Meiotic recombination dynamics in plants with repeat-based holocentromeres shed light on the primary drivers of crossover patterning

André Marques (amarques@mpipz.mpg.de)
Max Planck Institute for Plant Breeding Research
https://orcid.org/0000-0002-9567-2576

Marco Castellani
Max Planck Institute for Plant Breeding Research

Meng Zhang
Max Planck Institute for Plant Breeding Research
https://orcid.org/0000-0003-4792-9979

Gokilavani Thangavel
Max Planck Institute for Plant Breeding Research

Yennifer Mata Sucre
Federal University of Pernambuco

Thomas Lux
Helmholtz Zentrum München
https://orcid.org/0000-0002-5543-1911

José Campoy
Universidad Politécnica de Cartagena
https://orcid.org/0000-0002-6018-5698

Magdalena Marek
Max Planck Institute for Plant Breeding Research

Bruno Hüttel
Max Planck Institute for Plant Breeding Research
https://orcid.org/0000-0001-7165-1714

Hequan Sun
Max Planck Institute for Plant Breeding Research
https://orcid.org/0000-0003-2046-2109

Klaus Mayer
Helmholtz Center Munich
https://orcid.org/0000-0001-6484-1077

Korbinian Schneeberger
Ludwig-Maximilians-Universität München
https://orcid.org/0000-0002-5512-0443

Article

Keywords: centromere effect, crossover, meiosis, holocentromere, single-cell sequencing

Posted Date: September 14th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3266957/v1
Meiotic recombination dynamics in plants with repeat-based holocentromeres shed light on the primary drivers of crossover patterning

Authors:
Marco Castellani¹,¹,#, Meng Zhang¹,¹,#, Gokilavani Thangavel¹, Yennifer Mata Sucre¹,², Thomas Lux³, José A. Campoy¹,⁸, Magdalena Marek⁴, Bruno Huettel⁴, Hequan Sun¹,⁵,⁹, Klaus F. X. Mayer³,⁶, Korbinian Schneeberger¹,⁵,⁷ and André Marques¹,*

¹Department of Chromosome Biology, Max Planck Institute for Plant Breeding Research, Carl-von-Linné-Weg 10, NRW 50829 Cologne, Germany
²Laboratory of Plant Cytogenetics and Evolution, Department of Botany, Centre of Biosciences, Federal University of Pernambuco, 50670-901 Recife, Brazil
³Plant Genome and Systems Biology, German Research Centre for Environmental Health, Helmholtz Zentrum München, Ingolstädter Landstraße 1, 85764 Neuherberg, Germany
⁴Max Planck Genome-Centre Cologne, Max Planck Institute for Plant Breeding Research, Carl-von-Linné-Weg 10, NRW 50829 Cologne, Germany
⁵Ludwig-Maximilians-University of Munich, Faculty of Biology, Biocenter Großhaderner Str. 2, 82152 Planegg-Martinsried, Germany
⁶School of Life Sciences Weihenstephan, Technical University of Munich, Alte Akademie 8, 85354 Freising, Germany
⁷Cluster of Excellence on Plant Sciences (CEPLAS), Heinrich-Heine University, Universitätsstraße 1, 40225 Düsseldorf, Germany
⁸Current address: Departamento de Ingeniería Agronómica, Instituto de Biotecnología Vegetal, Universidad Politécnica de Cartagena, Paseo Alfonso XIII, 48, 30203, Cartagena, Spain
⁹Current address: School of Automation Science and Engineering, Faculty of Electronic and Information Engineering, Xi’an Jiaotong Univeristy, Xi’an, 710049 Shaanxi, China

#equal contribution
*corresponding author: amarques@mpipz.mpg.de
Abstract

Centromeres strongly affect (epi)genomic architecture and meiotic recombination dynamics influencing the overall distribution and frequency of crossovers. Here, we studied how recombination is regulated and distributed in the holocentric plant *Rhynchospora breviuscula*, a species lacking localised centromeres. Combining immunocytochemistry, chromatin analysis and high-throughput single-pollen sequencing, we discovered that crossover frequency is distally-biased, in sharp contrast with the diffused distribution of hundreds of centromeric units and (epi)genomic features. Remarkably, we found that crossovers were abolished inside centromeric units but not in their proximity indicating the absence of a canonical centromere effect. We further show that telomere-led synopsis of homologs is the feature that best explains the observed recombination landscape. Our results hint at the primary influence of mechanistic features of meiotic pairing and synapsis rather than (epi)genomic features and centromere organisation in determining the distally-biased crossover distribution in *R. breviuscula*. While centromeres and (epi)genetic properties only affect crossover positioning locally.

Keywords: centromere effect, crossover, meiosis, holocentromere, single-cell sequencing
During meiosis, homologous chromosomes (homologs) undergo meiotic recombination, in which genomic material is exchanged between homologs. This exchange is initiated by physiologically induced DNA double-strand breaks (DSBs)\textsuperscript{1,2}. The formation of meiotic DSBs is commonly resolved via crossovers (COs) as well as other recombination outcomes, referred to as non-COs (NCOs)\textsuperscript{3}. COs can be divided into two classes, although the existence of alternative CO pathways cannot be excluded\textsuperscript{4,5}. Class I COs are the most prevalent and are sensitive to interference, i.e., they do not occur near each other along a chromosome. Class I COs result from the ZMM pathway that includes the key factor Human enhancer of invasion-10 (HEI10), involved in CO designation, and ZYP1, a key protein involved in synaptonemal complex (SC) assembly\textsuperscript{6–12}. Class II COs are insensitive to interference and accommodate around 10% of the total COs in \textit{Arabidopsis thaliana}\textsuperscript{13}. Although the repertoire of meiotic-specific proteins is largely conserved across eukaryotes\textsuperscript{14,15}, species-specific adaptations can occur\textsuperscript{16}.

The global distribution of COs is typically associated with the distribution of genetic and epigenetic [(epi)genetic] features\textsuperscript{17,18}. In most eukaryotes, gene/euchromatin density positively correlates with CO frequency\textsuperscript{19,20}. By contrast, CO frequency is typically lower in heterochromatic regions, including at (peri)centromeres\textsuperscript{21,22}. In monocentric species, centromeres are single, defined structural entities and are typically repeat-based. Recombination is largely suppressed at and in the proximity of centromeres in these species, a phenomenon known as the centromere effect\textsuperscript{23}. In plants with large chromosomes the centromere effect can extend several megabases (Mb) along pericentromeric regions, that can represent a large proportion of the chromosomes\textsuperscript{24–26}. The distal bias of CO distribution observed in many eukaryotes is typically assumed to be a result of such structural chromosome features (telomere and centromere effects) and association with (epi)genetic factors\textsuperscript{17,20,27–30}. However, how these factors influence meiotic recombination patterning at broad and local scales is still elusive.

Monocentricity is not the only centromeric organisation adopted by eukaryotes, however. For instance, holocentric species harbour multiple centromeric determinants over the entire length of their chromosomes\textsuperscript{31,32}. Holocentricity has evolved independently multiple times during the evolution of nematodes, insects and
plants\textsuperscript{33,34}. In the holocentric animal models \textit{Caenorhabditis elegans} and silk moth
\textit{(Bombyx mori)}, holocentromeres do not associate with a specific sequence and can
have variable dynamics\textsuperscript{35,36}. By contrast, holocentric plants of the \textit{Rhynchospora}
genus (beaksedges) display repeat-based holocentromeres in both mitosis and
meiosis\textsuperscript{37,38}. Recently, we sequenced the genomes of three beaksedges (\textit{R.
breviuscula}, \textit{R. pubera} and \textit{R. tenuis}) and determined that each chromosome
harbours multiple short arrays (~20 kb each) of the specific \textit{Tyba} tandem repeat, 
evenly spaced (every 400–500 kb) along the entire chromosomal length, and
specifically associated with centromeric histone H3 protein CENH3\textsuperscript{39}. This particular
chromosome organisation is associated with remarkably uniform distribution of genes,
repeats, and epigenetic features, in stark contrast to the compartmentalised
chromosome organisation of most eukaryotes\textsuperscript{39}. Remarkably, each individual
centromeric unit in \textit{R. pubera} showed very similar epigenetic regulation as found in
other plant monocentromeres\textsuperscript{30,39}. Thus, beaksedges offer an excellent model to study
the mechanisms of CO formation in the absence of the major effect of the
monocentromere, while having similar centromere chromatin (epi)genetic properties
at fine scale.

