



HAL
open science

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 Vega¹, Michael H. Perlin³, R. Margaret
5 Wallen³, Michael E. Hood^{#4}, Tatiana Giraud^{#*1}

6

7 # These authors co-supervised the study

8

9 ¹ Ecologie Systématique Evolution, Bâtiment 360, CNRS, AgroParisTech, Université Paris-Saclay, 91400 Orsay,
10 France

11 ² Université de Lille, CNRS, UMR 8198-Evo-Eco-Paleo F-59000, Lille, France

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
17 Orsay France

18 phone: +33 1 69 15 56 69 + 33 7 86 01 67 59 fax: +33 1 69 15 46 97

19 tatiana.giraud@universite-paris-saclay.fr

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

37

38 **Abstract**

39

40 Recombination is beneficial over the long term, allowing more effective selection. Despite
41 long-term advantages of recombination, local recombination suppression is known to evolve
42 and lead to genomic degeneration, in particular on sex and mating-type chromosomes,
43 sometimes linked to severe genetic diseases. Here, we investigated the tempo of degeneration
44 in non-recombining regions, i.e., the function curve for the accumulation of deleterious
45 mutations over time, taking advantage of 17 independent events of large recombination
46 suppression identified on mating-type chromosomes of anther-smut fungi, including five newly
47 identified in the present study. Using high-quality genomes assemblies of alternative mating
48 types of 13 *Microbotryum* species, we estimated the degeneration levels in terms of
49 accumulation of non-optimal codons and non-synonymous substitutions in non-recombining
50 regions. We found a reduced frequency of optimal codons in the non-recombining regions on
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
54 estimated that the frequency of optimal codon usage decreased linearly at a rate of 0.989 per
55 My. The non-synonymous over synonymous substitution rate (d_N/d_S) increased rapidly after
56 recombination suppression and then reached a plateau. To our knowledge this is the first study
57 to disentangle effects of reduced selection efficacy from GC-biased gene conversion in the
58 evolution of optimal codon usage to quantify the tempo of degeneration in non-recombining
59 regions, leveraging on multiple independent recombination suppression events. Understanding
60 the tempo of degeneration is important for our knowledge on genomic evolution, on the origin
61 of genetic diseases and on the maintenance of regions without recombination.

62

63

64

65

66

67

68

69

70

71

72

73

74

75

76

77 **Introduction**

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

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

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

110 some codons being preferentially used over their synonymous alternative codons. This
111 preference is thought to result from a selection on either the rate or the accuracy of translation
112 [37], such that the preferred codons are called optimal codons. Optimal codons tend to
113 correspond to the most abundant tRNAs [38, 39], and are most frequent in highly expressed
114 genes [40, 41].

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

132 While the accumulation of deleterious mutations in non-recombining regions has been
133 reported in numerous systems, only a few studies have addressed the question of the tempo of
134 degeneration *i.e.* the rate at which deleterious mutations accumulate with time. Theoretical
135 works based on the mammalian Y-chromosome predicted that degeneration through gene losses
136 should occur at a high speed in young non-recombining regions, mainly due to Muller's ratchet
137 and background selection, and should then slows down in older non-recombining regions in
138 which genetic hitchhiking is a major driving force [16, 51]. In the plant *Silene latifolia*, the sex
139 chromosomes underwent two successive recombination suppression events (*i.e.*, evolutionary
140 strata; [52]), which have been losing genes at estimated rates of 4.64% and 6.92% per Myr for
141 the older and younger evolutionary strata, respectively [53]. Other observations have similarly
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
144 the tempo of degeneration requires estimating the time since recombination cessation and
145 gathering data for multiple independent events of recombination suppression of various ages.
146 Recombination suppression evolved numerous times in a wide range of organisms, even
147 sometimes within the same genus and at different times [56, 57]. Moreover, the evolutionary
148 strata present in some sex chromosomes represent stepwise extension events of recombination
149 suppression that occurred at different time points; *e.g.*, the four successive steps of
150 recombination suppression on the mammalian Y chromosome [52, 58, 59].

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

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

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

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

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

222

223

224 **Results**

225

226 **PR-HD linkage in three new *Microbotryum* genomes**

227 We sequenced (long-read Pacific Bioscience technology) and assembled the genomes from
228 haploid cells bearing opposite mating types (called a_1 and a_2) of a single diploid individual, for
229 the following species: *Microbotryum violaceum* (*s.l.*) parasitizing *Silene melanantha* (called *M.*
230 *v. melanantha*), *M. violaceum* (*s.l.*) parasitizing *S. tatarinowii* (called *M. v. tatarinowii*) and *M.*
231 *coronariae* parasitizing *Silene flos-cuculi* (Syn. *Lychnis flos-cuculi*) (Figure 1; Table S1). We
232 reconstructed the evolutionary history of their mating-type chromosomes by comparing their
233 genome structure to those of *M. intermedium*, taken as a proxy for the genomic ancestral state,
234 following previous studies [60, 61]. In *M. intermedium*, the PR and the HD mating-type loci are
235 located on distinct mating-type chromosomes (Figure S1).

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

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

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

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

274

275 **Evolutionary strata**

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

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

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

316

317 **Absolute dating of recombination suppression events**

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

327

328 **Identification of optimal codons**

329 To identify optimal codons in *Microbotryum* fungi, we investigated which codons were
330 preferentially used in highly expressed autosomal genes in the two distantly related species, *M.*
331 *lychnidis-dioicae* and *M. intermedium* (Figure 1). Selection for codon usage is expected to be
332 strongest on most highly expressed genes and most effective on autosomes because
333 recombination occurs. We had expression data for haploid cells cultivated clonally on nutritive
334 medium (Table S4), for the two mating types separately. To study autosomal codon usage, we
335 performed a within-group correspondence analysis (WCA; [70]), *i.e.*, a multivariate statistical
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* a_1 mating

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

353

354 **Decrease in optimal codon odds in non-recombining regions**

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

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.

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

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

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

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

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

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

438

439 **Control for ancestral gene expression level**

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

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

465

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

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

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

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

493

494 **Discussion**

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

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

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

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

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

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

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

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

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

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

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

621

622 **Data access**

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

626

627 **Material and methods**

628 **Genome sequencing**

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

641

642 **Assembly of new genomes and assembly improvement for *M. scabiosae* infecting *Knautia*** 643 ***arvensis* (initially published in Branco et al 2018).**

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

653 The previous genome assemblies of the a₁ and a₂ genome of *M. scabiosae* were based on five
654 PacBio movies. We re-assembled the genome using three additional movies, from the same
655 strain, and generated using the same PacBio technology for a total coverage of 506x for the a₁
656 genome and 803x for the a₂ genome. The assemblies were substantially improved, as the contig

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

659

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

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

669

670 **Evolutionary scenarios of the mating-type chromosomes**

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

685 We generated RNAseq data for *M. intermedium*. Total RNA was isolated from haploid cells
686 using the Qiagen RNeasy Mini Kit. Haploid strains (a_1 or a_2) were streaked on Potato Dextrose

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

704 For *M. lychnidis-dioicae*, we used RNAseq data published previously [95]. List and statistics of
705 the RNAseq experiments used in this study are provided in Table S4 We controlled the quality
706 of the RNAseq experiment data using fastQC (<https://github.com/s-andrews/FastQC>). We
707 trimmed the sequences using trimmomatic (Bolger et al. 2014); parameters:
708 ILLUMINACLIP:TruSeq3-PE.fa:2:30:10:2:keepBothReads LEADING:3 TRAILING:3
709 MINLEN:36). We only considered the resulting trimmed paired-end reads for further analysis
710 (Table S3). For each RNAseq dataset, we pseudo-aligned the reads onto the corresponding
711 references and quantified the gene expression levels using Kallisto using the “kallisto quant”
712 command with the “--bias” option [96]. For the diploid RNAseq experiment (Table S4, we used
713 the concatenated coding sequence set from a₁ and a₂ haploid genomes as reference. For haploid
714 RNAseq experiment, we considered the haploid coding sequence set from the corresponding
715 mating type. For each RNAseq experiment, we removed all the genes with an expression level
716 above the 95th percentile of the expression level distribution to avoid spurious high expression
717 level.

718

719 **Identifying optimal codons in *Microbotryum* fungi**

720 We generated a codon usage table for each of the a_1 *M. lychnidis-dioicae*, a_2 *M. lychnidis-*
721 *dioicae*, a_1 *M. intermedium* and a_2 *M. intermedium* haploid genomes (in-house awk script) and
722 filtered the genes that overlapped transposable elements, that were previously identified [92].
723 We then performed a within-group correspondence analysis (WCA), following the procedure
724 available at <http://pbil.univ-lyon1.fr/datasets/charif04/> [70] on each codon usage table.
725 Correspondence analysis is a multivariate statistical method that summarizes codon count data
726 by reducing them to a limited number of variables. WCA more specifically allows to dissociate
727 the effects of different amino-acid compositions from the effects directly related to synonymous
728 codon usage [70]. This method adjusts the value for each codon by the average value of all the
729 codons encoding for the same amino acid. We only considered the two first principal
730 components which explained between 16.30% and 16.50% of the variance for the first principal
731 component and between 7.75% and 8.51% for the second principal component (Figure S5). The
732 coordinates of the genes projected onto the first (PC1) and second (PC2) axes were significantly
733 correlated with gene expression level (Pearson's correlation tests; Table S5. We considered the
734 first axis to better characterized the gene expression level variation because the R^2 calculated
735 between PC1 gene coordinates and gene expression level were higher by two orders of
736 magnitude compared to the R^2 calculated between PC2 gene coordinates and gene expression
737 level (see correlation's coefficient value in Table S05).

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

745

746 **Frequency of optimal codons and GC content statistics**

747 For each coding sequence with base pair numbers multiple of three, we calculated the frequency
748 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

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

753

754 **Non-synonymous and synonymous divergence estimation**

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

758

759 **Absolute dating of recombination suppressions**

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

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

782 *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
785 previous study [62]. Gene family alignments are available as supplementary material.

786

787 **Statistics**

788 We tested in each of the 26 *Microbotryum* genomes (one genome per mating type for 13
789 species) whether the genes in distinct genomic compartments (PARs and the different
790 evolutionary strata with different ages of recombination suppression) had lower frequencies of
791 optimal codons compared to autosomal genes. To take into account the effect of any process
792 other than genomic degeneration that could lower the GC level in the non-recombining regions
793 such as less frequent GC-biased gene conversion in the non-recombining regions, we included
794 in the analysis the difference in GC content between coding and intronic sequences across
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
798 content between coding and intronic sequences across genes and/or their interaction. Using the
799 glm function in R, we gave as input the frequency of optimal codons weighted by the total
800 number of codons per gene (using the “weight” option in the glm function). After performing
801 a logistic regression assuming binomial distribution and a logit link, we noticed that predicted
802 frequencies of optimal codons were all lower than the raw frequency of optimal codons which
803 could indicate overdispersion issues. We therefore performed the logistic regression assuming
804 a quasibinomial distribution, the only difference being in the estimation of the dispersion
805 parameter to correct, among other things, the p-values. The estimated dispersion parameter was
806 high for the logistic regression model of each genome ($\Phi > 2$), so we choose the quasibinomial
807 distribution to interpret the data. Choosing a quasibinomial distribution did not change however
808 the log(odds) estimated by the model. We tested whether the relationship between the frequency
809 of optimal codons and the time since recombination suppression was better explained with a
810 non-linear component using a non-linear model (polynomial with two degrees) than with only
811 a linear component (Table S10).

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

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

819 We inquired how much expression level changed for genes in recombining region across the
820 *Microbotryum* genus. To do so, we performed Pearson's correlation tests between *M.*
821 *intermedium* and *M. lychnidis-dioicae* orthologous genes for all combinations of RNA-seq
822 experiment replicates that were comparable between the two species (*i.e.* cells grown on rich
823 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**

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

842 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
844 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.

- 886 19. Stolle E, Pracana R, Howard P, Paris CI, Brown SJ, Castillo-Carrillo C, Rossiter SJ, Wurm Y:
887 **Degenerative expansion of a young supergene.** *Molecular Biology and Evolution* 2019,
888 **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.
- 891 21. Brion C, Caradec C, Pflieger D, Friedrich A, Schacherer J: **Pervasive phenotypic impact of a**
892 **large non-recombining introgressed region in yeast.** *Molecular Biology and Evolution* 2020,
893 **37**:2520-2530.
- 894 22. Hough J, Hollister JD, Wang W, Barrett SCH, Wright SI: **Genetic degeneration of old and young**
895 **y chromosomes in the flowering plant *Rumex hastatulus*.** *Proceedings of the National*
896 *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.
- 900 24. Papadopulos AST, Chester M, Ridout K, Filatov DA: **Rapid Y degeneration and dosage**
901 **compensation in plant sex chromosomes.** *Proceedings of the National Academy of Sciences*
902 2015, **112**:201508454.
- 903 25. Whittle CA, Sun Y, Johannesson H: **Degeneration in codon usage within the region of**
904 **suppressed recombination in the mating-type chromosomes of *Neurospora tetrasperma*.**
905 *Eukaryotic Cell* 2011, **10**:594-603.
- 906 26. Bachtrog D: **Sex chromosome evolution: Molecular aspects of Y-chromosome degeneration**
907 **in *Drosophila*.** *Genome Research* 2005, **15**:1393-1401.
- 908 27. Bartolome C, Charlesworth B: **Evolution of amino-acid sequences and codon usage on the**
909 ***Drosophila miranda* neo-sex chromosomes.** *Genetics* 2006, **174**:2033-2044.
- 910 28. Steinemann M, Steinemann S: **Degenerating Y chromosome of *Drosophila miranda*: A trap**
911 **for retrotransposons (chromosome structure/larval cuticle protein genes).** *Genetics* 1992,
912 **89**:7591-7595.
- 913 29. Bianchi NO: **Y chromosome structural and functional changes in human malignant diseases.**
914 *Mutation research* 2009, **682**:21-27.
- 915 30. Lee Y, Kim C, Park Y, Pyun J-A, Kwack K: **Next generation sequencing identifies abnormal Y**
916 **chromosome and candidate causal variants in premature ovarian failure patients.** *Genomics*
917 2016, **108**:209-215.
- 918 31. Badouin H, Hood ME, Gouzy J, Aguilera G, Siguenza S, Perlin MH, Cuomo CA, Fairhead C, Branca
919 A, Giraud T: **Chaos of rearrangements in the mating-type chromosomes of the anther-smut**
920 **fungus *Microbotryum lychnidis-dioicae*.** *Genetics* 2015, **200**:1275-1284.
- 921 32. Yan H, Jin W, Nagaki K, Tian S, Ouyang S, Buell CR, Talbert PB, Henikoff S, Jiang J: **Transcription**
922 **and histone modifications in the recombination-free region spanning a rice centromere.**
923 *Plant Cell* 2005, **17**:3227-3238.
- 924 33. Hollister JD, Gaut BS: **Epigenetic silencing of transposable elements: A trade-off between**
925 **reduced transposition and deleterious effects on neighboring gene expression.** *Genome*
926 *Research* 2009, **19**:1419-1428.
- 927 34. Li SF, Zhang GJ, Yuan JH, Deng CL, Gao WJ: **Repetitive sequences and epigenetic modification:**
928 **inseparable partners play important roles in the evolution of plant sex chromosomes.** *Planta*
929 2016, **243**:1083-1095.
- 930 35. Duret L: **Evolution of synonymous codon usage in metazoans.** *Current Opinion in Genetics and*
931 *Development* 2002, **12**:640-649.
- 932 36. Sharp PM, Averof M, Lloyd AT, Matassi G, Peden JF: **DNA sequence evolution: the sounds of**
933 **silence.** *Philosophical transactions of the Royal Society of London Series B, Biological sciences*
934 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.**
- 947 41. Zhou Z, Dang Y, Zhou M, Li L, Yu C-h, Fu J, Chen S, Liu Y: **Codon usage is an important**
948 **determinant 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.**
- 951 43. Betancourt AJ, Welch JJ: **Report reduced effectiveness of selection caused by a lack of**
952 **recombination.** *Current Biology* 2009, **19:655-660.**
- 953 44. Vicario S, Moriyama EN, Powell JR: **Codon usage in twelve species of *Drosophila*.** *BMC*
954 *Evolutionary Biology* 2007, **7:1-17.**
- 955 45. Duret L, Galtier N: **Biased gene conversion and the evolution of mammalian genomic**
956 **landscapes.** *Annual Review of Genomics and Human Genetics* 2009, **10:285-311.**
- 957 46. Lassalle F, Périan S, Bataillon T, Nesme X: **GC-content evolution in bacterial genomes : the**
958 **biased gene conversion hypothesis expands.** *PLoS Genet* 2015, **11:e1004941.**
- 959 47. Marais G: **Biased gene conversion: implications for genome and sex evolution.** *TRENDS in*
960 *Genetics* 2003, **19:330-338.**
- 961 48. Pessia E, Popa A, Mousset S, Rezvoy C, Duret L, Marais GAB: **Evidence for widespread GC-**
962 **biased gene conversion in eukaryotes.** *Genome Biol Evol* 2012, **4:675-682.**
- 963 49. Weber CC, Boussau B, Romiguier J, Jarvis ED, Ellegren H: **Evidence for GC-biased gene**
964 **conversion as a driver of between-lineage differences in avian base composition.** *Genome*
965 *biology* 2014, **15:549.**
- 966 50. Kostka D, Hubisz MJ, Siepel A, Pollard KS: **The role of GC-biased gene conversion in shaping**
967 **the fastest evolving regions of the human genome.** *Molecular Biology and Evolution* 2012,
968 **29:1047-1057.**
- 969 51. Bachtrog D: **The temporal dynamics of processes underlying Y chromosome degeneration.**
970 *Genetics* 2008, **179:1513-1525.**
- 971 52. Lahn BT, Page DC: **Four evolutionary strata on the human X chromosome.** *Science* 1999,
972 **286:964-967.**
- 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, Zaghul S,
976 Graves T, Rock S, et al: **Mammalian Y chromosomes retain widely expressed dosage-sensitive**
977 **regulators.** *Nature* 2014, **508:494-499.**
- 978 55. Hughes JF, Rozen S: **Genomics and genetics of human and primate Y chromosomes.** *Annual*
979 *Review 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.**

- 988 59. Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, Hillier LD, Brown LG, Reppng S,
989 Pyntikova T, All J, Blerl T, et al: **The male-specific region of the human Y chromosome is a**
990 **mosaic of discrete sequence classes.** *Nature* 2003, **423**:825-837.
- 991 60. Branco S, Badouin H, Rodríguez de la Vega RC, Gouzy J, Carpentier F, Aguilera G, Siguenza S,
992 Brandenburg J-T, Coelho MA, Hood ME, Giraud T: **Evolutionary strata on young mating-type**
993 **chromosomes despite the lack of sexual antagonism.** *Proceedings of the National Academy*
994 *of Sciences* 2017, **114**:7067-7072.
- 995 61. Branco S, Carpentier F, Rodríguez de la Vega RC, Badouin H, Snirc A, Le Prieur S, Coelho MA,
996 de Vienne DM, Hartmann FE, Begerow D, et al: **Multiple convergent supergene evolution**
997 **events in mating-type chromosomes.** *Nature Communications* 2018, **9**:2000.
- 998 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.
- 1001 63. Hood ME, Antonovics J: **Intratetrad mating, heterozygosity, and the maintenance of**
1002 **deleterious alleles in *Microbotryum violaceum* (= *Ustilago violacea*).** *Heredity* 2000, **85**:231-
1003 241.
- 1004 64. Coelho MA, Bakkeren G, Sun S, Hood ME, Giraud T: **Fungal Sex: The Basidiomycota.**
1005 *Microbiology Spectrum* 2017, **5**:1-30.
- 1006 65. Fontanillas E, Hood ME, Badouin H, Petit E, Barbe V, Gouzy J, De Vienne DM, Aguilera G, Poulain
1007 J, Wincker P, et al: **Degeneration of the nonrecombining regions in the mating-type**
1008 **chromosomes of the anther-smut fungi.** *Molecular Biology and Evolution* 2015, **32**:928-943.
- 1009 66. Ma W-J, Carpentier F, Giraud T, Hood M: **Differential gene expression between fungal mating**
1010 **types 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.
- 1017 69. Sun Y, Corcoran P, Menkis A, Whittle CA, Andersson SGE, Johannesson H: **Large-scale**
1018 **Introgression shapes the evolution of the mating-type chromosomes of the filamentous**
1019 **Ascomycete *Neurospora tetrasperma*.** *PLoS Genetics* 2012, **8**:e1002820.
- 1020 70. Charif D, Thioulouse J, Lobry JR, Perrière G: **Online synonymous codon usage analyses with**
1021 **the *ade4* and *seqinR* packages.** *Bioinformatics* 2005, **21**:545-547.
- 1022 71. Suzuki H, Brown CJ, Forney LJ, Top EM: **Comparison of correspondence analysis methods for**
1023 **synonymous 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.
- 1029 74. Lutz M, Piatek M, Kemler M, Chlebicki A, Oberwinkler F: **Anther smuts of *Caryophyllaceae*:**
1030 **molecular 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.
- 1034 76. Lampson BL, Pershing NLK, Prinz JA, Lacsina JR, Marzluff WF, Nicchitta CV, MacAlpine DM,
1035 Counter CM: **Rare codons regulate *KRas* oncogenesis.** *Current Biology* 2013, **23**:70-75.
- 1036 77. Carlini DB: **Experimental reduction of codon bias in the *Drosophila* alcohol dehydrogenase**
1037 **gene results in decreased ethanol tolerance of adult flies.** *Journal of Evolutionary Biology*
1038 2004, **17**:779-785.

- 1039 78. Qiu S, Bergero R, Zeng K, Charlesworth D: **Patterns of codon usage bias in *Silene latifolia*.** *Molecular Biology and Evolution* 2011, **28**:771-780.
- 1040
- 1041 79. Chibalina MV, Filatov DA: **Plant y chromosome degeneration is retarded by haploid purifying**
1042 **selection.** *Current Biology* 2011, **21**:1475-1479.
- 1043 80. Bergero R, Charlesworth D: **Preservation of the Y transcriptome in a 10-Million-year-old plant**
1044 **sex 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.
- 1050 83. Lenormand T, Roze D: **Y recombination arrest and degeneration in the absence of sexual**
1051 **dimorphism.** *bioRxiv* 2021:2021.2005.2018.444606.
- 1052 84. Jay P, Tezenas E, Giraud T: **A deleterious mutation-sheltering theory for the evolution of sex**
1053 **chromosomes 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 long-**
1062 **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.
- 1066 89. The UniProt C: **Ongoing and future developments at the universal protein resource.** *Nucleic*
1067 *Acids Research* 2011, **39**:214-219.
- 1068 90. Ekseth OK, Kuiper M, Mironov V: **OrthAgoque: an agile tool for the rapid prediction of**
1069 **orthology relations.** *Bioinformatics* 2014, **30**:734-736.
- 1070 91. Dongen SV: **A cluster algorithm for graphs.** *Information Systems [INS]* 2000:1-40.
- 1071 92. Hartmann FE, Rodríguez de la Vega RC, Brandenburg J-T, Carpentier F, Giraud T: **Gene**
1072 **presence-absence polymorphism in castrating anther-smut fungi: recent gene gains and**
1073 **phylogeographic structure.** *Genome Biology and Evolution* 2018, **10**:1298-1314.
- 1074 93. Krzywinski aMe: **Circos: an information aesthetic for comparative genomics.** *Genome Res*
1075 2009, **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.
- 1078 95. Perlin MH, Amselem J, Fontanillas E, Toh SS, Chen Z, Goldberg J, Duplessis S, Henrissat B, Young
1079 S, Zeng Q, et al: **Sex and parasites: Genomic and transcriptomic analysis of *Microbotryum***
1080 ***lychnidis-dioicae*, the biotrophic and plant-castrating anther smut fungus.** *BMC genomics*
1081 2015, **16**:1-24.
- 1082 96. Bray NL, Pimentel H, Melsted P, Pachter L: **Near-optimal probabilistic RNA-seq quantification.**
1083 *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.
- 1086 98. Yang Z, Nielsen R: **Estimating synonymous and nonsynonymous substitution rates under**
1087 **realistic evolutionary models.** *Molecular Biology and Evolution* 2000, **17**:32-43.
- 1088 99. Bouckaert R, Vaughan TG, Barido-Sottani J, Duchêne S, Fourment M, Gavryushkina A, Heled J,
1089 Jones G, Kühnert D, De Maio N, et al: **BEAST 2.5: An advanced software platform for Bayesian**
1090 **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**
 1092 **fungal pathogen species that are specialized to different hosts: allopatric divergence and**
 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 102. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R:
 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.

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

Amino acid	Optimal 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				
Ile	ATC	ATT	ATA			
Leu	CTC	CTT	CTA	CTG	TTA	TTG
Lys	AAG	AAA				
Phe	TTC	TTT				
Pro	CCC	CCT	CCA	CCG		
Ser	TCG	TCT	TCC	TCA	AGC	AGT
Thr	ACC	ACT	ACA	ACG		
Tyr	TAC	TAT				
Val	GTC	GTT	GTA	GTG		
Stp	TAA	TGA	TAG			

1103

1104

1105 **Figures Legends**

1106 **Figure 1: Timetrees and rearrangement scenarios of recombination suppression events in**

1107 **the *Microbotryum* genus.** A) Densitree plot of 1,000 random trees with the best posterior

1108 probabilities based on the concatenation of single-copy autosomal genes (green traces) and

1109 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
1111 its mating-type loci) and *M. silenes-acaulis* (involving a very different rearrangement of the PR
1112 chromosome, with therefore very few gene shared with other species in its non-recombining
1113 region). Tips are labelled with scientific names of *Microbotryum* and of their hosts. Photos of
1114 diseased flowers by M.E. Hood and Eugene Popov @iNatutalist, cropped. CC BY 4.0. Inset:
1115 density of estimates of time since recombination suppression for *M. v. paradoxa* strata. Note
1116 the different scales for young (pink and white) and old strata (black, brown and light brown).
1117 Approximate position of the corresponding regions in the ancestral chromosome order is shown
1118 (HD chromosome in light blue and PR chromosome in light purple, HD locus in dark blue and
1119 PR locus in dark purple) and the succession of recombination suppression events is pictured
1120 below the rearrangement scenario (see Figure S2). Symbols key is also valid for panel B. B)
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
1123 chromosomal rearrangement and/or recombination suppression are separated by white lines.
1124 Scenarios shared by contiguous species on the tree are depicted only once. Chromosome
1125 arrangement in *M. intermedium* corresponds to the ancestral state [60, 61]. Symbol key is the
1126 same as in the inset of panel A. Rearrangements are indicated by blue shades. C) Consensus
1127 timetrees of single-copy autosomal genes (green trace) and evolutionary strata (with their
1128 corresponding color). Timetrees in the upper panel are identical but for the different dates
1129 estimated for the suppression of recombination in different strata of *M. v. paradoxa*. The
1130 timetree in the lower panel is the consensus of 32 concatenated single-copy genes in the non-
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**
1135 **sequenced species.** For each species (*Microbotryum coronariae*, *M. v. melanantha*, *M. v.*
1136 *tatarinowii* and *M. lychnidis-dioicae*), we show the synonymous divergence calculated between
1137 alternative mating types and plotted along the *M. intermedium* mating-type chromosome gene
1138 order. Zero d_s between mating-type chromosomes indicate recombination in these selfing
1139 species and the level of d_s between mating-type chromosomes increases with time since
1140 recombination suppression. Different evolutionary strata, i.e., different steps extending
1141 recombination suppression, are indicated by different colours. Divergence between the a_1 and

1142 a_2 pheromone receptor (PR) genes was too extensive, and d_S could not be calculated (depicted
1143 as “Un” for unalignable).

1144 **Figure 3: Analysis of the frequency of optimal codons in the a_1 genome of *Microbotryum***
1145 *silenes-dioicae*. A) Distribution of the frequency of optimal codons across genomic
1146 compartments. B) Distribution of the frequency of optimal codons predicted by the logistic
1147 regression, accounting for genomic compartment, difference of GC in coding and intronic
1148 sequences and their interactions as factor. In A) and B), boxplot grey shade refers to the
1149 genomic compartment. For each genomic compartment, the red dot indicates the mean
1150 frequency of optimal codons, the sample size (N) is labelled on top of the corresponding
1151 boxplot, as well as the significance level of the difference between a given genomic
1152 compartment and the autosomes (NS: non-significant; ***: p-value < 0.001) from the logistic
1153 regression. Note the difference in the ranges of Y axes for panels A and B. Similar analyses for
1154 all species studied here are presented in Figure S6 and S7 in the supplementary materials. C)
1155 Frequency of optimal codons predicted by the logistic regression plotted against the difference
1156 in GC content between coding and intronic sequences. From the logistic regression, significant
1157 interactions between a genomic compartment and the autosomes are indicated by solid line and
1158 filled dots, while non-significant interactions are indicated by dotted line and hollow dots.
1159 “Reference” refers to the reference considered in the logistic regression. D) Predicted effect of
1160 GC-biased gene conversion and selection on frequency of optimal codons. The figure shows
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
1164 frequency of optimal codons in autosomes (1) decreases with the increase of difference in GC
1165 content and the frequency of optimal codons are lower in non-recombining regions (2, 3, 4).

1166 **Figure 4: Tempo of degeneration in *Microbotryum* fungi.** A) Frequency of optimal codons
1167 as a function of time since recombination suppression (in million years). B) Non-synonymous
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
1171 coloured according to species and the dot fill colours correspond to genomic compartments.
1172 Predictions from the models are show as blue line and the 95% confidence interval of the
1173 prediction as grey background.

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**

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

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 a₁ and a₂ mating-type
1193 chromosomes (left top), synteny plots between a₁ mating-type chromosome and a₁ mating-type
1194 chromosome of *M. intermedium* (right panel).

1195 **Figure S3: Synteny and allelic divergence analysis between mating-type chromosomes of**
1196 **alternative mating types for the species *M. scabiosae* (A) and *Microbotryum v. caroliniana***
1197 **(B).** For each species, we show synteny plots between a₁ and a₂ mating-type chromosomes (left
1198 top panel), synteny plots between a₁ and a₂ mating-type chromosomes of *M. intermedium* (right
1199 top panel), and the synonymous divergence calculated between alternative mating types and
1200 plotted along the *M. intermedium* mating-type chromosome gene order (bottom panel). Zero d_s
1201 between mating-type chromosomes indicate recombination in these selfing species and the level
1202 of d_s between mating-type chromosomes increases with time since recombination suppression.
1203 On the top left figure in B), the light blue evolutionary stratum is indicated by a light blue box
1204 on the outer track.

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
1208 alternative mating types is plotted along the *M. intermedium* mating-type chromosome
1209 ancestral-like gene order. Zero d_s between mating-type chromosomes indicate recombination
1210 in these selfing species and the level of d_s between mating-type chromosomes increases with
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 a_1 and a_2 pheromone receptor (PR) genes was too extensive, and d_s could not be calculated
1214 (depicted as “Un” for unalignable).

1215

1216 **Figure S5: Within-correspondence analysis plots.** A) WCA plots for the a_1 and a_2 genomes of
1217 *Microbotryum intermedium* (resp. left and right top panel) and a_1 and a_2 genomes of *M.*
1218 *lychnidis-dioicae* (resp. left and right bottom panel). B) Percentage of variance explained for
1219 each correspondence analysis axis for the a_1 and a_2 genomes of *M. intermedium* (resp. left and
1220 right top panel) and a_1 and a_2 genomes of *M. lychnidis-dioicae* (resp. left and right bottom
1221 panel).

1222 **Figure S6 Data distribution across genomic compartments in *Microbotryum*.** A)
1223 Distribution of the frequency of optimal codons. B) Distribution of the difference in GC content
1224 between coding sequences and introns. C) Distribution of the expression level of *M.*
1225 *intermedium* orthologous genes. The gene expression level distribution was truncated at the 90th
1226 percentile for visualisation purpose and was not represented for the genomes a_1 and a_2 *M.*
1227 *intermedium*. Boxplot colours refer to the genomic compartment.

1228

1229 **Figure S7: Analysis of the frequency of optimal codons in all *Microbotryum* genomes using**
1230 **the genomic compartments considered in the main text.** A) Distribution of the frequency of
1231 optimal codons across genomic compartments. B) Distribution of the predicted frequency of
1232 optimal codons predicted by the logistic regression. In A) and B), boxplot colours refer to the
1233 genomic compartment. For each genomic compartment, the red dot indicates the mean
1234 frequency of optimal codons, the sample size (N) is labelled on top of the corresponding
1235 boxplot, as well as the significance level of the difference between a given genomic

1236 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**
1243 **all the genomic compartments.** A) Distribution of the frequency of optimal codons across
1244 genomic compartments. B) Distribution of the predicted frequency of optimal codons predicted
1245 by the logistic regression. In A) and B), boxplot colours refer to the genomic compartment. For
1246 each genomic compartment, the red dot indicates the mean frequency of optimal codons, the
1247 sample size (N) is labelled on top of the corresponding boxplot, as well as the significance level
1248 of the difference between a given genomic compartment and the autosomes (NS: non-
1249 significant; “.”: p-value < 0.1; *: p-value < 0.05; **: p-value < 0.01; ***: p-value < 0.001). C)
1250 Frequency of optimal codons predicted by the logistic regression along the difference of GC
1251 content between coding and intronic sequences. Significant interactions between a genomic
1252 compartment and the autosomes are indicated by a solid line and a filled dot, while non-
1253 significant interactions are indicated by a dotted line and a hollow dot.

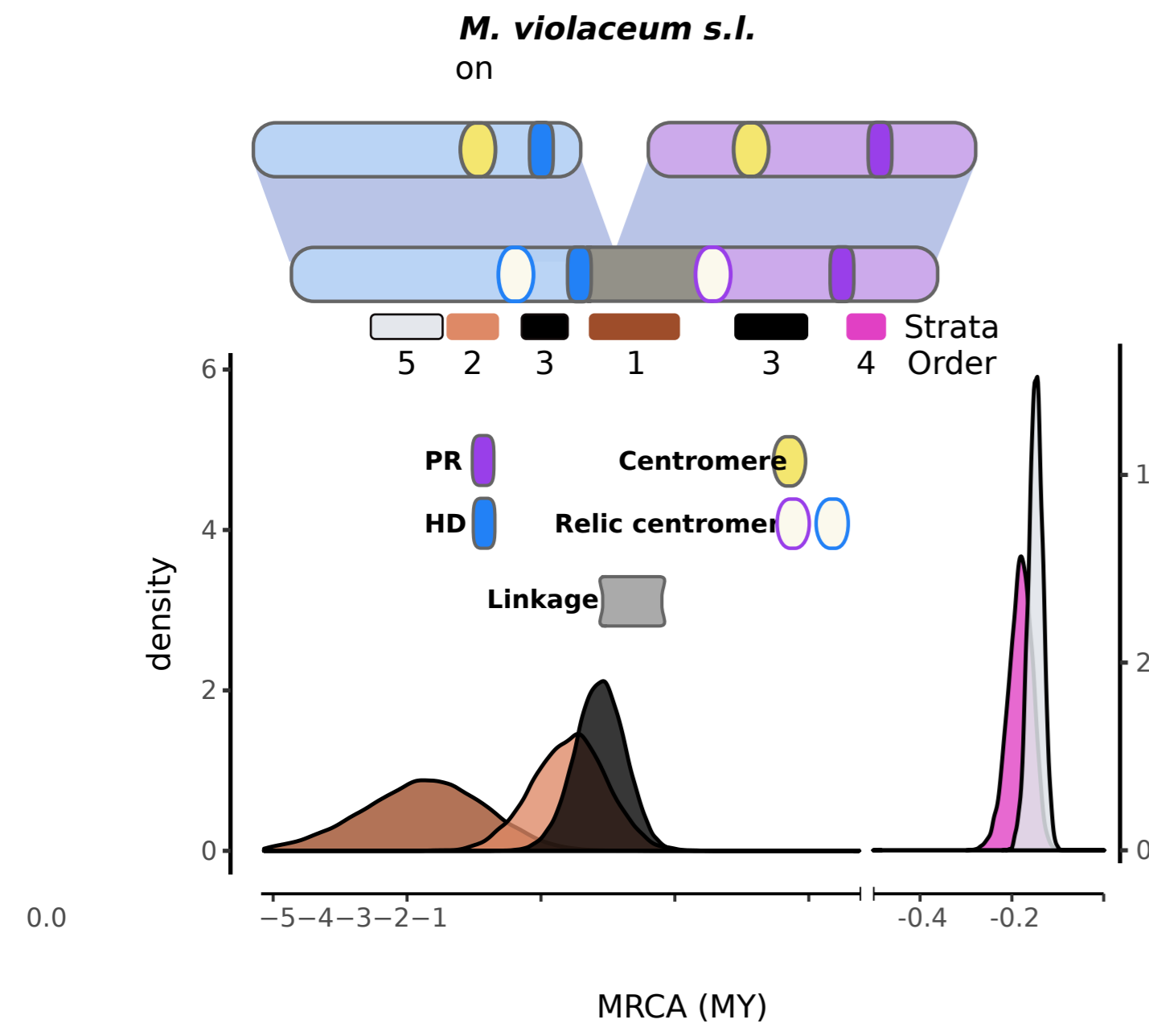
1254

1255

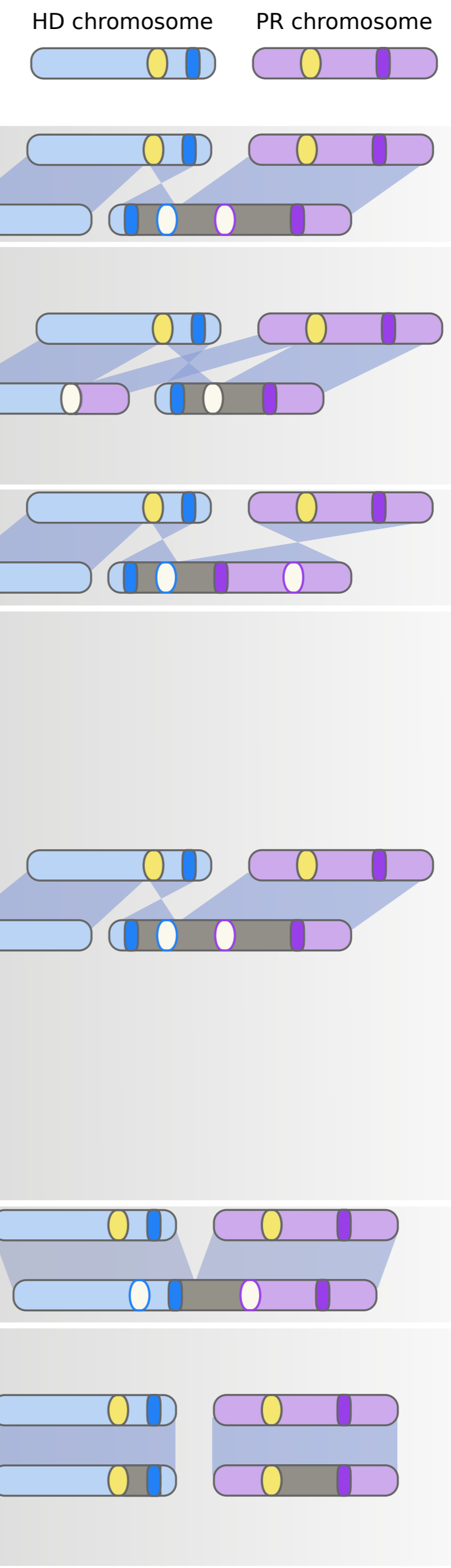
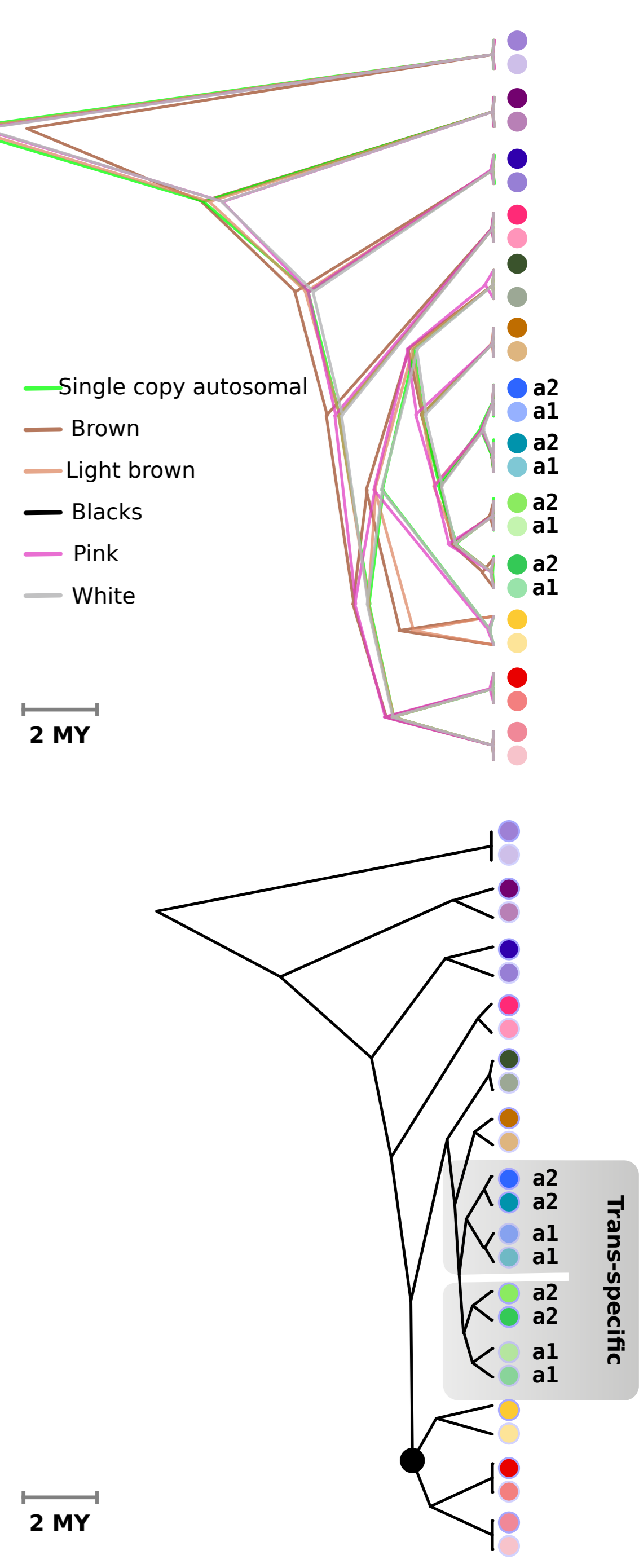
1256

A

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.



