SI Appendix

SI Materials and Methods

Plant material. Plants of *Rhynchospora pubera* (Vahl) Boeckler, *R. tenuis* Link and *R. ciliata* (Vahl) Kükenthal were cultivated at humid conditions in a greenhouse. Additionally, leaf material from other Cyperaceae species (*Carex flacca* Schreb.*, Cyperus aggregatus* (Willd.) Endl., *Scleria bracteata* Cav. and *Scirpoides holoschoenus* (L.) Soják) (SI Appendix, Table S3) was collected for DNA isolation.

Somatic and meiotic meristem preparation. Chromosome preparations for in situ hybridization analysis were conducted as described by Ruban*, et al.* (1), with modifications. First, young roots (pre-treated with 8-hydroxychinolin for 24 h at 10°C) and anthers were fixed in ice-cold methanol for 30 min. After that the fixative was changed to ice-cold 3:1 (methanol:acetic acid) for 2 h. The fixed tissues were treated with an enzyme mixture (0.7% cellulase R10, 0.7% cellulase, 1.0% pectolyase, and 1.0% cytohelicase in 1× citric buffer) for 30 min at 37ºC. Material was then washed in 1× citric buffer, twice in ice-cold water and fragmented in 7 µl of 60% freshly prepared acetic acid into smaller pieces with the help of a needle on a slide. After another 7 µl of 60% acetic acid was added, and the specimen was kept for 2 min at room temperature. Next, a homogenization step was performed with an additional 7 µl 60% acetic acid and the slide was placed on a 55ºC hot plate for 2 min. The material was spread by hovering a needle over the drop without touching the hot slide. After spreading of cells, the drop was surrounded by 200 µl of ice-cold, freshly prepared 3:1 (ethanol:acetic acid) fixative. More fixative was added and the slide was briefly washed in fixative, then dipped in 60% acetic acid for 10 min and rinsed 5 times in 96% ethanol. A quality check of the air dried slides was performed by phase contrast microscopy. The slides were stored until use in 96% ethanol at 4ºC. Chromosome preparations for immunolabelling analysis were made as described by Marques*, et al.* (2).

Probe preparation and fluorescence *in situ* hybridization. FISH probes were obtained as 5′-Cy3 or 5′-FAM-labeled oligonucleotides (Eurofins MWG Operon, http://www.eurofinsdna.com), or were PCR-amplified. All DNA probes, except oligonucleotides, were labelled with Cy3-, Texas Red- or Alexa 488-dUTP by nick translation, as described by Kato, Albert, Vega and Birchler (3). The sequences of all oligonucleotides and primers are listed in SI Appendix, Table S4. FISH was performed as described in Ma*, et al.* (4). Probes were then mixed with the hybridization mixture (50% formamide and 20% dextran sulfate in 2× SSC), dropped onto slides, covered with a cover slip and sealed. After denaturation on a heating plate at 80°C for 7 min, slides were hybridized at 37°C overnight. Post-hybridization washing was performed in 2× SSC for 20 min at 58°C. After dehydration in an ethanol series, 4′,6–diamidino-2–phenylindole (DAPI) in Vectashield (Vector Laboratories, http://www.vectorlabs.com) was applied. Microscopic images were recorded using an Olympus BX61 microscope equipped with an ORCA-ER CCD and a deconvolution system. Images were analyzed using the SIS software (Olympus).

PCR amplification of Tyba fragments. Tyba fragments for probe labelling were amplified using gDNA from *R. pubera*, *R. tenuis* and *R. ciliata* for all members using the forward primer Tyba1F: CTAAGTCATTTCATCACAATAATCTAC and the reverse primer Tyba1R: AATCCAGAAACGATTGAAATGCTC for Tyba1 and Tyba2F: GTGCAAATAATGCAATTCTGAGCATC and Tyba2R: ATATGCGTAATTACCATGTATAATCC for Tyba2. PCR reactions were performed in 25 µL reaction volume containing 100 ng of gDNA, 1 µM primers, 1× PCR buffer, 0.2 mM dNTPs, and 1 U of Taq polymerase (Qiagen). Thirty-five amplification cycles (45 s at 95°C, 45 s at 57 °C annealing temperature and 45 s at 72°C) were run.

Expression analysis of Tyba and RpCENH3 by semi-quantitative RT-PCR. Total DNase treated RNA was isolated from root, leaf and anther tissue of *R. pubera* using the SpectrumTM plant total RNA kit (Sigma) according to manufacturer's instructions. The cDNA was synthesized from 1µg of total RNA using the RevertAid First Strand cDNA Synthesis Kit (Thermo Scientific). PCR reactions were performed as described above. Primers sequences are listed in SI Appendix, Table S4, specific primers for the constitutively expressed GAPDH gene (5), GAPDH-F CAATGATAGCTGCACCACCAACTG and GAPDH-R CTAGCTGCCCTTCCACCTCTCCA, were used as control for equal amount of gDNA and cDNA. Amplified fragments of RpCENH3 were cloned into the StrataClone PCR Cloning Vector pSC-A-amp/kan (Agilent Technologies). Consensus sequences derived from sequencing of 10 randomly selected clones revealed two minor CENH3 variants which have been deposited in GenBank under accession numbers KR029618 and KR029619.

RNAseq and *de novo* assembly for identification of CENH3 gene. Total RNA was isolated from *R. pubera* pollen mother cells using the Spectrum^{TM} plant total RNA kit (Sigma) according to manufacturer's instructions followed by cDNA library preparation. Final libraries were paired-end sequenced 2x100bp on Illumina HiSeq 2000. D*e novo* assembly was performed using Velvet assembler (6), running a total of 87 million of paired-end reads, accession number PRJEB9645 at the Sequence Read Archive (http://www.ebi.ac.uk/ena/). Assembled contigs were finally submitted to a consensus assembly using Geneious assembler, Geneious version 7.1.7 [http://www.geneious.com/, (7)] producing 75,353 contigs. The Velvet assembly was performed using the Geneious provided plugin (Geneious v. 7.1.7). The assembly summary is shown on SI Appendix, Table S5.

Generation of a CENH3 antibody. The peptide ARTKHFSVRSGKKSASRTK was used to generate a *R. pubera* CENH3-specific (RpCENH3) polyclonal antibody. Peptide synthesis, immunization of rabbits and peptide affinity purification of antisera was performed by LifeTein (http://www.lifetein.com).

Preparation of extended fibers and immuno-FISH. Extended DNA fibers were obtained by first isolating leaf nuclei according to Li, Yang, Tong, Zhao and Song (8). Briefly, nuclei were obtained by chopping leaves according to the following method: 100 mg of fresh young leaves were collected for the preparation of 5-10 slides, and leaves were chopped with a sharp sterile scalpel in a Petri dish that contained 1 ml of ice-cold nucleus isolation buffer (0.01 M MgSO4, 0.005 M KCl, 0.0005 M HEPES, 1 mg/ml dithiothreitol, and 0.25% Triton X-100) (9, 10). The materials obtained by chopping were filtered through 33-µm nylon mesh, filtrates were centrifuged at high speed (16,000g) for 40 s, and the supernatant was discarded. The sediment was resuspended in 10 µl of nucleus isolation buffer. DNA fibers were obtained by dropping 2 µl of the suspension on one end of a coated slide and air dried for 5 to 10 min at room temperature. Thirty microliters of nucleus lysis buffer (0.5% sodium dodecylsulfate, 5 mM ethylenediaminetetraacetic acid, 100 mMTris, pH7.0) was added to the nuclei and incubated at room temperature for 9 min. DNA fibers were dragged and extended with a clean coverslip followed by fixation in 4% paraformaldehyde for immuno-FISH.

Immuno-FISH on extended fibers and somatic cells was performed as described in Ishii, Sunamura, Matsumoto, Eltayeb and Tsujimoto (11). Finally, preparations were stained with DAPI/Vectashield (Vector Laboratories, Burlingame, CA, USA).

Flow-sorting of nuclei for immuno-FISH signal overlapping quantification. Young leave tissue of *R. pubera* was fixed for 20 min under vacuum in 4% formaldehyde in Tris.Cl buffer (100 mM TRIS-HCl, pH 7, 5 mM $MgCl_2$, 85 mM NaCl, 0.1% Triton X100), washed twice for 10 min in TRIS buffer and chopped with a sharp razor blade in about 1 ml of ice-cold nuclei

isolation buffer LB01 (12). Resulting suspension was filtered through a 35µm mesh and nuclei were stained with 4′,6-diamidino-2-phenylindole (DAPI) (1 µg/ml) and flow-sorted using a FACSAria (BD Biosciences). 12 µl of sorted nuclei were mixed with equal amounts of sucrose buffer (100mM Tris, 50mM KCl, 2mM MgCl, 0.05% Tween-20, 5% sucrose) on slides. The slides were dried at room temperature and either used immediately or stored at -20°C until use.

Super-resolution microscopy. To analyze the structures and spatial arrangement of immunosignals and chromatin at a lateral optical resolution of ~120 nm (super-resolution, achieved with a 488 nm laser), 3D structured illumination microscopy (3D-SIM) was applied using a C-Apo 63×/1.2W Korr objective of an Elyra PS.1 microscope system and the software ZEN (Zeiss, Germany). Image stacks were captured separately for each fluorochrome using the 561, 488, and 405 nm laser lines for excitation and appropriate emission filters, and then merged using the ZEN software. The degree of co-localization between Tyba and CENH3 was measured in a single representative slice of each image stack and calculated by the ZEN software.