Studies of meiosis in holocentric plants have been mainly focused on the intriguing
phenomenon of “inverted meiosis”\textsuperscript{33,40–42}. Moreover, meiotic chromosomes in
\textit{Rhynchospora} maintain the repeat-based holocentromere organisation\textsuperscript{38}, suggesting
that COs can be formed in proximity to centromeric chromatin. However, no direct
evidence of meiotic recombination frequency and distribution has yet been reported
for any holocentric plant. It is still unknown whether and how plant holocentromeres
interact or interfere with meiotic recombination. Understanding whether the uniform
distribution of (epi)genetic features and hundreds of repeat-based centromeric units
distributed chromosome-wide in \textit{Rhynchospora} influence meiotic recombination
patterning will allow us to explore conserved and adapted mechanisms among
eukaryotes.

Here, we use \textit{R. breviuscula} as a model to study meiotic recombination dynamics in
the absence of both a localised centromere and a compartmentalised chromosome
organisation, features that potentially mask underlying factors affecting the CO
distribution in most genomes. Using a combination of immunocytochemistry,
chromatin and DNA analysis, and CO calling from high-throughput single-pollen
sequencing, we develop a comprehensive overview of meiotic recombination
dynamics and distribution for a species with repeat-based holocentromeres. We show
that despite this unique chromosome organisation, COs distribution is biased towards
the distal regions of chromosomes. Remarkably, the megabase-scale CO distribution
did not correlate with any (epi)genomic feature analysed. We found that COs are
suppressed inside repeat-based centromeric units but not in their proximity, indicating
the absence of a canonical centromere effect. We show that CO distal bias is achieved
even in the absence of both a monocentromere and correlation with (epi)genomic
features. In fact, our cytological data suggest that synapsis dynamics starting from
chromosomal ends exert a major influence in determining the broad-scale
recombination landscape in *R. breviuscula*, regardless of differential centromere
organization. We propose that centromere and (epi)genetic features play a role in CO
positioning only at fine scale.

**Results**

The molecular dynamics of early meiosis I is conserved in *R. breviuscula*

Chromosome spreads on male meiocytes of *R. breviuscula* allowed us to conclude
that prophase I progression is conserved in this species. We observed all the classical
prophase I stages, e.g., leptotene, zygotene, pachytene, diplotene, and diakinesis
(*Supplementary Fig. 1a*). It is important to note that at diakinesis we observe five
bivalents, hinting at a correct completion of recombination, with the formation of at
least one CO per homologs pair. Moreover, we confirmed the holocentric nature of *R.
breviuscula* chromosomes in mitosis and meiosis by showing the localisation of the
centromeric protein CENH3. Similar to what has been reported in *R. pubera*\(^ {38}\), CENH3
appears as lines during mitotic metaphase, but undergoes considerable centromere
restructuring into more irregular clusters during meiosis (*Supplementary Fig. 1b–c*).

We then investigated the immunolocalisation of ASY1\(^ {43,44}\) and ZYP1\(^ {6,7,11,12,45}\) as
indicators of a conserved and functional machinery for chromosome axis formation
and synapsis, respectively. The ASY1 signal was present along the entire length of
unsynapsed chromosomes in early prophase I, corresponding to leptotene (*Fig. 1a*).

During zygotene, the SC started to assemble and ZYP1 was gradually loaded onto
synapsed chromosomes. As ZYP1 was loaded, the two ASY1 linear signals could be
followed until they converged and lost intensity, after which the ZYP1 linear signal became clear and intense (Fig. 1b, insert). As meiosis progressed into pachytene, with complete synopsis and pairing, we detected the linear ZYP1 signal along the full length of chromosomes. The ZYP1 signal localised in the groove between the pairs of homologs. The combined behaviour of ASY1 and ZYP1 was consistent with the one observed in monocentric models. This hints at a conserved pairing and synopsis dynamics in *R. breviuscula*. We also tested whether the meiosis-specific alpha-kleisin REC8 is conserved in *R. breviuscula*. REC8 is responsible for sister chromatid cohesion and is important for chromosome segregation and recombination\(^46\). Indeed, we detected a conserved linear REC8 signal at pachytene, when REC8 co-localised with ZYP1 as a continuous linear signal along entirely synapsed chromosome (Fig. 1c). Thus, pairing and synapsis are conserved in the holocentric plant *R. breviuscula*, resembling those in monocentric models.

**Fig. 1:** Immunolocalisation of ASY1, REC8, and ZYP1 from leptotene to pachytene of meiotic prophase I. (a) ASY1 (green) appears as a linear signal on unpaired chromosomes. (b) Synapsis is visualised as the loading of ZYP1 (magenta) while the ASY1 (magenta) signal disappears. The insert shows a magnification of two unpaired chromosomes, represented by ASY1 (green), coming together to synapse, with the loss of the ASY1 signal and the loading of ZYP1 (magenta). (c) Full co-localisation of cohesin protein REC8 (green) and ZYP1 (magenta) at pachytene. A maximum projection is shown; chromosomes were counterstained with DAPI. Scale bars, 5 µm.
We further studied the behaviour of HEI10, a RING-family E3 ligase that has been characterised in plants, fungi and animals. HEI10 has a conserved pattern, acting after synapsis has occurred in the ZMM pathway, but before the resolution of COs. HEI10 has been proposed to interact with both early and late recombination proteins, acting by stabilising recombination sites and promoting their maturation into class I COs. In *R. breviuscula*, when synapsis started in early zygotene, HEI10 was immediately loaded as patchy dispersed signals co-localising with the first ZYP1 signals (Fig. 2a). At pachytene, HEI10 progressed to form a linear signal co-localising with ZYP1 along the entire synapsed chromosomes (Fig. 2b). During pachytene, when synapsis is complete, the HEI10 linear signal started to become non-homogeneous along chromosomes, while a few foci increased in intensity (Fig. 2c). We think that these are putative class I CO sites. In diplotene and diakinesis, only high-intensity foci remained (Fig. 2d). Thus, the dynamic behaviour of HEI10 is conserved and most likely the recently proposed HEI10 “coarsening” model is acting similarly in *R. breviuscula*.

Another established marker for meiotic recombination is the mismatch repair protein MLH1 (MUTL-HOMOLOG 1), which is essential for meiosis and is believed to have a meiosis-specific resolvase activity to process double Holliday junctions (dHJs) into final class I COs. MLH1 interacts with MSH4 (MUTS HOMOLOG 4) and MSH5 in the dHJ resolution pathway, thus specifically marking class I COs in distantly related species. In *R. breviuscula*, MLH1 signal appears exactly at diplotene, when the SC is being disassembled, and persists as bright foci until the end of diakinesis (Fig. 2e and Supplementary Fig. 2). We observed a mean number of approximately six for both HEI10 (N=63) and MHL1 (N=83) foci (Fig. 2f), which is consistent with the formation of at least one CO per homolog pair. We also observed at least one homolog pair, i.e., ring bivalent, with more than one CO (Supplementary Fig. 3). Thus, in contrast to the holocentric *C. elegans*, where strictly only one CO is allowed per bivalent, *R. breviuscula* can properly form more than one CO per bivalent. Furthermore, our results confirm a canonical and conserved early meiosis program in *R. breviuscula*. However, we still do not know whether its repeat-based holocentromeres influence CO distribution.
Fig. 2: Immunolocalisation of HEI10, ZYP1 and MLH1 in late prophase I. (a) In early zygotene when synapsis starts, HEI10 (green) is immediately loaded as many closely spaced foci, forming an irregular and patchy signal and co-localising with ZYP1 (magenta). (b) In pachytene, HEI10 (green) is visible appearing as lines, which co-localise with ZYP1 (magenta). (c) In late pachytene, the linear signal of HEI10 still co-localises with ZYP1, but becomes weaker, except for a few highly intense foci. (d) During the diplotene and diakinesis stages, HEI10 appears only as foci on bivalents, with no linear signal. (e) MLH1 (green) appears in diplotene and diakinesis as foci on bivalents, representing chiasmata. (f) Count of HEI10 and MLH1 foci at late prophase I. Maximum projections are shown, with chromosomes counterstained with DAPI. Scale bars, 5 µm.
Genome phasing of *R. breviuscula* as a prerequisite for CO identification

We set out to determine whether recombination in *R. breviuscula* is affected by the genome-wide distribution of holocentromeres. To identify COs from a single *R. breviuscula* individual, the genome of the given organism must be heterozygous and a phased chromosome-level reference genome must be available. The recently reported unphased genome of *R. breviuscula* was reported to be 1% heterozygous\(^9\), suggesting the feasibility of phasing the genome. We took advantage of the recent development of the assembler software Hifiasm\(^5^9\), which enables the accurate phasing of assembled contigs from both haplotypes using a combination of HiFi reads and Hi-C (Fig. 3a–b; Methods). Further Hi-C scaffolding of each set of haplotype-phased contigs led to high-quality haplotype-phased chromosome-level genome assemblies (Fig. 3c and Supplementary Table 1). We performed a synteny analysis and detected the structural variants between the two haplotypes, revealing a high degree of synteny between the haplotypes with only a few inversions, translocations and duplications (Fig. 3d and Supplementary Table 2).
Fig. 3: Phasing and structural variation of the *R. breviuscula* heterozygous genome. (a–b) Assembly statistics of the phased contigs (a) and scaffolds (b) for haplotype 1 and haplotype 2. (c) Hi-C scaffolding of the five haplotype-phased pseudochromosomes. Homozygous regions between the haplotypes are seen as clear regions depleted of signals on the Hi-C map. (d) Synteny blocks and structural variants (>10 kb) identified between the two haploid assemblies. Note the overall high synteny between the two haplotypes. Synteny blocks were computed with SyR\textsuperscript{60} and plotted with plotsr\textsuperscript{61}.

