

Tempo of degeneration across independently evolved non-recombining regions

Fantin Carpentier, Ricardo Rodríguez de La Vega, Michael Perlin, R. Margaret Wallen, Michael Hood, Tatiana E Giraud

▶ To cite this version:

Fantin Carpentier, Ricardo Rodríguez de La Vega, Michael Perlin, R. Margaret Wallen, Michael Hood, et al.. Tempo of degeneration across independently evolved non-recombining regions. 2021. hal-03365752

HAL Id: hal-03365752 https://hal.science/hal-03365752

Preprint submitted on 5 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1	Tempo of degeneration across independently evolved non-recombining							
2	regions							
3								
4	Fantin Carpentier ^{1,2} , Ricardo C. Rodríguez de la Vega ¹ , Michael H. Perlin ³ , R. Margaret							
5	Wallen ³ , Michael E. Hood# ⁴ , Tatiana Giraud#* ¹							
6								
7	# These authors co-supervised the study							
8 9 10 11	¹ Ecologie Systématique Evolution, Bâtiment 360, CNRS, AgroParisTech, Université Paris-Saclay, 91400 Orsay, France ² Université de Lille, CNRS, LIMR 8198-Evo-Eco-Paleo E-59000, Lille, Erance							
12	³ Department of Biology, Program on Disease Evolution, University of Louisville, Louisville, KY 40292, USA.							
13	⁴ Department of Biology, Amherst College, 01002-5000 Amherst, Massachusetts, United State of America.							
14								
15	*Corresponding author: Tatiana Giraud							
16	Laboratoire Ecologie, Systématique et Evolution, Bâtiment 360, Université Paris-Saclay, 91400							
1/	Orsay France							
18	phone: $+35109153009 + 35780010739$ 1ax: $+35109154097$							
19	tatiana.gnauu@universite-paris-saciay.n							
20	Running title:							
21	Key words: synonymous codon usage, genomic degeneration, degeneration tempo, mating-							
22	type chromosomes, recombination suppression, frequency of optimal codon, fungi							
23								
24 25 26 27 28 29 30 31 32 33 34 35 36 27								

38 Abstract

39

Recombination is beneficial over the long term, allowing more effective selection. Despite 40 long-term advantages of recombination, local recombination suppression is known to evolve 41 42 and lead to genomic degeneration, in particular on sex and mating-type chromosomes, sometimes linked to severe genetic diseases. Here, we investigated the tempo of degeneration 43 in non-recombining regions, i.e., the function curve for the accumulation of deleterious 44 mutations over time, taking advantage of 17 independent events of large recombination 45 suppression identified on mating-type chromosomes of anther-smut fungi, including five newly 46 identified in the present study. Using high-quality genomes assemblies of alternative mating 47 types of 13 Microbotryum species, we estimated the degeneration levels in terms of 48 49 accumulation of non-optimal codons and non-synonymous substitutions in non-recombining regions. We found a reduced frequency of optimal codons in the non-recombining regions on 50 51 mating-type chromosomes compared to autosomes. We showed that the lower frequency of 52 optimal codons in non-recombining regions was not due to less frequent GC-biased gene 53 conversion or lower ancestral expression levels compared to recombining regions. We estimated that the frequency of optimal codon usage decreased linearly at a rate of 0.989 per 54 55 My. The non-synonymous over synonymous substitution rate (d_N/d_S) increased rapidly after recombination suppression and then reached a plateau. To our knowledge this is the first study 56 57 to disentangle effects of reduced selection efficacy from GC-biased gene conversion in the evolution of optimal codon usage to quantify the tempo of degeneration in non-recombining 58 regions, leveraging on multiple independent recombination suppression events. Understanding 59 the tempo of degeneration is important for our knowledge on genomic evolution, on the origin 60 of genetic diseases and on the maintenance of regions without recombination. 61

- 62 63 64 65 66 67 68 69 70 71 72 73 74 75
- 76

77 Introduction

Recombination is beneficial over the long term in most situations, allowing higher 78 selection efficacy and therefore more rapid adaptation. Meiotic recombination, through 79 crossing-over resulting in the reciprocal exchange of DNA segments between two homologous 80 chromosomes, shuffles allelic combinations, such that more beneficial combinations might be 81 formed [1, 2] and selection at different loci can be decoupled [3, 4]. Across long time scales, 82 many crossover events indeed occur along chromosomes and selection can apply at each locus 83 independently. Recombination prevents selection interference between loci [5], thereby 84 facilitating the purge of deleterious mutations. In the absence of recombination, selection on 85 strongly beneficial alleles can drag along and increase the frequency of deleterious alleles at 86 linked loci through genetic hitchhiking [6]. Recombination also prevents the accumulation of 87 88 deleterious mutations by generating chromosomes with fewer deleterious mutations than parental ones, thus avoiding the increase in the frequency of genotypes with a higher number 89 of harmful mutations, known as Muller's ratchet [1, 2]. 90

Despite these long-term advantages of recombination, local recombination suppression 91 can be selected for, maintaining beneficial allelic combinations, and leading to a genetic 92 93 structure known as a supergene, where multiple genes are linked together and are transmitted as a single locus [7, 8]. Examples of complex phenotypes controlled by supergenes include 94 95 wing patterns in the Heliconius butterflies [9, 10], social systems in ants [11] and mating 96 compatibility systems, such as sex chromosomes or mating-type loci [12-15]. The suppression of recombination can thus arise and be maintained by selection, but the corresponding genomic 97 regions then accumulate deleterious mutations through interferences between loci and Muller's 98 99 ratchet [7, 16-19]. Typical deleterious mutations include non-synonymous substitutions [20-25] which alters the amino-acid sequences of a protein and may result in protein dysfunction. 100 Deleterious mutations also include frameshift mutations, gene losses and substitutions leading 101 102 to suboptimal gene expression level [26-28]. Such deleterious mutations in sex chromosomes may be responsible for genetic diseases [29, 30]. Genomic rearrangements [31], epigenetic 103 modifications [32] and transposable elements [19, 28] also accumulate in non-recombining 104 regions; they are not *per se* deleterious mutations but may disrupt gene function or expression 105 and impose strong energetic costs [33, 34]. 106

A less studied type of deleterious mutations concerns some synonymous substitutions.
Synonymous substitutions are usually considered neutral because they do not alter the aminoacid sequences of proteins. However, synonymous codons are often not used randomly [35, 36],

some codons being preferentially used over their synonymous alternative codons. This preference is thought to result from a selection on either the rate or the accuracy of translation [37], such that the preferred codons are called optimal codons. Optimal codons tend to correspond to the most abundant tRNAs [38, 39], and are most frequent in highly expressed genes [40, 41].

The existence of preferred codons among synonymous alternatives leads to a codon 115 usage bias, which has been found directly proportional to the recombination rate along genomes 116 [42]. Such deviations from preferred codon usage has been interpreted as resulting from the 117 decrease in selection efficacy in regions with low or no recombination [27]. Compared to 118 recombining regions, lower frequencies of optimal codons have been reported in the non-119 recombining region of the mating-type chromosomes in the fungus Neurospora tetrasperma 120 [25], in the non-recombining dot chromosome in *Drosophila americana* [43] and twelve other 121 Drosophila species [44]. However, codon biases could also result from the biased mutational 122 pattern caused by GC-biased gene conversion. Occurring in regions of frequent recombination, 123 124 gene conversion results from the resolution of base-pair mismatches in the heteroduplex of the recombination process [45]. GC-biased gene conversion has been shown to occur in bacteria 125 [46] and in eukaryotes [45, 47-49]. The effect of either fewer GC-biased gene conversion events 126 or less efficient selection may each lead to fewer optimal codons in non-recombining regions, 127 and their respective effects may be difficult to disentangle. Few studies have tried to distinguish 128 129 the roles of relaxed selection versus GC-biased gene conversion [50] and no study to our knowledge specifically attempted to control for the effect of GC-biased gene conversion in 130 131 assessing the effect of the lack of recombination on the frequency of optimal codons.

While the accumulation of deleterious mutations in non-recombining regions has been 132 reported in numerous systems, only a few studies have addressed the question of the tempo of 133 134 degeneration *i.e.* the rate at which deleterious mutations accumulate with time. Theoretical works based on the mammalian Y-chromosome predicted that degeneration through gene losses 135 should occur at a high speed in young non-recombining regions, mainly due to Muller's ratchet 136 and background selection, and should then slows down in older non-recombining regions in 137 which genetic hitchhiking is a major driving force [16, 51]. In the plant Silene latifolia, the sex 138 chromosomes underwent two successive recombination suppression events (i.e., evolutionary 139 140 strata; [52]), which have been losing genes at estimated rates of 4.64% and 6.92% per Myr for the older and younger evolutionary strata, respectively [53]. Other observations have similarly 141 142 indicated that the gene loss rate is negatively correlated with the age of recombination

143 suppression [54, 55], and these findings are in accordance with theoretical models [51]. Studying the tempo of degeneration requires estimating the time since recombination cessation and 144 145 gathering data for multiple independent events of recombination suppression of various ages. Recombination suppression evolved numerous times in a wide range of organisms, even 146 147 sometimes within the same genus and at different times [56, 57]. Moreover, the evolutionary strata present in some sex chromosomes represent stepwise extension events of recombination 148 149 suppression that occurred at different time points; e.g., the four successive steps of recombination suppression on the mammalian Y chromosome [52, 58, 59]. 150

In the fungal genus *Microbotryum*, multiple events of recombination suppression and 151 152 the formation of evolutionary strata occurred at different dates on mating-type chromosomes, 153 and independently in several species [60-62]. In these pathogenic anther-smut fungi, mating 154 occurs mostly between haploid cells resulting from the same meiosis event (*i.e.*, intra-tetrad mating or automixis [63]) and between cells bearing distinct alleles at the two mating-type loci, 155 *i.e.* the PR locus (carrying pheromone and pheromone receptor genes controlling gamete 156 fusion) and the HD locus (carrying homeodomain genes whose proteins undergo dimerization 157 158 to induce post-mating growth). Recombination suppression linking a few genes within each of the two mating-type loci is required for ensuring the mating-type functions, and is ancestral to 159 the basidiomycete fungi [13, 64]. In Microbotryum and some other selfing basidiomycetes, the 160 161 linkage of the two mating-type loci to each other or the linkage of each mating-type locus to their respective centromere has been favoured by selection [13, 60-62]. Indeed, under intra-tetrad 162 163 mating these linkage relationships increase the odds of gamete compatibility (producing only two inter-compatible mating-type phenotypes among haploid meiotic products) compared to 164 independent segregation (producing up to four different mating types among gametes following 165 a meiosis). Across the Microbotryum genus, five independent recombination suppression 166 events have been shown to link the PR and HD loci to each other [61]. The two mating-type 167 168 loci were ancestrally located on different chromosomes, and multiple PR-HD linkage events across lineages involved distinct rearrangements (fusion and/or fission) of the ancestral PR and 169 170 HD chromosomes [61]. Two additional recombination suppression events in other Microbotryum species independently linked each of the PR and HD loci to their respective 171 centromere [62]. Furthermore, extensions of recombination suppression occurred stepwise and 172 independently in several *Microbotryum* species, forming evolutionary strata of different ages 173 [60, 61]. The evolutionary strata linking HD and PR loci to each other or to centromeres were 174 called "black evolutionary strata" while extensions of recombination suppression to non-175

mating-type genes were called coloured evolutionary strata [60, 61]. Evolutionary strata formed
around each of the PR and HD locus (named "purple" and "blue" strata, respectively) and have
likely evolved before the radiation of the *Microbotryum* genus because they are shared by all
species studied so far. Other species-specific strata formed more recently and distal to the fused
PR-HD loci (*e.g.*, the pink and white strata in *M. v. paradoxa* and light blue stratum in *M. v. caroliniana*). All species retained a recombining region at least at one extremity of their matingtype chromosome, called the pseudo-autosomal region (PAR).

The non-recombining regions in the *Microbotryum* mating-type chromosomes seem to 183 have accumulated multiple footprints of degeneration, in the form of higher rate of non-184 synonymous substitutions, transposable element accumulation, gene losses and reduced gene 185 expression [65]. However, estimates in that early study were not optimal because of poor 186 assembly quality and the wrong assumption that the non-recombining regions were homologous 187 across species, having evolved once in the genus. Degeneration was later found in five 188 189 *Microbotryum* species with recombination suppression using high-quality assemblies, including gene gain/loss and transposable element accumulation [61]. The degeneration level 190 was found to be stronger in regions with older recombination suppression [61], but the tempo 191 of degeneration was not studied. Degeneration in terms of gene expression level was also found 192 based on the high-quality assembly of *M. lychnidis-dioicae*, with reduced expression being 193 194 associated with different types of degenerative mutations [66].

