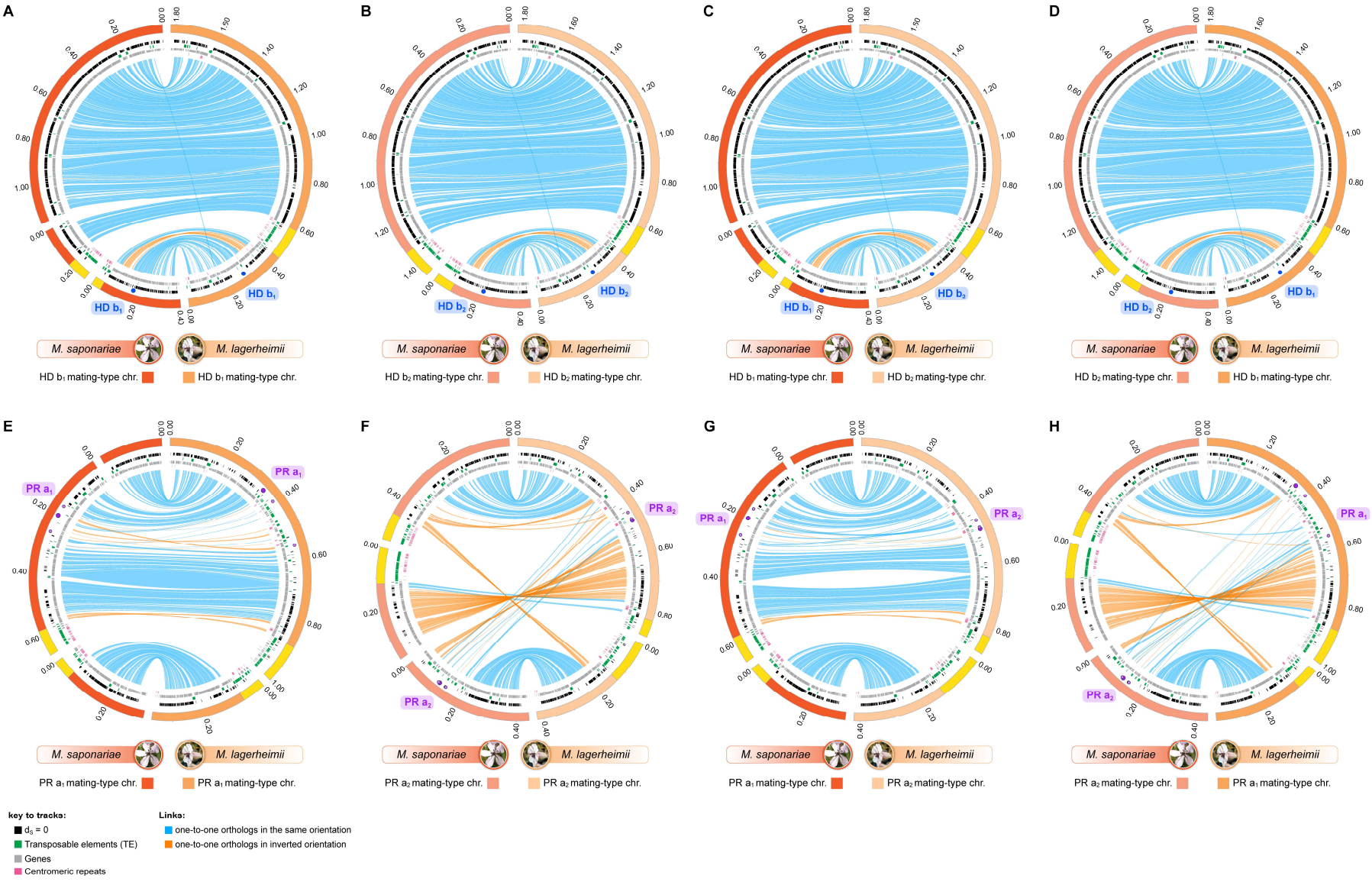
Supplemental Material

Convergent recombination cessation between mating-type genes and centromeres in selfing anther-smut fungi

Fantin Carpentier*†[1], Ricardo C. Rodríguez de la Vega†[1], Sara Branco[1,2], Alodie Snirc[1], Marco A. Coelho[3,4], Michael E. Hood#[5], Tatiana Giraud#[1]