**Single-cell RNA sequencing of pollen allows the high-throughput identification of genome-wide COs**

*R. breviuscula* is an outbred wild species with high levels of self-incompatibility, which hampers the standard detection of COs, typically involving the time-consuming generation of segregating offspring. We obtained only 63 F\textsubscript{1} plants by manual selfing-pollination that were sequenced at 3-fold genome coverage. Although, we detected 378 CO events at a very high resolution (CO located interval size: median 334 bp, mean ~ 2 kb), this low number of COs is not enough to draw its recombination landscape.

We set out to provide a shortcut for constructing broad-scale CO landscape for poorly studied non-model species. Gametes carry the outcome of meiotic recombination and can be obtained in large numbers in a relatively inexpensive manner from pollen grains. Thus, we adapted a strategy based on the gamete-binning method described by Campoy et al.\textsuperscript{62}, which was based on the discontinued 10X Genomics Chromium Single Cell CNV product. Here, instead, we identified genome-wide CO events by conducting 10X Genomics single-cell RNA sequencing (scRNA-seq) on extracted nuclei from pollen grains (male gametes) of *R. breviuscula* (**Fig. 4a; Methods**), with the caveat that pollen nuclei show relatively low abundance of transcripts (**Supplementary Fig. 4**), limiting the resolution of CO detection.

To genotype the haploid gamete genomes and determine from which haplotype a genomic segment is derived, genome-wide markers are needed to distinguish the two haplotypes. By aligning the ~26-Gb Illumina whole-genome short reads of *R. breviuscula* to the haplotype 1 phased genome, we detected 820,601 haplotype-specific single nucleotide polymorphisms (SNPs, ~1 SNP/449 bp) and used them as markers for genotyping (**Fig. 4b and Supplementary Figs. 5 and 7a**).
We pre-processed 10X scRNA sequences by correcting barcodes, demultiplexing, removing doublets and cells with a low number of reads (Supplementary Fig. 6; Methods), resulting in a final set of 1,641 pollen nuclei with at least 400 markers (~1 marker/Mb). These markers (median resolution ~1 marker/542 kb) covered almost the entire length of all five chromosomes (Supplementary Fig. 7b), ensuring genome-wide CO detection. We detected 4,047 COs in the 1,641 pollen nuclei by inspecting genotype conversions, as indicated in Fig. 4c–d. Overall, we delineated a complete and detailed pipeline to detect COs in an economical way by high-throughput scRNA-seq of gametes from a single heterozygous individual (Fig. 4).

**Fig. 4:** Overview of CO calling by adapting scRNA sequencing to *R. breviuscula* gametes. (a) Pollen sampling, library preparation and scRNA sequencing pipeline. FACS, fluorescence activated cell sorting. (b) Diagram of the strategy for identifying genotyping markers on the reference genome by mapping short reads and markers in gametes by mapping scRNA-seq reads across a large number of gametes to the reference genome. GMRs, genotyping markers on reference genome. (c) Diagram of the identification of potential CO events after the alignment of the scRNA reads from each gamete to
Chromosomes with evenly distributed hundreds of centromeres show uneven recombination landscape

Counting the occurrence of COs in chromosome-wide genomic intervals across 1641 pollen nuclei, we computed the CO rates along chromosomes and established the first recombination map for *R. breviuscula*, the first in any species with known repeat-based holocentromeres (Fig. 5a and Supplementary Fig. 8). Unexpectedly, the megabase-scale recombination landscape revealed an apparent trend of increasing CO rates towards most distal chromosomal ends, while the centre regions maintain rather low CO rates, mostly lower than the mean CO rate genome-wide. Remarkably, chromosomes 3, 4 and 5 showed distal bias of CO distribution at both ends (Fig. 5a). The distal bias was more conspicuous in both chromosomes 3 and 4, while less conspicuous in chromosome 5 due to a homozygous region with low marker density close to the right end, hampering the detection of COs in this particular region (Fig. 5a, orange bars). In contrast, chromosomes 1 and 2 showed a prominent increase of CO rate at only one chromosomal end (CO rate above the genome-wide mean), while the other end showed comparatively low CO rates similar to centred regions. Interestingly, these low recombination ends were correlated with the localization of 35S rDNA foci (nucleolar organizing regions) (Figure 5a–b). We thus revealed an uneven distribution of CO rates at megabase-scale, despite the decompartmentalized chromosome configuration reported in *R. breviuscula*.

Overall, we obtained similar results of CO number and distribution between our single-pollen sequencing strategy and our F₁ selfed offspring (Supplementary Fig. 9). However, we observed a shorter genetic map length in the pollen data compared to our F₁ offspring, likely resulted from missing some COs in our single-pollen sequencing (Supplementary Fig. 9). Using the set of 378 COs in our F₁ offspring the computed total linkage map length was 300 cM (Fig. 5b–c) compared to 250 cM computed from the pollen data (Supplementary Fig. 9a).

We further compared the number of genetically identified CO events in our pollen and F₁ offspring data. To have a precise estimation of CO number in our pollen data, we
counted only those COs from pollen nuclei with more than 2,000 markers (N=81). On average, we detected around three COs per haploid gamete, or 0.6 CO per chromatid (Fig. 5d–e). As gametes only have one chromatid from each recombined chromosome, the number of pollen-detected COs should be approximately half of COs detected in our segregating offspring and HEI10/MLH1 foci. Indeed, we found similar number of MLH1 foci and COs detected in our genetic analysis (Fig. 2f and 5d), suggesting that most COs formed in R. breviuscula are of class I. Furthermore, all single chromatids had in average one CO in half of these gametes (N=81), while double COs appeared in approximately 5% of the 81 gametes considered (Fig. 5e).

Chromosome 3 showed the highest frequency of double COs (9%; Fig. 5e and Supplementary Fig. 8c). Remarkably, chromosomes 3, 4 and 5 had longer genetic length among all R. breviuscula chromosomes (Fig. 5b–c). This is evident considering these three chromosomes have distal CO bias at both ends compared to only one in chr1 and chr2.

We also tested whether CO interference occurred in R. breviuscula. We used a Chi-square goodness-of-fit test to investigate whether the CO number on each chromosome follows a Poisson distribution, which revealed a significant discrepancy between observed and expected CO numbers (Supplementary Fig. 10a). This result shows that COs are not randomly distributed but under-dispersed, based on the negative alpha values from dispersion tests, that could be the effect of CO interference. We also computed the coefficient of coincidence (CoC) of COs across the genome, which measures the observed frequency of double COs over their expected frequency (see Methods). The CoC curve of all chromosomes showed that the coefficients are below 1 for genomic intervals with distances less than around 60 Mb (Fig. 5f and Supplementary Fig. 10b), showing that the frequency of double COs is lower than expected. This result indicates the presence of strong CO interference in R. breviuscula.
Fig. 5: Meiotic recombination dynamics in *R. breviuscula*. (a) The first recombination landscape of the five chromosomes in *R. breviuscula* achieved by computing COs from 1,641 pollen nuclei. Black line displays the CO rate. Median size of intervals within which a CO localised is \( \sim 1.5 \) Mb, mean \( \sim 2.24 \) Mb. Shadow ribbons indicate one standard deviation from mean CO rates. Blue dashed vertical lines: start and end of confident CO rate computation (Supplementary Fig. 8). Blue solid vertical lines: chromosomal end. Magenta horizontal lines: genome-wide average CO rate. Green horizontal lines: chromosome-wide average CO rate. Orange bars: large (>2 Mb) homozygous regions with a limited number of markers. (b) Genetic linkage map computed from 378 COs from 63 F\textsubscript{1} selfed offspring. Genetic length density is indicated by the colouring scale. A set of 705 markers was selected using a 500-kb sliding window through all markers defined against the reference (see Methods). The terminal locations of the 35S rDNA locus on chromosomes 1 and 2 are indicated by asterisks in a and b. (c) Marey map calculated from the linkage map in b. Marey maps for each chromosome (colour lines) show genetic position as a function of physical position. (d) CO number derived by counting CO events from the genetic analysis in pollen compared to the one extrapolated from the F\textsubscript{1} offspring. (e) Distribution of CO number for each single chromatid in gametes. Note the higher incidence of double COs on chromosome 3. (f) CoC curve in pollen nuclei (N=1,641). Chromosomes were divided into 15 intervals, with random sampling at CO intervals, to calculate the mean coefficient of coincidence of each pair of intervals.
Megabase-scale recombination landscape is independent of holocentromere
distribution and (epi)genomic features