Here, we investigated the tempo of degeneration in non-recombining regions, *i.e.*, the 195 curve of the relationship between deleterious mutation accumulation and time since 196 recombination cessation, taking advantage of the multiple independent linkage events and of 197 evolutionary strata of different ages in the anther-smut fungi. We first identified additional 198 independent events of recombination suppression, linking the PR and HD loci, on the mating-199 type chromosomes of the M. v. tatarinowii, M. v. melanantha and M. coronariae species, 200 leading to a total of eight independent events of recombination suppression between mating-201 202 type loci across *Microbotryum* species. We also identified two additional old strata in the large non-recombining mating-type chromosome of M. v. paradoxa. Using high-quality genomes 203 assemblies of alternative mating types of 13 Microbotryum species and a total of 17 independent 204 evolutionary strata of different ages, we estimated the degeneration levels in terms of the 205 accumulation of non-optimal codons and non-synonymous substitutions in non-recombining 206 regions. We identified optimal codons as those enriched in highly expressed autosomal genes 207 and tested whether they were less frequent in non-recombining regions controlling for possible 208

effects: i) of lower GC-biased gene conversion in non-recombining regions, by comparing GC 209 content in coding versus non-coding regions, as selection on codon usage acts only in exons 210 while GC-biased gene conversion also impacts introns and inter-genic sequences, and ii) of the 211 ancestral state of expression levels in non-recombining regions, taking as a proxy the expression 212 213 level in a closely related outgroup without recombination suppression. We then investigated the relationship between degeneration level, in terms of non-synonymous substitutions 214 215 accumulation (d_N/d_S) and of decrease in frequency of optimal codons, and the time since 216 recombination suppression in the different evolutionary strata and species. We found that the 217 frequency of optimal codons decreased nearly linearly and that the non-synonymous over synonymous substitution rate accumulated initially rapidly but then reached a plateau. 218 Understanding the tempo of degeneration is important for our knowledge on genomic evolution, 219 in particular for the maintenance of regions without recombination, that can be associated with 220 lower fitness and human genetic diseases. 221

- 222
- 223
- 224 **Results**

225

226 PR-HD linkage in three new Microbotryum genomes

We sequenced (long-read Pacific Bioscience technology) and assembled the genomes from 227 haploid cells bearing opposite mating types (called a₁ and a₂) of a single diploid individual, for 228 229 the following species: *Microbotryum violaceum* (s.l.) parasitizing *Silene melanantha* (called M. v. melanantha), M. violaceum (s.l.) parasitizing S. tatarinowii (called M. v. tatarinowii) and M. 230 coronariae parasitizing Silene flos-cuculi (Syn. Lychnis flos-cuculi) (Figure 1; Table S1). We 231 reconstructed the evolutionary history of their mating-type chromosomes by comparing their 232 genome structure to those of *M. intermedium*, taken as a proxy for the genomic ancestral state, 233 following previous studies [60, 61]. In M. intermedium, the PR and the HD mating-type loci are 234 235 located on distinct mating-type chromosomes (Figure S1).

In *M. v. melanantha* and *M. coronariae*, the mating-type chromosomes carried both PR and HD loci and were found to result from the fusion of the whole ancestral PR chromosome with the short arm of the ancestral HD chromosome (Figure S2), as was previously found in *M. lychnidis-dioicae*, *M. silenes-dioicae* and *M. violaceum sensu stricto* [60, 61]. Although the mating-type chromosomes of the five species in this clade resulted from the same chromosomal

rearrangements, trans-specific polymorphism analyses suggested that there were three events 241 of complete recombination cessation in this clade (Figure 1). Indeed, as soon as recombination 242 cessation links a gene to the mating-type locus, the alleles of this gene associated to the 243 alternative mating-type alleles start diverging and accumulate mutations independently. The 244 245 alleles of this non-mating-type gene will thus cluster per mating type, even across speciation events. The node to which trans-specific polymorphism extends therefore indicates the date of 246 247 complete recombination cessation [13]. Trans-specific polymorphism in the genes present in the non-recombining regions in all species but M. intermedium and M. silenes-acaulis 248 suggested the existence of three independent events of complete recombination cessation, one 249 in M. v. melanantha, one in the lineage leading to M. lychnidis-dioicae and M. silenes-dioicae, 250 and one in the lineage leading to *M. coronariae* and *M. violaceum* (s.s.) (Figure 1C). There 251 could have been three independent similar chromosomal rearrangements, or a single one but 252 253 without initially complete recombination cessation, which would have only occurred later, independently in the three lineages. 254

255 In M. v. tatarinowii, the mating-type chromosomes resulted from the fusion of a part of the long 256 arm of the PR chromosome with the short arm of the HD chromosome (Figure 1; Figure S2). We found the remaining part of the ancestral PR chromosome on an autosomal contig in the 257 258 two haploid genomes of M. v. tatarinowii. This represents the same type of chromosomal rearrangement leading to PR-HD linkage as found previously in M. v. caroliniana [61]. 259 260 However, the phylogenetic placement on the two species show that they represent independent events (Figure 1). In addition, we improved the assemblies of M. scabiosae (Table S2) and 261 refined the chromosomal rearrangement scenario for M. v. caroliniana. We could thus refine 262 the breakpoint locations for the fissions and fusion generating their mating-type chromosomes 263 (Figure S3). 264

265 The tree of single-copy genes belonging to non-recombining regions (black strata) in all species (but *M. intermedium* and *M. silenes-acaulis*) displayed a different placement of *M. v. paradoxa* 266 compared to the species tree: M. v. paradoxa is sister species to the clade encompassing M. 267 coronarie, M. lychnidis-dioicae and M. violaceum s.s. in the species tree (Fig. 1A), while it is 268 269 placed as a sister species to the *M. lagerheimii-M. saponariae* clade in the black stratum tree (Figs 1A and C). This discrepancy strongly suggests an event of introgression of the black 270 stratum in M. v. paradoxa shortly after speciation. This reinforces the view that mating-type 271 chromosomes in fungi may be particularly prone to introgression, as this process may 272 273 counteract the effect of degeneration following recombination suppression [67-69].

274

275 Evolutionary strata

We investigated the existence of evolutionary strata in M. v. melanantha, M. v. tatarinowii and 276 *M. coronariae* following the methods used previously [60, 61], *i.e.*, by plotting the synonymous 277 divergence calculated between the alleles from the genomes of opposite mating types along the 278 ancestral-like gene order of *M. intermedium* (Figure 2). The rationale of this method is that, as 279 soon as some genes become linked to the mating-type genes, the alleles associated with the two 280 mating-type alleles gradually accumulate substitutions independently, so that their 281 282 differentiation increases with the time since recombination suppression. Stepwise extension of recombination therefore yields a pattern of discrete strata with different levels of differentiation 283 284 between alleles on the mating-type chromosomes. In contrast, we expect high levels of homozygosity in these selfing species in recombining regions, including pseudo-autosomal 285 286 regions, and therefore no differentiation between the two genomes of a diploid individual. We identified four evolutionary strata in each of M. v. melanantha, M. v. tatarinowii and M. 287 288 coronariae. Three strata are ancient and shared among most *Microbotryum* species, *i.e.*, the purple stratum around the PR locus, the blue stratum around the HD locus, and the orange 289 290 stratum adjacent to the purple stratum gene (Figure 2). In addition, recombination suppression evolved independently, at different times, for linking the HD and PR loci together [60, 61]. We 291 called all these regions involved in the different events of HD-PR linkage the black evolutionary 292 strata, but they resulted from different types of chromosomal fissions and fusions and therefore 293 encompassed distinct gene sets (Figure 1). We did not identify additional evolutionary strata 294 that would be younger than the PR-HD linkage events in any of the species that we sequenced 295 296 in the present study. To run statistical analyses on independent events of recombination suppression, we only analysed the genomic regions corresponding to the black strata (linking 297 HD and PR loci one to each other or to their centromere) and to the species-specific strata, and 298 not the older strata shared by multiple species. 299

We however did detect here, in the process of building the single-copy gene trees, that the fusion of the whole HD and PR mating-type chromosomes in *M. v. paradoxa* to form a single mating-type chromosome led to a higher number of successive events of recombination cessation than previously thought. The two most recent strata in *M. v. paradoxa* had been identified previously [61], are distal to the mating-type loci and were called pink and white strata (Figure 1A). The oldest event of recombination suppression corresponds to the telomere-totelomere fusion region and includes genes ancestrally located between the HD locus and its

short arm telomere, as well as genes in the short arm of the PR chromosome ("brown" stratum, 307 Figure 1A). Recombination then stopped left to the centromere on the long arm of the ancient 308 HD chromosome ("light-brown" stratum), and shortly after in the regions between the HD and 309 PR and the centromeres in the ancestral chromosomes ("black" strata of M. v. paradoxa, Figure 310 1A). The independence of these three events of recombination cessation is supported by i) the 311 non-overlapping distributions of the estimated times of recombination suppression for the 312 313 brown stratum and for the light-brown or black strata (inset Figure 1A) and ii) the different tree topologies for the genes in the brown or light-brown strata versus the black stratum (Figure 1C, 314 Figure S4). 315

316

317 Absolute dating of recombination suppression events

As soon as a gene is permanently linked to the mating-type loci, its alleles will independently 318 accumulate substitutions, that will remain associated with the mating-type allele near which 319 they appeared. We therefore estimated the dates of recombination suppression for the 13 strata 320 (among the 17 used) for which we did not had estimated the date previously, by inferring the 321 322 divergence time between the alleles associated with the a_1 and a_2 mating types by molecular dating. The estimated times ranged 0.17 to 2.5 million years before present for the black strata 323 (Table S3). We estimated divergence times ranging from 3.82 (brown) to 0.15 (white) million 324 325 years before present (Table S3) in M. v. paradoxa strata and in the range of 0.47 to 0.1 million years for the *M*. *v*. *caroliniana* light-blue stratum. 326

327

328 Identification of optimal codons

To identify optimal codons in Microbotryum fungi, we investigated which codons were 329 preferentially used in highly expressed autosomal genes in the two distantly related species, M. 330 lychnidis-dioicae and M. intermedium (Figure 1). Selection for codon usage is expected to be 331 strongest on most highly expressed genes and most effective on autosomes because 332 recombination occurs. We had expression data for haploid cells cultivated clonally on nutritive 333 medium (Table S4), for the two mating types separately. To study autosomal codon usage, we 334 performed a within-group correspondence analysis (WCA; [70]), *i.e.*, a multivariate statistical 335 336 method summarizing codon count data, normalizing the count of each codon by the average 337 count of all the codons encoding for the same amino acid [71]. For the *M. intermedium* a₁ mating

type, the first principal component axis summarized 16.55% of the codon usage variation 338 (Figure S5) and was significantly negatively correlated with the gene expression level (Table 339 S5). Therefore, the genes with the lowest coordinates on the first principal component axis were 340 those with the codon usage most strongly shaped by selection, *i.e.*, with optimal codons. We 341 342 performed chi-square tests to compare the codon counts per amino acid of the genes with the 10% highest coordinates and the 10% lowest coordinates on the first axis. For each amino acid, 343 344 we inferred the optimal codon to be the one with the highest adjusted standardized residuals (*i.e.*, larger counts compared to expected under the hypothesis of random use of synonymous 345 codons) in highly expressed genes from chi-square tests (Table 1). We inferred the same 346 optimal codons when using the autosomes from the a_2 mating type of *M*. *intermedium* or from 347 either of the two *M. lychnidis-dioicae* mating types (Table S6). Because *M. intermedium* and 348 M. lychnidis-dioicae are very distantly related species in our dataset (last common ancestor at 349 350 ca. 20 million years before present, Figure 1), and have yet the same optimal codons, we assumed the optimal codons to be the same in all other *Microbotryum* species. The optimal 351 codons were more GC rich than random codons (Table 1). 352

353

354 Decrease in optimal codon odds in non-recombining regions

Using this set of optimal codons, we calculated the frequency of optimal codons per gene for 355 356 each haploid genome of the 13 Microbotryum species included in this study, on the autosomes and the different compartments of the mating-type chromosomes, i.e. the recombining pseudo-357 autosomal regions (PARs) and distinct evolutionary strata with different ages of recombination 358 suppression. Several evolutionary strata are small and contain few genes which impedes 359 statistical analysis of degeneration features. We therefore only considered the largest non-360 recombining regions, i.e., eight different black strata independently linking HD and PR, as well 361 as the light-blue, white, pink, light brown and brown strata extending recombination 362 suppression away from mating-type loci, and the strata corresponding to the PR-centromere 363 and HD-centromere linkage events in M. lagerheimii and M. saponariae. We tested whether 364 the genes in these non-recombining regions had lower frequencies of optimal codons compared 365 366 to autosomal genes. We found that optimal codons were enriched in GC in *Microbotryum*, thus a decrease in the frequency of optimal codons in the non-recombining regions compared to 367 recombining regions alternatively could be due to a reduction of GC-biased gene conversion in 368 non-recombining regions. However, GC-biased gene conversion should affect both coding and 369

370 non-coding sequences while lower efficacy of selection on codons should only affect coding 371 sequences (i.e. on exons but not on introns). In order to distinguish the effects of relaxed 372 selection and of GC-biased conversion on the frequency of optimal codons, we conducted an 373 analysis testing both the effect of the genomic compartment (i.e., autosomes, PAR and 374 evolutionary strata) and the effect of a footprint of GC-biased conversion on the frequency of 375 optimal codons on the mating-type chromosomes.