Southern blot hybridization. The Southern hybridization procedure was performed according to Sambrook, Fritsch and Maniatis (13) with modifications. Total genomic DNA was isolated from leaf tissue of *R. pubera*, *R. ciliata*, *R. tenuis*, *Carex flacca, Cyperus aggregatus*, *Scleria bracteata* and *Scirpoides holoschoenus,* using the DNeasy plant maxi kit (Qiagen) according to manufactures' instructions. The genomic DNAs of all species were further digested with the enzyme *Dra*I, which recognize only one restriction site within the Tyba monomer, size-fractionated by 1.8% agarose gel electrophoresis and transferred to Hybond N+ nylon membranes (Amersham). Probes for Tyba 1 and Tyba 2 were prepared after PCR amplification from genomic DNA of *R. pubera* or *R. tenuis* and labelling by random primer with α-^{32}P-dATP (Thermo Scientific). Hybridization was done overnight at 65 °C in Church and Gilbert hybridization buffer and post-hybridization

washes carried out at 65 °C in 2× SSC, 0.5% SDS for 20 min followed by 1× SSC, 0.5% SDS at 65 °C for 20 min and 0.5× SSC, 0.5% SDS at 65 °C for 20 min for high stringency and 2× SSC, 0.5% SDS at 65 °C for 20 min followed by 1× SSC, 0.5% SDS at 65 °C for 20 min for low stringency, respectively.

BAC library construction and screening. Cell nuclei of *R. pubera* were isolated from young leaves following the protocol of Doležel, Číhalíková and Lucretti (14). Briefly, the leaves were fixed for 20 min in 2% (v/v) formaldehyde and immediately afterwards chopped by a razor blade in ice-cold isolation buffer (15). The suspension of released nuclei was stained by DAPI (2 µg/ml). The nuclei were purified by flow cytometry and used to prepare high-molecular-weight (HMW) DNA as described in Šimková, Číhalíková, Vrána, Lysák and Doležel (15). HMW DNA of 1.2 million nuclei of *R. pubera* (~4.2 µg DNA) were used to construct a large insert library. *Hin*dIII digested HMW DNA was cloned in pIndigoBAC-5 vector (Epicentre, Madison, WI, USA) as described in Šimková*, et al.* (16). The *R. pubera* BAC library is composed of 3,840 clones of 120 kb insert size, which cover 0.25x of *R. pubera* genome (2C = 3.3 pg).

Screening of BACs containing Tyba was carried out by hybridization with PCR-amplified Tyba probes using the procedure described in Ming*, et al.* (17). BACs showing a wide range of hybridization intensity were chosen for sequencing.

Illumina HiSeq sequencing of genomic DNA and BACs. Library preparation was carried out by using ~1 µg of genomic DNA or BAC-DNA. Following random shearing by ultra-sonication (Covaris S220; Covaris Inc.) fragmented DNA was end-repaired, adapter-ligated, barcoded and amplified as previously described by Meyer and Kircher (18). Adapter-ligated DNA was size-selected in a range of 400 – 500 bp for sequencing 2x100 bp on Illumina HiSeq2000. The original Illumina sequencing data for the genomic DNA and BACs are available under study accession number PRJEB9643 and PRJEB9649 at the Sequence Read Archive (http://www.ebi.ac.uk/ena/), respectively.

Repeat identification and ChIP-seq analysis. Identification and characterization of moderately to highly repeated genomic sequences was achieved by graph-based clustering (19) of genomic Illumina reads using RepeatExplorer pipeline (20). A total of 8 million reads (SI Appendix, Table S1), representing 3.6× genome coverage, were used for the clustering and 369 largest clusters with genome proportions of at least 0.01% were examined in detail. Clusters containing satellite repeats were identified based on the presence of tandem sub-repeats within their read or assembled contig sequences. These satellite repeats were characterized using oligomer frequency analysis of the reads within their clusters as described previously (21). To identify repeats associated with CENH3-containing chromatin, reads from the ChIP-seq experiment obtained by sequencing DNA from isolated chromatin prior to (the input control sample) and after immunoprecipitation with the CENH3 antibody (the ChIP sample) were separately mapped to the repeat clusters. The mapping was based on read similarities to contigs representing individual clusters, using BLASTn (22) with parameters "-m 8 -b 1 -e 1e-20 -W 9 -r 2 -q -3 -G 5 -E 2 -F F" and custom Perl scripts for parsing the results. Each read was mapped to a maximum of one cluster, based on its best similarity detected among the contigs. Ratio of ChIP/input reads assigned to individual clusters was then used to identify repeats enriched in the ChIP sample as compared to the input.

CENH3-ChIP, ChIP-qPCR and ChIPseq. Immunoprecipitation experiments were done as described in Kuhlmann and Mette (23). First, young leaves and buds were collected and cross-linked with formaldehyde 1% for 30 min on ice. Leaves and buds were then ground in liquid nitrogen and sonicated using a Diagenode Sonicator. Sonicated chromatin-DNA ranging from ~400-800 bp was immunoprecipitated using anti-RpCENH3. To verify the quality of our IP-DNA we have performed real-time quantitative PCR using Tyba primers as putative positive markers and the 26S ribosomal primers (SI Appendix, Table S4) as negative control in three different samples: input chromatin isolated DNA, immunoprecipitated DNA and no antibody control (noAB).

Immunoprecipitated DNA and input samples (3-7ng for each sample) were used for library preparation following manufacturer's recommendations (Illumina TruSeq ChIP Sample Preparation Kit #IP-202-1012). Subsequently, prepared libraries were paired-end sequenced 2x100bp on Illumina HiSeq 2000. The original ChIPseq sample data are available under study accession number PRJEB9647 at the Sequence Read Archive (http://www.ebi.ac.uk/ena/).

BAC assembly and annotation. Seven positive clones from the *R. pubera* library were sequenced, producing millions of HiSeq paired-end reads (100bp, with >500× coverage). Next, reads were assembled using MIRA (24) and Velvet (6) assemblers. Contigs obtained from both assemblers were then submitted to several rounds of assembly using the Geneious assembler (Geneious v. 7.1.7) and manually edited. Contigs best-matching the estimated BAC length as measured by pulse-field-gel-electrophoresis (PFGE) were then used for annotation. For quantification and annotation of repetitive sequences we performed clustering analysis on RepeatExplorer on each batch of BAC paired-end reads. This approach helped mainly in the correct quantification and annotation of transposable elements and Tyba arrays. In parallel, Phobos, a tandem repeat search tool (Phobos 3.3.11, 2006-2010, http://www.rub.de/spezzoo/cm/cm_phobos.htm), was used for correct localization and annotation of Tyba arrays. Phobos was also used for the identification of Tyba high order repeat structures. Coding sequences were identified using the gene prediction tools Augustus and Glimmer and manually annotated by BLAST searches. All the analyses were performed inside Geneious v. 7.1.7 with the provided plugins, except for the clustering analyses. BAC annotation is described in SI Appendix, Table S2. To infer whether coding sequences presents in the BACs are transcriptionally active we isolated and blasted each individual predicted coding sequence against our PMC transcriptome database. Assembled BAC sequences are available through iPlant Data Store and can be accessed via iPlant Discovery Environment or at http://de.iplantcollaborative.org/dl/d/8258A143-C5F5-4DF1-84F2-88C94BE8EA8F/R_pubera_holocentromeres_data.rar).

Phylogenetic analysis. Reference IDs for all CENH3 sequences used in this study are available in SI Appendix, Table S6. Multiple alignment of protein sequences encoding the entire CENH3 sequences was generated using MUSCLE (25) and refined manually. Evolutionary analyses were conducted with IQ-TREE (26) using ultrafast bootstrap (27). Phylogenetic history was inferred using the Maximum Likelihood method. The analysis involved 113 protein sequences. All positions containing gaps and missing data were eliminated. There were a total of 101 positions in the final dataset.

Phylogenetic analysis of CRRh was done as previously and using the same alignment matrix from Neumann*, et al.* (28).