-25 -20 -15 -10 -5 MY

B**C**

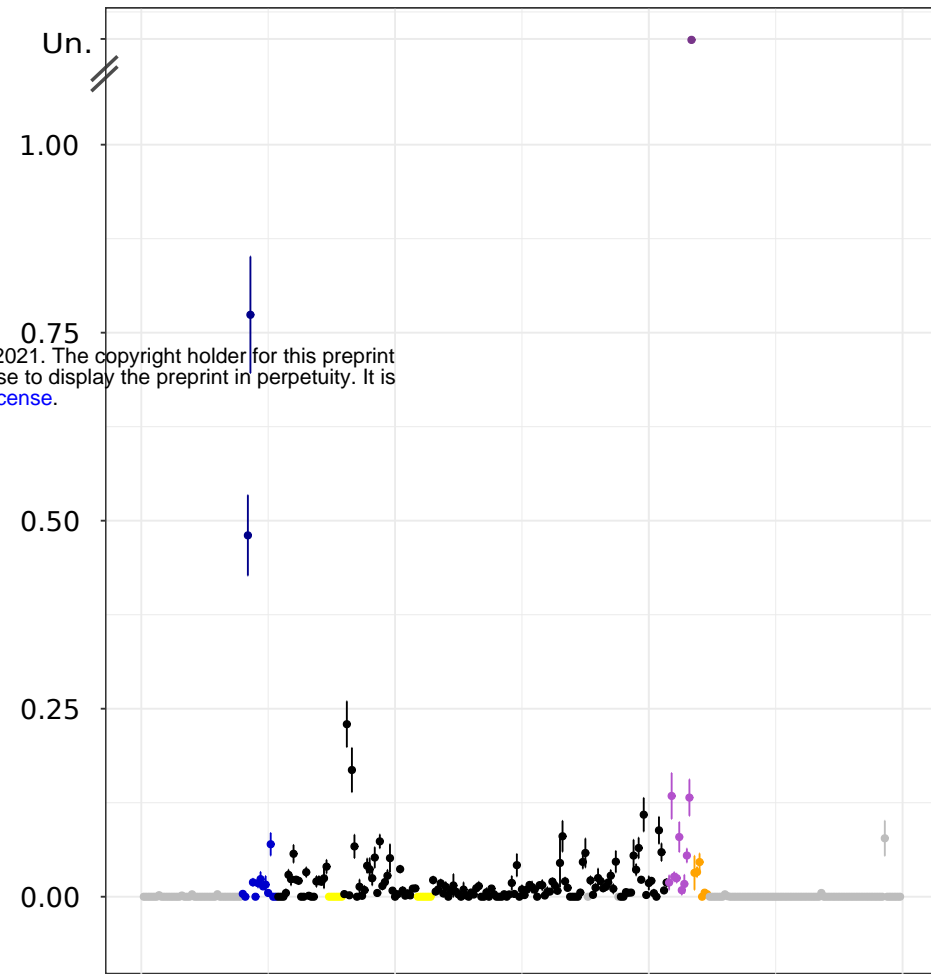
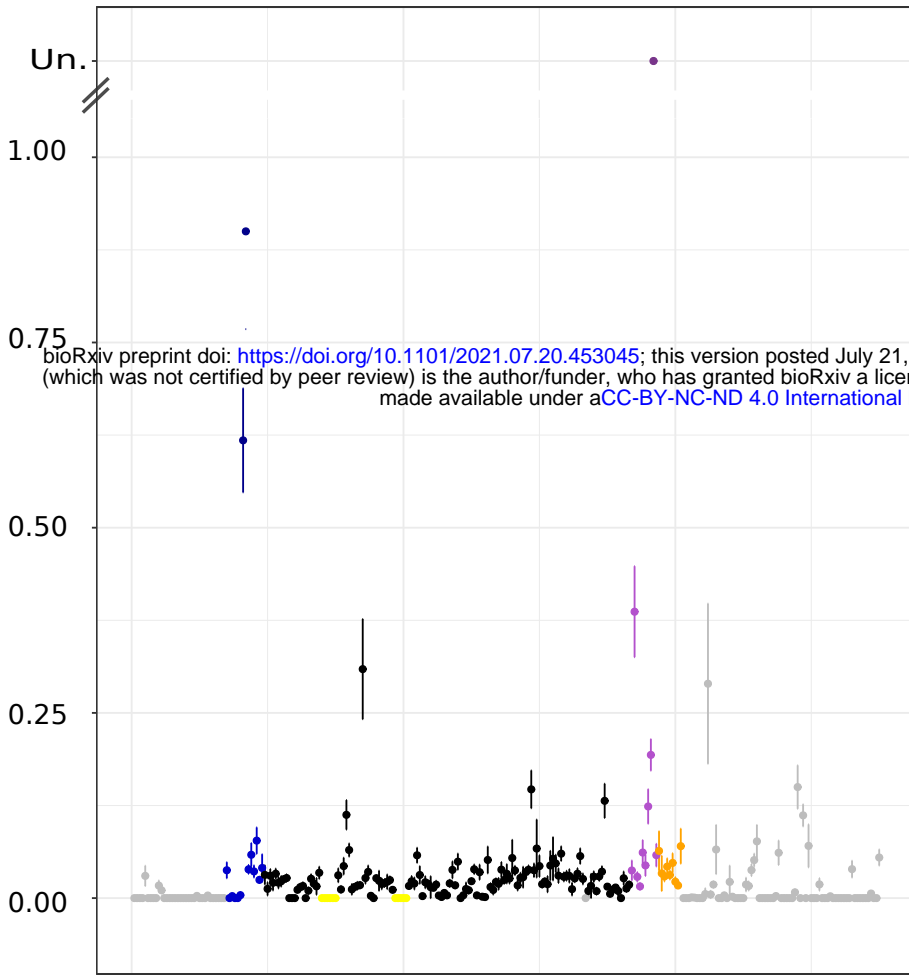


M. coronariae
on *Silene flos-cuculi*



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

$a_1 - a_2$ synonymous divergence

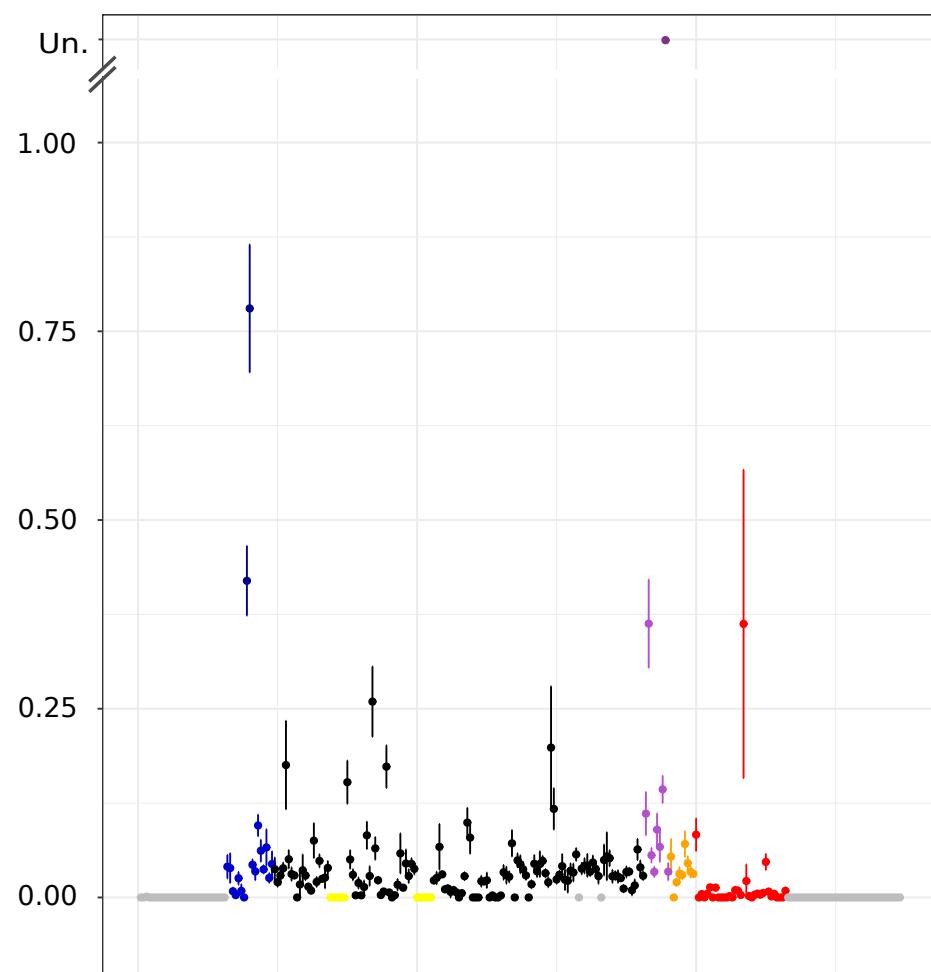
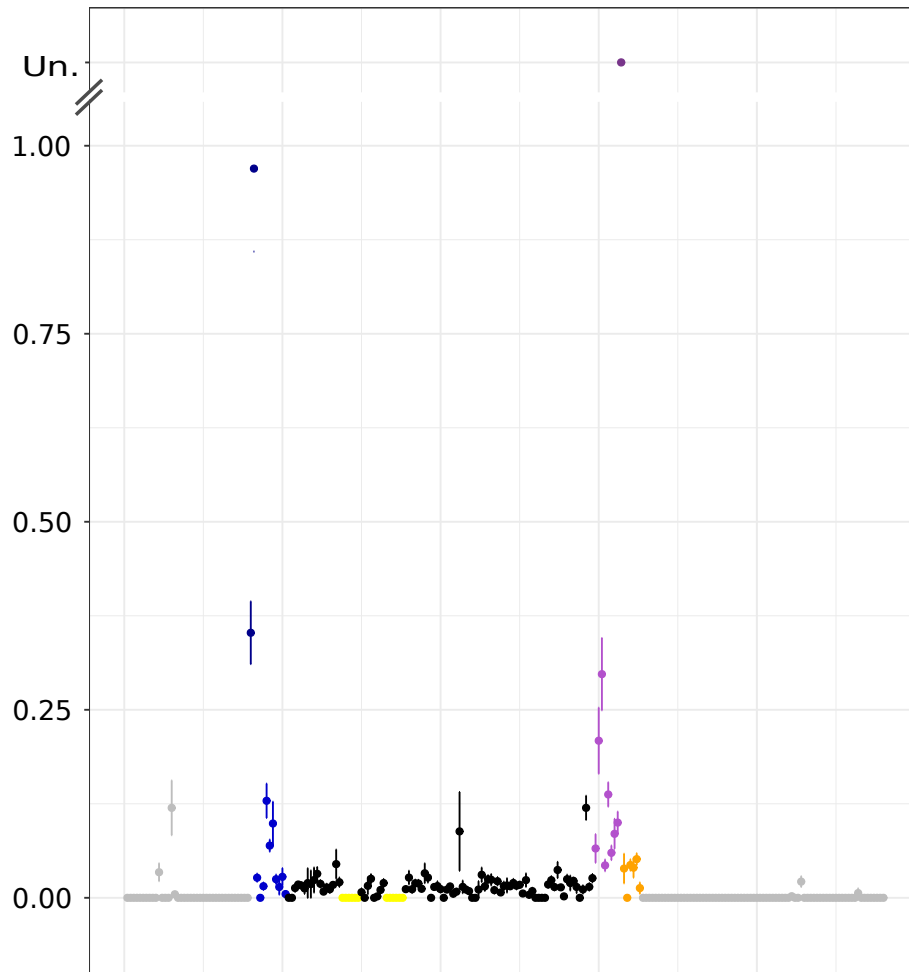


M. violaceum s.l.
on *Silene tatarinowii*

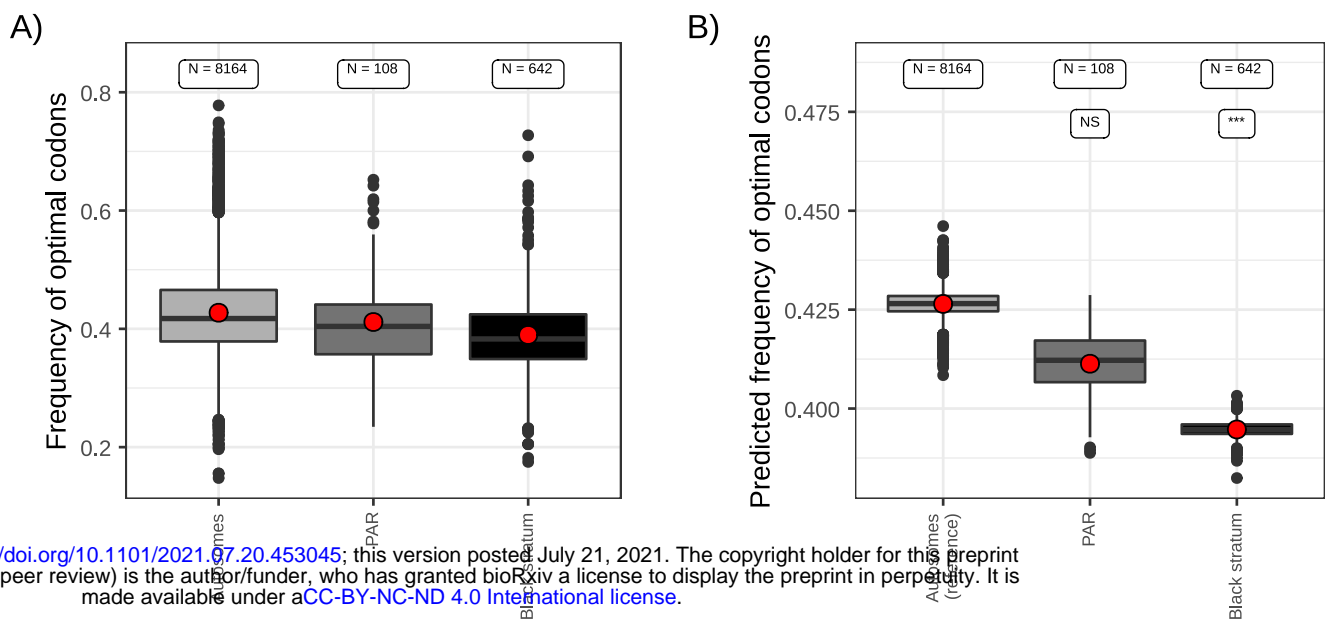


M. lychnidis-diocae
on *Silene latifolia*

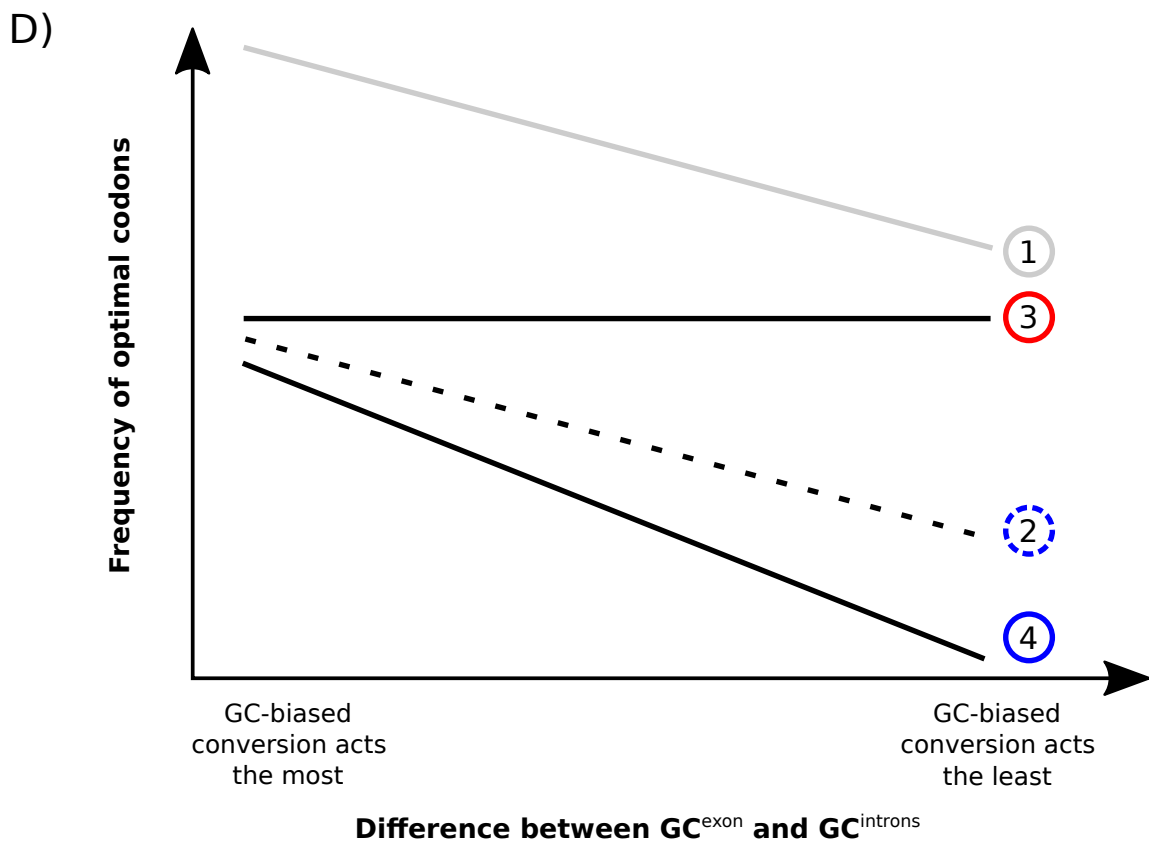
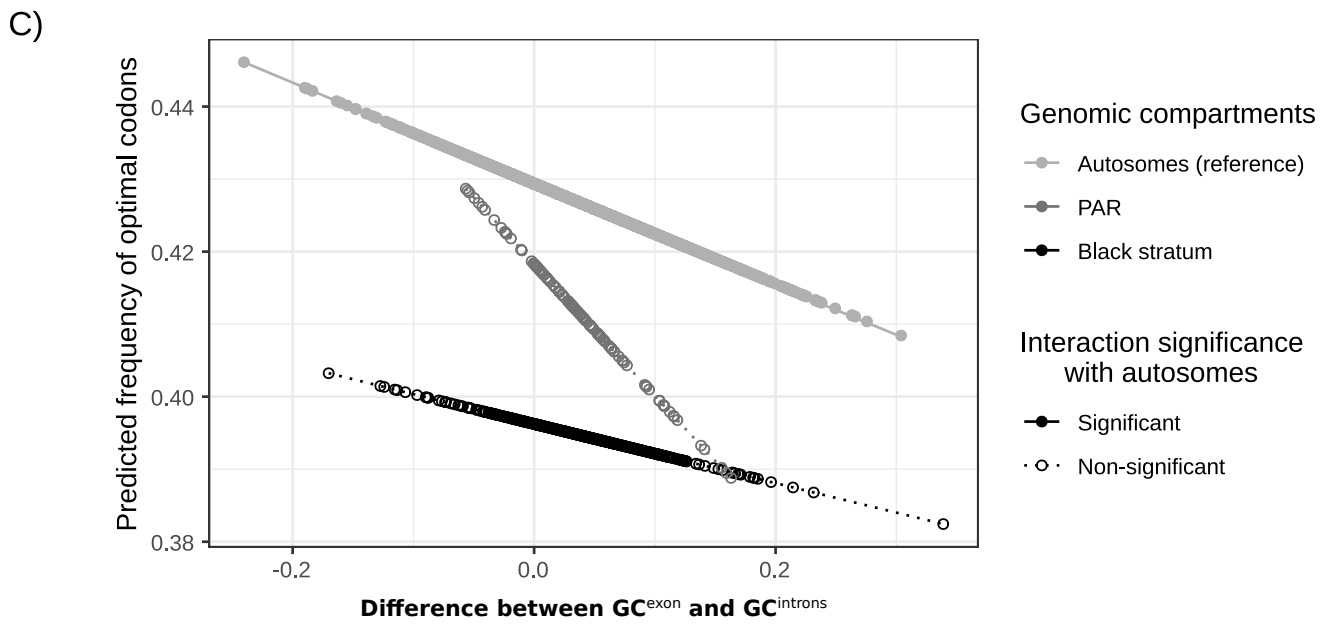
$a_1 - a_2$ synonymous divergence



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



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 a [CC-BY-NC-ND 4.0 International license](https://creativecommons.org/licenses/by-nc-nd/4.0/).



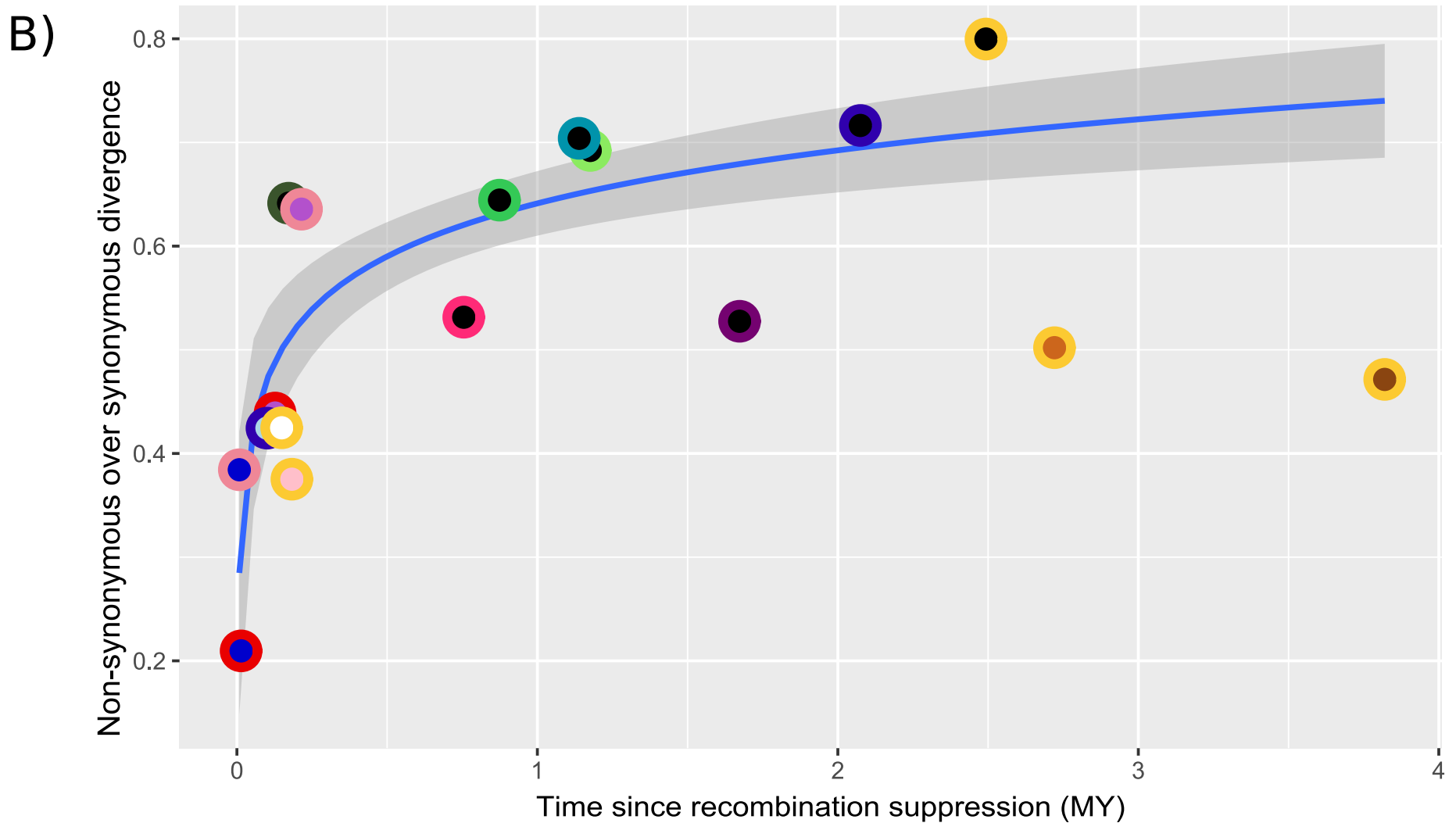
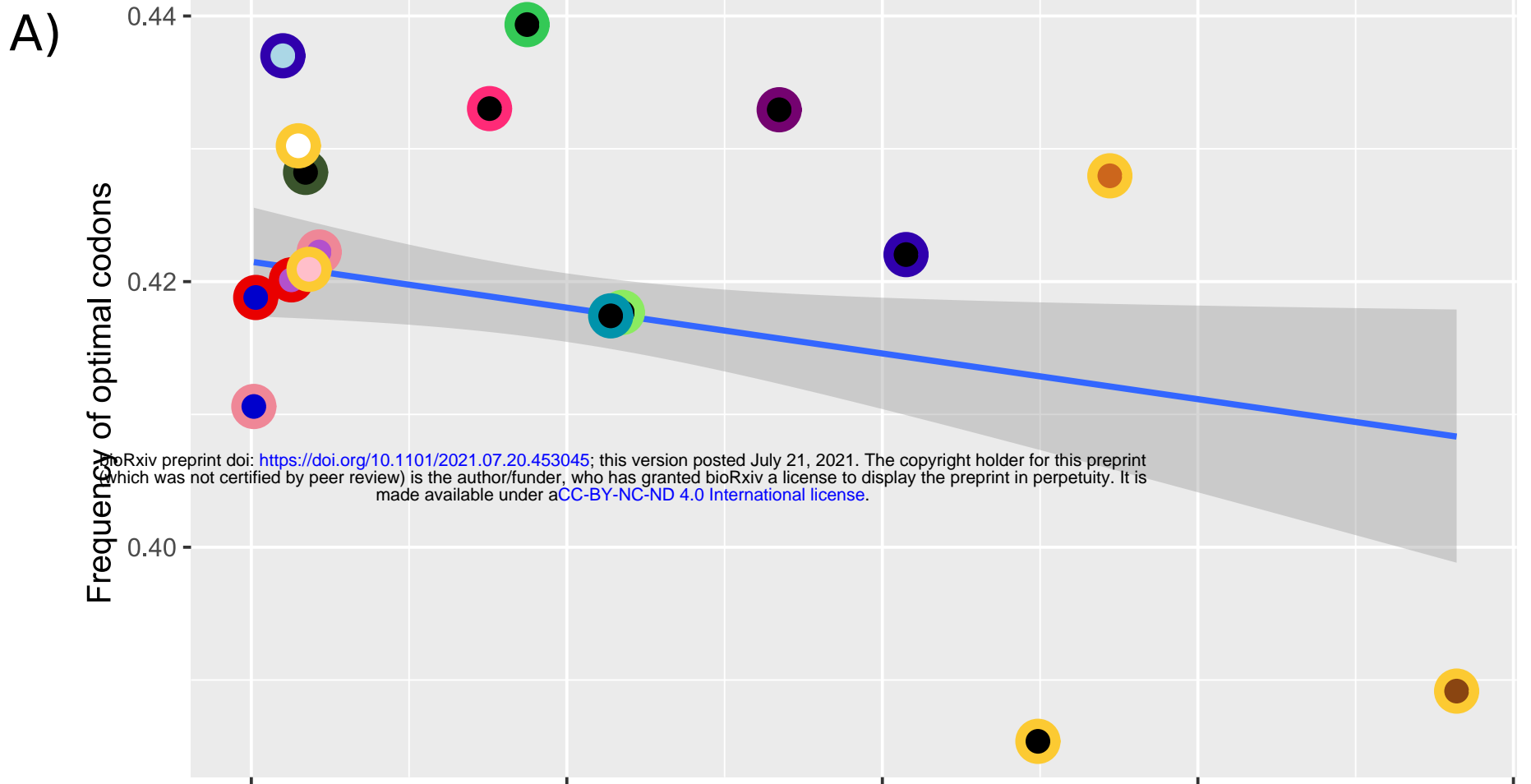
① **In autosomes**, the reduction of GC-biased gene conversion results in a decrease in frequency of optimal codons.

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

② **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.

③ **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.

④ **Significant and negative interaction.** Less GC-biased gene conversion in non-recombining regions alone can explain the reduced frequency of optimal codons compared to autosomes. Less efficient selection may also acts.



Legend

- Black stratum formed prior to *M. lychnidis-dioicae* - *M. silenes-dioicae* speciation
 - Black stratum formed prior to *M. coronariae* - *M. violaceum* (s.s.) speciation
 - Non-recombining region linking HD locus and centromere in *M. lagerheimii*
 - Non-recombining region linking PR locus and centromere in *M. lagerheimii*
 - Black stratum in *M. silenes-acaulis*
 - Non-recombining region linking HD locus and centromere in *M. saponariae*
 - Non-recombining region linking PR locus and centromere in *M. saponariae*
 - Black stratum in *M. scabiosae*
 - Black stratum in *M. melanantha*
 - Lightblue stratum in *M. v. caroliniana*
 - Black stratum in *M. v. caroliniana*
 - Black stratum in *M. v. paradoxa*
 - Lightbrown stratum in *M. v. paradoxa*
 - Brown stratum in *M. v. paradoxa*
 - Pink stratum in *M. v. paradoxa*
 - White stratum in *M. v. paradoxa*
 - Black stratum in *M. v. tatarinowii*
- Predicted value from statistical model and the corresponding 95% confidence interval

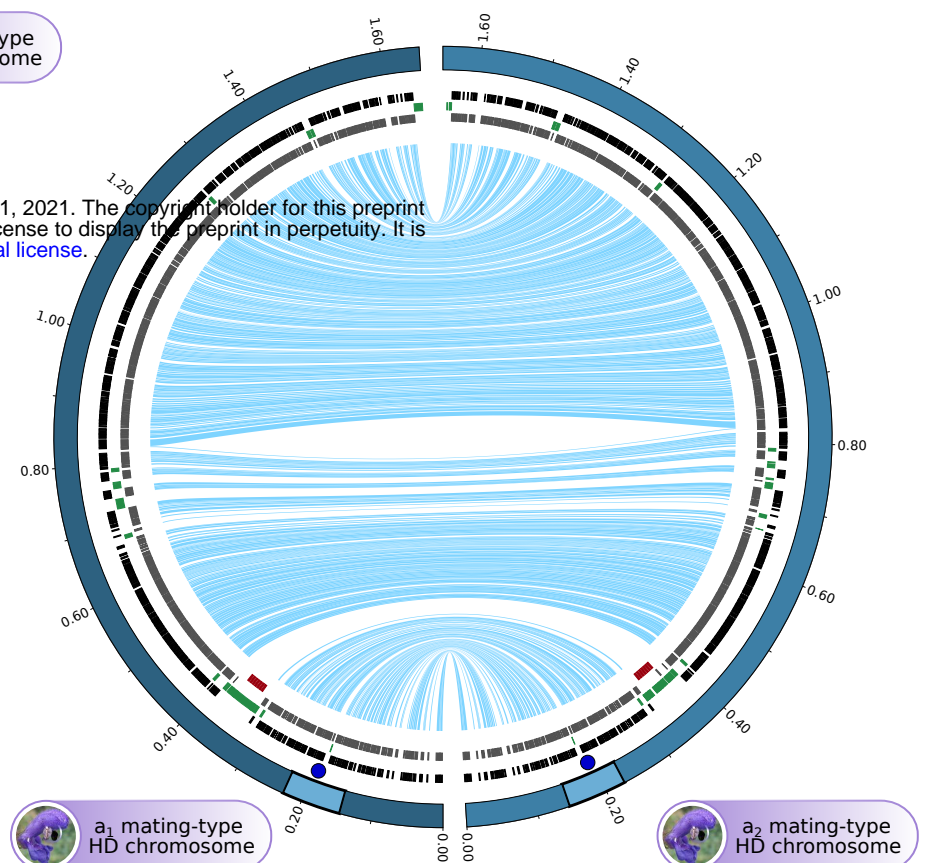
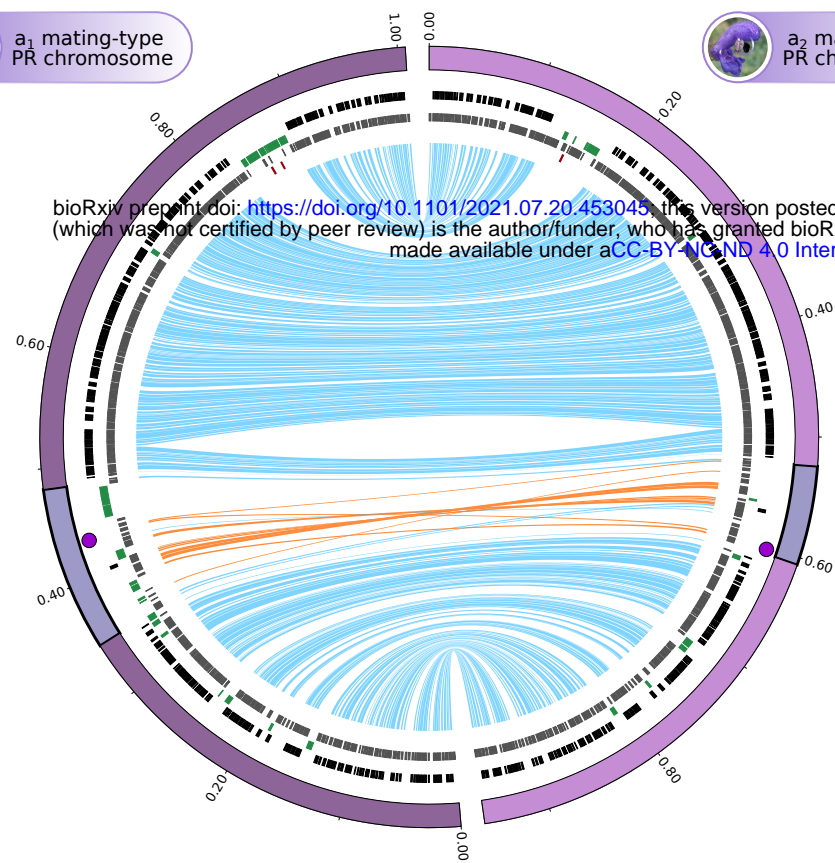
Figure S1



M. intermedium on *Salvia pratensis*

*a*₁ mating-type
PR chromosome

*a*₂ mating-type
PR chromosome



*a*₁ mating-type
HD chromosome

*a*₂ mating-type
HD chromosome

Key to tracks

- Genes with $d_s = 0$
- Transposable elements
- Genes
- Centromeric repeats

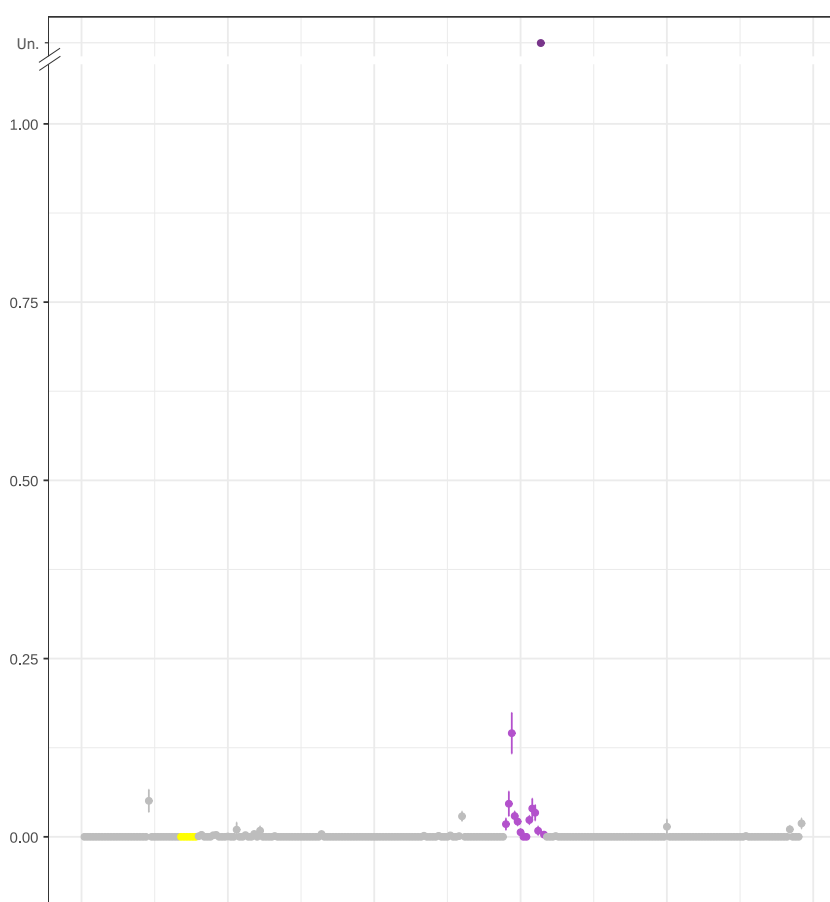
Key to links

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

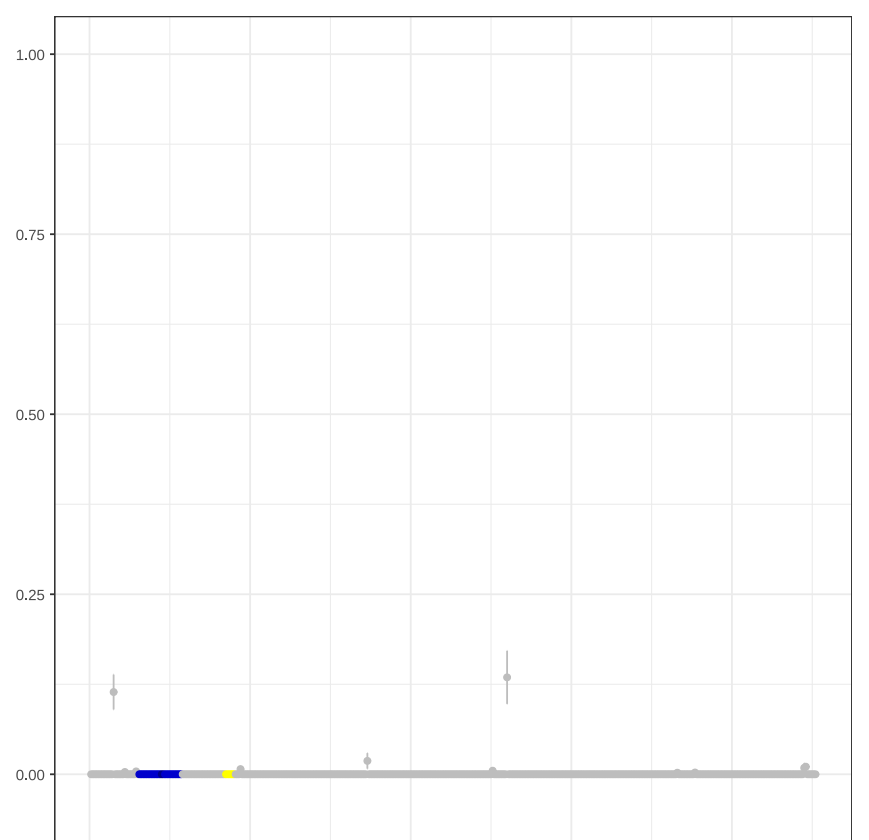
Evolutionary strata

- Purple stratum
- Blue stratum

*a*₁ - *a*₂ synonymous divergence

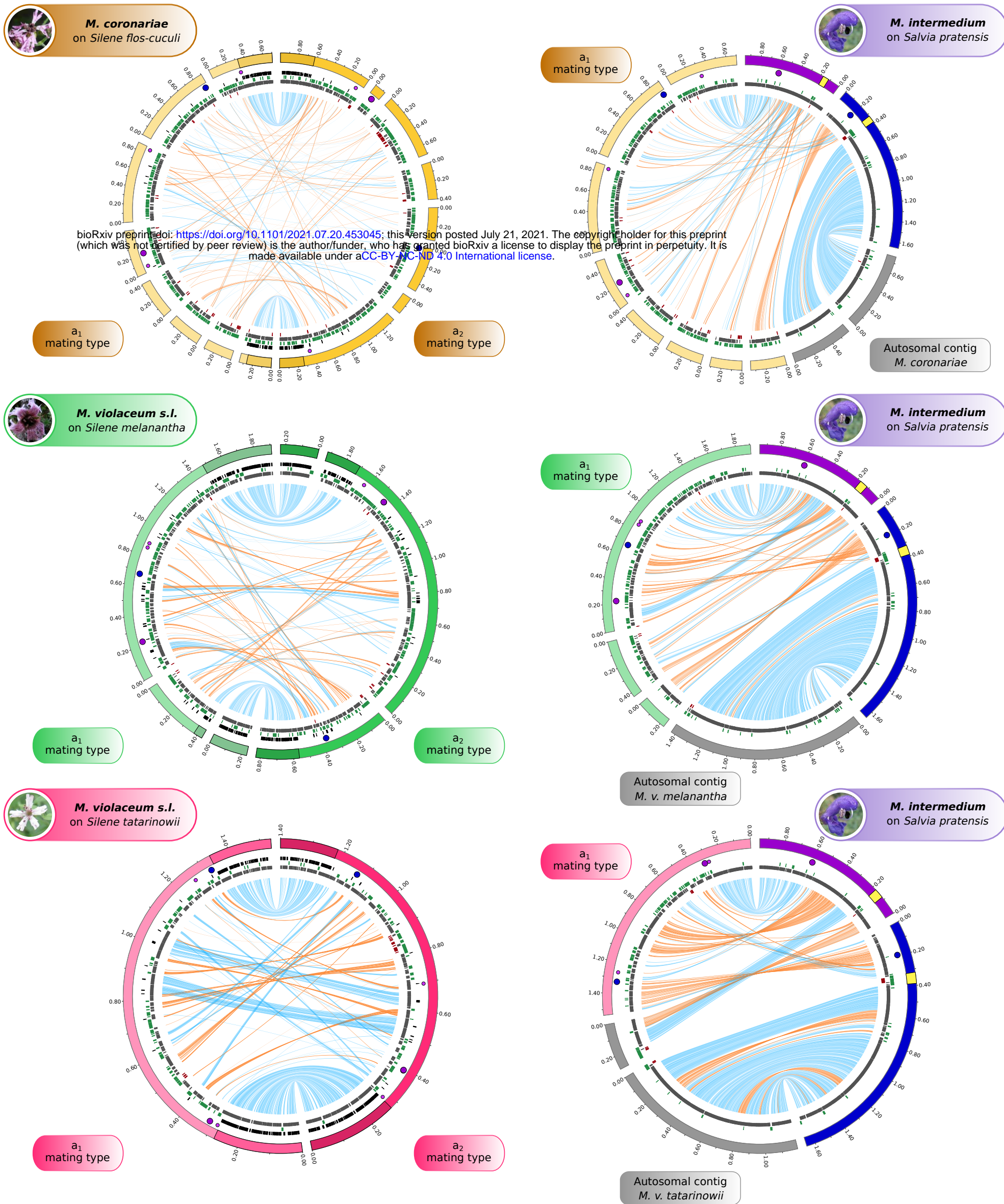


Gene order along
the PR mating-type chromosome
M. intermedium



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

Figure S2



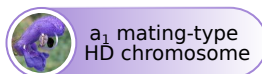
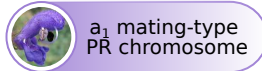
Key to tracks

- Genes with $d_s = 0$
- Transposable elements
- Genes
- Centromeric repeats

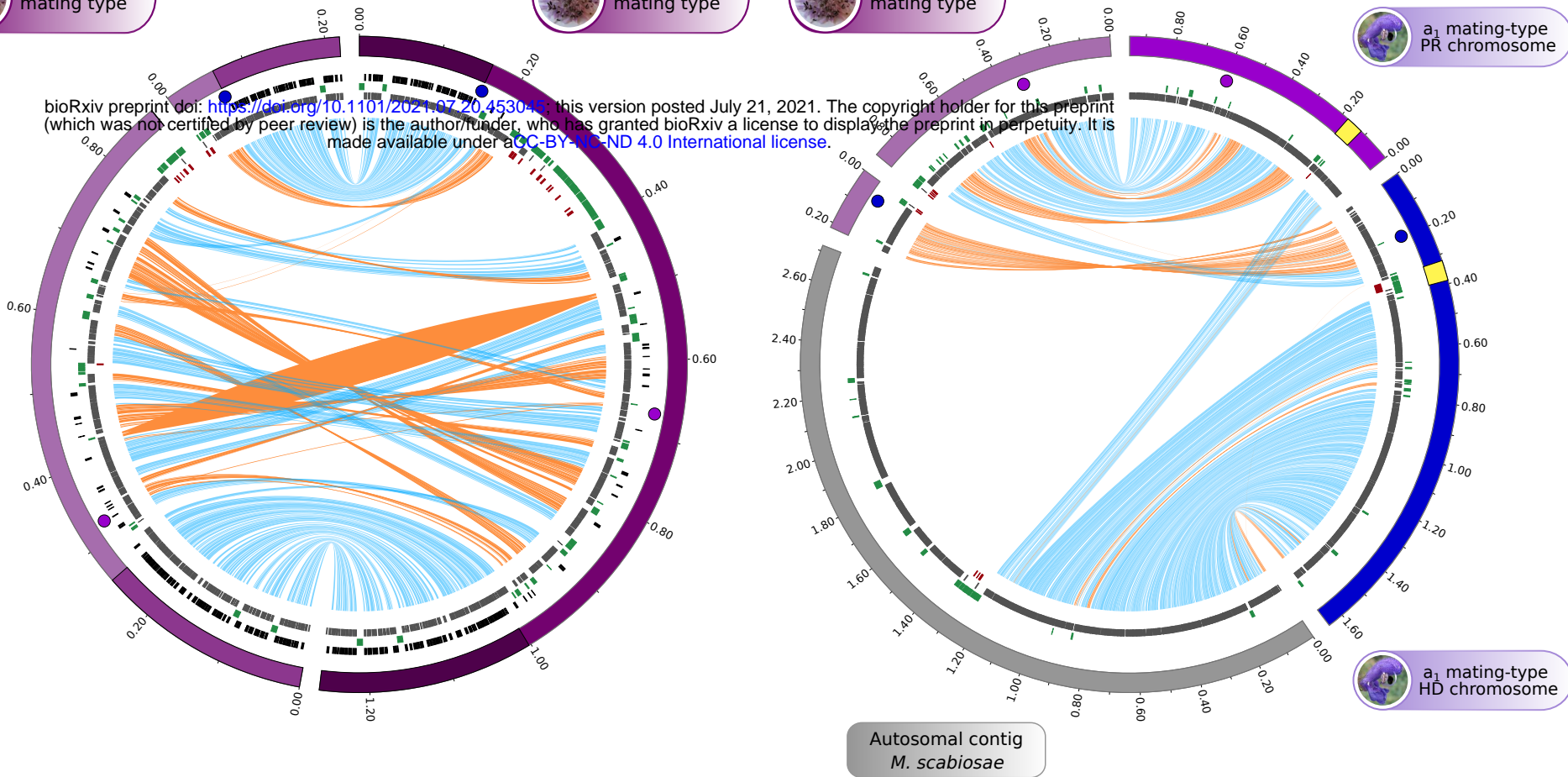
Key to links

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

Figure S3a



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.