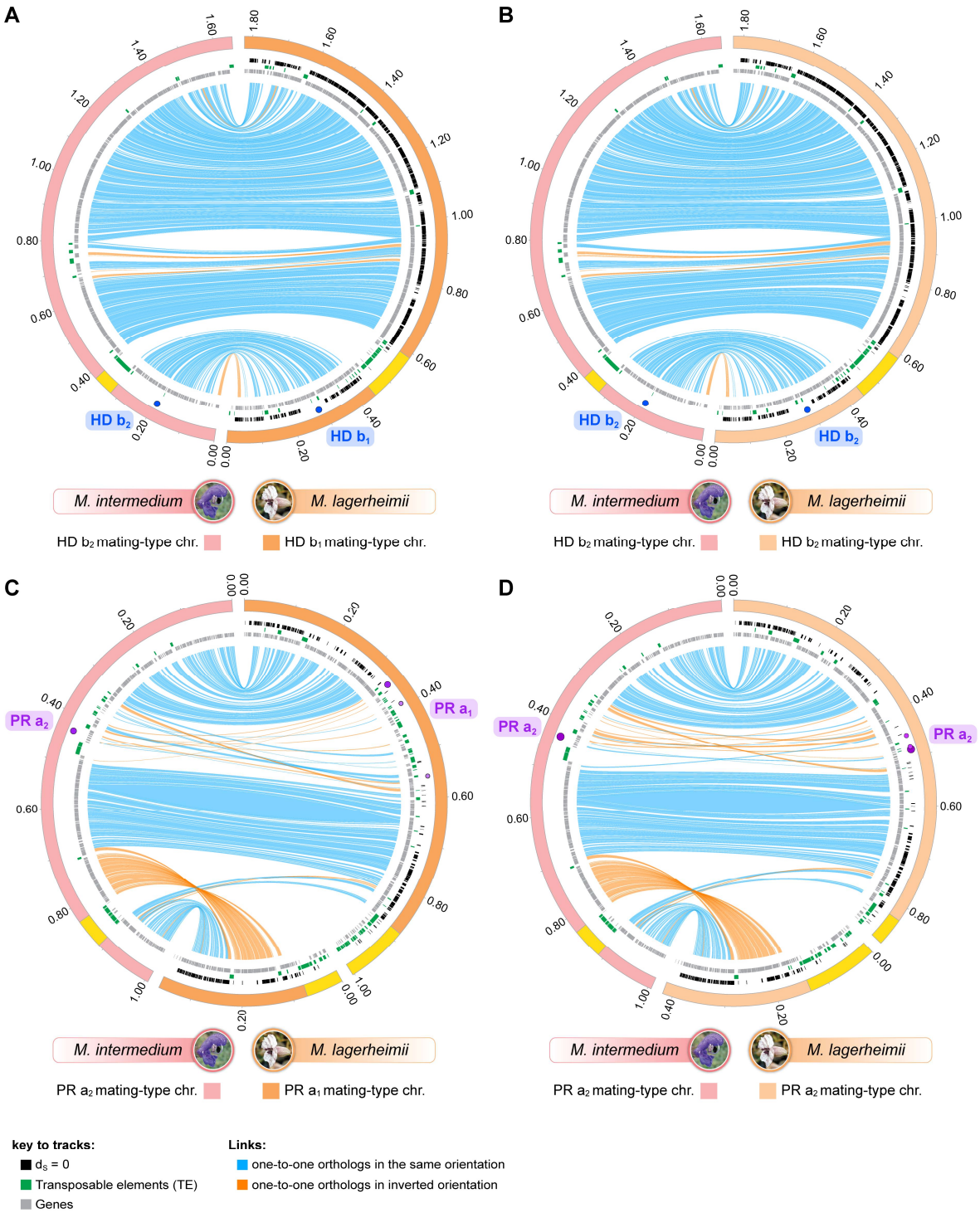
Table of contents

Supplemental Fig. S1



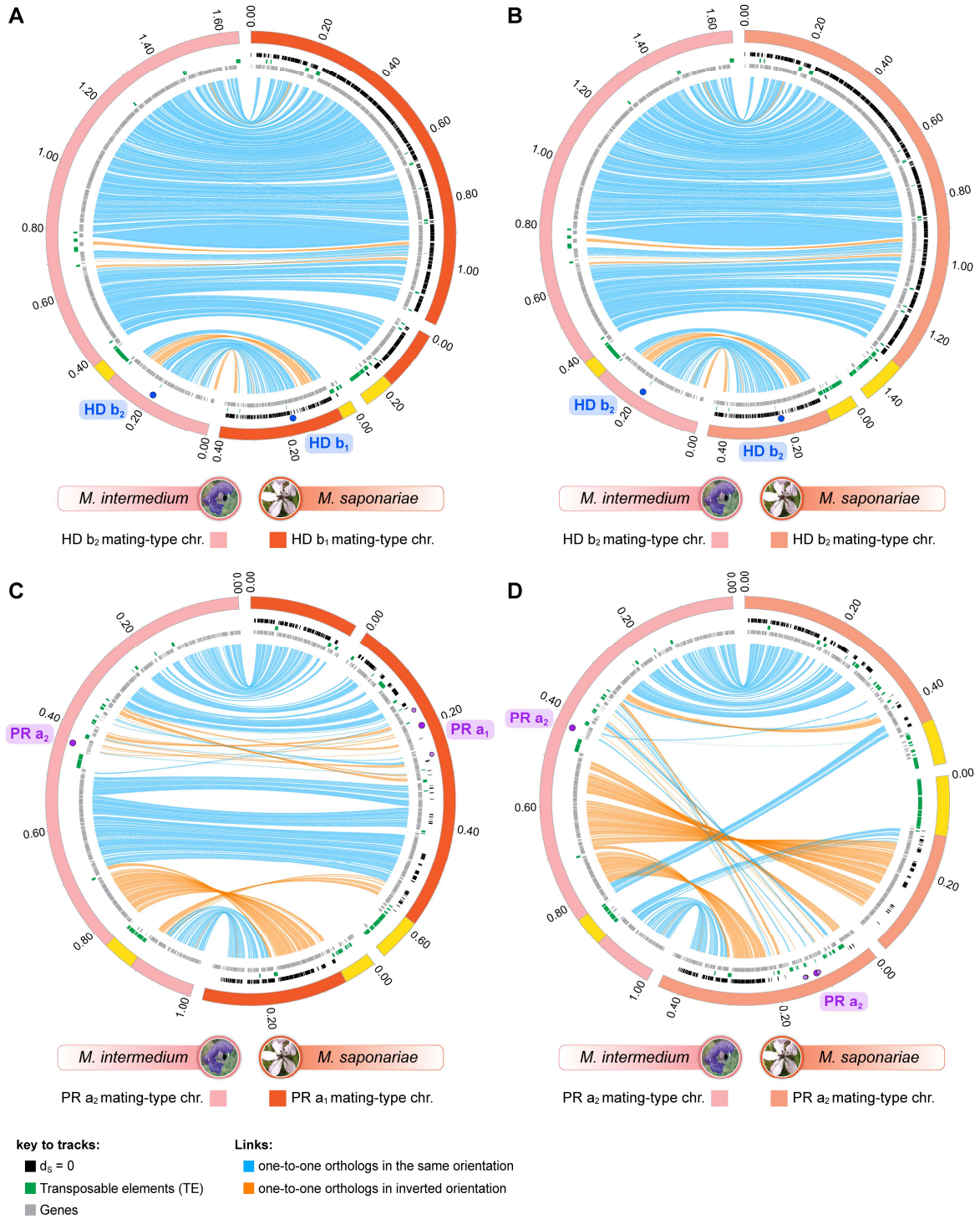
Supplemental Fig. S1. Interspecific gene order comparison of PR and HD mating-type chromosomes of *Microbotryum lagerheimii* and *M. saponariae*. The outer track represents contigs, staggered every 200 kilobases. The HD, PR and pheromone genes are indicated by blue, dark purple and small light-purple circles, respectively. Blue and orange lines link single-copy orthologs, the latter corresponding to inversions. The link width is proportional to the corresponding gene length. Yellow regions on the contig track indicate centromeres, *i.e.* regions with low gene density, high TE density and enriched in tandem repeats (pink marks). The black marks along the contigs track indicate genes that have no synonymous substitutions between mating types within individuals (d_S=0). Green marks indicate the transposable elements (TEs) and grey marks non-TE genes. Pink tracks indicate the position of *de novo* detected tandem repeats. **(A)** Comparison of the *M. saponariae* (left, red) and *M. lagerheimii* (right, orange) b_1 HD chromosomes. (**B)** Comparison of the *M. saponariae* (left, light red) and *M. lagerheimii* (right, light orange) b_2 HD chromosomes. **(C)** Comparison of the *M. saponariae* b_1 (left, red) and *M. lagerheimii* b_2 (right, light orange) HD chromosomes. **(D)** Comparison of the *M. saponariae* b_2 (left, light red) and *M. lagerheimii* b_1 (right, orange) HD chromosomes. **(E)** Comparison of the *M. saponariae* (left, red) and *M. lagerheimii* (right, orange) a_1 PR mating-type chromosomes. **(F)** Comparison of the *M. saponariae* (left, light red) and *M. lagerheimii* (right, light orange) PR a_2 mating-type chromosomes. **(G)** Comparison of the *M. saponariae* a_1 (left, red) and *M. lagerheimii* a_2 (right, light orange) PR mating-type chromosomes. **(H)** Comparison of the *M. saponariae* a_2 (left, light red) and *M. lagerheimii* a_1 (right, orange) PR mating-type chromosomes.