Supplementary Figures

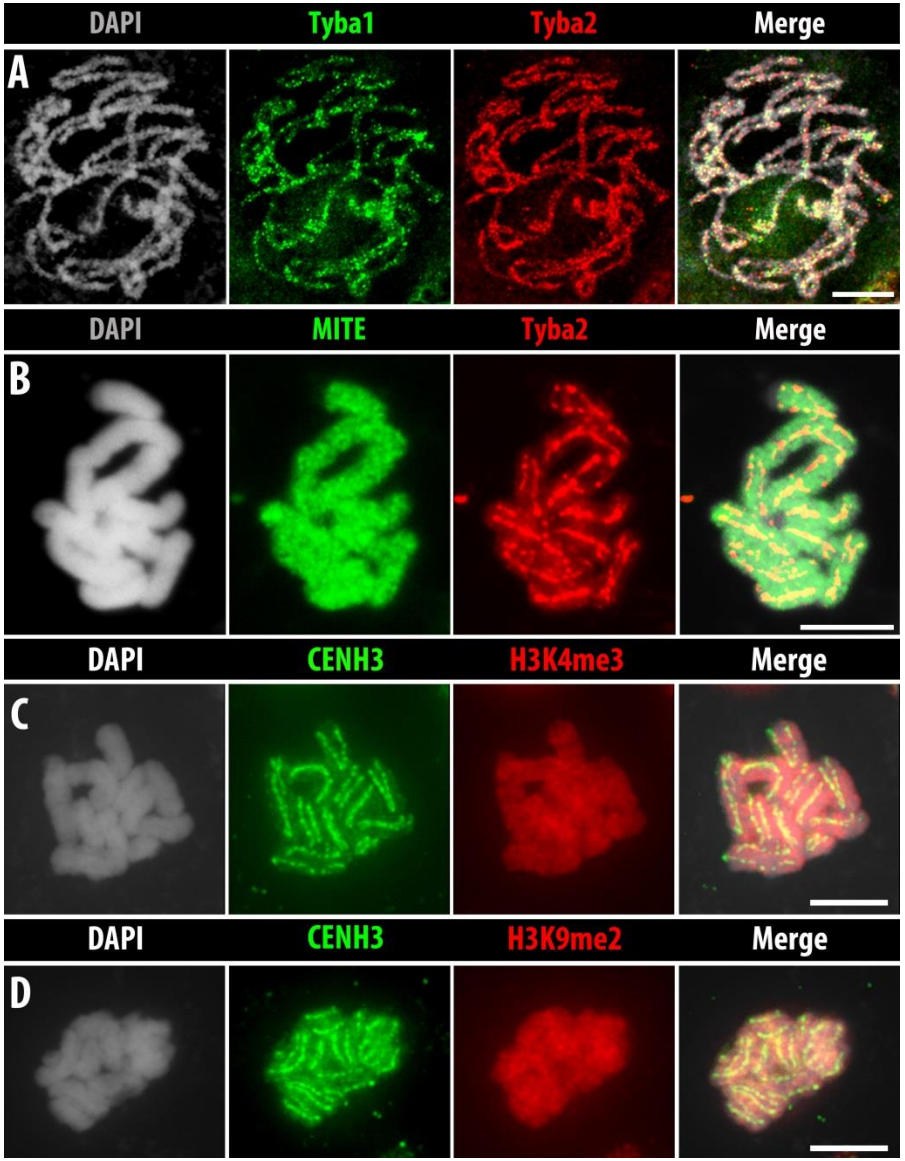

Fig. S1. Overall chromosomal chromatin organization in *R. pubera*. (**A-B**) FISH localization of Tyba and MITE repeats in *R. pubera* chromosomes. (**A**) Hybridization signals of both Tyba members in prophase chromosomes showing a line-like labeling on the poleward surface of each chromatid. (**B**) MITE signals are dispersed while Tyba2 displays a holocentromere-like pattern in metaphase chromosomes. (**C-D**) Metaphase chromosomes of *R. pubera* immunostained with antibodies recognizing H3K4me3 (**C**) and H3K9me2 (**D**) histone modifications in combination with anti-CENH3. Note, the disperse distribution of both H3 modifications. Scale bars: 5 µm.

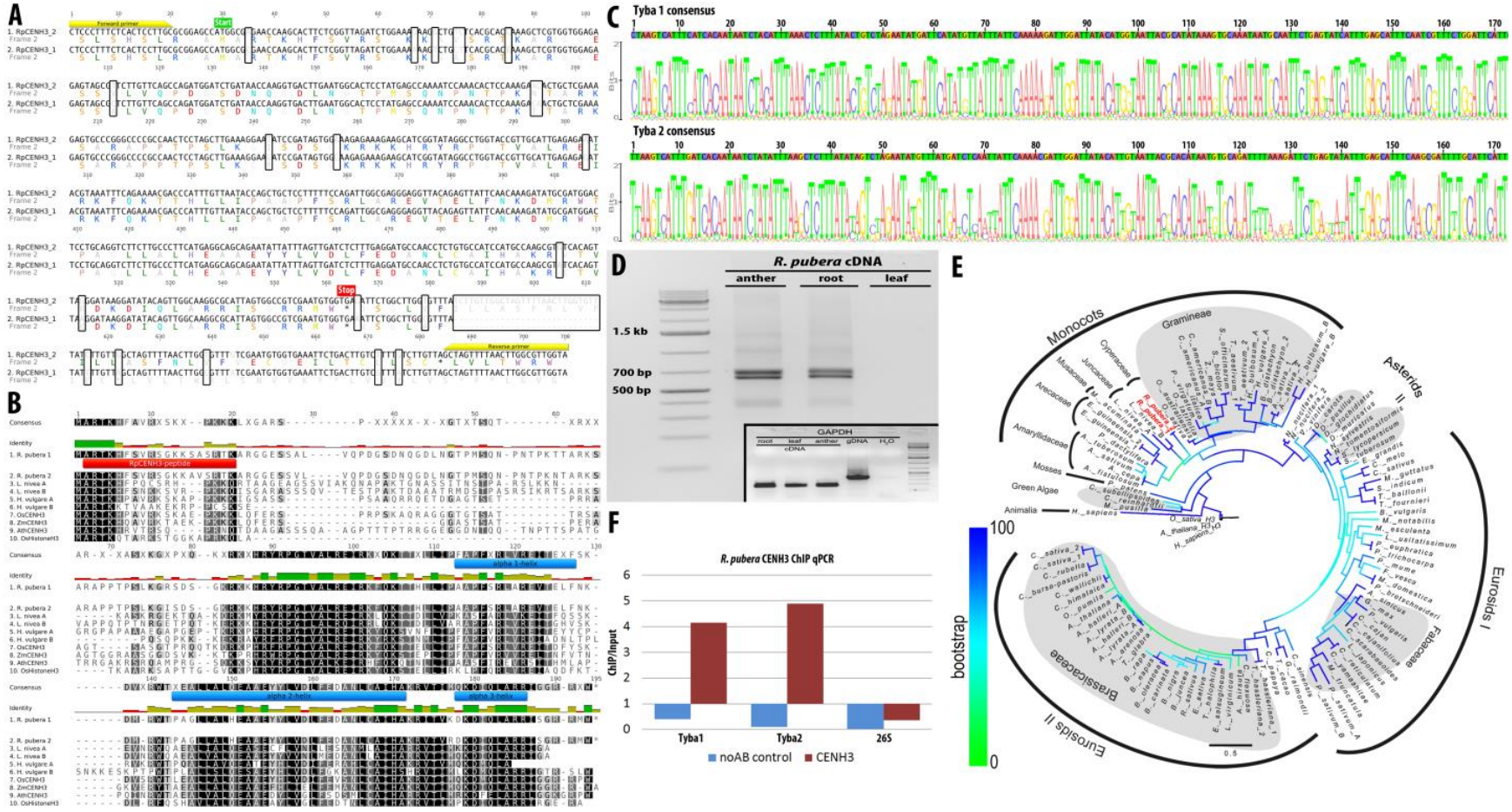

Fig. S2. CENH3 sequence characterization, Tyba monomer reconstruction and CENH3-ChIP analysis. (**A**) DNA and amino acid alignment of *R. pubera* CENH3 variants. Yellow boxes indicate the primer-binding sites used to amplify the fragments; green and red boxes indicate start and stop codons, respectively. Nucleotide disagreements between the variants are high-lightened by black-lined boxes. (**B**) Amino acid alignment of *R. pubera* CENH3 variants and other plant

CENH3 sequences. Red box and blue boxes indicate the amino acid residues used for generation of anti-RpCENH3 and histone alpha helixes fold domains, respectively. (**C**) Monomer reconstruction of Tyba1and Tyba2 using base frequency logo representation. (**D**) RT-PCR analysis of RpCENH3s in different tissues. (**E**) Analysis of evolutionary divergence in plant CENH3 sequences. The evolutionary history was inferred by using the Maximum Likelihood method based on the LG matrix-based model (79). The tree with the highest log likelihood (-5988.401) is shown. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. (**F**) Quantitative real-time PCR of *R. pubera* CENH3-ChIP using Tyba1 and 2 specific primers. As negative control we used a set of primers to specifically amplify a short region of the 26S ribosomal RNA gene. No antibody control (noAB) was used as a negative control for amplification.

