Chaos of Rearrangements in the Mating-Type Chromosomes of the Anther-Smut Fungus *Microbotryum lychnidis-dioicae*

Hélène Badouin,*† Michael E. Hood,† Michael H. Perlin,‡† Christina A. Cuomo,‡‡ Cécile Fairhead,§§ Antoine Branca,*† and Tatiana Giraud*†,2

*Ecologie, Systématique et Evolution, Univ Paris-Sud, 91405 Orsay, France, †Centre National de la Recherche Scientifique, F-91405 Orsay, France, ‡Department of Biology, Amherst College, Amherst, Massachusetts 02142, §§INRA and **Centre National de la Recherche Scientifique, Laboratoire des Interactions Plantes-Microorganismes, 31326 Castanet-Tolosan, France, ††Department of Biology, Program on Disease Evolution, University of Louisville, Louisville, Kentucky 40292. ‡‡Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, and §§Institut de Génétique et Microbiologie, Université Paris-Sud, F-91405 Orsay cedex, France

ORCID IDs: 0000-0002-2456-5968 (H.B); 0000-0002-5778-960X (C.A.C).

**ABSTRACT** Sex chromosomes in plants and animals and fungal mating-type chromosomes often show exceptional genome features, with extensive suppression of homologous recombination and cytological differentiation between members of the diploid chromosome pair. Despite strong interest in the genetics of these chromosomes, their large regions of suppressed recombination often are enriched in transposable elements and therefore can be challenging to assemble. Here we show that the latest improvements of the PacBio sequencing yield assembly of the whole genome of the anther-smut fungus, *Microbotryum lychnidis-dioicae* (the pathogenic fungus causing anther-smut disease of *Silene latifolia*), into finished chromosomes or chromosome arms, even for the repeat-rich mating-type chromosomes and centromeres. Suppressed recombination of the mating-type chromosomes is revealed to span nearly 90% of their lengths, with extreme levels of rearrangements, transposable element accumulation, and differentiation between the two mating types. We observed no correlation between allelic divergence and physical position in the nonrecombining regions of the mating-type chromosomes. This may result from gene conversion or from rearrangements of ancient evolutionary strata, i.e., successive steps of suppressed recombination. Centromeres were found to be composed mainly of copia-like transposable elements and to possess specific minisatellite repeats identical between the different chromosomes. We also identified subtelomeric motifs. In addition, extensive signs of degeneration were detected in the nonrecombining regions in the form of transposable element accumulation and of hundreds of gene losses on each mating-type chromosome. Furthermore, our study highlights the potential of the latest breakthrough PacBio chemistry to resolve complex genome architectures.

**KEYWORDS** *Microbotryum violaceum*; finished genome assembly; intratetrad mating; MAT; basidiomycete; selfing; bipolarity

Sex chromosomes often show exceptional genome features with extensive suppression of homologous recombination and cytological differentiation between members of the diploid chromosome pair (Bergero and Charlesworth 2009). Determination of male vs. female development by such nonrecombining sex chromosomes has evolved independently in diverse lineages, including mammals, birds, fishes, insects, vascular plants, bryophytes, and macroalgae (Ohno 1967; Charlesworth 1991; Itoh et al. 2006; Yamato et al. 2007; Marais et al. 2008; Bergero and Charlesworth 2009; Cock et al. 2010; Kaiser and Bachtrog 2010; Page et al. 2010; Bachtrog et al. 2011; McDaniel et al. 2013), where incorporation of multiple genes responsible for sexually antagonistic traits is a common explanation for expansion of recombination suppression (Rice 1984; Lahn and Page 1999). Despite the fundamental roles that sex
chromosomes play, their genic content is particularly prone to degenerative influences of recombination suppression and enforced heterozygosity (Bergero and Charlesworth 2009).

In fungi, mating-type chromosomes contain the genes for molecular mechanisms of mating compatibility (e.g., via pheromones and receptors), and they can display recombination suppression and size dimorphism analogous to sex chromosomes (Bakkeren and Kronstad 1994; Hood 2002; Fraser et al. 2004; Fraser and Heitman 2004; Hood et al. 2004, 2013; Fraser and Heitman 2005; Menkis et al. 2008; Ellison et al. 2011; Grogné et al. 2014), as well as degeneration (Hood et al. 2004; Whittle and Johannesson 2011; Whittle et al. 2011; Fontanillas et al. 2015). Grouped together, heteromorphic sex chromosomes and mating-type chromosomes can therefore be considered allosome pairs (Montgomery 1911), which are similarly maintained and influenced by their central role in regulating gamete fusion, contrasted by the homomorphic autosomes. Nonrecombining fungal mating-type chromosomes can also show footprints of degeneration (Hood et al. 2004; Whittle and Johannesson 2011; Whittle et al. 2011; Fontanillas et al. 2015). However, important differences exist between sex and mating-type chromosomes: because mating types are expressed as non-self-recognition at the haploid stage, they are found only in the heterogametic condition in diploids (similar to macroalgae and bryophytes), and thus with symmetrical roles, and mating types are not associated with male/female functions in fungi (Billiard et al. 2011). These common features and differences help make fungal mating-type chromosomes valuable models for the broader phenomenon of allosome evolution, with some shared and some distinguishing expectations compared to sex chromosomes, e.g., on the asymmetry of allosome degeneration (Bull 1978) and on the role of sexually antagonistic selection in driving suppressed recombination (Bergero and Charlesworth 2009).