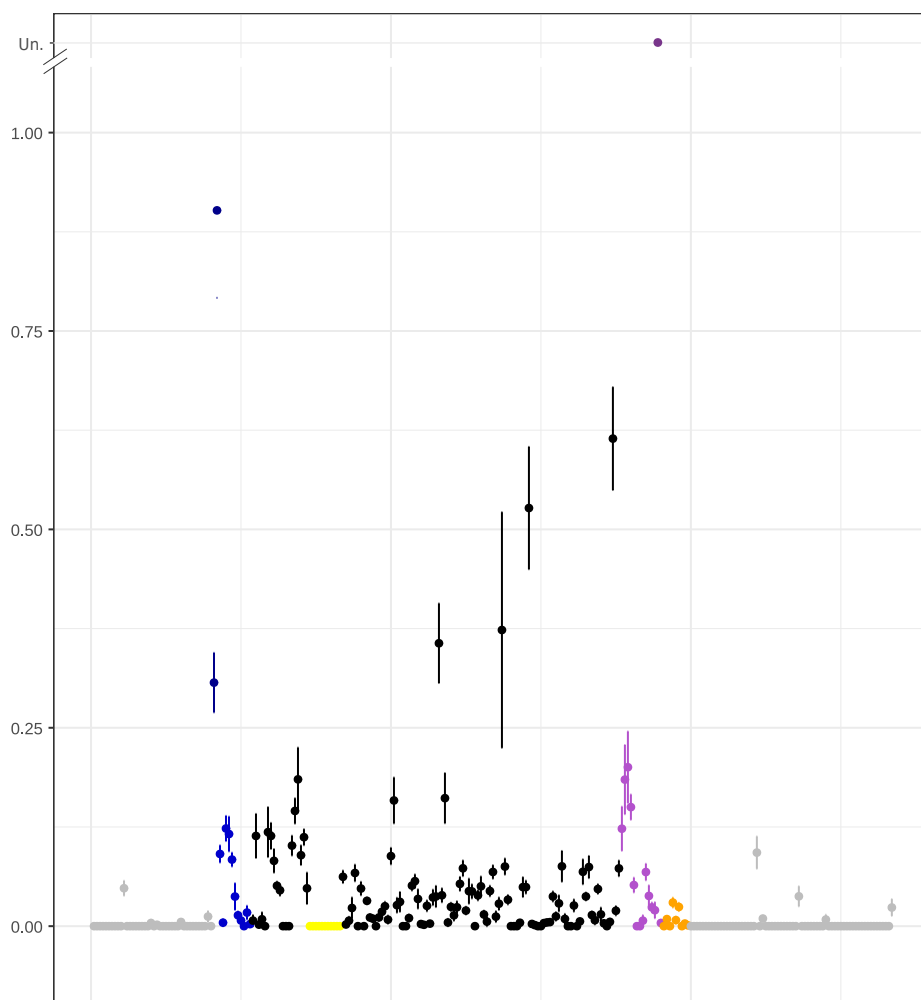
Key to tracks

- Genes with $d_s = 0$
- Transposable elements
- Genes
- Centromeric repeats

Key to links

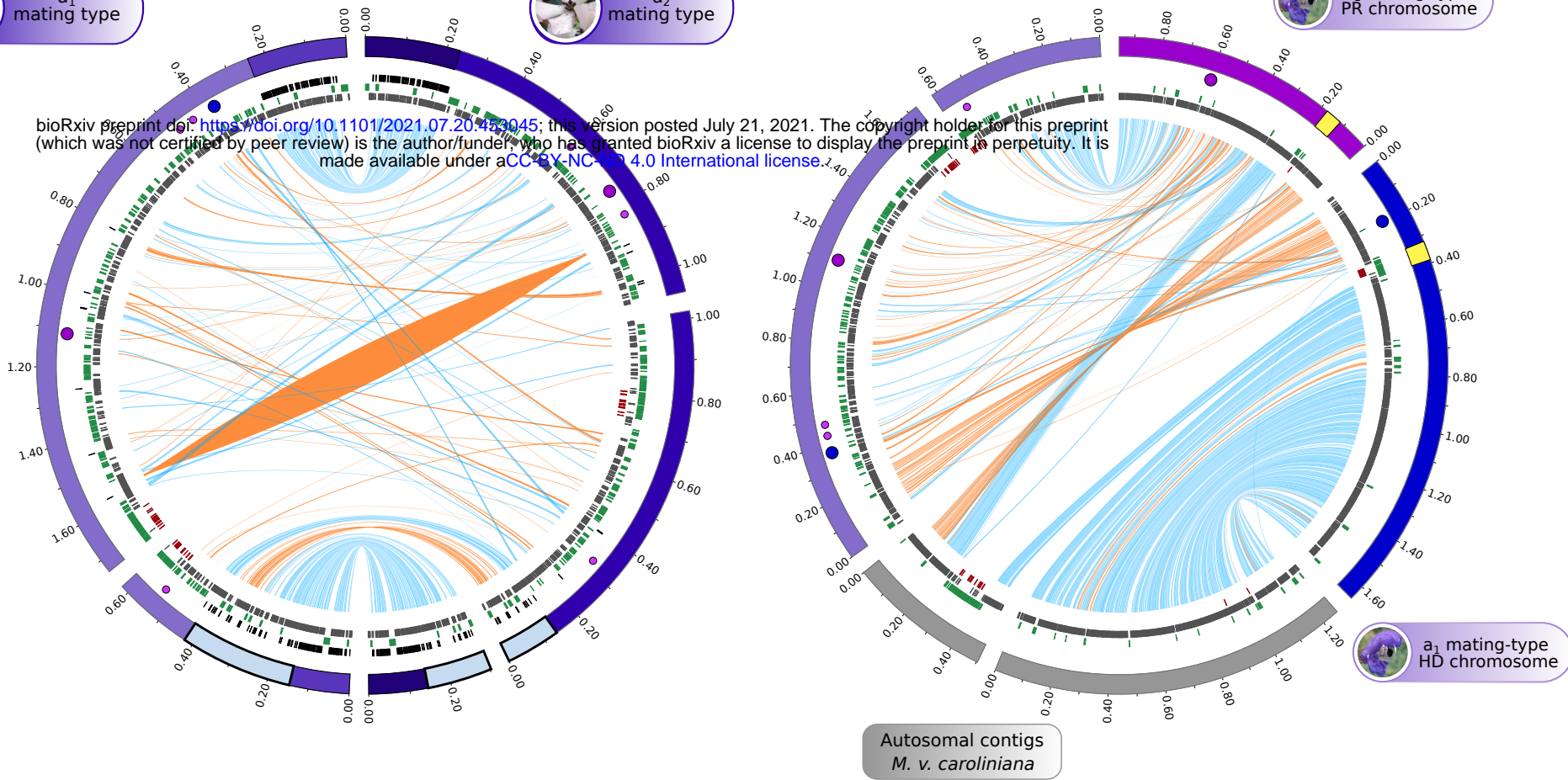
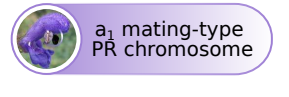
- One-to-one orthologs in the same orientation
- One-to-one orthologs in inverted orientation

a₁ - a₂ synonymous divergence



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

Figure S3b



Key to tracks

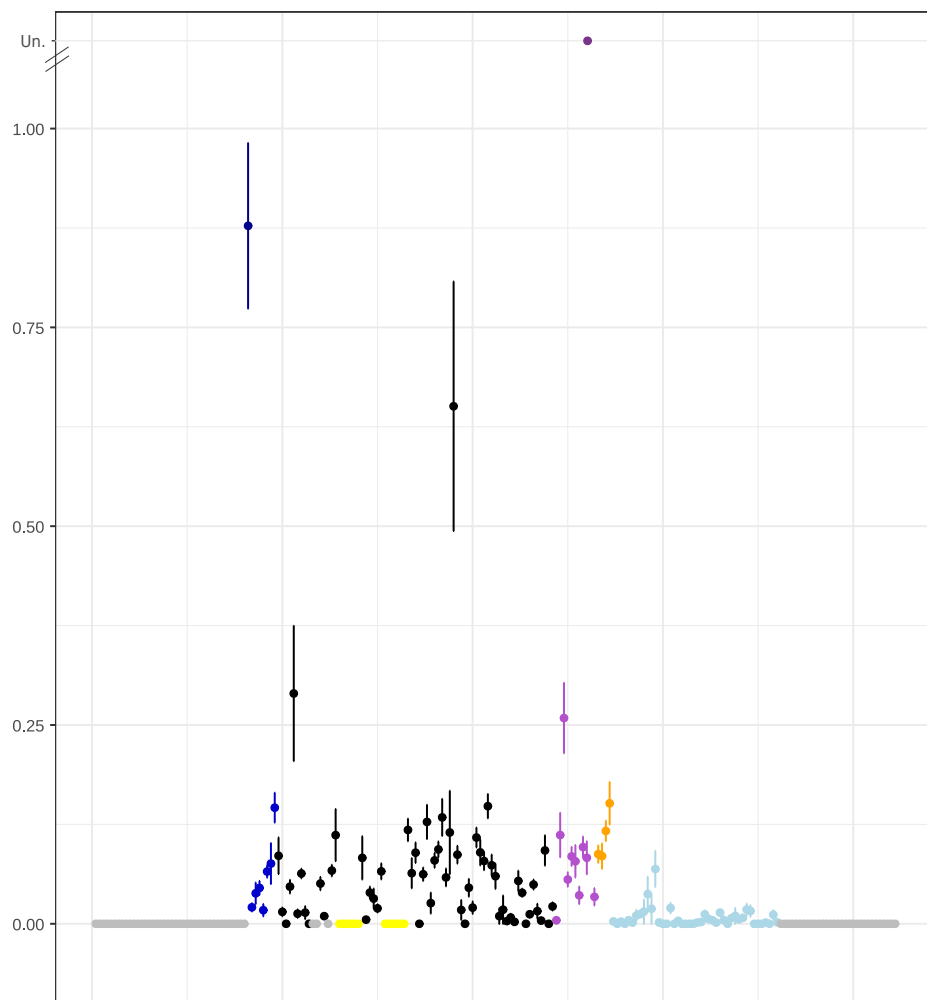
- Genes with $d_s = 0$
- Transposable elements
- Genes
- Centromeric repeats

Key to links

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

Evolutionary strata

- Lightblue stratum



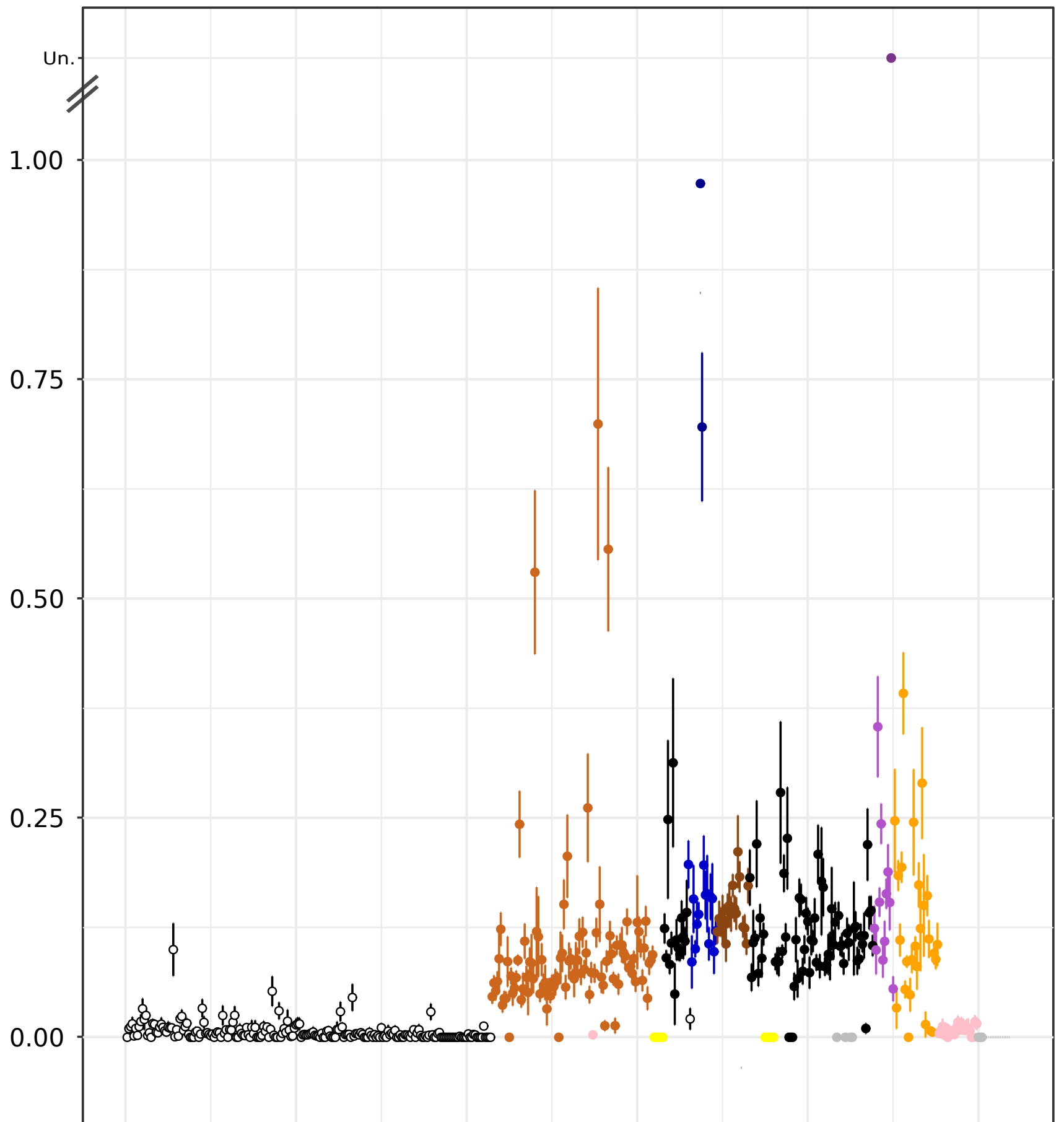
Gene order along the mating-type chromosomes of *M. intermedium*

Figure S4



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

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.



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

$a_1 - a_2$ synonymous divergence



Figure S5a

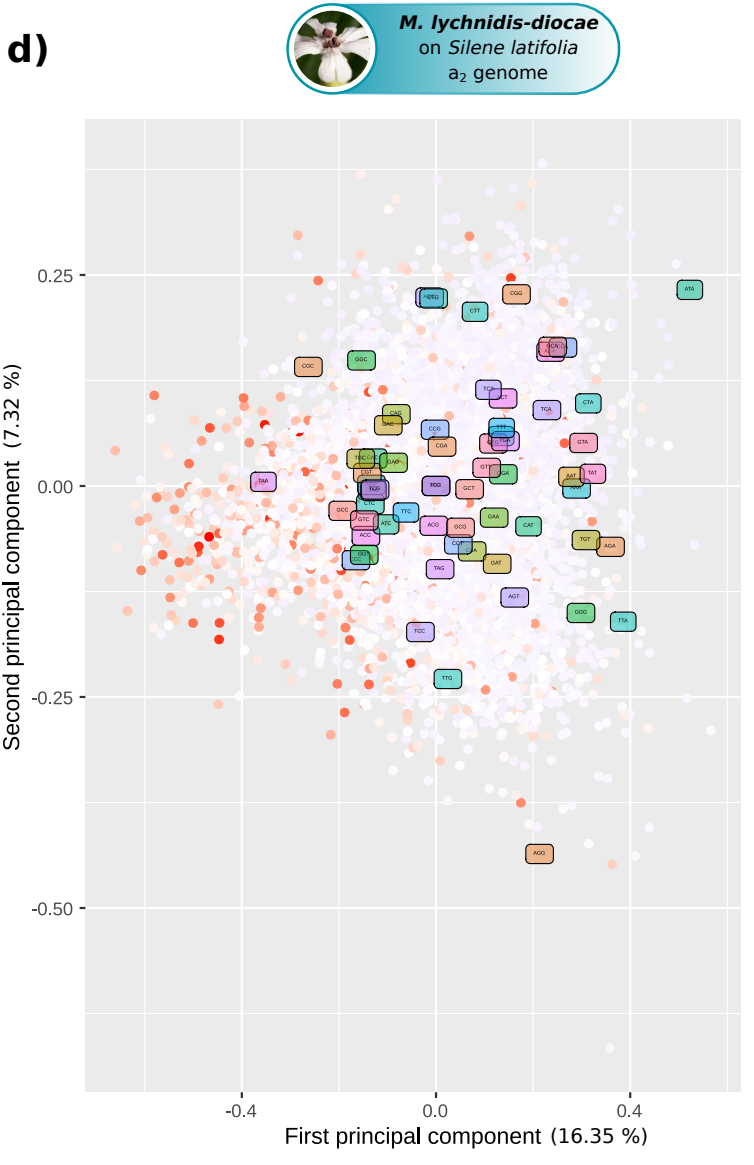
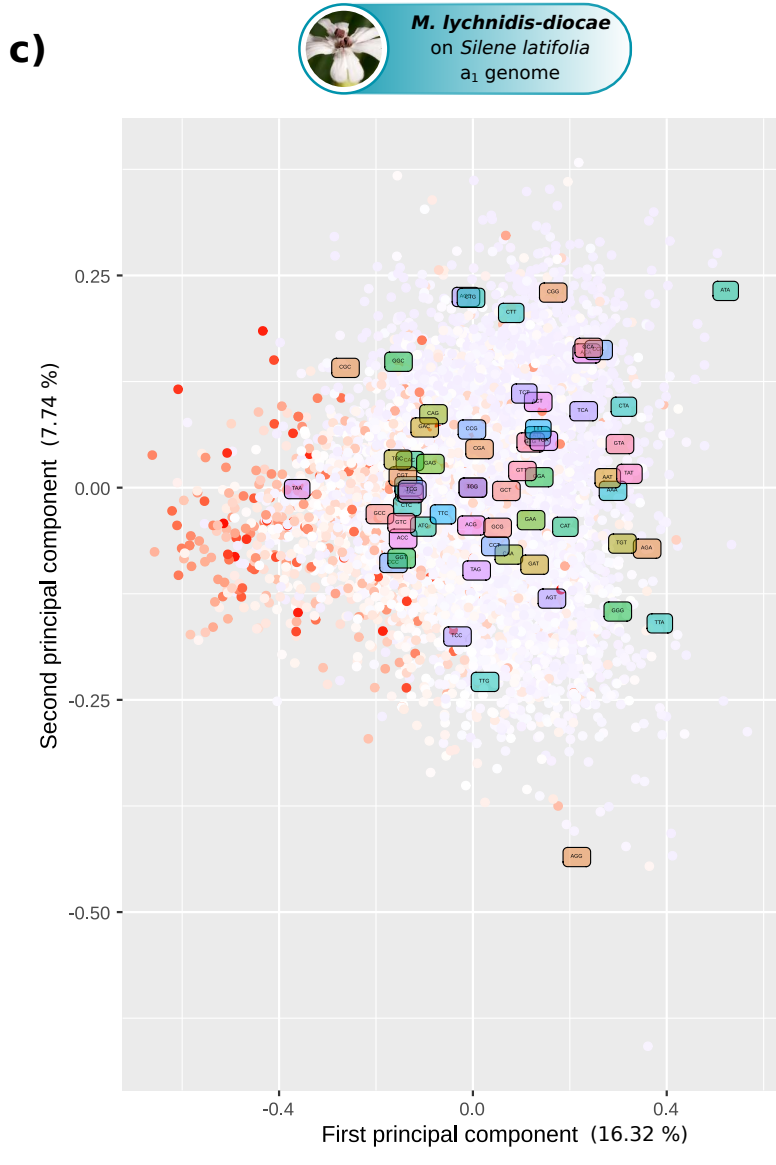
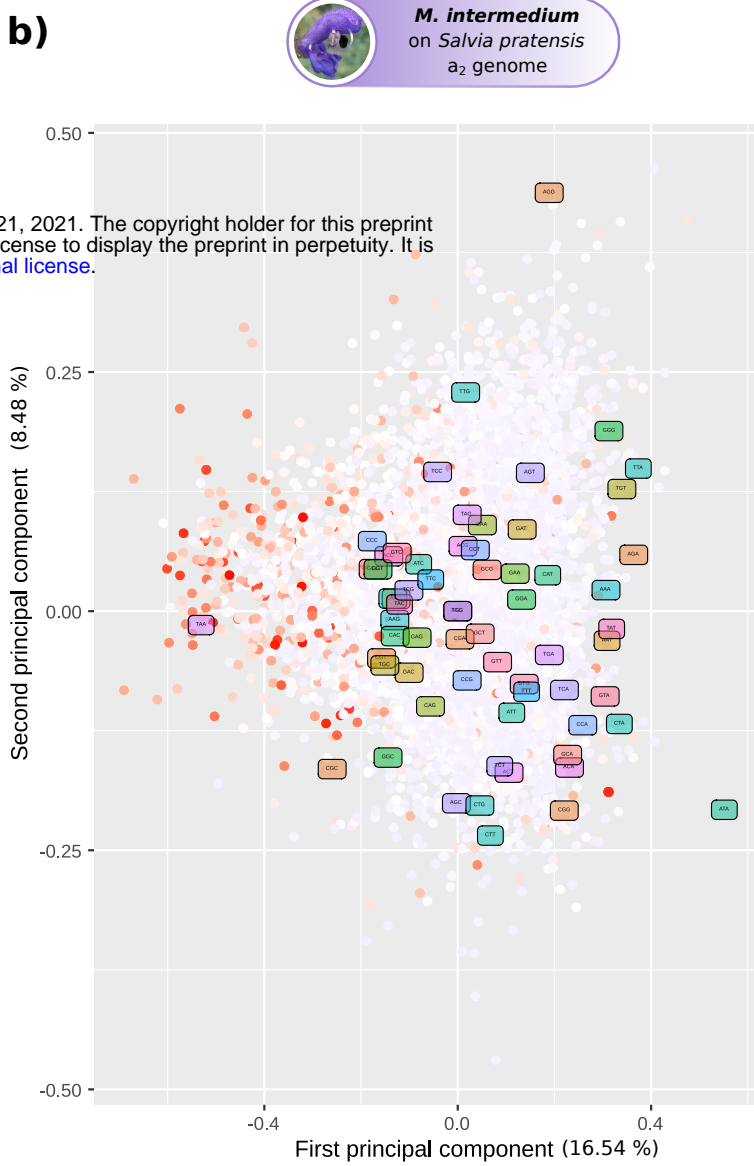
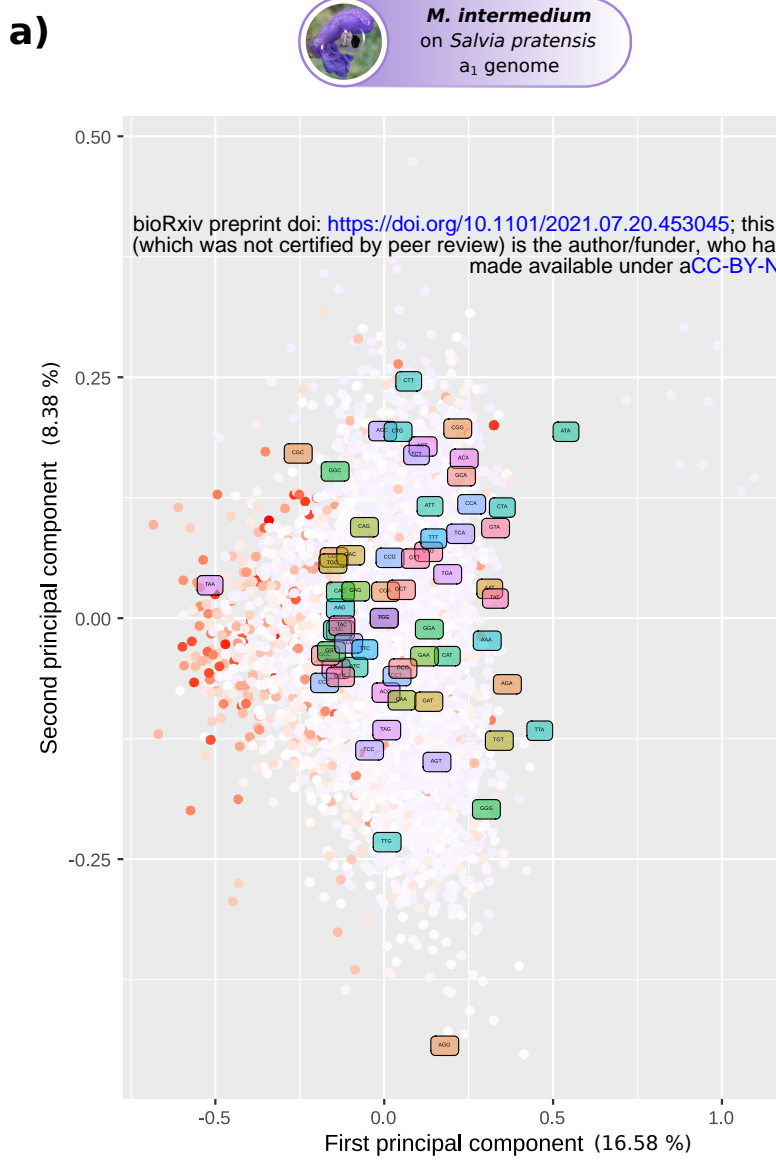
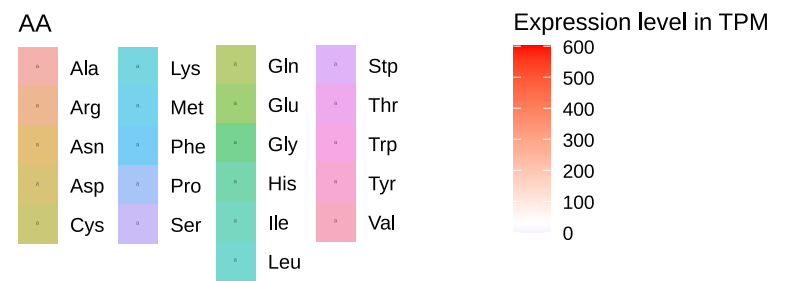
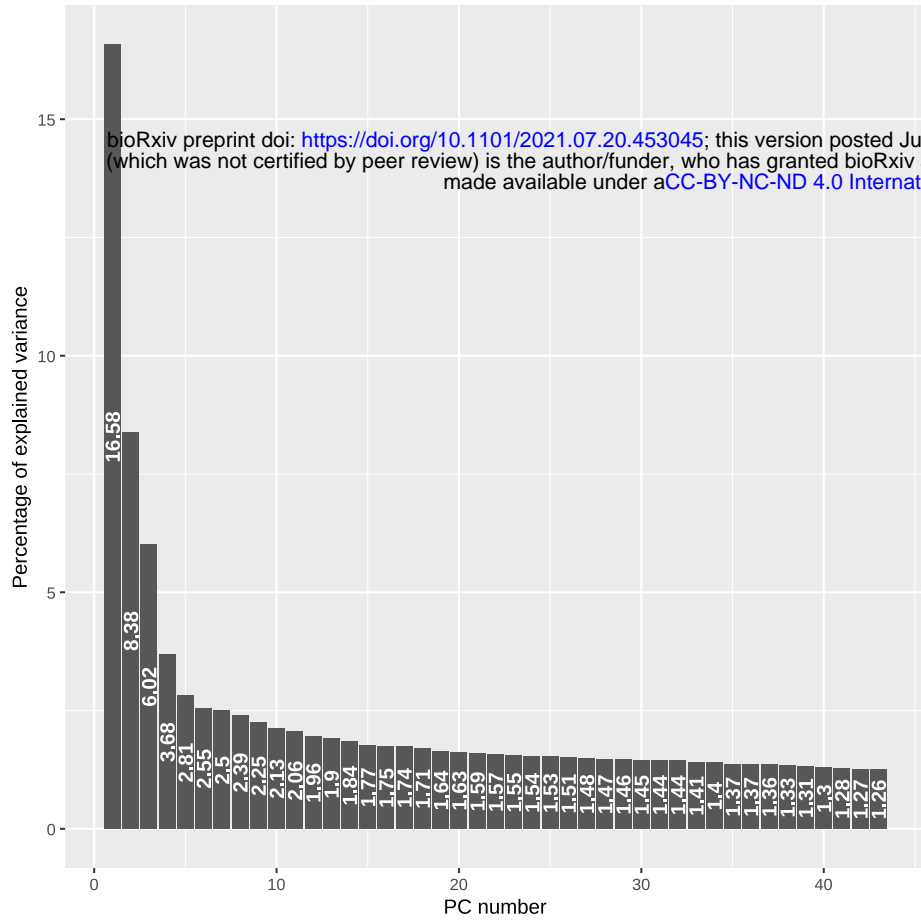
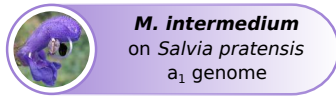
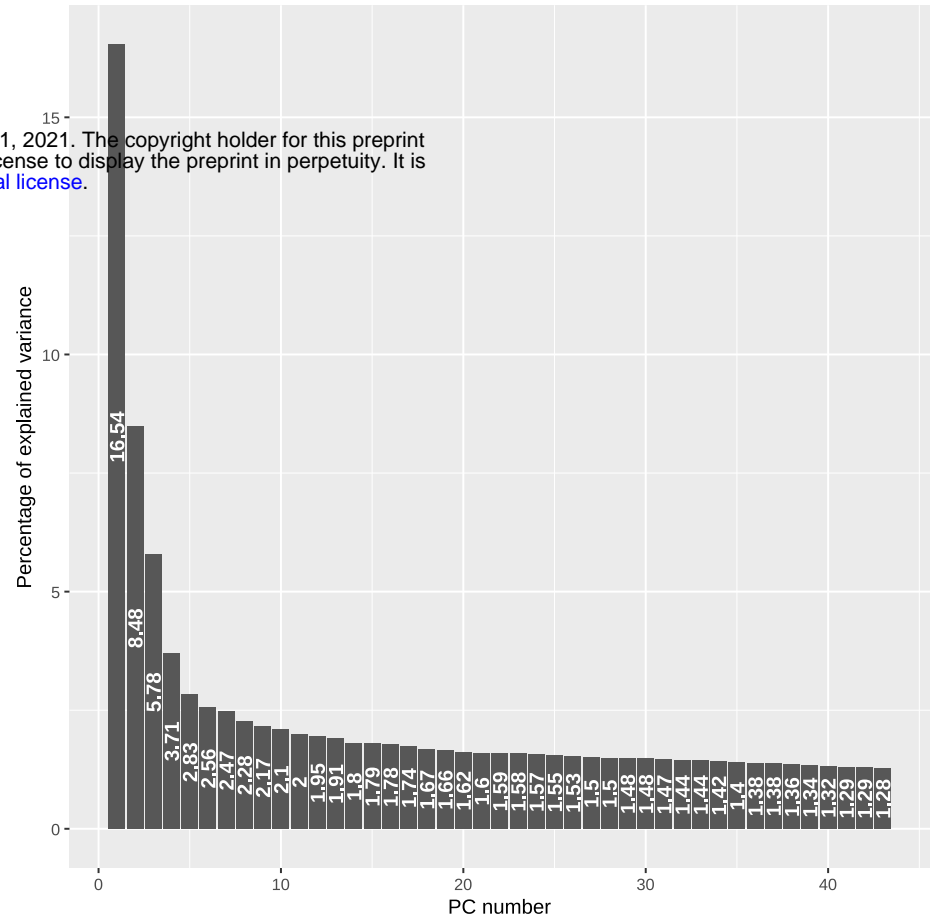
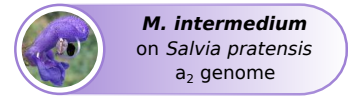


Figure S5b

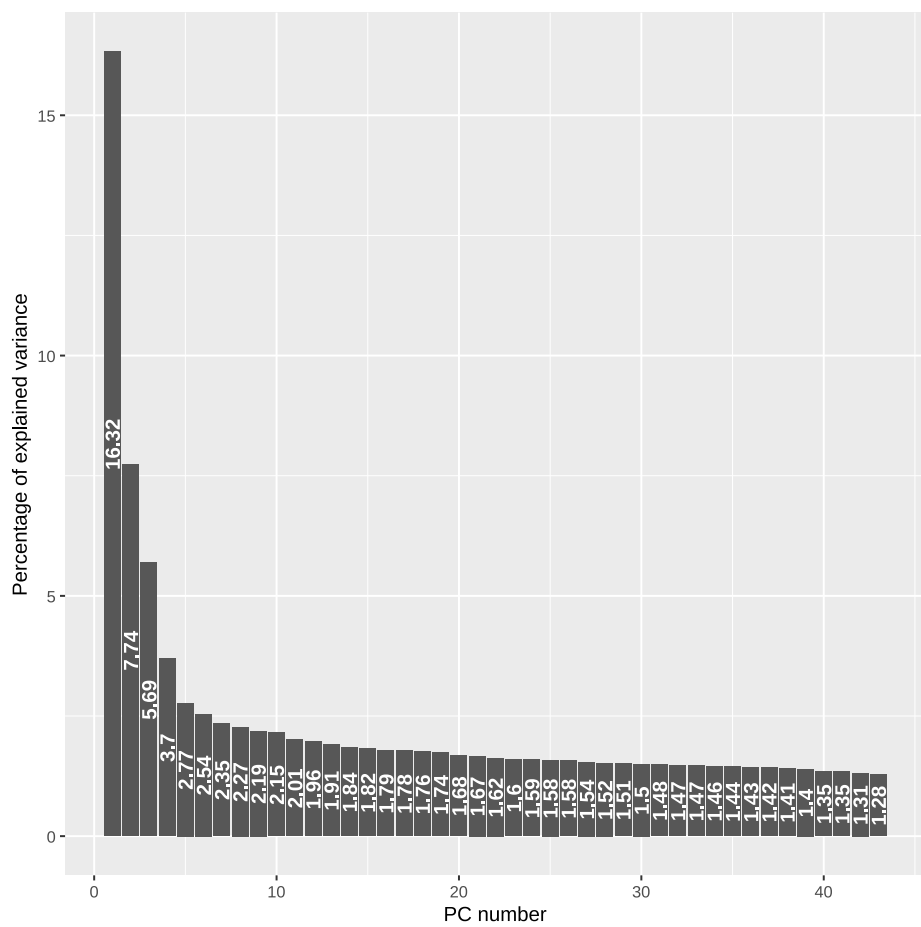
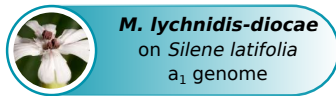
a)



b)



c)



d)

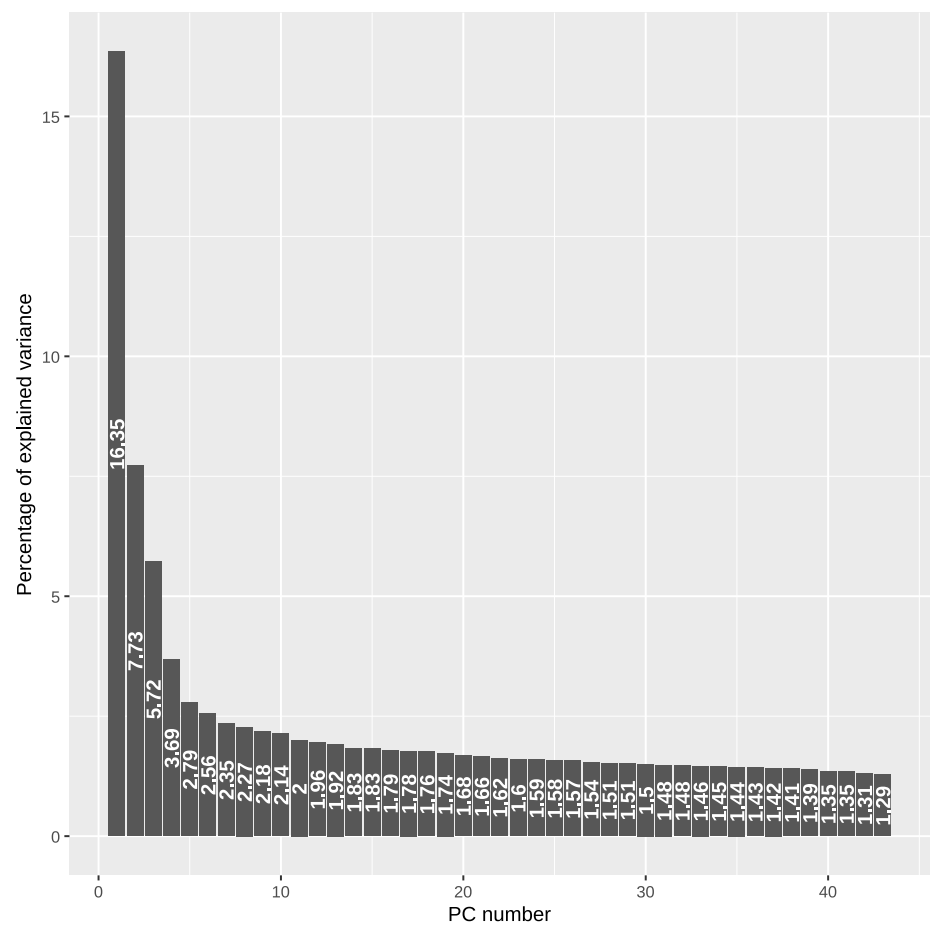
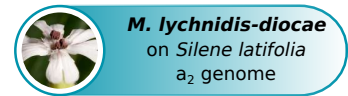
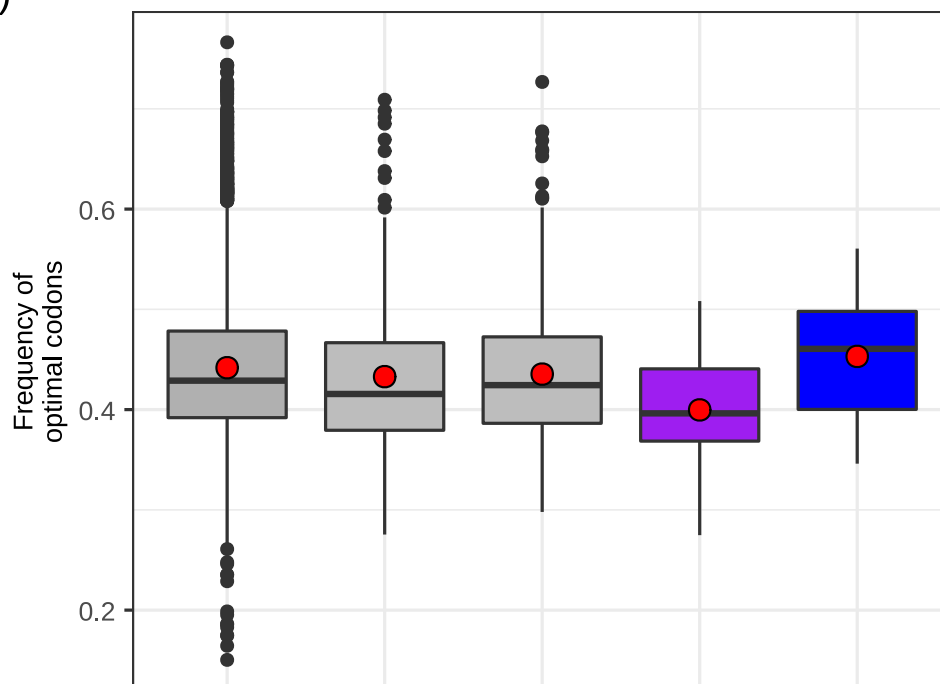


