchardt et al., 2020; Ribeiro et al., 2018) (Figure S2A). One explanation for this pronounced genome expansion would be a sudden and massive proliferation of repeat elements. However, we observed no accumulation of repeats when comparing repeat abundance profiles among closely related *Rhynchospora* species (Figure S2A). Thus, a different process must be responsible for the large genome size in *R. pubera*.

Completeness assessment of the *R. pubera* genome by calculating the benchmarking universal single-copy orthologs (BUSCOs) score revealed a surprisingly high level of gene duplications (96.0% duplicated BUSCOs) (Figure S2B). Annotation of the genome yielded far more high-confidence gene models (91,363) in *R. pubera* compared with the other species (Figures 1D and 1E; Table S1), confirming the high level of gene duplication (Figure S2C). Self-synteny analysis revealed that the *R. pubera* genome comprises two large syntenic blocks in four copies across the five chromosomes (Figure S3A). The larger syntenic block, named Block1, corresponded to the entire *Rp*Chr4 and *Rp*Chr5 and contributed to a large fraction of both *Rp*Chr1 and *Rp*Chr2. We identified the smaller block, named Block2, twice in an inverted arrangement in *Rp*Chr3, as well as in *Rp*Chr1 and *Rp*Chr2 (Figure S3A).

The distribution of synonymous substitutions per synonymous site (Ks) for coding sequences over the intragenomic syntenic blocks in *R. pubera* had a large peak indicative of recent and successive whole-genome duplication (WGD) events. An additional small peak was also observed, indicating an ancient WGD (Figure S3B). By filtering out the sequences showing the lowest Ks values, we determined that Block1 from *Rp*Chr1 shows higher sequence identity to *Rp*Chr4, which we renamed Block1A1 and Block1A2, respectively. Similarly, Block1 from *Rp*Chr2 showed higher sequence identity to *Rp*Chr5, which were thus named Block1B1 and Block1B2, respectively (Figures S3B and S3C). We confirmed the relationships of the four Block1 copies by comparative phylogenetic analysis (Figure S3D). A similar analysis of Block2 copies was inconclusive (Figures S3C and S3E). Using k-mer analysis, which provides information on genome size, ploidy, and genome structure through scrutiny of heterozygous k-mer pairs (Ranallo-Benavidez et al., 2020), we detected a higher incidence of homozygous and duplicated k-mers, favoring an autopolyploidy genome model for *R. pubera* (Figures S3F and S3G). Importantly, this analysis accurately determined the diploid heterozygous state of *R. breviuscula* and *R. tenuis* (Figure S4). Thus, *R. pubera* has an auto-octoploid genome shaped by two rounds of genome doubling explaining its large genome size. Post-polyploid genome shuffling events considerably reduced the chromosome number to n = 5.

Chromosome fusions explain karyotype evolution in beak-sedges

To explore the genome duplications seen in *R. pubera*, we compared its genome with its close relative *R. breviuscula*, which has the same chromosome number but a genome that is one-quarter the size (415 Mb) (Figure S2A). Assessment of the *R. breviuscula* genome revealed a high level of completeness, with a BUSCO score of 98.3%, and little gene duplication (2.1%) (Figures S2B and S2C), confirmed by the absence of self-synteny (Figure S1D). Gene annotation yielded 24,354 high-confidence gene models (Figures 1D and 1E; Table S1), 4 times fewer than in the *R. pubera* genome, as expected. Synteny analysis between both genomes illustrated how each *R. breviuscula* chromosome (*Rb*) is present in four copies in the *R. pubera* genome (Figure 5A). Remarkably, *Rp*Chr1 and *Rp*Chr2 contained all five putative *Rb*s in end-to-end configurations. *Rp*Chr3 contained *Rb*3 and *Rb*4 copied twice in an inverted order, comprising Block2, while *Rp*Chr4 and *Rp*Chr5 contained *Rb*1, *Rb*2, and *Rb*5, comprising Block1A and Block1B, respectively (Figure 5A). Thus, *R. breviuscula* likely conserved the ancestral karyotype, while *R. pubera* restored the ACN (*x* = 5) of its clade due to descending dysploidy, which was mediated by a complex chain of chromosome fusions, e.g., end-to-end fusions (EEFs), with 15 EEF junctions detected. Remarkably, each chromosome pair had a unique combination of ancestral chromosomes. We conclude that descending dysploidy involving a unique combination of chromosomes may be a strategy to avoid meiotic pairing issues that could potentially arise from autopolyploidy, thereby acting as a rapid route to diploidization facilitated by holocentricity.

---

respectively. Yellow, conserved *Helitron* sequence motifs in the alignment of *TCR1* and *Helitron-27* terminal sequences. Light gray, noncoding regions. Green triangles, *Tyba* array in *TCR1*. Yellow and dark gray, putative exons and introns in the *Helitron-27* coding region, respectively.

(K) Dot-plot comparison of a typical *TCR1* element (vertical sequences) with two other elements (horizontal sequences) that have insertions of *TCR1*-unrelated sequences marked as red lines and triangles.

See also Figure S1 and Tables S2 and S3.

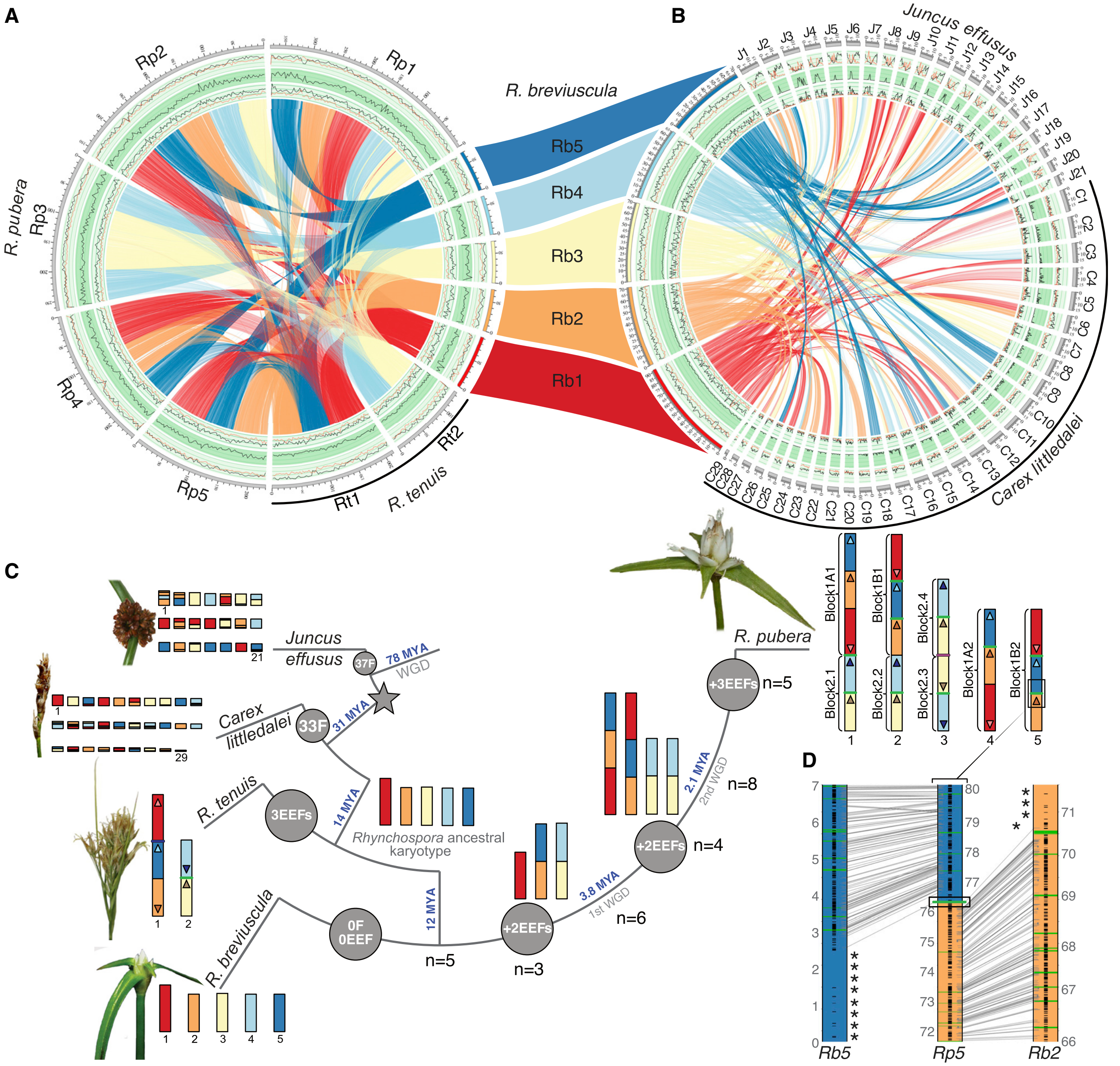


Figure 5. Genome organization and evolution of sedges and common rush

(A and B) Circos plots of *R. breviuscula* synteny to *R. pubera* and *R. tenuis* (A) and circos plots of *R. breviuscula* synteny to *J. effusus* and *C. littledalei* (B). Tracks from outside to inside: 1. genes (black line) and TEs (red line), 2. *Tyba*/tandem repeats (black line), and 3. LTR *Ty1/Copia* (black line) and *Ty3/Gypsy* (red line) retroelement distribution. Distribution of the main sequence classes was calculated in 3-Mb windows for *R. pubera*, *R. tenuis*, and *R. breviuscula* (A), in a 1-Mb window for *R. breviuscula*, and in 500-kb windows for *C. littledalei* and *J. effusus* (B).

(C) Karyotype evolution and synteny conservation in sedges and common rush. Transition to holocentricity is indicated by a star. Hypothetical ancestral karyotype for *Rhynchospora* based on the simplest karyotype of *R. breviuscula* illustrates frequent end-to-end fusions (EEFs) in beak-sedges. For reconstruction of karyotype evolution in *R. pubera* see also Figures S4 and S5. Arrow heads, orientation of the *R. breviuscula* chromosomes in the *R. pubera* and *R. tenuis* ideograms. For both *J. effusus* and *C. littledalei*, ideograms indicate the syntenic blocks to *R. breviuscula* chromosomes. Numbers of putative EEFs or fission (F) events necessary to transform the hypothetical *Rhynchospora* ancestral karyotype into the extant genomes are within the gray circles. Repeat sequences at the junctions between *Rb* blocks are indicated by colored bars (Tyba, green; rDNA, purple; telomeric DNA, blue) in *R. tenuis* and *R. pubera* ideograms.

(D) *R. pubera* Chr5 showing a *Tyba* array (black rectangle) at the junction between syntenic Rb2 and *Rb*5 blocks. Synteny from *Rp*Chr5 to Rb2 and *Rb*5 stops close to the last *Tyba* array, which is followed by a gene-poor, TE-enriched region, mainly LTR *Ty3/Gypsy* of the *Athila* clade (indicated by asterisks) that are frequently within *R. breviuscula* subtelomeric regions but absent in the fused chromosomes. Genes and *Tyba* arrays are annotated as black stripes and green lines, respectively.

See also Figures S2, S3, S4, S5, S6, and Table S4.

Because we detected several EEFs in multiple copies in *R. pubera*, we assessed whether they were derived from the same rearrangement or if they arose from multiple independent events. All duplicated EEFs in the *R. pubera* genome, e.g., Rb2/*Rb*5, *Rb*3/*Rb*4, *Rb*1/*Rb*2, and *Rb*1/*Rb*5 EEFs, share a fusion signature involving the same regions. This observation suggested that the Rb2/*Rb*5 and *Rb*3/*Rb*4 EEFs, which are present 4 times in the *R. pubera* genome emerged only once—before the first WGD event (Figures 5C and S5). The *Rb*1/*Rb*2 and *Rb*1/*Rb*5 EEFs, which were found twice, likely emerged after the first WGD event. Finally, we found the *Rb*1/*Rb*4, Rb2/*Rb*4, and *Rb*3/*Rb*3 EEFs only once, suggesting that they occurred after the second WGD (Figures 5C and S5).

The *Rb*3/*Rb*4 EEF, which forms Block2 in the *R. pubera* genome, was likely maintained as a duplicated fused chromosome after the first WGD, which might have allowed a longer period of tetrasomic inheritance. This hypothesis might explain the fact that the sequences of the four copies from Block2 cannot be distinguished from each other, in contrast to Block1.

We attempted to date the duplication events using a set of conserved genes shared among the four copies of Block1, which revealed the first WGD event as occurring around 3.8 million years ago (Mya) followed by a second WGD event around 2.1 Mya (Figure S3D). Based on this analysis, we deduced the origin and evolution of the *R. pubera* karyotype (Figure 5C). These results further support an autopolyploid origin for *R. pubera* and confirm a short interval between the two rounds of WGDs, indicating rapid chromosome number reduction in this species.

We carried out a number of analyses to determine the origin of the reduced karyotype in *R. tenuis* ($n = 2$). BUSCO analysis of its genome revealed high completeness (98.5% against the viridiplantae_odb10 dataset) and little duplication (3.7%) (Figures S2B and S2C). Gene annotation yielded 23,215 high-confidence gene models (Figures 1D and 1E). The absence of self-synteny in the *R. tenuis* genome ruled out large duplications (Figure S2E). Synteny comparison between *R. tenuis* and *R. breviuscula* genomes showed that again all *Rb*s were present in simple end-to-end configurations in the *R. tenuis* genome, explaining its karyotype by descending dysploidy from $n = 5$ to $n = 2$ (Figure 5A). Strikingly, we observed similar associations of syntenic *Rb* blocks as found in Block1 and 2 in both *R. pubera* and *R. tenuis*, where *Rt*Chr1 resembled Block1B and was composed of Rb2, *Rb*5, and *Rb*1, while *Rt*Chr2 resembled Block2, consisting of *Rb*3 and *Rb*4 (Figure 5A). However, the orientation of chromosome ends involved in the EEFs differed in the two instances, suggesting that the EEFs occurred independently (Figure 5C).

Despite their high chromosome number and centromere-type differences, *J. effusus* and the previously available genome for the sedge *Carex littledalei* (its homotypic synonym, *Kobresia littledalei*) (Can et al., 2020) showed a typical diploid gene content and no evidence of any recent WGD, outside of a shared ancient WGD between sedges and rushes (Figure S4). The *J. effusus* genome also revealed high completeness (100% viridiplantae_odb10 dataset) and little duplication (1.6%) (Figures S2B and S2C). Annotation of its genome yielded 18,942 high-confidence gene models (Figures 1D and 1E; Table S1). Synteny analysis further revealed that most *J. effusus* and *C. littledalei* chromosomes are present as highly collinear blocks across the five chromosomes of *R. breviuscula*, suggesting a high conservation of synteny although the group is ancient (78 Mya) (Figures 5B and 5C). Thus, neither the low nor the high chromosome numbers observed in many holocentric species necessarily reflect the absence or presence of recent polyploidy, respectively, and these numbers should be interpreted with caution in the absence of detailed genomic studies.

Tyba repeats are frequently present at the junctions of end-to-end fusions

TEs can influence chromosomal rearrangements (Lonnig and Saedler, 2002). To assess their possible role in the EEFs observed in *Rhynchospora* genomes, we looked for enrichment of specific repeats at the ends of *Rb*s and near the junctions of EEFs in *R. pubera* and *R. tenuis*. We detected a high density of TEs in almost all subtelomeric regions of *Rb*s. These repeat-rich regions varied from 500 kb to 3 Mb in size, were mainly enriched for an LTR *Ty3/Gypsy* element of the *Athila* clade, were poorly enriched for genes, and lacked *Tyba* repeats (Figures 5A, 5B, and S1D). Notably, the *R. breviuscula* subtelomeric repeat-rich regions were largely missing at the junctions of fused chromosomes in both *R. pubera* and *R. tenuis* (Figures 5D and S6). Remarkably, we detected *Tyba* repeats exactly at the EEF junctions in 10 out of the 15 EEFs of *R. pubera*, while we observed a small 45S rDNA remnant array (with only five 18S-5.8S-26S units) in one EEF (Figures 5C, 5D, and S6A–S6G). In *R. tenuis*, we also identified a *Tyba* repeat array in one out of three EEF junctions (*Rb*3/*Rb*4 junction on *Rt*Chr2), while an interstitial telomeric site (536 bp) was detected at the *Rb*5/*Rb*1 junction (Figures 5C and S6H).

Emergence and loss of CENH3 domains related to *Tyba*

We used the duplicated genome copies of *R. pubera* to study cases of paralogous CENH3 domains and *Tyba* arrays. Of 660 groups of paralogous regions, 66% of the CENH3 domains were present in all four copies (Table S5). We also identified 50 groups of paralogous regions in which the CENH3 domain was lost in one of the paralogs. Most cases (88%) were associated with *Tyba* loss (Figures 6A and S7A) or a reduced size of the *Tyba* region in loci devoid of CENH3 signal compared with their paralogous regions bound by CENH3. We observed the likely inverse event, i.e., the gain of a new CENH3 domain, in groups of paralogous regions where we only identified the CENH3 domain in only one of the four paralogous regions. In the newly acquired CENH3 domain, there was either a new *Tyba* insertion, likely due to a new insertion of *TCR1* (Figures 6B and S7B), a *Tyba* expansion, or the insertion of a new TE (which was most frequent). However, the ChIP/input ratios within these potentially new CENH3 domains containing a new TE insertion (1.5) were significantly lower ($p < 2.2.e{-}16$) than the ChIP/input ratios in potentially new CENH3 domains associated with a new insertion of a *Tyba* element (4.1).