Examples of nonrecombining fungal allosomes include the mating-type chromosomes of Microbotryum lycnthis-dioicae (Hood 2002) and the fungus-causing anther-smut disease on Silene latifolia (Votintseva and Filatov 2009; Hood et al. 2013; Fontanillas et al. 2015; Perlin et al. 2015; Whittle et al. 2015). These nonrecombining fungal mating-type chromosomes have been extensively studied, and degeneration has been reported in the form of transposable element (TE) accumulation and high rates of nonsynonymous substitutions (Hood et al. 2004; Fontanillas et al. 2015). The buildup of TEs in regions of suppressed recombination has, however, prevented genome assembly for these mating-type chromosomes. Therefore, the exact content in genes and transposable elements, the extent of the nonrecombining regions (NRRs), and the existence of evolutionary strata of recombination suppression have remained elusive (Votintseva and Filatov 2009; Hood et al. 2013; Fontanillas et al. 2015; Perlin et al. 2015; Whittle et al. 2015).

The spread of recombination suppression outward from key compatibility-determining loci to form evolutionary strata of differentiation is described for plant and animal sex chromosomes (Bergero and Charlesworth 2009) and may also occur on fungal mating-type chromosomes (Fraser et al. 2004; Fraser and Heitman 2004, 2005; Menkis et al. 2008; Votintseva and Filatov 2009). Suppressed recombination has been shown to link the two mating-type-determining loci (i.e., encoding mating pheromone/receptor and homeodomain proteins, respectively), including basidiomycete fungal pathogens of plants and humans (Ustilago hordei and Cryptococcus neoformans, respectively) (Bakkeren and Kronstad 1994; Fraser et al. 2004). This linkage of two mating-type-determining loci may be advantageous under selfing mating systems, where it increases the compatibility of gamete combinations (Billiard et al. 2011; Nieuwenhuis et al. 2013). In other fungi, suppressed recombination links mating-type loci to the centromere, causing segregation of mating types in the first meiotic division. This is associated with automictic reproduction (matting within the meiotic tetrad), e.g., in M. lycnthis-dioicae and Neurospora tetrasperma (Zakharov 1987; Hood and Antonovics 2000; Giraud et al. 2008; Menkis et al. 2008). Such intratetrad mating may favor successive linkage to mating type for a suite of genes that experience deleterious, recessive mutations, as shown by theoretical models (Antonovics and Abrams 2004; Johnson et al. 2005) (Supporting Information, Figure S1). Indeed, deleterious mutation occurring at the margin of the nonrecombining regions may be partially sheltered in a heterozygous state due to less frequent recombination, and extension of the region of suppressed recombination would then be selected for permanent sheltering of these deleterious mutations (Ironside 2010). This could lead to the formation of evolutionary strata, reflected in increasing allelic differentiation at physical positions along the chromosome when coming closer to the mating-type genes. Although in plants and in animals evolutionary strata are generally thought to be due to sexually antagonistic selection, the definitive linkage of partially linked deleterious alleles to the sex-determining region has also been proposed as an alternative explanation (Ironside 2010). Elucidating the existence of evolutionary strata in fungal mating-type chromosomes may thus generally provide insights on the shared evolutionary processes acting on allosomes.

In this research, we aimed to produce a finished assembly of the mating-type chromosomes of the anther-smut fungus M. lycnthis-dioicae for elucidating the extent of suppressed recombination, transposable element accumulation, gene content, and rearrangements and whether evolutionary strata are present. For this goal, we took advantage of the recent developments of the PacBio sequencing technology, which allowed full-length physical mapping of the mating-type chromosomes and the resolution of these important genomic properties.

