Multiple convergent supergene evolution events in mating-type chromosomes
Sara Branco, Fantin Carpentier, Ricardo Rodriguez de La Vega, Hélène Badouin, Alodie Snirc, Stéphanie Le Prieur, Marco Coelho, Damien De Vienne, Fanny Hartmann, Dominik Begerow, et al.

To cite this version:
Multiple convergent supergene evolution events in mating-type chromosomes

Sara Branco1,2, Fantin Carpentier1, Ricardo C. Rodríguez de la Vega1, Hélène Badouin1,3, Alodie Snirc1, Stéphanie Le Prieur1, Marco A. Coelho4, Damien M. de Vienne3, Fanny E. Hartmann1, Dominik Begerow5, Michael E. Hood6 & Tatiana Giraud1

Convergent adaptation provides unique insights into the predictability of evolution and ultimately into processes of biological diversification. Supergenes (beneficial gene linkage) are striking examples of adaptation, but little is known about their prevalence or evolution. A recent study on anther-smut fungi documented supergene formation by rearrangements linking two key mating-type loci, controlling pre- and post-mating compatibility. Here further high-quality genome assemblies reveal four additional independent cases of chromosomal rearrangements leading to regions of suppressed recombination linking these mating-type loci in closely related species. Such convergent transitions in genomic architecture of mating-type determination indicate strong selection favoring linkage of mating-type loci into cosegregating supergenes. We find independent evolutionary strata (stepwise recombination suppression) in several species, with extensive rearrangements, gene losses, and transposable element accumulation. We thus show remarkable convergence in mating-type chromosome evolution, recurrent supergene formation, and repeated evolution of similar phenotypes through different genomic changes.
Gould’s view that evolution is “utterly unpredictable and quite unrepeatable”1 has long prevailed. It is difficult to test the repeatability of evolution, but such tests are essential for understanding biological diversification in response to selection2,3. Cases of convergent evolution following similar selective pressures provide ideal opportunities for assessing the repeatability of evolutionary processes and unraveling the proximal and ultimate mechanisms generating diversity2,4. Examples of convergent evolution include ecological morphs in Nicaraguan crater lake cichlid fishes5, cave morphs in Mexican cavefishes6, resistance to toxic compounds in animals7, and lactase persistence in humans8. However, few examples have been studied in detail and many unresolved questions remain, including the frequency of convergent evolution, the genetic mechanisms underlying convergent trait evolution, whether convergence is widespread in organisms other than plants and animals, and the phylogenetic scales at which it occurs.

Supergenes (the beneficial linkage of genes controlling different ecological traits by recombination suppression) are striking cases of adaptation, arising by conspicuous changes in genomic architecture. As such, supergenes can be good models for assessing the predictability and proximate/ultimate causes of convergent evolution. Although we still know little about supergene prevalence and evolutionary importance9–11, interesting cases have been reported, including the non-recombining genomic regions controlling multiple wing color patterns in butterflies12 and polymorphic social behavior in ants10,11. Chromosomes involved in sexual compatibility often also have large non-recombining regions, and the early stages in development of these regions can be considered to constitute supergenes13. Recombination suppression linking different traits involved in sexual compatibility has been documented in the sex chromosomes of animals and plants13,14, the mating-type chromosomes of algae15 and fungi16–21, and self-incompatibility loci in plants22. Recombination cessation not only maintains beneficial allelic combinations but also reduces selection efficacy, leading to genomic decay and the accumulation of transposable elements (TEs)23. The frequency and proximal mechanisms of recombination suppression and the tempo of genomic degeneration remain unclear13,14,24. Chromosomes with recent recombination suppression events are ideal for investigating the initial steps of supergene formation and degeneration25–27.

Fungi provide excellent systems for studying the causes, consequences, and frequency of recombination cessation, as they have diverse mating-type-determining systems involving multiple genes28 and mating-type chromosomes with recent recombination suppression events16–21. Most basidiomycetes (mushrooms, rusts, and smut fungi) have two independently segregating loci controlling mating type at the haploid stage: (1) the PR locus, containing a pheromone receptor gene and one to several mating pheromone genes involved in pre-fertilization compatibility, and (2) the HD locus, encoding two homeodomain transcription factors responsible for post-fertilization compatibility28. Linkage between PR and HD loci underlies a major transition determining reproductive compatibility in these fungi29. Such linkage was long thought to be rare but is beneficial in selfing mating systems30,31 (Supplementary Fig. 1). Linkage between PR and HD loci results in large regions controlling multiple mating-type functions (pre- and post-fertilization compatibility), which can be described as supergenes, as they represent a beneficial allelic combination where linkage increases fitness. Here we describe a remarkable case of multiple convergent events of beneficial linkage between the PR and HD mating-type loci, corresponding to the repeated formation of supergenes in multiple young mating-type chromosomes across closely related fungi. We studied species of anther-smut fungi (Microbotryum violaceum complex; Fig. 1), a group of selfing pathogens. The mating-type chromosomes in this group were first described in Microbotryum lychnidis-dioicae32, in which the mating-type loci are linked by a large region without recombination resulting from...
the fusion of the entire ancestral PR chromosome and one arm of the ancestral HD chromosome (Fig. 2)\textsuperscript{16}. The two mating-type loci were precisely the ancestral limits of the initial recombination suppression event, with their linkage resulting in a supergene. Initial mating-type loci linkage was followed by further stepwise expansions of suppressed recombination beyond mating-type genes\textsuperscript{16}. These successive steps expanded the linked region and led to the formation of evolutionary strata with decreasing allelic divergence between mating-type chromosomes at increasing distance from the mating-type-determining genes\textsuperscript{16}, as observed