We performed a logistic regression for each genome to explain the variation in the frequency 376 of optimal codons by the genomic compartment, the difference in GC content between coding 377 sequences (GC^{exon}) and introns (GC^{intron}) as a measure of the relative effect of GC-biased gene 378 conversion and selection on codon usage, and their interactions (see Figure S6A and S6B for 379 raw data distribution). These factors allowed testing (i) whether a given genomic compartment 380 had a frequency of optimal codons significantly different from that in autosomes, (ii) the effect 381 of GC-biased gene conversion on the frequency of optimal codons in autosomes, and (iii) 382 whether this effect of GC-biased gene conversion was similar, stronger or weaker in non-383 384 recombining regions than in autosomes, by testing the significance of the interaction between the genomic compartment and the difference between GC^{exon} and GC^{intron}. We wanted to assess 385 whether the frequency of optimal codons was lower in non-recombining regions (tested through 386 387 the genomic compartment effect), and to assess whether this was due to a lower frequency of GC-biased conversion in non-recombining regions, which was tested through the interaction 388 389 between the genomic compartment and a measure of the relative effects of selection on codon usage and GC-biased conversion, i.e. the difference between GC^{exon} and GC^{intron} (Figure 3D). 390

We found that the recombining regions located on mating-type chromosomes, i.e., PARs or on 391 HD mating-type chromosome, had generally similar frequencies of optimal codons as 392 autosomes (p-values > 0.05; Table S7a). The frequencies of optimal codons were, however, 393 significantly lower in genes located in the recombining regions of the PR mating-type 394 chromosome in M. intermedium, M. lagerheimii and M. saponariae and of the PAR of M. 395 violaceum (s.s.), M. v. scopulorum, a₁ M. coronarie and a₂ M. silenes-acaulis (Table S7a). The 396 frequencies of optimal codons in the young pink, white, light blue, brown evolutionary strata 397 were not significantly different from autosomes in any species / mating type (p-value > 0.05; 398 Table S7a). 399

Across all species and mating types, the frequencies of optimal codons in black strata linking
HR and PR loci were in contrast significantly lower than in autosomes (p-values < 0.01; Table

S7a, Figure 3, Figure S6 et S7). The frequencies of optimal codons were higher than in 402 autosomes in the light brown in the a_2 but not in the a_1 genomes in *M*. *v*. paradoxa (Table S7a). 403 The frequencies of optimal codons in the non-recombining regions in the *M. lagerheimii* and 404 M. saponariae mating-type chromosomes (between the PR or HD loci and their respective 405 centromere) were lower than in autosomes (p-values < 0.05; Table S7a) except for the very 406 recently evolved non-recombining region in the a₂ HD M. lagerheimii mating-type 407 chromosome (Table S7a). The frequencies of optimal codons in the oldest non-recombining 408 evolutionary strata were therefore generally lower than in recombining regions. However, lower 409 frequencies of optimal codons in non-recombining regions could be due to a lower frequency 410 of recombination-mediated gene conversion because gene conversion is typically GC-biased 411 and optimal codons were found GC rich. To disentangle these effects, we compared GC^{exon} and 412 GC^{intron}, as the effect of low efficacy of selection on codon usage should act only on coding 413 sequences while GC-biased conversion should act on all sequences. We found that, when the 414 difference between GC^{exon} and GC^{intron} decreased, indicative of possible influence of GC-biased 415 conversion, the frequency of optimal codons significantly increased in autosomes for all the 416 genomes (p-values < 0.05; Table S7a) except the a_1 genomes of *M. lychnidis-dioicae*, *M. v.* 417 melanantha and M. violaceum (s.s.) for which the effect was not significant (Table S7a). This 418 suggests that GC-biased gene conversion increases the frequency of optimal codons in 419 autosomes and therefore that the generally lower frequency of optimal codons in non-420 recombining regions could be due to less frequent GC-biased gene conversion in these regions 421 compared to autosomes. 422

423 However, the interaction between the genomic compartment (particular recombining or nonrecombining regions) and the difference between GC^{exon} and GC^{intron} in predicting the 424 425 frequency of optimum codon usage was generally not significant (p-values > 0.05; Table S7; Figure 3C). This means that the effect of the difference between GC^{exon} and GC^{intron} in 426 predicting the frequency of optimal codons was not different between recombining and non-427 recombining regions. Therefore, the lower frequencies of optimal codons in the black strata 428 were not due to a lower GC-biased gene conversion in these regions (Figure 3D). The 429 interaction between the genomic compartment and the difference between GC^{exon} and GC^{intron} 430 was significant in only six non-recombining regions in five distinct haploid genomes out of 26. 431 The direction of the effect was the same for the black stratum of the a₂ genome of *M. lychnidis*-432 *dioicae*, *M. v. caroliniana* and *M. v. paradoxa* (p-values < 0.05; Table S7) but opposite for the 433 light brown stratum in the a_2 genome of *M*. *v*. *paradoxa* (p-value < 0.05; Table S7). We also 434

performed the same statistics using all evolutionary strata, even the shortest ones. While the
number of genes were too low to be confident in the results, it did not change the conclusions
we had for the largest ones (Figure S8, Table S7b).

438

439 Control for ancestral gene expression level

The lower frequency of optimal codons in non-recombining regions could also be due to a lack 440 of strong selection for a biased codon usage if gene expression in these regions was, by chance, 441 442 low in these large genomic regions even before recombination was suppressed. To investigate 443 whether genomic compartments differed in the average gene expression level before 444 recombination suppression, we retrieved the ancestral-like expression level of each gene taking as a proxy the expression level of its ortholog in *M. intermedium* (Figure S6), which has no 445 recombination suppression on its mating-type chromosomes, except in small regions around 446 the mating-type loci themselves. Of course, some gene-specific expression levels may have 447 448 changed in *M. intermedium* since the common ancestor, but, without recombination suppression in this species, there has likely been no major change in gene expression across large genomic 449 regions since the last common ancestor of the studied *Microbotryum* species. That gene 450 expression level changed little across the Microbotryum genus in recombining regions was 451 supported by significant correlations for all the 12 replicate combinations of gene expression 452 453 levels of shared autosomal genes measured in similar culture conditions between M. intermedium and M. lychnidis-dioicae (Table S8). 454

For each of the two haploid genomes across the 13 species, we compared the mean and variance 455 of the ancestral-like gene expression level (i.e. the gene expression level in *M. intermedium*) 456 457 between the current genomic compartments using a Kruskal-Wallis' test and a Levene's test, respectively. We thus analysed for each species whether there were differences in the ancestral-458 like gene expression between the current genomic compartments, *i.e.*, autosomes, PARs and 459 the various evolutionary strata that lack recombination. For none of genomes there was a 460 difference in the mean or variance of the ancestral-like gene expression level between the 461 current recombining and non-recombining regions (Table S9). Therefore, differences in 462 frequencies of optimal codons across the genomic compartments are unlikely to be explained 463 by differences in ancestral gene expression level between genomic compartments. 464

465

466 Tempo of degeneration in codon usage and in protein-coding genes

The multiple independent events of recombination suppression in the *Microbotryum* genus 467 provides an excellent case study for investigating the tempo of degeneration, as it allows 468 analysing the relationship between degeneration level and time using independent data points. 469 470 Here, we investigated the tempo of degeneration using the genes present in black strata, *i.e.*, the non-recombining regions linking the PR and HD loci, as well as the large species-specific strata, 471 as these represented independent events and large genomic regions. We did not consider the 472 strata that were shared by most species and therefore cannot provide independent data points 473 (purple, blue and orange strata), nor those that contained insufficient number of genes to 474 estimate properly the time since recombination suppression (e.g., the red and green strata in M. 475 lychnidis-dioicae). For the black strata shared by multiple species, we took the mean value of 476 variables for species descendent from the same recombination suppression event to avoid 477 pseudo-replication. 478

We sought to test whether the odds for a codon to be optimal and the non-synonymous versus synonymous substitution rates (d_N/d_S) ratio of the genes located in the independently evolved non-recombining regions are predicted by the time since recombination suppression. Using a logistic regression, we found that the odds for a codon to be optimal at the time when recombination suppression initially linked the PR and the HD loci *Microbotryum* was 0.725 (pvalue < 2e-16; Figure 4; Table S10), and that the odds significantly decreased, nearly linearly, by a factor of 0.989 every million year (p-value = 0.0503; Figure 4A and Table S10).

Using a linear regression, we found that the change in the d_N/d_S ratio was better explained by the log of the time since recombination suppression than by the raw time since recombination suppression ($R^2_{time} = 0.0016 \text{ vs } R^2_{log(time)} = 0.0096$; Table S10). The d_N/d_S ratio indeed increased rapidly after recombination suppression, but then in a decelerating manner until reaching a plateau, with its maximum value of 0.706 at 3.82 million years. It took only 9,400 years for the d_N/d_S ratio to reach 0.353 which corresponds to half of the maximum d_N/d_S value observed in our dataset.

493

494 **Discussion**

This study takes advantage of multiple events of recombination suppression on the mating-type chromosomes of closely related fungal species to document the tempo of genomic degeneration

in non-recombining regions (*i.e.*, the pace at which deleterious mutations accumulate with 497 time). We quantified the rate of degeneration by estimating, for each non-recombining region, 498 the time since recombination suppression and its relationship with measures of degeneration, 499 i.e., non-synonymous substitution accumulation and the reduction in the odds of optimal 500 501 codons. Few studies have focused on the degeneration of codon usage. This study further proposes a method to disentangle the effect of less efficient selection and of other factors on 502 503 codon usage in non-recombining regions. Indeed, a barrier in studying the consequences of relaxed selection on codon usage in non-recombining regions is that it may be confounded with 504 the effect of processes such as GC-biased gene conversion or ancestral expression level. We 505 found that genes in non-recombining regions had fewer optimal codons than in recombining 506 regions, and that this was due to a decrease in selection efficacy rather than lower GC-biased 507 gene conversion or lower ancestral expression level in non-recombining regions. Furthermore, 508 509 we found that the decrease in the frequency of optimal codon usage has been nearly linear during the last 2.5 million years of suppressed recombination on *Microbotryum* mating-type 510 chromosomes. The non-synonymous over synonymous substitution rate, in contrast, increased 511 initially rapidly, but then saturated, reaching a plateau. 512

Multiple independent events of recombination suppression involving mating-type loci were 513 514 reported in anther-smut fungi [60-62]. By sequencing and analysing the genomes of three additional *Microbotryum* species, we document three new cases of independent events of 515 516 mating-type chromosome rearrangements and fusions leading to PR-HD linkage, amounting to a total of eight such independent PR-HD linkage event. Linking mating-type loci is beneficial 517 under selfing [60, 72]. The high number of independent events of PR-HD linkage through 518 distinct genomic rearrangements shows the power of natural selection and the high degree of 519 evolution repeatability under strong selection, as well as the range of evolutionary paths that 520 521 can lead to the same phenotype. In addition, there have been independent recombination suppression events between centromeres and mating-type loci [62]. Recombination suppression 522 523 further extended at different times in a stepwise manner, forming evolutionary strata, at multiple instances in Microbotryum species and we document here additional evolutionary strata 524 unidentified so far. In total, we could use 17 genomic regions corresponding to independent 525 events of recombination suppression. The lack of availability of such multiple independent 526 events of recombination suppression across closely related species in other groups of organisms 527 had prevented so far empirical comparative studies of genomic degeneration. As the 528 529 *Microbotryum* genus contains more than a hundred species [73, 74], these results suggest that

there remains a rich resource of genomic diversity into the evolution of suppressed recombination in linkage to reproductive compatibility loci and the tempo of degeneration. The *Microbotryum* genus thus constitutes a unique model to provide insights on recombination suppression and the tempo of degeneration.