Supplemental Fig. S2



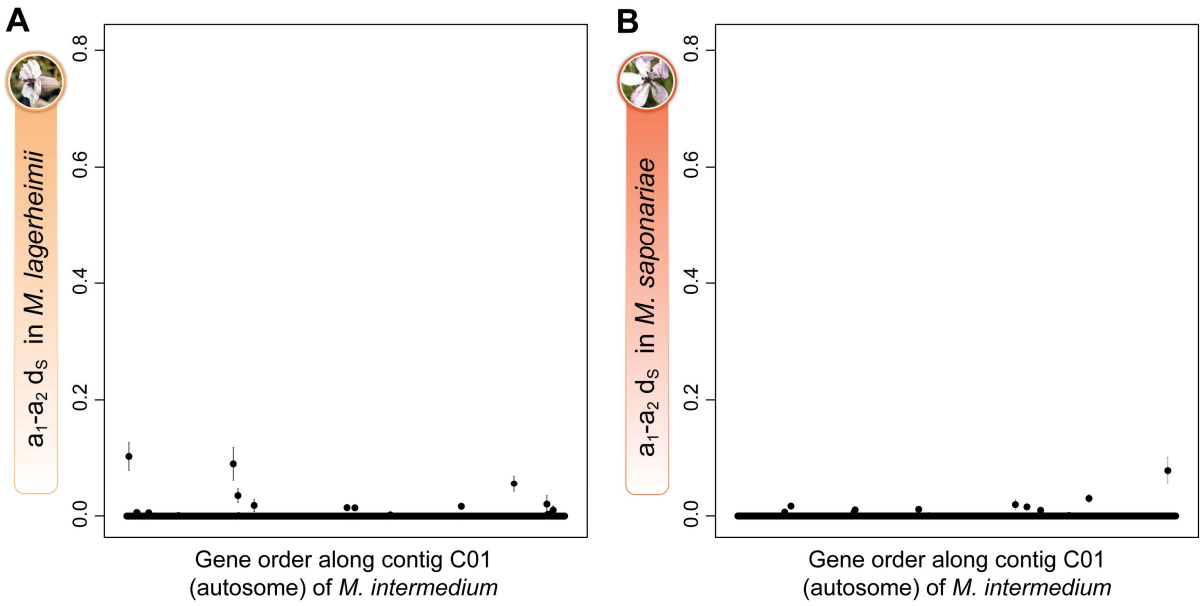
Supplemental Fig. S2. Comparison of gene order between the mating-type chromosomes of *Microbotryum lagerheimii* and *M. intermedium*. The outer track represents contigs, staggered every 200 kilobases. The HD, PR and pheromone genes are indicated by blue, dark purple and small light-purple circles, respectively. Blue and orange lines link single-copy orthologs, the latter corresponding to inversions. The link width is proportional to the corresponding gene length. Yellow regions on the contig track indicate centromeres. The black marks along the contigs track indicate genes with no synonymous substitutions between mating-type alleles within individuals (d_S=0). Because only one haploid genome was available for *M. intermedium*, no d_S values were computed. Green marks indicate transposable elements (TEs) and grey marks non-TE genes. **(A)** Comparison of the *M. intermedium* b_2 (left, pink) and *M. lagerheimii* b_1 (right, orange) HD chromosomes. **(B)** Comparison of the *M. intermedium* b_2 (left, pink) and *M. lagerheimii* b_2 (right, light orange) HD chromosomes. **(C)** Comparison of the *M. intermedium* a_2 (left, pink) and *M. lagerheimii* a_1 (right, orange) PR chromosomes. **(D)** Comparison of *M. intermedium* a_2 (left, pink) and *M. lagerheimii* a_2 (right, light orange) PR chromosomes.