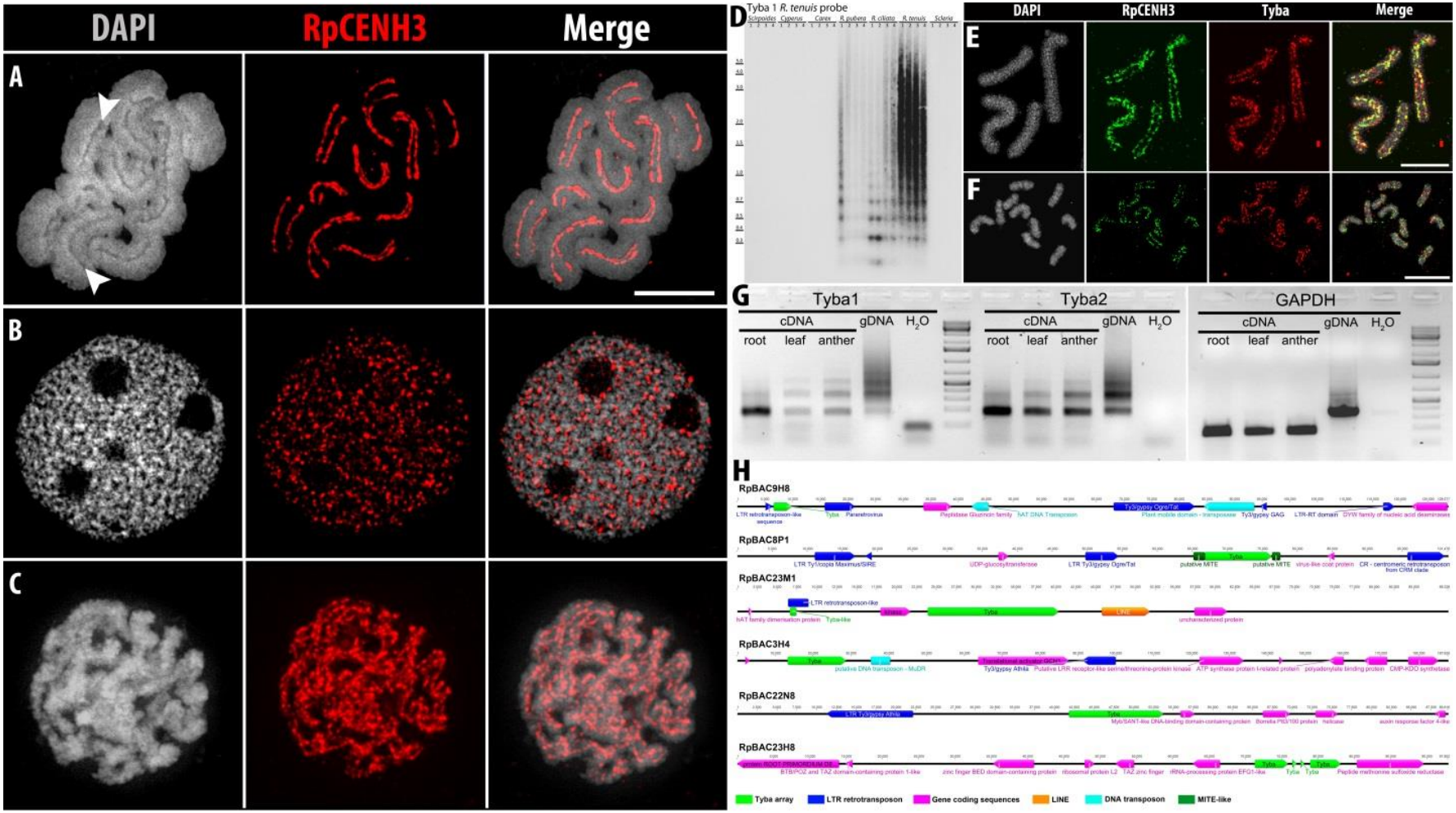

Fig. S3. Localization of anti-RpCENH3 during the cell cycle of *R. pubera*, characterization of the satellite features of Tyba, its presence in the Cyperaceae family, and analysis of Tyba-containing BACs. (**A-C**) CENH3 immunostaining in *R. pubera*, (**A**) metaphase, (**B**) interphase and (**C**) prophase. In metaphase CENH3 is only present densely within the centromere

groove (arrowheads). (**A-B**) SIM images. Scale bar: 5µm. (**D**) Southern blot hybridization of different *Rhynchospora* species (*R. ciliata*, *R. pubera* and *R. tenuis*) and other genera of Cyperaceae (*Cyperus aggregatus*, *Scleria bracteata*, *Scirpoides holoschoenus* and *Carex flacca*) with Tyba1 repeat amplified from *R. tenuis*. Numbers 1-4 on top represent different enzyme concentrations (0.3 U, 0.6 U, 1U and 5U of *Dra*I, respectevely). (**E-F**) Immuno-FISH colocalization of both CENH3 and Tyba in metaphase chromosomes of *Rhynchospora* species, SIM images of *R. tenuis* (2n = 4) (**E**) and of *R. ciliata* (2n = 10) (**F**) (scale bar: 5 µm). (**G**) Semi-quantitative RT-PCR reveals the transcriptional activity of Tyba1 and 2 in all tissues analyzed. GAPDH was used as control. (**H**) Annotation of *R. pubera* BACs containing centromeric repeats. RpBAC9H8 shows a ~3 kb Tyba array very close to the protein domain region of a Pararetrovirus, as well as a hAT DNA transposon, a Ty3/gypsy LTR retrotransposon of Ogre/Tat clade, other TE related proteins and single copy coding sequences. RpBAC8P1 shows a ~10 kb Tyba array flanked on both sides by MITE-like sequences with a centromeric retrotransposon on the neighborhood, a Ty1/copia LTR retrotransposon of Maximus/SIRE clade, a Ty3/gypsy LTR retrotransposon of Ogre/Tat clade, and other single copy coding sequences. RpBAC23M1 shows a ~17 kb Tyba array, a putative LTR-related region with a Tyba-like insertion and other single copy coding sequences. RpBAC3H4 shows a ~16 kb Tyba array, a MuDR-like DNA transposon, a Ty3/gypsy LTR retrotransposon of Athila clade and additional single copy coding sequences. RpBAC22N8 shows a ~12 kb Tyba array, a Ty3/gypsy LTR retrotransposon of Athila clade and additional single copy coding sequences. RpBAC23H8 shows a ~12 kb Tyba array with an apparently degenerated region and additional single copy coding sequences.

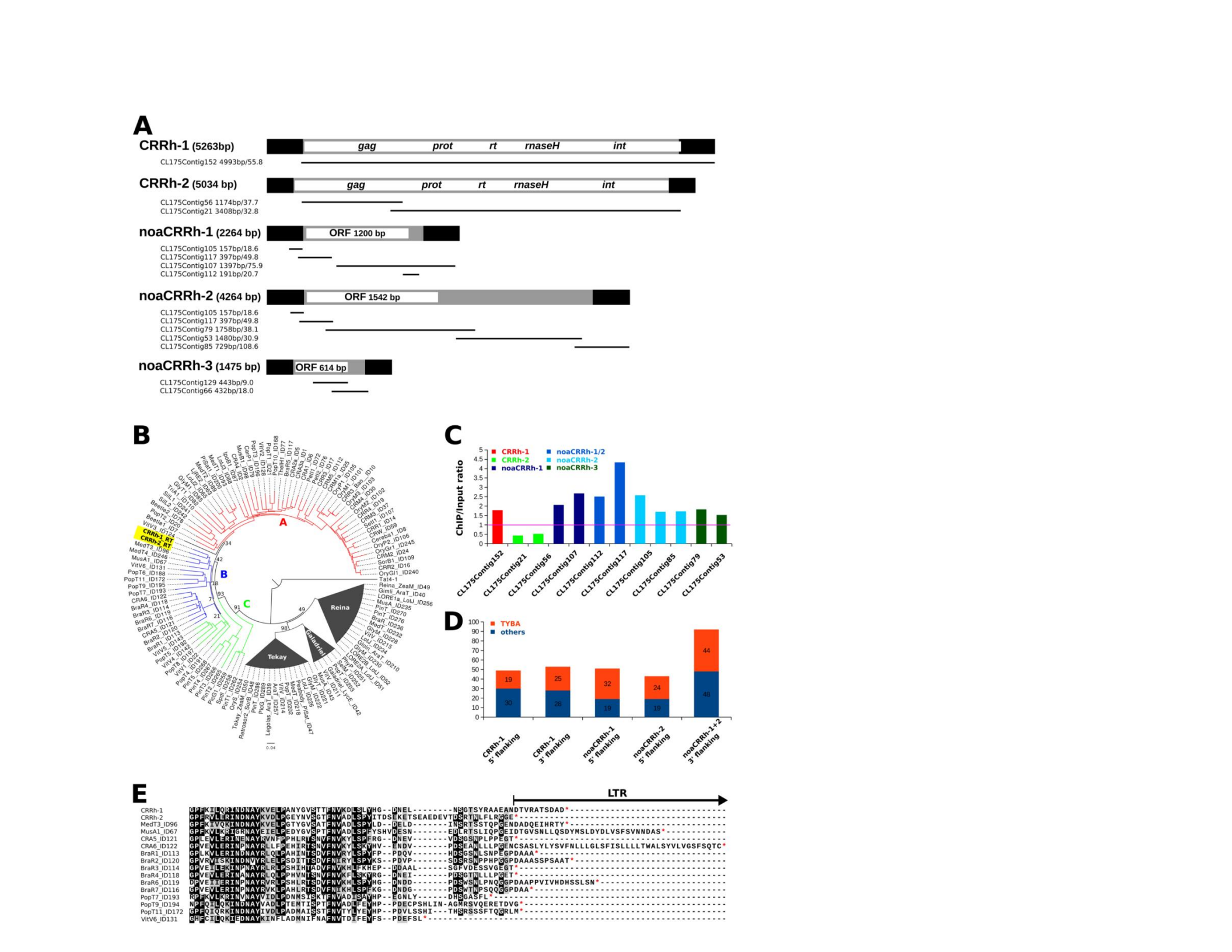
A
CRRh-1 (5263bp)
gag
prot
rt
rnaseH
int
CL175Contig152 4993bp/55.8
CRRh-2 (5034 bp)
gag
prot
rt
rnaseH
int
CL175Contig56 1174bp/37.7
CL175Contig21 3408bp/32.8
noaCRRh-1 (2264 bp)
ORF 1200 bp
CL175Contig105 157bp/18.6
CL175Contig117 397bp/49.8
CL175Contig107 1397bp/75.9
CL175Contig112 191bp/20.7
noaCRRh-2 (4264 bp)
ORF 1542 bp
CL175Contig105 157bp/18.6
CL175Contig117 397bp/49.8
CL175Contig79 1758bp/38.1
CL175Contig53 1480bp/30.9
CL175Contig85 729bp/108.6
noaCRRh-3 (1475 bp)
ORF 614 bp
CL175Contig129 443bp/9.0
CL175Contig66 432bp/18.0
B
A
B
C
Reina
Tekay
Galadriel
CRRh-1_RT
CRRh-2_RT
C
CRRh-1
CRRh-2
noaCRRh-1
noaCRRh-1/2
noaCRRh-2
noaCRRh-3
ChIP/Input ratio
CL175Contig152
CL175Contig21
CL175Contig56
CL175Contig107
CL175Contig112
CL175Contig117
CL175Contig105
CL175Contig85
CL175Contig79
CL175Contig53
D
TYBA
others
CRRh-1 5' flanking
CRRh-1 3' flanking
noaCRRh-1 5' flanking
noaCRRh-2 5' flanking
noaCRRh-1+2 3' flanking
E
LTR
CRRh-1
CRRh-2
MedT3_ID96
MusA1_ID67
CRA5_ID121
CRA6_ID122
BraR1_ID113
BraR2_ID120
BraR3_ID114
BraR4_ID118
BraR6_ID119
BraR7_ID116
PopT7_ID193
PopT9_ID194
PopT11_ID172
VitV6_ID131