534 We document here decreased optimal codon usage as a degenerative feature of recombination suppression in multiple *Microbotryum* species. Using a logistic model, we found that optimal 535 codons were less frequently used in non-recombining regions compared to autosomes. Because 536 optimal codons in Microbotryum are GC-rich, any process inducing a lower GC content in non-537 recombining regions could be responsible for the difference in optimal codon frequencies 538 between non-recombining regions and autosomes. We therefore tested whether less frequent 539 GC-biased gene conversion in non-recombining regions could be the cause of the smaller 540 frequencies of optimal codons in non-recombining regions compared to autosomes. Decreased 541 GC-biased gene conversion and less efficient selection on codon usage impacts similarly coding 542 sequences, while the GC content of introns is only affected by GC-biased gene conversion, not 543 544 by selection on codon usage. Therefore, we used the difference in GC content between coding 545 sequences and introns to infer the relative impacts of GC-biased gene conversion and selection on codon usage, and we incorporated this measure in the logistic model explaining the variance 546 547 of frequency optimal codons between genomic compartments (i.e., autosomes, pseudoautosomal regions and non-recombining regions). We found lower odds of optimal codons in 548 549 non-recombining regions, and not only when GC-biased gene conversion was high (*i.e.*, at the lower differences in GC content between coding sequences and introns; Figure 3D), but also 550 551 when selection for optimal codons was high (*i.e.*, at the higher differences in GC content 552 between coding sequences and introns; Figure 3D). Thus, we identified less effective selection 553 as the cause of the decrease in the odds of optimal codons in non-recombining regions in Microbotryum mating-type chromosomes. To our knowledge, this study constitutes the first 554 attempt to analyse codon usage by controlling for confounding processes in statistical analyses. 555

We also checked that the differences in codon usage between autosomes and nonrecombining regions were not due to differences in gene expression levels between these regions before the evolution of recombination suppression, as gene expression level affects codon usage [40, 41]. Indeed, if expression levels were ancestrally lower for genes now located in non-recombining regions, selection on codon usage would have been weaker than in other autosomal regions, and smaller frequencies of optimal codons would have been expected, regardless of the effect of recombination suppression. However, we found no difference in the

ancestral expression levels between autosomal genes and those in current non-recombining regions. Altogether, we thus found robust evidence showing that relaxed selection due to recombination suppression resulted in less optimal codon usage in *Microbotryum* species, as suggested in the non-recombining mating-type chromosome of the fungus *Neurospora tetrasperma* [25].

It has often been assumed that only weak selection would act on codon usage [37, 75], 568 because synonymous substitutions should not greatly impact phenotypes. Contradicting this 569 view, several studies, focusing at the gene level, showed substantial impacts of synonymous 570 substitutions on phenotypes [37, 76, 77], sometimes leading to a strong selection on codon 571 usage [37]. In Drosophila species, codon usage was shown to be quite stable over long time 572 frames with 11/12 species having the same preferred codons except those coding for serine 573 [44]. In the Microbotryum genus, optimal codons were the same in M. intermedium and M. 574 lychnidis-dioicae, two species having their last common ancestor at the basis of the studied 575 Microbotryum clade, ca. 20 million years ago (Figure 1). The same optimal codons have 576 therefore been maintained for several millions years, suggesting strong and consistent selection. 577 The preferred codons in S. latifolia are almost identical to those in Arabidopsis thaliana, despite 578 their long divergence time [78]. In S. latifolia, no decrease in optimal codon frequencies was 579 found in non-recombining regions of the Y-chromosome [27]. In Drosophila in contrast, a faster 580 rate of accumulation of unpreferred synonymous mutations was higher for the neo-Y 581 chromosome than for the neo-X chromosome even for highly expressed genes [27]. The codon 582 usage in non-recombining regions therefore needs more studies to help draw generalities about 583 its evolution. 584

585 In contrast, degeneration of protein-coding sequences on non-recombining regions of sex chromosomes has been much more studied than codon usage and was consistently found in a 586 587 variety of organisms, from animals and fungi to plants (e.g., [16, 25, 65, 79-81]. However, very few datasets have so far allowed the study of the tempo of non-synonymous mutation 588 accumulation [82]. Theoretical models made predictions on the tempo of gene losses, but 589 empirical studies on the tempo of degeneration are lacking [51]. Tendencies were suggested 590 from reporting higher degeneration in older evolutionary strata, e.g., on sex chromosomes in 591 papaya [82]. However, only two independent recombination suppression events in that case are 592 insufficient to study the temporal dynamics of genomic degradation. 593

The existence of multiple independent events of recombination suppression in *Microbotryum* 594 fungi allowed estimating the tempo of degeneration in codon usage in protein-coding 595 sequences. Using a logistic regression, we inferred that the decrease in the frequency of optimal 596 codons was nearly linear in the range of time since recombination suppression found in 597 *Microbotryum*, at a rate of 0.989 per My. The non-synonymous over synonymous substitution 598 rate in contrast increased non-linearly, initially increasing rapidly with the time since 599 600 recombination suppression, and then in a decelerating manner until reaching a plateau. The early burst of accumulation of non-synonymous substitution relatively to synonymous 601 602 substitution is likely caused by the sudden decrease in the efficacy of selection due to recombination suppression. A substantial number of positions underwent non-synonymous 603 change, which induced a saturation of the d_N/d_S ratio, with its maximum value of 0.706 at 3.82 604 million years. This is the first study to our knowledge that disentangles the effects of reduced 605 606 selection efficacy and GC-biased gene conversion on the frequency of optimal codons and that investigates the tempo of degeneration using multiple independents events of recombination 607 suppression. 608

609 Understanding the tempo of degeneration is important for our knowledge on genomic evolution, on the origin of genetic diseases, and on the maintenance of regions without 610 611 recombination. For example, it is important to note that the seven youngest evolutionary strata used in the study of the degeneration tempo are still collinear between mating types [60-62] and 612 613 correspond to low levels of degeneration. In contrast, the evolutionary strata with strong signs of degeneration are also highly rearranged between mating types [60-62]. This implies in 614 particular that the hypothesis postulating that inversions linked to a mating-type locus or a sex-615 determining regions could easily reverse when they accumulate too much load [83] likely does 616 not hold, as the region would then have already accumulated further rearrangements, preventing 617 618 reversion to recombination. This would in turn supports the hypothesis that the extension of recombination suppression could then simply be the result of the selection of recombination 619 620 cessation for sheltering deleterious alleles [84].

621

622 Data access

The dataset(s) supporting the conclusions of this article is(are) available in the [repository name]
repository, [unique persistent identifier and hyperlink to dataset(s) in http:// format]-will be added upon
manuscript acceptance.

626

627 Material and methods

628 Genome sequencing

We sequenced haploid genomes of opposite mating types for one diploid strain of each of the 629 following species: M. v. tatarinowii parasitizing S. tatarinowii (#1400, GPS 39°59'45.4" 630 116°11'37.3", Xiangshan, Beijing, collected in September 2015), M. v. melanantha parasitizing 631 S. melanantha (#1296; 27°01'50.6" 100°10'41.6" First Peak, Lijiang, date collected in July 632 2014), and M. coronariae parasitizing L. flos cuculi (#1247; GPS 55.7 -4.9, Great Cumbrae 633 Island, UK, collected in 2000). Additionally, we sequenced the a₁ genome of *M. intermedium* 634 parasitizing Salvia pratensis (#1389, GPS 44.3353 N 7.13637 E, Cuneo, Italy, collected in July 635 636 2011). Samples were collected before the publication of laws regarding the Nagoya protocol in the countries of collection. DNA extraction and sequencing based on Pacific Bioscience long-637 read sequencing was performed as described previously [60, 61]. Sequencing was carried out at 638 UCSD IGM Genomics Center (San Diego, USA) with the P6 C4 chemistry. Coverage was 639 between 24x and 35x for all genomes, except for *M. intermedium* for which it was 199x. 640

641

Assembly of new genomes and assembly improvement for *M. scabiosae* infecting *Knautia arvensis* (initially published in Branco et al 2018).

We converted the bax.h5 files from the each smart cell into one bam file using bax2bam with 644 the "--subread" option (https://github.com/PacificBiosciences/bax2bam), and all bam files from 645 646 the same sequencing into fastq file using bam2fastx genome one (https://github.com/PacificBiosciences/bam2fastx). We generated the assembly using canu [85] 647 with the parameters "genomeSize=30m" and "-pacbio-raw". We used pbalign (version 0.3.0) 648 with the blasr algorithm [86] to realign the raw reads onto the assembly (indexed with samtools 649 650 faidx [75] and then used the output bam file into quiver [87] to polish the assembly basecalling (https://github.com/PacificBiosciences/GenomicConsensus). Default parameters were used 651 when not detailed in the text (see the Table S1 for further assembly statistics). 652

The previous genome assemblies of the a_1 and a_2 genome of *M. scabiosae* were based on five PacBio movies. We re-assembled the genome using three additional movies, from the same strain, and generated using the same PacBio technology for a total coverage of 506x for the a_1 genome and 803x for the a_2 genome. The assemblies were substantially improved, as the contig

numbers were reduced from 123 and 147 to 26 and 20 for the a_1 and a_2 mating type, respectively (see the Table S2 for further assembly statistics).

659

660 Gene models, orthologous group reconstruction, transposable elements, centromeres

As for the previously published *Microbotryum* genomes [31, 60, 61], the protein-coding gene 661 662 models were predicted with EuGene (version 4.2a; [88], trained for *Microbotryum*. Similarities to the fungal subset of the uniprot database [89] plus the M. lychnidis-dioicae Lamole proteome 663 [31] were integrated into EuGene for the prediction of gene models. We obtained the 664 orthologous groups based on all-vs-all blastp high scoring pairs parsed with orthAgogue -u 665 option [90] and clustered with mcl [91] setting to 1.5 the inflation parameter. Transposable 666 667 elements were obtained from a previous study [92]. Centromeres were identified as regions with high density of the centromeric repeats described in *M. lychnidis-dioicae* [31]. 668

669

670 Evolutionary scenarios of the mating-type chromosomes

We represented genomic data using Circos [93] with the following tracks: (i) contig, (ii) dS=0 671 (indicating recombination in these selfing species as selfing leads to homozygosity), (iii) all 672 genes, (iv) transposable elements, centromeric repeats from *M. lychnidis-dioicae* [31]. Both the 673 dS=0 and all gene tracks were filtered for TEs and centromeric repeats, using their merged 674 675 coordinates with bedtools [94]. Homozygosity (dS=0) was used to infer recombining regions on mating-type chromosomes following previous studies [60, 61]. The comparisons of mating-676 type chromosomes of the species with linked HD and PR loci to the ancestral state with HD 677 and PR loci on different chromosomes (using the *M. intermedium* genome as a proxy for the 678 ancestral state) allowed to infer the chromosomal rearrangements having linked HD and PR 679 loci, as previously done [60, 61]. Evolutionary strata were identified as genomic regions with 680 681 different levels of differentiation between mating-type chromosomes in the ancestral gene order 682 as previously done [60, 61].