Supplemental Fig. S3



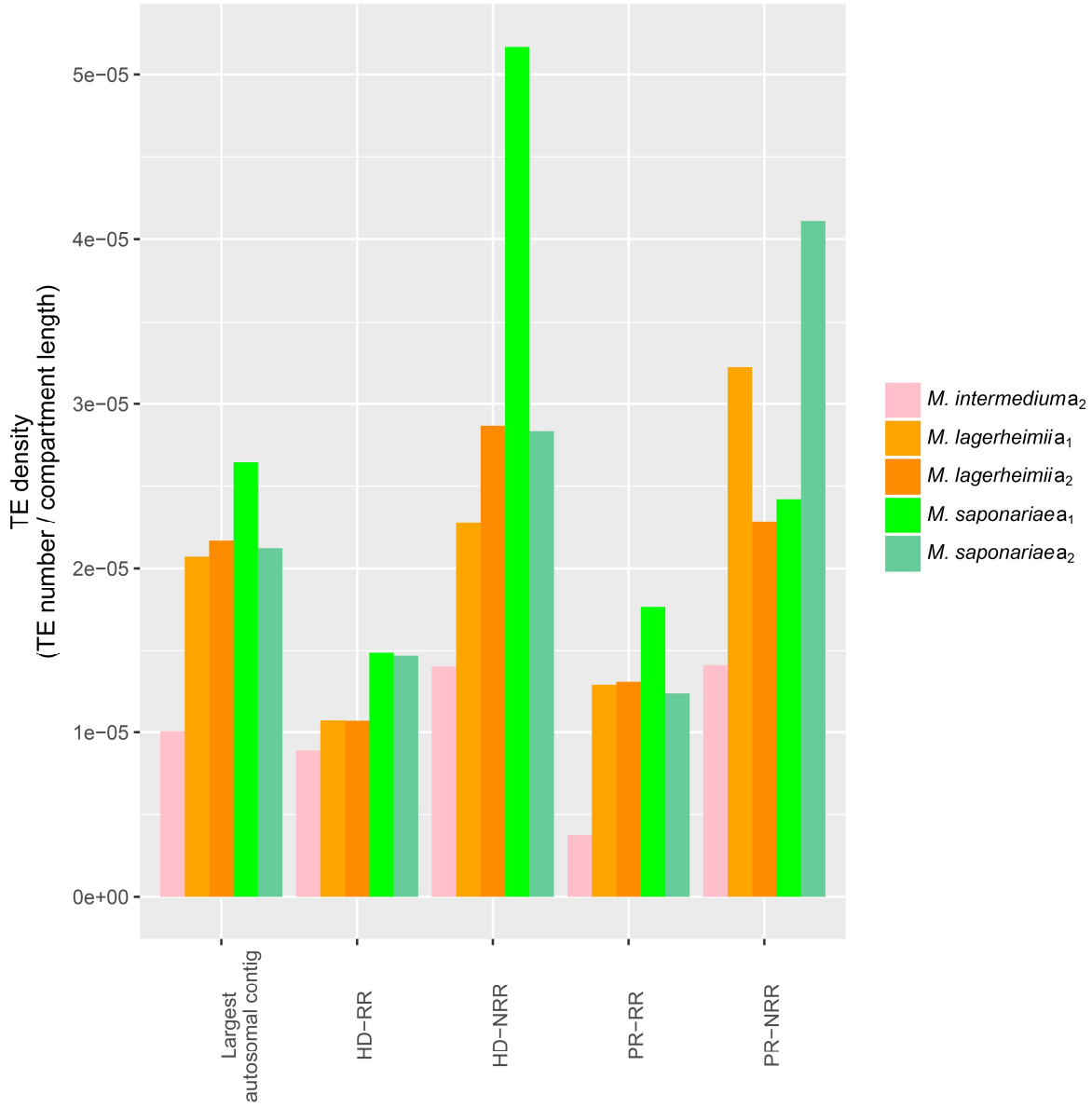
Supplemental Fig. S3. Comparison of gene order between the mating-type chromosomes of *Microbotryum saponariae* and *M. intermedium*. The outer track represents contigs, staggered every 200 kilobases. The HD, PR and pheromone genes are indicated by blue, dark purple and small light-purple circles, respectively. Blue and orange lines link single-copy orthologs, the latter corresponding to inversions. The link width is proportional to the corresponding gene length. Yellow regions on the contig track indicate centromeres. The black marks along the contigs track indicate genes with no synonymous substitutions between mating-type alleles within individuals (d_S=0). Because only one haploid genome was available for *M. intermedium*, no d_S values were computed. Green marks indicate transposable elements (TEs) and grey marks non-TE genes. **(A)** Comparison of the *M. intermedium* b_2 (left, pink) and *M. saponariae* b_1 (right, red) HD chromosomes. **(B)** Comparison of the *M. intermedium* (left, pink) and *M. saponariae* (right, light red) b_2 HD chromosomes. **(C)** Comparison of the *M. intermedium* a_2 (left, pink) and *M. saponariae* a_1 (right, red) PR chromosomes. **(D)** Comparison of the *M. intermedium* (left, pink) and *M. saponariae* (right, light red) a_2 PR chromosomes.

Supplemental Fig. S4



Supplemental Fig. S4. Per-gene synonymous divergence between mating types and its respective standard error (dS ± SE) between autosomal alleles of (A) *Microbotryum lagerheimii* and (B) *M. saponariae*, along the ancestral gene order of a *M. intermedium* autosome. Synonymous divergence is plotted against the genomic coordinates of an autosome of *M. intermedium* for all single-copy genes shared by this autosome, as a proxy for ancestral gene order. Almost all the autosomal genes show a null synonymous divergence between mating types within individuals, as expected in highly selfing organisms such as anther-smut fungi.