Fig. S4. Features of CRRh elements of *R. pubera*. (**A**) Schematics of CRRh elements. LTRs are shown in black, internal fragments in gray and ORF in white. Lines below the schemes show positions of the most representative contigs that were used to reconstruct sequences of full length elements. (**B**) Neighbor-joining tree inferred from a comparison of RT domain sequences. It demonstrates that CRRh-1 and CRRh-2 elements belong to CRM clade of chromoviruses, being most similar to those that form the group B. Classification of CRM into groups A, B, and C is based on differences at the C-terminus of integrase (3). The non-chromovirus element Tat4-1 was used as an outgroup, while members of the Tekay, Reina, and Galadriel clades were included as representatives of non-CRM clade chromoviruses. (**C**) A chart showing ChIP-enrichment calculated for contigs representing fragments of different CRRh elements. (**D**) Analysis of insertion sites sequences of CRRh elements revealed that CRRh-1, noaCRRh-1 and noaCRRh-2 integrates frequently into Tyba. Sequences at insertion sites of CRRh-2 and noaCRRh-3 could not be analyzed because LTR sequences were variable and shared similarity with other repeats. (**E**) Alignment of sequences at the C-terminus of integrase. Only sequences from CRM group B elements are included (3). Characteristic feature of this group of CRM elements is the absence of PTD domain at integrase C-terminus and termination of the coding region around the start of 3' LTR (3). This is in contrast to CRM group A elements having the coding region extended deeply into 3’ LTR and encoding for PTD domain (3). Stop codons at the end of each open reading frame are indicated by red asterisks. Beginning of 3’ LTR is depicted as an arrow above the alignment. Arrow shows a part of integrase which is encoded by sequence in the 3’ LTR. RT-domain sequences used for the phylogenetic analysis were obtained from the study of (3).

Supplementary Tables

Table S1. Repetitive DNA composition of the *R. pubera* genome

			Analyzed reads:	8,032,451
Repeat class	Subclass	Family	Genome [%]	(group sum)
LTR retrotransposons				17.13
	Ty1/copia	Angela	5.79	8.68
		Maximus	1.17	
		AleII	0.91	
		Tork	0.21	
		TAR	0.20	
		Bianca	0.19	
		Ivana	0.19	
		AleI	0.01	
	Ty3/gypsy	Athila	2.54	5.14
		Tat/Ogre	1.86	
		chromovirus	0.74	
	unclassified		3.32	
Non-LTR retrotransposons	LINE		0.55	0.72
	SINE		0.17	
Pararetrovirus			0.51	
DNA transposons	MITE		5.10	8.81
	hAT		0.84	
	unclassified		2.87	

Helitron	0.59
TEs unclassified	2.69
Others	
Satellite DNA	3.60
rDNA	0.70
Classified repeats	34.75
Total repeats >= 0.01%	41.16

Table S2. BAC annotation

RpBAC3H4				
Name	Type	Start base	End base	Length
not annotated	gene	2,138	3,256	1,119
Translational activator GCN1	gene	63,575	87,529	23,955
Putative LRR receptor-like serine/threonine-protein kinase	gene	122,120	133,775	11,656
not annotated	gene	143,214	144,067	854
ATP synthase protein I-related protein	gene	156,388	160,107	3,720
polyadenylate binding protein	gene	165,600	171,622	6,023
CMP-KDO synthetase	gene	177,295	185,236	7,942
Ty3/gypsy Athila	LTR retrotransposon	91,290	99,919	8,630
putative DNA transposon - MuDR	transposable element protein	35,290	40,298	5,009
Tyba	Tyba_array	13,333	28,723	15,391
862-nucleotide Repeat (Tyba HOR pentamer)	Tyba_array	14,456	16,813	2,358
844-nucleotide Repeat (Tyba HOR pentamer)	Tyba_array	19,976	22,426	2,451
862-nucleotide Repeat (Tyba HOR pentamer)	Tyba_array	23,488	25,286	1,799
863-nucleotide Repeat (Tyba HOR pentamer)	Tyba_array	26,126	28,113	1,988
Total length (kb) obtained in the assembly				187

Total length (kb) obtained by PFGE				140
RpBAC8P1				
Name	Type	Start base	End base	Length
UDP-glucosyltransferase	gene	37,320	38,735	1,416
virus-like coat protein	gene	84,288	85,301	1,014
LTR Ty1/copia Maximus/SIRE	LTR retrotransposon	11,053	16,873	5,821
LTR	LTR retrotransposon	18,270	19,178	909
LTR Ty3/gypsy Ogre/Tat	LTR retrotransposon	49,742	54,452	4,711
CR - chromovirus from CRM clade	LTR retrotransposon	95,761	101,062	5,302
putative MITE	MITE	65,171	66,805	1,635
putative MITE	MITE	76,391	77,519	1,129
Tyba	Tyba_array	66,822	76,329	9,508
334-nucleotide Repeat (Tyba HOR dimer)	Tyba_array	67,857	69,106	1,250
837-nucleotide Repeat (Tyba HOR pentamer)	Tyba_array	70,311	72,011	1,701
843-nucleotide Repeat (Tyba HOR pentamer)	Tyba_array	72,000	75,112	3,113
Total length (kb) obtained in the assembly				101
Total length (kb) obtained by PFGE				110
RpBAC9H8				
Name	Type	Start base	End base	Length
Peptidase Gluzincin family	gene	33,747	38,766	5,020

DYW family of nucleic acid deaminases	gene	122,112	128,542	6,431
LTR	LTR retrotransposon	5,126	6,494	1,369
Ty3/gypsy Ogre/Tat	LTR retrotransposon	68,135	82,856	14,722
Ty3/gypsy GAG	LTR retrotransposon	94,476	95,861	1,386
RT domain	LTR retrotransposon	116,924	118,979	2,056
Pararetrovirus	Pararetrovirus	15,855	21,257	5,403
hAT DNA Transposon	transposable element protein	42,447	45,458	3,012
Plant mobile domain - transposase	transposable element protein	84,413	93,547	9,135
Tyba	Tyba_array	6,473	9,853	3,381
851-nucleotide Repeat (Tyba HOR pentamer)	Tyba_array	6,473	9,853	3,381
Total length (kb) obtained in the assembly				128
Total length (kb) obtained by PFGE				140

RpBAC17C8

Name	Type	Start base	End base	Length
YTH domain family protein 2	gene	<1	5,601	>5601
glucose-6-phosphate dehydrogenase	gene	7,692	22,403	14,712
putative LINE protein	LINE	22,544	26,598	4,055
LINE	LINE	67,335	101,028	33,694
retrotransposon sequence	LTR retrotransposon	26,738	28,937	2,200

Tekay-chromovirus-related	LTR retrotransposon	50,272	58,624	8,353
LTR Ty1/copia Angela	LTR retrotransposon	105,313	108,382	3,070
TE protein - endoribonuclease	transposable element protein	67,335	82,808	15,474
RT domain	transposable element protein	85,380	89,144	3,765
GAG domain	transposable element protein	89,519	91,219	1,701
TE protein	transposable element protein	92,671	101,028	8,358
Tyba	Tyba_array	29,272	32,400	3,129
Tyba	Tyba_array	33,983	36,060	2,078
Tyba	Tyba_array	38,483	41,922	3,440
862-nucleotide Repeat (Tyba HOR pentamer)	Tyba_array	29,281	32,353	3,073
870-nucleotide Repeat (Tyba HOR pentamer)	Tyba_array	34,002	35,967	1,966
837-nucleotide Repeat (Tyba HOR pentamer)	Tyba_array	38,511	41,887	3,377
Total length (kb) obtained in the assembly				108
Total length (kb) obtained by PFGE				120

RpBAC22N8

Name	Type	Start base	End base	Length
Myb/SANT-like DNA-binding domain-containing protein	gene	55,932	57,708	1,777

Borrelia P83/100 protein	gene	66,313	69,706	3,394
helicase	gene	72,958	75,765	2,808
auxin response factor 4-like	gene	88,036	89,341	1,306
LTR Ty3/gypsy Athila	LTR retrotransposon	11,254	22,129	10,876
Tyba	Tyba_array	41,857	53,749	11,893
860-nucleotide Repeat (Tyba HOR pentamer)	Tyba_array	41,977	53,539	11,563
Total length (kb) obtained in the assembly				89
Total length (kb) obtained by PFGE				90

RpBAC23H8

Name	Type	Start base	End base	Length
protein ROOT PRIMORDIUM DEFECTIVE 1 isoform X2	gene	1	13,987	13,987
BTB/POZ and TAZ domain-containing protein 1-like	gene	14,836	15,883	1,048
zinc finger BED domain-containing protein	gene	35,222	40,847	5,626
ribosomal protein L2	gene	47,828	49,201	1,374
TAZ zinc finger	gene	52,087	54,622	2,536
rRNA-processing protein EFG1-like	gene	62,711	66,483	3,773
Peptide methionine sulfoxide reductase	gene	85,345	94,589	9,245
Tyba	Tyba_array	71,297	75,867	4,571
Tyba	Tyba_array	76,479	77,066	588
Tyba	Tyba_array	77,622	78,096	475

Tyba	Tyba_array	78,945	83,144	4,200
858-nucleotide Repeat (Tyba HOR pentamer)	Tyba_array	71,313	75,859	4,547
172-nucleotide Repeat (Tyba monomer)	Tyba_array	76,483	77,060	578
170-nucleotide Repeat (Tyba monomer)	Tyba_array	77,631	78,084	454
860-nucleotide Repeat (Tyba HOR pentamer)	Tyba_array	78,971	83,055	4,085
Total length (kb) obtained in the assembly				97
Total length (kb) obtained by PFGE				110