**Materials and Methods**

**DNA extraction and sequencing**

DNA was extracted with the Qiagen Kit 10243 (Courtaboeuf, France) following manufacturer instructions and using a Carver hydraulic press (reference 3968, Wabash, IN) for breaking cell
walls. Haploid a₁ and a₂ genomes of the Lamole reference strain of *M. lycnhdidis-dioicae* were sequenced separately using the Pacific Bioscience (PacBio) method (Institute for Genomic Medicine, University of California, San Diego, La Jolla, CA).

**Assembly and annotation**

The PacBio P5/C3 sequencing of haploid a₁ and a₂ genomes produced 658,501 and 648,664 reads, respectively (mean read length: 7.925/7.569; coverage: 180×/160×). Independent assemblies of the a₁ and a₂ genomes were generated with the wgs-8.2beta version of the PBcR assembler (Koren et al. 2012), with the following parameters: -pBCNS -length 500 -partitions 200 genomeSize = 27000000 -maxCoverage 150. Contigs were aligned with optical maps of the two mating-type chromosomes (Hood et al. 2013) with MapSolver software (OpGen). The alignment of the contig corresponding to the a₂ mating-type chromosome with the corresponding optical map was highly significant. The a₁ optical map was used to create an oriented a₁ pseudomolecule composed of two contigs for each chromosome arm, aligned on the optical map, plus a 50-kb unanchored contig in the centromeric region. The two contigs corresponding to the two a₁ chromosome arms produced full-length alignments that matched only with the a₁ mating-type chromosome optical map; alignment scores of the contigs were 46 and 143, exceeding the default MapSolver threshold range for significance of 3 to 6.

Mauve was used to compare the two assemblies (Darling et al. 2010) (Figure S6). In this genome comparison, a contig belonging to autosomes in one assembly was found to be a subsequence of a contig in the second assembly (and vice versa). Manual curation was performed to select the largest contigs and remove the few short degenerate contigs, probably resulting from the incorrect editing of reads (Figure S6). Comparison with the reference sequence (GenBank no. NC_020353) revealed that the mitochondrial genome was not correctly solved by the assembler, and this genome sequence was therefore removed from the assembly. The PacBio assembly was compared with the Illumina-based assembly available from the Broad Institute and the Illumina reads generated in a previous study (Perlin et al. 2015). The PacBio assembly was edited at positions for which both the genome alignment obtained with Mugsy (Angiuoli and Salzberg 2011) and a standard variant-calling-method based on read mapping (samtools, bcftools, and VarScan) (Li et al. 2009; Koboldt et al. 2012) gave concordant findings for the polymorphism. This conservative approach led to the editing of 1332 SNPs, 456 insertions, and 19,528 deletions. The 11,057 protein-coding gene models were predicted with EuGene (Foissac et al. 2008), trained for *Microbotryum*. Available *Microbotryum* EST data (Aguilera et al. 2010) and similarities to the fungi subset of the uniprot database (Consortium 2011) were integrated into EuGene for gene prediction. InterPro (Mitchell et al. 2014) was used to identify protein domains and families and to infer an automatic functional annotation from protein domain content.

**Analysis of genome structure**

Repetitive DNA content was analyzed with RepeatMasker (Smit 2013), using REPBASE v19.11 (Jurka 1998). We searched for subtelomeric motifs with Meme (Bailey et al. 2006) (Figure S2). Figure 1, Figure 2, and Figure S5 were prepared with Circos (Krzywyinski et al. 2009). We identified syntenic blocks from the alignment of the two chromosomes with Mauve, using standard parameters (Darling et al. 2010). We analyzed gene order after filtering out transposable elements (TEs) to identify larger blocks of synteny. Alleles were assigned between the two mating-type chromosomes by applying orthomcl (Li et al. 2003) to the protein data sets for unique a₁ and a₂ orthologs. The signed permutations were used to calculate minimal reversal distances with Grimm (Tesler 2002). The sequences of TE-like copies and minisatellites specific for the centromeres are provided in File S1.

**Divergence between a₁ and a₂ alleles**

Alleles were identified as indicated above, and MACSE (Ranwez et al. 2011) was used to align DNA-coding sequences, respecting phase. Synonymous divergence and its standard error were estimated with the yN00 program of the PAML package (Yang 2007). The correlation between dₛ and physical distance was assessed with JMP (SAS Institute). As we have here a single individual genome sequence, some of the substitutions between a₁ and a₂ identified here could theoretically include within-species polymorphism rather than fixed differences between the a₁ and a₂ alleles. However, this should not be a problem for identifying the nonrecombining regions from the pseudo-autosomal regions (PARs). Indeed, given its highly selfing mating system, *M. lycnhdidis-dioicae* is largely homozygous in regions unlinked to mating type (Giraud et al. 2008; Vercken et al. 2010), so that high levels of synonymous divergence between two a₁ and a₂ alleles from a given diploid individual reliably indicate suppressed recombination.