683

684 RNAseq experiment and expression level quantification.

We generated RNAseq data for *M. intermedium*. Total RNA was isolated from haploid cells
using the Qiagen RNeasy Mini Kit. Haploid strains (a₁ or a₂) were streaked on Potato Dextrose

Agar (PDA) and grown for 48 hours at 22°C. Cells were scraped, ground in liquid nitrogen, 687 and total RNA extracted following the manufacturer's protocol. For RNAseq analysis, equal 688 amounts of total RNA individually isolated from a₁ and a₂ cultures were pooled. For RNA 689 isolation from mated pairs, haploid a1 and a2 cultures were first grown separately in yeast extract 690 peptone dextrose broth overnight. Cell density was measured spectrophotometrically. The 691 concentration of each culture was adjusted to O.D.₆₀₀ 1.0. Equal volumes of the two haploid 692 cultures were mixed and plated on water agar. These plates were incubated for four days at 693 14°C, after which wet mounts were prepared from each mated plate to verify the presence of 694 conjugation tubes, indicating active mating behaviour. Cells were scraped and total RNA 695 harvested as previously described. After total RNA isolation, several quality control measures 696 697 were taken. Concentration and purity were assessed using a NanoDrop 2000 spectrophotometer (Thermo Scientific); 260/280 and 260/230 ratios greater than 1.8 were considered satisfactory 698 699 for RNAseq application. Additionally, cDNA was prepared and used as template for intronspanning primers in PCR reactions to verify the lack of genomic DNA contamination. 700 701 Bioanalyzer analysis was completed to detect intact 18S and 23S RNA as a measure of overall RNA quality. After passing all quality control measurements, RNA was sent for RNAseq 702 analysis to CD Genomics (Shirley, New York). 703

704 For M. lychnidis-dioicae, we used RNAseq data published previously [95]. List and statistics of the RNAseq experiments used in this study are provided in Table S4 We controlled the quality 705 706 of the RNAseq experiment data using fastQC (https://github.com/s-andrews/FastQC). We 707 trimmed the sequences using trimmomatic (Bolger et 2014); al. parameters: ILLUMINACLIP:TruSeq3-PE.fa:2:30:10:2:keepBothReads LEADING:3 TRAILING:3 708 MINLEN:36). We only considered the resulting trimmed paired-end reads for further analysis 709 (Table S3). For each RNAseq dataset, we pseudo-aligned the reads onto the corresponding 710 711 references and quantified the gene expression levels using Kallisto using the "kallisto quant" command with the "--bias" option [96]. For the diploid RNAseq experiment (Table S4, we used 712 the concatenated coding sequence set from a_1 and a_2 haploid genomes as reference. For haploid 713 RNAseq experiment, we considered the haploid coding sequence set from the corresponding 714 mating type. For each RNAseq experiment, we removed all the genes with an expression level 715 above the 95th percentile of the expression level distribution to avoid spurious high expression 716 level. 717

718

719 Identifying optimal codons in *Microbotryum* fungi

We generated a codon usage table for each of the a₁ M. lychnidis-dioicae, a₂ M. lychnidis-720 721 dioicae, a₁ M. intermedium and a₂ M. intermedium haploid genomes (in-house awk script) and filtered the genes that overlapped transposable elements, that were previously identified [92]. 722 723 We then performed a within-group correspondence analysis (WCA), following the procedure available at http://pbil.univ-lyon1.fr/datasets/charif04/ [70] on each codon usage table. 724 Correspondence analysis is a multivariate statistical method that summarizes codon count data 725 by reducing them to a limited number of variables. WCA more specifically allows to dissociate 726 the effects of different amino-acid compositions from the effects directly related to synonymous 727 728 codon usage [70]. This method adjusts the value for each codon by the average value of all the codons encoding for the same amino acid. We only considered the two first principal 729 730 components which explained between 16.30% and 16.50% of the variance for the first principal component and between 7.75% and 8.51% for the second principal component (Figure S5). The 731 732 coordinates of the genes projected onto the first (PC1) and second (PC2) axes were significantly correlated with gene expression level (Pearson's correlation tests; Table S5. We considered the 733 first axis to better characterized the gene expression level variation because the R² calculated 734 between PC1 gene coordinates and gene expression level were higher by two orders of 735 magnitude compared to the R² calculated between PC2 gene coordinates and gene expression 736 level (see correlation's coefficient value in Table S05). 737

We performed chi-square tests to compare, per amino acid, the codon counts of the 10% genes having the lowest PC1 coordinates to the codon counts of the 10% genes having the highest PC1 coordinates, representing the genes with highest and lowest expression, respectively (Figure S3). When the chi-square test was significant (*p-value* < 0.05), we considered for each amino acid the optimal codon to be the one with the highest residual from the chi-square test (Table S6). We did not consider the ATG (methionine) and TGG (tryptophan) codons as they do not have synonymous codons.

745

746 Frequency of optimal codons and GC content statistics

For each coding sequence with base pair numbers multiple of three, we calculated the frequency

of optimal codons (FOP), the GC content on the third position (%GC₃) of each codon and the

- 749 GC content of the introns of each gene (%GC_{introns}). We removed from this gene set any gene
- 750 overlapping with a TE. While parsing the coding sequence fasta files for the frequency of

optimal codons calculation, we did not consider the ATG (methionine) or TGG (tryptophan)codons since they do not have synonymous codons.

753

754 Non-synonymous and synonymous divergence estimation

For each pair of allele from alternative mating type in each species, we aligned the protein sequences with muscle v3.8.31 [97] and obtained the codon-based CDS alignments with TranslatorX; dN and dS values were computed in the yn00 program [98].

758

759 Absolute dating of recombination suppressions

We estimated the species and genome compartment divergence times under a calibrated Yule 760 model and a HKY substitution model (kappa=2, estimated frequencies) with 10,000,000 or 761 762 20,000,000 mcmc generations in beast v2.5.0 [99]. We used three unlinked site models (one per codon position). Clock model and tree parameters were kept linked. We used the split of M. 763 *lychnidis-dioicae* and *M. silenes-dioicae* as a calibration point, with a normal prior with a mean 764 0.42 MYA and sigma 0.04 [100]. Beast input files are available as supplementary material. 765 Concatenated alignments were obtained from codon-based alignments produced by macse 766 767 v2.03 -prog alignSequences [101] protocol on fully conserved single copy genes. We reconstructed the species and genome compartment maximum likelihood trees with IQTree 768 v2.0.4 (1000 rapid bootstraps and automatic model selection [102]). 769

770 For the species divergence we used 3,955 fully conserved single copy autosomal genes (2,732,071 aligned codons). For estimating the dates of recombination suppression, we 771 estimated the divergence time at the genes linked to mating-type loci between alleles associated 772 with a₁ and a₂ mating types in ML trees. For the black strata in species with HD-PR linkage, 773 we used a conserved set of 32 genes (28,402 aligned codons) consistently present in the non-774 recombining region between the PR and HD loci in all species but in M. silenes-acaulis. For 775 the black stratum of *M. silenes-acaulis* we used 59 genes (44,055 aligned codons) present in its 776 specific black stratum. For the strata of M. v. paradoxa we used 12 genes (9,165 aligned 777 codons), 21 genes (13,246 aligned codons), 92 genes (78,627 aligned codons) and 179 genes 778 (129,122 aligned codons) for the brown, pink, light-brown and white strata, respectively. We 779 estimated the recombination suppression in the light-blue stratum of M. v. caroliniana based 780 on the estimates of the divergence between the a_1 versus a_2 associated alleles in M. v. 781

caroliniana in the trees used for estimating the ages of the black stratum of *M. silenes-acaulis*

783 (21 genes in common) and the pink stratum of *M. v. paradoxa* (10 genes in common). Absolute

784 dates for the non-recombining regions in *M. lagerheimii* and *M. saponariae* were taken from a

previous study [62]. Gene family alignments are available as supplementary material.

786

787 Statistics

We tested in each of the 26 Microbotryum genomes (one genome per mating type for 13 788 species) whether the genes in distinct genomic compartments (PARs and the different 789 790 evolutionary strata with different ages of recombination suppression) had lower frequencies of optimal codons compared to autosomal genes. To take into account the effect of any process 791 other than genomic degeneration that could lower the GC level in the non-recombining regions 792 793 such as less frequent GC-biased gene conversion in the non-recombining regions, we included in the analysis the difference in GC content between coding and intronic sequences across 794 795 genes.

796 We therefore performed a logistic regression to assess whether the variation in the frequency 797 of optimal codons could be explained by the genomic compartment, the difference in GC content between coding and intronic sequences across genes and/or their interaction. Using the 798 glm function in R, we gave as input the frequency of optimal codons weighted by the total 799 number of codons per gene (using the "weight" option in the glm function). After performing 800 a logistic regression assuming binomial distribution and a logit link, we noticed that predicted 801 frequencies of optimal codons were all lower than the raw frequency of optimal codons which 802 could indicate overdispersion issues. We therefore performed the logistic regression assuming 803 a quasibinomial distribution, the only difference being in the estimation of the dispersion 804 parameter to correct, among other things, the p-values. The estimated dispersion parameter was 805 high for the logistic regression model of each genome ($\Phi > 2$), so we choose the quasibinomial 806 distribution to interpret the data. Choosing a quasibinomial distribution did not change however 807 808 the log(odds) estimated by the model. We tested whether the relationship between the frequency of optimal codons and the time since recombination suppression was better explained with a 809 810 non-linear component using a non-linear model (polynomial with two degrees) than with only 811 a linear component (Table S10).

We therefore analysed this relationship through a logistic regression and used this model to estimate the frequency of optimal codons using a sequence of time value and calculated 95% confidence interval using the linkinv function.

We built a linear model to analyse the relationship between d_N/d_S and the time since recombination suppression. As d_N/d_S seemed to saturate at high times, we also tested whether a linear regression with the log of the time since recombination suppression explained better the relationship.

We inquired how much expression level changed for genes in recombining region across the *Microbotryum* genus. To do so, we performed Pearson's correlation tests between *M. intermedium* and *M. lychnidis-dioicae* orthologous genes for all combinations of RNA-seq experiment replicates that were comparable between the two species (*i.e.* cells grown on rich media). This led to 12 correlation tests summarized in Table S8.

824

825 Author contribution

826 Conception and design, TG, MEH, FC; Formal analysis, FC, RCRV; Interpretation, FC, TG,

827 MEH, RCRV; Data acquisition, MHP, RMW; Original draft, FC, TG; Final draft, FC, TG,

828 MEH, RCRV; Revision, FC, TG, RCRV, MEH, MHP, RMW. All authors read and approved 829 the final version.

830

831 Competing interests

832 None of the authors have any competing interests.

833

834 Acknowledgements

We thank Hui Tang and Janis Antonovics for help with sample collection. We thank Jacqui Shykoff, Maxime Dubart and Sylvain Billiard for great help and discussion about statistical analyses. This work was supported by the National Institute of Health (NIH) grant R15GM119092 to M. E. H., and the Louis D. Foundation award and EvolSexChrom ERC advanced grant #832352 to T. G. We thank Hector Mendoza for help with RNA isolation. We thank Kurt Hasselman, John Bain and Hui Tang who collected the strains (involving a confrontation with a grizzly bear for John Bain). Support for R.M.W. and RNA isolation was

- additionally from NIH [sub-award #OGMB131493D1] to M.H.P. from [P20GM103436] to
- 843 [Nigel Cooper, PI] and also NSF/IRES Award (1824851) to M.H.P. The contents of this work
- are solely the responsibility of the authors and do not represent the official views of the NIH.

845

846 Cited literature

847

848	1.	Fisher RA: The genetical theory of natural selection. 1930.
849	2.	Muller HJ: Some genetic aspects of sex. The American Naturalist 1932, 66:118-138.
850	3.	Bernstein H, Byers GS, Michod RE: Evolution of sexual reproduction: importance of DNA
851		repair, complementation, and variation. Am Nat 1981, 117:537-549.
852	4.	Otto SP, Lenormand T: Resolving the paradox of sex and recombination. Nature Reviews
853		Genetics 2002, 3: 252-261.
854	5.	Hill WG, Robertson A: The effect of linkage on limits to artificial selection. Genetics Research
855		1966, 8: 269-294.
856	6.	Maynard Smith J, Haigh J: The hitch-hiking effect of a favourable gene. Genet Res 1974, 23:22-
857		35.
858	7.	Gutiérrez-Valencia J, Hughes PW, Berdan EL, Slotte T: The genomic architecture and
859		evolutionary fates of supergenes. Genome biology and evolution 2021, 13.
860	8.	Schwander T, Libbrecht R, Keller L: Supergenes and complex phenotypes. Current Biology
861		2014, 24: R288-R294.
862	9.	Nadeau NJ: Genes controlling mimetic colour pattern variation in butterflies. Current Opinion
863		in Insect Science 2016, 17: 24-31.
864	10.	Saenko VS, Chouteau M, Prunier FP, Blugeon C, Joron M: Unravelling the genes forming the
865		wing pattern supergene in the polymorphic butterfly Heliconius numata. EvoDevo 2019,
866		10: 16.
867	11.	Wang J, Wurm Y, Nipitwattanaphon M, Riba-grognuz O, Huang Y-c, Shoemaker D, Keller L: A
868		Y-like social chromosome causes alternative colony organization in fire ants. Nature 2013,
869		493: 664.
870	12.	Charlesworth B, Charlesworth D: A model for the evolution of dioecy and gynodioecy. The
871		American Naturalist 1978, 112: 975-997.
872	13.	Hartmann FE, Duhamel M, Carpentier F, Hood ME, Foulongne-Oriol M, Silar P, Malagnac F,
873		Grognet P, Giraud T: Recombination suppression and evolutionary strata around mating-type
874		loci in fungi: documenting patterns and understanding evolutionary and mechanistic causes.
875		New Phytologist 2021, 229: 2470-2491.
876	14.	Ohno S: Sex chromosomes and sex-linked genes. Springer-V edn; 1967.
877	15.	Westergaard M: The mechanism of sex determination in dioecious flowering plants. Adv
878		Genet 1958, 9: 217-281.
879	16.	Bachtrog D: Y-chromosome evolution : emerging insights into processes of Y-chromosome
880		degeneration. Nature Reviews Genetics 2013, 14:113-124.
881	17.	Charlesworth B, Charlesworth D: The degeneration of Y chromosomes. Philosophical
882		Transactions of the Royal Society B-Biological Sciences 2000, 355: 1563-1572.
883	18.	Jay P, Chouteau M, Whibley A, Bastide H, Parrinello H, Llaurens V, Joron M: Mutation load at
884		a mimicry supergene sheds new light on the evolution of inversion polymorphisms. Nature
885		Genetics 2021, 53 :288-293.