RpBAC23M1

Name	Type	Start base	End base	Length
hAT family dimerisation protein	gene	1,423	1,854	432
kinase	gene	17,963	21,750	3,788
uncharacterized protein	gene	57,436	61,561	4,126
LINE	LINE	45,783	51,849	6,067
LTR retrotransposon-like	LTR	6,368	8,900	2,533
367-nucleotide Repeat (Tyba-like dimer)	repeat_region	6,632	7,370	739
Tyba	Tyba_array	23,918	40,387	16,470
362-nucleotide Repeat (Tyba HOR dimer)	Tyba_array	24,009	25,718	1,710
869-nucleotide Repeat (Tyba HOR pentamer)	Tyba_array	26,047	28,660	2,614
364-nucleotide Repeat (Tyba HOR dimer)	Tyba_array	28,809	29,822	1,014
836-nucleotide Repeat (Tyba HOR pentamer)	Tyba_array	29,964	33,250	3,287
361-nucleotide Repeat (Tyba HOR dimer)	Tyba_array	33,701	38,275	4,575

827-nucleotide Repeat (Tyba HOR pentamer)	Tyba_array	38,716	40,380	1,665
Total length (kb) obtained in the assembly				89
Total length (kb) obtained by PFGE				120

Table S3. Species name and collecting places

Species name	Collected places
Rhynchospora pubera (Vahl) Boeckler	Curado, Recife, Brazil
Rhynchospora tenuis Willd. ex Link	Curado, Recife, Brazil
Rhynchospora ciliata (Vahl) Kükenthal	Curado, Recife, Brazil
Carex flacca Schreb.	Gatersleben, Germany
Cyperus aggregatus (Willd.) Endl.	Curado, Recife, Brazil
	Dois Irmãos, Recife,
Scleria bracteata Cav.	Brazil
Scirpoides holoschoenus (L.) Soják	Gatersleben, Germany

Table S4. List of oligonucleotide probes and primer sequences used

Target	oligo/primer name	oligo/primer sequence	fluorescence
Tyba1	Tyba1 oligo probe	ATTGGATTATACATGGTAATTACGCATATAAAGTGCAAATAATGCAATTC	FAM
Tyba2	Tyba2 oligo probe	ACAGATTCTGAGTATATTTGAGCATTTCAAGCGATTTTGCATT	Cy3
MITE	MITE oligo probe	AATTTATTATAAACAATCCAAACTCTTCACAAAGTTACACACTTCCCAAT	FAM
Tyba primer1	Tyba1F	CTAAGTCATTTCATCACAATAATCTAC	none
	Tyba1R	AATCCAGAAACGATTGAAATGCTC	none
Tyba primer2	Tyba2F	GTGCAAATAATGCAATTCTGAGCATC	none
	Tyba2R	ATATGCGTAATTACCATGTATAATCC	none
RpCENH3	RpCENH3F	CTCCCTTTCTCACTCCTTGC	none
	RpCENH3R	CGATCAAAATTGAACCGCAACCAT	none
CRRh-1	CRRh-1F	GACTAATCATCCCAGCCATGT	none
	CRRh-1R	GTGGCTCGAACGGTGTC	none
CRRh-2	CRRh-2F	TATTTTACTTTTGTGCACGGTAGAC	none
	CRRh-2R	GTTAAAGCCCATGTTATGTTCG	none
		control primers	
GAPDH	GAPDH-F	CAATGATAGCTGCACCACCAACTG	none
	GAPDH-R	CTAGCTGCCCTTCCACCTCTCCA	none
26S ribosomal RNA	26S-F	CCTCCGAAGTTTCCCTCAG	none
	26s-R	GCCTCTAATCATTGGCTTTACCT	none

Table S5. Summary of Velvet assembly from the cDNA library of the pollen mother cell transcriptome of *R. pubera*

Statistics	**All Contigs**	**Contigs >=100 bp**	**Contigs >=1000 bp**
Number of contigs	75.353	74.120	20.290
Min Length (bp)	62	100	1.000
Median Length (bp)	491	502	1.594
Mean Length (bp)	794	806	1.847
Max Length (bp)	15.473	15.473	15.473
N50 Length (bp)	1.341	1.344	1.908
Number of contigs >= N50	13.783	13.747	6.770
Length Sum (bp)	59.848.133	59.751.873	37.478.675

Table S6. List of sequence identifiers and description of plant CENH3 sequences used

Sequence identifier	Database	Description
KR029618	Genbank	>Rhynchospora pubera CENH3_1
KR029619	Genbank	>Rhynchospora pubera CENH3_2
AF465801	Genbank	>gi|19338703|gb|AF465801.1| Arabidopsis arenosa centromeric histone H3 HTR12 (HTR11) gene, complete cds
AB081501	Genbank	>gi|33146135|dbj|AB081501.1| Arabidopsis halleri subsp. gemmifera gene for histone H3 like protein, complete cds, histone H2 like-a
AB081503	Genbank	>gi|33146139|dbj|AB081503.1| Arabidopsis halleri subsp. gemmifera gene for histone H3 like protein, complete cds, histone H2 like-b
DQ450587	Genbank	>gi|91179111|gb|DQ450587.1| Arabidopsis lyrata subsp. petraea strain Mt. Esja11 histone H3 (HTR11A) gene, complete cds
DQ450557	Genbank	>gi|91178276|gb|DQ450557.1| Arabidopsis lyrata subsp. petraea strain Mt. Esja11 histone H3 (HTR11B) gene, complete cds
AB081500	Genbank	>gi|33146133|dbj|AB081500.1| Arabidopsis thaliana mRNA for histone H2 like protein, complete cds
AB299169	Genbank	>gi|134152506|dbj|AB299169.1| Arabis hirsuta cenp-A gene for putative centromeric histone H3-like protein_0, complete cds
GU166738	Genbank	>gi|268376515|gb|GU166738.1| Brassica nigra isolate BrCENH3-2 centromere-specific H3 variant protein (CENH2) mRNA, complete cds
GU166739	Genbank	>gi|268376517|gb|GU166739.1| Brassica oleracea isolate BrCENH3-3 centromere-specific H3 variant protein (CENH2) mRNA, complete cds
GU166737	Genbank	>gi|268376513|gb|GU166737.1| Brassica rapa isolate BrCENH3-1 centromere-specific H3 variant protein (CENH2) mRNA, complete cds
AB299175	Genbank	>gi|134152518|dbj|AB299175.1| Capsella bursa-pastoris cenp-A gene for centromeric histone H3-like protein-0, complete cds
XM_00630418	Genbank	>gi|565493209|ref|XM_006304182.1| Capsella rubella hypothetical protein (CARUB_v10010435mg) mRNA,

2		complete cds
AB299171	Genbank	>gi|134152510|dbj|AB299171.1| Cardamine flexuosa cenp-A gene for centromeric histone H3-like protein-0, complete cds
AY612790	Genbank	>gi|51103316|gb|AY612790.1| Crucihimalaya himalaica centromeric histone (HTR11) gene, complete cds
AB299177	Genbank	>gi|134152522|dbj|AB299177.1| Crucihimalaya wallichii cenp-A gene for centromeric histone H2-like protein, complete cds
AB299180	Genbank	>gi|134152528|dbj|AB299180.1| Eruca sativa cenp-A gene for centromeric histone H3-like protein-0, complete cds
AB299181	Genbank	>gi|134152530|dbj|AB299181.1| Lepidium virginicum cenp-A gene for centromeric histone H3-like protein_0, complete cds
AB299167	Genbank	>gi|134152502|dbj|AB299167.1| Olimarabidopsis pumila cenp-A gene for putative centromeric histone H3-like protein_0, complete cds
AB299183	Genbank	>gi|134152534|dbj|AB299183.1| Raphanus sativus cenp-A gene for centromeric histone H3-like protein_0, complete cds
Thhalv10008889m	Phytozome	>Thhalv10008888m
AB081505	Genbank	>gi|33146143|dbj|AB081505.1| Turritis glabra gene for histone H2 like protein, complete cds
AB649144	Genbank	>gi|365799494|dbj|AB649144.1| Astragalus sinicus AsCENH3 mRNA for centromere specific histone H2 variant, complete cds
EX259948	Genbank	>gi|186738465|gb|EX259948.1|EX259948 1440421_5_E19_076 PY05 Carica papaya cDNA, mRNA sequence
XM_006491184	Genbank	>gi|568876356|ref|XM_006491184.1| PREDICTED: Citrus sinensis histone H3-like centromeric protein HTR12-like (LOC102614120), transcript variant X1, mRNA
XM_004139613	Genbank	>gi|449443791|ref|XM_004139613.1| PREDICTED: Cucumis sativus histone H3-like centromeric protein HTR12-like (LOC101210010), mRNA
Eucgr.D00189.1	Phytozome	>Eucgr.D00189.0
XM_004306639	Genbank	>gi|470141951|ref|XM_004306639.1| PREDICTED: Fragaria vesca subsp. vesca histone H3-like centromeric protein HTR12-like (LOC101294589), mRNA