---

**Key:**
- Centromere
- Mating-type loci (PR or HD)
- Chromosome fission point
- Suppressed recombination and rearrangements
in animal and plant sex chromosomes and likely other fungal mating-type chromosomes. In contrast to the initial mating-type loci linkage event, these additional evolutionary strata did not control any traits for which linkage to mating-type would be beneficial. Instead, the recombination suppression events occurring after mating-type locus linkage probably evolved to shelter deleterious alleles or through neutral rearrangements.

The majority of Microbotryum anther-smut fungi species display linked mating-type loci, but M. lagerheimii and M. intermedium have mating-type loci located on separate chromosomes (Fig. 1). Given the phylogeny of anther-smut fungi (Fig. 1), a previous study based on parsimony inferred ancestral linkage between the PR and HD mating-type loci in the Microbotryum clade, with a reversal to unlinked mating-type loci in M. lagerheimii. However, the distantly related species M. lagerheimii and M. intermedium have highly collinear mating-type chromosomes, whereas gene order is rearranged in non-recombining regions across species with linked PR and HD mating-type loci. The collinearity between M. lagerheimii and M. intermedium mating-type chromosomes suggests that the ancestral state and gene order have been retained in these species. This raises the alternative hypothesis of a remarkable number of independent transitions linking the mating-type loci across anther-smut fungi (Fig. 1).

Using high-quality assemblies of closely related species and the ancestral gene order retained in M. lagerheimii, we uncovered four independent events of mating-type locus linkage in addition to the previously identified supergene. The various Microbotryum species achieved mating-type locus linkage through different chromosomal rearrangements and have non-recombining regions of different sizes, ages, and gene contents. Our results show that supergenes can evolve frequently and that natural selection can repeatedly lead to similar phenotypes through multiple evolutionary trajectories and different genomic changes, consistent with repeatable evolution. We also document repeated and independent formation of evolutionary strata, with stepwise expansions of non-recombining regions beyond mating-type genes and provide evidence for increasing genomic decay in regions with a longer history of recombination suppression.

Results

Five independent routes for linking mating-type loci. We inferred the evolutionary histories of mating-type chromosomes in multiple anther-smut fungi by comparing high-quality genome assemblies of eight Microbotryum species (Fig. 1). We obtained haploid genome assemblies for both mating types (a1 and a2) of four Microbotryum species with full linkage of PR and HD mating-type loci, as previously shown by progeny segregation. We also studied available haploid genome sequences of four additional Microbotryum species (Fig. 1; Supplementary Table 1). We used M. lagerheimii genome as a proxy for ancestral gene order due to its unlinked PR and HD loci and very few rearrangements relative to the distantly related M. intermedium species. Whole-genome BLAST comparisons revealed five different chromosomal rearrangements and fusions underlying the linkage between the HD and PR loci, one in each of the four newly assembled genomes and the fifth at the base of the previously analyzed clade containing M. lychnids-dioicae and M. silenes-dioicae (Figs. 1 and 2).

The different rearrangements led to variation across species in mating-type chromosome size and composition, as well as in non-recombining region length and captured gene content (Supplementary Table 1). Recombination ceased at different times, as shown by the different levels of synonymous divergence (dS) between alleles associated with the a1 and a2 mating types for genes ancestrally located between the HD and PR loci (Figs. 3 and 4). For genes linked to mating-type loci, alleles associated with alternative mating types accumulate differences over time since the linkage event, whereas genes unlinked to mating-type loci are highly homozygous in these selfing fungi (with virtually no divergence between alleles present in the a1 or a2 haploid genomes within diploid individuals; Supplementary Figs. 2a–c). The absence of trans-specific polymorphism in genes ancestrally located between the PR and HD loci following chromosome fusion provided further evidence for the existence of five independent fusion events. Specifically, alleles of genes between mating-type loci clustered by species and not by mating type, demonstrating that their linkage to mating-type loci occurred after speciation events (Fig. 4; Supplementary Fig. 3). Only M. lychnids-dioicae and M. silenes-dioicae displayed trans-specific polymorphism in the genomic regions ancestrally located between the PR and HD loci. Alleles associated with the a2 mating type of both species consistently clustered together, as did alleles associated with the a2 mating type, indicating that PR–HD linkage predated the speciation event in this clade (Fig. 4; Supplementary Fig. 3). Unlike mating-type chromosomes, autosomes were highly collinear between mating types and with no evidence of widespread interchromosomal rearrangements across species (Supplementary Fig. 4) or trans-specific polymorphism (only 1 of the 4229 single-copy shared autosomal genes displayed trans-specific polymorphism, and even then, only between M. lychnids-dioicae and M. silenes-dioicae). The interchromosomal rearrangements and recombination suppression were thus restricted to the mating-type chromosomes and repeatedly led to regions of suppressed recombination bordered by the HD and PR loci. This indicates that the mating-type chromosome rearrangements linking mating-type genes were selected for, forming adaptive supergenes.