Stolle E, Pracana R, Howard P, Paris CI, Brown SJ, Castillo-Carrillo C, Rossiter SJ, Wurm Y:
 Degenerative expansion of a young supergene. *Molecular Biology and Evolution* 2019,
 36:553-561.

889 20. Berlin S, Ellegren H: Fast accumulation of nonsynonymous mutations on the female-specific
890 W chromosome in birds. J Mol Evol 2006, 62:66-72.

- Brion C, Caradec C, Pflieger D, Friedrich A, Schacherer J: Pervasive phenotypic impact of a
 large non-recombining introgressed region in yeast. *Molecular Biology and Evolution* 2020,
 37:2520-2530.
- Hough J, Hollister JD, Wang W, Barrett SCH, Wright SI: Genetic degeneration of old and young
 y chromosomes in the flowering plant Rumex hastatulus. Proceedings of the National
 Academy of Sciences of the United States of America 2014, 111:7713-7718.
- 897 23. Nicolas M, Marais GAB, Hykelova V, Janousek B, Laporte V, Vyskot B, Mouchiroud D, Negrutiu
 898 I, Charlesworth D, Monéger F: A gradual process of recombination restriction in the
 899 evolutionary history of the sex chromosomes in dioecious plants. *PLoS Biology* 2004, 3:e4.
- Papadopulos AST, Chester M, Ridout K, Filatov DA: Rapid Y degeneration and dosage
 compensation in plant sex chromosomes. Proceedings of the National Academy of Sciences
 2015, 112:201508454.
- Whittle CA, Sun Y, Johannesson H: Degeneration in codon usage within the region of
 suppressed recombination in the mating-type chromosomes of *Neurospora tetrasperma*.
 Eukaryotic Cell 2011, **10**:594-603.
- 90626.Bachtrog D: Sex chromosome evolution: Molecular aspects of Y-chromosome degeneration907in Drosophila. Genome Research 2005, 15:1393-1401.
- 90827.Bartolome C, Charlesworth B: Evolution of amino-acid sequences and codon usage on the909Drosophila miranda neo-sex chromosomes. Genetics 2006, 174:2033-2044.
- Steinemann M, Steinemann S: Degenerating Y chromosome of Drosophila miranda: A trap
 for retrotransposons (chromosome structure/larval cuticle protein genes). Genetics 1992,
 89:7591-7595.
- 913 29. Bianchi NO: Y chromosome structural and functional changes in human malignant diseases.
 914 Mutation research 2009, 682:21-27.
- So. Lee Y, Kim C, Park Y, Pyun J-A, Kwack K: Next generation sequencing identifies abnormal Y
 chromosome and candidate causal variants in premature ovarian failure patients. *Genomics* 2016, 108:209-215.
- Badouin H, Hood ME, Gouzy J, Aguileta G, Siguenza S, Perlin MH, Cuomo CA, Fairhead C, Branca
 A, Giraud T: Chaos of rearrangements in the mating-type chromosomes of the anther-smut
 fungus *Microbotryum lychnidis-dioicae*. *Genetics* 2015, 200:1275-1284.
- Yan H, Jin W, Nagaki K, Tian S, Ouyang S, Buell CR, Talbert PB, Henikoff S, Jiang J: Transcription
 and histone modifications in the recombination-free region spanning a rice centromere.
 Plant Cell 2005, 17:3227-3238.
- 92433.Hollister JD, Gaut BS: Epigenetic silencing of transposable elements: A trade-off between925reduced transposition and deleterious effects on neighboring gene expression. Genome926Research 2009, 19:1419-1428.
- Li SF, Zhang GJ, Yuan JH, Deng CL, Gao WJ: Repetitive sequences and epigenetic modification:
 inseparable partners play important roles in the evolution of plant sex chromosomes. *Planta* 2016, 243:1083-1095.
- 93035.Duret L: Evolution of synonymous codon usage in metazoans. Current Opinion in Genetics and931Development 2002, 12:640-649.
- Sharp PM, Averof M, Lloyd AT, Matassi G, Peden JF: DNA sequence evolution: the sounds of
 silence. Philosophical transactions of the Royal Society of London Series B, Biological sciences
 1995, 349:241-247.
- 935 37. Machado HE, Lawrie DS, Petrov DA: Pervasive strong selection at the level of codon usage
 936 bias in Drosophila melanogaster. Genetics 2020:genetics.302542.302019.

937 38. Ikemura T: Correlation between the abundance of *Escherichia coli* transfer RNAs and the
 938 occurrence of the respective codons in its protein genes: A proposal for a synonymous codon
 939 choice that is optimal for the *E. coli* translational system. *Journal of Molecular Biology* 1981,
 940 151:389-409.

- 941 39. Post LE, Strycharz GD, Nomura M, Lewis H, Dennis PP: Nucleotide sequence of the ribosomal
 942 protein gene cluster adjacent to the gene for RNA polymerase subunit β in Escherichia coli.
 943 Proceedings of the National Academy of Sciences of the United States of America 1979,
 944 76:1697-1701.
- 945 40. Novoa EM, Ribas de Pouplana L: Speeding with control: Codon usage, tRNAs, and ribosomes.
 946 Trends in Genetics 2012, 28:574-581.
- 94741.Zhou Z, Dang Y, Zhou M, Li L, Yu C-h, Fu J, Chen S, Liu Y: Codon usage is an important948determinant of gene expression levels largely through its effects on transcription. 2016:9.

949 42. Kliman RM, Hey J: Reduced natural selection associated with low recombination in
 950 Drosophila melanogaster. Mol Biol Evol 1993, 10:1239-1258.

- 95143.Betancourt AJ, Welch JJ: Report reduced effectiveness of selection caused by a lack of952recombination. Current Biology 2009, 19:655-660.
- 44. Vicario S, Moriyama EN, Powell JR: Codon usage in twelve species of Drosophila. BMC
 Evolutionary Biology 2007, 7:1-17.
- 95545.Duret L, Galtier N: Biased gene conversion and the evolution of mammalian genomic956landscapes. Annual Review of Genomics and Human Genetics 2009, 10:285-311.

46. Lassalle F, Périan S, Bataillon T, Nesme X: GC-content evolution in bacterial genomes : the
biased gene conversion hypothesis expands. *PLoS Genet* 2015, 11:e1004941.

- 95947.Marais G: Biased gene conversion: implications for genome and sex evolution. TRENDS in960Genetics 2003, 19:330-338.
- 96148.Pessia E, Popa A, Mousset S, Rezvoy C, Duret L, Marais GAB: Evidence for widespread GC-
biased gene conversion in eukaryotes. Genome Biol Evol 2012, 4:675-682.
- Weber CC, Boussau B, Romiguier J, Jarvis ED, Ellegren H: Evidence for GC-biased gene
 conversion as a driver of between-lineage differences in avian base composition. Genome
 biology 2014, 15:549.
- So. Kostka D, Hubisz MJ, Siepel A, Pollard KS: The role of GC-biased gene conversion in shaping
 the fastest evolving regions of the human genome. *Molecular Biology and Evolution* 2012,
 29:1047-1057.
- 969 51. Bachtrog D: The temporal dynamics of processes underlying Y chromosome degeneration.
 970 *Genetics* 2008, **179**:1513-1525.
- 97152.Lahn BT, Page DC: Four evolutionary strata on the human X chromosome. Science 1999,972286:964-967.
- 973 53. Krasovec M, Chester M, Ridout K, Filatov DA: The mutation rate and the age of the sex
 974 chromosomes in *Silene latifolia*. *Current Biology* 2018, 28:1832-1838.e1834.
- 975 54. Bellott DW, Hughes JF, Skaletsky H, Brown LG, Pyntikova T, Cho TJ, Koutseva N, Zaghlul S,
 976 Graves T, Rock S, et al: Mammalian y chromosomes retain widely expressed dosage-sensitive
 977 regulators. Nature 2014, 508:494-499.
- 97855.Hughes JF, Rozen S: Genomics and genetics of human and primate Y chromosomes. Annual979Review of Genomics and Human Genetics 2012, 13:83-108.
- 980 56. Mrackova M, Nicolas M, Hobza R, Negrutiu I, Monéger F, Widmer A, Vyskot B, Janousek B:
 981 Independent origin of sex chromosomes in two species of the genus *Silene*. *Genetics* 2008,
 982 179:1129-1133.
- 983 57. Ma W-J, Veltsos P: The diversity and evolution of sex chromosomes in frogs. *Genes* 2021,
 984 12:483.
- 985 58. Ross MT, Grafham VD, Coffey AJ, Scherer S, McLay K, Muzny D, Platzer M, Howell GR, Burrows
 986 C, P C, et al: UKPMC Funders Group The DNA sequence of the human X chromosome. *Genome*987 2009, 434:325-337.

Skaletsky H, Kuroda-Kawaguchl T, Minx PJ, Cordum HS, Hlllier LD, Brown LG, ReppIng S,
Pyntikova T, All J, Blerl T, et al: The male-specific region of the human Y chromosome is a
mosaic of discrete sequence classes. *Nature* 2003, 423:825-837.

- Branco S, Badouin H, Rodríguez de la Vega RC, Gouzy J, Carpentier F, Aguileta G, Siguenza S,
 Brandenburg J-T, Coelho MA, Hood ME, Giraud T: Evolutionary strata on young mating-type
 chromosomes despite the lack of sexual antagonism. *Proceedings of the National Academy*of Sciences 2017, 114:7067-7072.
- 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: Multiple convergent supergene evolution
 events in mating-type chromosomes. Nature Communications 2018, 9:2000.
- 62. Carpentier F, Rodríguez RC, Vega D, Branco S, Snirc A, Coelho MA, Hood ME, Giraud T,
 999 Evolution ES, Paris-saclay U, et al: Convergent recombination cessation between mating-type
 1000 genes and centromeres in selfing anther-smut fungi. *Genome Research* 2019, 29:944-953.
- 100163.Hood ME, Antonovics J: Intratetrad mating, heterozygosity, and the maintenance of1002deleterious alleles in Microbotryum violaceum (= Ustilago violacea). Heredity 2000, 85:231-1003241.
- 100464.Coelho MA, Bakkeren G, Sun S, Hood ME, Giraud T: Fungal Sex: The Basidiomycota.1005Microbiology Spectrum 2017, 5:1-30.
- Fontanillas E, Hood ME, Badouin H, Petit E, Barbe V, Gouzy J, De Vienne DM, Aguileta G, Poulain
 J, Wincker P, et al: Degeneration of the nonrecombining regions in the mating-type
 chromosomes of the anther-smut fungi. *Molecular Biology and Evolution* 2015, **32**:928-943.
- 100966.Ma W-J, Carpentier F, Giraud T, Hood M: Differential gene expression between fungal mating1010types is associated with sequence degeneration. Genome Biology and Evolution 2020.
- 1011 67. Corcoran P, Anderson JL, Jacobson DJ, Sun Y, Ni P, Lascoux M, Johannesson H: Introgression
 1012 maintains the genetic integrity of the mating-type determining chromosome of the fungus
 1013 *Neurospora tetrasperma. Genome Research* 2016, **26:**486-498.
- 1014 68. Hartmann FE, Rodríguez de la Vega RC, Gladieux P, Ma W-J, Hood ME, Giraud T: Higher gene
 1015 flow in sex-related chromosomes than in autosomes during fungal divergence. *Molecular*1016 *Biology and Evolution* 2020, 37:668-682.
- Sun Y, Corcoran P, Menkis A, Whittle CA, Andersson SGE, Johannesson H: Large-scale
 Introgression shapes the evolution of the mating-type chromosomes of the filamentous
 Ascomycete Neurospora tetrasperma. PLoS Genetics 2012, 8:e1002820.
- 102070.Charif D, Thioulouse J, Lobry JR, Perrière G: Online synonymous codon usage analyses with1021the ade4 and seqinR packages. Bioinformatics 2005, 21:545-547.
- 102271.Suzuki H, Brown CJ, Forney LJ, Top EM: Comparison of correspondence analysis methods for1023synonymous codon usage in bacteria. DNA Research 2008, 15:357-365.
- 1024 72. Nieuwenhuis BPS, Billiard S, Vuilleumier S, Petit E, Hood ME, Giraud T: Evolution of uni- and
 1025 bifactorial sexual compatibility systems in fungi. *Heredity* 2013, 111:445-455.
- 1026 73. Hood ME, Mena-Alí JI, Gibson AK, Oxelman B, Giraud T, Yockteng R, Arroyo MTK, Conti F,
 1027 Pedersen AB, Gladieux P, Antonovics J: Distribution of the anther-smut pathogen n on species
 1028 of the Caryophyllaceae. New Phytologist 2010, 187:217-229.
- 102974.Lutz M, Piatek M, Kemler M, Chlebicki A, Oberwinkler F: Anther smuts of Caryophyllaceae:1030molecular analyses reveal further new species. Mycological research 2008, 112:1280-1296.
- 1031 75. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R,
 1032 Genome Project Data Processing Subgroup, et al: The Sequence alignment/map (SAM) format
 1033 and SAMtools. *Bioinformatics* 2009, 25:2078-2079.
- 103476.Lampson BL, Pershing NLK, Prinz JA, Lacsina JR, Marzluff WF, Nicchitta CV, MacAlpine DM,1035Counter CM: Rare codons regulate KRas oncogenesis. Current Biology 2013, 23:70-75.