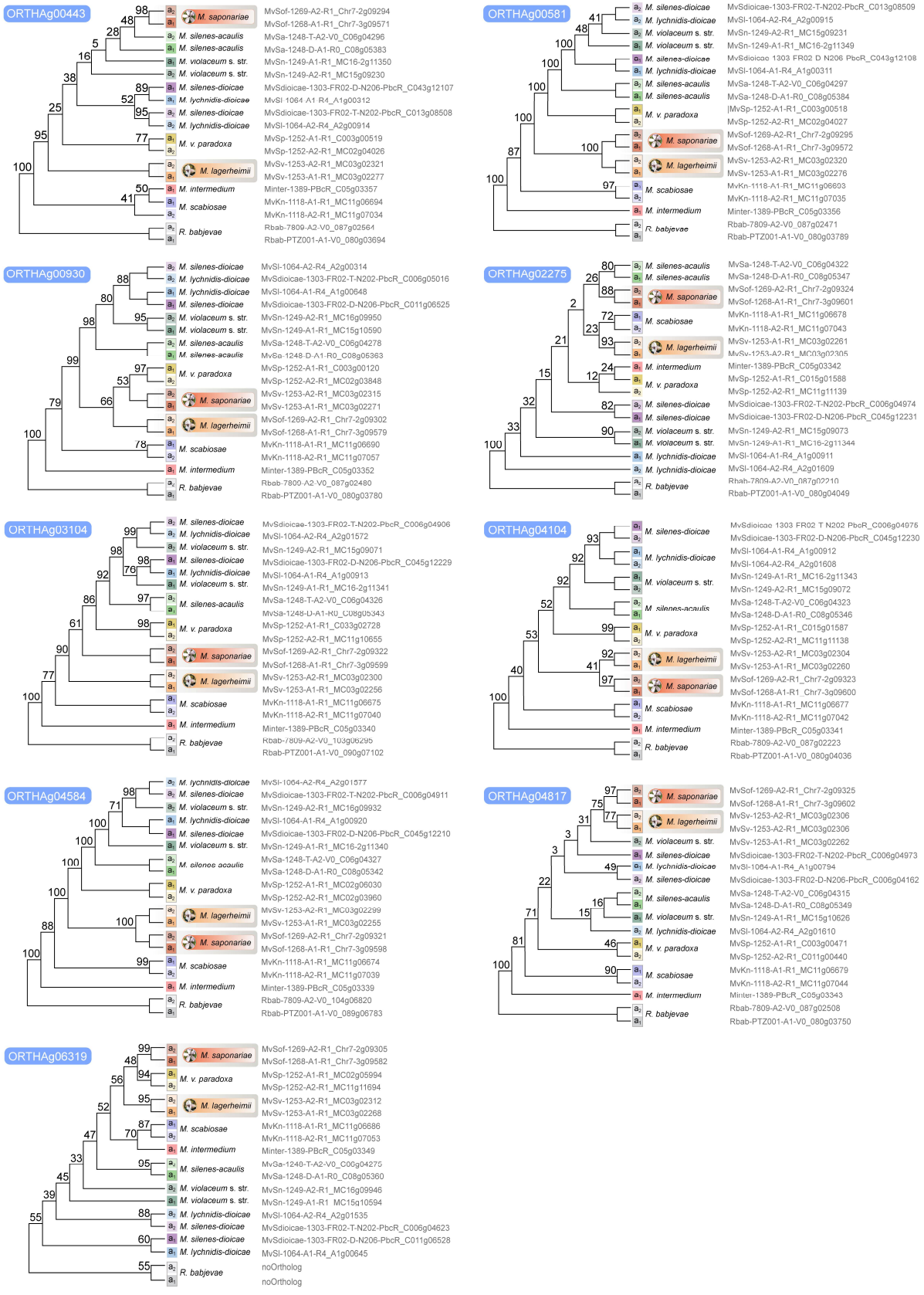
Supplemental Fig. S5



Supplemental Fig. S5. Density of transposable elements (TEs) in *Microbotryum* species. The TE density (number of TEs detected divided by the compartment length) is plotted per genomic compartment, i.e. for the largest autosomal contig, the recombining region, RR, the non-recombining region, NRR, of either the PR or HD mating-type chromosomes, for each available haploid genome of the three studied species with mating types segregating independently: *Microbotryum intermedium* (a_2 genome), *M. lagerheimii* and *M. saponariae* (a_1 and a_2 genomes).

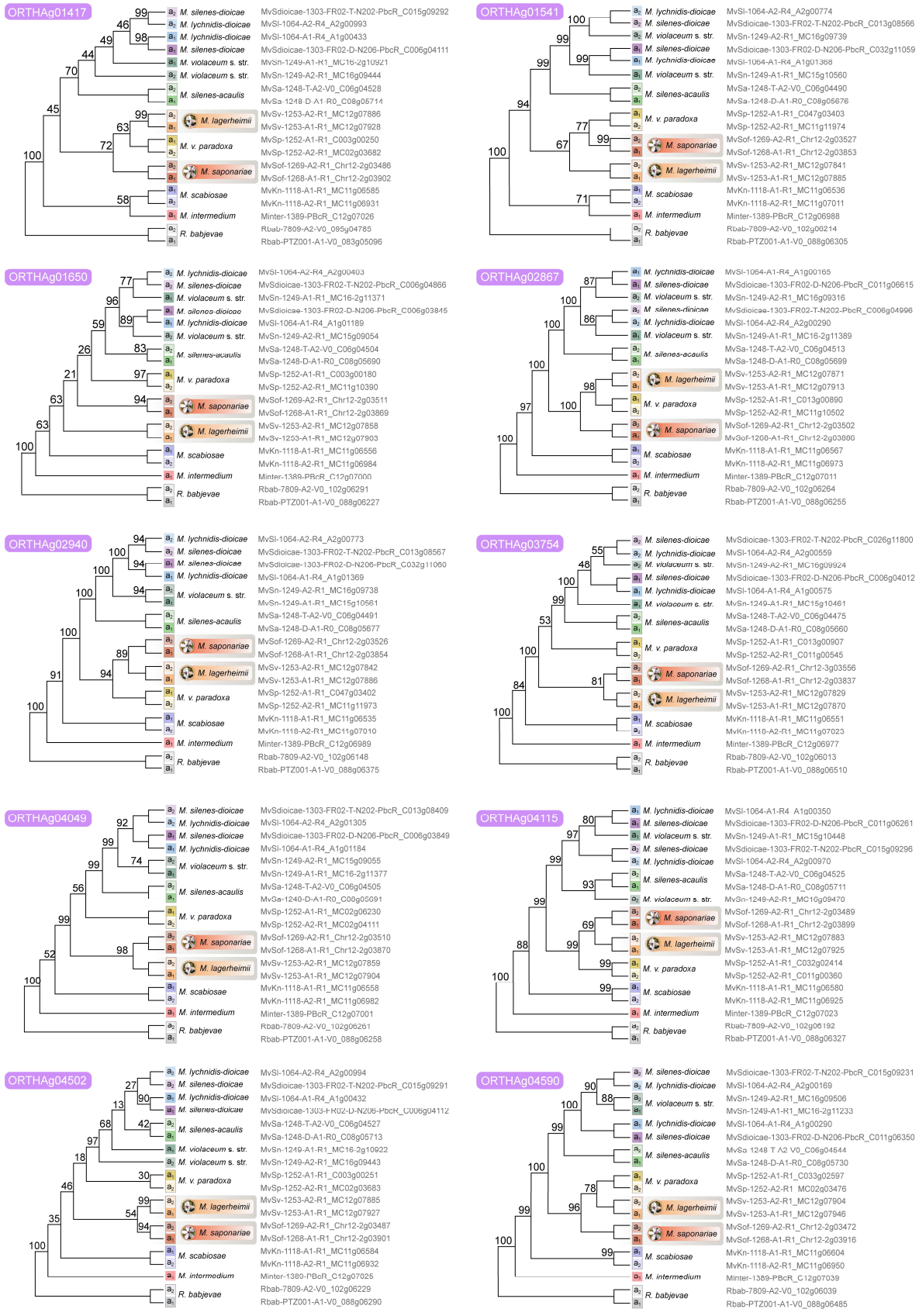
Supplemental Fig. S6A

A



Supplemental Fig. S6B

B



Supplemental Fig. S6. Individual genealogies for the 19 of the genes used for dating the linkage between mating-type loci and centromeres. The individual genealogies of these genes located between the centromeres and the HD-proximal or PR-proximal strata containing the mating-type loci illustrate the lack of trans-specific polymorphism in this genomic region between *Microbotryum lagerheimii* and *M. saponariae*, supporting that complete recombination cessation between the mating-type loci and the centromeres occurred independently in the two species. **(A)** Gene genealogies of nine genes located between the HD-proximal stratum and the centromere. **(B)** Gene genealogies of ten genes located between PR-proximal and the centromere.

Supplemental Table S1. Statistics on the genome assemblies of the *Microbotryum saponariae* genomes analysed in this study

Sample	Accession Numbers	Assembly statistics									
		# Contigs	Length of the smallest contig (bp)	Length of the largest contig (bp)	N50 (bp)	L50 (# contigs)	N90 (bp)	L90 (# contigs)	Mean Length (bp)	Median length (bp)	Assembly size (bp)
Microbotryum saponariae from *Saponaria officinalis* (1268) a_1	GCA_900015975	161	3,858	2,799,476	945,666	10	195,491	32	183,285.323	20,702	29,508,937
Microbotryum saponariae from *Saponaria officinalis* (1269) a_2	GCA_900015475	72	9,639	2,733,711	1,443,984	9	441,166	22	397,966.3472	35,381.5	28,653,577

Supplemental Table S2. Statistics of the mating-type chromosomes of the *Microbotryum saponariae* genomes analysed in this study, and in the different genomic partitions of the mating-type chromosomes: recombining regions (RR), non-recombining regions (NRR), pseudo-autosomal regions (PAR), homeodomain gene (HD) mating-type chromosome and pheromone receptor gene (PR) mating-type chromosome.