In M. lychnids-dioicae, mating-type locus linkage was achieved by the fusion of the putative centromere end of the HD chromosome short arm to the distal end of the PR chromosome short arm (Fig. 2a). This evolutionary transition occurred before the divergence of M. lychnids-dioicae and M. silenes-dioicae, as shown by their similar chromosome structures and trans-specific polymorphism, as previously reported. Genealogies of genes ancestrally located between the HD and PR loci and calibrated using the date of speciation between M. lychnids-dioicae and M. silenes-dioicae indicated that the mating-type loci in this clade became linked 1.2 million years (MY) ago (Fig. 4). The timing of

---

Fig. 2 Routes of mating-type chromosome evolution in Microbotryum. Model for mating-type chromosomal rearrangement events, as inferred from comparisons with the two mating-type chromosomes of M. lagerheimii (used as a proxy for the ancestral mating-type chromosomes in the genus M. lagerheimii). Matting-type chromosome content across Microbotryum species is illustrated by colors referring to different parts of the two M. lagerheimii mating-type chromosomes (Supplementary Figs. 5–8). The inferred ancestral locations of putative centromeres and mating-type loci are indicated in yellow and black, respectively, and the regions of suppressed recombination are dashed. Chromosome sizes are indicated by their relative scales; the last stage in the evolution of recombination suppression often involves increases in chromosome size due to the accumulation of repetitive elements. Mating-type chromosome evolution in a M. lychnids-dioicae, b M. silenes-acaulis, c M. violaceum caroliniana, d M. scabiosae, and e M. v. paradoxa, in which the top edge of the mating-type chromosome corresponds to a rearrangement from the middle of the chromosome, supporting complete recombination suppression up to the end of the chromosome.
PR–HD linkage was inferred from the divergence between alleles associated with the a1 and a2 mating types at 17 genes (red vertical arrows in Fig. 3). In these regions linked to mating-type loci, the alleles remained associated with the a1 or a2 mating type and diverged progressively with time since recombination suppression (Fig. 3). In *M. lychnidis-dioicae*, autosomes were highly syntenic and without differentiation between alleles in the sequenced a1 and a2 genomes derived from a single diploid individual.

In *M. silenes-acaulis*, PR–HD linkage resulted from a different and more recent chromosomal rearrangement. As in *M. lychnidis-dioicae*, the short arm of the ancestral HD chromosome

---

**Fig. 3** Divergence between a1- and a2-associated alleles. Per-gene synonymous divergence and standard error ($d_S$ ± SE) between alleles associated with the a1 and a2 mating types within *Microbotryum* diploid individuals, following the ancestral gene order for the mating-type chromosome. Synonymous divergence is plotted against the genomic coordinates of the a1 mating-type chromosomes of *M. lagerheimii* for all single-copy genes common to both mating-type chromosomes. The limits of the PR and HD *M. lagerheimii* mating-type chromosomes are indicated and oriented according to the fusion in each species (i.e., not in the same orientation in all species). Divergence between the a1 and a2 pheromone receptor (PR) genes was too extensive and $d_S$ could not be calculated (depicted as “Un” for unalignable). The yellow boxes indicate the positions of *M. lagerheimii* putative centromeres. The red vertical arrows at the bottom indicate the 17 genes used for inferring HD–PR linkage dates in all species except for *M. silenes-acaulis*, for which we used a restricted set of 13 genes ancestrally located between the HD locus and the putative centromere (blue vertical arrows). Ancient evolutionary strata that evolved at the base of the *Microbotryum* clade are indicated in purple (around PR) and blue (around HD), as in the previous study in which they were discovered. The genes involved in the more recent evolutionary strata previously identified in *M. lychnidis-dioicae* are indicated with red and green bars at the bottom. a *M. silenes-acaulis*, b *M. v. caroliniana*, with a recent stratum indicated in light blue and enlarged in an inset; the current location of these genes is indicated in Supplementary Fig. 6a; c *M. scabiosae*, d *M. v. paradoxa*, with recent strata depicted in pink and white (the current location of these genes is indicated in Supplementary Fig. 8a). The light blue bar at the bottom indicates the genes involved in the young evolutionary stratum of *M. v. caroliniana*.
fused with the entire ancestral PR chromosome, but the putative centromere end of the HD chromosome arm fused to the opposite end of the PR chromosome (Fig. 2b; Supplementary Fig. 5b, c). The much lower levels of synonymous divergence (dS) between the alleles of genes positioned between the PR and HD mating-type loci (Fig. 3a; Supplementary Fig. 5a and Supplementary Table 2) indicated a more recent fusion event. Because not all the 17 genes used for dating mating-type loci linkage in the other species were ancestrally located between the mating-type loci in M. silenes-acaulis (due to the PR fusion in the opposite direction in this species, Fig. 2), we used a restricted set of 13 genes ancestrally located between the HD locus and the putative centromere (blue vertical arrows in Fig. 3) for dating HD–PR linkage in this species. We estimated that mating-type loci linkage occurred ca. 0.2 MYs ago in M. silenes-acaulis (Figs. 3 and 4). Further evidence for a more recent chromosome fusion was also provided by the small number of rearrangements between the a1 and a2 mating-type chromosomes in M. silenes-acaulis (Supplementary Fig. 5a, b), contrasting with the extensive rearrangements observed in the M. lychnitis-dioicae non-recombining region 19. In M. silenes-acaulis, a large inversion encompassing the region between the PR and HD loci (Supplementary Fig. 5a) may have contributed directly to the recombination suppression linking the two mating-type loci. All M. silenes-acaulis autosomes displayed high levels of synteny and dS values of zero between the sequenced a1 and a2 genomes originating from a single diploid individual (Supplementary Figs. 2a and 4a).