The distally-biased recombination landscape is ubiquitously seen across many
eukaryotic species, which is typically explained by suppression of CO formation at
pericentromeric regions and association with large-scale epigenetic regulation\textsuperscript{27,28}.
Hence, we set out to correlate the recombination landscape of \textit{R. breviuscula} with
several (epi)genetic features. Surprisingly, a chromosome-wide comparison revealed
no apparent correlation of CO rates with the uniform holocentromere distribution and
other genomic (genes, TEs, SNP densities, or with GC content) and epigenomic (such
as H3K4me3, H3K27me3, H3K9me2 or DNA methylation) features. To estimate
whether fast-evolving genes correlated with the regions showing higher recombination
frequencies, we also compared the Ka/Ks ratio (measurement of the relative rates of
synonymous and nonsynonymous substitutions at each gene). Notably, the Ka/Ks
ratio across the chromosomes was rather uniform, and we did not find any bias
towards the chromosome ends (\textbf{Supplementary Fig. 11}). Quantification at
megabase-scale of (epi)genomic features revealed no strong correlation with CO
distribution as well (\textbf{Fig. 6b}). These results indicate that, at broad scale, meiotic
recombination occurs independently of any (epi)genomic feature and chromosome-
wide distribution of repeat-based holocentromeres.
Fig. 6: Broad- and fine-scale correlations between CO patterning and (epi)genomic features in *R. breviuscula*. (a–b) Absence of correlation at megabase scale: (a) Chromosome distribution of the CO rate coupled with different (epi)genetic features. Top: recombination landscape (black line) created with COs detected in all single-pollen nuclei (N=1,641), coupled with Omni-C chromosome conformation capture contacts. For the x-axes, the coordinates were based on the haplotype 1 assembly *R. breviuscula*. For the y-axes, all features were scaled [0,1], with 1 indicating a maximum of 2.34 for recombination frequency (cM/Mb), 5 for Tyba density, 6 for CENH3 density, 7205 for SNP density, 88 for gene density, and 227 for TE density. GC [33.3, 46.6], H3K4me3 [−1.494, 0.231], H3K9me2 [−1.20, 1.84], and H3K27me3 [−0.671, 0.491] are scaled to [0,1] by their minima and maxima. mCG, mCHG and CHH are original values (0 to 100%). All features except for Hi-C contact map were smoothed by 1Mbp sliding window and 250kbp step size. COs are almost completely absent in a large inversion in chr2:30–35 Mb, while in homozygous regions we could not confidently call COs, for example in chr4:25–35 Mb. The large variants were confirmed within Hi-C contact maps (Fig. 3c). (b) Quantification at megabase scale confirms the absence of correlation between CO rate and (epi)genomic features. Correlation matrix illustrates the correlation coefficients of 378 high-resolution COs detected in 63 selfed F1 offspring with all available (epi)genomic features. Positive correlations are displayed in blue and negative correlations in red. Colour intensity and the size of the circle are proportional to the correlation coefficients. Pearson correlation coefficients for each pair of all features
under 1-Mb smoothing window and 250-kb step size: specifically, mean CO rates, mean GC contents, CENH3 peak density, Tyba array density, SNP density, TE density, H3K4me3 RPKM, H3K9me2 RPKM, H3K27me3 RPKM, mean CpG, mean CHG, and mean CHH. (c–g) Identification of correlation at fine scale: (c) Z-score of the overlapped CO numbers with different (epi)genetic features to the 5,000 simulations of randomly distributed COs. Positive z-score indicates that COs overlap with H3K4me3 and genes more frequently than expected under the hypothesis of random distributed COs along chromosomes. Negative z-score implies the contrary. The higher the absolute z-score, the more deviation is observed. (d) Within genic regions, CO frequency (blue line) is higher in promoter regions or after the transcription termination site (TTS), but lower in gene bodies, independent of marker density (grey line). TSS, transcription start site. (e) Within CENH3 domains CO frequency is remarkably suppressed, despite relative high marker density. (f) Random distribution of the relative distance of CO positions to the end of the left and to the start of the right CENH3 domain. Correlation analysis performed using data from 63 F1 recombinant offspring and a total of 378 COs. Pink-bordered triangles schematically represent CENH3 domains. (g) Magnified view of one of the five COs placed within a region containing CENH3-positive chromatin and Tyba repeats. CO resolution in this case 200 bp. CO is indicated by the grey dashed line showing the haplotype switch (blue to orange) in the marker density track.

Absence of centromere effect sheds light on the fine-scale epigenetic CO regulation

Next, we tested whether COs are epigenetically affected at a fine scale and whether individual centromeric units have an effect on CO designation in *R. breviuscula*. For that we only used the set of 378 COs resolved at high resolution (median 334 bp, mean ~ 2 kb) from our selfing offspring experiment.

The holocentromeres in *R. breviuscula* are repeat-based, i.e., each centromeric unit is based on a specific array of the holocentromeric repeat Tyba associated with CENH3, with average sizes of ~20 kb and average spacings of ~400 kb, where each chromosome harbours hundreds of individual centromeric units (*Supplementary Fig. 12a–b*). Remarkably, we found the same epigenetic centromere identity in *R. breviuscula* (*Supplementary Fig. 12c*) as reported for *R. pubera*39. This organisation makes it possible to identify centromeric units at the DNA level by annotating Tyba repeat arrays (*Supplementary Fig. 12b*). We then computed the observed versus expected by random distribution fine-scale CO position across all available chromatin marks and genetic features. We found that COs are more frequently formed at H3K4me3 peaks and genes than what expected by random distribution (*Fig. 6c and
Supplementary Fig. 13). Within genic regions COs were preferentially formed at the promoter regions (Fig. 6d). Remarkably, COs were mostly suppressed inside the core of centromeric units and heterochromatic regions, where high levels of DNA methylation are found (Fig. 6c–e and Supplementary Fig. 12c–d and 13). However, after computing the distances between the CO break intervals and the corresponding nearest CENH3 domains/Tyba arrays, the COs did not show a tendency to be positioned away from or close to centromeric units (Fig. 6e–f and Supplementary Fig. 12d–e), indicating that the proximity to a centromeric unit does not affect CO formation. Moreover, we found five cases of a CO being placed to a region containing reduced CENH3-positive chromatin and Tyba repeats (Fig. 6g), confirming the absence of a centromere effect in its proximity. Our results point to the exciting finding that local CO formation in *R. breviuscula* is abolished at repeat-based centromeric units but enriched at genic promoter regions, supporting the role of chromatin features at fine scale in contrast to the absence of correlation at broad scale.

**Telomere-led spatiotemporal dynamics of chromosome synapsis and HEI10 loading**

We hypothesised that pairing and synapsis progression might contribute to the distally-biased recombination landscape observed in *R. breviuscula*. To investigate this, we performed immunolocalisation with antibodies against ZYP1, ASY1, and HEI10 and fluorescence *in situ* hybridisation (FISH) for telomeres on meiocytes. Signals detected for ZYP1, ASY1 and telomere probes indicated a tendency for telomeric signals to cluster together in one location, forming a typical *bouquet* configuration. Near this structure, we observed ZYP1 signal, representing synapsed chromosomes elongating from telomeres. The linear signal of ASY1 was still present and representing unpaired chromosomes on the opposite cell side to the *bouquet* (Fig. 7a and Supplementary Fig. 14). Next, we asked whether HEI10 loading also shows telomere-led dynamics. Indeed, we could determine that the first synapsed regions (ZYP1-stained) were also first loaded with HEI10 in the proximity of chromosome ends, exhibiting a discrete segmented linear signal (Fig. 7b–c). These telomeric early synapsed regions showed a high degree of colocalization with early loaded linear HEI10 signals, but decreased in colocalization intensity as HE10 signals became less linear and more dot-like in pachytene (Fig. 7c-d and Supplementary Fig. 15) We consistently observed a few
telomeres that did not participate in the bouquet in all analysed meiocytes (N=44). These signals represent the terminal ends of chromosomes 1 and 2 that harbour the 35S rDNA loci and show low recombination rates; instead, these chromosome ends localised in the nucleolus (Fig. 5a–b and Supplementary Fig. 16). Remarkably, the nucleolus-positioned telomeres showed delayed or no ZYP1 loading compared to the telomeres involved in the bouquet (Fig. 7b and d). Indeed, these telomeric signals had an average of four, consistent with two unsynapped chromosome ends, while telomeric signals at the bouquet had an average of eight, consistent with eight synapsed chromosome ends (N=44). Thus, the broad-scale recombination landscape in R. breviuscula is better explained by the telomere-led HEI10 loading on early synapsed chromosome ends rather than by any association with a centromere effect or (epi)genetic features.
Fig. 7: Telomere-led dynamics of synopsis formation and HEI10-loading. (a) Telomeres cluster in a bouquet (white arrow) on one side of the cell, where ZYP1 elongating as the SC is being assembled. ASY1 represents unpaired chromosomes not yet reached by ZYP1. (b) As ZYP1 lines elongate from the bouquet (white arrow), HEI10 is quickly loaded onto synapsed chromosomes, while some telomeres localise to the nucleolus (white arrowhead) and lack the ZYP1 and HEI10 signals. (c) Detail of early synopsis initiation and intensity profile of HEI10, ZYP1 and telomere: as soon as the SC (ZYP1) is assembled from telomeres, HEI10 is loaded and shows high-colocalization profile with ZYP1 (N=44). (d) In late pachytene, ZYP1 occupies the whole chromosomal length, HEI10 signals become more dot-like and telomeres are still clustered in the bouquet (white arrow), while few telomeres remain at the nucleolus (white arrowhead). Telomeres (orange), ZYP1 (magenta), ASY1 (green) and HEI10 (green). Scale bars, 5 µm.
**Discussion**