**Data access**

The assembly is available at EMBL-ENA (accession no. PRJEB7910).

**Results and Discussion**

**Finished genome assembly and new insights about centromere and telomere structures**

The recent PacBio P5/C3 chemistry produced long reads, averaging 8 kb. The imperfect quality of PacBio reads was fully compensated by >100× sequencing depth, as illustrated by the almost complete assembly of the *M. lycnhdidis-dioicae* genome and including the mating-type chromosomes (Figure 1 and Figure 2), while several previous efforts and chemistries had been unsuccessful (Votintseva and Filatov 2009; Hood et al. 2013; Fontanillas et al. 2015). The PacBio genome assembly indeed yielded 22 contigs, including 18 autosomal contigs, 1 contig for the a₂ mating-type
chromosome, and 3 contigs for the \textit{a}_1 mating-type chromosome that could be assembled in a scaffold (see below) (Figure 2, Table 1). The total size of the assembly was 33.3 Mb, i.e., almost 7 Mb larger than that of the very same \textit{a}_1 haploid \textit{M. lychnidis-dioicae} strain sequenced at the Broad Institute using Illumina technology (Perlin \textit{et al.} 2015), as the long reads of the PacBio chemistry allowed assembly by more complete incorporation of repetitive DNA elements and included both mating-type chromosomes. In particular, putative centromeric regions could be assembled and were identified as the most repeat-rich regions, where we found two copia-like TEs and two minisatellite repeats specifically clustering in these regions at one location per contig (tracks E and F in Figure 2). Both copia-like TEs possessed two open reading frames, coding for an integrase and for a reverse transcriptase, but no long terminal repeats, and many copies were truncated. The minisatellites consisted of a 66-bp repeat motif of GGCCCA and a 110-bp repeat motif of CGACGG; the 66-bp repeat was specific for centromeric regions and identical in all iterations. The structures of the finished centromeres are shown in Figure 3. Eleven of the contigs had these specific repeats at one end, suggesting that some autosomes were split at their centromeres in the assembly (Figure 2). On finished chromosomes, the centromeric regions spanned 65–155 kb. Centromeric regions have been elucidated in a single other basidiomycete fungus so far, \textit{Cryptococcus neoformans} var. grubii, in which they were instead enriched in Tcn transposons (Janbon \textit{et al.} 2014).

We found no typical telomeric repeats (TTAGGG) while they were identified at the edge of five scaffolds in the Broad Institute \textit{M. lychnidis-dioicae} genome (Perlin \textit{et al.} 2015). Here, we identified instead a specific 50-bp motif repeated at some, but not all, chromosome ends (Figure 2, Figure S2), which thus likely represent a subtelomeric motif. This motif was in fact found near the TTAGGG telomeric repeats (80–1500 bp apart) in two of the five contigs from the Broad Institute containing TTAGGG repeats (the three other contigs being only 500–850 bp long). This suggests that PacBio technology does not sequence the very edges of telomeres well and that other motifs may be useful for detecting ends of chromosomes.

Previous pulse-field gel electrophoresis estimated 11 autosomes for the haploid stage (Perlin \textit{et al.} 2015), corresponding precisely to the 5 complete autosomes plus the 6 autosomes as two contigs split in their centromeric regions (the S18 contig corresponded only to a telomeric region). The autosomes were thus assembled into full or two-arm contigs (Figure 2). The \textit{a}_2 mating-type chromosome was
completely assembled from PacBio sequences, including the submetacentric centromeric region. Available optical maps (Hood et al. 2013), i.e., ordered, genome-wide restriction maps obtained from single DNA molecules, allowed orienting the two contigs covering the arms of the a1 mating-type chromosome and attributing a 50-kb contig composed of repeated DNA to its centromeric region. Collinearity between the optical maps and the contigs confirmed the high quality of the assembly (Figure S3). The combination of PacBio chemistry and optical maps thus revealed the right approach to finally resolve the mating-type chromosome structure and content.

**Identification of pseudo-autosomal regions and extensive nonrecombining regions on the mating-type chromosomes**

Assemblies of the dimorphic *M. lychnidis-dioicae* mating-type chromosomes (a1, 3.5 Mbp; a2, 4.0 Mbp) encompassed 614 and 683 predicted genes, respectively, filtering out TEs. Among these genes, only a few had putative functions known to be involved in mating (Petit et al. 2012; Fontanillas et al. 2015), i.e., the PR genes encoding the pheromone receptors, the pHp genes encoding the precursors of pheromones, the HD1 and HD2 homeodomain genes, STE20 encoding a protein kinase regulating mating (Smith et al. 2004), and STE12 encoding a transcription factor regulating mating and invasive hyphal growth in fungal pathogens (Hoi and Dumas 2010) (Figure 1). Actually, two homologs of STE12 were found, both present only on the a2 mating-type chromosome, while absent from the a1 mating-type chromosome; this seems consistent with previous observations that initiation of the conjugation tubes is earlier and to a greater extent from cells of the a2 mating type in *M. lychnidis-dioicae* (Day 1976; Xu et al. 2015). The pHp pheromone genes were recently characterized as encoding proteins with the typical C terminus CAAX motif and producing functional pheromone peptides triggering the mating response in several *Microbotryum* species (Xu et al. 2015).