Other specific rearrangements led to mating-type locus linkage in the remaining species. All these rearrangements appeared older than those occurring in the ancestor of M. lychnitis-dioicae and M. silenes-dioicae, as shown by the higher dS levels (Fig. 3; Supplementary Table 2) and the inferred earlier occurrence of recombination suppression based on the 17 gene set (Fig. 4). One of the oldest events was estimated to have occurred in M. v. caroliniana, about 2.3 MYs ago (Fig. 4). Unlike those of M. lychnitis-dioicae and M. silenes-acaulis, the M. v. caroliniana mating-type chromosome contained a single ancestral PR chromosome arm (Fig. 2c; Supplementary Fig. 6b). Higher dS values between the alleles of genes ancestrally positioned between the PR and HD loci (Fig. 3b, Supplementary Table 2) and massive rearrangements in the non-recombining region (Supplementary Fig. 6a) provided further evidence for an earlier onset of recombination cessation in M. v. caroliniana than in M. silenes-acaulis. All M. v. caroliniana autosomes were syntenic and with dS values of zero between the sequenced a1 and a2 genomes isolated from a single diploid individual (Supplementary Figs. 2b and 4b).

In M. scabiosae, a more recent event (1.7 MY old; Fig. 4) linked the mating-type loci following a chromosomal rearrangement similar to that in M. lychnitis-dioicae and M. silenes-dioicae (Fig. 2d) but with one extremity of the ancestral PR chromosome becoming incorporated into an autosome (black fragment in the outer track in Supplementary Fig. 7b; Fig. 2d). This particular configuration suggests ectopic recombination within a chromosome arm rather than rearrangement at putative centromeres as described above (Fig. 2). Consistent with the more recent recombination suppression, M. scabiosae displayed less extensive rearrangements between the a1 and a2 mating-type chromosomes than M. v. caroliniana (Supplementary Fig. 7a). Several large inversions nevertheless occurred between mating-type chromosomes. We were unable to sequence two meiotic products of a single diploid individual for M. scabiosae, which probably

Fig. 4 Dates of mating-type loci linkage. Reconstructed phylogenetic tree based on 17 concatenated genes ancestrally located between the mating-type loci after chromosomal fusion in all the studied species with linked mating-type loci but M. silenes-acaulis and including alleles from both a1 and a2 genomes. Numbers on tree nodes indicate the inferred dates of speciation (in black) and the events of mating-type loci linkage, either one to each other or to their respective putative centromeres (in red and orange, respectively). The blue bars correspond to 95% confidence intervals. The scale at the bottom indicates the time before present (million years). None of the individual genes showed trans-specific polymorphism, except between the sister species M. lychnitis-dioicae and M. silenes-dioicae (Supplementary Fig. 3). We used a restricted set of 13 genes (Fig. 3) for estimating the M. silenes-acaulis mating-type loci linkage because not all 17 genes were located in its non-recombining region.
explains the allelic variation observed between the sequenced a1 and a2 genomes even for pseudo-autosomal regions (PARs) and autosomes (Fig. 3c; Supplementary Figs. 2d and 7a). Nevertheless, the M. scabiosae mating-type chromosomes still appear to be exceptional in terms of rearrangements and divergence between the a1 and a2 genomes compared to autosomes (Fig. 3c; Supplementary Figs. 2d, 4c, and 7a).

Unlike those of all other species considered, M. v. paradoxa mating-type chromosomes resulted from the fusion of the entire ancestral PR and HD chromosomes (Fig. 2e; Supplementary Fig. 5k). This species experienced one of the earliest mating-type locus linkage events, with recombination suppression occurring about 2.3 MY ago (Fig. 4). This estimated age of recombination suppression is consistent with the high levels of rearrangements about 2.3 MY ago (Fig. 4). This estimated age of recombination suppression is consistent with the high levels of rearrangements about 2.3 MY ago (Fig. 4). This estimated age of recombination suppression is consistent with the high levels of rearrangements about 2.3 MY ago (Fig. 4).