DISCUSSION

Here, we report high-quality and contiguous chromosome-scale reference genomes for three species with repeat-based holocentromeres, *R. pubera*, *R. breviuscula*, and *R. tenuis*, and their closest monocentric relative, *J. effusus*. These newly assembled genomes provide a valuable resource for comparative biology

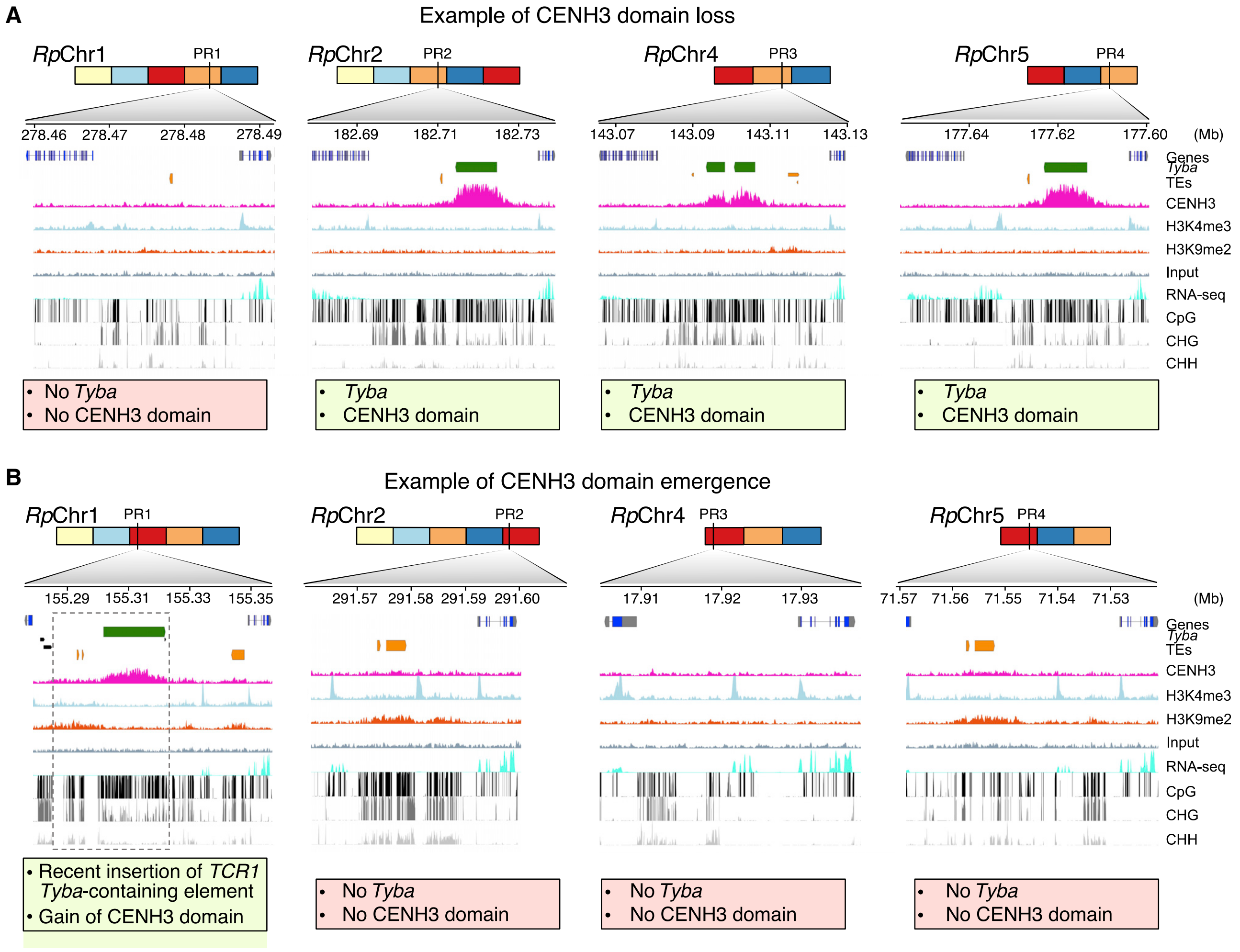


Figure 6. Emergence and loss of CENH3 domains in *R. pubera*

(A) CENH3 domain with *Tyba* array loss in one of the four paralogous regions, while the other three copies retain the *Tyba* array. Zoomed-in view of all four regions demonstrates the CENH3 domain loss only in the *Rp*Chr1 copy.

(B) CENH3 domain with *Tyba* array gain in one of four paralogous regions due to a transposition of a *Tyba*-containing *TCR1* in *Rp*Chr1, while the other three copies lack the *Tyba* array. The gained locus is indicated by the dashed box. Zoomed-in view of all four regions demonstrates the acquisition of a new CENH3 domain only in the *Rp*Chr1 copy. PR, paralogous region. Note that the four copies shared a similar chromatin composition.

See also Figures S5 and S7 and Table S5.

and studies related to genome adaptation to different centromere types.

Repeat-based holocentromeres influence genome organization and regulation

Repeat-based holocentromeres in beak-sedges comprise small islands (20–25 kb) of centromeric *Tyba* repeats, in which high mCpG, low H3K9me2, and depletion of H3K4me3 distinguish them from other holocentric genomes with and without repeat-based holocentromeres (Cortes-Silva et al., 2020; Despot-Slade et al., 2021; Nhim et al., 2022; Steiner and Henikoff, 2014). The association levels of H3K9me2 and mCHG at the core (low) and flanking (high) centromere units in *R. pubera* are strikingly similar to the recently reported *A. thaliana* centromeres (Naish et al., 2021). We also observed a similar pattern of mCHG methylation in monocentric *J. effusus*. Heterochromatinization of pericentromeres appears to be important for stabilizing the centromeric core, by preventing recombination between core repeats and stopping the spread of CENH3 into adjacent regions (Achrem et al., 2020; Wong et al., 2020). Thus, despite substantial genome restructuring, the epigenetic regulation of centromere units in beak-sedges resembles that in monocentric centromeres. This observation suggests an evolutionarily conserved epigenetic regulation of repeat-based centromeres in both mono- and holocentric organisms. We observed active genes close to and even within centromere units, which, although rare, is likely only possible with a plastic regulation of euchromatic and heterochromatic boundaries. We hypothesize that *R. pubera* achieves such a feat with a fine-scale epigenetic regulation of centromere units (Figures 7A and 7B).

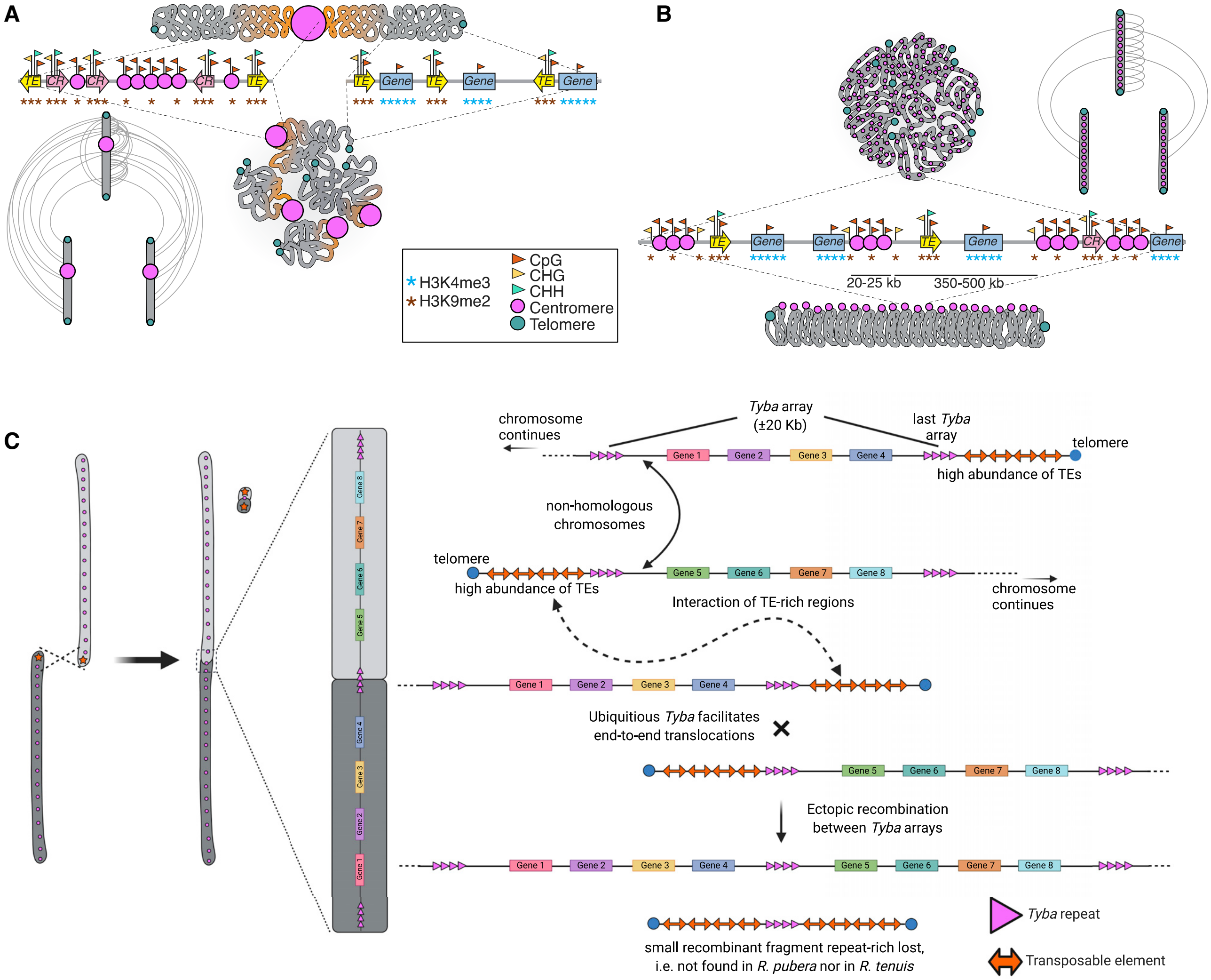


Figure 7. Genome organization in monocentric versus holocentric chromosomes and proposed model for end-to-end fusions

(A) Typically, in a monocentric chromosome, compartments of more compacted and silenced chromatin states extend along large megabase-long regions around centromeres and pericentromeres, while genes concentrate at subtelomeric regions. A telomere-to-centromere axis is frequently observed in genome contact maps in monocentric species due to the clustering of centromeres and telomeres, which increases the rate of interchromosomal chromatin contacts.

(B) *Rhynchospora* holocentric genome revealed uniform deposition of epigenetic marks at the macro scale and fine epigenetic regulation of repeat-based centromere units and silenced and active chromatin states at the micro scale. The regular spacing between centromere units (350–500 kb) appears to be the distance necessary to loop the chromatin back, aligning centromere units (20–25 kb) at the outer surface of the condensed chromosome. A telomere-to-centromere axis is absent in genome contact maps in holocentric species due to the lack of centromere clustering, affecting the spatial genome organization and decreases the rate of interchromosomal chromatin contacts. The model represents intra-/interchromosomal contacts among three different monocentric (A, bottom left) and holocentric (B, upper right) chromosomes.

(C) Possible mechanism for the involvement of centromeric *Tyba* repeats in end-to-end fusions (EEFs). Interaction of highly repetitive regions close to the telomere could facilitate ectopic recombinations of *Tyba* repeats.

Centromere units are regularly spaced (350–500 kb) in the *Rhynchospora* genomes, instead of randomly distributed. This specific spacing might point to a selection mechanism for establishing centromere units separated by an optimal spacing required to fold the chromatin during cell-cycle progression and for the recruitment of CENH3-positive nucleosomes to build the line-like holocentromere at metaphase (Figure 7B). *In silico* modeling based on polymer simulations of chromatin folding in holocentric chromosomes suggests that centromere units can act as anchors of loop extruders, facilitating the formation of line-like holocentromeres during chromosome condensation (Câmara et al., 2021).

A mechanism for the formation of repeat-based holocentromeres

The repeat-based holocentromeres of the *Rhynchospora* species analyzed here are almost exclusively composed of *Tyba* repeats. We cannot conclude from the available data whether the accumulation of such repeats triggered the transition to holocentricity or whether CENH3 spreading preceded and/or also facilitated the subsequent expansion of holocentromeric repeats. However, we did demonstrate that a portion of *Tyba* arrays in *R. pubera* emerged in the genome as a result of the amplification of *TCR1*-type *Helitron*s and that most (98.7%) full-length *TCR1* elements possessing *Tyba* are associated with CENH3-bound chromatin. This either indicates that centromere units are at least partially determined genetically by the nucleotide sequence of *Tyba* or that *TCR1* transposition involves the transfer of epigenetic centromere marks, e.g., CENH3, which remain associated with the new copy of the element. The presence of tandem repeats within *Helitron*s is common in both plants and animals (Thomas and Pritham, 2015), but *TCR1* is unique because it possesses a centromeric satellite. The lack of *Tyba* repeats in *TCR1*-related elements in *R. breviuscula* and *R. tenuis* suggests that *Tyba* was captured by *TCR1* after the ancestors of *R. pubera* and the two other *Rhynchospora* species diverged.

It is conceivable, however, that the amplification and dispersal of *Tyba* occurred via mobilization by TEs earlier in the evolution of *Rhynchospora* and that the signatures of such events have long since been lost due to the accumulation of mutations, insertions and deletions, and DNA rearrangements. We also observed such changes in many *TCR1* loci identified in the genome of *R. pubera* that contained either truncated *TCR1* elements or full-length elements with nested insertions of other sequences (Figures 4K and 4L). The existence of a single *TCR1*/*Helitron27*-related element in *R. breviuscula* that possesses *Tyba* but lacks overall similarity to *TCR1* suggests that *Tyba* capture by *Helitron*s occurs recurrently in the evolution of *Rhynchospora* species and may result in waves of *Tyba* amplification via *Helitron* transposition.

The effect of holocentricity on karyotype evolution and diploidization

Our results are consistent with dysploidy as the main driver of karyotype evolution in holocentric organisms (Guerra, 2016; Mayrose and Lysak, 2021), where strong descending dysploidy restored the ACN ($x = 5$) in *R. pubera* and reduced the chromosome number in *R. tenuis*. In both cases, the same ancestral chromosomes were fused independently either without (*R. tenuis*) or following WGDs (*R. pubera*). Such tolerance of extensive chromosomal rearrangements seems to underlie rapid karyotype evolution, eventually leading to chromosomal speciation (Lucek et al., 2022; Lukhtanov et al., 2018).

Robertsonian translocations and chromosome fusions leading to descending dysploidy have been reported in some holocentric butterflies (Cicconardi et al., 2021; Hill et al., 2019). However, the incidence of EEFs as the sole mechanism of descending dysploidy in *Rhynchospora* is intriguing. Remarkably, meiotic pairing and segregation are not disturbed in the *R. pubera* genome (Marques et al., 2016), suggesting that selection has produced a balanced set of fewer chromosomes. Since *R. pubera* underwent two rounds of WGD, descending dysploidy by EEFs would be a way to effectively create chromosomes with different combinations of ancestral syntenic blocks, reducing the risk of meiotic multivalent pairing without the need of rapid genome downsizing. EEFs in genomes with monocentric chromosomes are normally associated with the formation of typically unstable dicentric chromosomes but may represent a way for chromosomal rearrangements when coupled with concurrent centromere elimination as part of structural diploidization after WGDs (Mandáková et al., 2010; Mandáková and Lysak, 2018; Murat et al., 2010). We argue that the prevalence of EEFs observed in *R. pubera* was facilitated by holocentricity, avoiding the deleterious effect of two centromeres after EEFs in monocentric species and likely promoting rapid structural diploidization.

In *Rhynchospora*, homologous non-sister chromatids are linked by terminal chromatin threads during inverted meiosis (Cabral et al., 2014). EEFs may occur with high(er) frequency in scenarios where chromatids of nonhomologous chromosomes are erroneously connected via repeat-based chromatin threads. However, this notion does not exclude the possibility of EEFs occurring during interphase or mitosis. It is tempting to speculate that the repeat-rich regions observed at chromosome ends in *R. breviuscula* are involved in the formation of chromatin threads, which may act as substrates for ectopic recombination. *Tyba* repeats near these repeat-rich regions may be preferentially used as the site for recombination and may thus facilitate the occurrence of EEFs (Figure 7C). Alternatively, the recruitment of *Tyba* repeats as DNA templates to seal double-stranded breaks involved in EEFs may explain their pronounced association with EEFs (Vu et al., 2017).