We identified 305 predicted genes shared between the a1 and a2 mating-type chromosomes. Synonymous divergence between a1 and a2 alleles of these shared genes, when plotted along their physical positions on both mating-type chromosomes, clearly identified the nonrecombining regions; suppressed recombination can lead to divergence between alleles while recombination homogenizes allele sequences. Synonymous divergence values between alleles confirmed the existence of recombining regions at both ends of the mating-type chromosomes, called PARs (Figure 4) (Votintseva and Filatov 2009; Hood et al. 2013). The PARs displayed almost no substitutions between a1 and a2 alleles (Figure 4) and included loci that have previously been shown to be unlinked to the mating type (Votintseva and Filatov 2009) (Figure 1). PARs on the short arm (pPAR) and long arm (qPAR) of the mating-type chromosomes spanned 0.20 and 0.19 Mb,
respectively, and contained 102 genes. PAR boundaries corresponded precisely to the two distal regions of collinearity between a1 and a2 mating-type chromosomes on the restriction digestion optical maps (Hood et al. 2013).

The NRRs were characterized by high divergence between the a1 and a2 alleles of the single-copy genes (Figure 4 and Figure S4). Fourteen widely distributed loci known to cosegregate with mating-type (Votintseva and Filatov 2009; Abbate and Hood 2010; Petit et al. 2012) were found within these divergent regions, validating the NRRs (Figure 1). The a1 and a2 NRRs spanned the majority of the mating-type chromosomes (>88%), covering 3.08 Mb with 506 genes and 3.67 Mb with 580 genes for the a1 and a2 mating-type chromosomes, respectively (Figure 1), confirming conclusions based on optical maps and marker segregation (Hood et al. 2004, 2013; Fontanillas et al. 2015) and contrasting with reports of smaller regions of suppressed recombination (Votintseva and Filatov 2009; Whittle et al. 2015).

### Extensive rearrangements between mating-type chromosomes and degeneration

Comparison of aligned sequences between the a1 and a2 NRRs revealed extensive rearrangement (Figure 1 and Figure S5), probably reflecting and contributing to the current degree of suppressed recombination. We identified 40 syntenic blocks of shared genes in the NRRs, encompassing 2–18 genes. At least 210 inversion events were inferred to account for these rearrangements (Figure 1 and Figure S5). In contrast, all autosomes were collinear (Figure S6). Suppressed recombination on fungal mating-type chromosomes has been shown to be associated with rearrangements in *N. tetrasperma* (Ellison et al. 2011) and *C. neoformans* (Fraser et al. 2004; Fraser and Heitman 2004, 2005) or to occur despite conserved collinearity in *Podospora anserina* (Grognet et al. 2014). The lack of collinearity across several megabases and patterns of major inversions or translocations in *M. lychnidis-dioicae* is, however, to an exceptional degree among fungal mating-type chromosomes.

The two mating-type loci typical of basidiomycete fungi (i.e., the mating pheromone receptor determining premating recognition and the homeodomain proteins determining postmating compatibility) were 0.60 Mb apart on the a1 and a2 mating-type chromosomes, while inverted in orientation (Figure 1 and Figure S5). This distance is much greater than the one estimated recently based on Illumina sequencing (56 kb) (Whittle et al. 2015), again showing the great advantage of PacBio sequencing for chromosome assembly. These mating-type loci were not near the edges of the NRRs, unlike other basidiomycete fungi with linkage suppression in their mating-type regions, i.e., *U. hordei* (Bakkeren and Kronstad 1994) and *C. neoformans* (Fraser et al. 2004). The mating-type loci were instead on the same side of the centromere, on the long arm of the chromosomes. The pheromone precursor loci were linked to the receptor loci in the NRR, but were up to 1.4 Mb away, on opposite sides of the centromere on the a2 mating-type chromosome and the same side in the a1 mating-type chromosome. This distance between pheromone receptor and pheromone is also a novel observation among fungi.