We identified a young evolutionary stratum in M. v. caroliniana (light blue in Fig. 3b). This genomic region was located distally to the PR mating-type locus and had non-zero dS values significantly lower than the mean dS values for genes ancestrally located between the PR and HD loci (Fig. 3d, Supplementary Table 2). The limit of the light-blue stratum was set at the most distal gene with a non-zero dS value, as all autosomes had zero dS values in the sequenced M. v. caroliniana diploid individual (Supplementary Fig. 2a). The light-blue stratum extended farther into the PAR than the most recent evolutionary stratum in M. lychnidis-dioicae (red and green bars in Fig. 3b). The mean dS value in this evolutionary stratum was not significantly different from that in the PARs (Supplementary Table 2), indicating that mating-type locus linkage was recent. Such stretches of within-individual non-zero dS genes were restricted to non-recombining regions in the M. v. caroliniana diploid individual sequenced (Supplementary Fig. 2b), providing strong evidence for recombination suppression in the light-blue region. Gene order in this region was largely conserved between the a1 and a2 mating-type chromosomes (Supplementary Fig. 5b), demonstrating that recombination can be halted in the absence of inversions. A small localized inversion within this stratum (orange links in Supplementary Fig. 5d) provided further evidence for the lack of recombination in this region. The autosomes were completely collinear between the two haploid genomes in the sequenced diploid individual (Supplementary Fig. 4b).

Evolutionary strata extending beyond the genes involved in mating-type determination were also detected in M. v. paradoxa. In this species, dS values were also highest in the non-recombining region ancestrally located between the PAR and HD loci (Fig. 3d; Supplementary Table 2) and were lower, but non-zero, in the two distal regions. These regions thus likely constitute two additional young evolutionary strata (white and pink, Fig. 3d). The region of recombination suppression extended farther into the PARs than in any of the other studied species, to the extent that only a single, very small PAR was retained (Fig. 3d and Supplementary Fig. 8a). We confirmed the complete suppression of recombination suppression in the white stratum by identifying a small region at the extremity of the M. v. paradoxa mating-type chromosome corresponding to rearranged genes ancestrally located in the center of the chromosome (Fig. 2e, and shown in gray in the outer track in Supplementary Fig. 8b). The pink region on the other side of the mating-type chromosome, distal to the PR locus (Fig. 3d), had high dS values but not higher than those of autosomes in the sequenced M. v. paradoxa individual and without inversions or rearrangements.

We confirmed recombination suppression in the pink, white, and light blue regions by sequencing multiple genomes for M. v. caroliniana and M. v. paradoxa (Supplementary Table 3). For genes linked to the mating-type loci, a1- and a2-associated alleles will differentiate to the point of forming two distinct clades in gene genealogies within species. Gene genealogies revealed such pattern for genes in the pink, white, and light blue regions, where the alleles associated with a given mating type were significantly more clustered than for genes in the PARs or in autosomes (Supplementary Fig. 9, Supplementary Table 4). In all three regions, all a1 alleles branched in one clade and all a2 alleles in
another clade in multiple gene genealogies (Supplementary Fig. 9), strongly supporting full linkage to mating type. Furthermore, the mean levels of polymorphism per mating type and per species were significantly lower in all the evolutionary strata than in the PARs (Supplementary Fig. 10, Supplementary Table 5), as expected in regions without recombination due to the lower effective population size. These findings indicated that the non-zero $d_s$ values in the evolutionary strata within the sequenced individuals were due to recombination strata within the sequenced individuals were due to recombination suppression rather than higher polymorphism levels.

**Fig. 5** Differential degeneration across strata and species. We quantified the TE content and gene loss in genomes of both mating types ($a_1$ and $a_2$) of all species under study. For each species, we measured the TE accumulation separately for one fully assembled autosome (as a control), recombining regions (RR), and non-recombining regions (NRR) on mating-type chromosomes (MAT), separating the youngest evolutionary strata (light blue, red, green pink, and white strata) from the remaining of the NRR where applicable. Strata were ordered from the youngest to the oldest per species. In *M. lagerheimii*, the NRRs correspond to the regions between the mating-type loci and the putative centromeres, while in the other species they mostly correspond to the regions ancestrally between the HD and PR loci. The purple and blue strata were too rearranged within the large non-recombining region to quantify their specific gene loss or TE content except in *M. lagerheimii*. **a** Transposable element (TE) content (percent of base pairs); **b** Gene loss (genes with an allele present in the genome of one mating type but absent from the genome of the opposite mating type). Numbers at the top of the bars indicate the numbers of genes missing in the $a_2$ mating-type chromosome, present only in the $a_1$ mating-type chromosome.
Gene loss and TE accumulation. We found evidence of differential gene loss and TE accumulation across species and evolutionary strata. Levels of gene loss and TE content increased with the age of recombination suppression, both within and across species (Fig. 5). Even the youngest evolutionary strata showed footprints of genomic decay relative to recombining regions (Fig. 5). The regions ancestrally located between the PR and HD loci displayed higher levels of degeneration in species with older recombination suppression events than in species with more recent mating-type locus linkage. Higher TE loads resulted in chromosomes larger than those in the ancestral state and contributed to differences in size between mating-type chromosomes (Figs. 2 and 5). The PARs and youngest strata displayed little evidence of TE accumulation compared to autosomes and showed low but non-negligible levels of gene loss (Fig. 5). Mating type (a1 versus a2) had no significant effect on gene loss or TE content, while differences between species were significant (Supplementary Table 6). The onset of genomic degeneration is thus rapid, with further gradual accumulation of TEs. Gene loss was extensive on both mating-type chromosomes: the two species with the most ancient recombination suppression between the mating-type loci lost between 60 and 70% of genes in this region within 2.3 MY, and *M. silenes-acaulis* has already lost >20% of genes within 0.20 MY.