Limitations of the study

The three *Rhynchospora* species analyzed in this study are characterized by repeat-based holocentromeres associated with *Tyba* repeats. However, some *Rhynchospora* species lack *Tyba* repeats (Ribeiro et al., 2017); thus, it is not clear whether repeat-based holocentromeres evolved in all species of the genus. Extending our approach to other holocentric species lacking *Tyba*-like repeats will certainly reveal new insights into the evolution of repeat-based holocentromeres. In addition, the presence of holocentric chromosomes in multiple genera of sedges as well as in closely related rushes (e.g., *Luzula* species), but not in *Juncus*, suggests that the transition to holocentricity occurred a long time ago (>60 Mya), which makes temporal tracking challenging. Indeed, our analyses of orthogroups did not identify a clear pattern related to different centromere types.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Plant material
- METHOD DETAILS
  - Genome size measurement by flow cytometry
  - Library preparations and sequencing
  - Genome size estimation using k-mer frequency
  - Sequence-based ploidy assessment

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cell.2022.06.045.

ACKNOWLEDGMENTS

We thank Martin Mascher, Raphael Mercier, Korbinian Schneeberger, Martin Lysak, and Ingo Schubert for their comments on the manuscript. We thank Saurabh Pophaly for assistance with NCBI sequence submission. We thank Neysan Donnelly for extensive manuscript editing. We thank Kristin Krause and Hequan Sun for support with ChIP experiments and k-mer frequency analysis, respectively. We acknowledge the excellent technical assistance of Zdeňka Dubská, Jitka Weiserová, and Eva Jahnová in preparing HMW DNA for optical mapping. This work was supported by the Max Planck Society, Germany, a PhD fellowship awarded to G.T. from the DAAD/India, and grants awarded to A.M. (DFG MA 9363/2-1), A.H., A.P.-H., A.M. (CAPES/DAAD PROBRAL 88881.144086/2017-01), and J.M. (GACR 20-24252S). We also thank the Elixir CZ Research Infrastructure Project (LM2018131), Czech Republic, for providing computing and data-storage facilities. This manuscript is dedicated to Prof. Marcelo Guerra for his long-standing work on beak-sedges cytogenetics, which paved the way for this work.

AUTHOR CONTRIBUTIONS

Conceptualization, A.M.; funding and resources, A.P.-H., J.M., K.F.X.M., A.H., and A.M.; data production, P.G.H., G.T., B.H., H.T., H.Š., and A.M.; formal analyses, investigation, and visualization, P.G.H., G.T., T.L., P. Neumann, T.V., P. Novak, M.Z., L.C., M.C., A.S., H.T., J.F., Y.M.-S., Y.D., A.L.L.V., C.C.S.A., G.S., and A.M.; writing – original draft, P.G.H., G.T., and A.M.; writing – review & editing: all co-authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

One or more of the authors of this paper self-identifies as an underrepresented ethnic minority in science. One or more of the authors of this paper self-identifies as a member of the LGBTQ+ community.

Received: January 3, 2022
Revised: May 24, 2022
Accepted: June 24, 2022
Published: August 4, 2022

REFERENCES

Achrem, M., Szućko, I., and Kalinka, A. (2020). The epigenetic regulation of centromeres and telomeres in plants and animals. Comp. Cytogenet. *14*, 265–311. https://doi.org/10.3897/CompCytogen.v14i2.51895.

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

Bremer, K. (2002). Gondwanan evolution of the grass alliance of families (Poales). Evolution *56*, 1374–1387. https://doi.org/10.1111/j.0014-3820.2002.tb01451.x.

Burchardt, P., Buddenhagen, C.E., Gaeta, M.L., Souza, M.D., Marques, A., and Vanzela, A.L.L. (2020). Holocentric karyotype evolution in Rhynchospora is marked by intense numerical, structural, and genome size changes. Front. Plant Sci. *11*, 536507. https://doi.org/10.3389/fpls.2020.536507.

Cabral, G., Marques, A., Schubert, V., et al. (2014). Chiasmatic and achiasmatic inverted meiosis of plants with holocentric chromosomes. Nat. Commun. *5*, 5070. https://doi.org/10.1038/ncomms6070.

Câmara, A.S., Schubert, V., Mascher, M., and Houben, A. (2021). A simple model explains the cell cycle-dependent assembly of centromeric nucleosomes in holocentric species. Nucleic Acids Res. *49*, 9053–9065. https://doi.org/10.1093/nar/gkab648.

Can, M., Wei, W., Zi, H., Bai, M., Liu, Y., Gao, D., Tu, D., Bao, Y., Wang, L., Chen, S., et al. (2020). Genome sequence of Kobresia littledalei, the first chromosome-level genome in the family Cyperaceae. Sci. Data *7*, 175. https://doi.org/10.1038/s41597-020-0518-3.

Capella-Gutiérrez, S., Silla-Martínez, J.M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics *25*, 1972–1973. https://doi.org/10.1093/bioinformatics/btp348.

Castiglione, M.R., and Cremonini, R. (2012). A fascinating island: 2n=4. Plant Biosyst *146*, 711–726. https://doi.org/10.1080/11263504.2012.714806.

Cheng, H.Y., Concepcion, G.T., Feng, X.W., Zhang, H.W., and Li, H. (2021). Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nat. Methods *18*, 170–175. https://doi.org/10.1038/s41592-020-01056-5.

Cicconardi, F., Lewis, J.J., Martin, S.H., Reed, R.D., Danko, C.G., and Montgomery, S.H. (2021). Chromosome fusion affects genetic diversity and evolutionary turnover of functional loci but consistently depends on chromosome size. Mol. Biol. Evol. *38*, 4449–4462. https://doi.org/10.1093/molbev/msab185.

Conway, J.R., Lex, A., and Gehlenborg, N. (2017). UpSetR: an R package for the visualization of intersecting sets and their properties. Bioinformatics *33*, 2938–2940. https://doi.org/10.1093/bioinformatics/btx364.

Cortes-Silva, N., Ulmer, J., Kiuchi, T., Hsieh, E., Cornilleau, G., Ladid, I., Dingli, F., Loew, D., Katsuma, S., and Drinnenberg, I.A. (2020). CenH3-independent kinetochore assembly in Lepidoptera requires CCAN, including CENP-T. Curr. Biol. *30*, 561–572.e10. https://doi.org/10.1016/j.cub.2019.12.014.

Costa, L., Marques, A., Buddenhagen, C., Thomas, W.W., Huettel, B., Schubert, V., Dodsworth, S., Houben, A., Souza, G., and Pedrosa-Harand, A. (2021). Aiming off the target: recycling target capture sequencing reads for investigating repetitive DNA. Ann. Bot. *128*, 835–848. https://doi.org/10.1093/aob/mcab063.

Despot-Slade, E., Mravinac, B., Širca, S., Castagnone-Sereno, P., Plohl, M., and Meštrović, N. (2021). The Centromere histone Is Conserved and Associated with Tandem Repeats Sharing a Conserved 19-bp Box in the Holocentromere of meloidogyne Nematodes. Mol. Biol. Evol. *38*, 1943–1965. https://doi.org/10.1093/molbev/msaa336.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21. https://doi.org/10.1093/bioinformatics/bts635.

Dolezel, J., Bartos, J., Voglmayr, H., and Greilhuber, J. (2003). Nuclear DNA content and genome size of trout and human. Cytometry A *51*, 127–128. https://doi.org/10.1002/cyto.a.10013.

Drummond, A.J., and Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. *7*, 214. https://doi.org/10.1186/1471-2148-7-214.

Durand, N.C., Robinson, J.T., Shamim, M.S., Machol, I., Mesirov, J.P., Lander, E.S., and Aiden, E.L. (2016). Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. Cell Syst. *3*, 99–101. https://doi.org/10.1016/j.cels.2015.07.012.

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

Escudero, M., Márquez-Corro, J.I., and Hipp, A.L. (2016). The phylogenetic origins and evolutionary history of holocentric chromosomes. Syst. Bot. *41*, 580–585. https://doi.org/10.1600/036364416X692442.

Feng, S., Zhong, Z., Wang, M., and Jacobsen, S.E. (2020). Efficient and accurate determination of genome-wide DNA methylation patterns in Arabidopsis thaliana with enzymatic methyl sequencing. Epigenetics Chromatin *13*, 42. https://doi.org/10.1186/s13072-020-00361-9.

Fernandes, J.B., Wlodzimierz, P., and Henderson, I.R. (2019). Meiotic recombination within plant centromeres. Curr. Opin. Plant Biol. *48*, 26–35. https://doi.org/10.1016/j.pbi.2019.02.008.

Fuchs, J., Demidov, D., Houben, A., and Schubert, I. (2006). Chromosomal histone modification patterns–from conservation to diversity. Trends Plant Sci. *11*, 199–208. https://doi.org/10.1016/j.tplants.2006.02.008.

Gernhard, T. (2008). The conditioned reconstructed process. J. Theor. Biol. *253*, 769–778. https://doi.org/10.1016/j.jtbi.2008.04.005.

Ghosh, S., and Chan, C.K. (2016). Analysis of RNA-seq data using TopHat and cufflinks. Methods Mol. Biol. *1374*, 339–361. https://doi.org/10.1007/978-1-4939-3167-5_18.

Ghurye, J., Rhie, A., Walenz, B.P., Schmitt, A., Selvaraj, S., Pop, M., Phillippy, A.M., and Koren, S. (2019). Integrating Hi-C links with assembly graphs for chromosome-scale assembly. PLoS Comput. Biol. *15*, e1007273. https://doi.org/10.1371/journal.pcbi.1007273.

Gohard, F.H., Zhiteneva, A.A., and Earnshaw, W.C. (2014). Centromeres (eLS). Chichester. https://onlinelibrary.wiley.com/doi/epdf/10.1002/9780470015902.a0005785.pub2.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q.D., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. *29*, 644–652. U130. https://doi.org/10.1038/nbt.1883.

Gremme, G., Brendel, V., Sparks, M.E., and Kurtz, S. (2005). Engineering a software tool for gene structure prediction in higher organisms. Inf. Software Technol. *47*, 965–978. https://doi.org/10.1016/j.infsof.2005.09.005.

Guerra, M. (2016). Agmatoploidy and symploidy: a critical review. Genet. Mol. Biol. *39*, 492–496. https://doi.org/10.1590/1678-4685-GMB-2016-0103.

Guerra, M., Ribeiro, T., and Felix, L.P. (2019). Monocentric chromosomes in Juncus (Juncaceae) and implications for the chromosome evolution of the family. Bot. J. Linn. Soc. *191*, 475–483. https://doi.org/10.1093/botlinnean/boz065.

Haas, B.J., Delcher, A.L., Mount, S.M., Wortman, J.R., Smith, R.K., Jr., Hannick, L.I., Maiti, R., Ronning, C.M., Rusch, D.B., Town, C.D., et al. (2003). Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. Nucleic Acids Res. *31*, 5654–5666. https://doi.org/10.1093/nar/gkg770.

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 Biol. *9*, R7. https://doi.org/10.1186/gb-2008-9-1-r7.

Hao, Z., Lv, D., Ge, Y., Shi, J., Weijers, D., Yu, G., and Chen, J. (2020). RIdeogram: drawing SVG graphics to visualize and map genome-wide data on the idiograms. PeerJ Comput. Sci. *6*, e251. https://doi.org/10.7717/peerj-cs.251.

Haug-Baltzell, A., Stephens, S.A., Davey, S., Scheidegger, C.E., and Lyons, E. (2017). SynMap2 and SynMap3D: web-based whole-genome synteny browsers. Bioinformatics *33*, 2197–2198. https://doi.org/10.1093/bioinformatics/btx144.

Heckmann, S., Macas, J., Kumke, K., Fuchs, J., Schubert, V., Ma, L., Novák, P., Neumann, P., Taudien, S., Platzer, M., et al. (2013). The holocentric species *Luzula elegans* shows interplay between centromere and large-scale genome organization. Plant J. *73*, 555–565. https://doi.org/10.1111/tpj.12054.

Hill, J., Rastas, P., Hornett, E.A., Neethiraj, R., Clark, N., Morehouse, N., de la Paz Celorio-Mancera, M., Cols, J.C., Dircksen, H., Meslin, C., et al. (2019). Unprecedented reorganization of holocentric chromosomes provides insights into the enigma of lepidopteran chromosome evolution. Sci. Adv. *5*, eaau3648. https://doi.org/10.1126/sciadv.aau3648.

Hoencamp, C., Dudchenko, O., Elbatsh, A.M.O., Brahmachari, S., Raaijmakers, J.A., van Schaik, T., Sedeño Cacciatore, Á., Contessoto, V.G., van Heesbeen, R.G.H.P., van den Broek, B., et al. (2021). 3D genomics across the tree of life reveals condensin II as a determinant of architecture type. Science *372*, 984–989. https://doi.org/10.1126/science.abe2218.

Hoff, K.J., and Stanke, M. (2019). Predicting genes in single genomes with AUGUSTUS. Curr. Protoc. Bioinformatics *65*, e57. https://doi.org/10.1002/cpbi.57.

Kasinathan, S., and Henikoff, S. (2018). Non-B-form DNA is enriched at centromeres. Mol. Biol. Evol. *35*, 949–962. https://doi.org/10.1093/molbev/msy010.

Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. *30*, 772–780. https://doi.org/10.1093/molbev/mst010.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., et al. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics *28*, 1647–1649. https://doi.org/10.1093/bioinformatics/bts199.

Kovaka, S., Zimin, A.V., Pertea, G.M., Razaghi, R., Salzberg, S.L., and Pertea, M. (2019). Transcriptome assembly from long-read RNA-seq alignments with StringTie2. Genome Biol. *20*, 278. https://doi.org/10.1186/s13059-019-1910-1.

Krueger, F., and Andrews, S.R. (2011). Bismark: a flexible aligner and methylation caller for bisulfite-Seq applications. Bioinformatics *27*, 1571–1572. https://doi.org/10.1093/bioinformatics/btr167.

Krumsiek, J., Arnold, R., and Rattei, T. (2007). Gepard: a rapid and sensitive tool for creating dotplots on genome scale. Bioinformatics *23*, 1026–1028. https://doi.org/10.1093/bioinformatics/btm039.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., and Marra, M.A. (2009). Circos: an information aesthetic for comparative genomics. Genome Res. *19*, 1639–1645. https://doi.org/10.1101/gr.092759.109.

Kurtz, S., Narechania, A., Stein, J.C., and Ware, D. (2008). A new method to compute K-mer frequencies and its application to annotate large repetitive plant genomes. BMC Genomics *9*, 517. https://doi.org/10.1186/1471-2164-9-517.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359. https://doi.org/10.1038/nmeth.1923.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

Lonnig, W.E., and Saedler, H. (2002). Chromosome rearrangements and transposable elements. Annu. Rev. Genet. *36*, 389–410. https://doi.org/10.1146/annurev.genet.36.040202.092802.

Lopez-Delisle, L., Rabbani, L., Wolff, J., Bhardwaj, V., Backofen, R., Grüning, B., Ramírez, F., and Manke, T. (2021). pyGenomeTracks: reproducible plots for multivariate genomic datasets. Bioinformatics *37*, 422–423. https://doi.org/10.1093/bioinformatics/btaa692.

Lucek, K., Augustijnen, H., and Escudero, M. (2022). A holocentric twist to chromosomal speciation? Trends Ecol. Evol. https://doi.org/10.1016/j.tree.2022.04.002.

Lukhtanov, V.A., Dincă, V., Friberg, M., Šíchová, J., Olofsson, M., Vila, R., Marec, F., and Wiklund, C. (2018). Versatility of multivalent orientation, inverted meiosis, and rescued fitness in holocentric chromosomal hybrids. Proc. Natl. Acad. Sci. USA *115*, E9610–E9619. https://doi.org/10.1073/pnas.1802610115.

Mandáková, T., Joly, S., Krzywinski, M., Mummenhoff, K., and Lysak, M.A. (2010). Fast diploidization in close mesopolyploid relatives of Arabidopsis. Plant Cell *22*, 2277–2290. https://doi.org/10.1105/tpc.110.074526.

Mandáková, T., and Lysak, M.A. (2018). Post-polyploid diploidization and diversification through dysploid changes. Curr. Opin. Plant Biol. *42*, 55–65. https://doi.org/10.1016/j.pbi.2018.03.001.

Mandrioli, M., and Manicardi, G.C. (2020). Holocentric chromosomes. PLOS Genet. *16*, e1008918. https://doi.org/10.1371/journal.pgen.1008918.

Marçais, G., and Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics *27*, 764–770. https://doi.org/10.1093/bioinformatics/btr011.

Marques, A., Ribeiro, T., Neumann, P., Macas, J., Novák, P., Schubert, V., Pellino, M., Fuchs, J., Ma, W., Kuhlmann, M., et al. (2015). Holocentromeres in *Rhynchospora* are associated with genome-wide centromere-specific repeat arrays interspersed among euchromatin. Proc. Natl. Acad. Sci. USA *112*, 13633–13638. https://doi.org/10.1073/pnas.1512255112.