103677.Carlini DB: Experimental reduction of codon bias in the Drosophila alcohol dehydrogenase1037gene results in decreased ethanol tolerance of adult flies. Journal of Evolutionary Biology10382004, 17:779-785.

- 1039 78. Qiu S, Bergero R, Zeng K, Charlesworth D: Patterns of codon usage bias in *Silene latifolia*.
 1040 *Molecular Biology and Evolution* 2011, 28:771-780.
- 1041 79. Chibalina MV, Filatov DA: Plant y chromosome degeneration is retarded by haploid purifying
 1042 selection. Current Biology 2011, 21:1475-1479.

104380.Bergero R, Charlesworth D: Preservation of the Y transcriptome in a 10-Million-year-old plant1044sex chromosome system. Current Biology 2011, 21:1470-1474.

- 1045 81. Soojin Y, Charlesworth B: Contrasting patterns of molecular evolution of the genes on the
 1046 new and old sex chromosomes of *Drosophila miranda*. *Molecular Biology and Evolution* 2000,
 1047 17:703-717.
- 1048 82. Wu M, Moore RC: The evolutionary tempo of sex chromosome degradation in *Carica papaya*.
 1049 Journal of Molecular Evolution 2015, 80:265-277.
- 105083.Lenormand T, Roze D: Y recombination arrest and degeneration in the absence of sexual1051dimorphism. bioRxiv 2021:2021.2005.2018.444606.
- 105284.Jay P, Tezenas E, Giraud T: A deleterious mutation-sheltering theory for the evolution of sex1053chromosomes and supergenes. *bioRxiv* 2021:2021.2005.2017.444504.
- 1054 85. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM: Canu: scalable and accurate
 1055 long-read assembly via adaptive k-mer weighting and repeat separation. Genome Research
 1056 2017, 27:722-736.
- 1057 86. Chaisson MJ, Tesler G: Mapping single molecule sequencing reads using basic local alignment
 1058 with successive refinement (BLASR): application and theory. BMC Bioinformatics 2012,
 1059 13:238.
- 1060 87. Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A,
 1061 Huddleston J, Eichler EE, et al: Nonhybrid, finished microbial genome assemblies from long1062 read SMRT sequencing data. Nature Methods 2013, 10:536.
- 1063 88. Foissac S, Gouzy J, Rombauts S, Mathe C, Amselem J, Sterck L, de Peer VY, Rouze P, Schiex T:
 1064 Genome annotation in plants and fungi: EuGene as a model platform. Current Bioinformatics
 1065 2008, 3:87-97.
- 106689.The UniProt C: Ongoing and future developments at the universal protein resource. Nucleic1067Acids Research 2011, 39:214-219.
- 106890.Ekseth OK, Kuiper M, Mironov V: OrthAgogue: an agile tool for the rapid prediction of1069orthology relations. Bioinformatics 2014, 30:734-736.
- 1070 91. Dongen SV: A cluster algorithm for graphs. *Information Systems [INS]* 2000:1-40.
- 107192.Hartmann FE, Rodríguez de la Vega RC, Brandenburg J-T, Carpentier F, Giraud T: Gene1072presence-absence polymorphism in castrating anther-smut fungi: recent gene gains and1073phylogeographic structure. Genome Biology and Evolution 2018, 10:1298-1314.
- 107493.Krzywinski aMe: Circos: an information aesthetic for comparative genomics. Genome Res10752009, 19:1639-1645.
- 1076 94. Quinlan AR, Hall IM: BEDTools: a flexible suite of utilities for comparing genomic features.
 1077 Bioinformatics 2010, 26:841-842.
- Perlin MH, Amselem J, Fontanillas E, Toh SS, Chen Z, Goldberg J, Duplessis S, Henrissat B, Young
 S, Zeng Q, et al: Sex and parasites: Genomic and transcriptomic analysis of *Microbotryum Iychnidis-dioicae*, the biotrophic and plant-castrating anther smut fungus. *BMC genomics* 2015, 16:1-24.
- Bray NL, Pimentel H, Melsted P, Pachter L: Near-optimal probabilistic RNA-seq quantification.
 Nature Biotechnology 2016, 34:525-527.
- 1084 97. Edgar RC: MUSCLE: Multiple sequence alignment with high accuracy and high throughput.
 1085 Nucleic Acids Research 2004, 32:1792-1797.
- 108698.Yang Z, Nielsen R: Estimating synonymous and nonsynonymous substitution rates under1087realistic evolutionary models. Molecular Biology and Evolution 2000, 17:32-43.
- Bouckaert R, Vaughan TG, Barido-Sottani J, Duchêne S, Fourment M, Gavryushkina A, Heled J,
 Jones G, Kühnert D, De Maio N, et al: BEAST 2.5: An advanced software platform for Bayesian
 evolutionary analysis. PLOS Computational Biology 2019, 15:1-28.

1091 100. Gladieux P, Vercken E, Fontaine MC, Hood ME, Jonot O, Couloux A, Giraud T: Maintenance of fungal pathogen species that are specialized to different hosts: allopatric divergence and 1092 1093 introgression through secondary contact. Molecular Biology and Evolution 2011, 28:459-471. 1094 101. Ranwez V, Douzery EJP, Cambon C, Chantret N, Delsuc F: MACSE v2: Toolkit for the Alignment 1095 of Coding Sequences Accounting for Frameshifts and Stop Codons. Molecular Biology and 1096 Evolution 2018, 35:2582-2584. 1097 Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R: 102. 1098 IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era.

1099 *Molecular Biology and Evolution* 2020, **37:**1530-1534.

1100

1101 **Table 1: Optimal codons in** *Microbotryum*. All codons are represented for each amino acid.

Amino	Optimal					
acid	codon	Synonymous codons				
Ala	GCC	GCT	GCA	GCG		
Arg	CGC	CGT	CGA	CGG	AGA	AGG
Asn	AAC	AAT				
Asp	GAC	GAT				
Cys	TGC	TGT				
Gln	CAG	CAA				
Glu	GAG	GAA				
Gly	GGC	GGT	GGA	GGG		
His	CAC	CAT				
lle	ATC	ATT	ATA			
Leu	СТС	CTT	СТА	CTG	TTA	TTG
Lys	AAG	AAA				
Phe	TTC	TTT				
Pro	CCC	ССТ	CCA	CCG		
Ser	TCG	ТСТ	TCC	TCA	AGC	AGT
Thr	ACC	ACT	ACA	ACG		
Tyr	TAC	TAT				
Val	GTC	GTT	GTA	GTG		
Stp	TAA	TGA	TAG			

1102 Optimal codons found using WCA and expression level are shown for each amino acid.

1103

1104

1105 Figures Legends

Figure 1: Timetrees and rearrangement scenarios of recombination suppression events in the *Microbotryum* genus. A) Densitree plot of 1,000 random trees with the best posterior probabilities based on the concatenation of single-copy autosomal genes (green traces) and single-copy genes that are located in the non-recombining regions (black traces) of all species 1110 used here but *M. intermedium* (that has only very small regions without recombination around its mating-type loci) and *M. silenes-acaulis* (involving a very different rearrangement of the PR 1111 chromosome, with therefore very few gene shared with other species in its non-recombining 1112 region). Tips are labelled with scientific names of *Microbotryum* and of their hosts. Photos of 1113 diseased flowers by M.E. Hood and Eugene Popov @iNatutalist, croped. CC BY 4.0. Inset: 1114 density of estimates of time since recombination suppression for M. v. paradoxa strata. Note 1115 1116 the different scales for young (pink and white) and old strata (black, brown and light brown). Approximate position of the corresponding regions in the ancestral chromosome order is shown 1117 (HD chromosome in light blue and PR chromosome in light purple, HD locus in dark blue and 1118 PR locus in dark purple) and the succession of recombination suppression events is pictured 1119 below the rearrangement scenario (see Figure S2). Symbols key is also valid for panel B. B) 1120 1121 Rearrangement scenarios bringing the HD and PR loci into the same chromosome or linkage 1122 events to their centromeres (M. lagerheimii and M. saponariae). The ten independent events of chromosomal rearrangement and/or recombination suppression are separated by white lines. 1123 1124 Scenarios shared by contiguous species on the tree are depicted only once. Chromosome arrangement in *M. intermedium* corresponds to the ancestral state [60, 61]. Symbol key is the 1125 same as in the inset of panel A. Rearrangements are indicated by blue shades. C) Consensus 1126 1127 timetrees of single-copy autosomal genes (green trace) and evolutionary strata (with their corresponding color). Timetrees in the upper panel are identical but for the different dates 1128 1129 estimated for the suppression of recombination in different strata of M. v. paradoxa. The timetree in the lower panel is the consensus of 32 concatenated single-copy genes in the non-1130 1131 recombining region of all species but *M. intermedium* and *M. silenes-acaulis*. The shaded areas 1132 in this tree show the two trans-specific polymorphism occurrences and the big black dot 1133 indicates the different placement of M. v. paradoxa compared to the species tree

1134 Figure 2: Allelic divergence analysis between mating-type chromosomes for the newly sequenced species. For each species (Microbotryum coronariae, M. v. melanantha, M. v. 1135 tatarinowii and M. lychnidis-dioicae), we show the synonymous divergence calculated between 1136 1137 alternative mating types and plotted along the *M. intermedium* mating-type chromosome gene order. Zero d_S between mating-type chromosomes indicate recombination in these selfing 1138 species and the level of d_s between mating-type chromosomes increases with time since 1139 recombination suppression. Different evolutionary strata, i.e., different steps extending 1140 recombination suppression, are indicated by different colours. Divergence between the a1 and 1141

1142 a₂ pheromone receptor (PR) genes was too extensive, and d_s could not be calculated (depicted1143 as "Un" for unalignable).

Figure 3: Analysis of the frequency of optimal codons in the a₁ genome of *Microbotryum* 1144 silenes-dioicae. A) Distribution of the frequency of optimal codons across genomic 1145 1146 compartments. B) Distribution of the frequency of optimal codons predicted by the logistic regression, accounting for genomic compartment, difference of GC in coding and intronic 1147 sequences and their interactions as factor. In A) and B), boxplot grey shade refers to the 1148 genomic compartment. For each genomic compartment, the red dot indicates the mean 1149 frequency of optimal codons, the sample size (N) is labelled on top of the corresponding 1150 1151 boxplot, as well as the significance level of the difference between a given genomic compartment and the autosomes (NS: non-significant; ***: p-value < 0.001) from the logistic 1152 regression. Note the difference in the ranges of Y axes for panels A and B. Similar analyses for 1153 all species studied here are presented in Figure S6 and S7 in the supplementary materials. C) 1154 Frequency of optimal codons predicted by the logistic regression plotted against the difference 1155 1156 in GC content between coding and intronic sequences. From the logistic regression, significant interactions between a genomic compartment and the autosomes are indicated by solid line and 1157 filled dots, while non-significant interactions are indicated by dotted line and hollow dots. 1158 "Reference" refers to the reference considered in the logistic regression. D) Predicted effect of 1159 GC-biased gene conversion and selection on frequency of optimal codons. The figure shows 1160 1161 the interpretation of the absence of interaction (2) or the presence of a positive (3) or negative 1162 interaction (4) between the difference in GC content between coding sequences and introns and 1163 the genomic compartments regarding the logistic regression performed in this study. The frequency of optimal codons in autosomes (1) decreases with the increase of difference in GC 1164 1165 content and the frequency of optimal codons are lower in non-recombining regions (2, 3, 4).