**Discussion**

We report an unprecedented case of convergent evolution, with five parallel recombination suppression events independently linking PR and HD mating-type loci in anther-smut fungi and generating megabase-long supergenes beneficial under selfing mating systems. We also reveal the convergent evolution of young evolutionary strata in multiple closely related species. Furthermore, our unique dataset suggests a progression of genomic decay in non-recombining mating-type chromosomes, with older regions of suppressed recombination displaying higher levels of gene loss and TE accumulation.

The convergent evolution of megabase-long supergenes in *Microbotryum* mating-type chromosomes, through distinct genomic rearrangements, has repeatedly led to the beneficial co-segregation of different mating-type functions (pheromones and pheromone receptors controlling pre-mating compatibility encoded by the PR locus, and the homeodomain proteins controlling post-mating compatibility encoded by the HD locus). A previous parsimony analysis indicated a very low probability of independent mating-type loci linkage events in anther-smut fungi and instead inferred reversal to the ancestral state of unlinked PR and HD loci in *M. lagerheimii*34. However, our analyses based on well-assembled chromosomes contradict these earlier inferences and reveal the striking occurrence of repeated convergent events linking mating-type loci through distinct genomic rearrangements. Further evidence for convergent linkage events is provided by differences in synonymous divergence between alleles associated with alternative mating types, the absence of trans-specific polymorphism in genes ancestrally located between the PR and HD loci, and differences in inferred linkage dates, with linkage occurring after speciation in at least five lineages. The existence of numerous other species with linked PR and HD mating-type loci across the *Microbotryum* genus34,38 suggests the existence of many more independent events of mating-type locus linkage and convergent supergene formation in this genus. The occurrence of recombination suppression linking mating-type loci to each other or to centromere has been documented in a few interspersed lineages across fungi37,18,21,29,31, but we provide here the first report of repeated convergent mating-type locus linkage events in multiple closely related species.

Repeated evolution of mating-type locus linkage in multiple closely related species implies very strong selection, which is supported by the lack of similar large-scale rearrangements or recombination suppression in autosomes. In selfing-based mating systems, as observed in all species of *Microbotryum* studied to date39,40, HD–PR linkage increases the odds of compatibility between the gametes of a given diploid individual (Supplementary Figs. 1a, c) and is expected to be favored by selection. Interestingly, although the selfing species *M. lagerheimii* has unlinked mating-type loci34, the HD and PR loci are linked to the putative centromeres of their corresponding chromosomes34, which also increases gamete compatibility under selfing via intra-tetrad mating (automixis) (Supplementary Fig. 1b).

Our study provides invaluable insight into the frequency and importance of supergene evolution, a topic currently under explored4. Other examples of supergenes include the linkage of genes involved in mating types in plants and algae, mimicry wing patterns in butterflies, and complex social behavior in ants40,11,13,15,22. The repeated convergent evolution of non-recombining regions in anther-smut fungi provides strong support for the view that chromosomal rearrangements and the formation of supergenes are frequent and play an important role in evolution and adaptation10,13,41. The existence of repeated transitions in the genomic architecture of mating-type determination, following different chromosomal rearrangements, illustrates the power of natural selection and high genomic fluidity in shaping adaptation. Our findings suggest that natural selection drives evolution along trajectories leading to similar phenotypic outcomes through different genomic changes24. Very few instances of convergent evolution have been documented beyond the textbook examples of Nicaraguan crater lake cichlid fishes5, cave morphs in Mexican cavefishes6, and non-recombining chromosomes controlling polymorphic social behavior in ants10,11.

We also found striking convergence in the stepwise extension of recombination suppression beyond mating-type genes, generating independent evolutionary strata in several *Microbotryum* species. The evolutionary causes leading to such strata devoid of mating-type genes are more likely to be the sheltering of deleterious alleles or neutral rearrangements than beneficial gene linkage16,24,42. Finding independent evolutionary strata in multiple closely related species provides further support for the occurrence of repeated evolution towards similar chromosomal states.