Despite rearrangements in NRRs, several genes close to the pheromone receptor genes in the closely related fungus *Sporidiobolus salmonicolor* (Microbotryomycetes) (Coelho et al. 2010) were also found on the mating-type chromosomes of *M. lychnidis-dioicae* (KAP95, RNAPol, RIB, and ABC1, Figure 1). In contrast, the homologs of the 40 predicted genes in the scaffold containing the HD locus in *S. salmonicolor* with blast hits in *M. lychnidis-dioicae* were all found on autosomes but one, and 34 of them on a single autosome arm (S09). Interestingly, we also found on the S09 contig a sequence with similarity to one of the two homeodomain genes (HD1). The possibility that the HD locus may have duplicated to a location in linkage with the PhP/PR locus warrants further study as a potential transition to mating-type bipolarity within the Microbotryomycetes.

The high degree of rearrangements between the two mating-type chromosomes in *M. lychnidis-dioicae* prevents us from obtaining definitive evidence for the evolutionary cause of the expansive nature of recombination suppression; nevertheless, the linkage of mating-type loci and the centromere within NRRs indicates that first-division segregation of both mating-type loci during meiosis may have been a primary evolutionary factor in recombination suppression, as suggested for other automictic fungi (Zakharov 2005).

As expected for nonrecombining regions, the NRRs had a higher frequency of repetitive elements than the PARs and autosomes (Figure 2). Repetitive elements accounted for 16.0 and 18.7% of the a1 and a2 NRRs, respectively, explaining why they have remained unassembled until now. Most of the repetitive elements were long terminal repeat retro-elements of the copia and gypsy families, as found previously (Hood et al. 2005; Fontanillas et al. 2015). In addition, gene density was lower in the NRRs (25.8% were expressed) than in the autosomes (62.7%) or PARs (51.2%).

The almost finished assembly of the genome allowed detection of 245 and 289 genes in a hemizygous state on the mating-type chromosomes a1 and a2, respectively, suggesting that functional copies have been lost from a single one of the mating-type chromosomes or have degenerated to

<table>
<thead>
<tr>
<th>Table 1 Assembly statistics of the <em>M. lychnidis-dioicae</em> genome</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of contigs</td>
</tr>
<tr>
<td>Minimum contig size</td>
</tr>
<tr>
<td>Maximum contig size</td>
</tr>
<tr>
<td>N50 (bp)</td>
</tr>
<tr>
<td>N50 (no. of contigs)</td>
</tr>
<tr>
<td>N90 (bp)</td>
</tr>
<tr>
<td>N90 (no. of contigs)</td>
</tr>
<tr>
<td>Mean contig size</td>
</tr>
<tr>
<td>Median contig size</td>
</tr>
<tr>
<td>Total length of the assembly (bp)</td>
</tr>
</tbody>
</table>

The PacBio genome assembly yielded 22 contigs—18 for autosomes, 1 for the a2 mating-type chromosome, and 3 for the a1 mating-type chromosome—that could be assembled in a single scaffold.
a great extent. None of these hemizygous genes had putative functions known to be involved in mating except the STE12 homologs on the a2 mating-type chromosome. This degree of gene losses on mating-type chromosomes appears unprecedented among fungi. Together with the accumulation of TEs in NRRs, low gene density and hemizygous loss of genes constitute signs of advanced mating-type chromosome degeneration, as expected for long-term, nonrecombining regions (Bergero and Charlesworth 2009). Degeneration has also been reported for the mating-type chromosome NRRs of *M. lycnthis-dioicae* in terms of higher rates of non-synonymous substitutions than for recombining regions, and these degenerative characteristics were shown to occur in homologous sequences from several other *Microbotryum* species (Fontanillas et al. 2015).

The a2 mating-type chromosome was 590 kb larger than the a1 mating-type chromosome, where TEs accounted for 26.4% (154 kb) and non-TE genes for 6.8% (40 kb) of the difference. In particular, gypsy copies were more abundant on a2 than a1 mating-type chromosome (129 vs. 93 copies, Figure 2). However, intergenic, nonrepetitive DNA accounted for the majority of the difference in size (59%, or 0.347 kb). This finding supports previous conclusions that TE accumulation and degeneration are not asymmetrical between the mating-type chromosomes (Fontanillas et al. 2015), as expected from their symmetrical role in mating (Bull 1978).  

**Ancient recombination suppression without clear evolutionary strata despite heterogeneity in a1–a2 divergence**

Divergence between a1 and a2 alleles of predicted genes in the NRRs was substantial (mean synonymous substitution rates: \(d_S = 0.064\), Figure 4), indicating relatively ancient recombination suppression. However, this divergence was lower than for the pheromone receptor gene, which was previously shown to constitute the oldest trans-specific polymorphism in any organism so far (Devier et al. 2009).