Figure S6

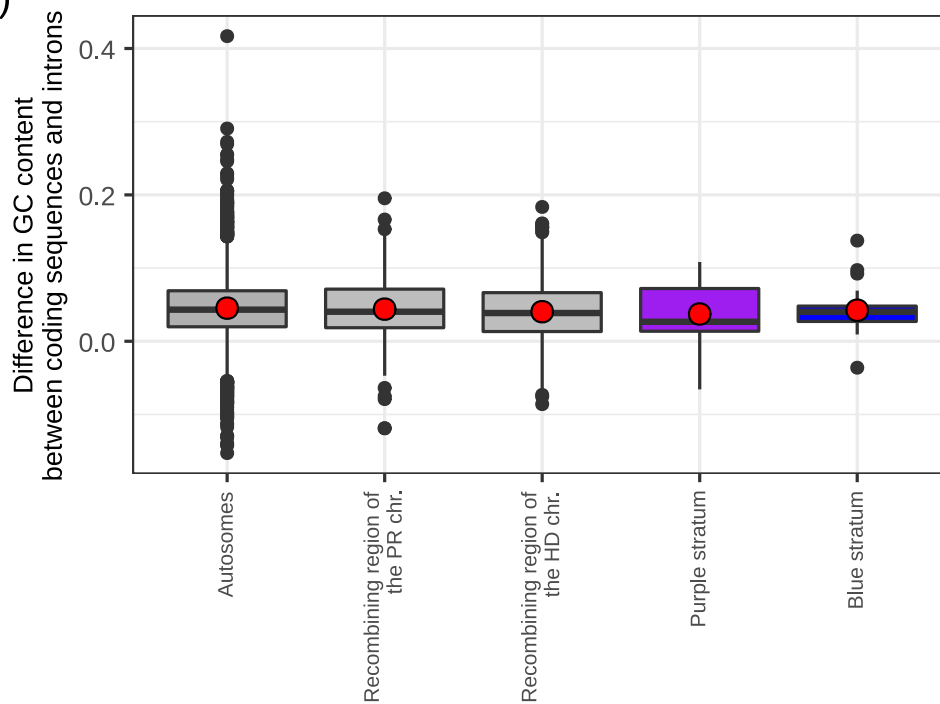


M. intermedium
on *Salvia pratensis*
a₁ genome

A)



B)



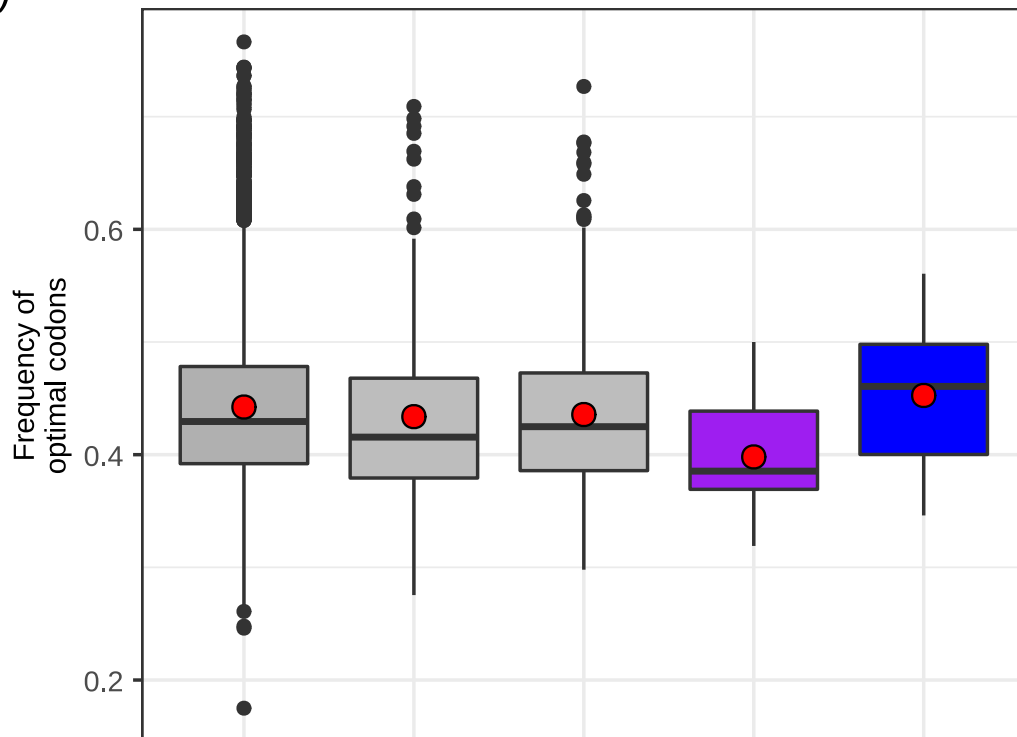
Genomic compartments

- Autosomes
- Recombining region of the PR chr.
- Recombining region of the HD chr.
- Purple stratum
- Blue stratum

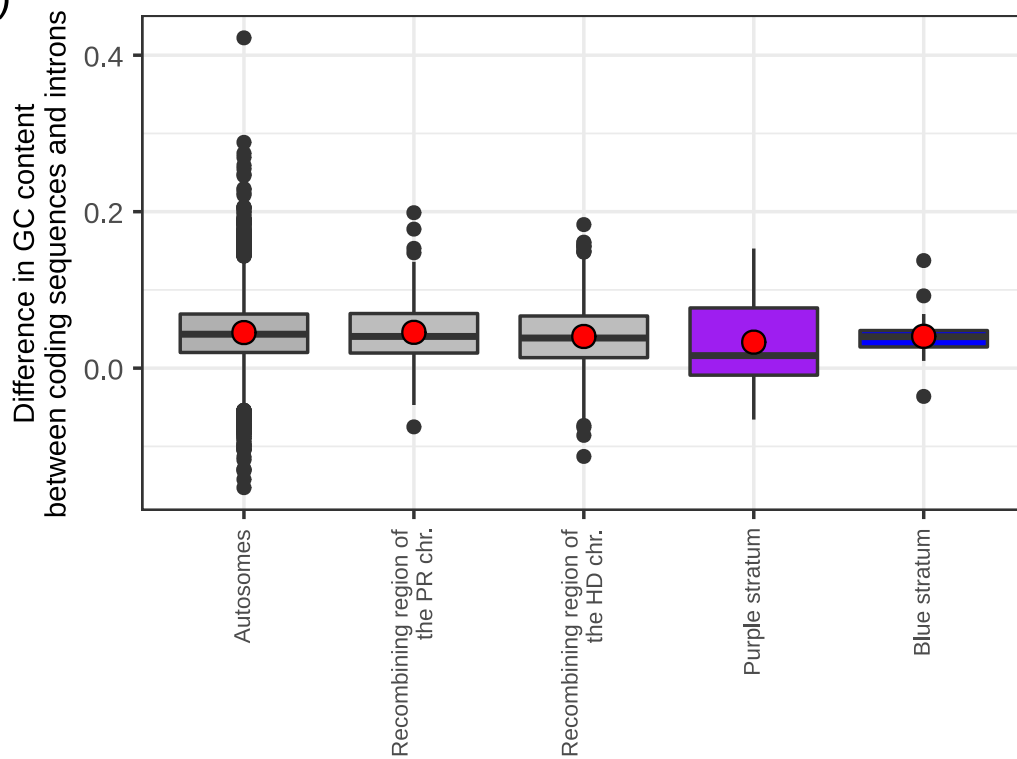


M. intermedium
on *Salvia pratensis*
a₂ genome

A)



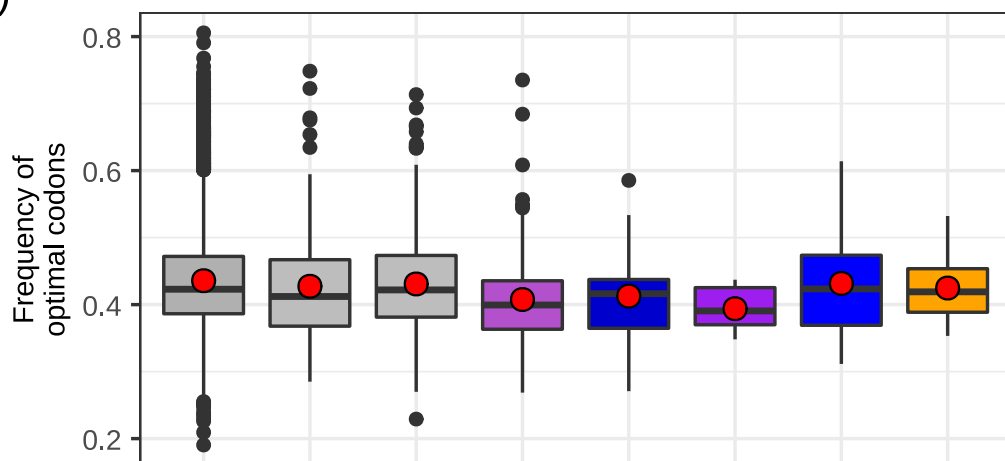
B)



Genomic compartments

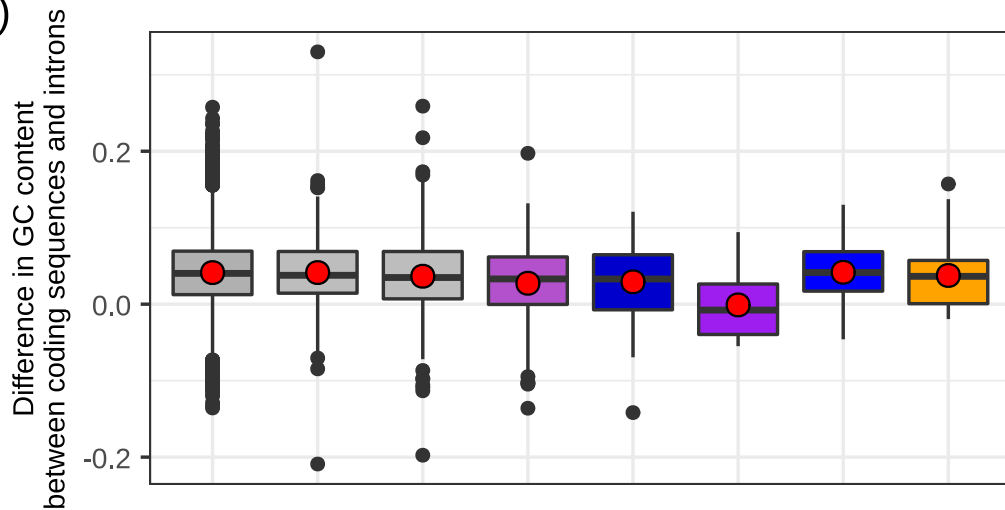
- Autosomes
- Recombining region of the PR chr.
- Recombining region of the HD chr.
- Purple stratum
- Blue stratum

A)

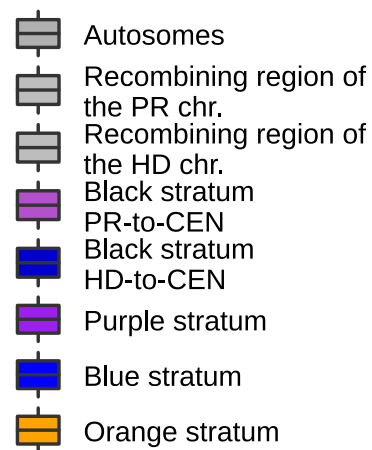


M. lagerheimii
on *Silene vulgaris*
a₁ genome

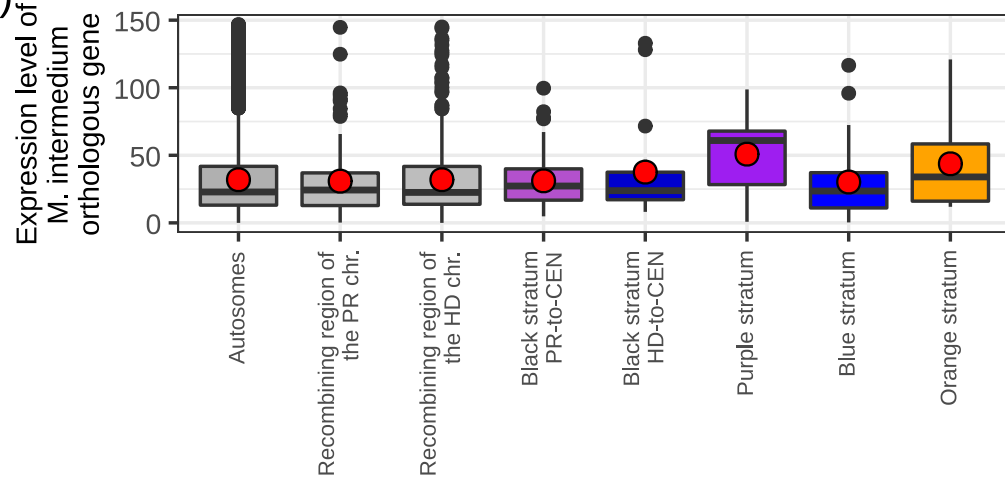
B)



Genomic compartments



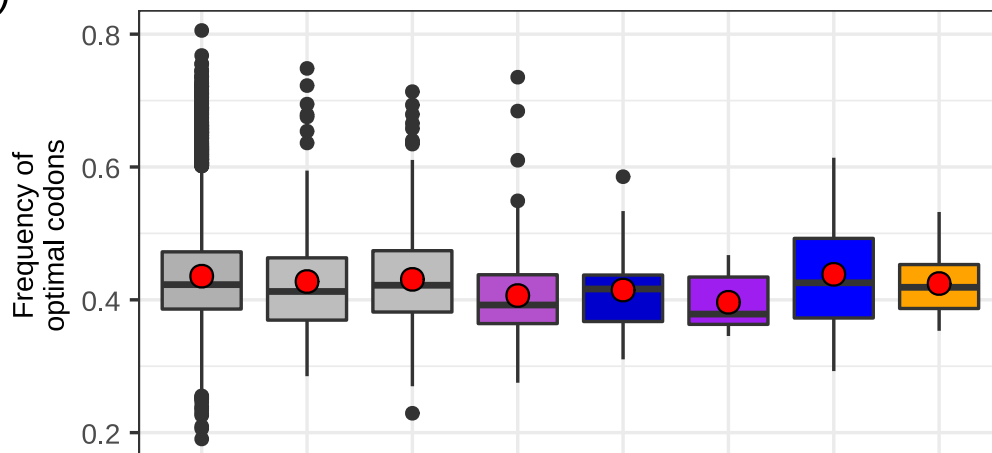
C)



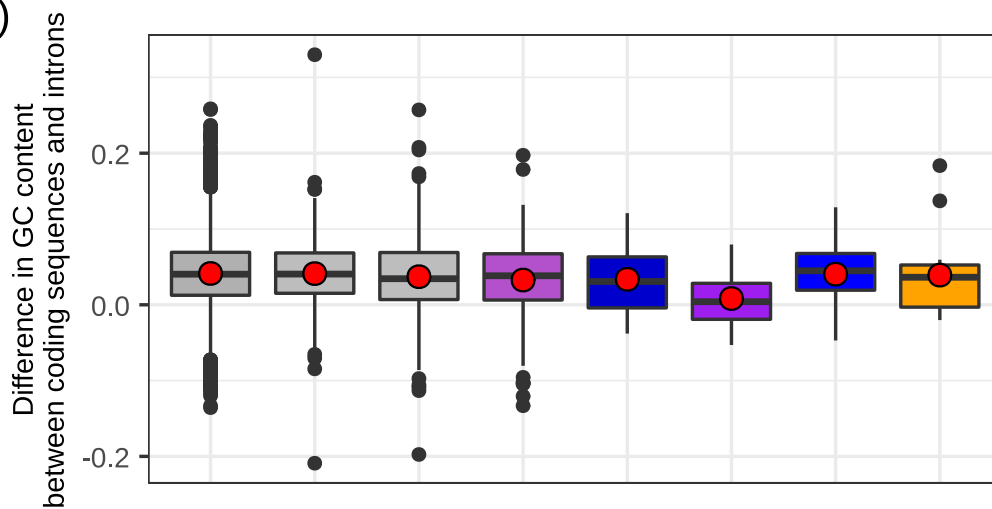


M. lagerheimii
on *Silene vulgaris*
a₂ genome

A)



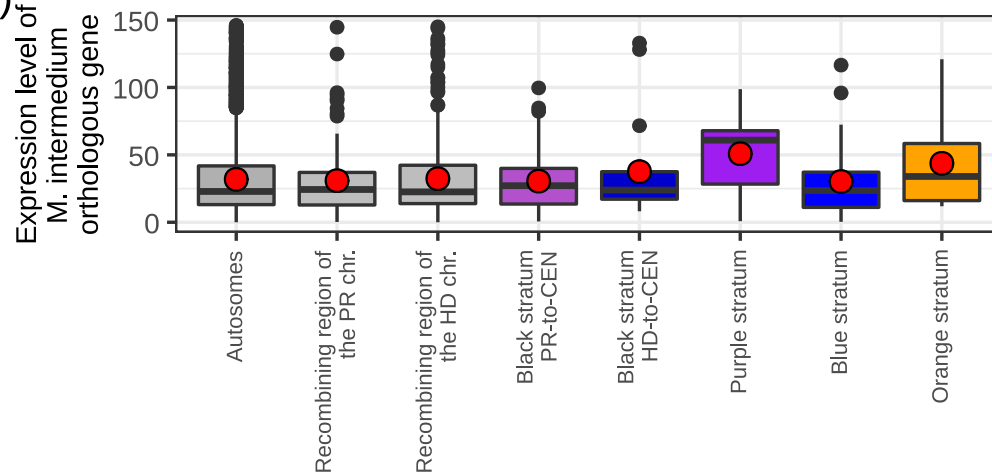
B)



Genomic compartments

- Autosomes
- Recombining region of the PR chr.
- Recombining region of the HD chr.
- Black stratum PR-to-CEN
- Black stratum HD-to-CEN
- Purple stratum
- Blue stratum
- Orange stratum

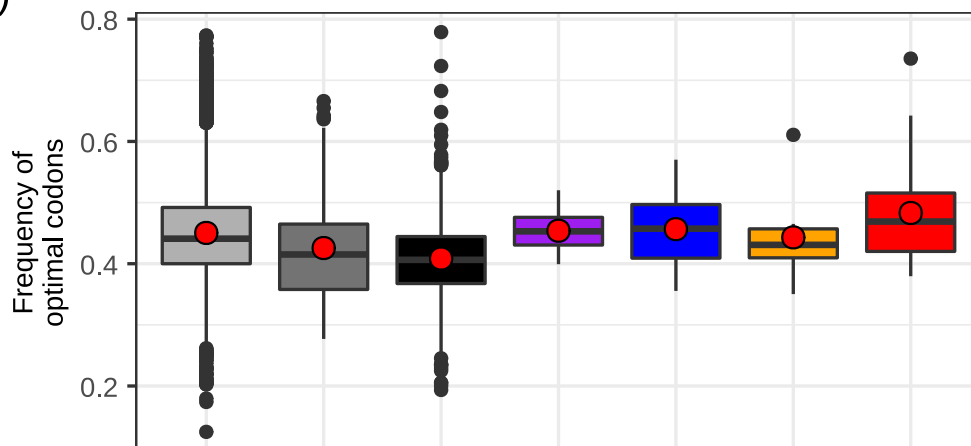
C)



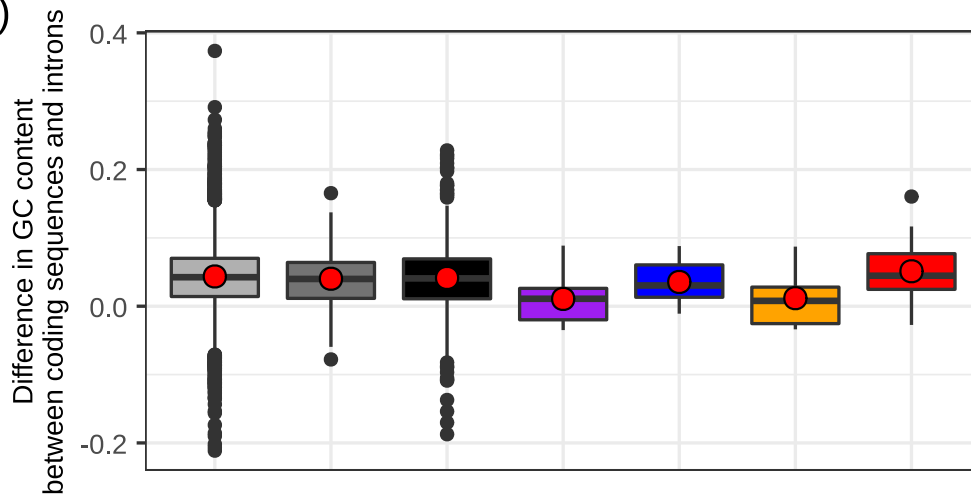


M. lychnidis-diocae
on *Silene latifolia*
a₁ genome

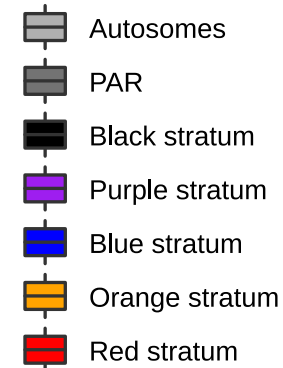
A)



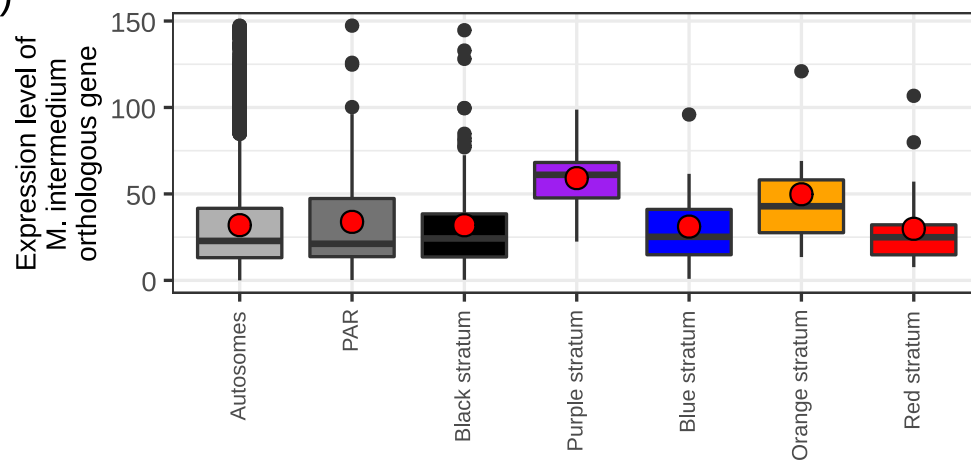
B)

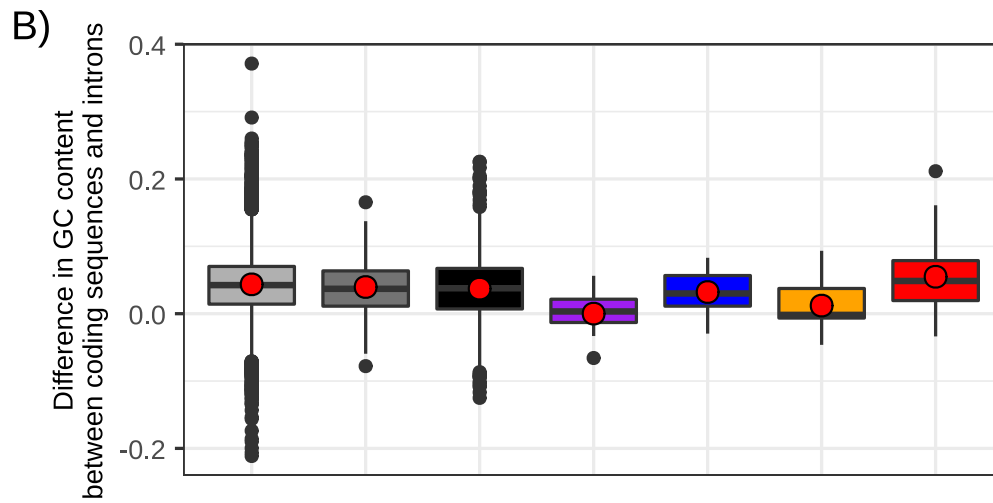
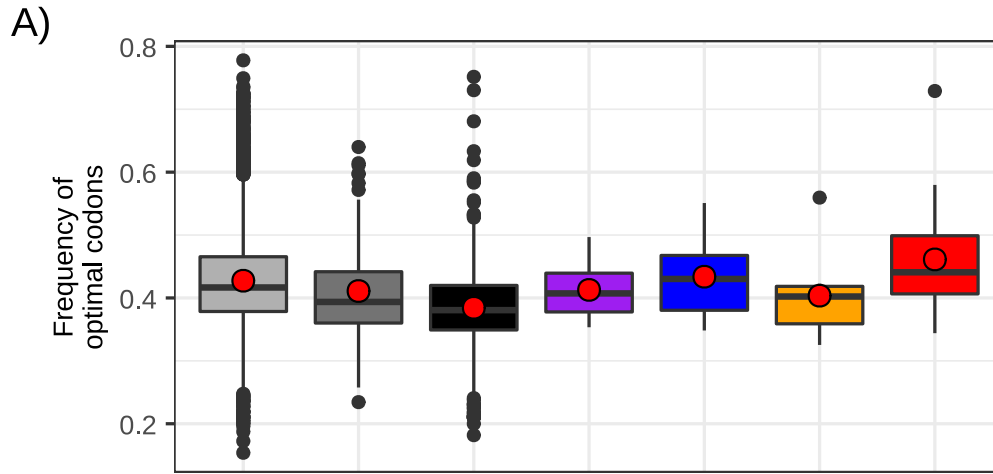


Genomic compartments

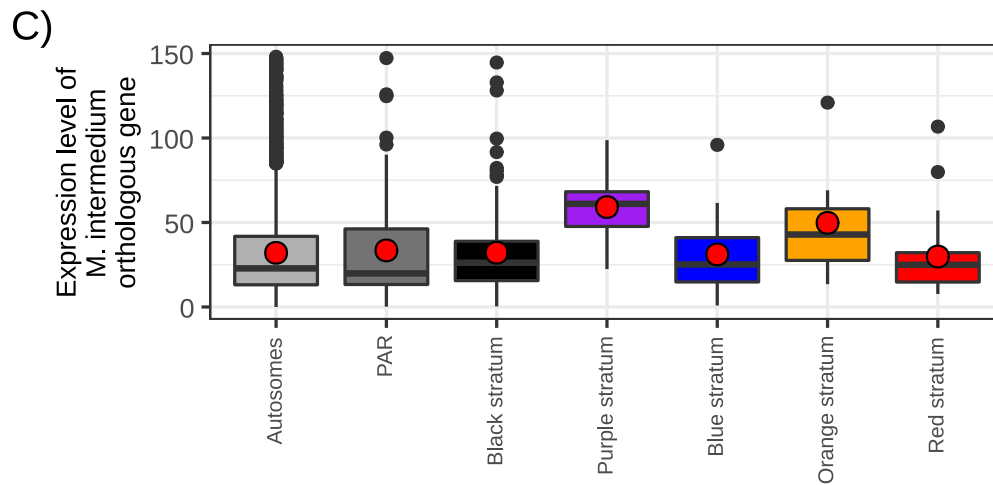
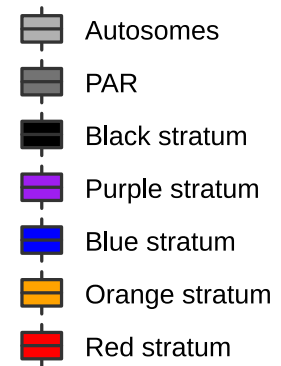


C)



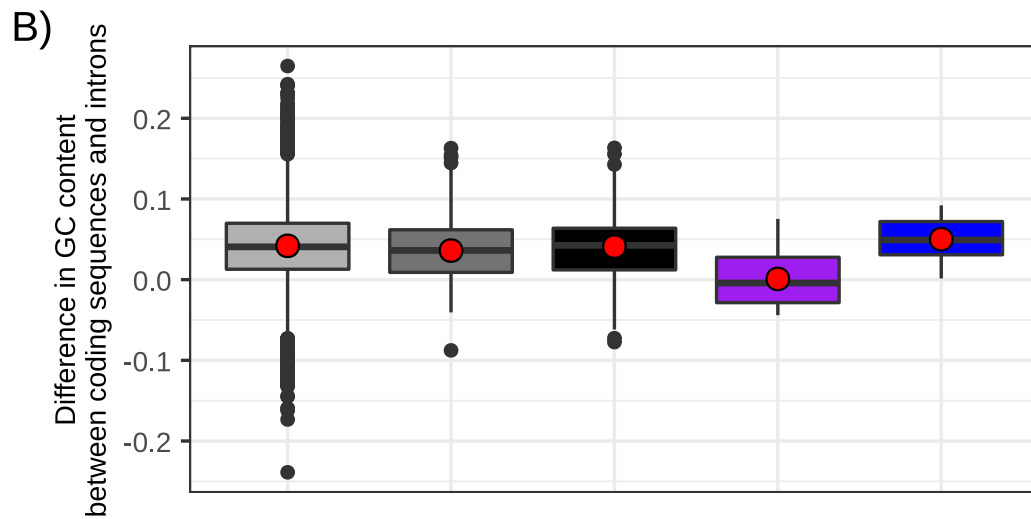
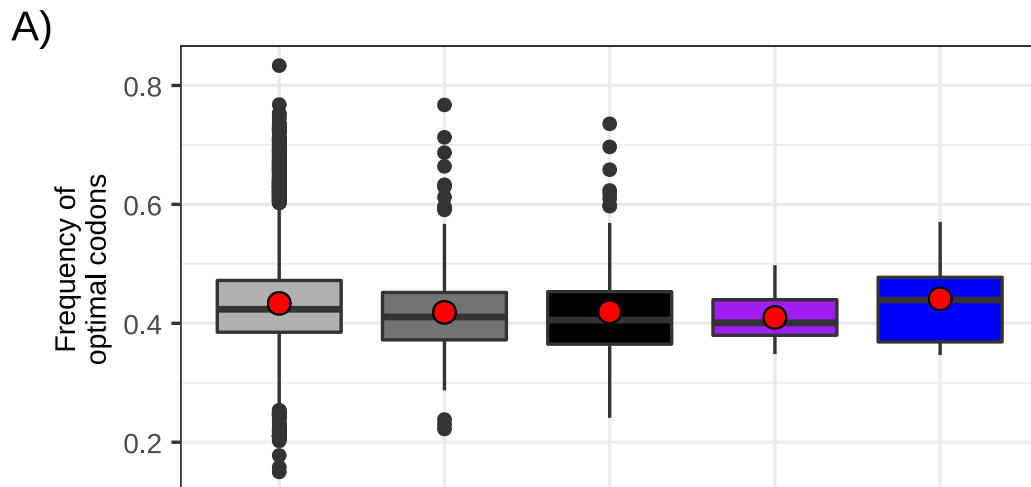


Genomic compartments

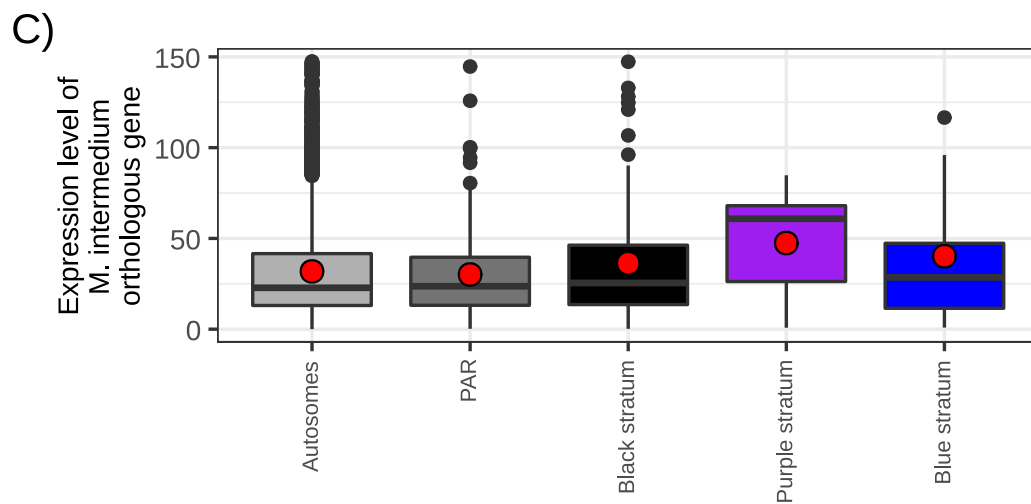
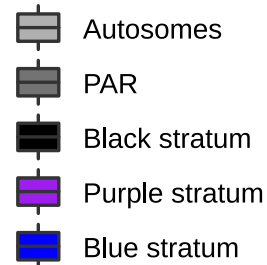




M. silenes-acaulis
on *Silene acaulis*
a₁ genome

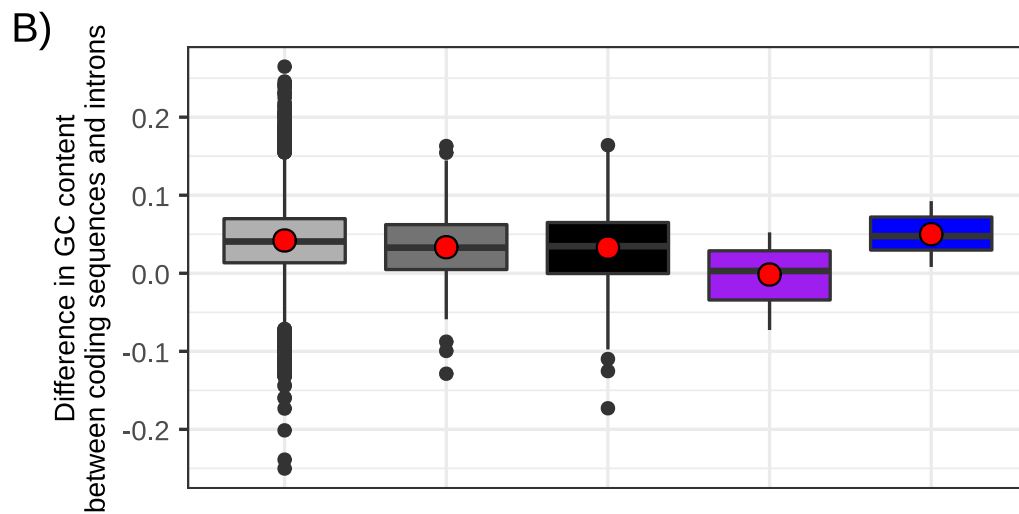
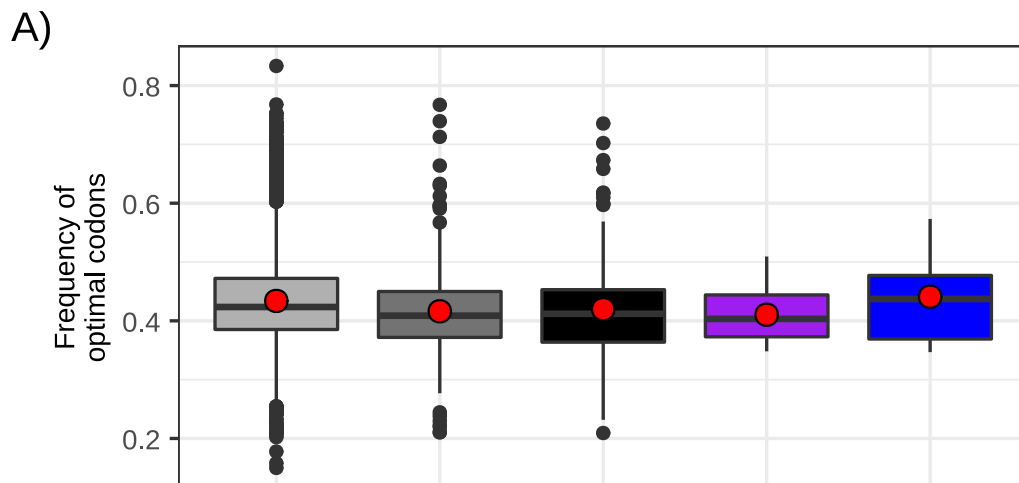


Genomic compartments

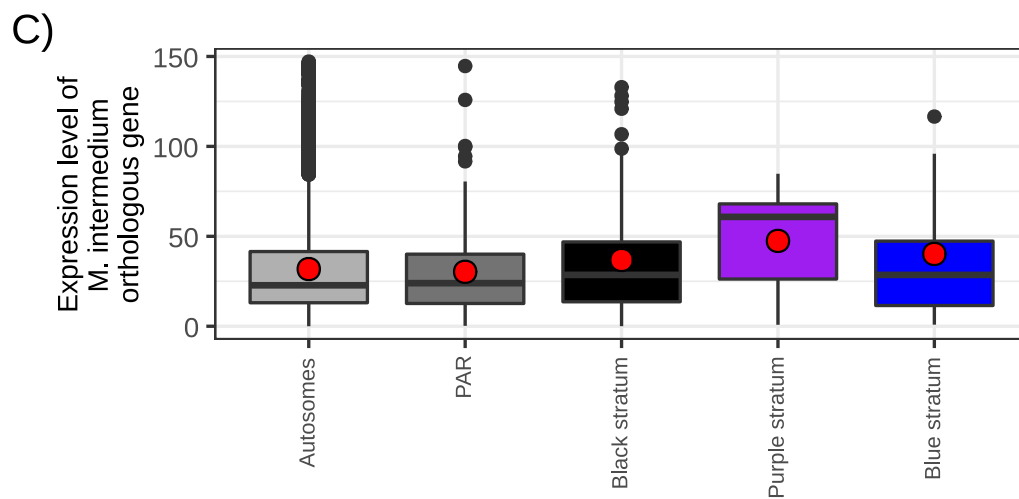
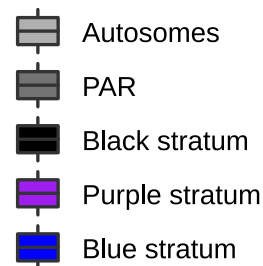


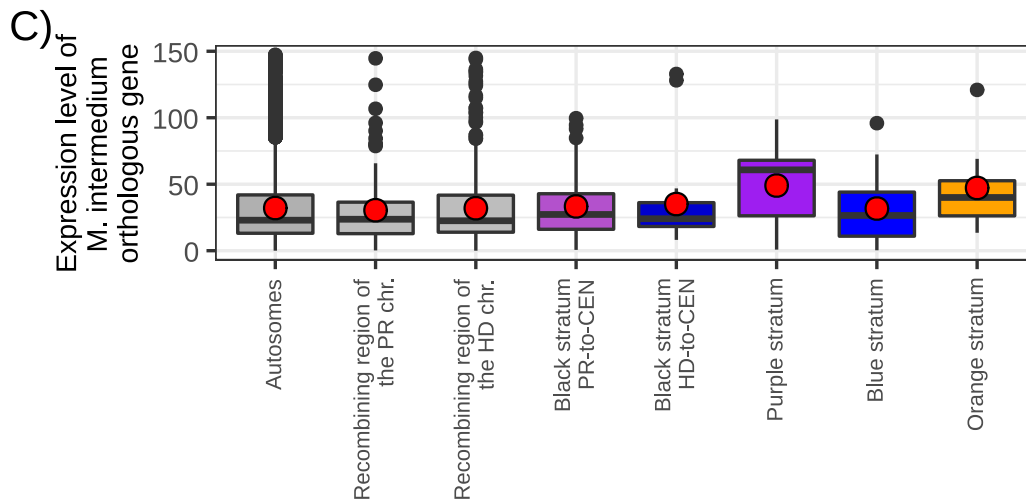
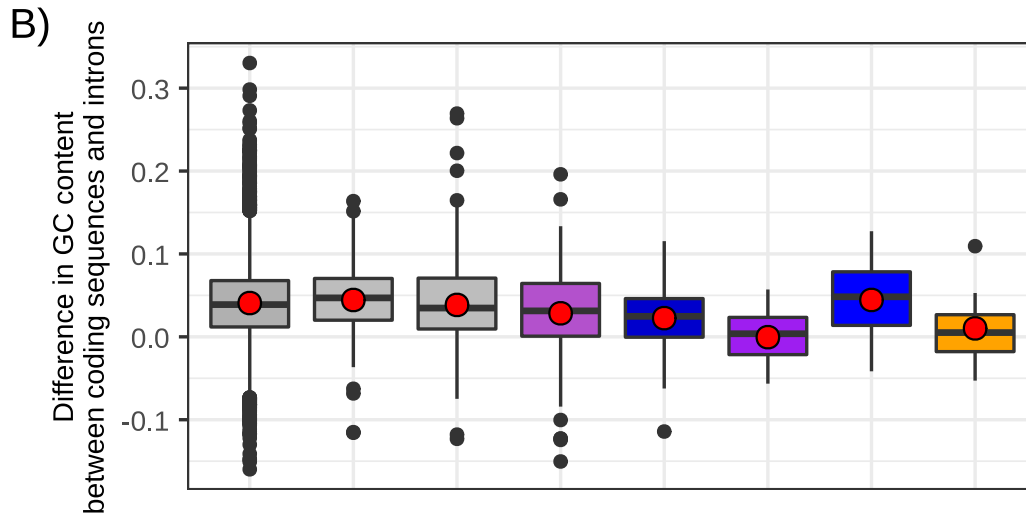
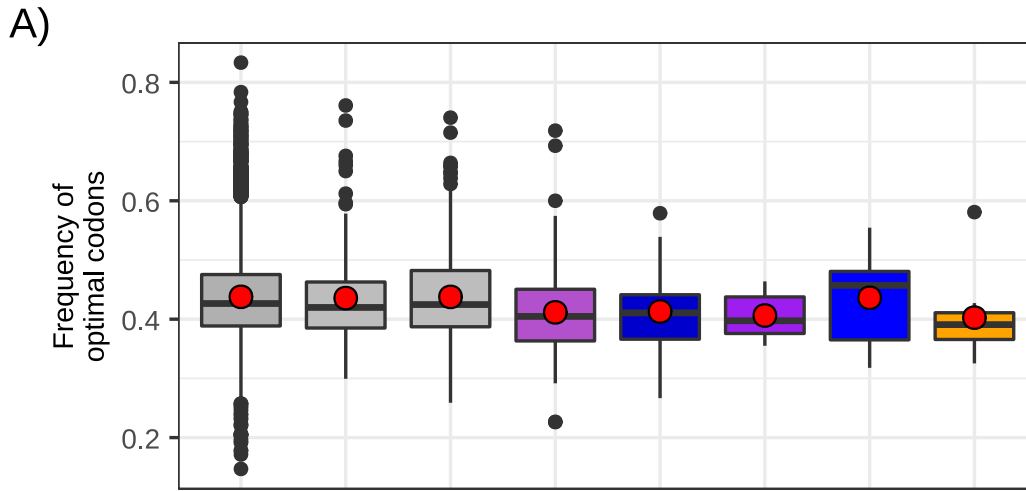