Our study provides important clues to the proximal mechanisms underlying the evolution of recombination suppression. Contrary to the common view that chromosomal inversions play a major role in preventing recombination14, we found that recombination cessation can occur with the conservation of collinearity, as previously documented in fungi16,20,21. The occurrence of mating-type locus linkage via different routes reveals a high degree of genomic fluidity. Chromosomal rearrangements occurred frequently at putative centromeres, as in a recently reported case in the human fungal pathogen *Cryptococcus neoformans*20. These repeat-rich regions are highly labile and some of the inferred fusions in anther-smut fungi encompassed two ancestral putative centromeres in the same chromosome, as in *M. v. paradoxa*.

By the examination of repeated supergene formation events of contrasting ages, our results add further insights on the tempo of genomic decay and TE accumulation after recombination suppression. We found that gene loss occurred more rapidly than rearrangements or repeat accumulation. TE accumulation rates differed between species, probably because the *Microbotryum* species-specific TE loads66 affect transposition rates. The high levels of degeneration observed in the two mating types probably
resulted from less efficient selection, due to lack of recombination and the sheltering of deleterious mutations in a permanently heterozygous state, with only very brief periods of haploid selection restricted to the meiotic tetrad stage. Gene loss in these fungal mating-type chromosomes was more rapid than in the Y chromosome of the plant Silene latifolia, likely because plant sex chromosome degeneration is delayed by haploid purifying selection, unlike in animals or Microbotryum fungi. Contrasting with sex chromosomes, genomic degeneration in Microbotryum mating-type chromosomes was not asymmetric (with no significant effect of mating-type on TE content or gene loss), as expected for organisms with an obligate heterozygous mating-type or sex chromosomes. Lethal alleles linked to mating type were found at relatively high frequencies in natural populations of several Microbotryum species, preventing haploid growth in vitro but maintained through high levels of intratetrad mating. We also detected non-negligible levels of gene loss in the PARS. This may reflect the existence of very recent, undetected evolutionary strata or lower rates of recombination in the PARS compared to fully recombining autosome. This second hypothesis is consistent with the suggestion that partial deleterious allele sheltering in the PARS may account for evolutionary strata. Low recombination rates in the PARS would allow the accumulation of deleterious alleles, leading to selection for further recombination suppression and the permanent sheltering of these deleterious alleles.

In conclusion, our findings reveal remarkable repeated convergence in young mating-type chromosomes in closely related species, with supergenes evolving rapidly and frequently. Furthermore, our study shows that natural selection can repeatedly lead to similar phenotypes through multiple different evolutionary trajectories and genomic changes, rendering evolution predictable. The very recent advances in sequencing technologies yielding high-quality genome assemblies are allowing in-depth studies of chromosomal architecture and documenting the importance and prevalence of supergenes, as well as the high degree of genomic fluidity and convergence. Future studies will certainly lead to the identification of many more cases of beneficial gene linkage, as predicted from evolutionary theory.

Methods

Strains, DNA extraction, and sequencing. Microbotryum violaceum is a plant pathogen species complex that includes recently recognized cryptic and host species, which have not all been formally named yet. For species without Latin names, we used M. violaceum (the terminology used to denote the whole species complex) followed by the name of the host plant, as is typically done in pathogen species complex that includes recently recognized cryptic and host species. Assembly and annotation. The PARS and the full genome were assembled with the wgs-8.2 version of the PBcR assembler with the following parameters: genomeSize = 300,000,000, assembleCoverage = 50. Assemblies were polished with quiver algorithm of smrtanalysis suite 2.3.0 (https://github.com/Paci).

Figures and statistical tests. Supplementary Figs. 4–8 were prepared using Circos. We analyzed gene order after removing TEs to identify larger blocks of orthology. We identified syntenic regions by searching all one-to-one gene correspondences between pairs of haploid genomes based on the orthologous groups reconstruction (see above). Statistical tests (Student’s t-test, analysis of variance, and Wilcoxon rank tests) were performed using JMP v7 (SAS Institute).

Date estimates for recombination suppression and genealogies. For dating HD/PR linkage, we used alignments including a1 and a2-associated alleles at 17 single-copy orthologous groups that were located between the PR and HD loci following chromosomal fusion in all species but M. silenes-acaulis and that had both alleles retained (red vertical arrows in Fig. 3). The divergence between alleles associated with the a1 versus a2 mating types in these genes corresponds to the date of their linkage to mating-type loci. Indeed, genes linked to mating-type loci maintain one allele associated with a1 and another allele associated with a2 over time and the differentiation between these two alleles increases with the time since linkage to mating type. We only used the genes that displayed trans-specific polymorphism between M. lychnidis-dioicae, M. silenes-dioicae and M. violaceum sense stricto to avoid biasing estimations to younger dates because of gene conversion. For M. silenes-acaulis, we used a restricted set of 13 genes ancestrally located between the HD locus and the putative centromere (blue vertical arrows in Fig. 3) because most genes located in its non-recombining region were in recombining regions in other species (Fig. 2). Divergence times were estimated using BEAST v2.4.0 with XML inputs generated using BEAUTi, and the default parameters except for unlinked substitution (HKY + G with empirical frequencies for each codon position) and clock models. Vale process to model speciation and 10,000,000 mcmc generations sampled every 1000. We used a single calibration prior at 0.42 MY for all runs, corresponding to the divergence between...
M. lycidios-dioicae and M. silenes-dioicae25, with a normal distribution and a sigma of 0.05. In some of the 17 individual gene trees ancestrally located between the HE and RR loci, some basal nodes were different from those in the species trees (Supplementary Fig. 3). However, the incongruent nodes were weakly supported so that gene genealogies were actually not significantly different from the species tree (P > 0.35, AU test26). We therefore forced the tree resulting from the concatenation of these 17 genes to the species tree topology for the date estimate analysis in BEAST. Genealogies for 17 genes were inferred for codon-based alignments of genes in the different strata using RAxML55 version 8.2.7, assuming the GTR+GCAT model.