In the evaluation of evolutionary strata within NRRs, we quantified synonymous divergence \(d_S\) along the mating-type chromosomes, and there was no correlation with physical distance to the PAR boundary or to the mating-type loci on either the a1 or a2 chromosomes (Figure 3). In particular, the correlations between \(d_S\) and gene order from PAR were non-significant (on a1, \(r = 0.06, P = 0.36\); on a2, \(r = 0.02, P = 0.77\)), as were correlations between \(d_S\) and gene order from the nearest mating-type gene (on a1, \(r = 0.04, P = 0.58\); on a2, \(r = 0.02, P = 0.77\)). There was a particularly sharp boundary between the pPARs, with zero synonymous divergence between a1 and a2 alleles, and the adjacent NRRs, with extensive divergence (mean \(d_S = 0.064\)). Thus, there was no evidence of evolutionary strata.

However, given that there is no homogametic diploid in heterothallic fungi (i.e., no a1a1 or a2a2 genotypes), which tends to retain chromosome structure in homogametic sex chromosomes of plants and animals (X or Z chromosomes), the mating-type chromosomes may more rapidly undergo rearrangements that obscure any history of evolutionary strata, had they existed. Actually, in XY or ZW systems, evolutionary strata can be inferred only by using as reference the less rearranged homogametic sex chromosome (Lahn and Page 1999).

It is possible that we may have detected one recent ministratum as a very small region (23 kb) at the edge of the NRR close to the qPAR that was collinear between a1 and a2 mating-type chromosomes but with a high rate of synonymous substitutions between mating types. This region was separated from the qPAR by a rearranged region rich with TEs (Figure S7). This pattern is consistent with a recent incorporation of a PAR fragment into the NRR, which would support the scenario of an extension of the NRR, and thus evolutionary strata.

Allele divergence between mating types in NRRs was characterized by heterogeneous values (Figure 4 and Figure S4), which may result from a continuing process of translocations between the chromosomes or the remnants of ancient evolutionary strata that have been scrambled by rearrangements. Alternatively, gene conversion may have played a role by recurrently resetting divergence to zero in some stretches of DNA.
Conclusions

The nearly finished assembly of the *M. lychnidis-dioicae* genome allowed the resolution of controversies over the extent of the nonrecombining regions of the mating-type chromosomes, the degree of divergence between the a1 and a2 mating-type chromosomes, the patterns of evolutionary strata, and the mechanisms of evolution to bipolarity (Votintseva and Filatov 2009; Hood et al. 2013; Fontanillas et al. 2015; Whittle et al. 2015). The great extent to which the mating-type chromosomes are consumed by the NRR and the chaos of rearrangements within this region indicate that these dimorphic chromosomes are the result of ancient and impactful evolutionary processes, approaching an ultimate state of genomic entropy, and thus showing striking convergence with some animals (Repping et al. 2006; Bachtrog 2013). Our study allows the broadening of evolutionary biology for nonrecombining chromosomes by contributing to evidence of convergence in genomic patterns between fungal mating-type chromosomes and sex chromosomes, as previously highlighted (Hood 2002; Fraser et al. 2004; Menkis et al. 2008). These distant systems share extensive suppressed recombination, a high level of rearrangements, numerous losses of genes, and accumulation of a high repetitive elements content. The high level of rearrangements and the possibility of gene conversion still make it difficult to reconstitute the full history of the recombination suppression of the mating-type chromosomes. Comparative genomics between closely related *Microbotryum* species and other members of the Microbotryomycetes with different degrees and/or ages of suppression of recombination between their mating-type chromosomes should help in resolving these questions in future studies.

Next generation sequencing is increasingly making it possible to obtain sequence data at low cost for nonmodel organisms. However, the short length of reads remains a major limitation, potentially preventing large-scale assembly, particularly for genomic regions rich in transposable elements (Treangen and Salzberg 2012). The lack of a complete physical assembly precludes investigation of some of the most fundamental evolutionary inferences. PacBio chemistry now makes it possible to obtain a complete and affordable assembly, even for notoriously challenging regions, such as highly repetitive centromeres and nonrecombining chromosomes (Treangen and Salzberg 2012). This breakthrough technology will now facilitate studies of the dynamics and evolutionary role of genome structures with unprecedented resolution and power.