Genomics regions	Statistics
Number of mating-type chromosomes (haploid)	2
Number of contigs for the a1 HD mating-type chromosome	3
Number of contigs for the a2 HD mating-type chromosome	2
Number of contigs for the a1 PR mating-type chromosome	3
Number of contigs for the a2 PR mating-type chromosome	3
Size of the a1 HD mating-type chromosome (bp)	1,836,444
Size of the a2 HD mating-type chromosome (bp)	1,903,821
Size of the a1 PR mating-type chromosome (bp)	1,179,386
Size of the a2 PR mating-type chromosome (bp)	1,346,385
Size (percentage) of the RR on the a1 HD mating-type chromosome (bp)	85.19% (1,564,501 bp)
Size (percentage) of the NRR on the a1 HD mating-type chromosome (bp)	7.59% (139,334 bp)
Size (percentage) of the RR on the a2 HD mating-type chromosome (bp)	81.37% (1,549,232 bp)
Size (percentage) of the NRR on the a2 HD mating-type chromosome (bp)	7.34% (139,779 bp)
Size(percentage) of the RR on the a1 PR mating-type chromosome (bp)	46.55% (548,957 bp)
Size (percentage) of the NRR on the a1 PR mating-type chromosome (bp)	41.69% (491,719 bp)
Size (percentage) of the RR on the a2 PR mating-type chromosome (bp)	39.33% (529,587 bp)
Size (percentage) of the NRR on the a2 PR mating-type chromosome (bp)	45.47% (612,166 bp)
Size (percentage) of the centromere on the a1 HD mating-type chromosome (bp)	7.22% (132,606 bp)
Size (percentage) of the centromere on the a2 HD mating-type chromosome (bp)	11.28% (214,807 bp)
Size(percentage) of the centromere on the a1 PR mating-type chromosome (bp)	11.76% (138,706 bp)
Size (percentage) of the centromere on the a2 PR mating-type chromosome (bp)	15.2% (204,628 bp)

Supplemental File S1. Centromeric repeats *de novo* detected (provided as a separate file).

Supplemental methods

Segregation analyses

For mating-type segregation analysis we used *M. lagerheimii* collected on *Lychnis flos-jovis* in Valle Pesio, Italy (GPS 44.188400, 7.670650). *Microbotryum* fungi undergo meiosis immediately following spore germination, and the septate basidium allows for the micromanipulation and isolation of post-meiotic yeast-like cells as linear tetrads (Hood et al. 2015). Haploid cells were isolated from opposite poles of meiosis I across replicate meioses from the same diploid parent and characterized for mating type segregation by PCR amplification of allele-specific markers. For the PR locus, primers 660 and 588 were used that discriminate between a_1 and a_2 alleles based on allele-specific amplification (Devier et al. 2009; Hood et al. 2015). For the HD locus, the primer pair HD4-F5 (5' CCATCGAGCTCCTTTTACCC) and HD-R1 (5' TCTAGGCAGCTCTTGCTC) was designed to produce PCR products of allele-specific size due to insertion/deletion mutations. Under centromere linkage, variation at heterozygous loci should segregate at the first meiotic division and thus always differing between meiotic products from separated at meiosis I; a crossing-over recombination between the locus and the centromere would result in different alleles between two meiotic products separated at meiosis I only fifty percent of the time. The observed proportion of instances where haploid product separated at meiosis I carried alternate alleles for both the PR and HD locus was calculated for 78 meioses, the corresponding 95% confidence interval was calculated using Vassarstats®.

DNA extraction and sequencing

We isolated haploid cells of opposite mating types from single tetrads using micromanipulation as described previously (Hood et al. 2015) for *M. saponariae* parasitizing *S. officinalis* (cell 1268, PRAT 47, a1 b1, and cell 1269, PRAT 48, a2 b2) collected near Chiusa di Pesio, (GPS coordinates 44.31713297, 7.622967437 on July 8th, 2012). DNA was extracted with the QIAGEN Genomic-tip 100/G (ref. 10243; Courtaboeuf, France) and Genomic DNA Buffer Set (ref. 19060) following manufacturer instructions and using a Carver hydraulic press (reference 3968, Wabash, IN, USA) for breaking cell walls. Haploid genomes were sequenced using the P6/C4 Pacific Biosciences SMRT technology (UCSD IGM Genomics Facility La Jolla, CA, USA).

Assembly and annotation

Assemblies of the genomes were generated with the wgs-8.2 version of the PBcR assembler (Koren et al. 2012) with the following parameters: genomeSize=30000000, assembleCoverage=50. Assemblies were polished with quiver software (https://github.com/PacificBiosciences/GenomicConsensus). A summary of raw data and assembly statistics is reported in Table S1 and S2. Contigs were aligned with optical maps of the two mating-type chromosomes obtained previously (Hood et al. 2015), with MapSolver software (OpGen), allowing generating oriented a_1 and a_2 pseudomolecules for each mating type chromosome. Mating-type chromosomes were identified by finding the contigs carrying the PR and HD mating-type genes using BLAST (Altschul et al. 1990).

Species tree and species incorporated in the study

To study the evolution of suppressed recombination in a phylogenetic context, we reconstructed the relationships between the *Microbotryum* species for which high-quality genomes were available (Fig. 2), with linked or unlinked mating-type loci: in addition to the newly sequenced *M. saponariae* genomes of opposite mating types, we used the high-quality a_1 and a_2 genomes of seven available *Microbotryum* species (*M. lychnidis-dioicae*, *M. silenes-dioicae*, *M. violaceum sensu stricto*, *M. lagerheimii*, *M. silenes-acaulis*, *M. scabiosae*, *M. violaceum paradoxa*), a high-quality a_2 genome of *M. intermedium*, and a high-quality outgroup genome of the red yeast *Rhodosporidium babjevae*.