M. silenes-acaulis
on *Silene acaulis*
a₂ genome











Genomic compartments



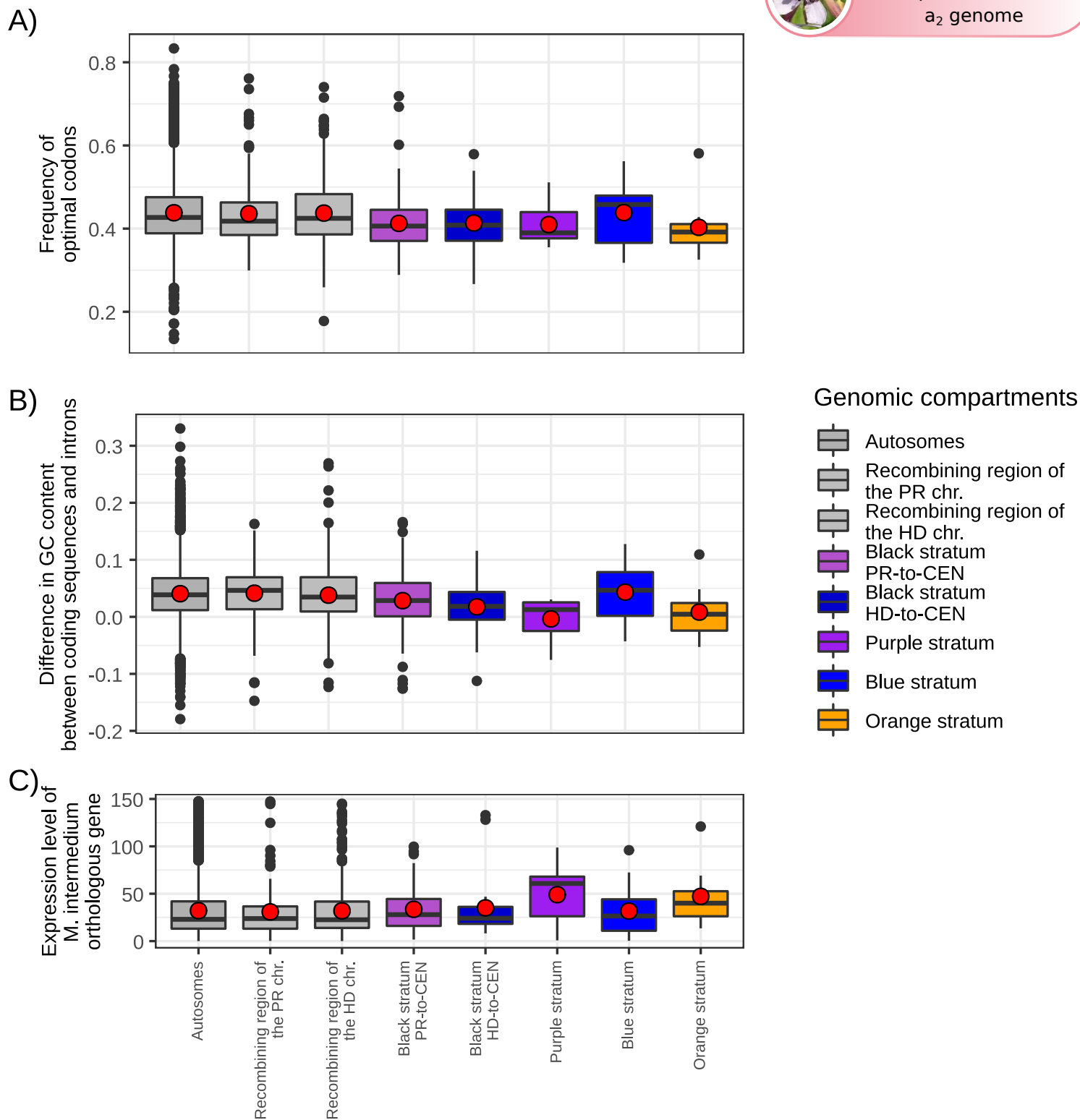


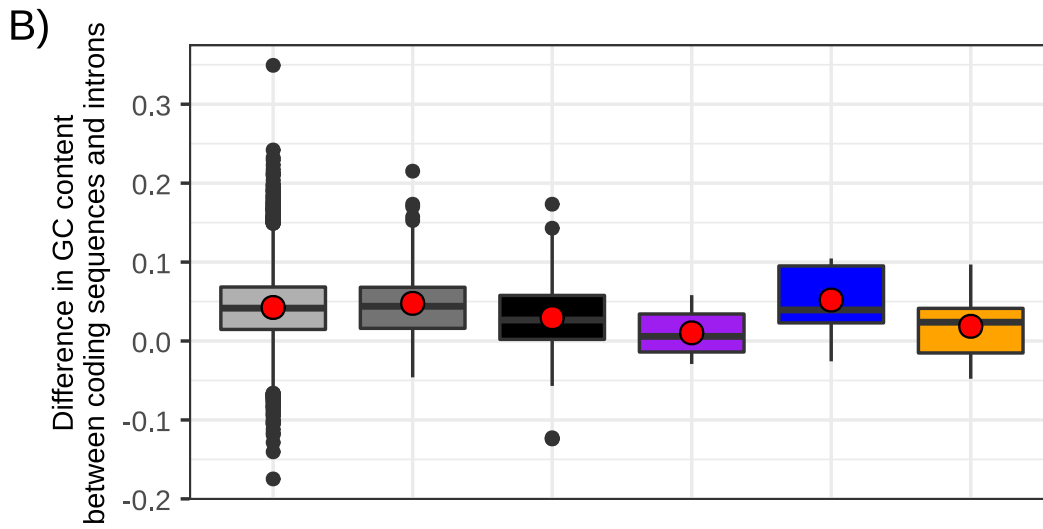
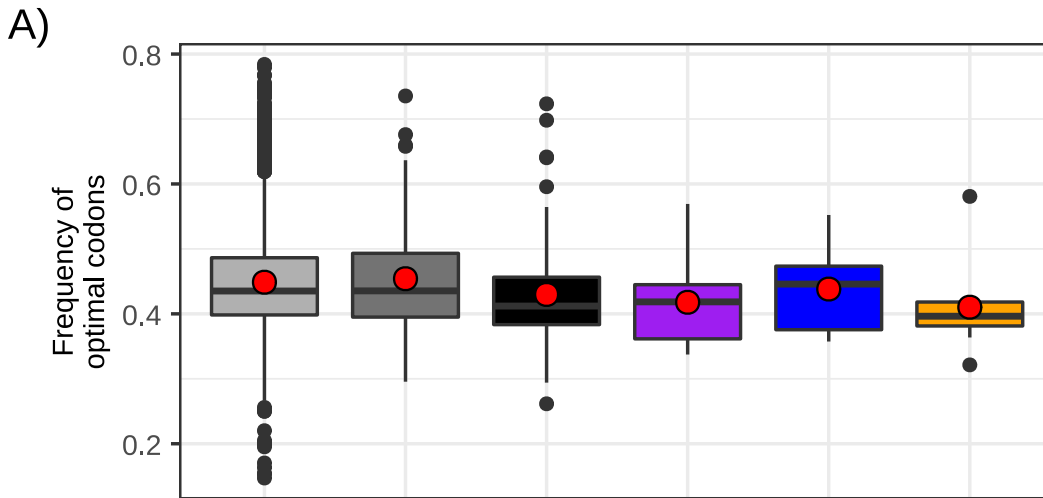
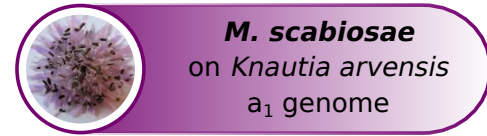
Genomic compartments

-  Autosomes
-  Recombining region of the PR chr.
-  Recombining region of the HD chr.
-  Black stratum PR-to-CEN
-  Black stratum HD-to-CEN
-  Purple stratum
-  Blue stratum
-  Orange stratum

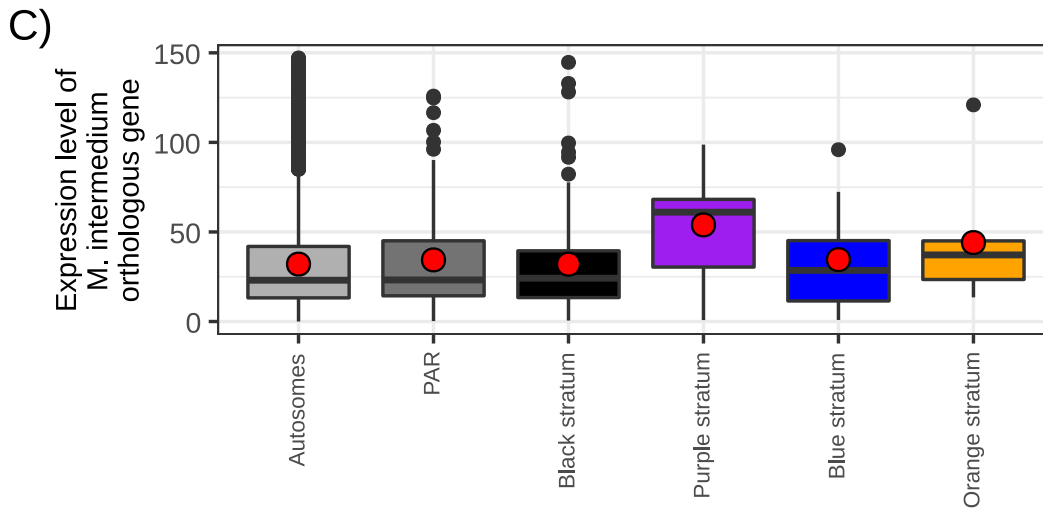
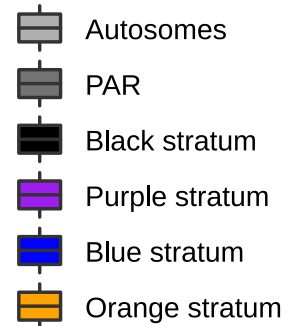


M. saponariae
on *Saponaria officinalis*
a₂ genome





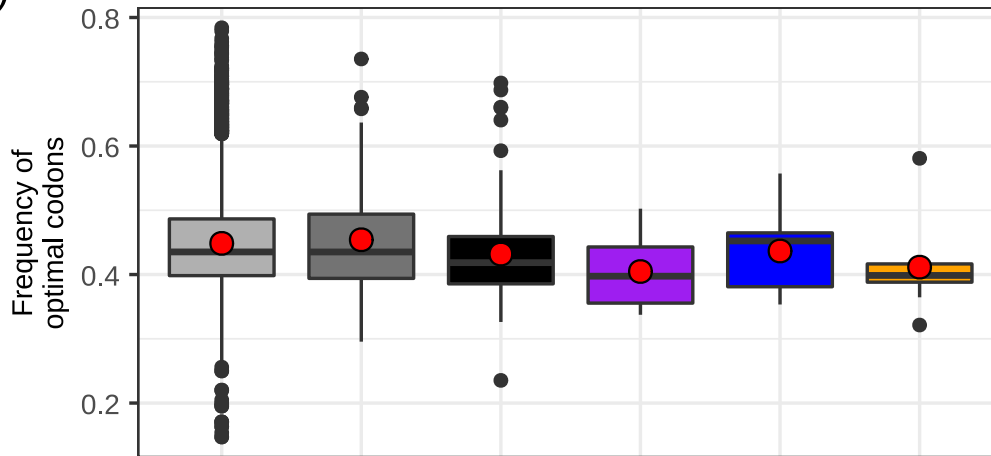
Genomic compartments



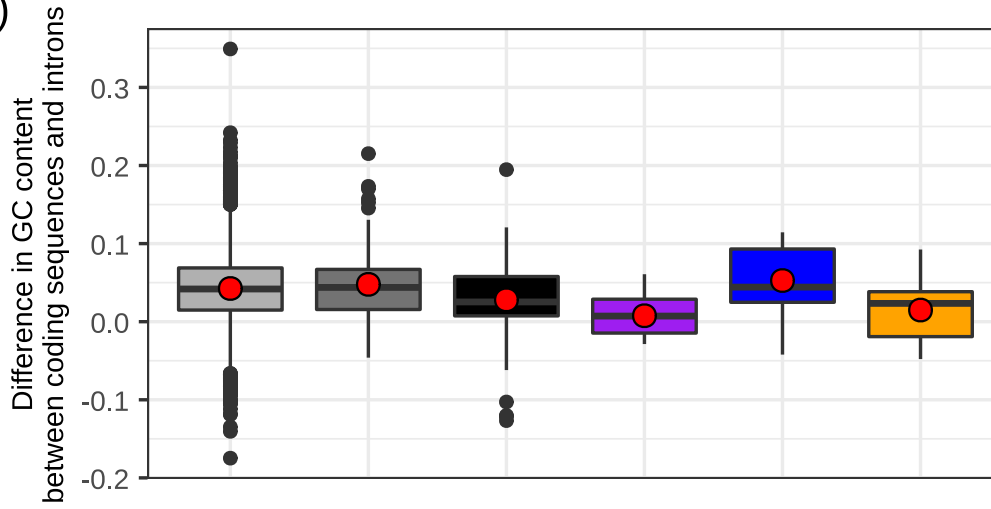


M. scabiosae
on *Knautia arvensis*
a₂ genome

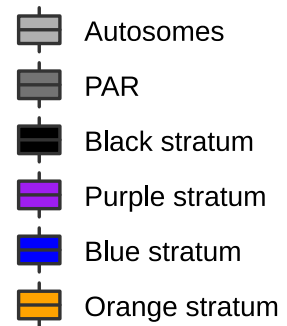
A)



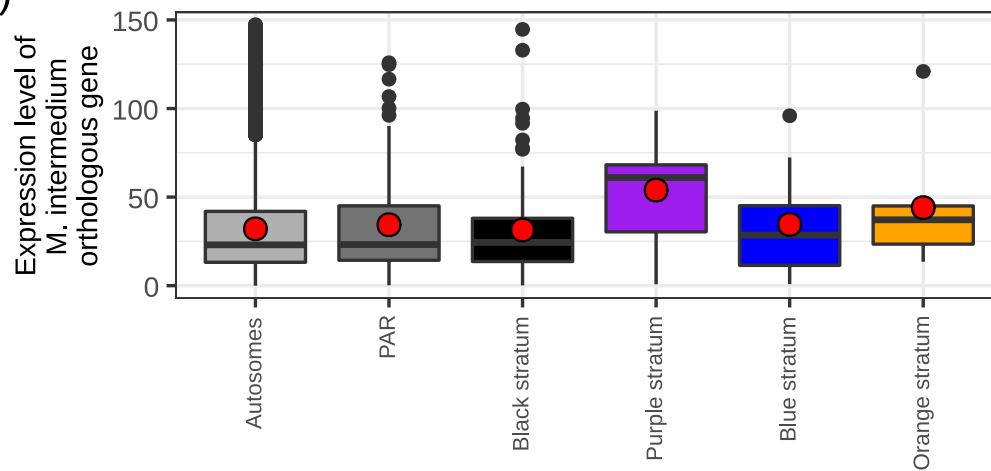
B)



Genomic compartments

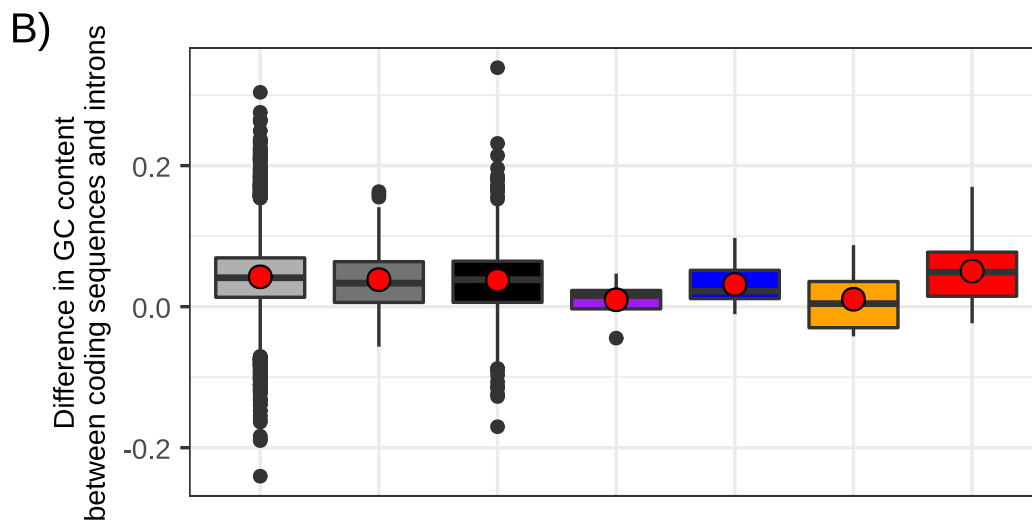
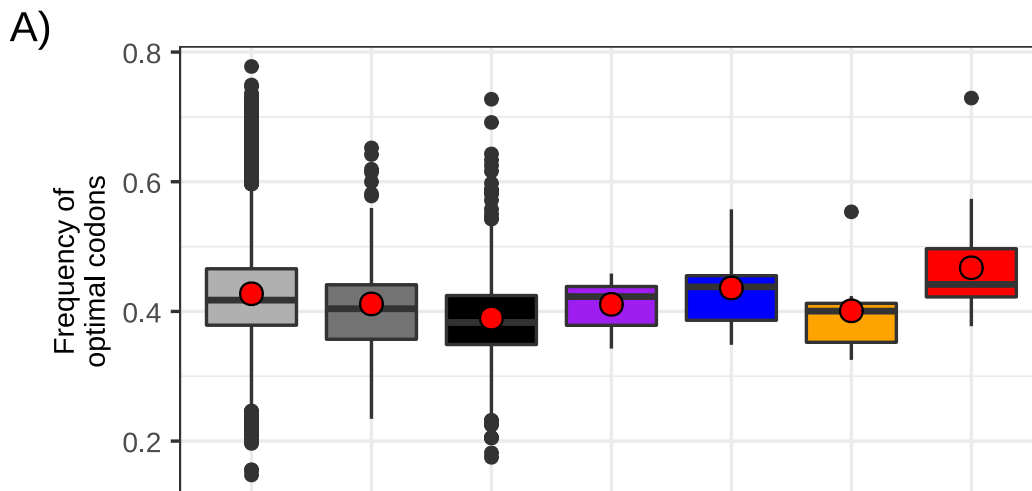


C)

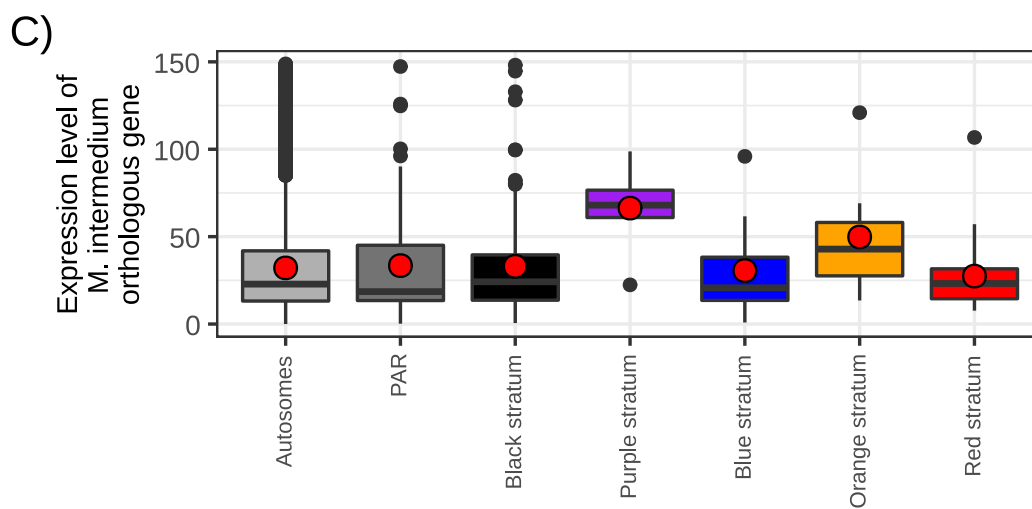
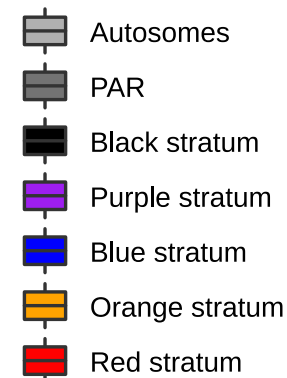




M. silenes-dioicae
on *Silene dioica*
a₁ genome



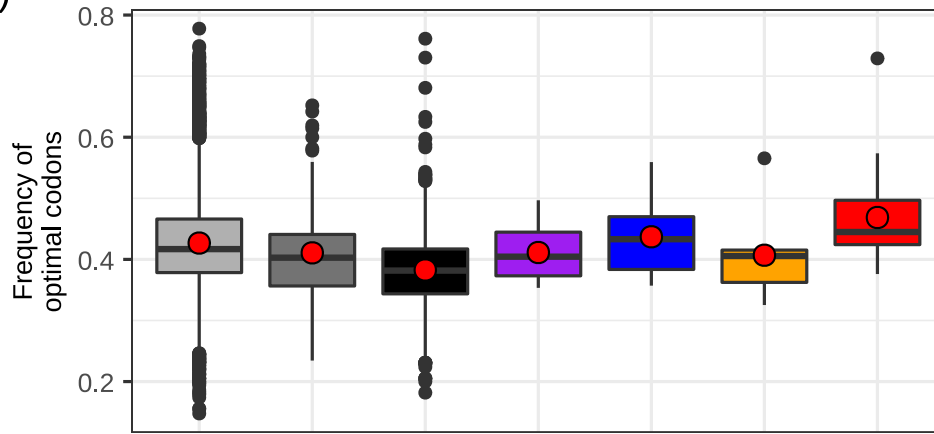
Genomic compartments



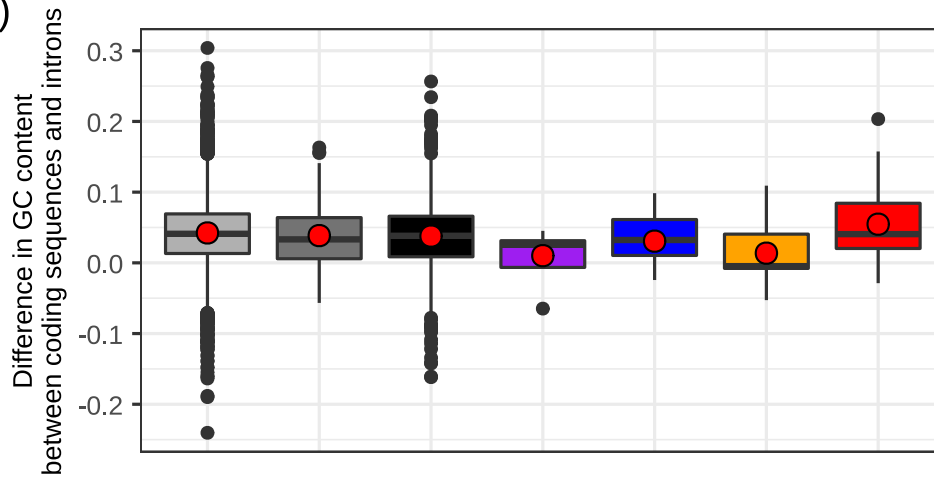


M. silenes-dioicae
on *Silene dioica*
a₂ genome

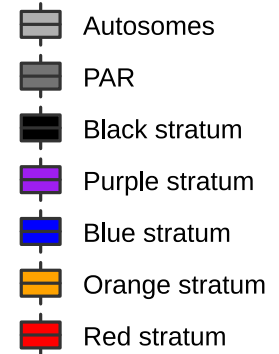
A)



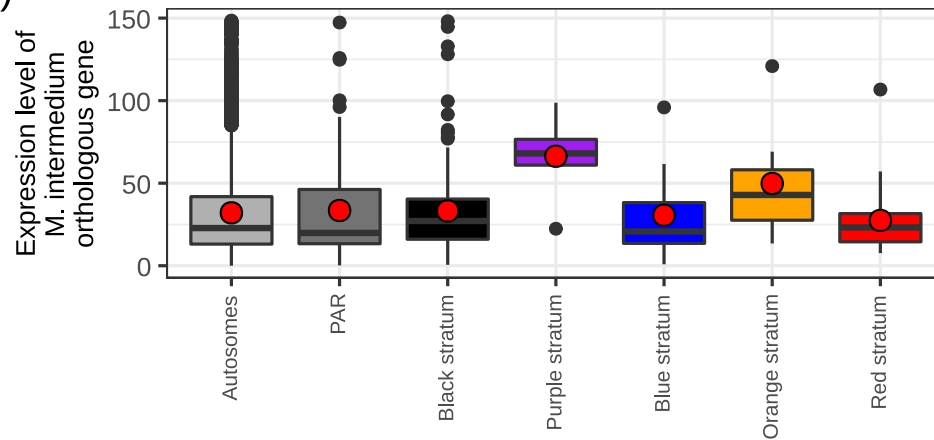
B)



Genomic compartments



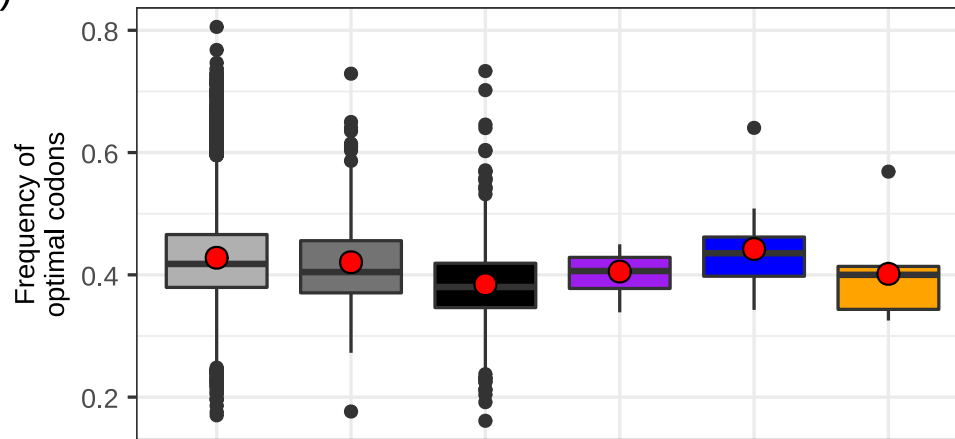
C)



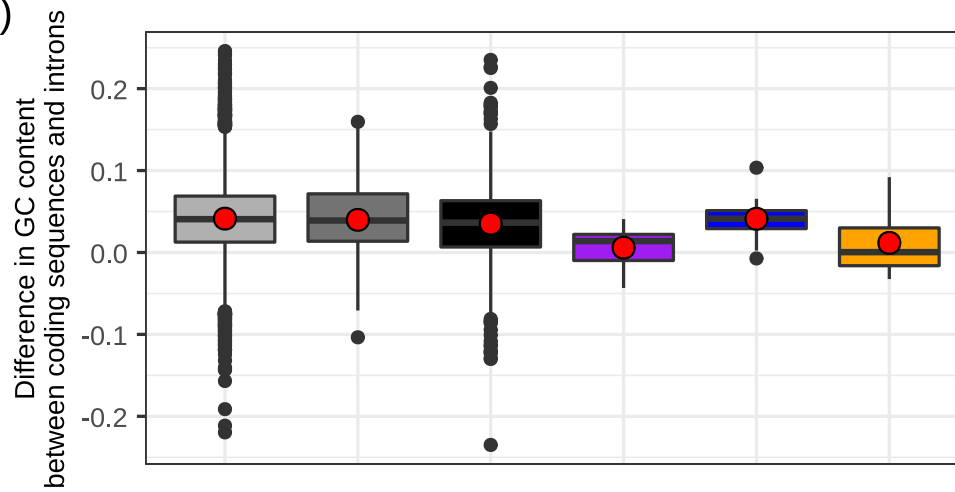


M. coronaria
on *Silene flos-cuculi*
a₁ genome

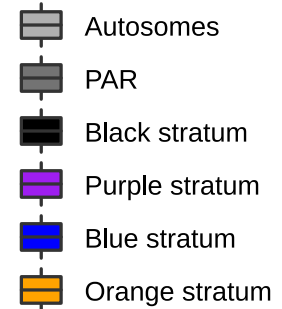
A)



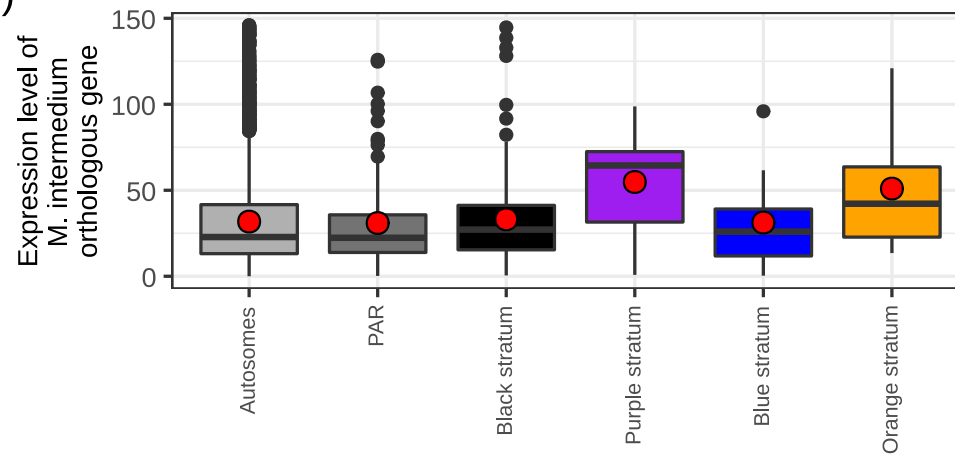
B)



Genomic compartments



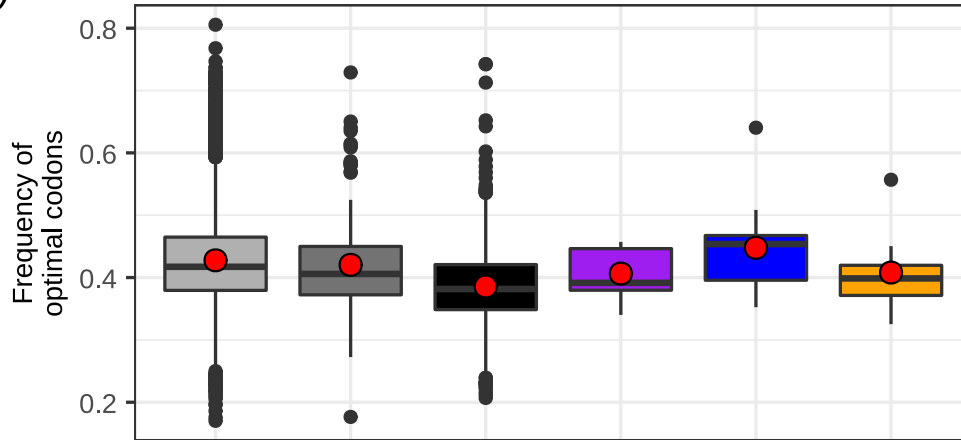
C)



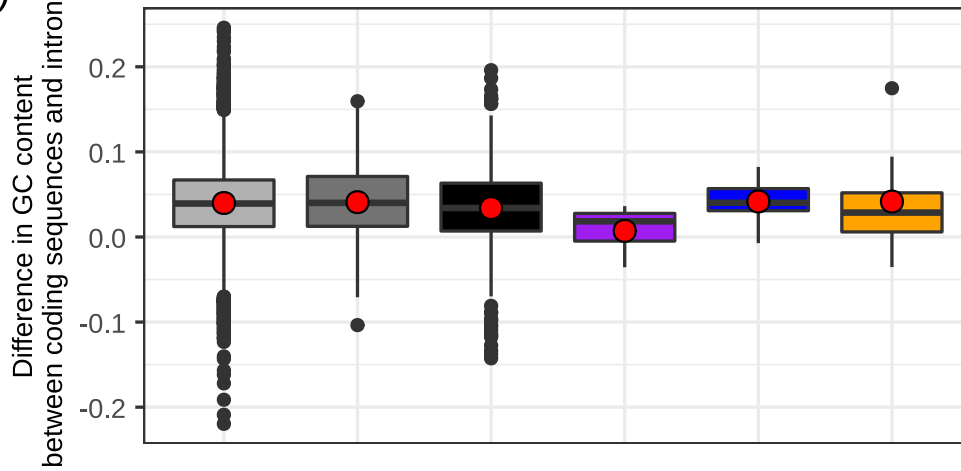


M. coronariae
on *Silene flos-cuculi*
a₂ genome

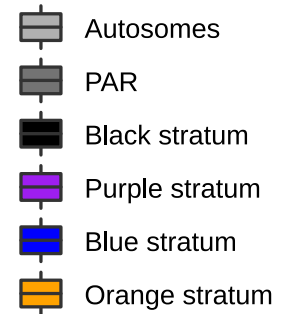
A)



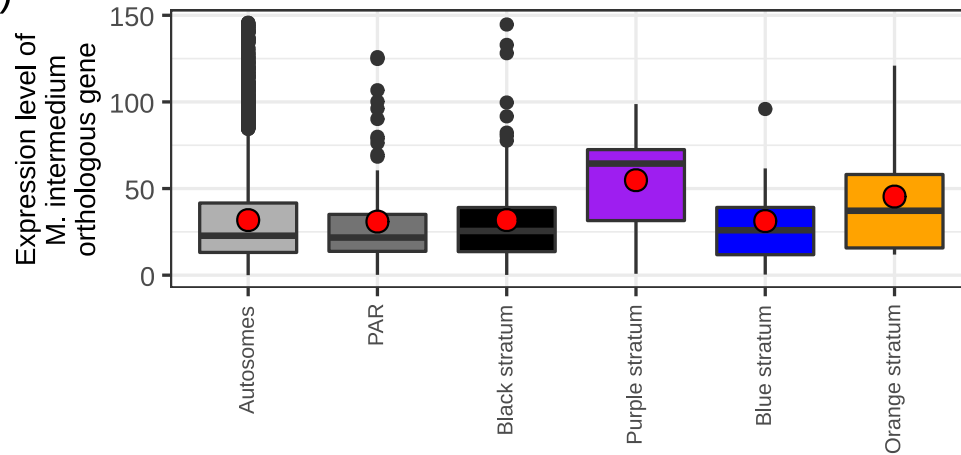
B)



Genomic compartments



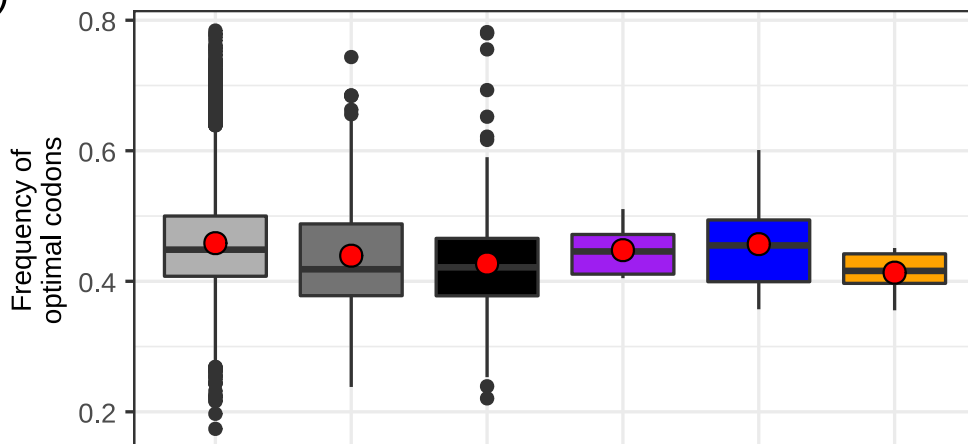
C)



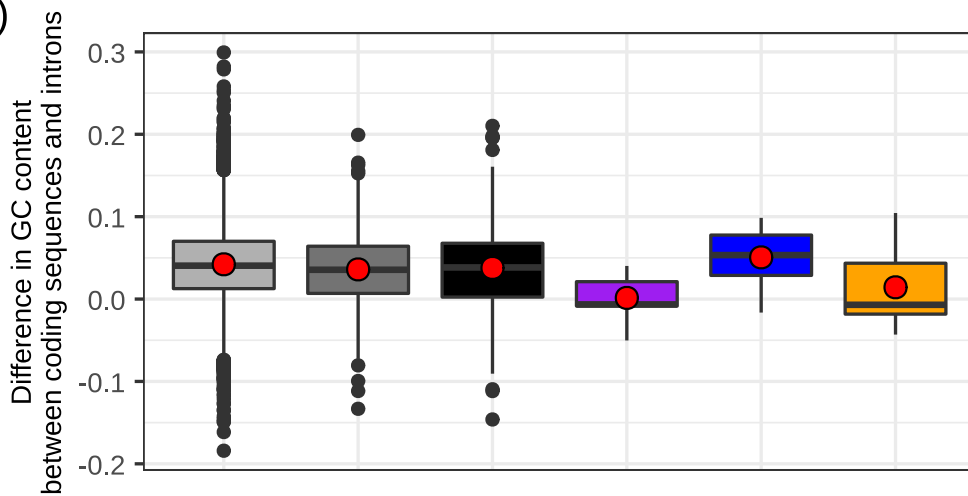


M. violaceum s.l.
on *Silene melanantha*
a₁ genome

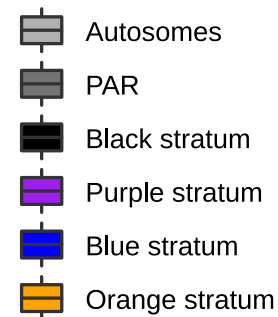
A)



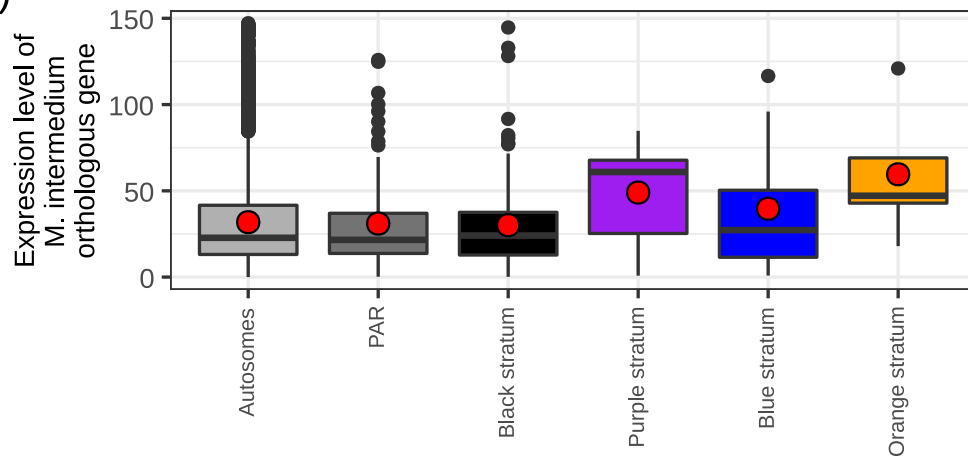
B)



Genomic compartments



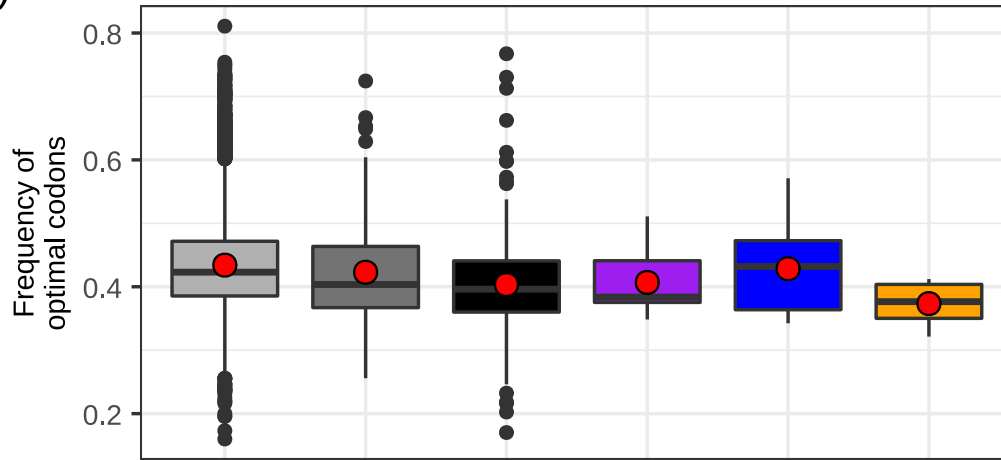
C)