Deciphering the mechanisms controlling CO formation and distribution is key to understanding one of the main driving forces for genetic diversity in eukaryotes: meiotic recombination. We show that by using *R. breviuscula*, a holocentric organism lacking both localized centromeres and compartmentalised chromosome organisation, features that can potentially mask the factors underlying CO patterning, we can reveal important insights into CO control mechanisms. By combining comprehensive immunocytochemistry, chromatin, and genetic analyses, we uncovered canonical recombination dynamics and distally-biased CO distribution, which is best explained by the telomere-led pairing and synapsis independent of (epi)genomic features. This result is consistent with the *bouquet* configuration reported in many organisms, where synapsis and DNA double-strand breaks (DSBs) required for COs are mostly initiated from the telomeres\textsuperscript{63,64}. Such telomere-led mechanisms have already been proposed to influence the location of COs to be more likely at the chromosome ends than the centres (see Haenel et al.\textsuperscript{29} and references therein). Considering the marked conservation for *bouquet* formation and synapsis progression in *R. breviuscula*, and the distal bias of COs observed, we propose that synapsis itself, and possibly the observed telomere-led HEI10 loading dynamics, are the driving force that shapes the recombination landscape in this species. This early loading at ends might create a bias that increases CO rates at the distal regions of the chromosomes, whether a centromere is present or not. Recently, a “coarsening” model for the behaviour of HEI10 has been proposed. In this model, enhanced loading of HEI10 at the chromosome ends leads to increased COs. As the amount of loaded HEI10 accounts for the increased coarsening over time, early loading at the chromosome ends would accelerate the maturation of recombination intermediates compared to the interstitial regions of the chromosomes\textsuperscript{53,54}.

Moreover, we observed a gradual reduction in CO rates from the regions directly adjacent to telomeres in *R. breviuscula*. Similar to the centromere effect, a telomere effect is proposed to commonly occur across eukaryotes\textsuperscript{20,29}. Interestingly, this phenomenon can also be explained by HEI10 coarsening dynamics. Telomeric regions would be subject to reduced unidirectional coarsening contribution, compared to adjacent regions. Additionally, the phenomenon of CO interference lowers the recombination frequencies at the centre of chromosomes, with the distal regions...
having already been designated for COs. The model described here would explain the
behaviour of chromosomes 3, 4 and 5; however, the 35S rDNA-harbouring distal
regions of chromosomes 1 and 2 do not participate in bouquet formation, as they stay
at the nucleolus. Remarkably, these two chromosomal ends are also characterised by
low recombination frequencies. For the model plant *A. thaliana*, it has been proposed
that ribosomal DNA localises in the nucleolus, where it is shielded from synapsis and
meiotic recombination. Indeed, we observed that the telomeres of *R. breviuscula*
situated in the nucleolus were involved later in synapsis than those that cluster in the
*bouquet*. This late involvement in synapsis means a delay in HEI10 loading, which is
consistent with the lower recombination frequency observed at the 35S rDNA-
harbouring ends of chromosomes 1 and 2 (Figure 8). Here, we show that the CO distal
bias is present even in the absence of compartmentalised chromosomal features,
suggesting that telomere-led pairing and synapsis initiation alone can impose CO bias
in *R. breviuscula*. The proximity of telomeres during early prophase has been
proposed to be responsible for telomere-led recombination in *A. thaliana asy1* mutants
and wheat. However, we cannot exclude that other factors, like a different density
of DSBs along chromosomes might contribute to the observed distribution of COs.
Moreover, a recent study has shown that the bouquet configuration is required for the
distal bias of DSB formation and repair, possibly playing an important role in the
observed reduced recombination rates at interstitial positions. Future experiments in
*Rhynchospora* will be important to identify conserved and adapted mechanisms about
the role of spatiotemporal dynamics of meiotic DSB formation and HEI10 loading in
shaping the recombination landscape.

---

**Fig. 8:** Model for CO regulation at broad and fine-scale in the holocentric plant *R. breviuscula*.

(a) Model for the role of telomere-led synapsis and HEI10 loading in shaping the broad-scale CO
landscape in *R. breviuscula*. While chromosomes 3, 4 and 5 have CO distal bias in both ends, chromosomes 1 and 2 show only bias at one end, which is opposite to the localization of 35S rDNA loci. (b) Model for CO formation at (left) broad and (right) fine scale. Telomere-led synapsis leads to an early loading of HEI10 at chromosomal ends that can potentially favour COs at distal regions, while 35S-rDNA harbouring chromosome ends do not show early synapsis and thus have less probability of making COs. At local scale, COs are suppressed at core centromeric units, but not at their vicinity, where COs can be placed anywhere between two centromeric units. Remarkably, COs were preferentially placed at gene promoter regions.

In the new era of highly-accurate long read genomics, haplotype-phased genomes are routinely available. By applying high-throughput single-cell RNA sequencing to individual pollen nuclei, we provide a powerful pipeline that can be used to investigate CO frequencies in any available gamete of any heterozygous individual with an available phased genome. Using haplotype-specific markers, we detected and mapped CO events from thousands of gametes for the first time in a species with repeat-based holocentromeres. Although the distal bias of CO found is similar to what is observed in numerous eukaryotes, including the holocentric *C. elegans*\(^{28,29,74}\), the lack of correlation with the even distribution of (epi)genetic features (this study; \(^{39}\)) in *R. breviuscula* is remarkable. In *C. elegans*, for instance, distinct chromosome domains (“centre” and “arms”) are characterized by differential gene density, repeat contents, and histone modifications\(^{28}\) and the unequal CO distribution corresponds to its chromosome domains – recombination rates are lower at centres than arms\(^{75}\). Furthermore, a recent study showed that the megabase-scale CO landscape in *A. thaliana* is mostly explained by association with (epi)genetic marks beyond a centromere effect, with open chromatin states showing the highest positive correlation with CO formation\(^{18}\), while nucleotide polymorphisms only played a rather local effect on CO positioning\(^ {18,76}\). In contrast to these organisms, we could only find correlation of CO positioning with centromere and (epi)genetic features at a very fine scale, where COs preferentially formed within gene promoter regions rather than in neighbouring transcribed gene bodies, TEs and centromeres (Figure 8b). This result appears to hold true for several eukaryotes and may be related to open chromatin states\(^ {17,18,77}\), suggesting that fine-scale CO regulation is associated with similar (epi)genetic factors independent of the chromosome organisation. In contrast, the absence centromere effect found in *R. breviuscula*, which seems to suppress CO formation only at the core
centromeric units but not at their vicinity (Figure 8b), is likely due to the closed chromatin state of centromeric chromatin in *R. breviuscula*, as marked by high DNA methylation levels, as also found within TEs. Our findings suggest that the pericentromeric inhibition of COs observed in many monocentric eukaryotes is likely a secondary effect of heterochromatin accumulation along large monocentric pericentromeres and not a direct effect of centromeric chromatin. Understanding the molecular mechanisms of CO control in holocentric organisms will potentially unveil new strategies to address meiotic recombination within centromere proximal regions of monocentric chromosomes that rarely recombine.

**Data and code availability**

All sequencing data used in this study have been deposited at NCBI under the Bioproject no. XXXXXXX and are publicly available as of the date of publication. The reference genomes, sequencing data, annotations and all tracks presented in this work are made available for download at DRYAD: 
https://datadryad.org/stash/share/EvB3PRNVph5iTkOM3jTZddgmS45cJhQYq2v3LI5InE.