Marques, A., Schubert, V., Houben, A., and Pedrosa-Harand, A. (2016). Restructuring of holocentric centromeres During meiosis in the plant Rhynchospora pubera. Genetics *204*, 555–568. https://doi.org/10.1534/genetics.116.191213.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal *17*, 10–12. https://doi.org/10.14806/ej.17.1.200.

Mayrose, I., and Lysak, M.A. (2021). The evolution of chromosome numbers: mechanistic models and experimental approaches. Genome Biol. Evol. *13*. https://doi.org/10.1093/gbe/evaa220.

Melters, D.P., Paliulis, L.V., Korf, I.F., and Chan, S.W. (2012). Holocentric chromosomes: convergent evolution, meiotic adaptations, and genomic analysis. Chromosome Res. *20*, 579–593. https://doi.org/10.1007/s10577-012-9292-1.

Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., and Lanfear, R. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol. *37*, 1530–1534. https://doi.org/10.1093/molbev/msaa015.

Mizuno, H., Kawahara, Y., Wu, J., Katayose, Y., Kanamori, H., Ikawa, H., Itoh, T., Sasaki, T., and Matsumoto, T. (2011). Asymmetric distribution of gene expression in the centromeric region of rice chromosome 5. Front. Plant Sci. *2*, 16. https://doi.org/10.3389/fpls.2011.00016.

Muller, H., Gil, J., Jr., and Drinnenberg, I.A. (2019). The impact of centromeres on spatial genome architecture. Trends Genet. *35*, 565–578. https://doi.org/10.1016/j.tig.2019.05.003.

Murat, F., Xu, J.H., Tannier, E., Abrouk, M., Guilhot, N., Pont, C., Messing, J., and Salse, J. (2010). Ancestral grass karyotype reconstruction unravels new mechanisms of genome shuffling as a source of plant evolution. Genome Res. *20*, 1545–1557. https://doi.org/10.1101/gr.109744.110.

Nachtweide, S., and Stanke, M. (2019). Multi-genome annotation with AUGUSTUS. Methods Mol. Biol. *1962*, 139–160. https://doi.org/10.1007/978-1-4939-9173-0_8.

Naish, M., Alonge, M., Wlodzimierz, P., Tock, A.J., Abramson, B.W., Schmücker, A., Mandáková, T., Jamge, B., Lambing, C., Kuo, P., et al. (2021). The genetic and epigenetic landscape of the Arabidopsis centromeres. Science *374*, eabi7489. https://doi.org/10.1126/science.abi7489.

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. https://doi.org/10.1186/s13100-018-0144-1.

Nhim, S., Gimenez, S., Nait-Saidi, R., Severac, D., Nam, K., d'Alençon, E., and Nègre, N. (2022). H3K9me2 genome-wide distribution in the holocentric insect Spodoptera frugiperda (Lepidoptera: Noctuidae). Genomics *114*, 384–397. https://doi.org/10.1016/j.ygeno.2021.12.014.

Novák, P., Neumann, P., and Macas, J. (2020). Global analysis of repetitive DNA from unassembled sequence reads using RepeatExplorer2. Nat. Protoc. *15*, 3745–3776. https://doi.org/10.1038/s41596-020-0400-y.

Nurk, S., Walenz, B.P., Rhie, A., Vollger, M.R., Logsdon, G.A., Grothe, R., Miga, K.H., Eichler, E.E., Phillippy, A.M., and Koren, S. (2020). HiCanu: accurate assembly of segmental duplications, satellites, and allelic variants from high-fidelity long reads. Genome Res *30*, 1291–1305. https://doi.org/10.1101/gr.263566.120.

Paten, B., Earl, D., Nguyen, N., Diekhans, M., Zerbino, D., and Haussler, D. (2011). Cactus: algorithms for genome multiple sequence alignment. Genome Res. *21*, 1512–1528. https://doi.org/10.1101/gr.123356.111.

Pedrosa, A., Sandal, N., Stougaard, J., Schweizer, D., and Bachmair, A. (2002). Chromosomal map of the model legume Lotus japonicus. Genetics *161*, 1661–1672. https://doi.org/10.1093/genetics/161.4.1661.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842. https://doi.org/10.1093/bioinformatics/btq033.

Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. *44*, W160–W165. https://doi.org/10.1093/nar/gkw257.

Ranallo-Benavidez, T.R., Jaron, K.S., and Schatz, M.C. (2020). GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. Nat. Commun. *11*, 1432. https://doi.org/10.1038/s41467-020-14998-3.

Reimer, J.J., and Turck, F. (2010). Genome-wide mapping of protein-DNA interaction by chromatin immunoprecipitation and DNA microarray hybridization (ChIP-chip). Part A: ChIP-chip molecular methods. Methods Mol. Biol. *631*, 139–160. https://doi.org/10.1007/978-1-60761-646-7_12.

Ribeiro, T., Buddenhagen, C.E., Thomas, W.W., Souza, G., and Pedrosa-Harand, A. (2018). Are holocentrics doomed to change? Limited chromosome number variation in Rhynchospora Vahl (Cyperaceae). Protoplasma *255*, 263–272. https://doi.org/10.1007/s00709-017-1154-4.

Ribeiro, T., Marques, A., Novák, P., Schubert, V., Vanzela, A.L., Macas, J., Houben, A., and Pedrosa-Harand, A. (2017). Centromeric and non-centromeric satellite DNA organisation differs in holocentric Rhynchospora species. Chromosoma *126*, 325–335. https://doi.org/10.1007/s00412-016-0616-3.

Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: the European molecular biology open software suite. Trends Genet. *16*, 276–277. https://doi.org/10.1016/s0168-9525(00)02024-2.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. *29*, 24–26. https://doi.org/10.1038/nbt.1754.

Schotanus, K., Yadav, V., and Heitman, J. (2021). Epigenetic dynamics of centromeres and neocentromeres in Cryptococcus deuterogattii. PLoS Genet. *17*, e1009743. https://doi.org/10.1371/journal.pgen.1009743.

Schubert, I., and Lysak, M.A. (2011). Interpretation of karyotype evolution should consider chromosome structural constraints. Trends Genet. *27*, 207–216. https://doi.org/10.1016/j.tig.2011.03.004.

Senaratne, A.P., Muller, H., Fryer, K.A., Kawamoto, M., Katsuma, S., and Drinnenberg, I.A. (2021). Formation of the CenH3-deficient holocentromere in

Lepidoptera avoids active chromatin. Curr. Biol. *31*, 173–181.e7. https://doi.org/10.1016/j.cub.2020.09.078.

Seppey, M., Manni, M., and Zdobnov, E.M. (2019). BUSCO: assessing genome assembly and annotation completeness. Methods Mol. Biol. *1962*, 227–245. https://doi.org/10.1007/978-1-4939-9173-0_14.

Šimková, H., Číhalíková, J., Vrána, J., Lysák, M.A., and Doležel, J. (2003). Preparation of HMW DNA from plant nuclei and chromosomes isolated from root tips. Biol. Plant. *46*, 369–373. https://doi.org/10.1023/A:1024322001786.

Smith, S., Collinson, M., Rudall, P., and Simpson, D. (2010). The Cretaceous and Palaeogene fossil record of Poales: review and current research. In Diversity, phylogeny, and evolution in monocotyledons. Proceedings of the fourth international conference on the comparative biology of the monocotyledons & the fifth international, O. Seberg, G. Petersen, A. Barfod, and J.I. Davis, eds. (Aarhus University Press), pp. 333–356.

Soderlund, C., Bomhoff, M., and Nelson, W.M. (2011). SyMAP v3.4: a turnkey synteny system with application to plant genomes. Nucleic Acids Res *39*, e68. https://doi.org/10.1093/nar/gkr123.

Soderlund, C., Nelson, W., Shoemaker, A., and Paterson, A. (2006). SyMAP: A system for discovering and viewing syntenic regions of FPC maps. Genome Res. *16*, 1159–1168. https://doi.org/10.1101/gr.5396706.

Steiner, F.A., and Henikoff, S. (2014). Holocentromeres are dispersed point centromeres localized at transcription factor hotspots. eLife *3*, e02025. https://doi.org/10.7554/eLife.02025.

Stovner, E.B., and Sætrom, P. (2019). epic2 efficiently finds diffuse domains in ChIP-seq data. Bioinformatics *35*, 4392–4393. https://doi.org/10.1093/bioinformatics/btz232.

Sun, H.Q., Ding, J., Piednoël, M., and Schneeberger, K. (2018). findGSE: estimating genome size variation within human and Arabidopsis using k-mer frequencies. Bioinformatics *34*, 550–557. https://doi.org/10.1093/bioinformatics/btx637.

Tang, H., Bowers, J.E., Wang, X., Ming, R., Alam, M., and Paterson, A.H. (2008). Synteny and collinearity in plant genomes. Science *320*, 486–488. https://doi.org/10.1126/science.1153917.

Ter-Hovhannisyan, V., Lomsadze, A., Chernoff, Y.O., and Borodovsky, M. (2008). Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. Genome Res. *18*, 1979–1990. https://doi.org/10.1101/gr.081612.108.

Thomas, J., and Pritham, E.J. (2015). Helitrons, the eukaryotic rolling-circle transposable elements. Microbiol. Spectr. *3*. https://doi.org/10.1128/microbiolspec.MDNA3-0049-2014.

Vanzela, A.L.L., Guerra, M., and Luceño, M. (1996). *Rhynchospora tenuis* Link (Cyperaceae), a species with the lowest number of holocentric chromosomes. Cytobios *88*, 219–228.

Vu, G.T.H., Cao, H.X., Fauser, F., Reiss, B., Puchta, H., and Schubert, I. (2017). Endogenous sequence patterns predispose the repair modes of CRISPR/Cas9-induced DNA double-stranded breaks in Arabidopsis thaliana. Plant J. *92*, 57–67. https://doi.org/10.1111/tpj.13634.

Wickham, H. (2016). ggplot2 : elegant graphics for data analysis. In Use R! Imprint (Springer International Publishing).

Wong, C.Y.Y., Lee, B.C.H., and Yuen, K.W.Y. (2020). Epigenetic regulation of centromere function. Cell. Mol. Life Sci. *77*, 2899–2917. https://doi.org/10.1007/s00018-020-03460-8.

Wu, T.D., and Watanabe, C.K. (2005). GMAP: a genomic mapping and alignment program for mRNA and EST sequences. Bioinformatics *21*, 1859–1875. https://doi.org/10.1093/bioinformatics/bti310.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137. https://doi.org/10.1186/gb-2008-9-9-r137.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
polyclonal rabbit anti-RpCENH3	Marques et al., 2015	anti-RpCENH3
Rabbit polyclonal to Histone H3 (tri methyl K4)	abcam	Cat# ab8580; RRID:AB_306649
Mouse monoclonal to Histone H3 (di methyl K9)	abcam	Cat# ab1220; RRID:AB_449854
Recombinant Rabbit IgG, monoclonal	abcam	Cat# ab172730; RRID:AB_2687931
Biological samples		
Rhynchospora pubera	Own greenhouse	N/A
Rhynchospora breviuscula	Own greenhouse	N/A
Rhynchospora tenuis	Own greenhouse	N/A
Juncus effusus var. spiralis	Own greenhouse (commercially acquired)	N/A
Critical commercial assays		
NucleoBond HMW DNA kit	Macherey Nagel	Cat# 740160.2
SMRTbell Express Template Prep Kit 2.0	Pacific Biosciences	Cat# 101-685-400
Dovetail® Omni-C® Kit	Dovetail	Cat# 21005
Ovation Ultralow V2 DNA-Seq library preparation kit	Tecan Genomics	Cat# 0344NB-08
Enzymatic Methyl-seq Kit	NEBNext®	Cat# E7120S
Poly(A) mRNA Magnetic Isolation Module	NEBNext®	Cat# E7490S
TeloPrime Version 2 kit	Lexogen	Cat# 013PF032V0200
Deposited data		
All sequence data	This study	PRJNA784789
Oligonucleotides		
(FAM)TTTAGGG(8)	Sigma-Aldrich	Telomeric FISH (FAM)oligo-labeled probe
(CY3)ATTGGATTATACATGGTAATTACGCATATAAAGTGCAAATAATGCAATTC	Sigma-Aldrich	*Tyba* repeat FISH oligo-labeled probe
(CY3)GCAAACCAAAATTTGTGTTCAATTTTAAAATATTTCTCCAC	Sigma-Aldrich	*Juncus effusus* cenDNA FISH oligo-labeled probe
Software and algorithms		
HiCanu 2.0	Nurk et al., 2020	https://github.com/marbl/canu
Hifiasm 0.16.1 (r375)	Cheng et al., 2021	https://github.com/chhylp123/hifiasm
BUSCO	Seppey et al., 2019	https://gitlab.com/ezlab/busco/
SALSA2	Ghurye et al., 2019	https://github.com/marbl/SALSA
Bowtie2	Langmead and Salzberg, 2012	https://github.com/BenLangmead/bowtie2
BWA	Li and Durbin, 2009	https://github.com/lh3/bwa
STAR (version 2.7.8a)	Dobin et al., 2013	https://github.com/alexdobin/STAR
DeepTools	Ramirez et al., 2016	https://github.com/deeptools/deepTools
PyGenomeTracks	Lopez-Delisle et al., 2021	https://github.com/deeptools/pyGenomeTracks
bedtools	Quinlan and Hall, 2010	https://github.com/arq5x/bedtools2
MCScan	Tang et al., 2008	https://github.com/tanghaibao/mcscan
SyMAP	Soderlund et al., 2011; Soderlund et al., 2006	https://github.com/csoderlund/SyMAP
Rideogram	Hao et al., 2020	https://github.com/TickingClock1992/RIdeogramsy
SynMap2	Haug-Baltzell et al., 2017	https://genomevolution.org/coge/SynMap.pl
FindGSE	Sun et al., 2018	https://github.com/schneebergerlab/findGSEsmudge

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Smudgeplot	Ranallo-Benavidez et al., 2020	https://github.com/KamilSJaron/smudgeplot
Juicer	Durand et al., 2016	https://github.com/aidenlab/juicer
Straw (strawC v0.0.9)	Durand et al., 2016	https://github.com/aidenlab/straw
Augustus (version 3.3.3)	Hoff and Stanke, 2019	https://github.com/Gaius-Augustus/Augustus
EVidenceModeller	Haas et al., 2008	https://github.com/EVidenceModeler
TRINITY (version 2.13.1)	Grabherr et al., 2011	https://github.com/trinityrnaseq/trinityrnaseq
BLAST+ (ncbi-blast-2.3.0+)	Altschul et al., 1990	ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/
circus	Krzywinski et al., 2009	http://circos.ca/
ggplot2	Wickham, 2016	https://github.com/tidyverse/ggplot2
Geneious	Kearse et al., 2012	https://www.geneious.com/
RepeatExplorer2	Novák et al., 2020	https://repeatexplorer-elixir.cerit-sc.cz/galaxy/
jellyfish	Marçais and Kingsford, 2011	https://github.com/gmarcais/Jellyfish
Bismark	Krueger and Andrews, 2011	https://github.com/FelixKrueger/Bismark
MAFFT	Katoh and Standley, 2013	https://github.com/GSLBiotech/mafft
TrimAl	Capella-Gutiérrez et al., 2009	https://github.com/inab/trimal
IQ-tree2	Minh et al., 2020	http://www.iqtree.org/
BEAST v.1.10.4	Drummond and Rambaut, 2007	https://beast.community/
Other		
Assemblies, predicted transcripts and proteins, ChIP and DNA methylation tracks, repeat annotation tracks	This study	https://data.cyverse.org/dav-anon/iplant/home/dabitz66/marquesLabTrackHub/
Assemblies	This study	https://genomevolution.org/coge
Genome Browser	This study	https://genome-euro.ucsc.edu/cgi-bin/hgGateway?genome=rhyPub2m&hubUrl=https://data.cyverse.org/dav-anon/iplant/home/dabitz66/marquesLabTrackHub/hub.txt

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, André Marques (amarques@mpipz.mpg.de).

Materials availability

This study did not generate new unique reagents.

Data and code availability

All sequencing data used in this study have been deposited at NCBI under the Bioproject no. PRJNA784789 and are publicly available as of the date of publication. The reference genomes, annotations and all tracks presented in this work are made available at https://data.cyverse.org/dav-anon/iplant/home/dabitz66/marquesLabTrackHub/, the CoGe platform (https://genomevolution.org/coge) and the following UCSC Genome Browser hosted by CyVerse. All other data needed to evaluate the conclusions in the paper are provided in the paper and/or the supplemental information.

This paper does not report original code.

Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Plant material

Plants from naturally occurring populations of *R. pubera* and *R. tenuis* growing in Curado (Recife), Northeastern Brazil, and *R. breviuscula* growing in Londrina (Paraná state), Southern Brazil were collected in 2013 and further cultivated under controlled greenhouse conditions (16h daylight, 26 °C, >70% humidity). As a monocentric outgroup an individual of the ornamental plant *Juncus effusus* var. *spiralis* was commercially obtained and cultivated under controlled greenhouse conditions (16h daylight, 20°C).