Acknowledgments

We thank Stéphanie Le Prieur, Alodie Snirc, and Gilles Deparis for help with DNA extraction. PacBio sequencing was conducted at the Institute for Genomic Medicine, University of California, San Diego, La Jolla, CA. This work was supported by a European Research Council starting grant, GenomeFun 309403 and ANR-12-ADAP-0009 (Gandalf project) (to T.G.); an Institut Diversité et Evolution du Vivant grant (to T.G. and C.F.); National Science Foundation (NSF) grant DEB-1115765 to (M.E.H.); and NSF award #0947963 (to M.H.P. and C.A.C.). We declare no conflict of interest.
**Literature Cited**


Chaos of Rearrangements in the Mating-Type Chromosomes of the Anther-Smut Fungus Microbotryum lychnidis-dioicae

Hélène Badouin, Michael E. Hood, Jérôme Gouzy, Gabriela Aguileta, Sophie Siguenza, Michael H. Perlin, Christina A. Cuomo, Cécile Fairhead, Antoine Branca, and Tatiana Giraud
Figure S1: Meiosis and intratetrad-mating (automixis) in Microbotryum lycnids-dioicae. A scenario with 2N number of 4 is shown, with a mating type (blue) and an autosomal (black) pair of homologous chromosomes. Mating type (MAT) is linked to the centromere. Mating type and linked loci (locus B) segregate at Meiosis I. Heterozygosity at loci linked to mating type, or that segregates a Meiosis I linked to other centromeres (locus C), is therefore restored by mating between cells from the same meiotic tetrad. This can shelters deleterious load loci linked to the mating type while purging heterozygosity at loci that segregate at Meiosis II due to crossing over (as shown for locus D).
Figure S2: Identification of sub-telomeric motifs in *Microbotryum lycnhdis-dioicae*. a) Consensus motif. b) Location in the 3000 bp of putative chromosomes edges. The genomic coordinates in bp are indicated in brackets.
Figure S3: Correspondence between the optical map and the PacBio assembly of the *Microbotryum lycnids-dioicae* mating-type chromosomes. The vertical lines within the chromosomes indicate restriction sites. Alignments compare restriction fragment sizes on single DNA molecules and from *in silico* sequences. Regions highlighted in blue indicate aligned optical map and *in silico* restriction site distributions, while regions in grey could not be aligned (i.e., a1 centromeres and a2 telomeres).
**Figure S4:** Divergence between the *Microbotryum lycnidis-dioicae* mating-type chromosomes. Synonymous divergence $dS \pm SE$ is plotted against the genomic coordinates of the $a_1$ (A) or $a_2$ (B) alleles, for predicted genes with a gap-free codon alignment of more than 1000 bp. The boundaries between the pseudo-autosomal region (PAR) and the non-recombining regions (NRR) are indicated, as well as the locations of the mating-type loci (*PR*: pheromone receptor gene; *HD*: homeodomain genes). The mean value of $dS$ in the NRR is shown as blue dotted lines.
Figure S5: Rearrangements between the main syntenic blocks of genes common to $a_1$ and $a_2$ Microbotryum lycnids-dioicae mating-type chromosomes. Tracks A to D show the location of different genomic elements, as follows: A – Structure of the chromosomes, with the pseudo-autosomal regions (PARs) in green, the non-recombining regions (NRRs) in blue, and the centromeres in yellow. B – Location of loci shown to be linked (blue circles) or unlinked (white circles) to mating-type by previous segregation analyses in M. lycnids-dioicae (Abbate, 2010 #87; Petit, 2012 #86; Votintseva, 2009 #84). C – Location of the genes related to the mating-type function: pheromone receptor and homeodomain genes (in red), the other genes likely involved in mating (STE12, STE20, and the precursors of pheromones, PhP) and the genes located around the pheromone receptor gene in the closely related Sporidiobolus salmonicolor (Coelho, 2010 #156) (KAP95, RNP, RIB and ABC1). D – Links between syntenic blocks of shared genes larger than 10 kb.
Figure S6: Autosomal contigs from the a₁ and a₂ assemblies, aligned with Mauve (Darling, 2010 #146), illustrating their synteny. The contigs of a₁ (A) and a₂ (B) are separated by vertical red lines and the locally collinear blocks (LCBs) are represented as colored blocs. A couple of a₁ contigs were split into two contigs in the a₂ assembly (i.e., at position 2MB and 6MB), and conversely (at 14MB and 18MB). In these cases where an autosomal contig was larger in the a₁ or in the a₂ assembly, the largest contig was selected to resolve the final reference assembly (Figure 2). Contig names are indicated as in Figure 2, except the S18 that is one of the small contigs at right. The few short degenerate contigs at right were disregarded, likely resulting from the incorrect editing of reads.
Figure S7: Map of the base-pair substitutions in pseudo-autosomal regions of the *Microbotryum lychnidis-dioicae* mating-type chromosomes. A collinear region of 23 kb in the non-recombining region is identified close to the qPAR boundary.
File S1

Available for download as a .txt file at www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.177709/-/DC1