XM_003528751	Genbank	>gi|571465058|ref|XM_003528751.2| PREDICTED: Glycine max histone H3-like centromeric protein HTR12-like (LOC100811871), mRNA
Gorai.001G155600.1	Phytozome	>Gorai.001G155600.0
Lus10008119	Phytozome	>Lus10008118
BT137822	Genbank	>gi|388499101|gb|BT137822.1| Lotus japonicus clone JCVI-FLLj-13P8 unknown mRNA
XP_008361919.1	Genbank	>gi|657953278|ref|XP_008361919.1| PREDICTED: histone H3-like centromeric protein cnp0 [Malus domestica]
FF379687	Genbank	>gi|182383948|gb|FF379687.1|FF379687 CASL069TF CASL Manihot esculenta cDNA 4', mRNA sequence
XM_003637685	Genbank	>gi|358347374|ref|XM_003637685.1| Medicago truncatula Histone H3 (MTR_100s0022) mRNA, complete cds
KC491791	Genbank	>gi|523371675|gb|KC491791.1| Phaseolus vulgaris centromere specific histone H3 variant (CENH2) mRNA, complete cds
JF739989	Genbank	>gi|371486399|gb|JF739989.1| Pisum sativum centromere-specific variant of histone H3 type 0 gene, complete cds
JF739990	Genbank	>gi|371486401|gb|JF739990.1| Pisum sativum centromere-specific variant of histone H3 type 1 gene, complete cds
XM_002320818	Genbank	>gi|224130507|ref|XM_002320818.1| Populus trichocarpa centromeric histone H3 HTR12 family protein (POPTR_0014s09209g) mRNA, complete cds
XM_007051531	Genbank	>gi|590721367|ref|XM_007051531.1| Theobroma cacao Histone superfamily protein, putative isoform 1 (TCM_005175) mRNA, complete cds
XM_002281037	Genbank	>gi|731421864|ref|XM_002281037.2| PREDICTED: Vitis vinifera histone H3-like centromeric protein HTR12 (LOC100260234), transcript variant X1, mRNA
GR117778	Genbank	>gi|238366551|gb|GR117778.1|GR117778 CCBG8245.g1 CCBG Mimulus guttatus IM62 floral buds (H+L) Erythranthe guttata cDNA clone CCBG8245 2', mRNA sequence
AB467328	Genbank	>gi|218744595|dbj|AB467328.1| Nicotiana sylvestris NsCENH3 mRNA for centromere specific histone H2 variant, complete cds
AB467329	Genbank	>gi|218744597|dbj|AB467329.1| Nicotiana tomentosiformis NtoCENH3 mRNA for centromere specific histone H2

		variant, complete cds
BG127218	Genbank	>gi|12627406|gb|BG127218.1|BG127218 EST472864 tomato shoot/meristem Solanum lycopersicum cDNA clone cTOF14D12 4' sequence, mRNA sequence
XM_006339625	Genbank	>gi|565345203|ref|XM_006339625.1| PREDICTED: Solanum tuberosum histone H3-like centromeric protein HTR12-like (LOC102603326), mRNA
AB600275	Genbank	>gi|371940020|dbj|AB600275.1| Allium cepa AceCENH3 mRNA for centromere specific histone H2 variant, complete cds
AB571555	Genbank	>gi|371940014|dbj|AB571555.1| Allium fistulosum AfiCENH3 mRNA for centromere specific histone H2 variant, complete cds
AB571556	Genbank	>gi|371940016|dbj|AB571556.1| Allium sativum AsaCENH3 mRNA for centromere specific histone H2 variant, complete cds
AB571557	Genbank	>gi|371940018|dbj|AB571557.1| Allium tuberosum AtuCENH3 mRNA for centromere specific histone H2 variant, complete cds
XM_003566059	Genbank	>gi|357128898|ref|XM_003566059.1| PREDICTED: Brachypodium distachyon uncharacterized LOC100830307 (LOC100830306), mRNA
GU245882	Genbank	>gi|282895619|gb|GU245882.1| Hordeum bulbosum centromeric histone H3 (CENH2) mRNA, partial cds
JF419330	Genbank	>gi|339836913|gb|JF419330.1| Hordeum bulbosum beta centromeric histone H3 (CENH2) mRNA, complete cds
JF419328	Genbank	>gi|339836909|gb|JF419328.1| Hordeum vulgare subsp. vulgare alpha centromeric histone H3 (CENH2) mRNA, partial cds
JF419329	Genbank	>gi|339836911|gb|JF419329.1| Hordeum vulgare subsp. vulgare beta centromeric histone H3 (CENH2) mRNA, complete cds
AB201356	Genbank	>gi|90652790|dbj|AB201356.2| Luzula nivea LnCENH3 mRNA for centromere specific histone H2 variant, complete cds
HM988988	Genbank	>gi|304277059|gb|HM988988.1| Luzula nivea centromeric histone H3 isoform B (CENH2-B) mRNA, complete cds
AY438639	Genbank	>gi|40365139|gb|AY438639.1| Oryza sativa (japonica cultivar-group) centromeric histone 3 (CenH2) mRNA, complete cds
FL730019	Genbank	>gi|197988596|gb|FL730019.1|FL730019 CCGB10819.g1 CCGB Panicum virgatum apex + stem (L) Panicum

		virgatum cDNA clone CCGB10819 2', mRNA sequence
CA127217	Genbank	>gi|35006880|gb|CA127217.1|CA127217 SCCCLR2004A05.g LR2 Saccharum hybrid cultivar SP80-3280 cDNA clone SCCCLR2004A05 4', mRNA sequence
XM_004961625	Genbank	>gi|514748700|ref|XM_004961625.1| PREDICTED: Setaria italica histone H3-like centromeric protein HTR12-like (LOC101784681), transcript variant X1, mRNA
XM_002441245	Genbank	>gi|242090914|ref|XM_002441245.0| Sorghum bicolor hypothetical protein, mRNA
NM_001112050	Genbank	>gi|162460347|ref|NM_001112050.1| Zea mays centromeric histone H3 (cenH2), mRNA
AEH95350.1	Genbank	>gi|336041546|gb|AEH95350.1| centromeric histone 3 [Triticum aestivum]
AEH95351.1	Genbank	>gi|336041548|gb|AEH95351.1| centromeric histone 3 [Triticum aestivum]
ACZ04985.1	Genbank	>gi|268376530|gb|ACZ04985.1| centromere-specific H3 variant protein [Brassica napus]
ACZ04980.1	Genbank	>gi|268376520|gb|ACZ04980.1| centromere-specific H3 variant protein [Brassica juncea]
ACZ04982.1	Genbank	>gi|268376524|gb|ACZ04982.1| centromere-specific H3 variant protein [Brassica carinata]
KJ201906.1	Genbank	>gi|656991661|gb|KJ201906.1| Daucus glochidiatus centromeric histone H3 (CENH3) mRNA, partial cds
AHW98233.1	Genbank	>gi|612176255|gb|AHW98233.1| centromeric histone 3, partial [Cicer reticulatum]
XM_009346944.1	Genbank	>gi|694436201|ref|XM_009346944.1| PREDICTED: Pyrus x bretschneideri histone H3-like centromeric protein cnp1 (LOC103937036), transcript variant X2, mRNA
XM_008234980.1	Genbank	>gi|645254791|ref|XM_008234980.1| PREDICTED: Prunus mume histone H3-like centromeric protein HTR12 (LOC103332255), mRNA
XM_010482040.1	Genbank	>gi|727429465|ref|XM_010482040.1| PREDICTED: Camelina sativa histone H3-like centromeric protein HTR12 (LOC104759068), transcript variant X1, mRNA
XM_010482047.1	Genbank	>gi|727429467|ref|XM_010482047.1| PREDICTED: Camelina sativa histone H3-like centromeric protein HTR12 (LOC104759068), transcript variant X2, mRNA
KJ507236.1	Genbank	>gi|612176248|gb|KJ507236.1| Cajanus scarabaeoides isolate Pigeonpea_ICP 15731 centromeric histone 3 (CenH3) mRNA, partial cds
XM_01055829	Genbank	>gi|729412033|ref|XM_010558290.1| PREDICTED: Tarenaya hassleriana histone H3-like centromeric protein

0.1		HTR12 (LOC104825875), transcript variant X2, mRNA
XM_010558289.1	Genbank	>gi|729412030|ref|XM_010558289.1| PREDICTED: Tarenaya hassleriana histone H3-like centromeric protein HTR12 (LOC104825875), transcript variant X1, mRNA
KJ507233.1	Genbank	>gi|612176242|gb|KJ507233.1| Cajanus cajan isolate Pigeonpea_ICPL 87119_Asha centromeric histone 3 (CenH3) mRNA, complete cds
KJ507235.1	Genbank	>gi|612176246|gb|KJ507235.1| Cajanus cajanifolius isolate Pigeonpea_ICP 15631 centromeric histone 3 (CenH3) mRNA, complete cds
XM_006418354.1	Genbank	>gi|567156644|ref|XM_006418354.1| Eutrema salsugineum hypothetical protein (EUTSA_v10008889mg) mRNA, complete cds
XM_011039624.1	Genbank	>gi|743790056|ref|XM_011039624.1| PREDICTED: Populus euphratica histone H3-like centromeric protein HTR12 (LOC105134977), mRNA
AB793503.1	Genbank	>gi|586941098|dbj|AB793503.1| Torenia fournieri TfCENH3 mRNA for centromere specific histone H3, complete cds
AB793504.1	Genbank	>gi|586941100|dbj|AB793504.1| Torenia baillonii TbCENH3 mRNA for centromere specific histone H3, complete cds
XM_011094744.1	Genbank	>gi|747090693|ref|XM_011094744.1| PREDICTED: Sesamum indicum histone H3-like centromeric protein HTR12 (LOC105173094), mRNA
XM_010263215.1	Genbank	>gi|720017605|ref|XM_010263215.1| PREDICTED: Nelumbo nucifera histone H3-like centromeric protein HTR12 (LOC104600331), transcript variant X2, mRNA
XM_010268168.1	Genbank	>gi|720033563|ref|XM_010268168.1| PREDICTED: Nelumbo nucifera histone H3-like centromeric protein HTR12 (LOC104603975), transcript variant X1, mRNA
KJ201904.1	Genbank	>gi|656991657|gb|KJ201904.1| Daucus muricatus centromeric histone H3 (CENH3) mRNA, partial cds
KJ201903.1	Genbank	>gi|656991655|gb|KJ201903.1| Daucus carota centromeric histone H3 (CENH3) mRNA, partial cds
KJ201905.1	Genbank	>gi|656991659|gb|KJ201905.1| Daucus pusillus centromeric histone H3 (CENH3) mRNA, partial cds
XM_010696392.1	Genbank	>gi|731365747|ref|XM_010696392.1| PREDICTED: Beta vulgaris subsp. vulgaris histone H3 (LOC104907459), mRNA
KJ507243.1	Genbank	>gi|612176262|gb|KJ507243.1| Cicer yamashitae isolate Chickpea_ICC 17117 centromeric histone 3 (CenH3)