The previously published genome assemblies of the species used in this study are available at the GenBank under the following accession numbers: GCA_900015445 for *M. lychnidis-dioicae* 1064 a_2; GCA_900015465 for *M. lychnidis-dioicae* 1064 a_1; GCA_900013405 for *M. lagerheimii* 1253 a_2; GCA_900015505 for *M. lagerheimii* 1253 a_1; GCA_900015455 for *M. violaceum s. str.* 1249 a_2; GCA_900015425 for *M. violaceum s. str.* 1249 a_1; GCA_900096595 for *M. intermedium*; SAMN09670553 for *M. silenes-dioicae* 1303 a_2; and GCA_900120095 for *M. silenes-dioicae* 1303 a_1; GCA_900015485 *M. violaceum paradoxa* 1252 a_2; GCA_900015495 for *M. v. paradoxa* 1252 a_1; PRJEB12080 and ERZ250708 for *M. scabiosae* 1118 a_2; PRJEB12080 and ERZ250707 for *M. scabiosae* 1118 a_1; GCA_900014955 for *M. v. caroliniana* 1250 a_2; GCA_900014965 for *M. v. caroliniana* 1250 a_1; SAMN09670554 for *M. silenes-acaulis* 1248 a_1; SAMN09670555 for *M. silenes-acaulis* 1248 a_2.

We used the translated gene models for the nine *Microbotryum* species and the outgroup *Rhodotorula babjevae* to obtain orthologous groups with orthAgogue (Ekseth et al. 2014) based on blastp+ 2.2.30 followed by Markov clustering (Van Dongen 2000). We aligned the protein sequences of 780 fully conserved single-copy genes with MAFFT v7.388 (Katoh and Standley 2013) and obtained the codon-based CDS alignments with TranslatorX (Abascal et al. 2010). We used RAxML 8.2.7 (Stamatakis 2006) to obtain maximum likelihood gene trees for all 780 fully conserved single-copy genes and a species tree with the concatenated alignment of 447,405 codons under the

GTRGAMMA substitution model. We estimated the branch support values by bootstrapping the species tree based on the concatenated alignment and by estimating the relative internode and tree certainty scores based on the frequency of conflicting bipartitions for each branch in the species tree among the fully conserved single-copy genes (Salichos et al. 2014).

We removed all transposable elements (see the *de novo* TE identification method) from the gene dataset used in all analysis, assuming a gene to be a TE if the gene sequence shares more than 50% of TE sequence.

We identified alleles as orthologous groups with a single sequence in each haploid genome for a given species. We used TranslatorX (Abascal et al. 2010) with the MUSCLE aligner v3.8.31 (Edgar 2004) to align nucleotide sequences of predicted genes. We estimated synonymous divergence (d_S) and its standard error with the yn00 program of the PAML package (Yang 2007) and we plotted d_S along the chromosomes using the ggplot2 R package (Wickham 2016). The gene assignment to the PR- or HD-proximal stratum was made using previously gene assignments (Branco et al. 2017, 2018), and by considering genes previously unassigned to a stratum and that are now identified as being in-between genes in the PR-proximal stratum. The latter genes were not assigned to the HD- or PR-proximal strata because the gene assignment was made based on d_S plot using the *M. lagerheimii* gene order, which was slightly different from the *M. intermedium* gene order used in this study.

Figures

We prepared the figures 3, S1, S2 and S3 using Circos (Krzywinski 2009). We linked alleles in Circos plots comparing contigs belonging to alternative mating-types of a given species, and the ortholog genes in Circos plots comparing contigs from distinct species.

Date estimates for recombination cessation

We used the allele codon-based alignments within each species based on the rationale that the divergence between alleles associated to the a_1 versus a_2 mating types is dependent on the time since recombination cessation. We used 9 orthologous groups (8,525 aligned codons) for dating the recombination cessation between the HD-proximal stratum and the centromere, and 10 orthologous (10,200 aligned codons) groups for dating recombination cessation between the PR-proximal stratum and the centromere (these were the genes for which a_1 versus a_2 alleles were available in all species studied and located between the ancient mating-type strata and the centromere and are indicated with red arrows on Figure 4). Divergence times were estimated using BEAST v2.4.0 (Drummond and Rambaut 2007), with the xml inputs being generated using BEAUTi (Drummond et al. 2012), and setting the following parameters (others left as default values): unlinked substitution (HKY+G with empirical frequencies for each codon position) and clock models, Yule process to model speciation, and 5,000,000 MCMC generations sampled every 1,000. For all runs, we used a single calibration prior at 0.42 million years, corresponding to the divergence between *M. lychnidis-dioicae* and *M. silenes-dioicae* (Gladieux et al. 2011), with a normal distribution and a sigma of 0.04. Time trees were annotated with BEAST's TreeAnnotator tool setting burnin to discard 10% of the trees. The divergence between *M. lychnidis-dioicae* and *M. silenes-dioicae* have been estimated earlier to 0.42 million years through the same approach, using another calibration point (Gladieux et al. 2011).

Gene genealogies

Gene genealogies were inferred for codon-based alignments of genes using RAxML (Stamatakis 2006) version 8.2.7, assuming the GTRGAMMA model and rapid bootstrap (options: -f a and -# 100). We analysed all single-copy genes for which we had both alleles in all species and were located between the HD-proximal or PR-proximal strata and the corresponding centromeres.

Transposable element identification, annotation and detection

Transposable elements were identified and annotated *de novo* in the *Microbotryum* high-quality genome assemblies, using both LTR-harvest (defaults parameters; Ellinghaus et al. 2008); and RepeatModeler (defaults parameters; Smit and Hubley 2015). Transposable element sequences were clustered per family to get a consensus sequence per annotation, using *usearch* (centroid method, id=0.7; Edgar 2010). These consensus sequences were compiled to form a *Microbotryum* transposable elements database (Hartmann et al. 2018), and were annotated with blast using Repbase (Bao et al. 2015) database (20.05) that we used as library in RepeatMasker (Smit et al. 2015) to retrieve their genomic location in all genomes.