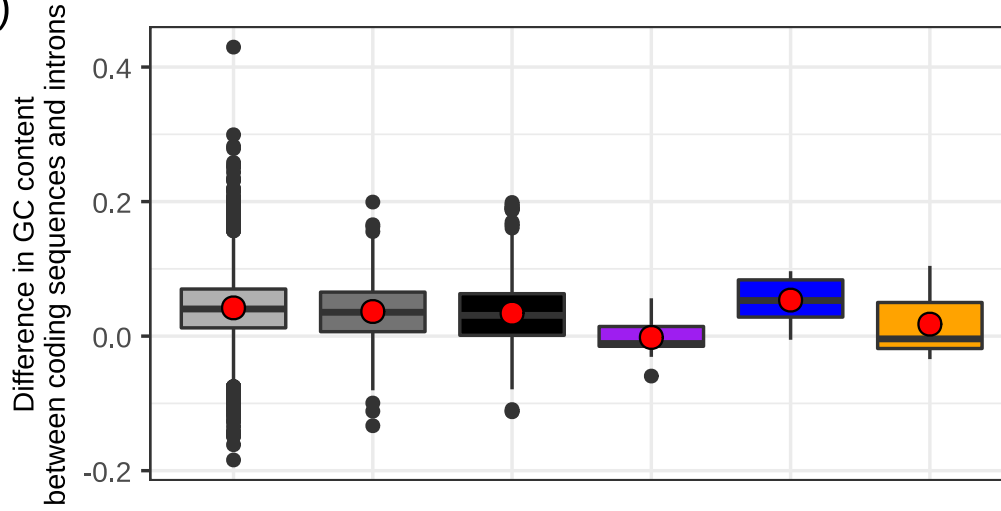


M. violaceum s.l.
on *Silene melanantha*
a₂ genome

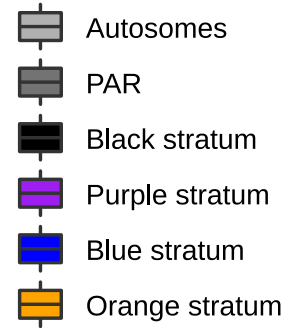
A)



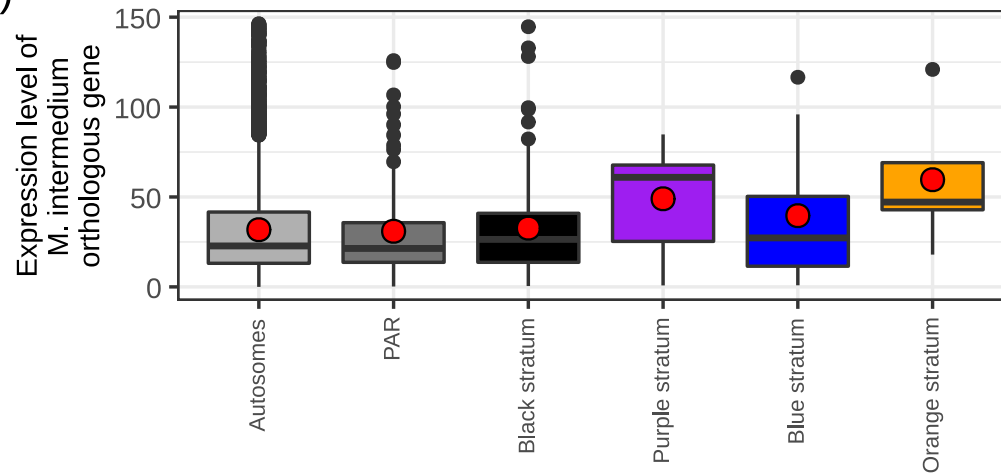
B)



Genomic compartments



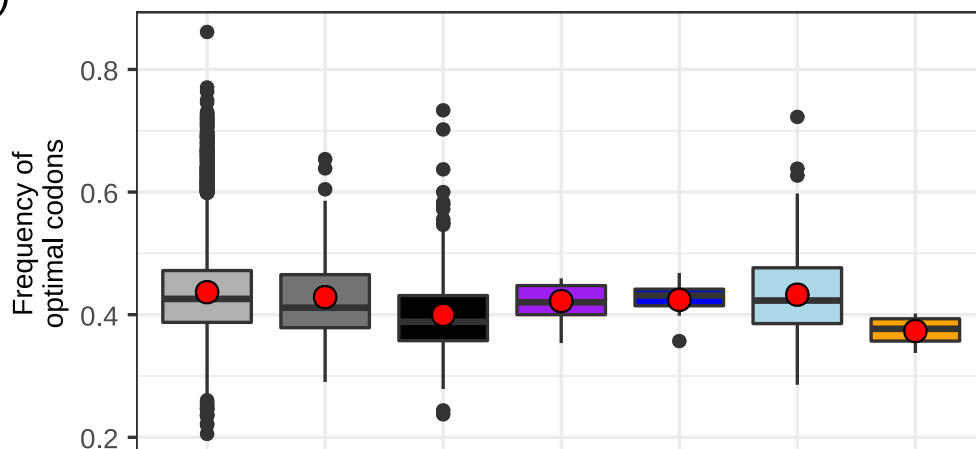
C)



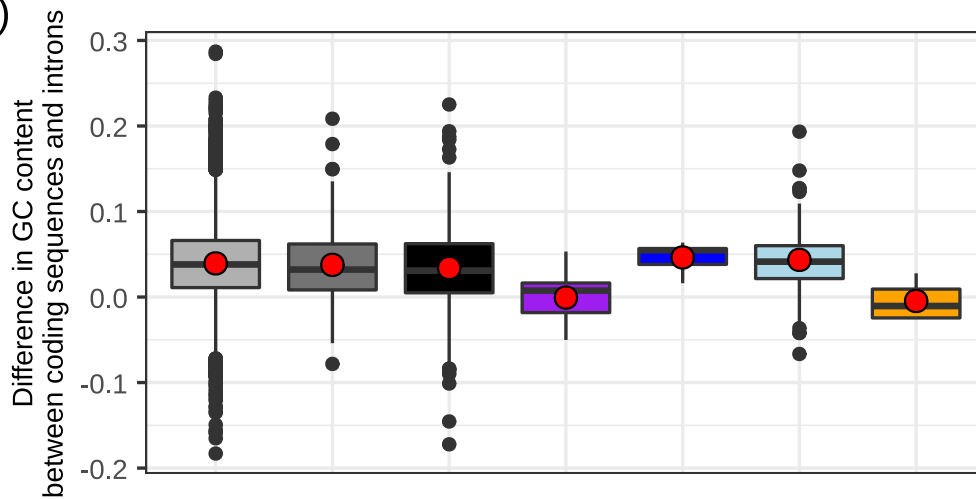


M. violaceum s.l.
on *Silene caroliniana*
a₁ genome

A)



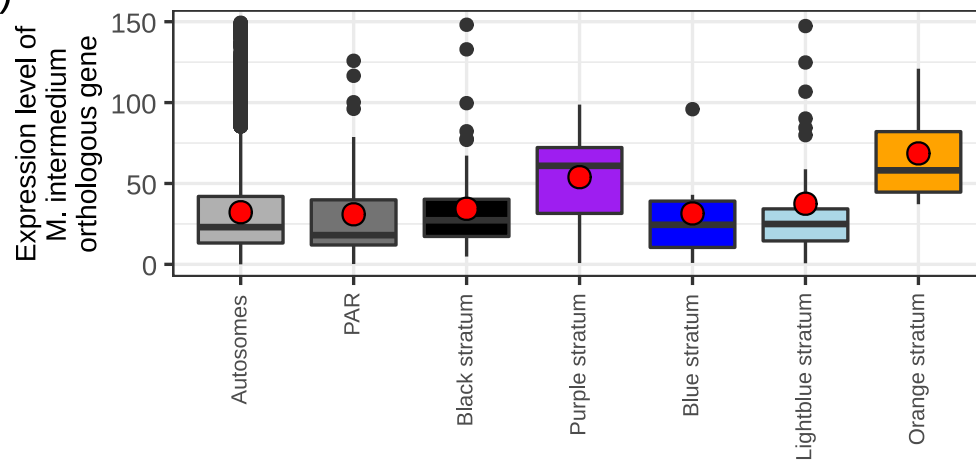
B)



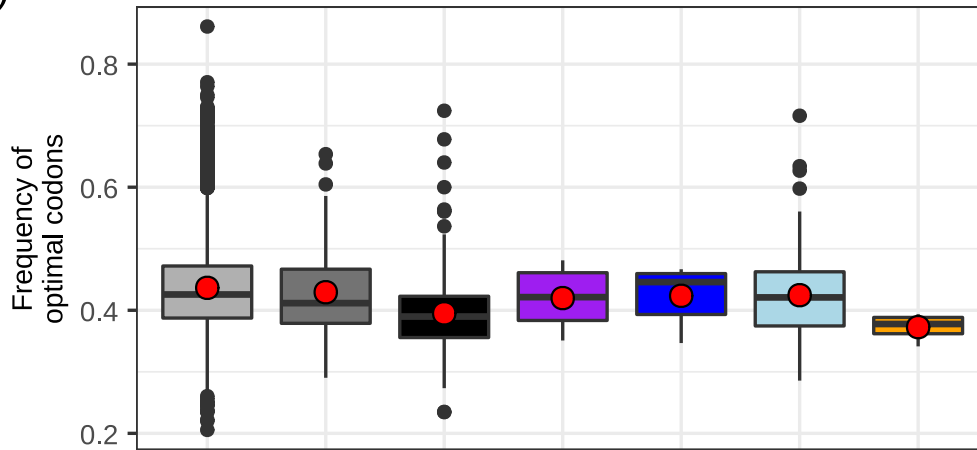
Genomic compartments

- Autosomes
- PAR
- Black stratum
- Purple stratum
- Blue stratum
- Lightblue stratum
- Orange stratum

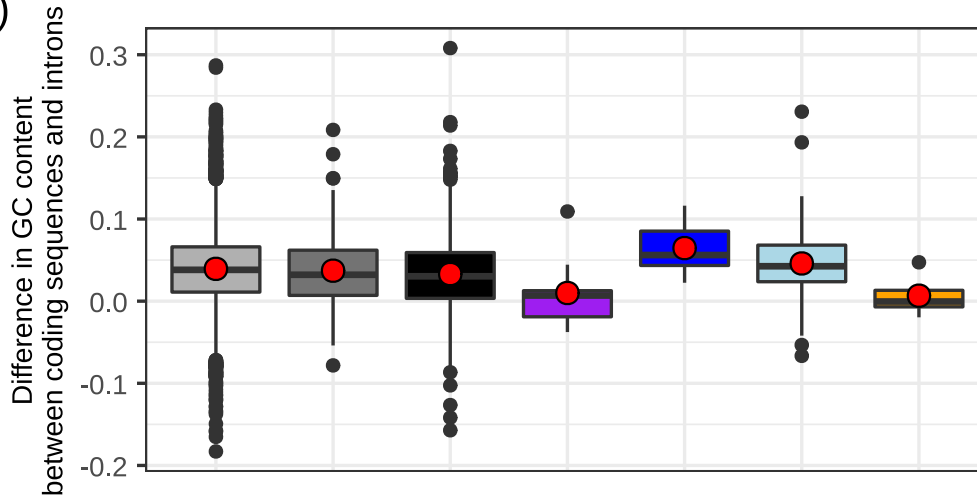
C)



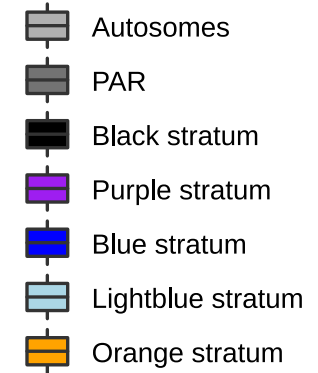
A)



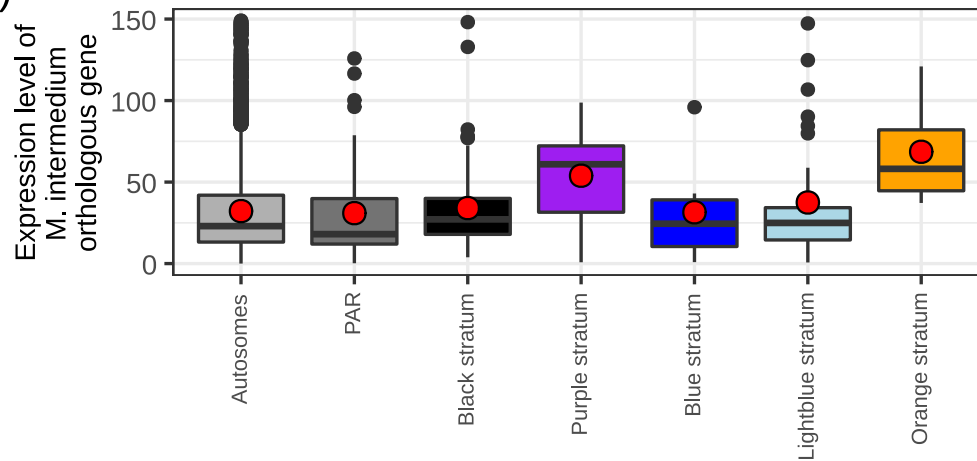
B)



Genomic compartments

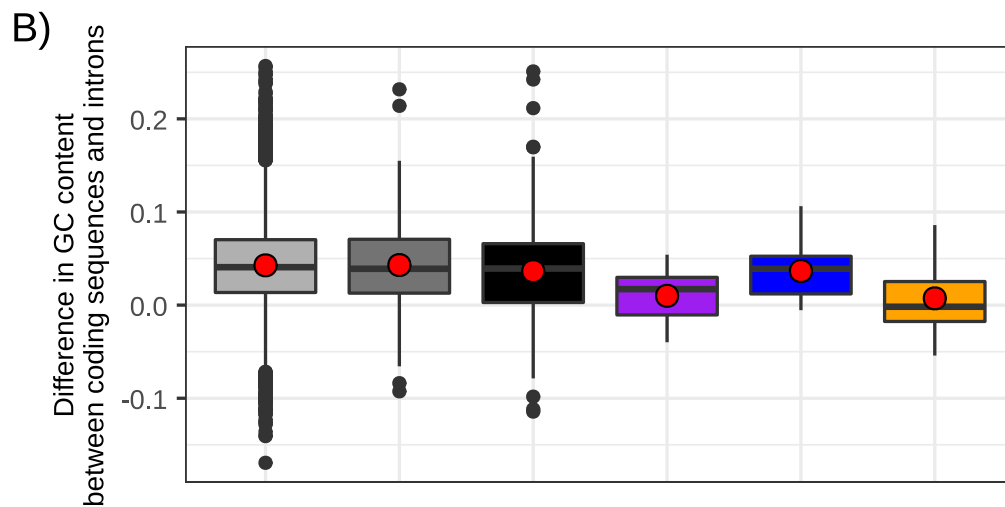
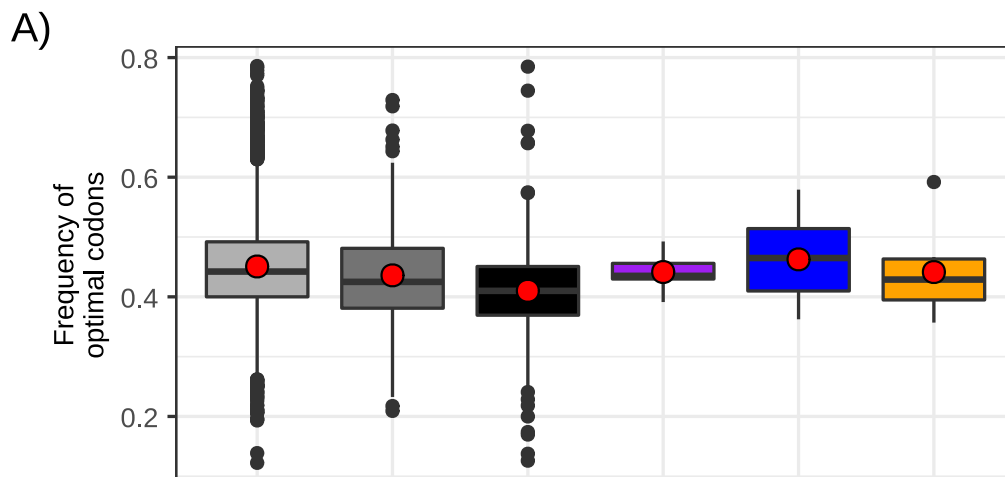


C)

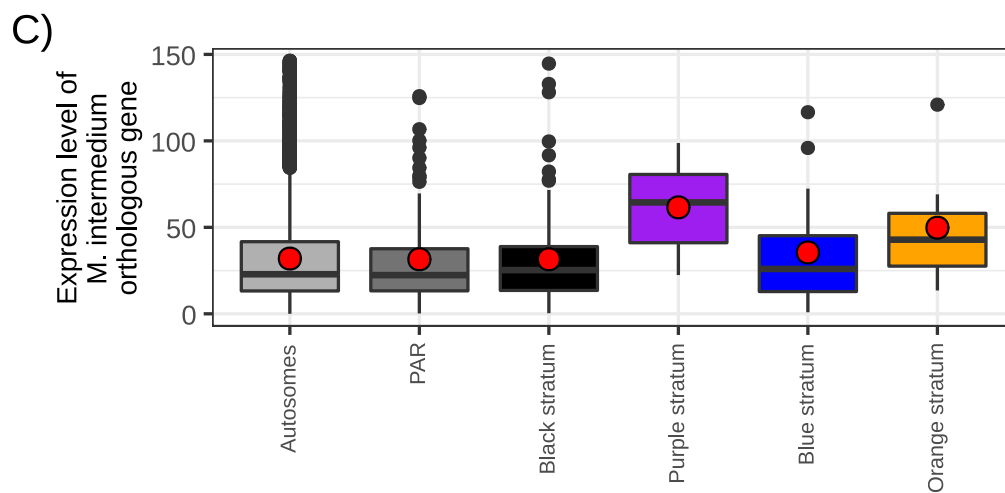
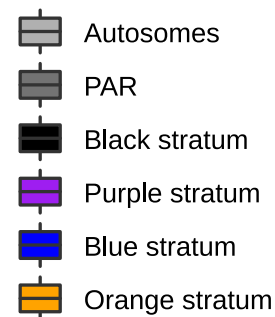




M. violaceum s.s.
on *Silene nutans*
a₁ genome

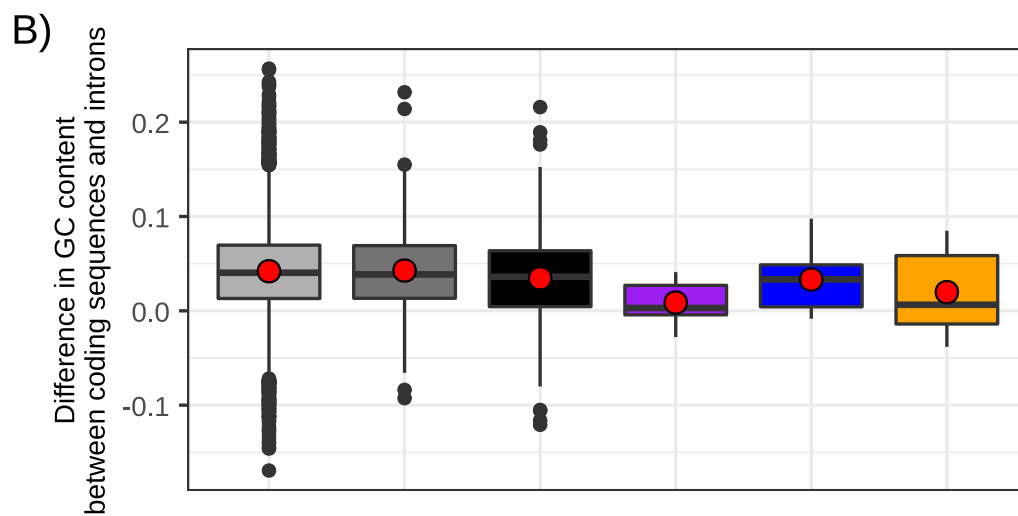
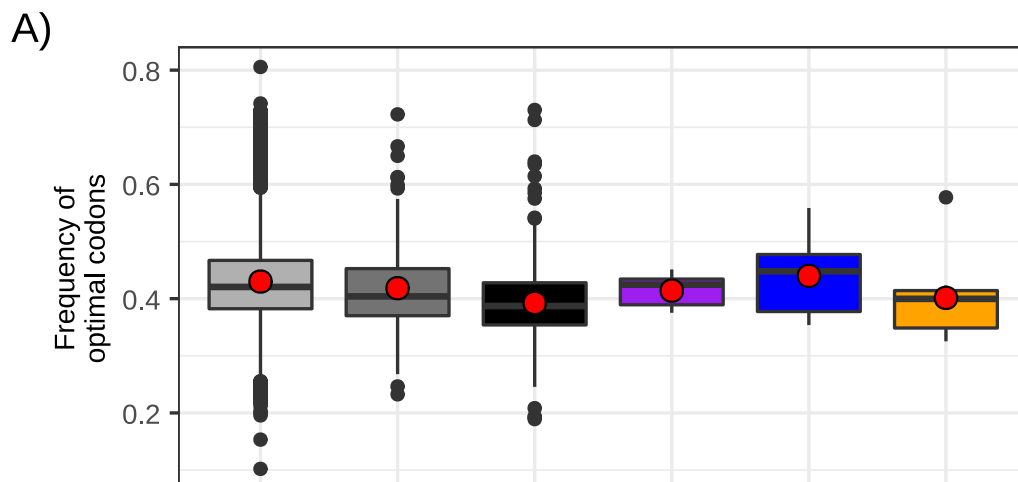


Genomic compartments

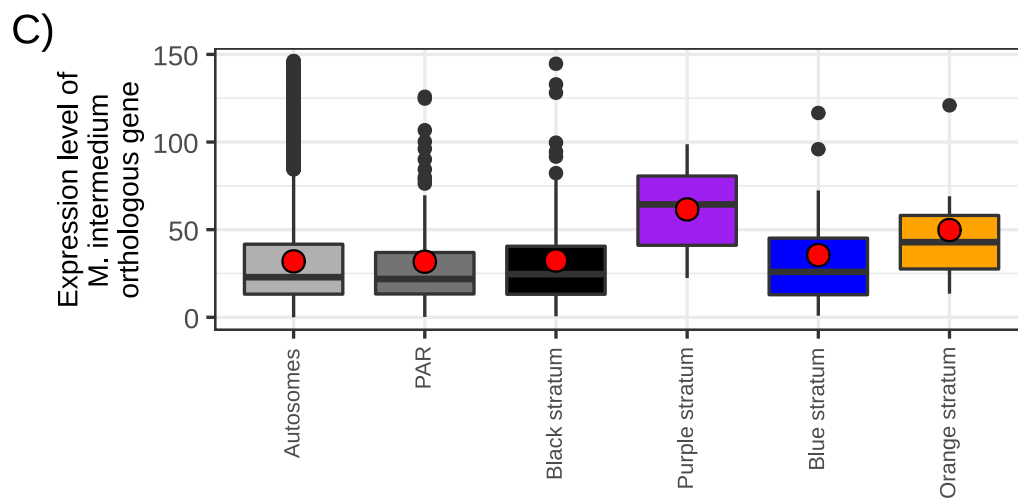
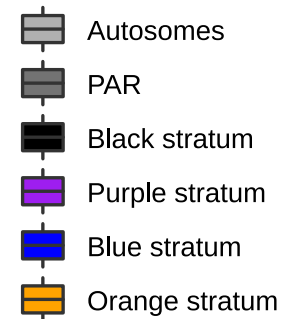




M. violaceum s.s.
on *Silene nutans*
a₂ genome



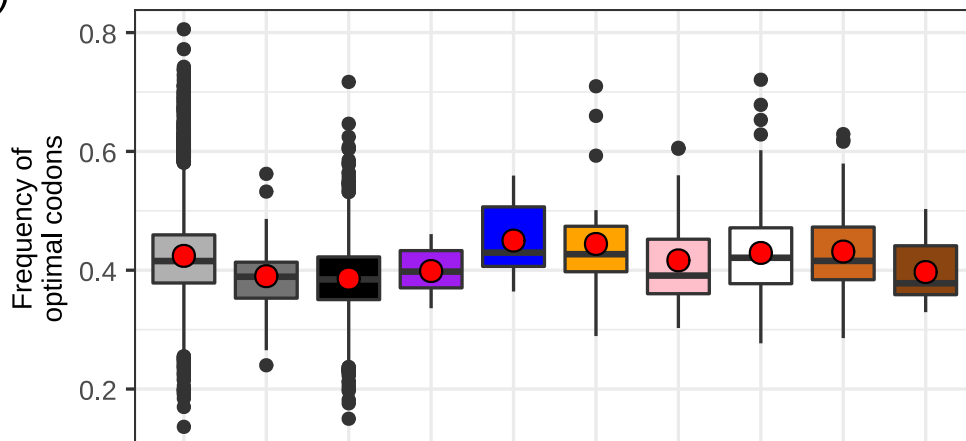
Genomic compartments



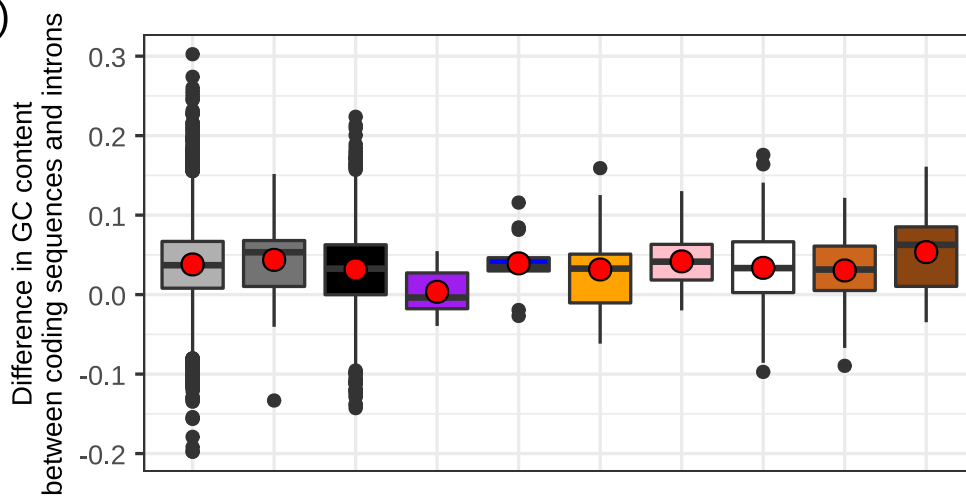


M. violaceum s.l.
on *Silene paradoxa*
a₁ genome

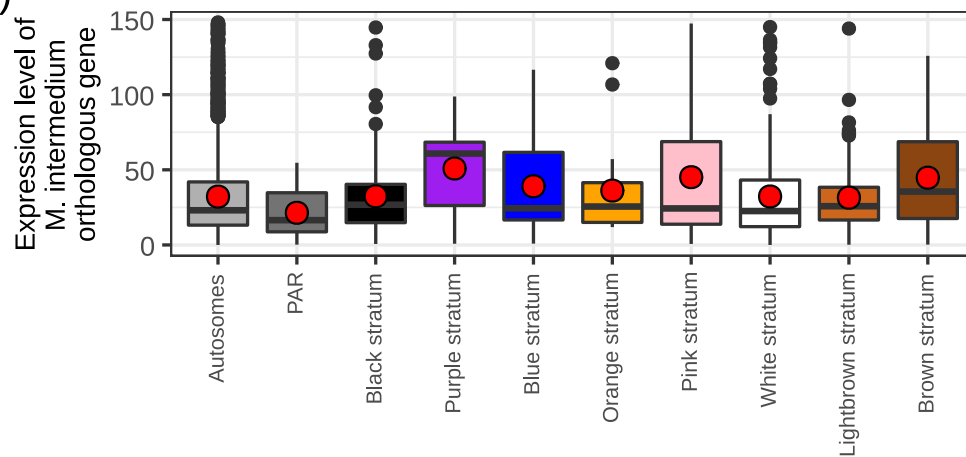
A)



B)



C)



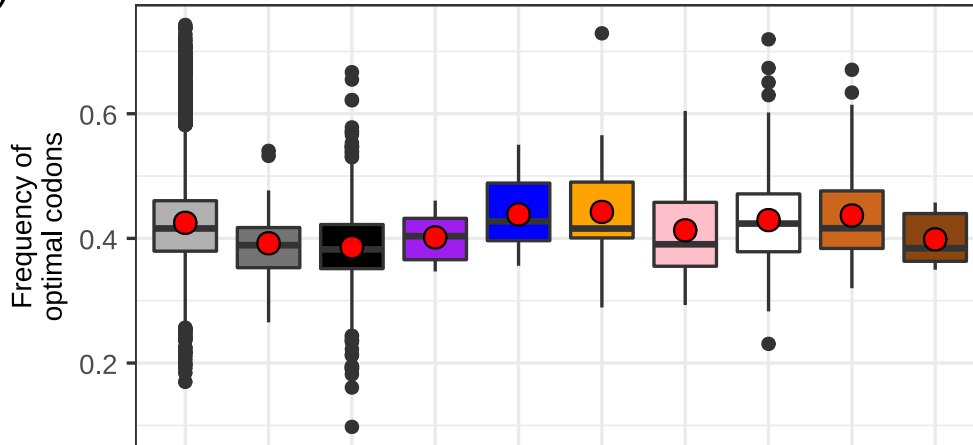
Genomic compartments

- Autosomes
- PAR
- Black stratum
- Purple stratum
- Blue stratum
- Orange stratum
- Pink stratum
- White stratum
- Lightbrown stratum
- Brown stratum

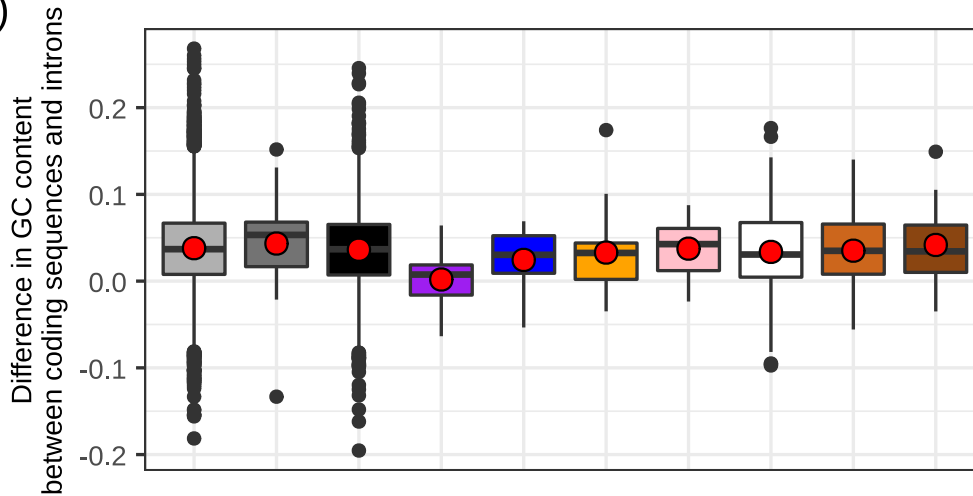


M. violaceum s.l.
on *Silene paradoxa*
a₂ genome

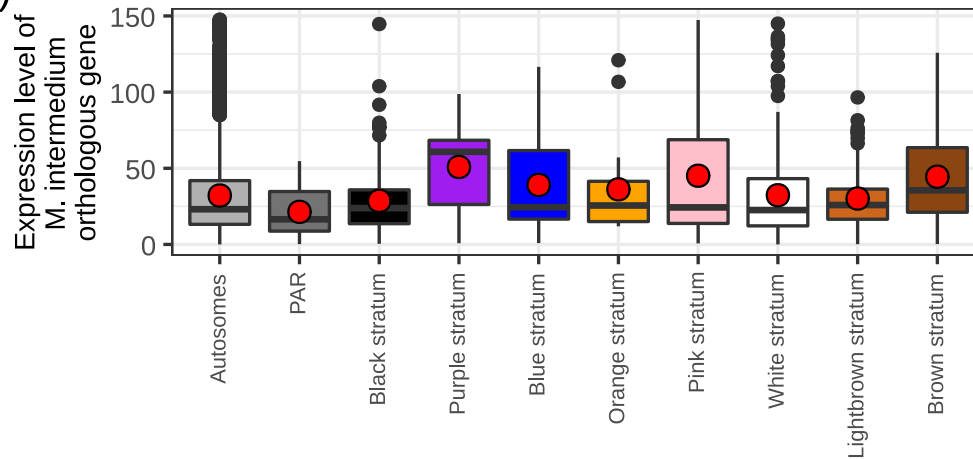
A)



B)



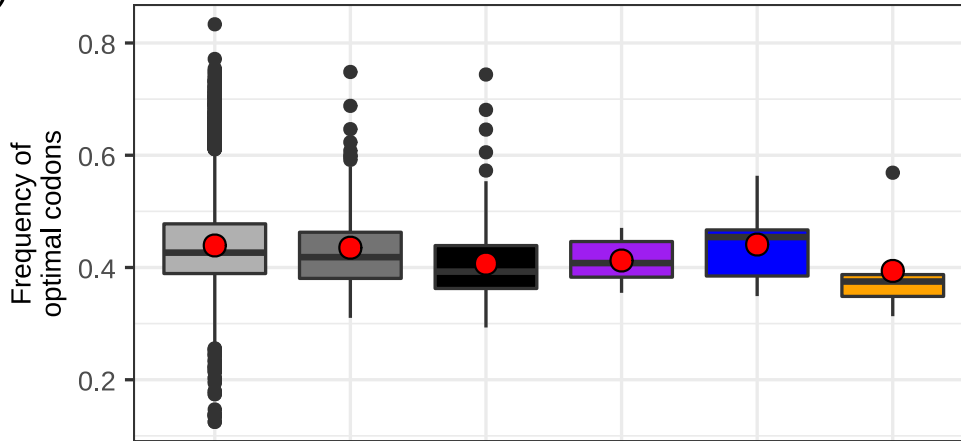
C)



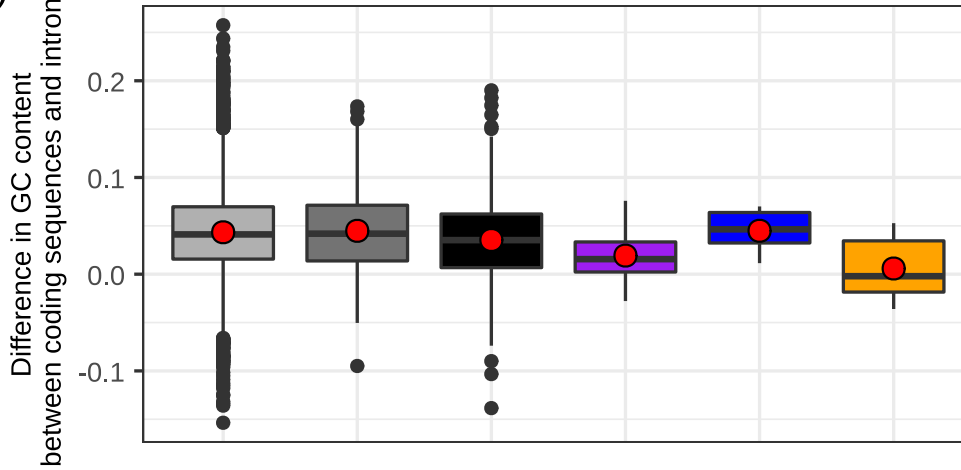
Genomic compartments

- Autosomes
- PAR
- Black stratum
- Purple stratum
- Blue stratum
- Orange stratum
- Pink stratum
- White stratum
- Lightbrown stratum
- Brown stratum

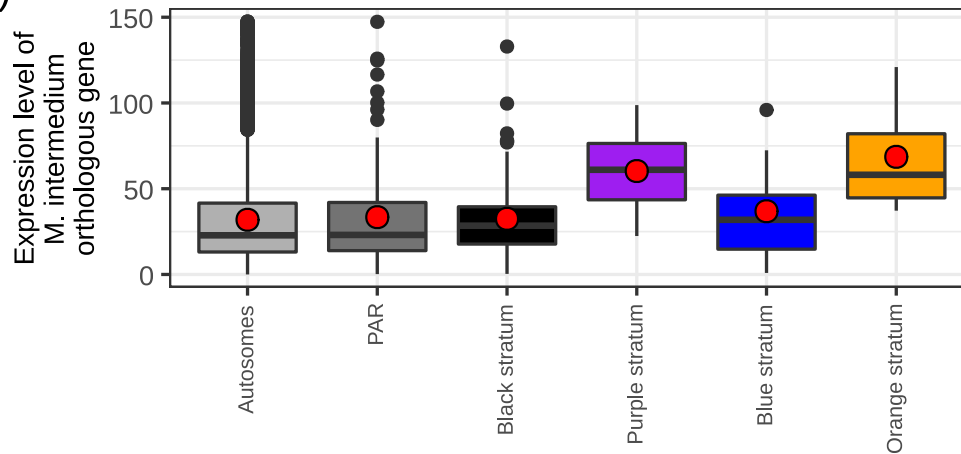
A)



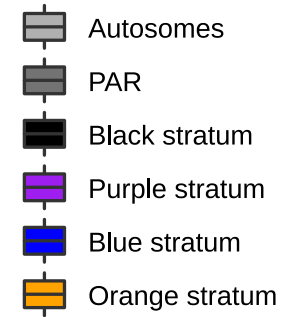
B)



C)



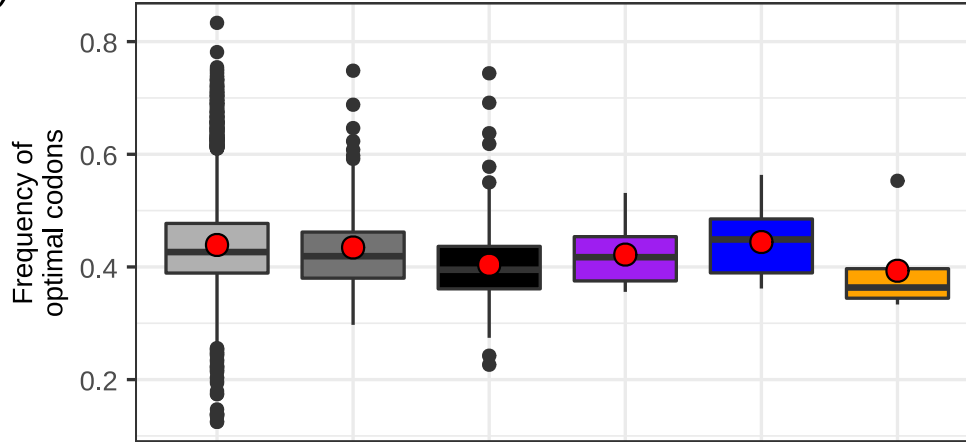
Genomic compartments



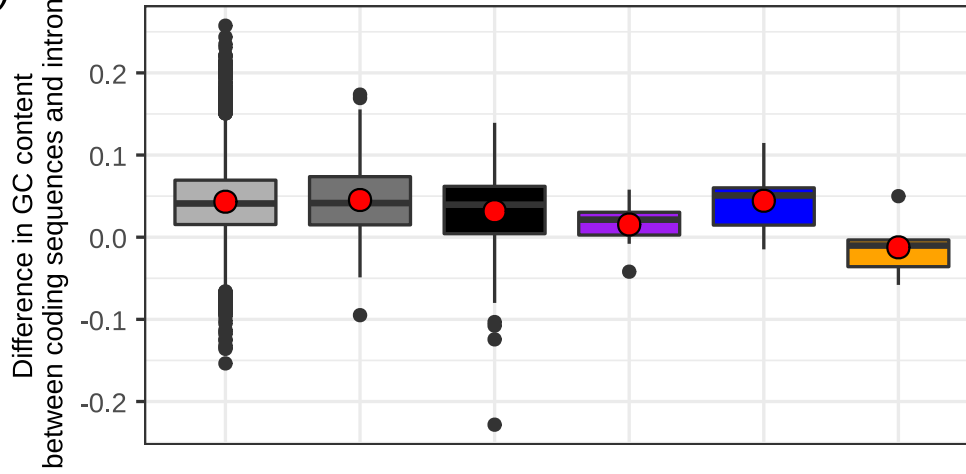


M. violaceum s.l.
on *Silene tatarinowii*
a₂ genome

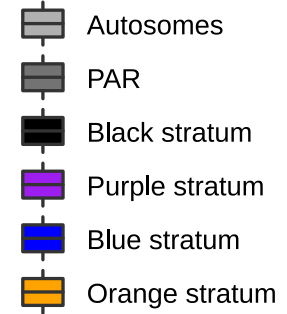
A)



B)



Genomic compartments



C)