METHOD DETAILS

Genome size measurement by flow cytometry

The genome size of 1C=1.6 Gb for the *R. pubera* accession sequenced here has been previously measured (Marques et al., 2015). Thus, genome size estimations by flow cytometry were performed for the accessions of *R. breviuscula* and *R. tenuis* as well as for *Juncus effusus* var. *spiralis*. For that, roughly 0.5 cm^2 of young leaf tissue was chopped with a sharp razorblade in a Petri dish together with appropriate amounts of leaf tissue of the internal reference standard *Raphanus sativus* cv. Voran (2C=1.11 pg; Genebank Gatersleben, accession number: RA 34) using the 'CyStain PI Absolute P' nuclei extraction and staining kit (Sysmex-Partec). The resulting nuclei suspension was filtered through a 50-μm filter (CellTrics, Sysmex-Partec) and measured on a CyFlow Space flow cytometer (Sysmex-Partec). The absolute DNA content (pg/2C) was calculated based on the values of the G1 peak means and the corresponding genome size (Mb/1C) according to Dolezel et al. (2003).

Library preparations and sequencing

DNA isolation

High-molecular-weight DNA was isolated from 1.5 g of material with a NucleoBond HMW DNA kit (Macherey Nagel). Quality was assessed with a FEMTO-pulse device (Agilent), and quantity was measured with a Quantus fluorometer (Promega).

Whole-genome shotgun sequencing (WGS)

Genomic DNA from *R. breviuscula* and *R. alba* were deep-sequenced with an Illumina HiSeq 3000 in 150-bp paired-end mode. Alternatively, DNBseq short read sequencing (BGI Genomics, Hong Kong) of genomic DNA was performed for *R. pubera*, *R. tenuis*, and *R. tenerrima*. Available WGS short reads from *R. cephalotes* (SRX9381225), *R. ciliata* (Ribeiro et al., 2017), *R exaltata* (SRX9381226), *R. globosa* (Ribeiro et al., 2017), and *C. littledalei* (SRX5833125, SRX5833124) were used.

PacBio

A HiFi library was then prepared according to the "*Procedure & Checklist - Preparing HiFi SMRTbell® Libraries using SMRTbell Express Template Prep Kit 2.0*" manual with an initial DNA fragmentation by g-Tubes (Covaris) and final library size binning into defined fractions by SageELF (Sage Science). Size distribution was again controlled by FEMTO-pulse (Agilent). Size-selected libraries were then sequenced on a Sequel II device with Binding kit 2.0 and Sequel II Sequencing Kit 2.0 for 30 h (*Pacific Biosciences*). The numbers of SMRT cells for each species were as follows: *R. pubera* (3 cells), *R. breviuscula* (1 cell), *R. tenuis* (2 cells), and *J. effusus* (1 cell).

Omni-C

For each species, a single chromatin-capture library was prepared from 0.5 g fresh weight material input. All treatments were according to the recommendations of the kit vendor for plants (Omni-C, Dovetail). As a final step, an Illumina-compatible library was prepared (Dovetail) and paired-end 2 x 150 bp deep-sequenced on a HiSeq 3000 (Illumina) device for *R. breviuscula*, *R. tenuis*, and *J. effusus*. Alternatively, the *R. pubera* library was paired-end 2 x 150 bp deep-sequenced using DNBseq technology (BGI Genomics, Hong Kong).

ChIPseq

ChIP DNA was quality-controlled using the NGS-assay on a FEMTO-pulse (Agilent); then, an Illumina-compatible library was prepared with the Ovation Ultralow V2 DNA-Seq library preparation kit (Tecan Genomics) and single-end 1 x 150-bp sequenced on a HiSeq 3000 (Illumina) device. For each library, an average of 20 millions reads were obtained.

Enzymatic Methyl-seq

To investigate the methylome space in *R. pubera* and *J. effusus*, the relatively non-destructive NEBNext® Enzymatic Methyl-seq Kit was employed to prepare an Illumina-compatible library, followed by paired-end sequencing (2 x 150 bp) on a HiSeq 3000 (Illumina) device. For each library, 10 Gb of reads were generated.

RNAseq

Total RNA from root, leaves, and flower buds was isolated from *R. breviuscula*. For *R. tenuis*, total RNA was isolated from flower buds only. For *J. effusus*, RNAseq data from the NCBI (accession numbers SRX2268676, SRX2268675, and SRX1639021) were used to complement its genome annotation. For *R. pubera*, total RNA was extracted from six different tissues (i.e., roots, young leaves, old leaves, stem, early flower buds, and late flower buds). Poly-A RNA was enriched from 1 μg total RNA using the NEBNext® Poly(A) mRNA Magnetic Isolation Module. RNAseq libraries were prepared as described in the NEBNext Ultra™ II Directional RNA Library Prep Kit for Illumina (New England Biolabs). A total of 11 cycles were applied to enrich library concentration. Sequencing was done at BGI Genomics (Hong Kong) with a BGISEQ-500 system in the DNBseq platform in paired-end mode 2 x 150 bp.

IsoSeq

For the proper annotation of the complex *R. pubera* genome, total RNA was extracted from six different tissues (i.e., roots, young leaves, old leaves, stem, early flower buds and late flower buds) and quality-assessed by a Nanochip (Agilent Bioanalyser, Santa Clara, U.S.A.). Next, cDNA was synthesized according to the TeloPrime Version 2 kit (Lexogen, Vienna, Austria). We exchanged the Lexogen first-strand synthesis oligo-dT primer for the (5'-AAGCAGTGGTATCAACGCAGAGTACT(30)VN-3') primer to introduce a 3' anchor base. Then, the optimal number of cycles was determined by qPCR (Viia7, Applied Biosystems) with the 1x Evagreen fluorochrome (Biotium, Fremont, U.S.A.), TeloPrime kit chemistry and 25% of the cDNA as input. The forward primer was FP from the TeloPrime kit, and the reverse primer was 5’-AAGCAGTGGTATCAACGCAGAGTAC-3’. The residual cDNA was mass-amplified with an extended Lexogen FP primer by adding 16mer barcodes as recommended by PacBio at the 5’ end and a cycle number by which 80% of the maximal fluorescence signal was reached. The PCR products were bead-purified (Pronex beads, Promega) followed by PacBio library preparation with the SMRTbell Express Template Prep Kit 2.0 (Pacific Biosciences, Menlo Park, U.S.A.), and then quantity- (Quantus, Promega) and quality-assessed (Agilent Bioanalyser). Long-read sequencing was performed on a Sequel II sequencer with a Sequel II Binding kit 2.1, Sequel II Sequencing Kit 2.0 sequencing chemistry 2.0, and a single 8M SMRT cell (*Pacific Biosciences*, Menlo Park, U.S.A.). The movie time was 30 h after a 2-h immobilization step and 2-h pre-extension step to adjust for high-fidelity (HiFi) sequencing.

Genome size estimation using k-mer frequency

Genome sizes of the three *Rhynchospora* species and *J. effusus* were also confirmed by k-mer frequency analysis with the findGSE tool (Sun et al., 2018), after counting k-mers with Jellyfish (Marçais and Kingsford, 2011). High-coverage short reads were used as follows: *R. pubera* (60x), *R. breviuscula* (50x), and *R. tenuis* (130x). Since for *J. effusus* we did not have short-read data, we used our high-coverage HiFi PacBio reads (70x).

Sequence-based ploidy assessment

We used Smudgeplot (Ranallo-Benavidez et al., 2020) to visualize and estimate the ploidy and structure of the sequenced genomes. This tool can infer ploidy directly from the k-mers present in sequencing reads by analyzing heterozygous k-mer pairs.

Genome assembly

HiFi reads obtained by the sequencing process were subjected to assembly using the HiCanu function of Canu (Nurk et al., 2020), available at https://github.com/marbl/canu, for *R. pubera* with the following command line:

```
canu -assemble -p output.asm -d run1 genomeSize=1.6g maxThreads=40 useGrid=false -pacbio-hifi *.fastq.
```

Alternatively, Hifiasm (Cheng et al., 2021), available at https://github.com/chhylp123/hifiasm, was used for the assembly of *R. breviuscula*, *R. tenuis*, and *J. effusus* with the following command:

```
hifiasm -o output.asm -t 40 reads.fq.gz.
```

Preliminary assemblies were evaluated for contiguity and completeness with BUSCO (Seppey et al., 2019).

Optical map and hybrid scaffolding

We developed an optical mapping strategy to help resolve the complexity of the *R. pubera* genome. High-molecular-weight DNA was prepared from young leaves of *R. pubera*. A total of 3.15 million cell nuclei were purified by flow cytometry, pelleted by centrifugation (30 min at 300 *g*), and embedded in four agarose plugs of 20-µL volume. The nuclear DNA was purified in the plugs as described by Šimková et al. (2003) with an increased concentration of proteinase K (1 mg/mL of lysis buffer). The proteinase- and RNase-treated DNA was isolated from the agarose gel, and the resulting 525 ng DNA was directly labeled at DLE-1 recognition sites following the standard Bionano Prep Direct Label and Stain (DLS) Protocol (Bionano Genomics, San Diego, USA) and analyzed on the Saphyr platform of Bionano Genomics. A total of 1.27 Tbp of single-molecule data with N50 of 236 kb, corresponding to effective coverage of 96.8x of the *R. pubera* genome, was used in *de novo* assembly by Bionano Solve 3.6.1_11162020, using a standard configuration file ‘‘optArguments_nonhaplotype_noES_noCut_DLE1_saphyr.xml’’ (Table S6). A p-value threshold of 1e–11 was used to build the initial assembly, a p-value of 1e–12 was used for extension and refinement steps (five rounds), and a p-value of 1e–16 was used for final map merging. To improve the contiguity of the sequence assembly, an automatic hybrid scaffold pipeline integrated in Bionano Solve 3.6.1_11162020 was run with the *de novo* optical map assembly. The default *DLE-1 Hybrid Scaffold* configuration file was used with the ‘‘Resolve conflict’’ option for conflict resolution. The conflicts between sequences and the optical map were manually curated, and the pipeline was re-run using the modified *conflict_cut_statu.txt* file (Table S7). The results obtained from the optical mapping scaffolding of the genome assembly of *R. pubera* were used as input for Omni-C scaffolding.

Omni-C scaffolding

Dovetail Omni-C reads were first mapped using BWA (Li and Durbin, 2009) following the hic-pipeline available at https://github.com/esrice/hic-pipeline. Hi-C scaffolding was performed using SALSA2 (Ghurye et al., 2019), available at https://github.com/marbl/SALSA, with default parameters. After testing several minimum mapping quality values of bam alignments, final scaffolding was performed with MAPQ10. Several rounds of assembly correction guided by Hi-C contact maps and manual curation of scaffolds were performed to obtain the pseudomolecules.

Assembly and scaffolding strategy

The rather homozygous genome of *J. effusus* was estimated to be close to 1C=271 Mb (Figure 1; Figure S4C). Sequencing of *J. effusus* var. *spiralis* yielded 19 Gb of reads and an initial assembly of 258 Mb (79× coverage, N_{50} = 11 Mb, Figure 1), where 18 contigs corresponded to complete chromosomes. The assembly was further scaffolded to the expected 21 pseudomolecules (240 Mb), and unplaced contigs contained 18 Mb, corresponding to the complete haploid chromosome set of the species

(Figures 1 and 2A). The sequencing of *R. pubera*, which is an inbred species, yielded 66 Gb of PacBio HiFi reads, and the initial assembly spanned 1.7 Gb (38× coverage, N_{50} = 11.2 Mb). After removing redundant sequences likely due to some small residual heterozygosity, the assembly closely matched its estimated haploid genome size (Figures 1 and S3F). A first scaffolding using optical mapping was followed by a second scaffolding using chromosome conformation capture (Omni-C, Dovetail™) of the genome assembly, which yielded five very large pseudomolecules (1.47 Gb, N_{50} = 361 Mb), while unplaced contigs contained 141 Mb (Figure 1). The sequencing of *R. breviuscula* yielded 30 Gb of PacBio HiFi reads, resulting in an initial assembly that was 813 Mb in length. In contrast to *R. pubera*, *R. breviuscula* is outbred, which resulted in an assembly of its diploid genome size showing a high level of heterozygosity confirmed by k-mer analysis (Figures 1 and S4A). We pruned the resulting large contigs to the single largest representative haplotype (75× coverage, 421 Mb, N_{50} = 11 Mb; Figure 1) and then oriented and ordered it into five pseudomolecules using Omni-C scaffolding comprising 370 Mb (N_{50} = 71 Mb; Figure S1A). Unplaced contigs contained 50 Mb (Figure 1). The sequenced genome of *R. tenuis* yielded 45.9 Gb of PacBio HiFi reads resulting in an assembly of 770 Mb, which closely corresponds to its diploid genome size, showing a high level of heterozygosity (Figures 1 and S4B). We pruned the resulting large contigs to the single largest representative haplotype (120× coverage, 395 Mb, N_{50} = 19 Mb, Figure 1), which was oriented and ordered into two pseudomolecules of about 350 Mb (N_{50} = 215 Mb; Figure S1B). Unplaced contigs contained 47 Mb (Figure 1).

GENERATION OF HI-C MAPS

Final Hi-C maps of *R. pubera*, *R. breviuscula*, *R. tenuis*, and *J. effusus* were generated by Juicer (v1.6) (Durand et al., 2016) using the sequencing data from DNase *in situ* Hi-C (Omni-C) experiments. Specifically, technical replicates were aligned and deduplicated and then the results of each replicate were merged by MEGA from Juicer.

Quantitative analysis of Hi-C contacts

The python version of Straw (strawC v0.0.9) (Durand et al., 2016) was used to extract Hi-C counts from the illustrated Hi-C maps (Figure 2; Figure S1) in 1-Mb resolution and with the normalization approach of Vanilla Coverage (VC). To represent the intra- and interchromosomal interactions in an intuitive manner, the *cis* Hi-C contact of a chromosome was quantified as the sum of all Hi-C counts within the chromosome *per se*, while *trans* Hi-C contacts referred to the sum of Hi-C counts between the designated chromosome and all other chromosomes. The final intra- and interchromosomal contacts for each single chromosome were normalized through the percentages of Hi-C counts over the sum of all Hi-C signals in the corresponding Hi-C map. It is also noteworthy that the infinite extracted Hi-C counts through Straw were replaced by the mean of all other finite counts within the extracted chromosomal pair.

ChIP

ChIP experiments were performed following Reimer and Turck (2010), with adjustments for *R. pubera* and *R. breviuscula*. Unopened flower buds were harvested and frozen in liquid nitrogen until sufficient material was obtained. The samples were fixed in 4% formaldehyde for 30 min and the chromatin was sonicated for 25 min. Then, 7–85 μL of sonicated chromatin was incubated with 2 ng of respective antibody overnight. Immunoprecipitation was carried out for rabbit anti-*Rp*CENH3, for *R. pubera* and *R. breviuscula*, and for rabbit anti-H3K4me3 (abcam, ab8580), and mouse anti-H3K9me2 (abcam, ab1220). Recombinant rabbit IgG (abcam, ab172730) and no-antibody inputs were used as controls. Two experimental replications were also maintained for all the combinations. After overnight incubation of chromatin with antibody, protein beads (anti-mouse: Protein G Sepharose 4 Fast Flow, anti-rabbit: rProtein A Sepharose Fast Flow) were added to the chromatin-antibody mixture. The bound chromatin was finally eluted, de-crosslinked, precipitated, and sent for sequencing.

Synteny and self-synteny analyses

The synteny analysis shown in Figure 5 was performed using the MCscan pipeline implemented in the JVCI utility libraries (Tang et al., 2008). For this analysis, CDS sequences of the longest transcript were used. Circular plots were drawn with the circos package (Krzywinski et al., 2009).

Self-synteny analyses shown in Figures S1D, S1E, and S3A were performed with SyMAP v. 5.0.6 (Soderlund et al., 2006, 2011). Circular self-synteny plots were obtained with SyMAP or RIdeogram software (Hao et al., 2020) using the synteny calculation blocks obtained from SyMAP.

Whole-genome alignment (WGA)

A whole-genome alignment (WGA) between *R. pubera*, *R. tenuis*, *R. breviuscula*, *J. effusus*, and *C. littledalei* was generated using the Cactus pipeline (Version 1.0) (Paten et al., 2011). Prior to the alignment step, all nucleotide sequences were 20-kmer-softmasked to reduce complexity and facilitate construction of the WGA using the tallymer subtools from the genome tools package (Version 1.6.1) (Kurtz et al., 2008). The Cactus pipeline was run stepwise with the default settings described at https://github.com/ComparativeGenomicsToolkit/cactus#running-step-by-step.

Characterization of end-to-end fusions

For the characterization of the regions involved in EEFs observed in *R. pubera* and *R. tenuis*, we first compared the synteny alignment between their genomes with *R. breviuscula* used as a reference in SyMAP. This allowed us to pin the putative regions around the borders of the fusion events. To gain insights about the order of fusion events in the complex genome of *R. pubera* we extracted all duplicated fusion regions and aligned them against themselves in SyMAP (Figures 5C and S5). This detailed analysis further allowed us to reconstruct the order of the fusion events and tracing the karyotype history of *R. pubera* based on the shared EEF signatures found in the genome.