All other data needed to evaluate the conclusions in the paper are provided in the paper and/or the supplemental information. The original code for the construction of recombination maps from single-cell RNA sequencing is available at https://github.com/Raina-M/detectCO_by_scRNAseq. Any additional information required to reanalyse the data reported in this paper is available from the corresponding author upon request.

**Acknowledgements**

We thank Raphaël Mercier for the insightful comments and critical reviewing on the manuscript. We thank Neysan Donnelly for reviewing the manuscript. This study was financially supported by the Max Planck Society (core funding to AM) and the Deutsche Forschungsgemeinschaft (DFG, grant number MA 9363/2-1 to AM). MZ is financially supported by the DFG (grant number MA 9363/2-1). We thank the PhD fellowship awarded to G.T. from the DAAD/India. YMS received financial support from PROBRAL (CAPES/DAAD) program (grant number 88881.144086/2017-01). JAC is financially supported by the Marie Skłodowska-Curie Individual Fellowship PrunMut.
We thank Ursula Pfordt and Christina Philipp for the excellent technical assistance.

**Author contributions**

AM conceived the research program and coordinated the analyses. MC performed all cytogenetic analyses and microscopy. MC isolated the pollen nuclei and generated sequencing libraries with assistance from JAC. MZ performed all single-cell RNA sequencing and recombination-related analyses with assistance from HS. GT performed the ChIP-seq analysis. YMS performed the immuno-FISH analysis. TL and KFXM performed the gene annotation and Ka/Ks ratio analysis. MM operated the FACS machine. BH performed all sequencing. KS supervised the single-cell analysis. MC, MZ and AM wrote the first manuscript draft with input from all authors. All authors approved the final version of the manuscript.

**Competing interests**

The authors declare no competing interests.
Methods

Plant Material and Specimens

A single specimen of *Rhynchospora breviuscula* was collected from a natural population (Jaguariaiva, State of Paraná, Brazil) and vegetatively propagated by stolon elongation and kept under cultivation in a greenhouse at the MPI for Plant Breeding Research in Cologne, Germany. Flower buds were used for screeing meiotic cells and pollen was collected for single-gamete sequencing (see below).

Despite the low selfing rate of *R. breviuscula* due to high self-incompatibility, a segregation offspring (n=63 plants) was generated by manual self-pollination and germination of seeds in soil.

Isolation of pollen nuclei, 10X Genomics scRNA-seq library preparation and sequencing

Protocols were adapted from Campoy et al.\textsuperscript{62}. Briefly, to release pollen grains, anthers from fully developed flowers of *R. breviuscula* and *R. tenuis* (for multiplex purposes) were harvested and submerged in woody pollen buffer (WPB)\textsuperscript{79}. The nuclei were extracted using a modified bursting method. The solution containing the pollen grains was pre-filtered with a 100-µm strainer, and the pollen was crushed on a 30-µm strainer (Celltrics). The isolated nuclei were gathered in WPB and stained with DAPI (1 µg/ml) before being sorted using a BD FACSARia Fusion sorter with a 70-µm nozzle and 483-kPa sheath pressure. A total of 10,000 nuclei were sorted into 23 µl of sheath fluid solution and loaded into a 10X Chromium controller, according to the manufacturer’s instructions. A library was created according to the chromium single-cell 3’ protocol. A CG000183 Rev A kit from 10X Genomics was used for library preparation. The library was sequenced (100 Gb) on an Illumina NOVAseq instrument in 150-bp paired-end mode.

Whole-genome sequencing (WGS) of F\textsubscript{1} recombinant offspring

To obtain a recombinant population of *R. breviuscula* plants, young inflorescences of the heterozygous reference *R. breviuscula* were bagged to force self-pollination. Due to its high self-incompatibility, only 63 F\textsubscript{1} plants were obtained, and they were
sequenced to 3X coverage (~2 Gb) using an Illumina NextSeq2000 instrument in 150-bp paired-end mode.

**Anther fixation and immunocytochemistry**

Immunostaining was performed as described by Marques et al. with some modifications. Anthers of *R. breviuscula* were harvested and fixed in ice-cold 4% (w/v) paraformaldehyde in phosphate buffered saline (pH 7.5; 1.3 M NaCl, 70 mM Na₂HPO₄, 30 mM NaH₂PO₄) for 90 min. The anthers were separated according to their size and were dissected to release the meiocytes onto glass slides. The meiocytes were squashed with a coverslip that was later removed using liquid nitrogen. The slides were stained with mounting solution (Vectashield + 0.2 µg DAPI) to select the meiotic stages of interest, after which they were blocked with a 1 h incubation in 3% (w/v) bovine serum albumin in PBS + 0.1% (v/v) Triton X-100 at 37°C. The antibodies used were anti-AtASY1 raised in rabbits (inventory code PAK006), anti-AtMLH1 raised in rabbits (PAK017) and anti-Verseck CENH3 raised in rabbits. The anti-ZYP1 was raised in chickens against the peptide EGSLNPYADDPYAFD of the C-terminal end of AtZYP1a/b (gene ID: At1g22260/At1g22275) and affinity-purified (Eurogentec) (PAK048). Remarkably, this peptide region showed 100% similarity with the *Rhynchospora* ZYP1 C-terminal (gene ID: RBREV_HAP1.r01.Chr2_h1G00222020.1). The anti-REC8 was a combination of two antibodies raised in rabbits against the *Rhynchospora*-specific REC8-peptides C-EYPYEIQISKGMNM and C-YNPDDSVVERMRDDPG (gene ID: RBREV_HAP1.r01.Chr4_h1G00395720.1) and affinity-purified (Eurogentec). The anti-HEI10 was a combination of two antibodies raised in rabbits against the *Rhynchospora*-specific HEI10-peptides C-NRPNQSARTMNFMQL and C-PVRQRNNKSMVSGGP (gene ID: RBREV_HAP1.r01.Chr4_h1G00387160.1) and affinity-purified (Eurogentec). Each primary antibody was diluted 1:200 in blocking solution. The slide-mounted samples were incubated with the primary antibodies overnight at 4°C, after which they were washed three times for 10 min with PBS + 0.1% (v/v) Triton X-100. The slides were incubated with the secondary antibodies for 2 h at room temperature. The secondary antibodies were conjugated with Abberior STAR ORANGE or Abberior STAR RED (1:250; Abberior) before being washed again three times for 10 min with PBS + 0.1% (v/v) Triton X-100 and allowed to dry. The samples were prepared with 10 µl of
mounting solution (Vectashield + 0.2 µg DAPI), covered with a coverslip, and sealed with nail polish for storage. Images were taken with a Zeiss Axio Imager Z2 with Apotome system for optical sectioning or with a Leica Microsystems Thunder Imager dMi8 with Computational Clearing. The images were deconvolved and processed with Zen 3.2 or LAS X software. Colocalization analysis of ZYP1 and HEI10 signals was performed using the colocalization function of ZEN v3.2 software (Zeiss) and auto-thresholding was done using the Costes function\textsuperscript{81,82}.

**Sequential immunostaining and fluorescence in situ hybridisation**

Immuno-FISH was performed following Baez et al.\textsuperscript{81}. The best slides obtained from immunostaining, as described above, were selected for FISH using a telomeric probe. The slides were washed with 1× PBS for 15 min, postfixed in 4% (w/v) paraformaldehyde in PBS for 10 min, dried with 70% (v/v) and 100% ethanol for 5 min each and probed with direct-labelled telomeric sequence (Cy3-[TTTAGGG]$_5$; MilliporeSigma). The hybridisation mixture contained formamide (50% w/v), dextran sulphate (10%, w/v), 2× SSC and 50 ng/µl of telomeric probe. The slides were denatured at 75°C for 5 min. Stringency washes were performed following\textsuperscript{82} to give a final stringency of approximately 72%. The slides were counterstained with 10 µl of mounting solution (Vectashield + 0.2 µg DAPI), and images were captured as described above.