		mRNA, complete cds
XM_008794232.1	Genbank	>gi|672137449|ref|XM_008794232.1| PREDICTED: Phoenix dactylifera histone H3-like (LOC103709056), mRNA
HQ123579.1	Genbank	>gi|313104721|gb|HQ123579.1| Oryza latifolia isolate DD centromeric histone 3 (CenH3) mRNA, complete cds
XM_009413514.1	Genbank	>gi|260072772|gb|GQ849341.1| Oryza australiensis isolate EE CENH3 (CenH3) mRNA, complete cds
AB649144.1	Genbank	>gi|695047791|ref|XM_009413514.1| PREDICTED: Musa acuminata subsp. malaccensis histone H3-like centromeric protein cnp1 (LOC103993441), mRNA
XM_010933196.1	Genbank	>gi|743819355|ref|XM_010933196.1| PREDICTED: Elaeis guineensis histone H3.3 type c-like (LOC105052399), transcript variant X1, mRNA
XM_010933197.1	Genbank	gi|743819358|ref|XM_010933197.1| PREDICTED: Elaeis guineensis histone H3.3 type c-like (LOC105052399), transcript variant X2, mRNA
AB770163.1	Genbank	>gi|670453236|dbj|AB770163.1| Cenchrus americanus CENH3 mRNA for centromeric histone H3 isoform a, complete cds
AB770164.1	Genbank	>gi|670453238|dbj|AB770164.1| Cenchrus americanus CENH3 mRNA for centromeric histone H3 isoform b, complete cds
AB981585.1	Genbank	>gi|745991781|dbj|AB981585.1| Avena sativa AsCENH3-2 mRNA for centromere specific histone H3, partial cds
AB981584.1	Genbank	>gi|745991779|dbj|AB981584.1| Avena sativa AsCENH3-1 mRNA for centromere specific histone H3, partial cds
XM_008463887.1	Genbank	>gi|659124344|ref|XM_008463887.1| PREDICTED: Cucumis melo histone H3-like centromeric protein HTR12 (LOC103500538), partial mRNA
XM_010101295.1	Genbank	>gi|703110467|ref|XM_010101295.1| Morus notabilis Histone H3-like centromeric protein partial mRNA
XM_010232934.1	Genbank	>gi|721637023|ref|XM_010232934.1| PREDICTED: Brachypodium distachyon histone H3-like (LOC100830307), mRNA
XM_001785914	Genbank	>gi|168068179|ref|XM_001785914.1| Physcomitrella patens subsp. patens histone H3 (HTR1515) mRNA, complete cds
XM_00169781	Genbank	>gi|159479581|ref|XM_001697817.1| Chlamydomonas reinhardtii strain CC-503 cw91 mt+

7		
XM_00564732	Genbank	>gi|545364669|ref|XM_005647320.1| Coccomyxa subellipsoidea C-169 histone-fold-containing protein
0		(COCSUDRAFT_24063) mRNA, complete cds
XM_00305646	Genbank	>gi|303274379|ref|XM_003056465.1| Micromonas pusilla CCMP1545 histone H2, mRNA
5		
NM_001809.3	Genbank	>gi|109637780|ref|NM_001809.3| Homo sapiens centromere protein A (CENPA), transcript variant 0, mRNA
AF093633.1	Genbank	>gi|3885889|gb|AF093633.1| Oryza sativa histone H2 mRNA, complete cds
NM_125934.2	Genbank	>gi|30698117|ref|NM_125934.2| Arabidopsis thaliana histone H3.0 mRNA, complete cds
X57128.1	Genbank	>gi|31981|emb|X57128.1| H.sapiens H3.1 gene for H2 histone

SI References

1. Ruban A*, et al.* (2014) B chromosomes of *Aegilops speltoides* are enriched in organelle genome-derived sequences. *PloS one* 9(2):e90214.
2. Marques A*, et al.* (2011) Characterization of eu- and heterochromatin of *Citrus* with a focus on the condensation behavior of 45S rDNA chromatin. *Cytogenetic and genome research* 134(1):72-82.
3. Kato A, Albert PS, Vega JM, & Birchler JA (2006) Sensitive fluorescence *in situ* hybridization signal detection in maize using directly labeled probes produced by high concentration DNA polymerase nick translation. *Biotechnic & histochemistry : official publication of the Biological Stain Commission* 81(2-3):71-78.
4. Ma L*, et al.* (2010) Synteny between *Brachypodium distachyon* and *Hordeum vulgare* as revealed by FISH. *Chromosome Research* 18(7):841-850.
5. Banaei-Moghaddam AM, Meier K, Karimi-Ashtiyani R, & Houben A (2013) Formation and expression of pseudogenes on the B chromosome of rye. *The Plant cell* 25(7):2536-2544.
6. Zerbino DR & Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome research* 18(5):821-829.
7. Kearse M*, et al.* (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28(12):1647-1649.
8. Li L, Yang J, Tong Q, Zhao L, & Song Y (2005) A novel approach to prepare extended DNA fibers in plants. *Cytometry. Part A : the journal of the International Society for Analytical Cytology* 63(2):114-117.
9. Galbraith DW*, et al.* (1983) Rapid flow cytometric analysis of the cell cycle in intact plant tissues. *Science* 220(4601):1049-1051.

10. Arumuganathan K, Slattery JP, Tanksley SD, & Earle ED (1991) Preparation and flow cytometric analysis of metaphase chromosomes of tomato. *Theoretical and Applied Genetics* 82(1):101-111.
11. Ishii T, Sunamura N, Matsumoto A, Eltayeb AE, & Tsujimoto H (2015) Preferential recruitment of the maternal centromere-specific histone H3 (CENH3) in oat (*Avena sativa* L.) x pearl millet (*Pennisetum glaucum* L.) hybrid embryos. *Chromosome Research*.
12. Doležel J, Binarova P, & Lucretti S (1989) Analysis of nuclear DNA content in plant cells by flow-cytometry. *Biologia Plantarum* 31(2):113-120.
13. Sambrook J, Fritsch EF, & Maniatis T (1989) Analysis of genomic DNA by southern hybridization and northern hybridization. *Molecular Cloning: A Laboratory Manual*, (Cold Spring Harbor Press: Cold Spring Harbor, NY, USA), 2nd Ed, pp E21–E25.
14. Doležel J, Číhalíková J, & Lucretti S (1992) A high-yield procedure for isolation of metaphase chromosomes from root tips of *Vicia faba* L. *Planta* 188(1):93-98.
15. Šimková H, Číhalíková J, Vrána J, Lysák MA, & Doležel J (2003) Preparation of HMW DNA from plant nuclei and chromosomes isolated from root tips. *Biologia Plantarum* 46(3):369-373.
16. Šimková H*, et al.* (2011) BAC libraries from wheat chromosome 7D: efficient tool for positional cloning of aphid resistance genes. *Journal of biomedicine & biotechnology* 2011:302543.
17. Ming R*, et al.* (2001) Construction and characterization of a papaya BAC library as a foundation for molecular dissection of a tree-fruit genome. *Theoretical and Applied Genetics* 102:892-899.
18. Meyer M & Kircher M (2010) Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor protocols* 2010(6):pdb prot5448.
19. Novak P, Neumann P, & Macas J (2010) Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. *BMC bioinformatics* 11:378.
20. Novak P, Neumann P, Pech J, Steinhaisl J, & Macas J (2013) RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. *Bioinformatics* 29(6):792-793.
21. Macas J, Neumann P, Novak P, & Jiang J (2010) Global sequence characterization of rice centromeric satellite based on oligomer frequency analysis in large-scale sequencing data. *Bioinformatics* 26(17):2101-2108.
22. Altschul SF*, et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research* 25(17):3389-3402.
23. Kuhlmann M & Mette MF (2012) Developmentally non-redundant SET domain proteins SUVH2 and SUVH9 are required for transcriptional gene silencing in Arabidopsis thaliana. *Plant Mol Biol* 79(6):623-633.
24. Chevreux B, Wetter T, & Suhai S (1999) Genome sequence assembly using trace signals and additional sequence information. *Computer Science and Biology: Proceedings of the German Conference on Bioinformatics* 99:45-56.
25. Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC bioinformatics* 5:113.
26. Nguyen LT, Schmidt HA, von Haeseler A, & Minh BQ (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular biology and evolution* 32(1):268-274.
27. Minh BQ, Nguyen MA, & von Haeseler A (2013) Ultrafast approximation for phylogenetic bootstrap. *Molecular biology and evolution* 30(5):1188-1195.
28. Neumann P*, et al.* (2011) Plant centromeric retrotransposons: a structural and cytogenetic perspective. *Mobile DNA* 2(1):4.