In order, to identify the underlying sequences at the fusion regions, we loaded annotation features for genes, TEs, *Tyba*, rDNA and telomeric repeats on SyMAP aligments. This allowed us to detect the sequence types in the putative translocated regions. In *R. pubera*, we counted 15 potential EEF regions, of which 11 regions had a *Tyba* array right in the middle between two ancestral syntenic chromosomes of *R. breviuscula*. Further inspection and characterization of such regions were done by checking the genome coordinates and annotation features with Geneious (Kearse et al., 2012), which revealed a remnant rDNA cluster involved in the EEF of two ancestral *Rb*3 in the *Rp*Chr3. The other three regions did not show any specific sequence (Figures 5C and S6A–S6G). Similar strategy was used to identify the underlying sequences within the three fusion events in *R. tenuis* chromosomes. In this species, we found a *Tyba* repeat underlying the fusion region between *Rb*3/*Rb*4 in *Rt*Chr2 and a telomeric repeat between *Rb*1/*Rb*5 in *Rt*Chr1, while the third fusion between had no specific sequence (Figures 5C and S6H). Common to most fusion events we found that the very ends of *R. breviuscula* chromosomes, which are enriched for TEs are mostly missing in the fused chromosomes (Figures 5D and S6).

Whole-genome duplication analysis

To identify ancient WGD events, we performed Synonymous (Ks) substitutions analysis on the fully annotated genomes with the SynMap2 (Haug-Baltzell et al., 2017) available at CoGe webportal (https://genomevolution.org/coge/SynMap.pl).

Gene annotation

Structural gene annotation was done combining *de novo* gene calling and homology-based approaches with RNAseq, IsoSeq, and protein datasets.

Using evidence derived from expression data, RNAseq data were first mapped using STAR (Dobin et al., 2013) (version 2.7.8a) and subsequently assembled into transcripts by StringTie (Kovaka et al., 2019) (version 2.1.5, parameters -m 150-t -f 0.3). *Triticeae* protein sequences from available public datasets (UniProt, https://www.uniprot.org, 05/10/2016) were aligned against the genome sequence using GenomeThreader (Gremme et al., 2005) (version 1.7.1; arguments -startcodon -finalstopcodon -species rice -gcmincoverage 70 -prseedlength 7 -prhdist 4). Isoseq datasets were aligned to the genome assembly using GMAP (Wu and Watanabe, 2005) (version 2018-07-04). All transcripts from RNAseq, IsoSeq, and aligned protein sequences were combined using Cuffcompare (Ghosh and Chan, 2016) (version 2.2.1) and subsequently merged with StringTie (version 2.1.5, parameters –merge -m150) into a pool of candidate transcripts. TransDecoder (version 5.5.0; http://transdecoder.github.io) was used to find potential open reading frames and to predict protein sequences within the candidate transcript set.

Ab initio annotation was initially done using Augustus (Hoff and Stanke, 2019) (version 3.3.3). GeneMark (Ter-Hovhannisyan et al., 2008) (version 4.35) was additionally employed to further improve structural gene annotation. To avoid potential over-prediction, we generated guiding hints using the above described RNAseq, protein, and IsoSeq datasets as described by Hoff and Stanke (2019). A specific Augustus model for *Rhynchospora* was built by generating a set of gene models with full support from RNAseq and IsoSeq. Augustus was trained and optimized using the steps detailed by Hoff and Stanke (2019).

To maximize uniformity across all annotated species, Augustus was also run in comparative annotation mode (Nachtweide and Stanke, 2019). The generated WGA served as sequence input together with the mapping of RNAseq data as described above.

All structural gene annotations were joined using EVidenceModeller (Haas et al., 2008) (version 1.1.1), and weights were adjusted according to the input source: *ab initio* (Augustus: 5, GeneMark: 2), homology-based (10), and comparative *ab initio* (7). Additionally, two rounds of PASA (Haas et al., 2003) (version 2.4.1) were run to identify untranslated regions and isoforms using transcripts generated by a genome-guided TRINITY (Grabherr et al., 2011) (version 2.13.1) assembly derived from *Rhynchospora* RNAseq data and the above described IsoSeq datasets.

We used BLASTP (Altschul et al., 1990) (ncbi-blast-2.3.0+, parameters -max_target_seqs 1 -evalue 1e–05) to compare potential protein sequences with a trusted set of reference proteins (Uniprot Magnoliophyta, reviewed/Swissprot, downloaded on 3 Aug 2016; https://www.uniprot.org). This differentiated candidates into complete and valid genes, non-coding transcripts, pseudogenes, and transposable elements. In addition, we used PTREP (Release 19; http://botserv2.uzh.ch/kelldata/trep-db/index.html), a database of hypothetical proteins containing deduced amino acid sequences in which internal frameshifts have been removed in many cases. This step is particularly useful for the identification of divergent transposable elements with no significant similarity at the DNA level. Best hits were selected for each predicted protein from each of the three databases. Only hits with an e-value below 10e–10 were considered. Furthermore, functional annotation of all predicted protein sequences was done using the AHRD pipeline (https://github.com/groupschoof/AHRD).

Proteins were further classified into two confidence classes: high and low. Hits with subject coverage (for protein references) or query coverage (transposon database) above 80% were considered significant and protein sequences were classified as high-confidence using the following criteria: protein sequence was complete and had a subject and query coverage above the threshold in the

UniMag database or no BLAST hit in UniMag but in UniPoa and not PTREP; a low-confidence protein sequence was incomplete and had a hit in the UniMag or UniPoa database but not in PTREP. Alternatively, it had no hit in UniMag, UniPoa, or PTREP, but the protein sequence was complete. In a second refinement step, low-confidence proteins with an AHRD-score of 3* were promoted to high-confidence.

BUSCO (Seppey et al., 2019) (version 5.1.2.) was used to evaluate the gene space completeness of the pseudomolecule assembly and structural gene annotation with the 'viridiplantae_odb10' database containing 425 single-copy genes.

Orthogroup analysis

Orthogroup assignments (Table S4) was performed with OrthoFinder (Emms and Kelly, 2019). For GO term enrichment, a GO annotation file (gaf; 2.1) was built using all GO terms assigned by the functional annotations of *R. pubera*, *R. breviuscula*, *R. tenuis*, and *J. effusus*. GO term enrichment was performed by feeding GO terms of the shared orthologos into Ontologiser (ontologiser.de). *P*-values were corrected using the Benjamini-Hochberg procedure. We used the UpSetR (Conway et al., 2017) package (http://gehlenborglab.org/research/projects/upsetr/) to analyze how many orthogroups are shared between the five species or are unique to a single species.

De novo repeat discovery and annotation

To identify the overall repetitiveness of genomes we performed *de novo* repeat discovery with RepeatExplorer2 (Novák et al., 2020) for nine species of *Rhynchospora*, *C. littledalei*, and *J. effusus*. We used a repeat library obtained from the RepeatExplorer2 analysis of Illumina paired-end reads. All clusters representing at least 0.005% of the genomes were manually checked, and the automated annotation was corrected if needed. Contigs from the annotated clusters were used to build a repeat library. To minimize potential conflicts due to the occasional presence of contaminating sequences in the clusters, only contigs with average read depths $\geq$ 5 were included and all regions in these contigs that had read depths < 5 were masked. Genome assemblies were then annotated using custom RepeatMasker (REF - Smit, AFA, Hubley, R & Green, P. RepeatMasker Open-4.0. 2013-2015 http://www.repeatmasker.org) search with options -xsmall -no_is -e ncbi -nolow. Output from RepeatMasker was parsed using custom scripts (https://github.com/kavonrtep/repeat_annotation_pipeline) to remove overlapping and conflicting annotations.

Transposable element protein domains (Neumann et al., 2019) found in the assembled genomes were annotated using the DANTE tool available from the RepeatExplorer2 Galaxy portal. To find master *Helitron* elements related to *TCR1*, we first searched the genome assembly for *Helitron* helicase-coding sequences using DANTE (https://repeatexplorer-elixir.cerit-sc.cz/galaxy/) exploiting the REXdb database (Neumann et al., 2019) (Viridiplantae_version_3.0) and then manually identified boundaries of full-length *Helitron* elements. We identified 111 putative autonomous *Helitron*s and compared their terminal sequences with *TCR1*. This revealed that *TCR1* is most similar to the *Helitron-27*, sharing 90% and 100% identity over 30-bp sequences at the 5' and 3' ends, respectively (Figures 4J and 4K), meeting the criteria for classification of *TCR1* and *Helitron-27* into the same family (Thomas and Pritham, 2015). To find *TCR1* insertions in the *R. pubera* genome, we performed iterative blastn searches using 30-bp sequences from their 5' and 3' termini and consensus sequences of *Tyba*.

To obtain the average number of *Tyba* arrays for each *Rhynchospora* genome, we first removed spurious low-quality *Tyba* monomer annotations with less than 500 bp and merged with bedtools (Quinlan and Hall, 2010) all adjacent *Tyba* monomers situated at a maximum distance of 50 kb into individual annotations to eliminate the gaps that arise because of fragmented *Tyba* arrays. Length and distance between *Tyba* arrays were then calculated using bedtools. Bar plots of the average distance and unit length used to compare the *Tyba* arrays among the three *Rhynchospora* species were made in RStudio using ggplot2 (Wickham, 2016) library available at https://github.com/tidyverse/ggplot2.

Detection of dyad symmetries in Tyba repeats

Dyad simmetries detection was performed as reported in Kasinathan and Henikoff (2018). We used EMBOSS palindrome (Rice et al., 2000) to detect perfect dyad symmetries in the *Tyba* consensus of the three *Rhynchospora* species with the following parameters:

-minpallen 4 -maxpallen 100 -gaplimit 20 -nummismatches 0 –overlap

ChIP-seq analysis

Raw sequencing reads were trimmed by Cutadapt (Martin, 2011) to remove low-quality nucleotides (with quality score less than 30) and adapters. Trimmed ChIPed 150-bp single-end reads were mapped to the respective reference genome with bowtie2 (Langmead and Salzberg, 2012) with default parameters, where all read duplicates were removed and only the single best matching read was kept on the final alignment BAM file. BAM files were converted into BIGWIG coverage tracks using the bamCompare tool from deeptools (Ramírez et al., 2016). The coverage was calculated as the number of reads per 50-bp bin and normalized by reads per kilobase per million mapped reads (RPKM). Plots of detailed chromosome regions showing multiple tracks presented in Figures 4 and 6 were done with pyGenomeTracks (Lopez-Delisle et al., 2021).

CENH3 domains were identified by comparing the ChIPed and input data using MACS3 (Zhang et al., 2008). The parameters for MACS3 included -B –broad –g 1470000000 –trackline. As an alternative method for detection of CENH3 domains, we compared input and ChIP using the epic2 program for detection of diffuse domains (Stovner and Sætrom, 2019). Parameters for epic2 included—bin-size 2000. Only CENH3 domains detected with both methods were kept for further analysis.

To determine the sizes and positions of centromere units, we merged with bedtools CENH3 peaks that were separated by less than 50 kb to eliminate the gaps that arise because of fragmented *Tyba* arrays or due to insertion of TEs. Small CENH3 domains of less than 1 kb were discarded. Length and distance between *Tyba* arrays and between CENH3 domains were then calculated using bedtools. Bar plots of the average distance and unit length used to compare CENH3 domains and *Tyba* arrays were made in RStudio using the ggplot library.

The obtained repeat annotation was used to evaluate the association of individual classes of repetitive sequences with the CENH3 domain in *R. pubera*. For each repeat type, we calculated the total abundance in the genome as a sum of repetitive element length and compared it with abundance of repetitive elements located within CENH3 domains. For each type of repetitive element, we calculated the observed/expected ratio using:

$$OE = \frac{\sum(R_{CENH3})}{\sum\left(\frac{L_{CENH3}}{L_G}\right)R_G}$$

where R_{CENH3} is length of repeat located within CENH3 domains, L_{CENH3} is the length of CENH3-binding regions, L_G is total genome size, and R_G is total length of repeat type in the genome.

Identification of paralogous CENH3 domains

To identify groups of paralogous CENH3 domains within the blocks of homologous regions of *R. pubera*, we identified the two nearest paralogous genes on both sides of each CENH3 domain. Subsequently, the groups of four genes surrounding CENH3 domains were used to identify corresponding regions on the other homologous blocks where we checked for the presence of the CENH3 domain. Resulting groups of four homologous regions were manually inspected using dotplot (Krumsiek et al., 2007) and the IGV browser (Robinson et al., 2011).

Methyl-seq analysis

To comparatively evaluate the DNA methylation context of a holocentric and monocentric genome, we applied enzymatic methyl-seq and used the Bismarck pipeline (Krueger and Andrews, 2011) to analyze the data using the standard pipeline described at https://rawgit.com/FelixKrueger/Bismark/master/Docs/Bismark_User_Guide.html. Individual methylation context files for CpG, CHG, and CHH were converted to BIGWIG format and used as input track for overall genome-wide DNA methylation visualization with pyGenomeTracks.

Metaplots

Analysis of the enrichment of all ChIP treatment files was performed as follows: BAM files of each ChIP treatment were normalized to the ChIP Input BAM file by RPKM using bamCompare available from deeptools. The generated normalized BIGWIG files were used to calculate the level of enrichment associated with gene bodies, *Tyba* repeats, CENH3 domains, and TEs using computeMatrix scale-regions (parameters: –regionBodyLength 4000 –beforeRegionStartLength 2000 –afterRegionStartLength 2000) also available from deeptools. Finally, metaplots for all ChIPseq treatment files were plotted with plotHeatmap available from deeptools (Ramírez et al., 2016). Additionally, coverage BIGWIG files of transcriptional activity (RNAseq) and all DNA methylation contexts were also used to calculated their enrichment on gene bodies, *Tyba* repeats, CENH3 domains, and TEs with computeMatrix and plotting with plotHeatmap.

Dating WGD events

To date the two rounds of duplication of the genome of *R. pubera*, a large tree of concatenated single copy genes was produced. For this analysis, each of the four homologous regions of *R. pubera* were separated and treated as a tip in the subsequent phylogeny reconstructions. Only coding sequences were used. We used BUSCO (*Poales* dataset) (Seppey et al., 2019) to look for conserved single-copy genes that are shared by all selected datasets. We performed this analysis in three different ways: solely the large syntenic block (Block1) of *R. pubera*, solely the smaller syntenic block (Block2) of *R. pubera*, and the two blocks combined. For the analyses, we included the following nine datasets: *J. effusus*, *C. littledalei*, *R. tenuis*, *R. breviuscula*, and the four homologous blocks of *R. pubera*. BUSCO analyses were run for all datasets; all the resulting single-copy genes were selected for each dataset. The single-copy genes shared among all datasets were used for the analyses: 841 for the larger Block1, 400 for the smaller Block2, and 1,204 for the two blocks combined. All genes were then aligned with MAFFT (Katoh and Standley, 2013), trimmed with Trimal (Capella-Gutiérrez et al., 2009), and concatenated into a single large multi-fasta alignment, and used as input for a ML tree built with IQ-tree2 (Minh et al., 2020).

A molecular clock analysis was performed to explore genome evolution in *Rhynchospora* and related genera. Divergence times were estimated using BEAST v.1.10.4 (Drummond and Rambaut, 2007) through the CIPRES Science Gateway fixing the tree topology from the Bayesian inference of the *Rhynchospora* concatenated 1,204 BUSCO gene alignment. Uncorrelated relaxed lognormal clock (Drummond and Rambaut, 2007) and Birth-Death speciation model (Gernhard, 2008) were applied. Two independent runs of 100,000,000 generations were performed, sampling every 10,000 generations. After removing 25% of samples as burn-in, the independent runs were combined and a maximum clade credibility (MCC) tree was constructed using TreeAnnotator v.1.10.4

(Drummond and Rambaut, 2007). To verify the effective sampling of all parameters and assess convergence of independent chains, we examined their posterior distributions in TRACER. The MCMC sampling was considered sufficient at effective sampling sizes (ESSs) equal to or higher than 200. The phylogeny was dated using both fossils and secondary calibration from published dated phylogenies. We chose three calibration points: i) Juncaceae/Cyperaceae divergence at 72.0 Mya (Bremer, 2002); ii) a fossil for *Carex* at 37.8 MYA (Smith et al., 2010), and iii) *R. pubera*/*R. tenuis* divergence at 32.0 Mya (Unpublished data).

Fluorescence in situ hybridization (FISH)

Interphase nuclei were prepared using the air-drying method, after enzymatic digestion with 2% cellulase Onozuka and 20% pectinase Sigma (Ribeiro et al., 2017). Roots were fixed in Carnoy ethanol:acetic acid 3:1 (v/v) for 2 h and stored at –20 °C. The best slides were selected for FISH, performed as described by Pedrosa et al. (2002) and the slides were counterstained with 2 μg/mL DAPI in Vectashield (Vector) mounting buffer. *Juncus effusus* interphase nucleus was hybridized with directly labeled (FAM)TTTAGGG(8)-telomeric probe and a directly labeled (CY3) probe for its most abundant satellite repeat, while *R. breviuscula* nucleus was hybridized with the same telomeric probe and directly labeled *Tyba* (CY3) oligo-probe.