De novo detection of centromeric repeats

Centromeric regions are poor in gene and rich in transposable and repetitive elements. As several regions fulfilled these criteria in each contig, we identified *de novo* centromeric-specific repeats (Melters et al. 2013) using Tandem-Repeat Finder (TRF v. 4.07b; Benson 1999) on assembled Illumina reads of the very same strains as those sequenced using the Pacific Bioscience technology. The mate-pair reads used to detect centromeric repeats in *M. lagerheimii* are available in the sequence read archive, with the accession number SRR7047936; the mate-pair reads used to detect centromeric repeats in *M. saponariae* were previously published (Fortuna et al. 2016). For both *M. lagerheimii* and *M. saponariae* genomes, we performed the assemblies as follows: we randomly chose 500,000 Illumina reads that we assembled with PRICE (v1.2; Ruby et al. 2013) using a random set of 1,000,000 reads as seed file, and using the following command line arguments: -mpp inputFile_R1 inputFile_R2 650 90 -picf 20000 seedFile 500 2 25 -nc 10 -mpi 85 -MPI 95 – tpi 85 -TPI 95 -logf logfile -o outputFile. PRICE works by round of assembly: in the first round, it maps randomly picked reads onto contigs (provided by the seedFile), assembles the reads that did not mapped, and then extends the contig with the unmapped assembled sequences. For the second and following rounds, PRICE considers the extended contigs as the reference to restart the process of picking, mapping reads, assembling the unmapped reads and extending the reference contigs. We analysed the presence of tandem repeats in each of the 10 assembly cycle output, using the following parameters in a TRF wrapper perl script (Melters et al. 2013): match=1, mismatch=1, indel=2, probability of match=80, probability of indel=5, min score=200, max period=2000. We performed these steps 15 times, picking randomly 500,000 input reads and 1,000,000 reads for the seed file. The repeats detected in the Illumina genomes were blasted against the corresponding high-quality genomes of *M. lagerheimii* and *M. saponariae*. We defined the centromeric regions of the mating-type chromosomes by identifying at the largest TE-rich, gene-poor regions containing the greatest density of tandem repeats. The delimitations of the centromeric regions using this method were congruent with those using BLAST of the *M. lychnidis-dioicae* centromeric repeats identified previously *in M. lychnidis-*

dioicae (Badouin et al. 2015). To identify centromeric repeats in *M. intermedium*, we therefore blasted the centromeric repeats identified in *M. lychnidis-dioicae* (Badouin et al. 2015), as no Illumina reads were available for this species. FASTA files containing the *de novo* identified centromeric repeats are provided in Supplemental File S1.

References

Abascal F, Zardoya R, Telford MJ. 2010. TranslatorX: Multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Res* **38**(Web Server Issue): W7-13.

Altschul SF, Gish W, Pennsylvania T, Park U. 1990. Basic Local Alignment Search Tool. : 403–410.

Badouin H, Hood ME, Gouzy J, Aguileta G, Siguenza S, Perlin MH, Cuomo CA, Fairhead C, Branca A, Giraud T. 2015. Chaos of rearrangements in the mating-type chromosomes of the anther-smut fungus *Microbotryum lychnidis-dioicae*. *Genetics* **200**(4): 1275–1284.

Bao W, Kojima KK, Kohany O. 2015. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob DNA* **6**(1): 4–9.

Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* **27**(2): 573–580.

Branco S, Badouin H, Rodríguez de la Vega RC, Gouzy J, Carpentier F, Aguileta G, Siguenza S, Brandenburg J-T, Coelho MA, Hood ME, et al. 2017. Evolutionary strata on young mating-type chromosomes despite the lack of sexual antagonism. *Proc Natl Acad Sci* **114**(27): 7067–7072.

Branco S, Carpentier F, Rodríguez de la Vega RC, Badouin H, Snirc A, Le Prieur S, Coelho MA, de Vienne DM, Hartmann FE, Begerow D, et al. 2018. Multiple convergent supergene evolution events in mating-type chromosomes. *Nat Commun* **9**(1): 2000.

Devier B, Aguileta G, Hood ME, Giraud T. 2009. Ancient trans-specific polymorphism at pheromone receptor genes in basidiomycetes. *Genetics* **181**(1): 209–223.

Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* **7**(1): 1–8.

Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* **29**(8): 1969–1973.

Edgar RC. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput.

Nucleic Acids Res **32**(5): 1792–1797.

Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**(19): 2460–2461.

Ekseth OK, Kuiper M, Mironov V. 2014. OrthAgogue: an agile tool for the rapid prediction of orthology relations. *Bioinformatics* **30**(5): 734–736.

Ellinghaus D, Kurtz S, Willhoeft U. 2008. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinformatics* **9**(18): 1–14.

Fortuna TM, Snirc A, Badouin H, Gouzy J, Siguenza S, Esquerre D, Le Prieur S, Shykoff JA, Giraud T. 2016. Polymorphic microsatellite markers for the tetrapolar anther-smut fungus *Microbotryum saponariae* based on genome sequencing. *PLoS One* **11**(11): e0165656.

Gladieux P, Vercken E, Fontaine MC, Hood ME, Jonot O, Couloux A, Giraud T. 2011. Maintenance of fungal pathogen species that are specialized to different hosts: allopatric divergence and introgression through secondary contact. *Mol Biol Evol* **28**(1): 459–471.

Hartmann FE, Rodríguez de la Vega RC, Brandenburg J-T, Carpentier F, Giraud T, Van De Peer Y. 2018. Gene presence–absence polymorphism in castrating anther-smut fungi: recent gene gains and phylogeographic structure. *Genome Biol Evol* **10**(5): 1298–1314.

Hood ME, Scott M, Hwang M. 2015. Breaking linkage between mating compatibility factors: tetrapolarity in *Microbotryum*. *Evolution (N Y)* **69**(10): 2561–2572.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**(4): 772–780.

Koren S, Schatz MC, Walenz BP, Martin J, Howard JT, Ganapathy G, Wang Z, Rasko DA, McCombie RW, Jarvis ED, et al. 2012. Hybrid error correction and de novo assembly of single-molecule sequencing reads. *Nat Biotechnol* **30**(7): 693–700.

Krzywinski M et al. 2009. Circos: an information aesthetic for comparative genomics. *Genome Res* **19**(604): 1639–1645.

Melters DP, Bradnam KR, Young HA, Telis N, May MR, Ruby JG, Sebra R, Peluso P, Eid J, Rank

D, et al. 2013. Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. *Genome Biol* **14**(1): R10.

Ruby JG, Bellare P, DeRisi JL. 2013. PRICE: Software for the targeted assembly of components of (meta) genomic sequence data. *G3 Genes | Genomes | Genet* **3**(5): 865–880.

Salichos L, Stamatakis A, Rokas A. 2014. Novel information theory-based measures for quantifying incongruence among phylogenetic trees. *Mol Biol Evol* **31**(5): 1261–1271.

Smit AF, Hubley R, Green P. 2015. RepeatMasker Open-4.0. *Repeat Masker Website*.

Smit AF, Hubley RR. 2015. RepeatModeler Open-1.0. *Repeat Masker Website*.

Stamatakis A. 2006. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**(21): 2688–2690.

Van Dongen SM. 2000. Graph clustering by flow simulation. PhD thesis. (https://dspace.library.uu.nl/handle/1874/848).

Wickham H. 2016. *ggplot2: elegant graphics for data analysis*. Springer.

Yang Z. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**(8): 1586–1591.