Mitotic and meiotic chromosome spreads were performed as described by Ruban et al.\textsuperscript{83}, with some modifications. Briefly, tissue samples were fixed in 3:1 (ethanol:acetic acid, v/v) solution for 2 h with gentle shaking. The samples were washed with water twice for 5 min and treated with an enzyme mixture (0.7% [w/v] cellulase R10, 0.7% [w/v] cellulase R10, 1.0% [w/v] pectolyase, and 1.0% [w/v] cytohelicase in citric buffer) for 30 min at 37°C. The material was immersed in freshly prepared 60% (v/v) acetic acid, and the samples were dissected on slides under a binocular microscope. The slides were placed on a hot plate at 50°C and the samples were spread by hovering a needle over the drop of acetic acid without touching the slide. After spreading the cells, the fixation was completed by dropping fresh 3:1 (v/v) fixative on the slides and immersing them in 60% (v/v) acetic acid for 10 min. The slides were dehydrated in 100% ethanol and air-dried, ready for future applications.
Haplotype phasing and scaffolding

A phased chromosome-level genome of *R. breviuscula* was assembled using PacBio HiFi and Hi-C data available from Hofstatter et al. under the NCBI Bioproject no. PRJNA784789. First, a phased primary assembly was obtained by running Hifiasm using as inputs the 30 Gb of PacBio HiFi reads (~35X coverage per haplotype) in combination with Dovetail Omni-C reads, using the following command: hifiasm -o Rbrevi.phased.asm.hic --h1 hic.R1.fastq.gz --h2 hic.R2.fastq.gz hifi.reads.fastq.gz. The phased assemblies of each individual haplotype were further scaffolded to chromosome scale using Salsa, followed by successive rounds of manual curation and re-scaffolding. The genome sizes of haplotypes 1 and 2 were 418,624,405 and 390,890,712 bp, respectively. Both haplotypes comprise five chromosomes with a length of ~370 Mb in total, as well as other unplaced sequences (Supplementary Table 1).

Hi-C map generation and haplotype comparison

Hi-C heatmap (Fig. 3c) was generated with juicer (v1.6) by aligning OmniC reads that used for genome assembly to the phased *R. breviuscula* genome. The Hi-C triangle for each chromosome in Fig. 6 was plot by fancplot (v0.9.1) with 500kbp resolution and KR normalization. Synteny blocks and structural rearrangements between two haplotypes (Fig. 3d) were computed by SyRi (v1.5.3) after aligning two haplotypes by minimap2 (v2.20).

Definition of allelic SNPs as genotyping markers on the phased reference genome

To define genotyping markers for *R. breviuscula*, all available (NCBI Bioproject no. PRJNA784789) raw Illumina HiSeq3000 150-bp paired-end reads (25,899,503,075 bases, ~54X coverage) were first mapped to the five pseudochromosome scaffolds in haplotype 1 of the phased reference genome using bowtie2 (v2.4.4). The alignment file was further sorted with SAMtools (v1.9). The alignments of short reads to the reference genome were used for SNP calling by `bcftools mpileup` and `bcftools call` (v1.9) (with the --keep-alts, --variants-only, and --multiallelic-caller flags enabled). A total of 1,404,927 SNPs excluding indels were derived. To distinguish the two haplotypes using these SNPs, only allelic SNPs were selected as markers for
genotyping; therefore, variant information was collected, including mapping quality, alternative base coverage, and allele frequency resulting from SHOREmap conversion (v3.6)\textsuperscript{87}, which converts SNP files (.vcf) into a read-friendly, tab-delimited text file. A final set of 820,601 alleles fulfilling certain thresholds (mapping quality > 50; 5 ≤ alternative base coverage ≤ 30, 0.4 ≤ allele frequency ≤ 0.6) was selected as markers (Fig. 4b; Supplementary Fig. 5).

Pre-processing single-cell RNA sequencing data from pollen nuclei

Raw scRNA-seq data usually include barcode errors and contaminants such as doublets and ambient RNA. In the present study, cell barcodes (CBs) were first corrected in these data using `bcctools correct` (v0.0.1) based on 10X v3 library complete barcode list with options “--alts 16 --spacer 12” because of the 16-bp CB and 12-bp unique molecular identifier (UMI). After correction, 952,535 viable CBs were detected. This step also truncated the CBs and UMIs from every pair of scRNA-seq reads. After counting the occurrence of CBs, the number of read pairs under each CB was determined. To ensure a sufficient number of reads for SNP calling, only CBs appearing more than 5,000 times were used for the subsequent analyses. Finally, each CB was seen as one viable cell, and reads corresponding to the CB were assigned to this cell (demultiplexing). A total of 8,001 viable cells were ultimately identified, with 365,771,748 (77.25% of all raw scRNA-seq) read pairs included. We also input the scRNA-seq data to 10X standard analysis pipeline - cellranger (v7.1.0) to check the statistics. The clustering analysis and the gene number in each cell (Supplementary Fig. 4) were based on cellranger count results.

Alignments of single-pollen RNA sequences to genome and deduplication

To identify genotyping markers in the \textit{R. breviuscula} gametes, scRNA reads of the pollen nuclei were first mapped to the haplotype 1 chromosomes (Fig. 4b) using hisat2 (v2.1.0)\textsuperscript{88}. Specifically, each cell-specific pair of reads was merged as one single-end FASTQ file, and hisat2 was run under single-end mode (-U) because the SNP-calling approach does not detect SNPs on reads whose mated reads are not mapped. Before further analyses of the alignment results, UMIs were previously extracted from the read alongside the CBs; hence, a fast UMI deduplication tool, UMIcollapse\textsuperscript{89}, was employed to remove the PCR duplicates by collapsing reads with the same UMIs.
The sequencing library was prepared for mixed pollen nuclei of *R. breviuscula* and *R. tenuis* to enable multiple-potential analyses. The addition of gametes from *R. tenuis* was done for multiplexing purposes, and they will be analysed in another study. To discriminate the single-cell data between the two species, we used a straightforward approach without gene expression profiling: For each cell, a) the DNA sequences were mapped to both the *R. breviuscula* and *R. tenuis* chromosomal genomes; and b) the alignment rates between the two species were compared to decide the cell identity (Supplementary Fig. 6). The alignment rates to *R. breviuscula* and *R. tenuis* were both bimodal distributions (Supplementary Fig. 6a–b); therefore, these cells can be grouped solely based on their mapping rates. It was estimated that 4,733 cells were from *R. breviuscula* and 2,709 cells were from *R. tenuis* (Supplementary Fig. 6c) based on the alignment fractions. The remaining 559 cells presented very similar alignment rates, which were potential doublets. Among the 4,733 *R. breviuscula* cells, those whose alignment rates were lower than 25% were discarded, leaving 4,392 cells from *R. breviuscula* available for the next stage of the analysis.

**SNP calling and selection of markers in gametes**

SNP calling in all gametes adopted the same methods as the reference genome SNP calling, e.g., via ‘bcftools mpileup’ and ‘bcftools call’ (v1.9), with the difference that the “--variants-only” flag was not applied. After acquiring SNPs for every gamete, the SNP positions, allele counts of the reference, and alternative bases were extracted through the ‘bcftools query’. Comparing SNPs in every gamete with markers defined on the reference resulted in reliable genotyping markers in this gamete.

Not all cells were suitable for CO calling due to insufficient markers or doublets generated during the 10X library construction; hence, filtering is necessary before CO calling. A total of 2,338 cells with fewer than 400 markers were first discarded to ensure accurate genotyping by sufficient markers. To remove doublets, the frequency of marker genotype switches across the remaining 2,054 cells was estimated. Cells with frequent switches, i.e., a switching rate (genotype switching times/number of markers) greater than 0.07, were taken as doublets (Supplementary Fig. 6e). Ultimately, 402 doublets were identified, with the remaining 1,652 cells proving suitable for subsequent CO calling.
CO identification

The chromosome genotyping was performed by adapting the haplotype phasing method proposed by Campoy et al.\textsuperscript{62}. The original approach was designed based on a scDNA-seq library, which is commonly able to examine more SNPs than scRNA-seq data. Therefore, the smoothing function and parameters were adjusted accordingly to genotype genomic blocks with relatively sparse markers. Specifically, the markers were first smoothed by allele frequencies of neighbouring markers (two ahead and two behind), and then smoothed by genotypes of surrounding markers. After smoothing, the genomic segments harbouring the markers in the same genotype are merged to genotype blocks, and the genotype blocks containing at least five markers within 1 Mb length were qualified to be assigned a final genotype. The genomic regions that saw the conversion of the genotypes at the flanks were taken as CO break positions (Fig. 4c–d). Finally, the CO numbers in each cell were counted and manually assessed, and those with double COs were corrected.

Recombination landscape and CO interference from single-gamete sequencing

To gain an overview of the CO rates across the chromosomes of \textit{R. breviuscula}, the CO positions in all viable cells (1,641 cells remaining after manual correction) were summarised, and the recombination landscape for each chromosome was plotted (Fig. 5a). Recombination rate (cM/Mb) was computed by 1-Mb sliding window and 100-kb step size.

\[
\text{Recombination rate} = \frac{\text{Number of COs within This Window} \times 100 \times 1M}{\text{Number of Cells} \times \text{Window Size}}
\]

CO interference was analysed with MADpattern (v.1.1)\textsuperscript{90}, using 1,641 confident singleton pollen nuclei. Chromosome 1 was divided into 18 intervals and chromosomes 2–5 were divided into 15 intervals to compute the mean CoC of every pair of intervals.