Immunostaining

Immunostaining was performed as described before by Marques et al. (2016) with some modifications. Roots or Anthers were fixed in ice-cold 4% paraformaldehyde in 1× PBS buffer pH 7.5 (1.3 M NaCl, 70 mM Na2HPO4, 30 mM NaH2PO4) for 1 hr and 30 min and squashed in a drop of the same buffer. Then, the slides were washed in 1× PBS and blocked with 3% BSA for 30 min at 37 °C. Rabbit anti-H3K4me3 (abcam, ab8580), mouse anti-H3K9me2 (abcam, ab1220), and previously generated *R. pubera* rabbit anti-CENH3 (Marques et al., 2015) were used for immunostaining. The slides were counterstained with 2 mg/ml 49,6-diamidino-2-phenylindole (DAPI) in Vectashield H-1000.

QUANTIFICATION AND STATISTICAL ANALYSIS

Comparison of Hi-C contacts

The chromosomal interactions between holo- and monocentric plant species were compared by the ratios of *cis* and *trans* Hi-C contacts, i.e., for each species, we quatified the ratios of *cis* and *trans* Hi-C counts for every chromosome and tested if they were significantly different across distinct species. For grouped comparison, we adopted the mutiple testing method of one-way ANOVA (Analysis of Variance), specifically the Kruskal-Wallis ranked test with Holm-Bonferroni correction, because the compared values and ratios of intra- and inter-chromosomal contacts were different in length among various species and were not supported by evidence such as normality. Pair-wise significance analysis was conducted using Dunn's *post hoc* test.

Tyba array and CENH3 domain size and spacing

The Dunn's test was used to compare pairwise distributions of values of interest between *Tyba* arrays and CENH3 domains size and spacing.

Supplemental figures

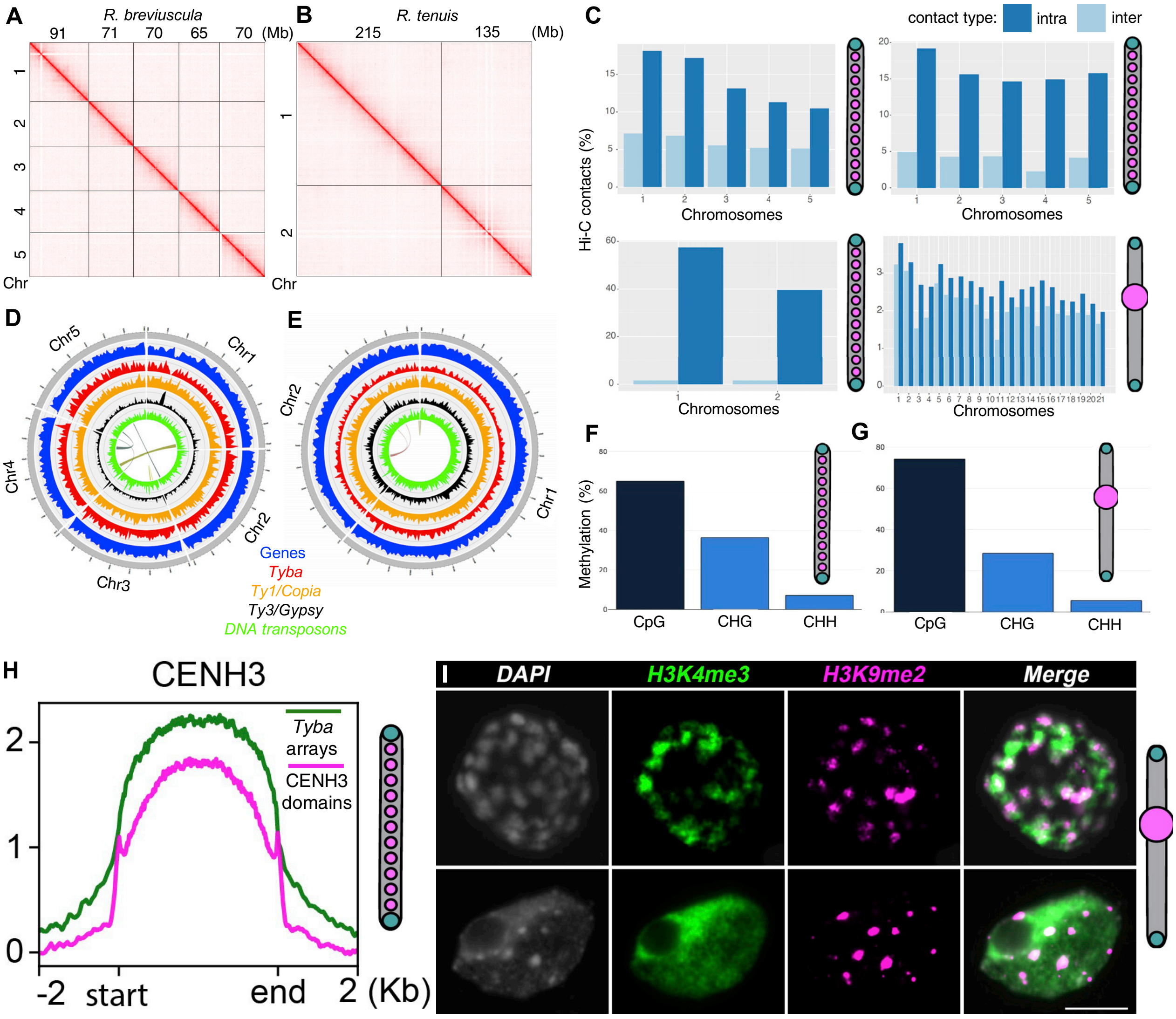


Figure S1. Characterization of the *Rhynchospora* and *J. effusus* genomes, related to Figures 2 and 4

(A and B) Contact maps for the five assembled pseudochromosomes of *R. breviuscula* (A) and the two assembled pseudochromosomes of *R. tenuis* (B). The intensity of pixels represents the normalized count of Hi-C links between 500-kb windows on a log scale.

(C) Hi-C contact counts (bin size, 1 Mb, normalization, VC) of intra- (*cis*) and interchromosomal (*trans*) chromatin contacts in the four species showing a significantly higher ratio ($p < 4.04e-05$) in holocentric compared with monocentric species, which implies relatively enriched *trans* interactions in the latter species.

(D and E) Distribution of the main classes of sequence types in *R. breviuscula* (D) and *R. tenuis* (E) with a 1-Mb sliding window. Note the high peaks of LTR *Ty3/Gypsy* density at most subtelomeric regions in *R. breviuscula* chromosomes. Self-synteny of *R. breviuscula* (D) and *R. tenuis* (E) genomes is shown in the inner circle.

(F and G) Summary of genome-wide DNA methylation contexts in *R. pubera* (F) and *J. effusus* (G).

(H) Metaplot showing the enrichment of CENH3 on *Tyba* repeat arrays (green) and CENH3 domains (magenta) in *R. breviuscula*.

(I) Immunostaining of metaphase chromosomes and an interphase nucleus of *J. effusus* for H3K4me3 and H3K9me2. Scale bars, 5 μm.

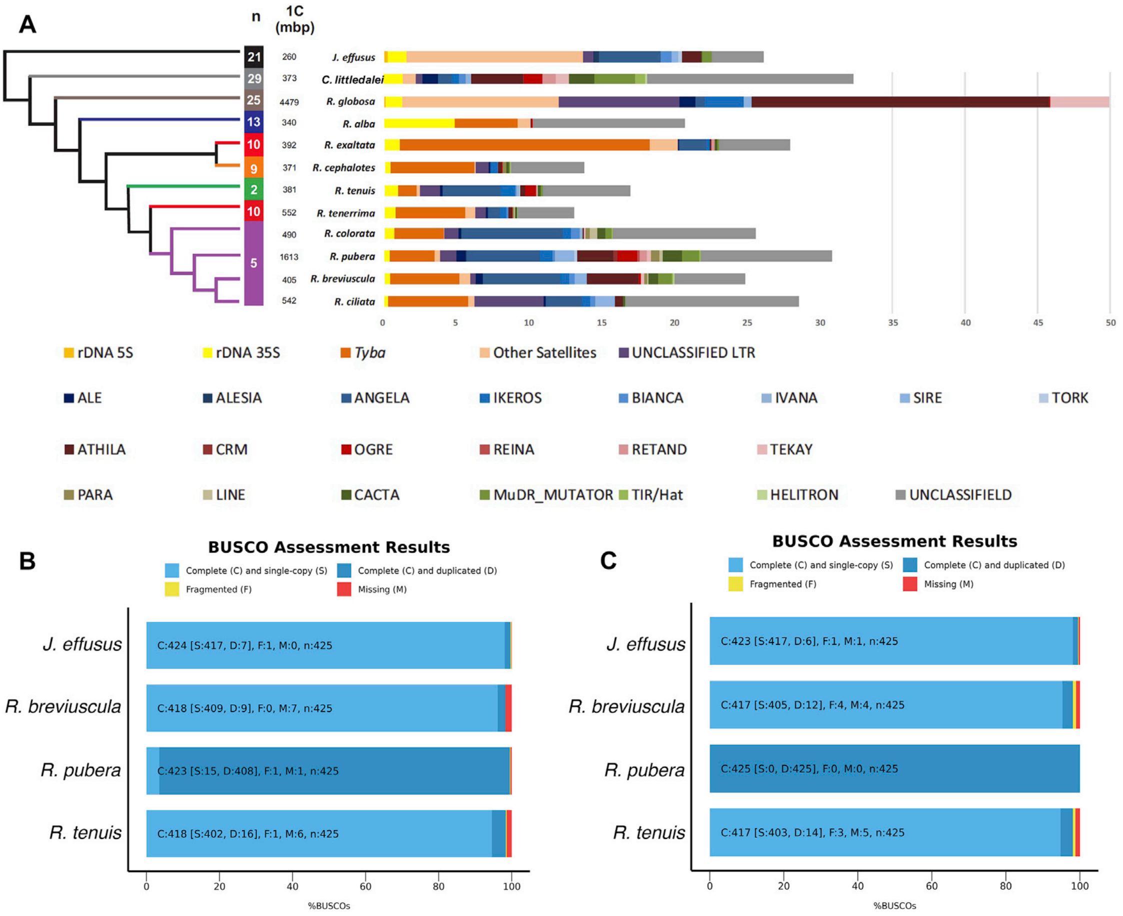


Figure S2. Composition and evolution of sedges and rush genomes, related to Figure 5

(A) Schematic phylogenetic tree and repeat composition of beak-sedge genomes and comparison with *C. littledalei* and *J. effusus*.

(B and C) BUSCO assessment for completeness of genic space with the viridiplantae_odb10 dataset, using the entire genome assembly (B) or the longest transcript (C).

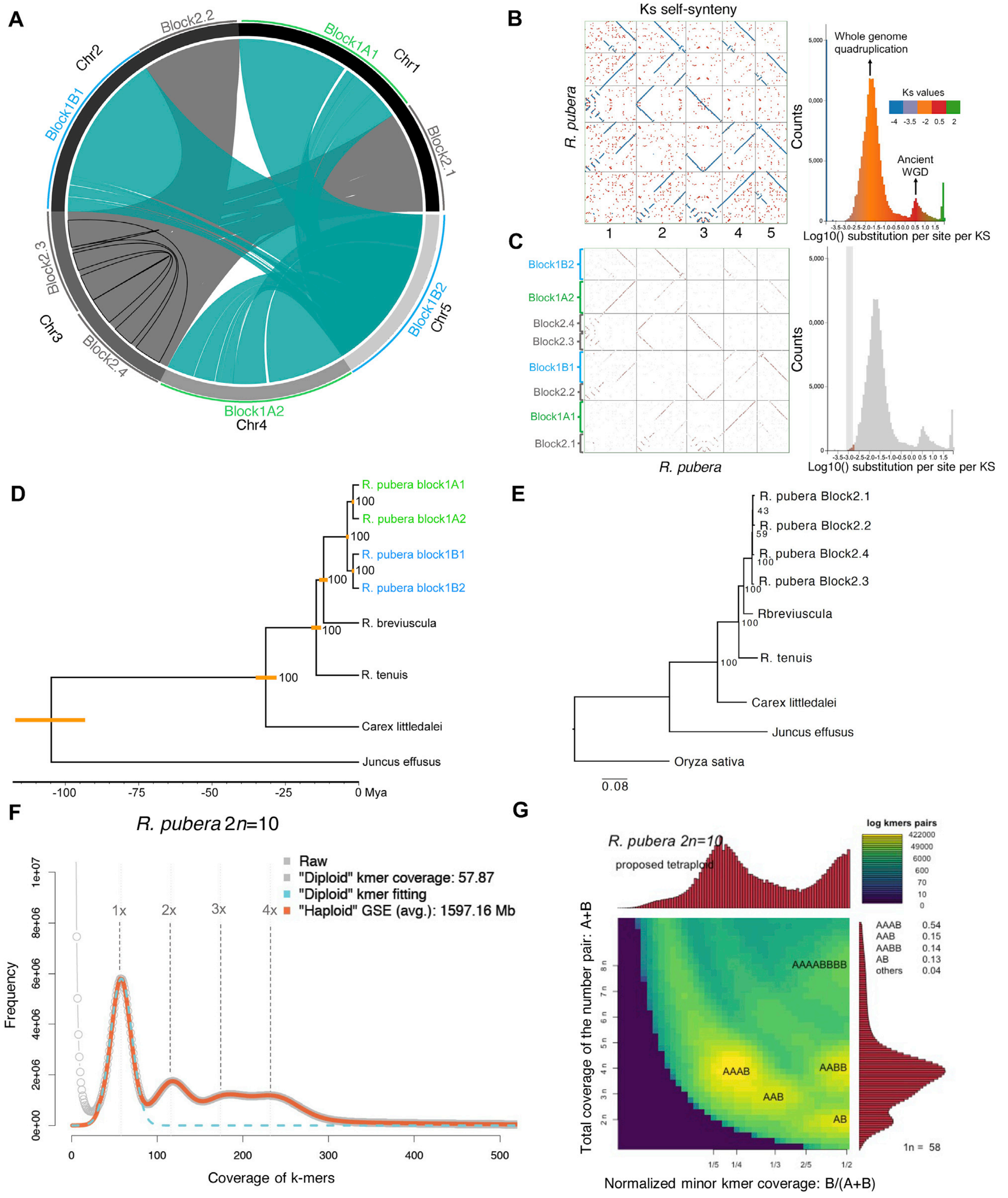


Figure S3. Identification, characterization, and dating of WGDs in *R. pubera*, related to Figure 5

(A) SyMap self-synteny plot of *R. pubera*. Block structure is indicated by outer arcs.

(B) SyMap self-synteny dot plot colored based on Ks values. Ks values on a log scale are shown to the right of the dot plot. Note the large peak that correlates with the large duplication events in *R. pubera* and a second small peak most likely representing an ancient WGD.

(legend continued on next page)

(C) Same plot as (B) but selecting only the sequences with the lowest number of synonymous substitutions, allowing the identification of intragenomic syntenic block relationships (Block1A and Block1B). We were unable to detect any relationships for Block2. The small colored block within the vertical gray bar represents the sequences with the lowest number of synonymous substitutions used in the dot-plot to the left. Ks values are indicated by the color scale in (B).

(D) Based on the assessment of the relationships among the syntenic blocks of *R. pubera*, we selected 1,204 BUSCO genes (Poales dataset) uniquely present in each block and also shared with *R. breviuscula*, *R. tenuis*, *C. littledalei*, and *J. effusus* to build a phylogenetic tree from a concatenated alignment, which was further used for dating the duplication events in *R. pubera*. We confirmed the Block1A and Block1B relationships with 100% bootstrap support and also determined that a first WGD occurred around 3.8 Mya, followed by a second event around 2.1 Mya. Note that the second WGD closely overlaps in both Block1A and Block1B branches. Yellow bars indicate the dating time interval.

(E) Phylogenetic analysis of Block2 genes did not resolve the relationships for this particular block and was not used for dating.

(F and G) K-mer based estimation of genome size and heterozygosity (F) and smudgeplot analysis of k-mer-based ploidy inference for *R. pubera* using 21-mers (G). GSE, genome size estimation. Smudgeplot infers ploidy directly from the k-mers present in sequencing reads by analyzing heterozygous k-mer pairs.