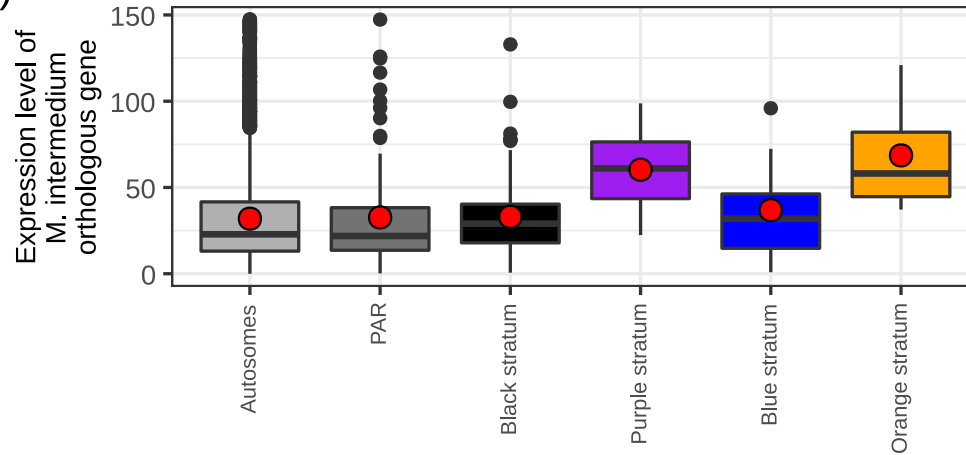
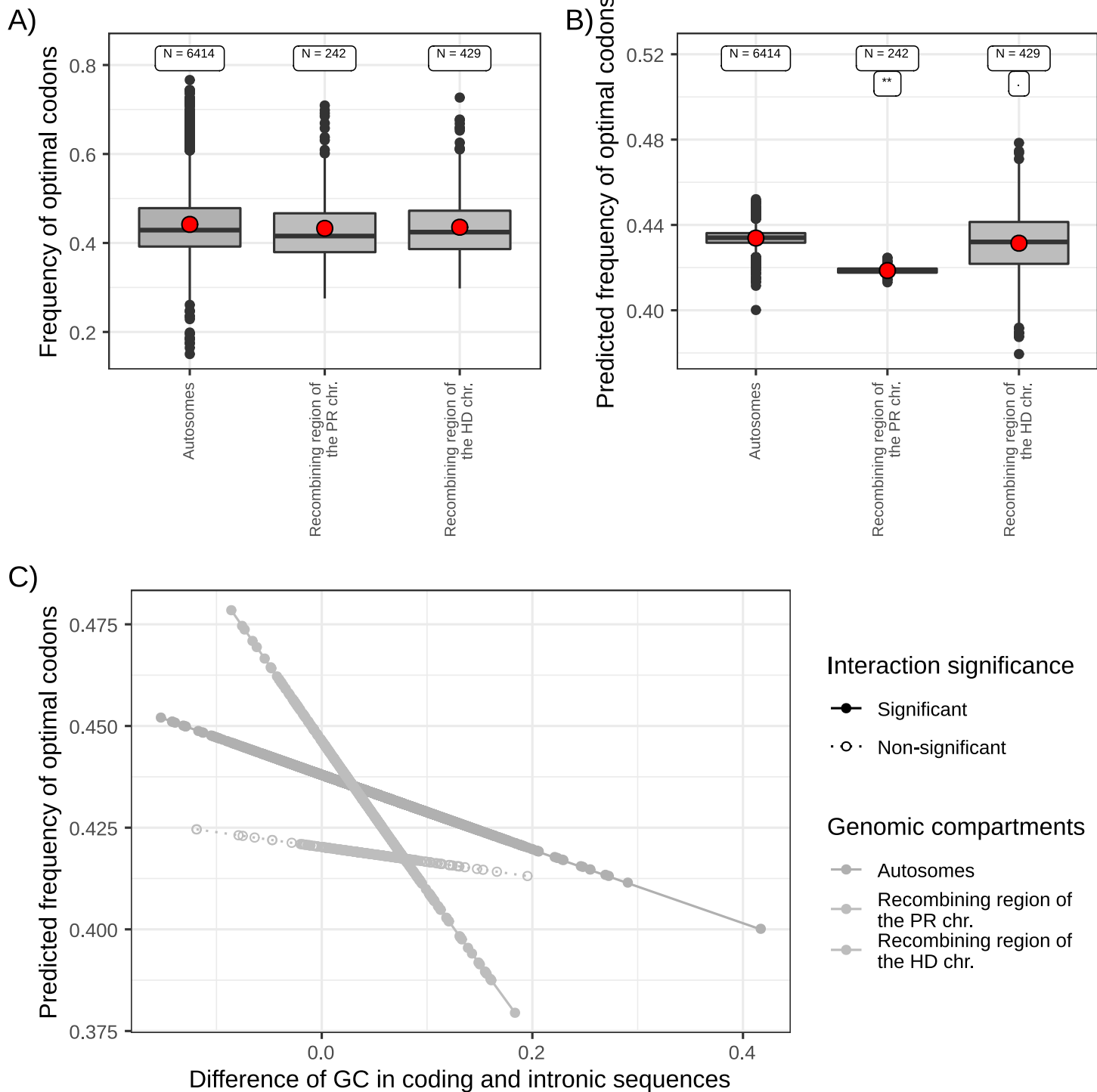


Figure S7

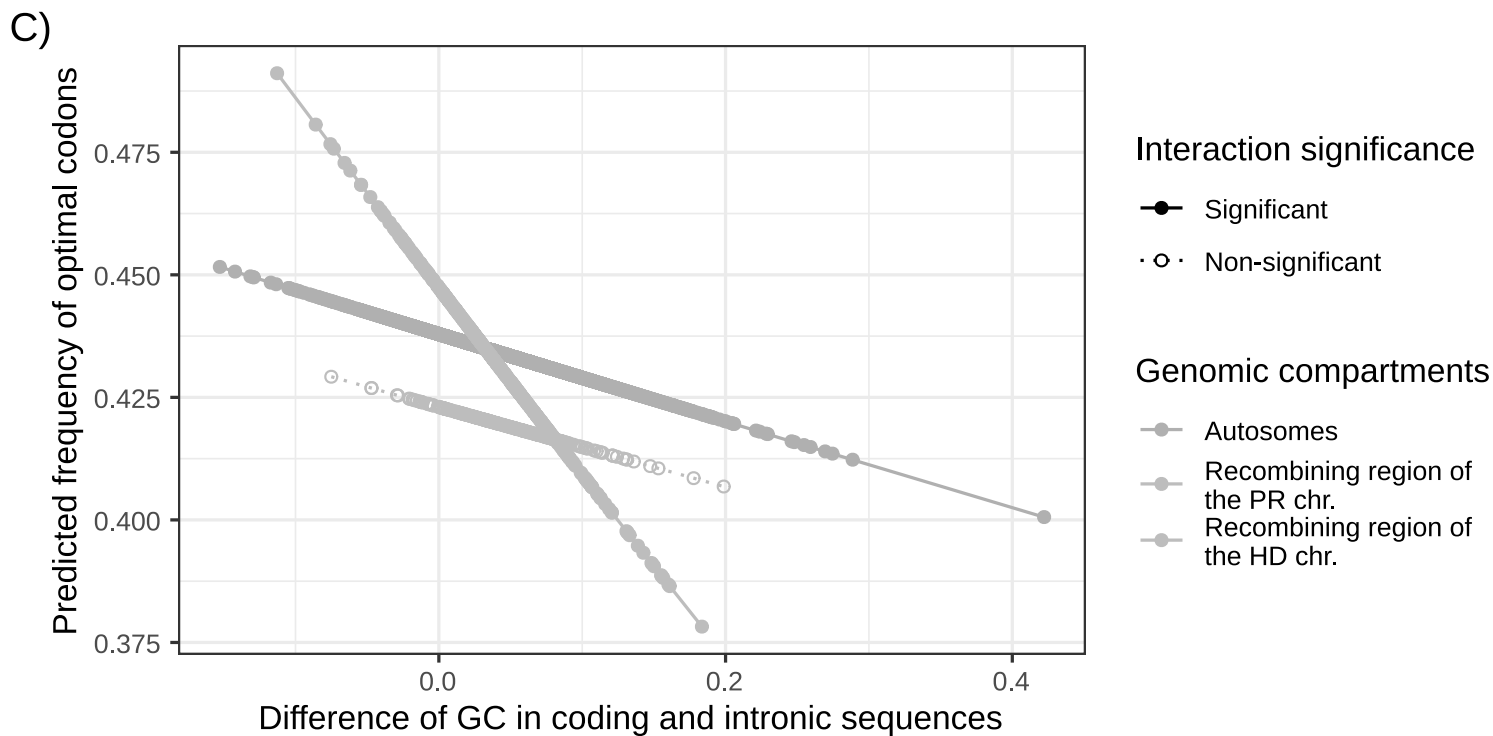
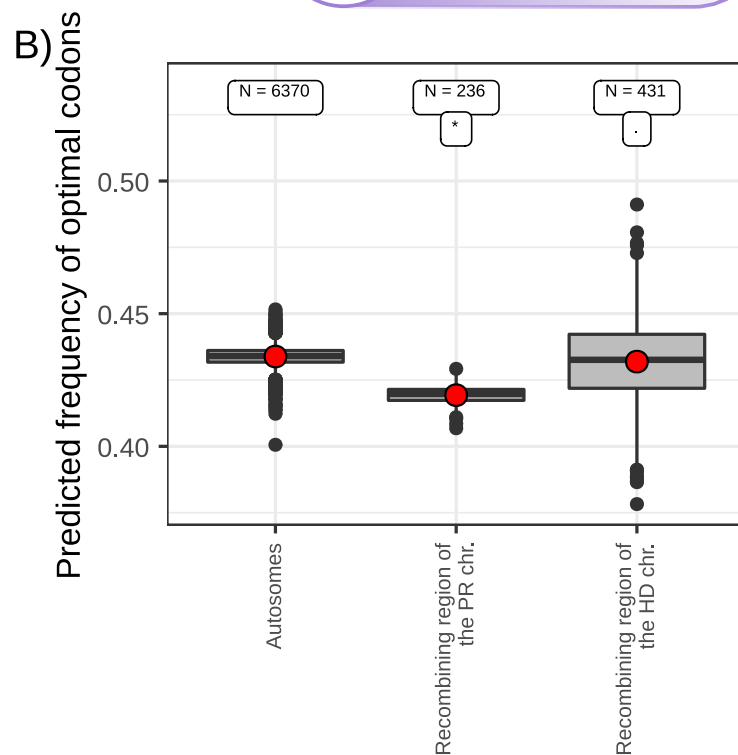
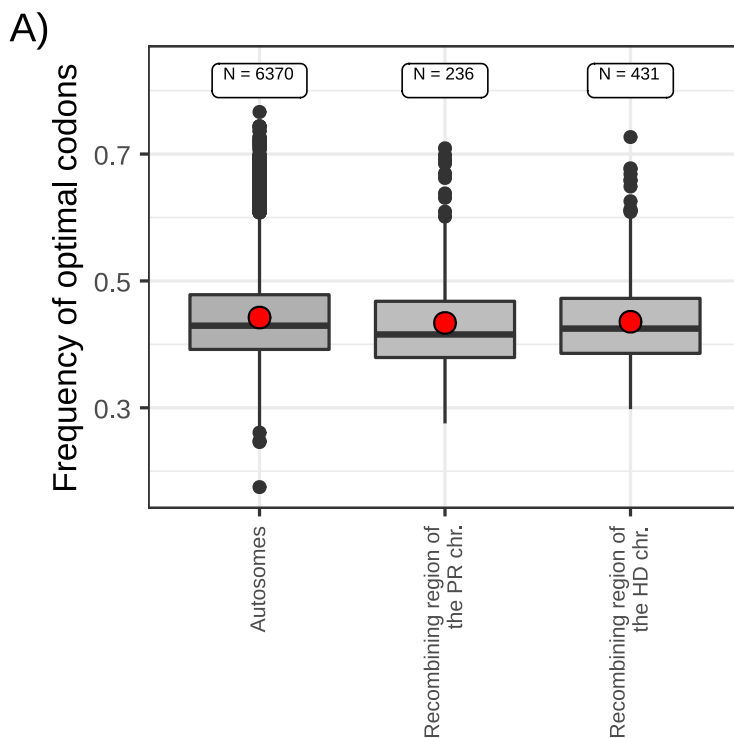


M. intermedius
on *Salvia pratensis*
a₁ genome



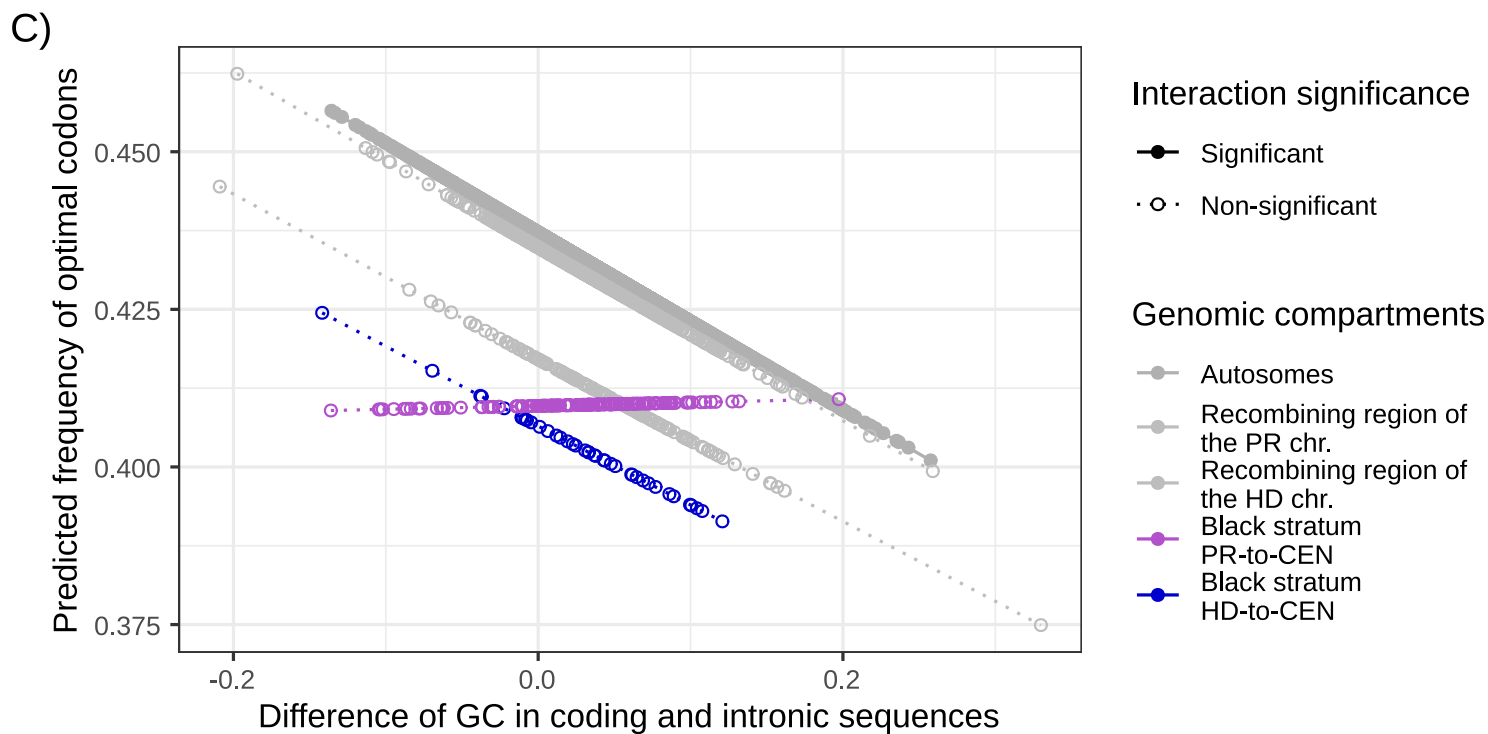
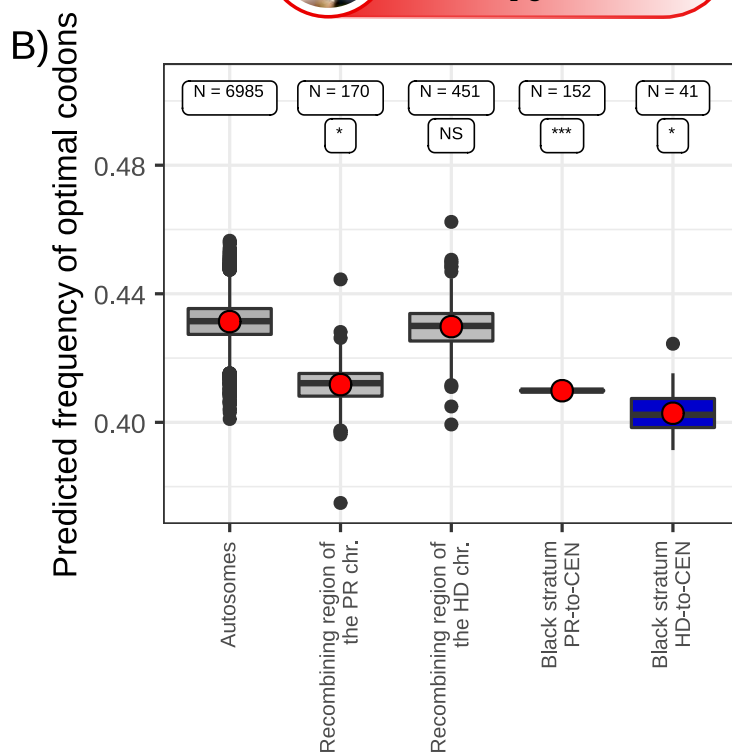
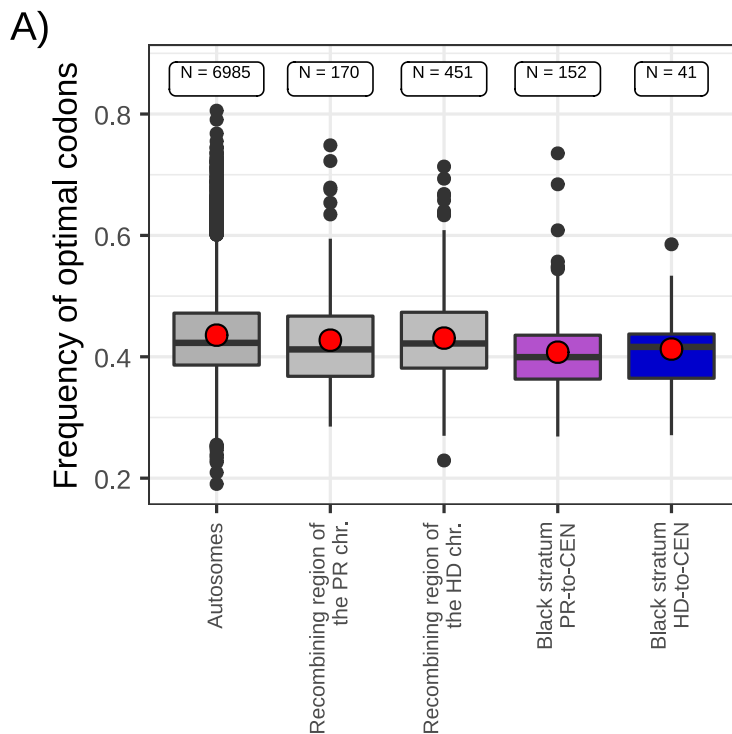


M. intermedium
on *Salvia pratensis*
a₂ genome



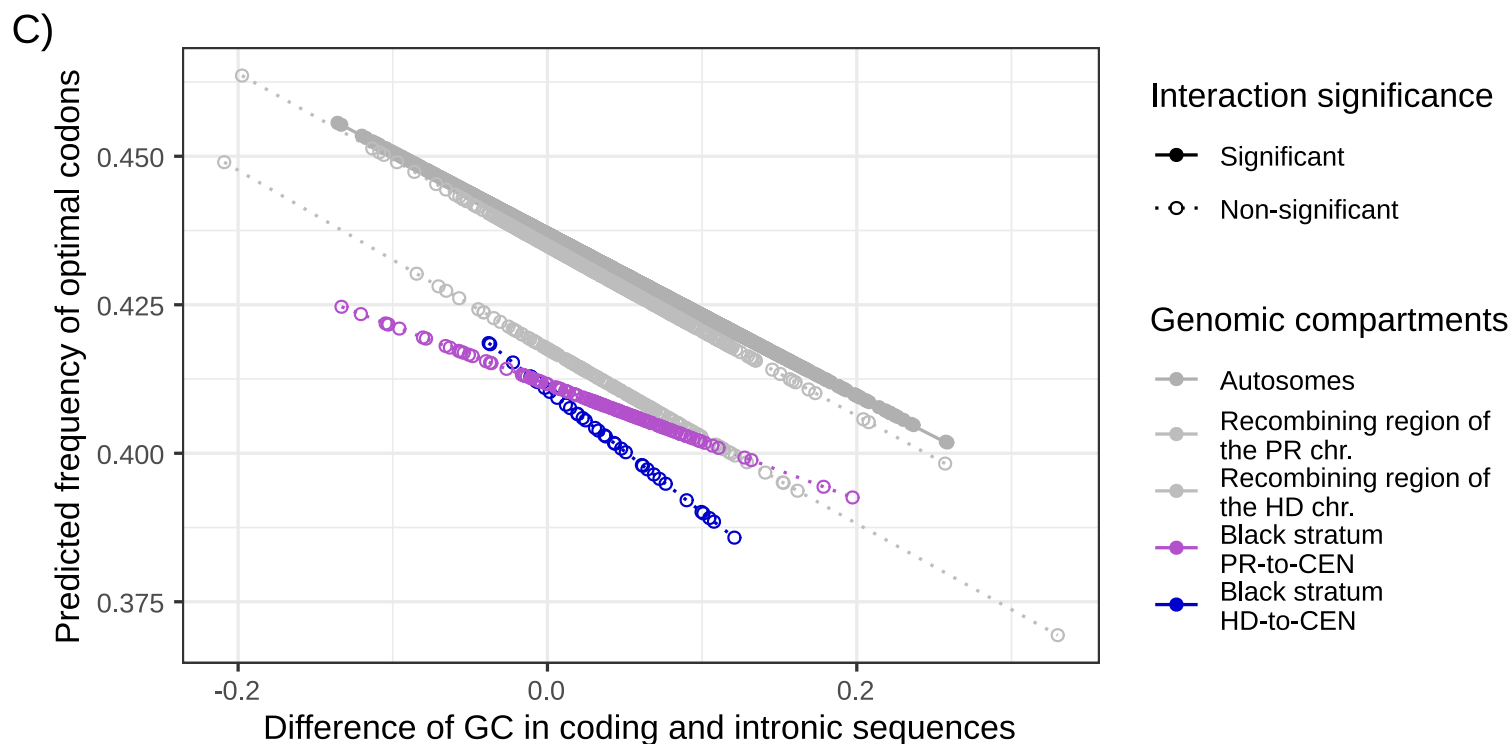
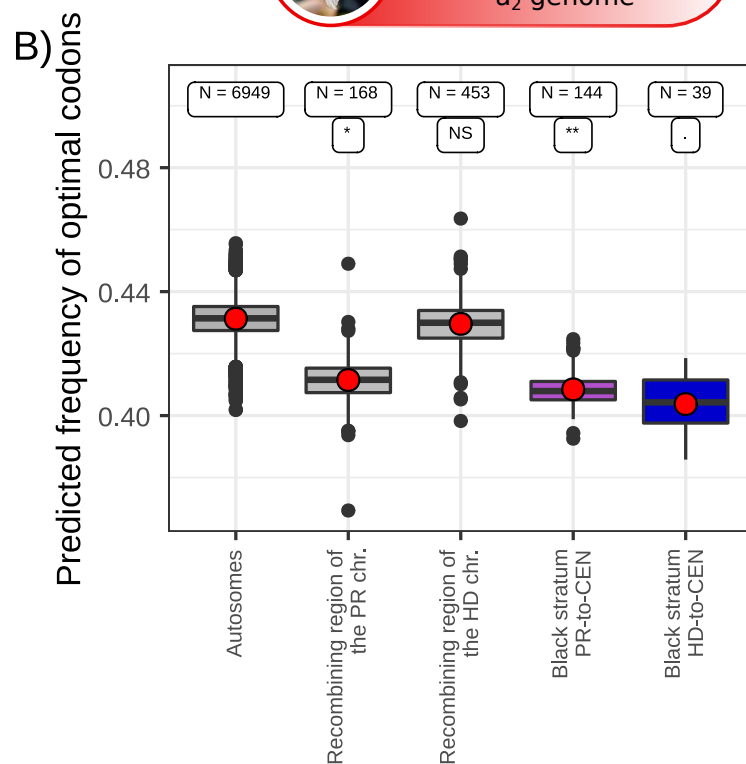
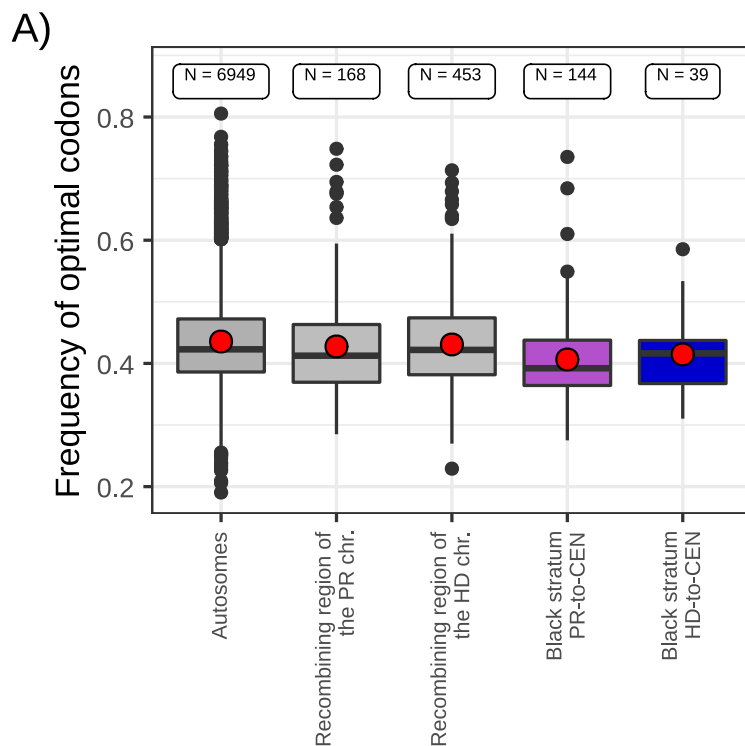


M. lagerheimii
on *Silene vulgaris*
a₁ genome





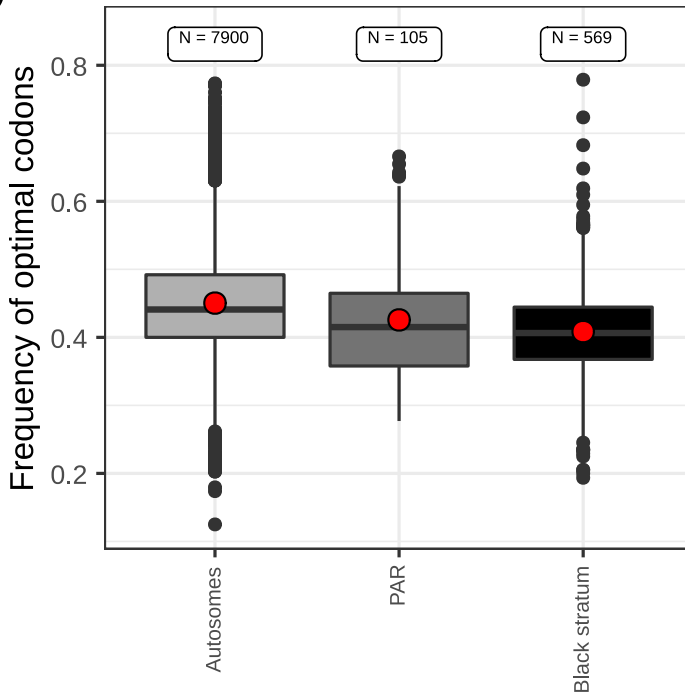
M. lagerheimii
on *Silene vulgaris*
a₂ genome



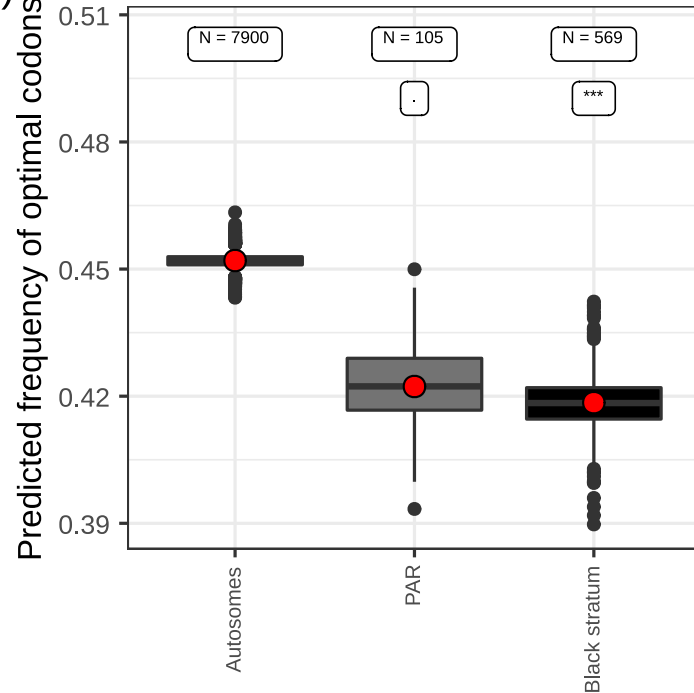


M. lychnidis-dioecae
on *Silene latifolia*
a₁ genome

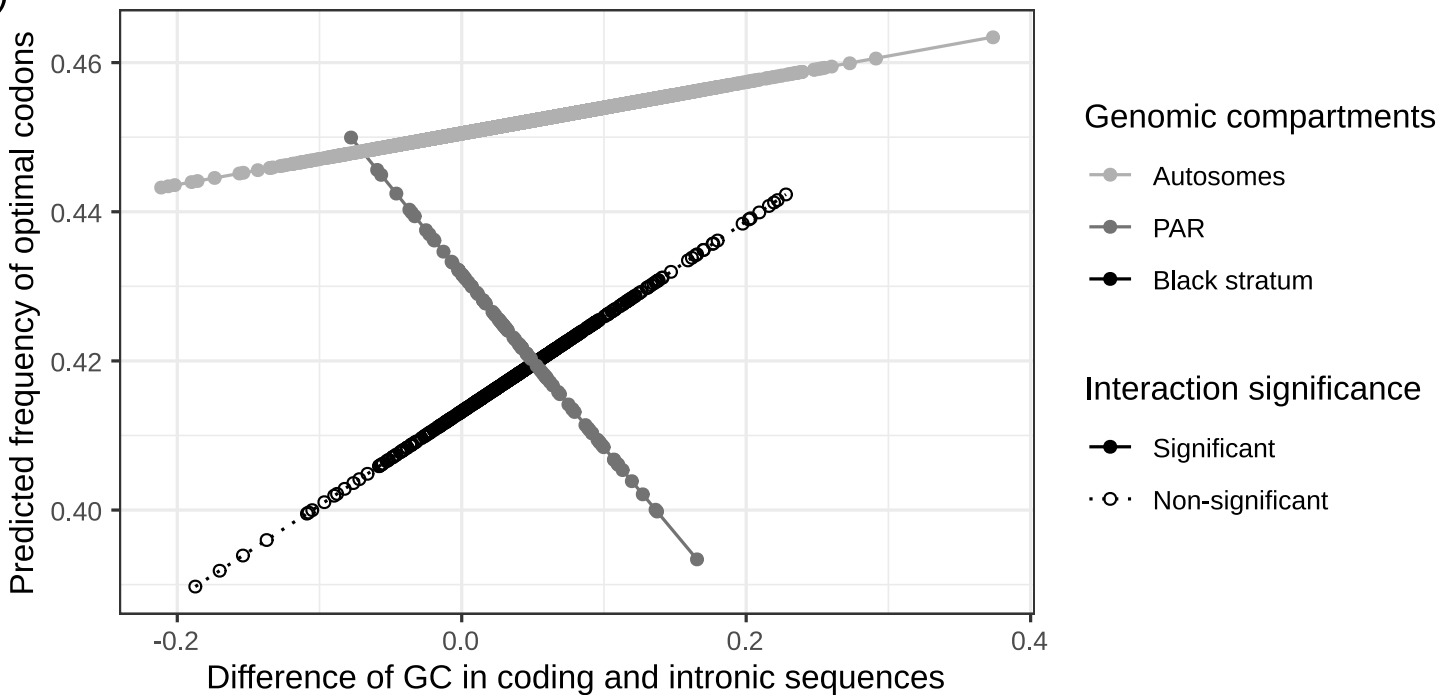
A)



B)



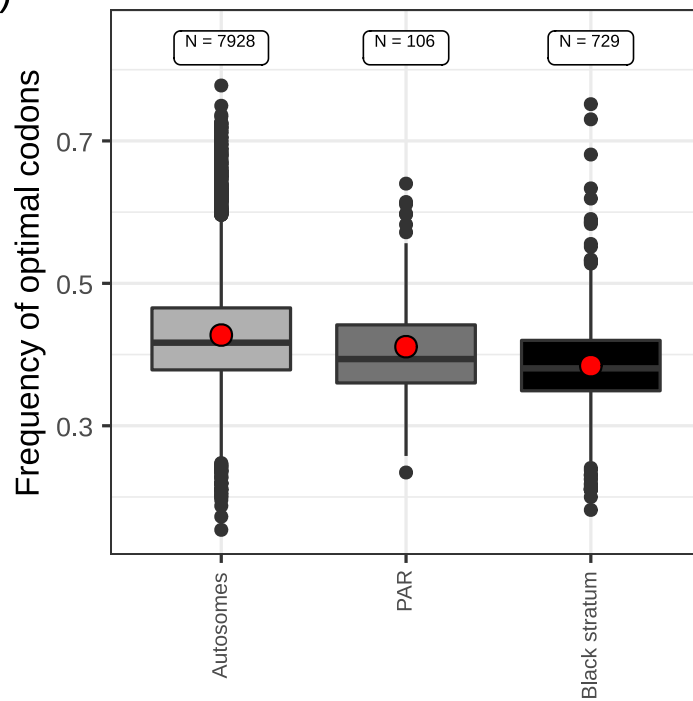
C)



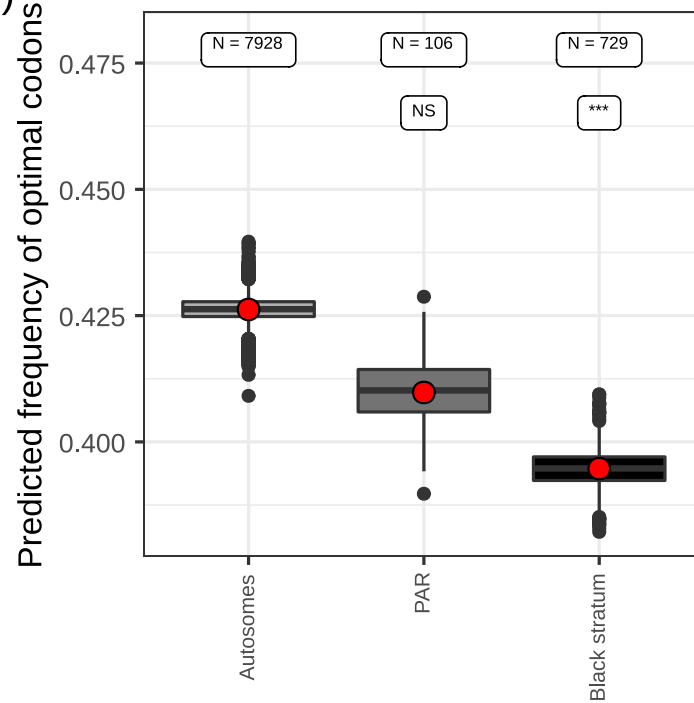


M. lychnidis-dioicae
on *Silene latifolia*
a₂ genome

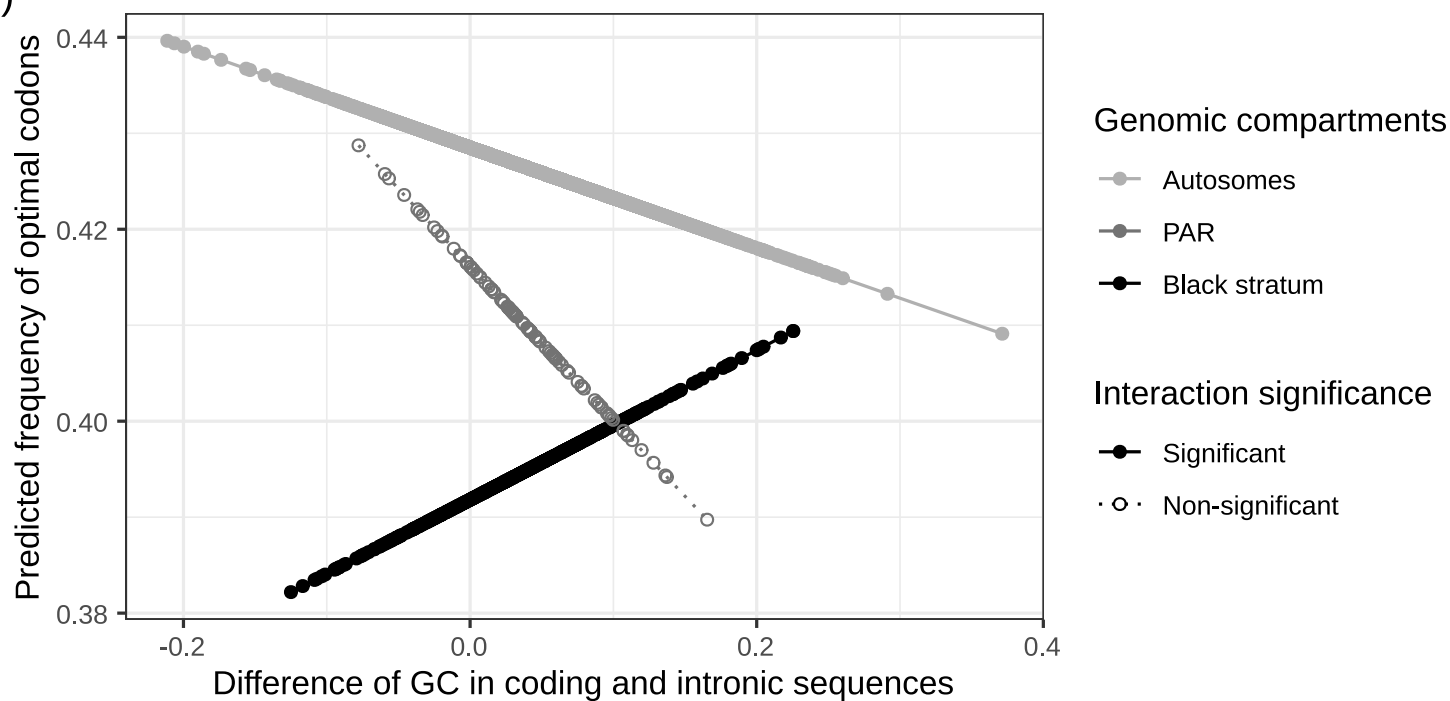
A)



B)



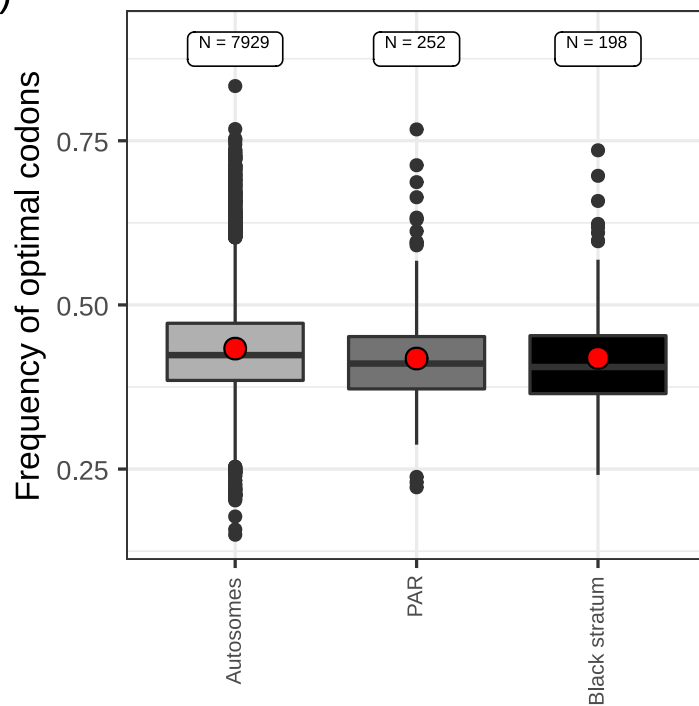
C)



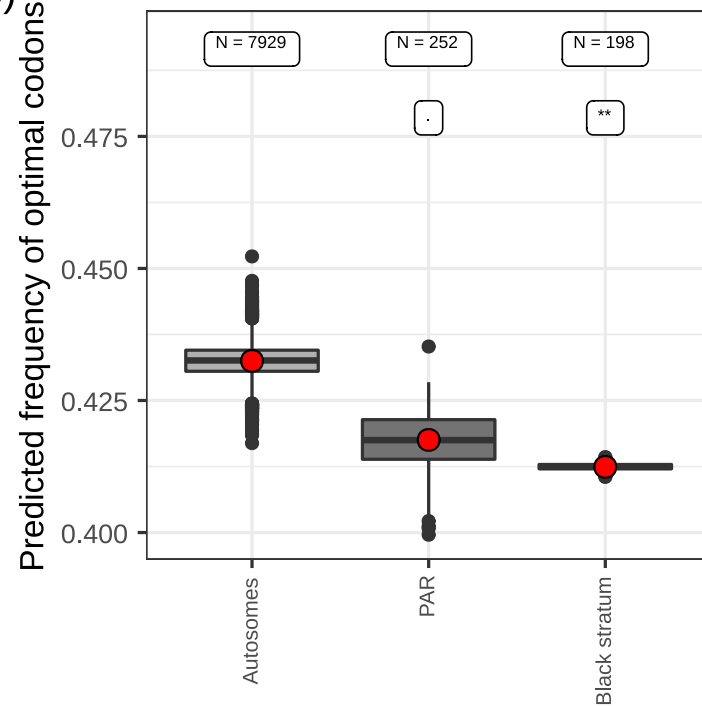


M. silenes-acaulis
on *Silene acaulis*
a₁ genome

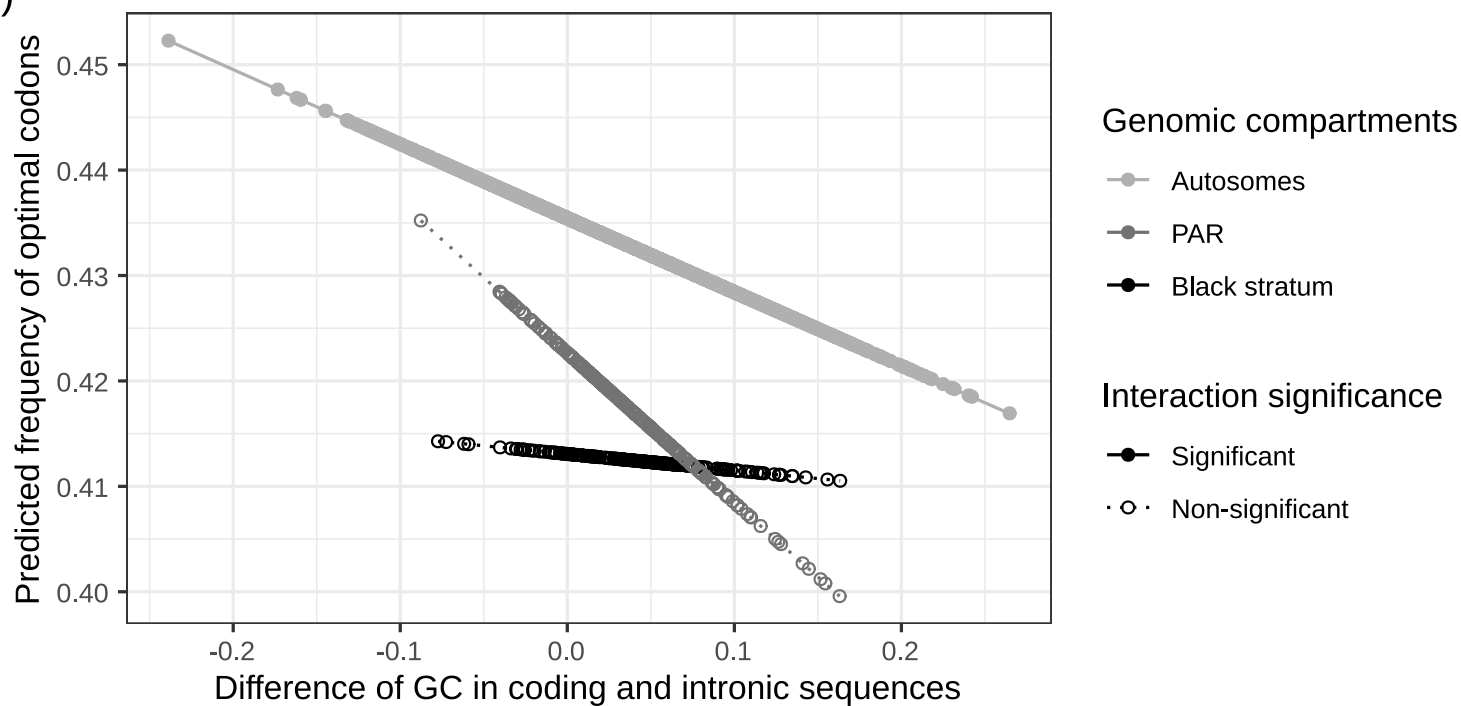
A)