F1 offspring mapping and CO analysis

Sixty-three F1 offspring were reproduced from selfed \textit{R. breviuscula}. Each F1 plant was sequenced with ~3X Illumina WGS data. To genotype F1 offspring, WGS Illumina sequences of each plant were first mapped to rhyBreHap1 reference genome with
bowtie2 (v2.4.4) paired-end mode, and then SNPs were called by ‘bcftools mpileup’ and ‘bcftools call’ (v1.9) (with --keep-alts, --variants-only, and --multiallelic-caller flags enabled). Next, SNPs of each F1 sample were input to TIGER\textsuperscript{91} for genotyping and generating potential CO positions. In addition, RTIGER\textsuperscript{92} was also used to identify the genotypes of chromosomal segments by utilizing the corrected markers resulted from TIGER. Only the COs that agreed by both tools were kept. The recombination landscape from F1 COs (Supplementary Fig. 9b) was plotted using the same strategy and sliding window as illustrated for pollen nuclei.

**Genetic linkage maps**

To plot the genetic linkage maps (Fig. 5b and Supplementary Fig. 9a), 743 markers were extracted from the 820,601 reference markers by selecting the median marker within each 500-kb sliding window (step size was also 500 kb) from the first present marker until the last. The linkage map was then plotted by R package LinkageMapView. The Marey map (Fig. 5c) was plotted by the genetic length computed based on F1 offspring against physical length for each chromosome.

**ChIP**

CENH3 ChIP-seq data were obtained from Hofstatter et al.\textsuperscript{39}. Further ChIP experiments were performed for H3K4me3 (rabbit polyclonal to Histone H3 tri-methyl K4; Abcam ab8580), H3K9me2 (mouse monoclonal to Histone H3 di-methyl K9, Abcam ab1220), H3K27me3 (mouse monoclonal to Histone H3 tri-methyl K27, Abcam ab6002), and the IgG control (recombinant rabbit IgG, monoclonal Abcam ab172730) using the same protocol described by Hofstatter et al.\textsuperscript{39}.

**ChIP-seq and analysis**

ChIP DNA was quality-controlled using the next-generation sequencing assay on a FEMTO pulse (Agilent Technologies). An Illumina-compatible library was prepared with the Ovation Ultralow V2 DNA-Seq library preparation kit (Tecan Genomics) and sequenced as single-end 150-bp reads on a NextSeq2000 (Illumina) instrument. For each library, an average of 20 million reads were obtained.

Raw sequencing reads were trimmed using Cutadapt\textsuperscript{93} to remove low-quality nucleotides (with a quality score less than 30) and the adapters. Trimmed ChIPed 150-bp single-end reads were mapped to their respective reference genome using
bowtie2\textsuperscript{85} with default parameters. All read duplicates were removed and only the single best matching read was kept on the final alignment BAM file. The BAM files were converted into BIGWIG coverage tracks using the bamCompare tool from deeptools\textsuperscript{84}. The coverage was calculated as the number of reads per 50-bp bin and normalised as reads per kilobase per million mapped reads (RPKM). The magnified chromosome regions showing multiple tracks presented in Fig. 6g were plotted with pyGenomeTracks\textsuperscript{95}.

**Tyba array and CENH3 domain annotation**

Tyba repeats were annotated using a BLAST search with a consensus Tyba sequence, allowing a minimum of 70\% similarity. Further annotation of the Tyba arrays was performed by removing spurious low-quality Tyba monomer annotations shorter than 500 bp. Bedtools\textsuperscript{96} was used to merge all adjacent Tyba monomers situated at a maximum distance of 25 kb into individual annotations to eliminate the gaps that arise because of fragmented Tyba arrays, and those smaller than 2 kb were discarded.

CENH3 peaks were called with MACS3\textsuperscript{97} using the broad peak calling mode:

\begin{verbatim}
macs3 callpeak -t ChIP.bam -c Control.bam --broad -g 380000000 --broad-cutoff 0.1
\end{verbatim}

The identified peaks were further merged using a stepwise progressive merging approach. CENH3 domains were generated by 1) merging CENH3 peaks with a spacing distance less than 25 kb using bedtools to eliminate the gaps that arise because of fragmented Tyba arrays or due to the insertion of TEs; and 2) removing CENH3 domains less than 1 kb in size.

**Transposable element annotation**

Transposable element protein domains and complete LTR retrotransposons were annotated in the reference haplotype genome using the REXdb database (Viridiplantae\_version\_3.0)\textsuperscript{98} and the DANTE tool available from the RepeatExplorer2 Galaxy portal\textsuperscript{99}.

**Enzymatic methyl-seq and analysis**

To investigate the methylome space in *R. breviuscula*, the relatively non-destructive NEBNext Enzymatic Methyl-seq Kit was employed to prepare an Illumina-compatible
library, followed by paired-end sequencing (2 × 150 bp) on a NextSeq2000 (Illumina) instrument. For each library, 10 Gb of reads was generated.

Enzymatic methyl-seq data were analysed using the Bismarck pipeline following 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 into BIGWIG format and used as input tracks for the overall genome-wide DNA methylation visualisation with pyGenomeTracks and R plots.

**Quantitative correlation of COs and (epi)genetic features**

The distribution and accumulation of all the different classes of (epi)genetic features were correlated with the distribution of the COs. Correlation matrix (Fig. 6b) was calculated for all pairwise features by Pearson correlation coefficient using sliding window: specifically, mean CO rates, mean GC contents, CENH3 peak density, Tyba array density, SNP density, TE density, H3K4me3 RPKM, H3K9me2 RPKM, H3K27me3 RPKM, mean CpG, mean CHG, and mean CHH.

To inspect a possible centromere effect on CO positioning, the relative distance from the CO site was calculated to the closest left and right centromeric unit, i.e., the CENH3 domain or Tyba array, across the 378 COs in the F_1 offspring and normalised all distances to 0–1 such that all neighbouring centromeric units were displayed in the same scale (Fig. 6f and Supplementary Fig. 12e). Crossover and marker positions over the transcript bodies, CENH3 domain or Tyba array were normalised by their distance to start sites and end sites and then counted by binning (Fig. 6d–e and Supplementary Fig. 12d).

To see the association of CO designations with a variety of (epi)genetic features at a local scale, we first counted the number of COs that overlap with CENH3, Tyba arrays, genes, TEs, LTRs, H3K4me3 peaks, H3K9me2 peaks, and H3K27me3 peaks by ‘bedtools intersect’ (v2.29.0). Next, we assigned 378 pseudo-COs genome-wide at random. The number of COs on each chromosome was the same as that was detected by F1 individuals (e.g., 72 COs on chr1, 69 on chr2, 76 on chr3, 84 on chr4, and 77 on chr5), while the CO break gap length was picked up from the 378 real F1 CO gaps randomly. For each simulation round, the pseudo-COs were overlapped with
(epi)genetic features again with ‘bedtools intersect’. Five thousand of these simulations were done, and the results were then plotted as the distribution of overlapped CO numbers for each feature (Supplementary Fig. 13). Finally, to evaluate the deviation of real overlapped COs with each feature to the expected overlapped CO number under the hypothesis of randomly distributed COs, Z-scores were calculated by the mean values and standard deviations of the simulated number of overlapped CO distribution (Fig. 6c).

**Gene annotation**

Structural gene annotation was done combining de novo gene calling and homology-based approaches with *Rhynchospora* RNAseq, IsoSeq, and protein datasets already available. Using evidence derived from expression data, RNAseq data were first mapped using STAR (version 2.7.8a) and subsequently assembled into transcripts by StringTie (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 (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 (version 2018-07-04). All assembled transcripts from RNAseq, IsoSeq, and aligned protein sequences were combined using Cuffcompare (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 identify potential open reading frames and to predict protein sequences within the candidate transcript set.

Ab initio annotation was initially done using Augustus (version 3.3.3). GeneMark (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 Nachtweide and Stanke. 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 Nachtweide and Stanke.
All structural gene annotations were joined using EVidenceModeller\textsuperscript{108} (version 1.1.1), and weights were adjusted according to the input source: ab initio (Augustus: 5, GeneMark: 2), homology-based (10). Additionally, two rounds of PASA\textsuperscript{109} (version 2.4.1) were run to identify untranslated regions and isoforms using the above described IsoSeq datasets.

We used DIAMOND\textsuperscript{110} (v2.0.5) 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; https://trep-db.uzh.ch), 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 \(10^{-10}\) 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 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\textsuperscript{111} (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.

\textbf{Ka/Ks ratio calculation}
We identified homologs between *Brachypodium distachyon* (v3.0) (downloaded from ensembl plants.ensembl.org) and *Juncus effesus* using the ortholog module from JCVI python library. Subsequently, pairwise alignments were generated with ParaAT (v2) and the Ka/Ks ratio was calculated using KaKs_Calculator (v3) using the YN method. Plots were generated using karyoplateR.
References


Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- CastellaniandZhangetalSupplementary.15.08.2023revised.pdf