Identification of TE repeats. Repetitive DNA content was analyzed with RepeatMasker60 using RepBASE v19.11.167. We used the RepeatMasker output to compute the percentage of base pairs occupied by TEs across the different evolutionary strata and for each species. The putative centromeric repeats were filtered out with Bowtie267 along chromosomes, repeats were removed. Further filtering of repeats was performed by blasting (BLASTX), with removal of repeats matching to more than five locations in the genome.

Detection of centromeric repeats. We identified centromeric-specific repeats using a method specifically designed for this purpose26, based on the observation that in most species studied to date putative centromeres contain the most abundant tandem repeats, are gene poor, and repeat rich. For identifying centromeric repeats, we used Tandem-Repeat Finder (TRF v. 0.47b5) on assembled Illumina reads of the M. lagerheimii strain as the one sequenced using the Pacific Bioscience technology. We performed the assemblies as follows: we randomly chose 500,000 Illumina reads that we assembled with PRICE v1.27 using a random seed of 1,000,000 reads as seed file and using the following command line arguments: -fпп (or -mрр when using mate-pair reads) inputF1, R1 inputF1, R2 650 90 -pсf20 20000 seedF1 500 25 -nc 150 -pсf 20000 seedF2 500 25 -nc 10 -mpi 85 -TPI 95 -psi 85 -TPI 95 -logf logf -o outputF1. PRICE works by rounds of assembly: in the first round, it maps randomly chosen 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, mappingreads, assembling the unmapped reads, and extending the reference contigs. We analyzed the presence of tandem repeats in each of the 10 assembly cycle outputs using the following parameters in a TRF wrapper perl script26: 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. We identified the putative centromeres in M. lagerheimii as the most gene-poor and repeat-rich regions and with the most abundant tandem repeats. The deletions of the centromeric regions using this method yielded a single region per contig and were congruent with those using BLAST of the centromeric repeat identifications previously26. Putative centromeres in A. a2 were identified in the other species by blasting the identified centromeric-specific repeats (the tandem repeats identified here and previously19 gave congruent results, see Supplementary Figure S4b).

Gene loss. Alleles were identified by applying orthomcl26 to the protein data sets for unique a1 and a2 orthologs, discarding orthologous groups containing more than one protein-coding gene per mating type on the karyotype. The putative centromeres were identified in the other species by blasting the identified orthologs in the opposite mating type within a diploid genome. We computed the number of gene losses across the PARs and the evolutionary strata that were not rear ranged to be delimited.

Polymorphism data and analyses. To rule out high levels of polymorphism as the cause for the observed high dS values in the youngest strata of the Microbotryum mating types, we computed the level of polymorphism and a1–a2 allele segregation in gene genealogies. When genes are linked to mating-type loci, alleles associated with the a1 versus a2 alleles accumulate differences until completely segregating according to mating-type allele rather than according to strain in gene genealogies. To test for this pattern, we re-sequenced multiple strains of M. v. paradoxa and M. v. caroliniana. Alleles were identified when a protein-coding gene in one mating type did not have any match in the orthomcl output in the opposite mating type within a diploid genome. We computed the number of gene losses across the PARs and the evolutionary strata that were not rearranged to be delimited.


Acknowledgements

This work was supported by the ERC starting grant GenomeFun 309403 to T.G., the Louis D. foundation award to T.G., the NSF DEB-1115765 and NIH R15GM119092 grants to M.E.H., the Marie Curie European grant 70164 to S.B., a postdoctoral fellowship (SFRH/BPD/79198/2011) from Fundação para a Ciência e a Tecnologia, Portugal to M.A.C, and DAAD, the Marie Curie European grant PRESTIGE-2016-4-0013 to F.H. and Campus France (PPP 57211753) to T.G. and D.B., and the Montana State University Agricultural Experiment Station. F.H. received the Young Biological Researcher Prize from the Fondation des Treilles, created by Anne Gruner Schlumberger, which supports research in Science and Art (http://www.les-treilles.com). We thank Cécile Fairhead for help with DNA extraction. PacBio sequencing was conducted at the IGM Genomics Center, University of California, San Diego, La Jolla, CA. We thank Jérôme Gouzy and Jean-Tristan Brandenburg for invaluable help with genomic analyses.

Author contributions


Additional information

Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-018-04380-9.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s) 2018