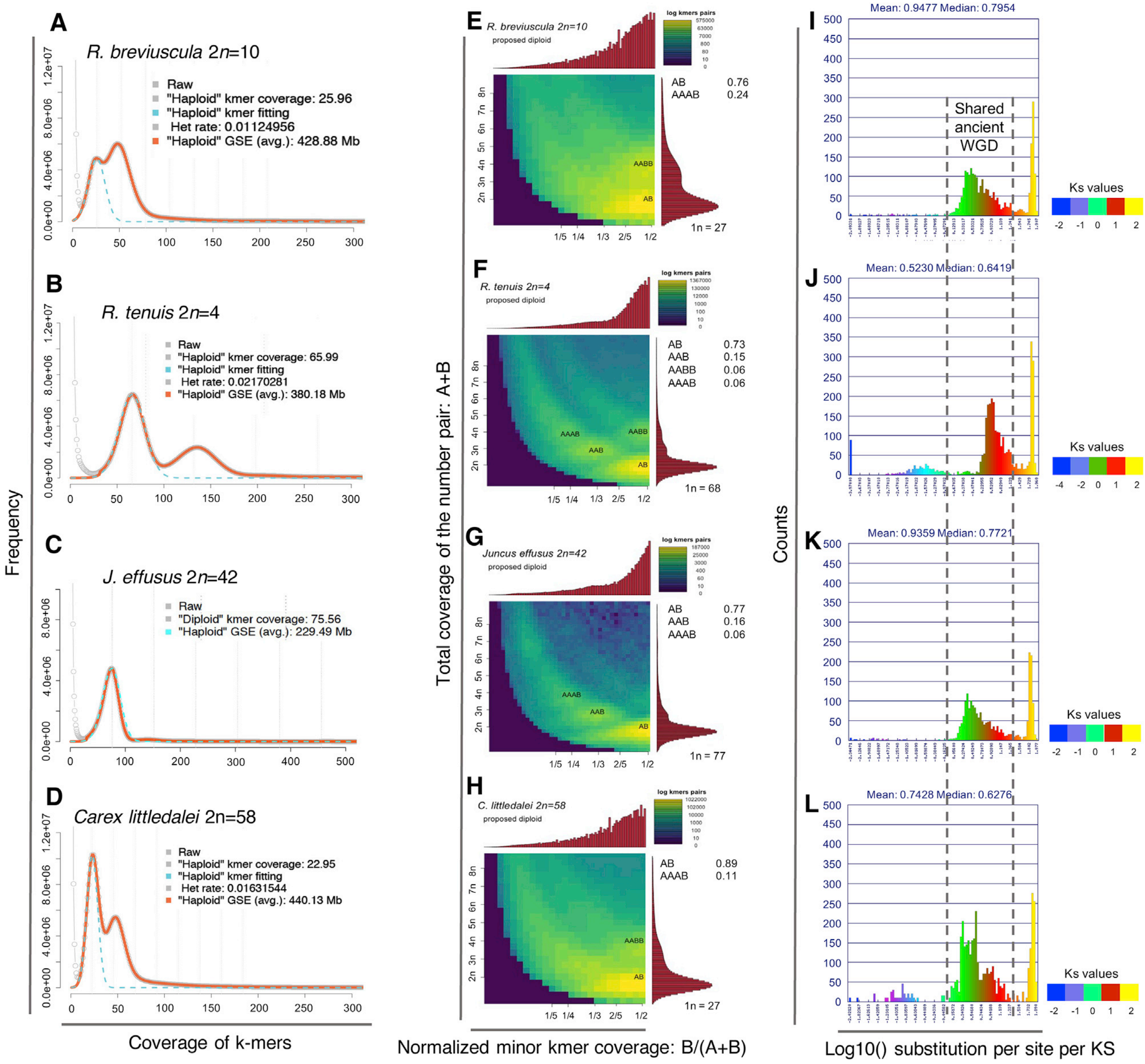


Figure S4. K-mer based genome size estimation and ploidy inference and WGD identification in sedges and rushes, related to Figure 5
(A–D) 21-mer based estimation of genome size and heterozygosity. GSE, genome size estimation.
(E–H) Ploidy and genome structure inference based on 21-mer Smudgeplot analysis.
(I–L) Ks values of coding sequences for each genome; a shared ancient WGD peak was observed for all species.

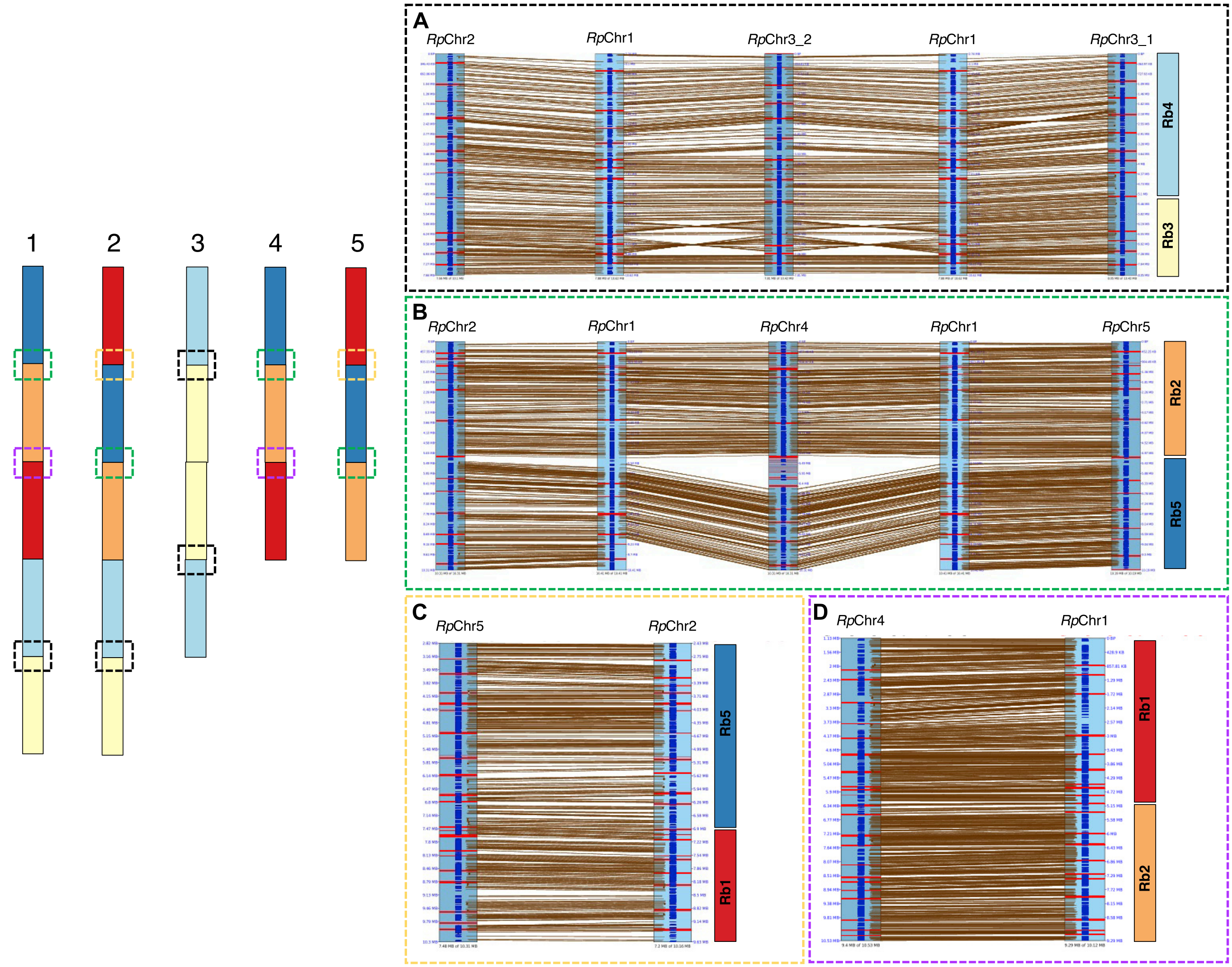


Figure S5. Comparative alignment of the duplicated end-to-end fusion (EEF) transition regions in the *R. pubera* genome, related to Figure 5
(Left) Ideogram model of *R. pubera* chromosomes, with the dashed boxes indicating the extracted and compared regions on the right.
(A) Alignment of the EEF of *Rb*3 and *Rb*4 found once on *Rp*Chr1 and *Rp*Chr2 and twice on *Rp*Chr3, showing the same fusion signature.
(B) Alignment of the EEF of Rb2 and *Rb*5, found on *Rp*Chr1, *Rp*Chr2, *Rp*Chr4, and *Rp*Chr5, also showing the same fusion signature.
(C) Alignment of the EEF of *Rb*1 and *Rb*5, found on *Rp*Chr2 and *Rp*Chr5 with the same fusion signature.
(D) Alignment of the EEF of *Rb*1 and Rb2, found on *Rp*Chr1 and *Rp*Chr4 with the same fusion signature. Colored boxes assign the synteny to *R. breviuscula* chromosomes. Red stripes on the synteny alignments depict *Tyba* repeats, while genes are annotated in dark blue.

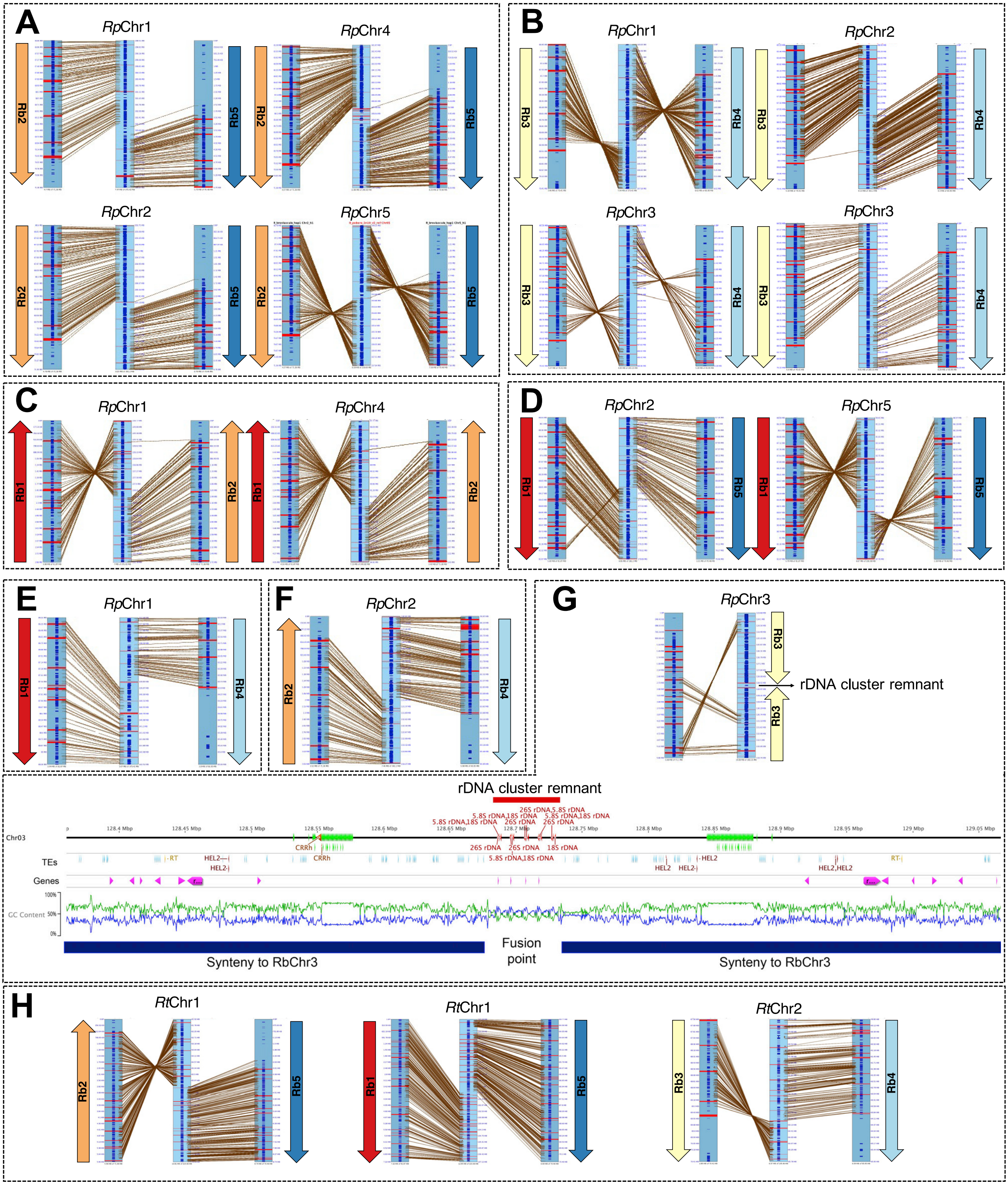


Figure S6. Identification of the sequences underlying the transitions between the syntenic regions to *R. breviuscula* chromosomes in the end-to-end fusions found in the *R. pubera* and *R. tenuis* genomes, related to Figure 5

(A) EEF of Rb2 and *Rb*5 found on *Rp*Chr1, *Rp*Chr2, *Rp*Chr4, and *Rp*Chr5. Similar fusion signatures are shared among the four chromosomes. In three of them, a *Tyba* repeat is found between them.

(legend continued on next page)

(B) EEF of *Rb*3 and *Rb*4 found on *Rp*Chr1 and *Rp*Chr2 and twice on *Rp*Chr3 with the same fusion signature. A *Tyba* repeat array is found between the transitions in all cases.

(C) EEF of *Rb*1 and Rb2 found on *Rp*Chr1 and *Rp*Chr4 with the same fusion signature, without a *Tyba* repeat in between.

(D) EEF of *Rb*1 and *Rb*5 found on *Rp*Chr2 and *Rp*Chr5 with the same fusion signature, with a *Tyba* repeat in between.

(E) EEF of *Rb*1 and *Rb*4 found only on *Rp*Chr1 with a *Tyba* repeat array in between.

(F) EEF of *Rb2* and *Rb*4 found only on *Rp*Chr2 with no *Tyba* repeat in between.

(G) EEF of *Rb*3 and *Rb*3 found only on *Rp*Chr3 and with a remnant of a rDNA cluster in the transition region (with detailed annotation shown to the right).

(H) Characterization of the three EEFs responsible for the chromosome reduction in *R. tenuis*. On *Rt*Chr1 we found an EEF involving Rb2 and *Rb*5, and a second event involving *Rb*5 and *Rb*1, while on *Rt*Chr2, we found a single EEF involving *Rb*3 and *Rb*4. Colored arrows indicate the *R. breviuscula* chromosomes and point to the telomeric region involved in the fusion event. Remarkably, although similar ancestral chromosome associations are found in *R. pubera* and *R. tenuis*, the chromosomal ends involved in the fusions are different. Red stripes on the synteny alignments depict *Tyba* repeats, while genes are annotated in dark blue.

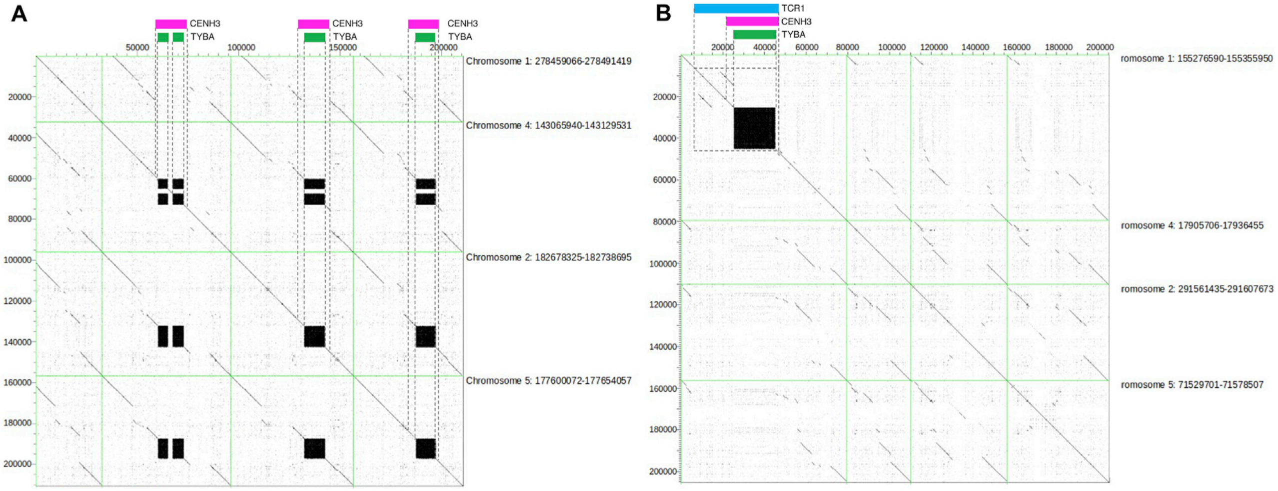


Figure S7. Characterization of emergence and loss of CENH3-binding regions in *R. pubera*, related to Figure 6

(A) Example of CENH3-binding region and *Tyba* array lost in one of four paralogous regions, while the other three copies retained the *Tyba* array and CENH3 binding. The conserved locus is indicated by the dashed box, along the x axis of the dot plot, with rectangles marking the area associated with CENH3 (magenta) and the *Tyba* array (green). The genome positions of the extracted regions are given to the right.

(B) Example of CENH3-binding region and *Tyba* array gain in one of four paralogous regions due to a transposition of *Tyba*-containing *TCR1* in *Rp*Chr1, while the other three copies lack the *Tyba* array. The gained locus is indicated by the dashed box, along the x axis of the dot plot, with rectangles marking the *TCR1* element (blue), the area associated with CENH3 (magenta), and the *Tyba* array (green). The genome positions of the extracted regions are given to the right.