Figure 4: Tempo of degeneration in *Microbotryum* fungi. A) Frequency of optimal codons 1166 as a function of time since recombination suppression (in million years). B) Non-synonymous 1167 1168 over synonymous divergence (d_N/d_S) as a function of time since recombination suppression (in 1169 million years). The dots represent the data used in the regressions, plotted as the frequency of 1170 optimal codons or d_N/d_S against the time since recombination suppression. Dot edges are coloured according to species and the dot fill colours correspond to genomic compartments. 1171 Predictions from the models are show as blue line and the 95% confidence interval of the 1172 prediction as grey background. 1173

1174

1175 Additional files:

1176 1-All supplementary Tables as a single xls file (different Tables as sheets)

1177 2-All supplementary Figures as a single PDF file

1178 Supplemental Figure legends

Figure S1: Synteny and allelic divergence analysis between mating-type chromosomes of 1179 1180 alternative mating types for the species *Microbotryum intermedium*. We show synteny plots between a₁ and a₂ PR mating-type chromosome (left top panel), synteny plots between a₁ and 1181 1182 a₂ HD mating-type chromosomes (right top panel), and the synonymous divergence calculated between alternative mating types and plotted along the M. intermedium PR and HD mating-1183 type chromosome gene order (bottom left and right panel, resp.). Zero d_S between mating-type 1184 chromosomes indicate recombination in these selfing species and the level of d_s between 1185 mating-type chromosomes increases with time since recombination suppression. Different 1186 evolutionary strata, i.e., different steps extending recombination suppression, are indicated by 1187 different colours. Divergence between the a_1 and a_2 pheromone receptor (PR) genes was too 1188 extensive, and d_s could not be calculated (depicted as "Un" for unalignable). 1189

1190 Figure S2: Synteny between mating-type chromosomes of alternative mating types for the 1191 *Microbotryum* species A) *M. lychnis flos-cuculi*, B) *M. v. melanantha* and C) *M. v.* 1192 *tatarinowii*. For each species, we show synteny plots between a_1 and a_2 mating-type 1193 chromosomes (left top), synteny plots between a_1 mating-type chromosome and a_1 mating-type 1194 chromosome of *M. intermedium* (right panel).

Figure S3: Synteny and allelic divergence analysis between mating-type chromosomes of 1195 1196 alternative mating types for the species *M. scabiosae* (A) and *Microbotryum v. caroliniana* (B). For each species, we show synteny plots between a_1 and a_2 mating-type chromosomes (left 1197 1198 top panel), synteny plots between a_1 and a_2 mating-type chromosomes of *M. intermedium* (right top panel), and the synonymous divergence calculated between alternative mating types and 1199 plotted along the *M. intermedium* mating-type chromosome gene order (bottom panel). Zero d_S 1200 between mating-type chromosomes indicate recombination in these selfing species and the level 1201 of d_s between mating-type chromosomes increases with time since recombination suppression. 1202 On the top left figure in B), the light blue evolutionary stratum is indicated by a light blue box 1203 1204 on the outer track.
bioRxiv preprint doi: https://doi.org/10.1101/2021.07.20.453045; this version posted July 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1205

1206 Figure S4: Allelic divergence analysis between mating-type chromosomes of alternative 1207 mating types for *Microbotryum v. paradoxa*. The synonymous divergence calculated between alternative mating types is plotted along the *M. intermedium* mating-type chromosome 1208 1209 ancestral-like gene order. Zero d_s between mating-type chromosomes indicate recombination in these selfing species and the level of d_s between mating-type chromosomes increases with 1210 1211 time since recombination suppression. Different evolutionary strata, i.e., different steps 1212 extending recombination suppression, are indicated by different colours. Divergence between 1213 the a₁ and a₂ pheromone receptor (PR) genes was too extensive, and d_s could not be calculated (depicted as "Un" for unalignable). 1214

1215

Figure S5: Within-correspondence analysis plots. A) WCA plots for the a₁ and a₂ genomes of *Microbotryum intermedium* (resp. left and right top panel) and a₁ and a₂ genomes of *M. lychnidis-dioicae* (resp. left and right bottom panel). B) Percentage of variance explained for each correspondence analysis axis for the a₁ and a₂ genomes of *M. intermedium* (resp. left and right top panel) and a₁ and a₂ genomes of *M. intermedium* (resp. left and right top panel) and a₁ and a₂ genomes of *M. intermedium* (resp. left and right bottom panel).

Figure S6 Data distribution across genomic compartments in *Microbotryum*. A) Distribution of the frequency of optimal codons. B) Distribution of the difference in GC content between coding sequences and introns. C) Distribution of the expression level of *M*. *intermedium* orthologous genes. The gene expression level distribution was truncated at the 90th percentile for visualisation purpose and was not represented for the genomes a_1 and a_2 *M*. *intermedium*. Boxplot colours refer to the genomic compartment.

1228

Figure S7: Analysis of the frequency of optimal codons in all *Microbotryum* genomes using the genomic compartments considered in the main text. A) Distribution of the frequency of optimal codons across genomic compartments. B) Distribution of the predicted frequency of optimal codons predicted by the logistic regression. In A) and B), boxplot colours refer to the genomic compartment. For each genomic compartment, the red dot indicates the mean frequency of optimal codons, the sample size (N) is labelled on top of the corresponding boxplot, as well as the significance level of the difference between a given genomic compartment and the autosomes (NS: non-significant; ".": p-value < 0.1; *: p-value < 0.05; **:
p-value < 0.01; ***: p-value < 0.001). C) Frequency of optimal codons predicted by the logistic
regression along the difference of GC content between coding and intronic sequences.
Significant interactions between a genomic compartment and the autosomes are indicated by a
solid line and a filled dot, while non-significant interactions are indicated by a dotted line and
a hollow dot.

1242 Figure S8: Analysis of the frequency of optimal codons in all *Microbotryum* genomes using

all the genomic compartments. A) Distribution of the frequency of optimal codons across 1243 genomic compartments. B) Distribution of the predicted frequency of optimal codons predicted 1244 by the logistic regression. In A) and B), boxplot colours refer to the genomic compartment. For 1245 1246 each genomic compartment, the red dot indicates the mean frequency of optimal codons, the sample size (N) is labelled on top of the corresponding boxplot, as well as the significance level 1247 1248 of the difference between a given genomic compartment and the autosomes (NS: nonsignificant; ".": p-value < 0.1; *: p-value < 0.05; **: p-value < 0.01; ***: p-value < 0.001). C) 1249 1250 Frequency of optimal codons predicted by the logistic regression along the difference of GC content between coding and intronic sequences. Significant interactions between a genomic 1251 compartment and the autosomes are indicated by a solid line and a filled dot, while non-1252 1253 significant interactions are indicated by a dotted line and a hollow dot.

1254

1255

1256

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.20.453045; this version posted July 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



MRCA (MY)

0.0

Α

	1			
-25	-20	-15	-10	-5



frans-specific





Gene order along the mating-type chromosomes of *M. intermedium*



In non-recombining regions, the reduction of GC-biased gene conversion results in a significant decrease in frequency of optimal codons.



No significant interaction. The difference in frequencies of optimal codons between autosomes and non-recombining region is only due to relaxed selection in non-recombining regions.



Significant and positive interaction. Selection on optimal codons is not relaxed in non-recombining regions and the reduced frequency of optimal codons compared to autosomes can only be explained by less GC-biased gene conversion.



Significant and negative interaction. Less GC-biased gene conversion in nonrecombining regions alone can explain the reduced frequency of optimal codons compared to autosomes. Less efficient selection may also acts.





Legend



Figure S1





Key to tracks

Genes

synonymous divergence

 a_2

a1

Genes with $d_{\rm S}=0$

Transposable elements

Centromeric repeats

Key to links

- One-to-one orthologs in the same orientation
 - One-to-one orthologs in inverted orientation

Evolutionary strata

- Purple stratum
- Blue stratum





Gene order along the PR mating-type chromosome *M. intermedium* Gene order along the HD mating-type chromosome *M. intermedium*

Figure S2



Key to tracks

Key to links

- One-to-one orthologs in the same orientation
- One-to-one orthologs in inverted orientation
- Genes

Centromeric repeats

Genes with $d_s = 0$

Transposable elements

Figure S3a



Key to tracks

Key to links

- Genes with $d_s = 0$
 - Transposable elements On
- Genes
- Centromeric repeats
- One-to-one orthologs in the same orientationOne-to-one orthologs in inverted orientation



Gene order along the mating-type chromosomes of *M. intermedium*

Figure S3b



Key to tracks

Genes

Genes with $d_s = 0$

Transposable elements

Centromeric repeats

Key to links

One-to-one orthologs in the same orientation

One-to-one orthologs in inverted orientation

Evolutionary strata

Lightblue stratum





Gene order along the mating-type chromosomes of *M. intermedium*

Figure S4



M. violaceum s.l. on *Silene paradoxa*



Gene order along the mating-type chromsomes of *M. intermedium*



Figure S5b







Figure S6











Recombining region of the PR chr. Recombining region of the HD chr.

- Purple stratum
- Blue stratum





Genomic compartments



M. lagerheimii on *Silene vulgaris* a₁ genome





M. lagerheimii

on Silene vulgaris a₂ genome







M. lychnidis-diocae on *Silene latifolia* a₁ genome



M. lychnidis-diocae on *Silene latifolia* a₂ genome



M. silenes-acaulis on Silene acaulis a₁ genome



Genomic compartments













M. saponariae on Saponaria officinalis a₁ genome







M. saponariae on Saponaria officinalis a₂ genome

























M. silenes-diocae on Silene dioica a₁ genome



M. silenes-diocae on *Silene dioica* a₂ genome





M. coronariae on Silene flos-cuculi a1 genome























M. violaceum s.l. on *Silene caroliniana* a₁ genome





M. violaceum s.l. on Silene caroliniana a₂ genome











Blue stratum

Orange stratum











M. violaceum s.s. on Silene nutans a_2 genome



Genomic compartments





M. violaceum s.l. on *Silene paradoxa* a₁ genome




M. violaceum s.l. on Silene tatarinowii a₁ genome



Figure S7

or

M. intermedium on *Salvia pratensis* a₁ genome





M. intermedium on Salvia pratensis a₂ genome





Difference of GC in coding and intronic sequences



















M. saponariae on *Saponaria officinalis* a₂ genome













M. silenes-diocae on *Silene dioica* a₁ genome





M. silenes-diocae on Silene dioica a₂ genome

























M. violaceum s.s. on *Silene nutans* a₁ genome





M. violaceum s.s. on Silene nutans a₂ genome





M. violaceum s.l. on *Silene paradoxa* a₁ genome







M. violaceum s.l. on Silene paradoxa a₂ genome



0.1

Difference of GC in coding and intronic sequences



M. violaceum s.l. on Silene tatarinowii a₂ genome





M. violaceum s.l. on Silene tatarinowii a_2 genome



Figure S8

M. intermedium on Salvia pratensis a1 genome





M. intermedium on Salvia pratensis a_2 genome





M. lagerheimii on Silene vulgaris a₁ genome





Genomic compartments





M. lagerheimii on Silene vulgaris a₂ genome





Genomic compartments









M. lychnidis-diocae on Silene latifolia a2 genome



²000.

0

. 0.2

σ

ð

Difference of GC in coding and intronic sequences

0.0

0.39

0.36

-0.2

Interaction significance

- Significant
- Non-significant ·о •

0.4






M. saponariae on Saponaria officinalis a₁ genome





Genomic compartments







Genomic compartments











M. scabiosae on *Knautia arvensis* a₂ genome





M. silenes-diocae on Silene dioica a₁ genome







0.000000000

-0.1

0.375

-0.2

0

Difference of GC in coding and intronic sequences

0.1

0.0

Interaction significance

Significant

· O· ·O

0.3

0.2

Non-significant · O ·



M. coronariae on Silene flos-cuculi a₁ genome





M. coronariae on Silene flos-cuculi a₂ genome





M. violaceum s.l. on Silene melanantha a_1 genome





M. violaceum s.l. on *Silene melanantha* a₂ genome





M. violaceum s.l. on Silene caroliniana a₁ genome

NS

N = 4

*





Genomic compartments

Lightblue stratum

Orange stratum



Interaction significance

- Significant
- Non-significant



M. violaceum s.l. on Silene caroliniana a_2 genome







M. violaceum s.s. on *Silene nutans* a₁ genome





Genomic compartments



Frequency of optimal codons 0.75 0.50 0.25

A)





M. violaceum s.l. on Silene paradoxa a₁ genome

N = 12

NS

rown stratum





Genomic compartments



Interaction significance with autosomes

	Significant
· o ·	Non-significant



M. violaceum s.l. on *Silene paradoxa* a₂ genome





Interaction significance with autosomes

	Significant
· o ·	Non-significant





M. violaceum s.l. on *Silene tatarinowii* a₁ genome





M. violaceum s.l. on Silene tatarinowii a₂ genome