B)



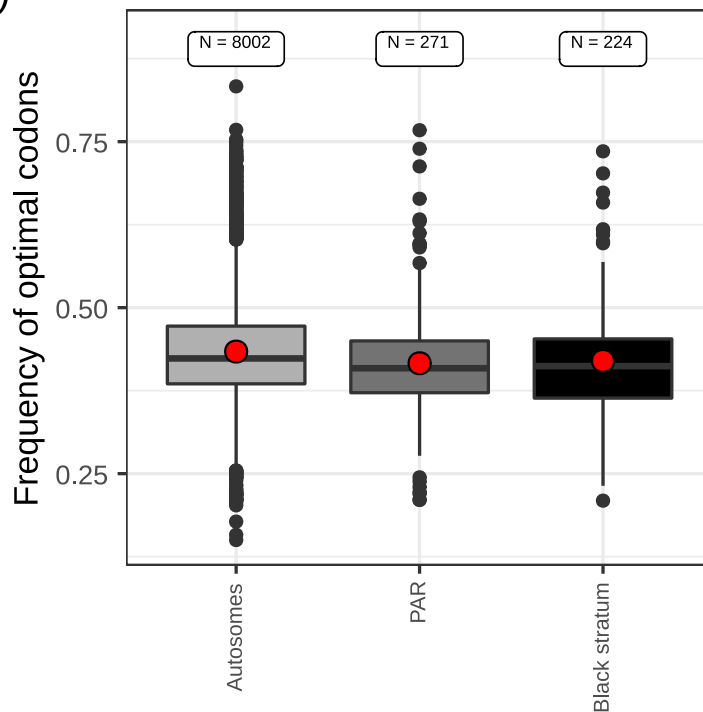
C)



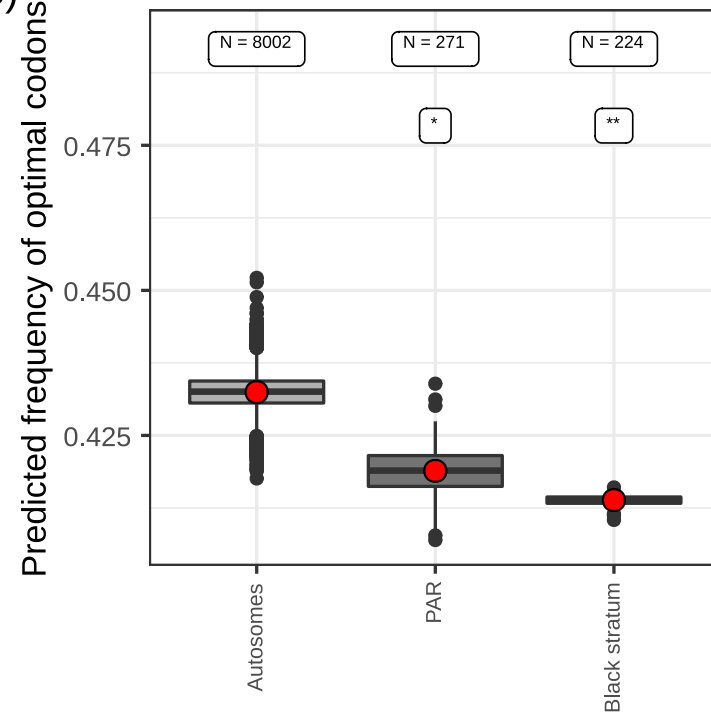


M. silenes-acaulis
on *Silene acaulis*
a₂ genome

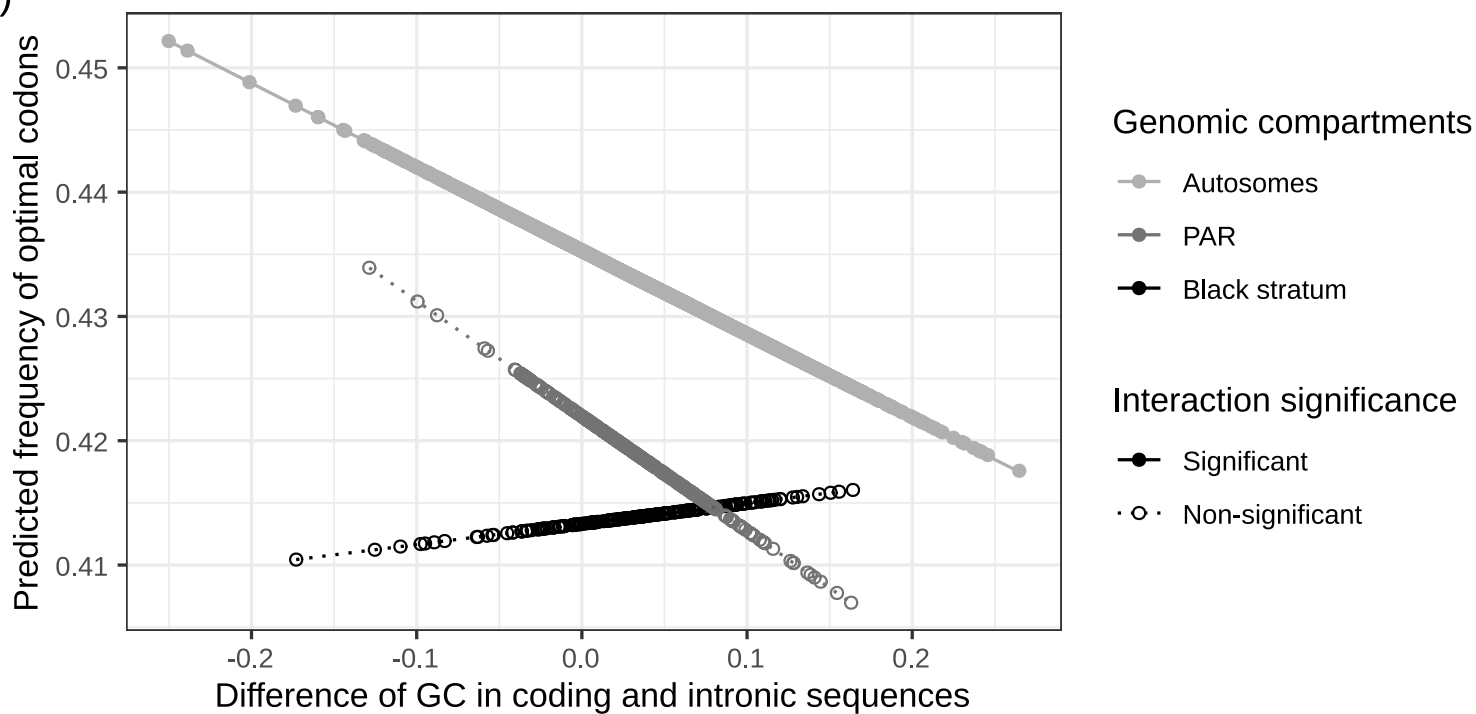
A)



B)

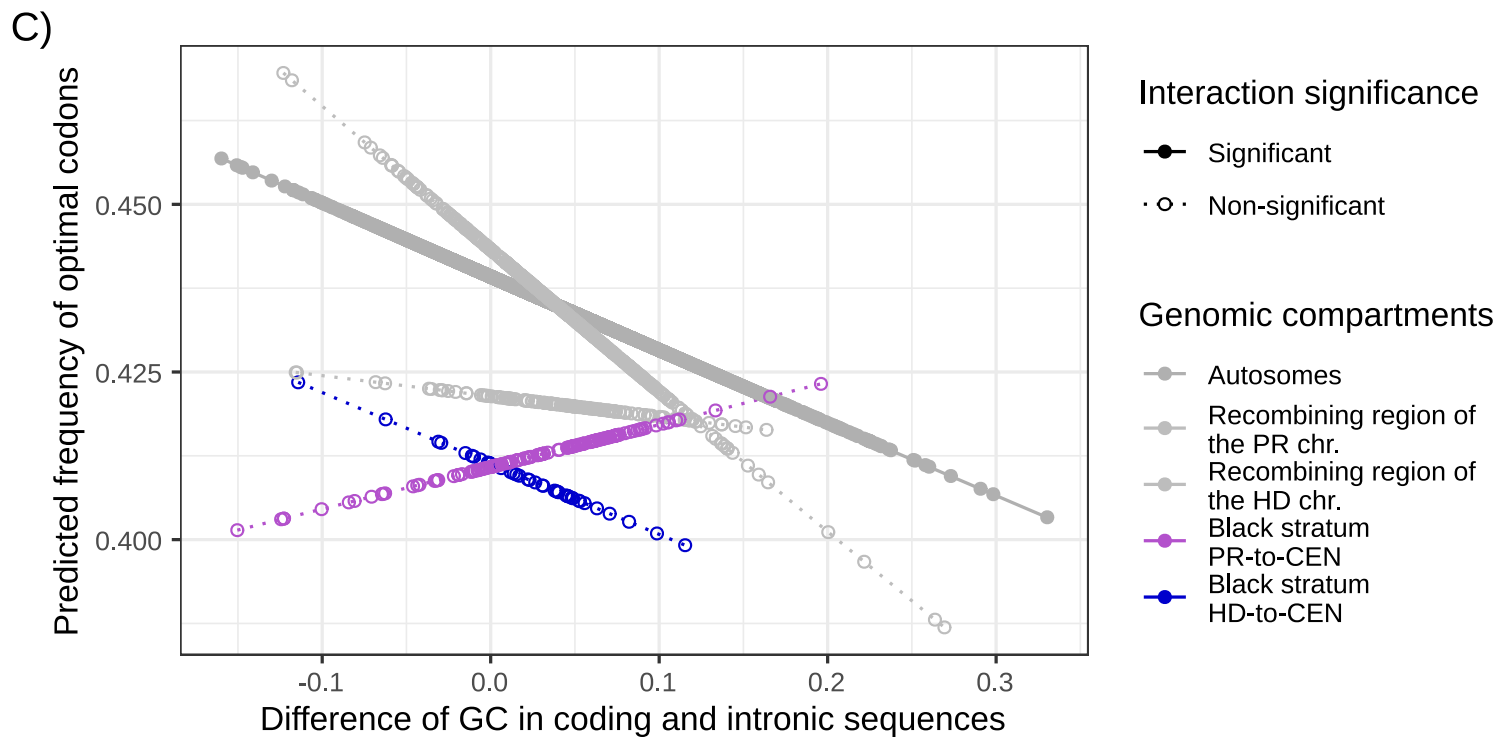
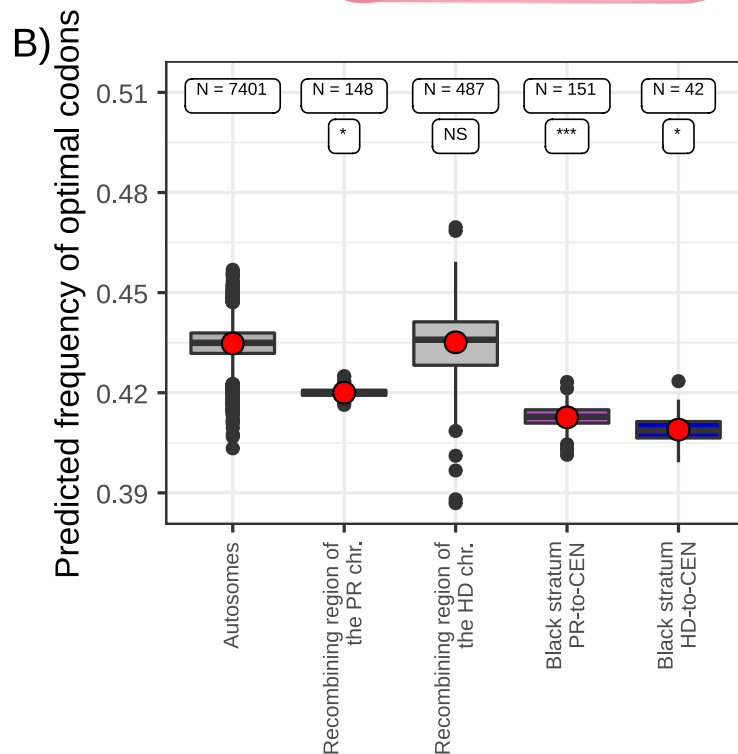
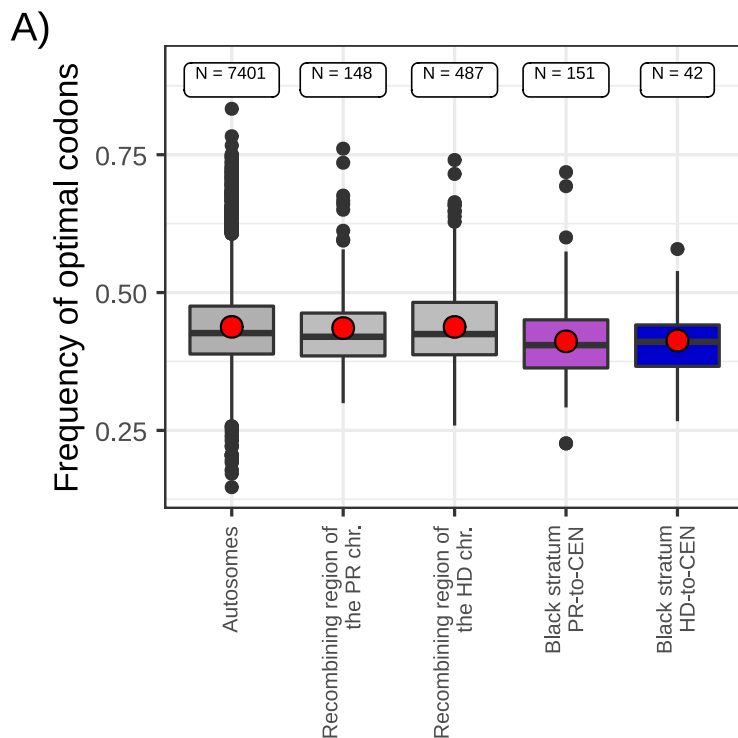


C)





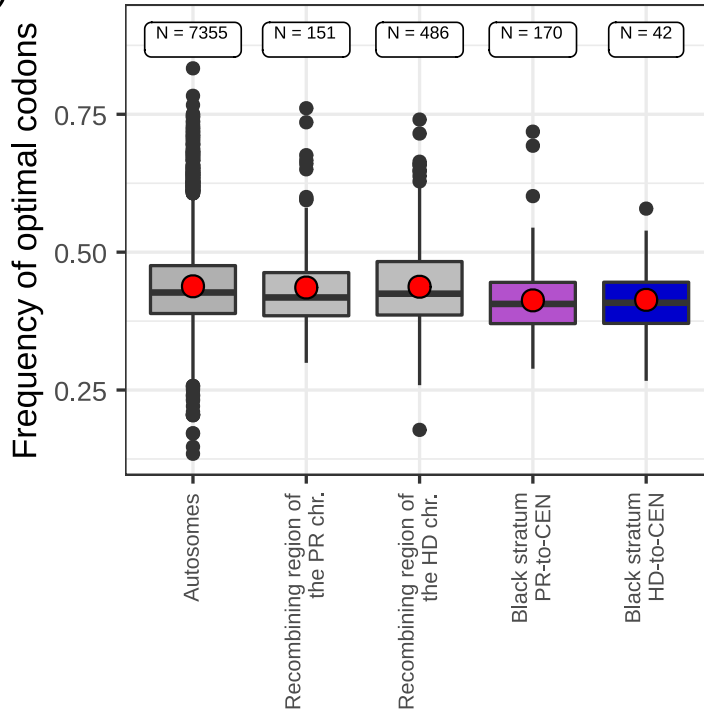
M. saponariae
on *Saponaria officinalis*
a₁ genome



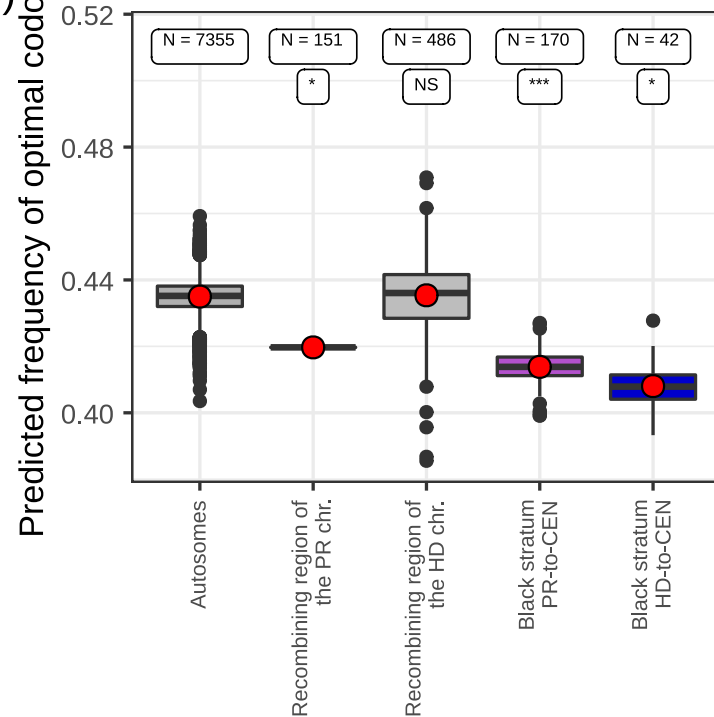


M. saponaria
on *Saponaria officinalis*
a₂ genome

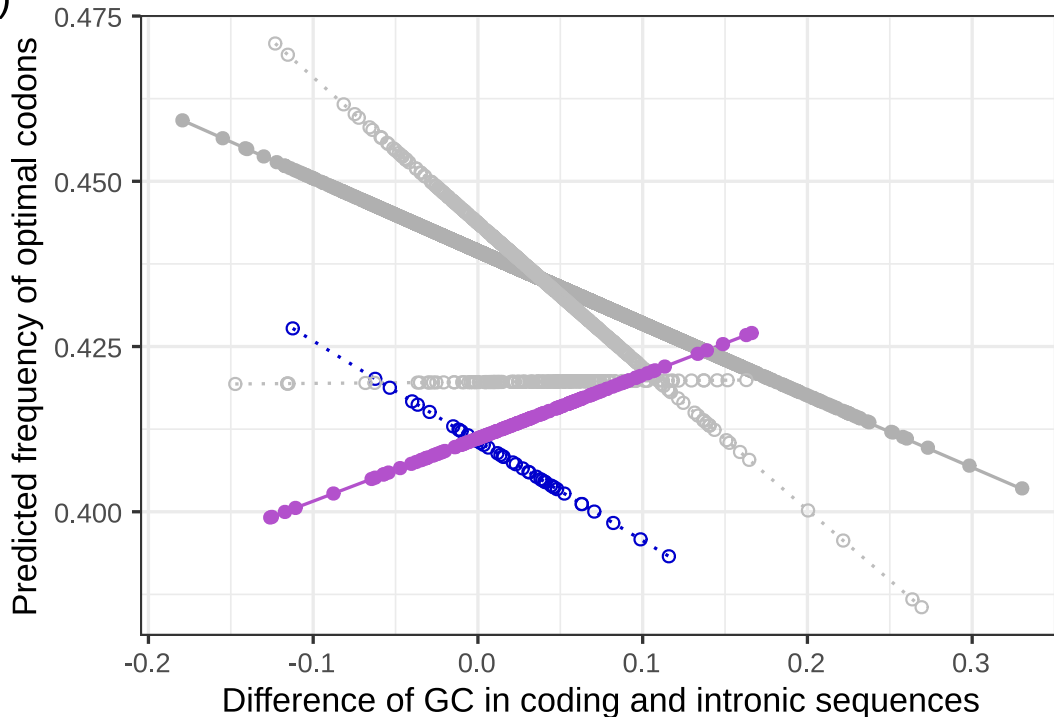
A)



B)



C)

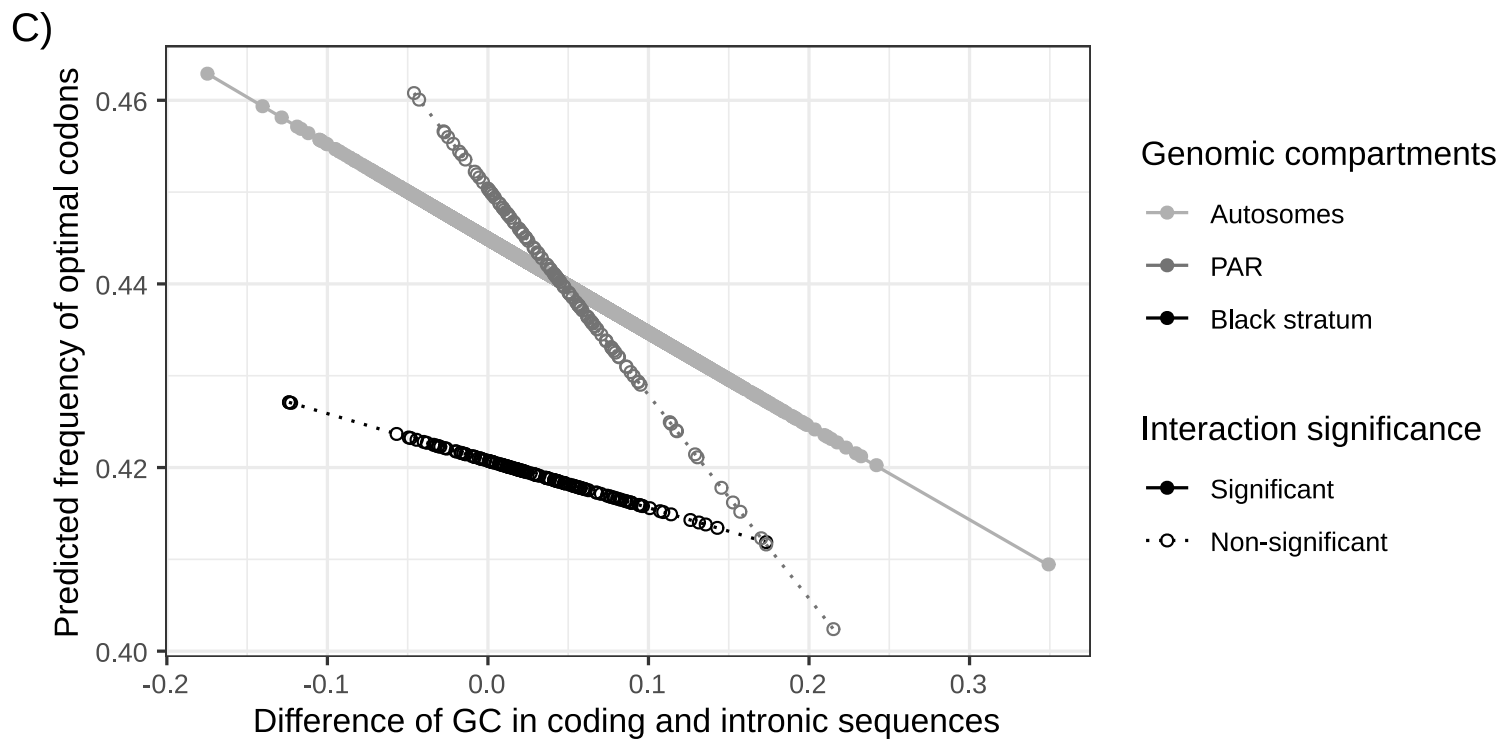
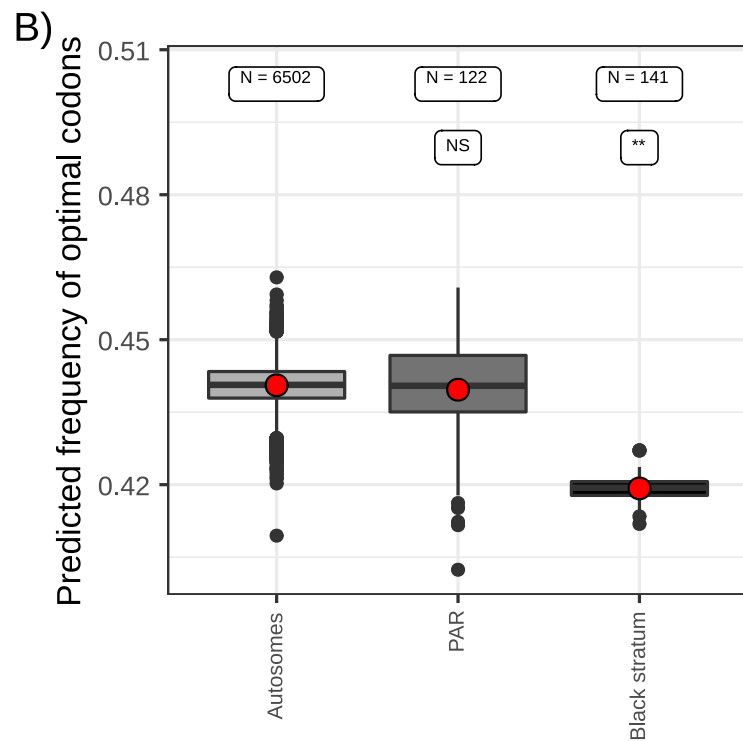
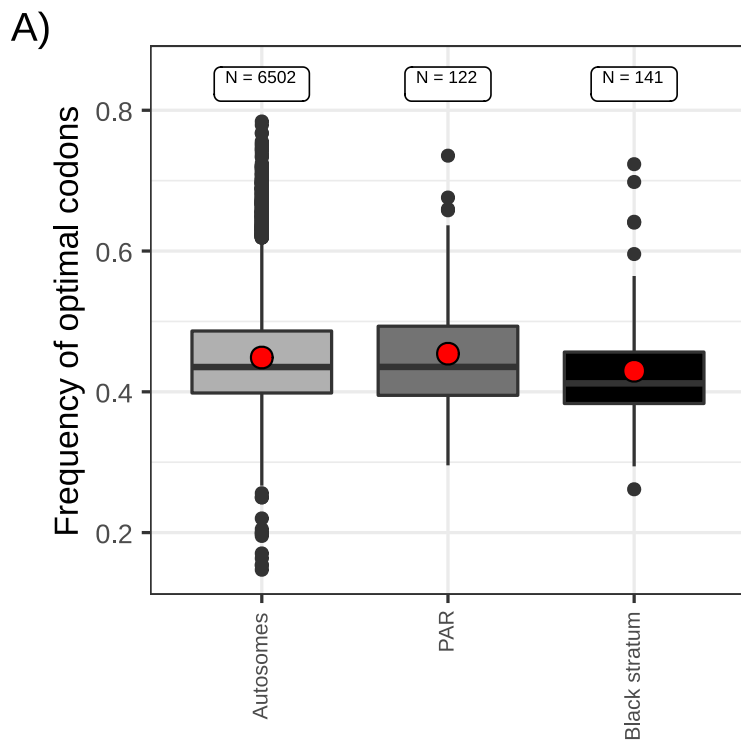


Interaction significance

- Significant
- · · Non-significant

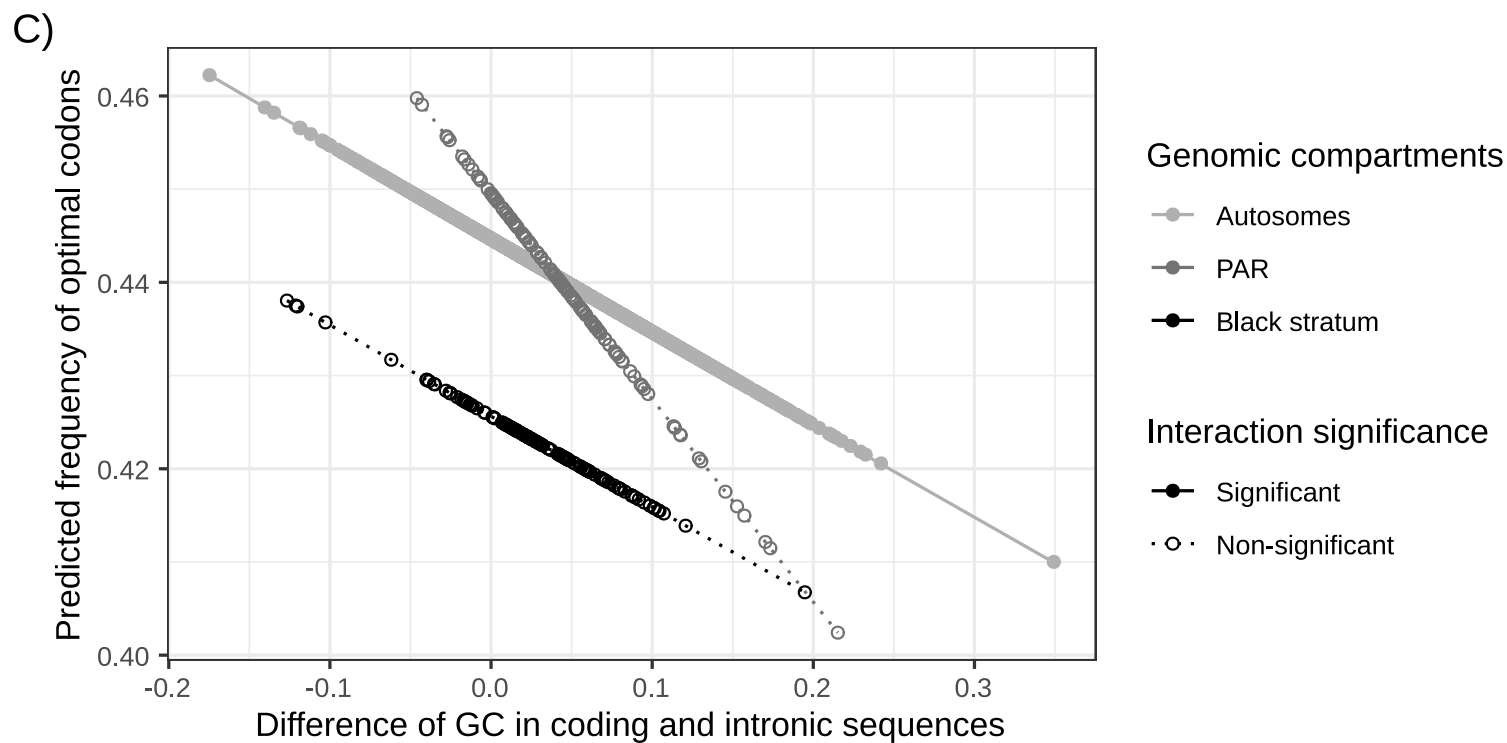
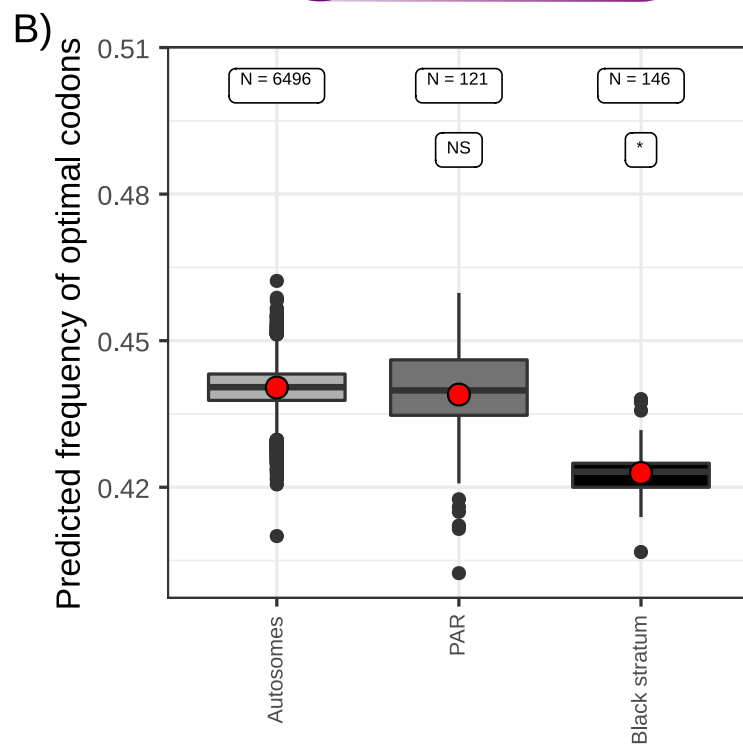
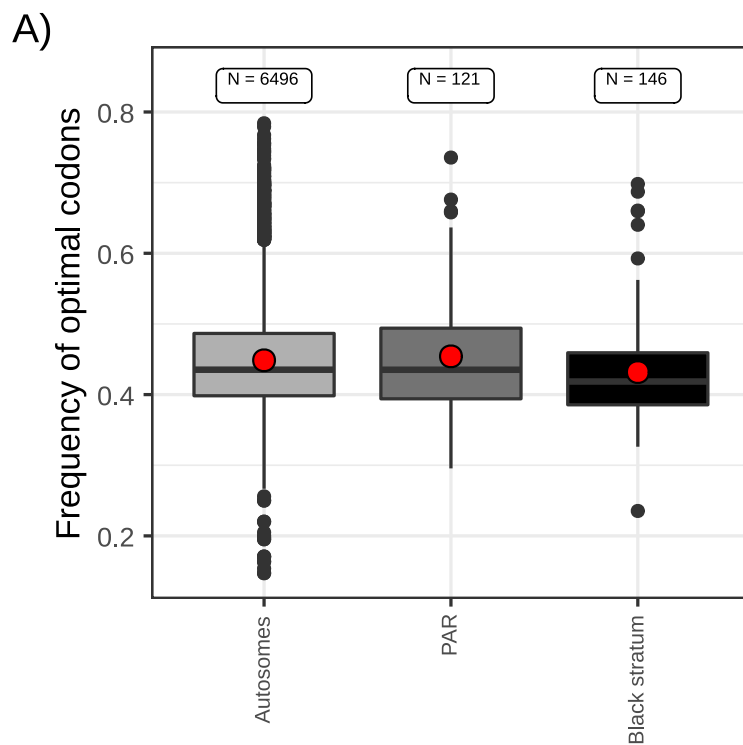
Genomic compartments

- Autosomes
- Recombining region of the PR chr.
- Recombining region of the HD chr.
- Black stratum PR-to-CEN
- Black stratum HD-to-CEN



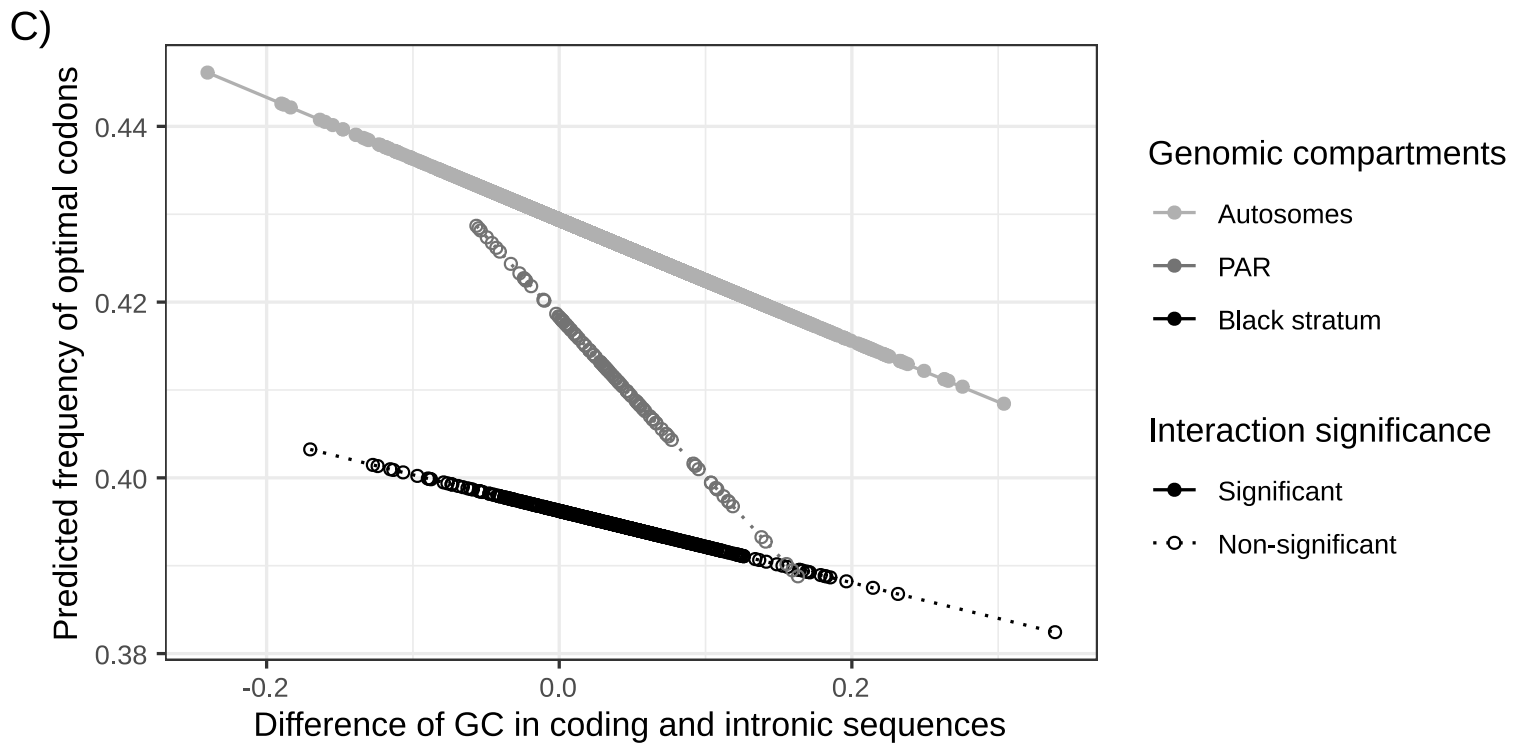
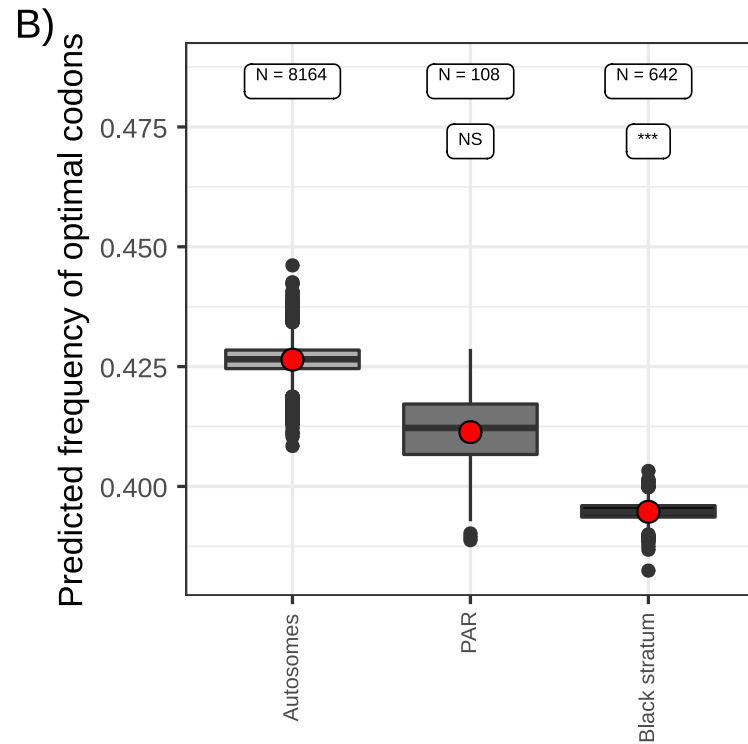
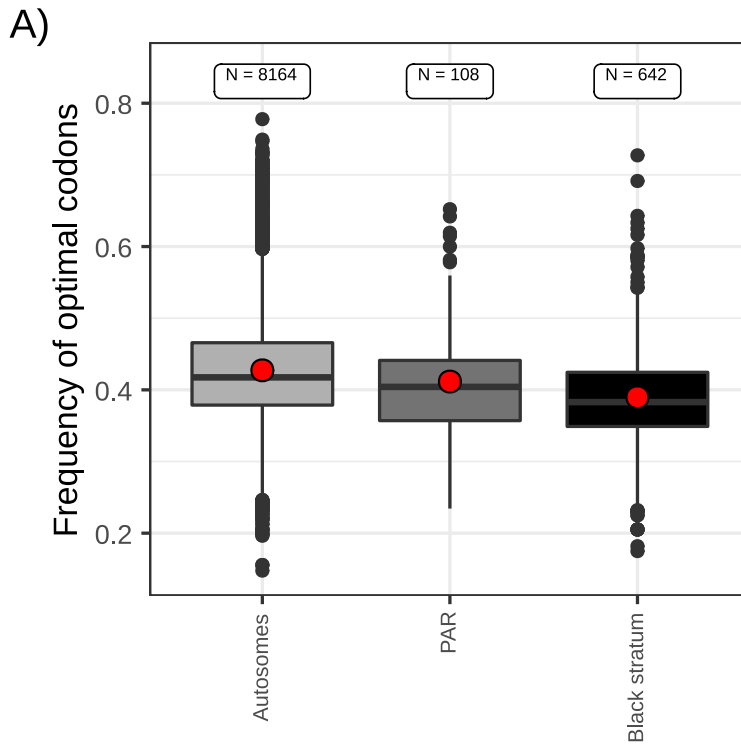


M. scabiosa
on *Knautia arvensis*
a₂ genome



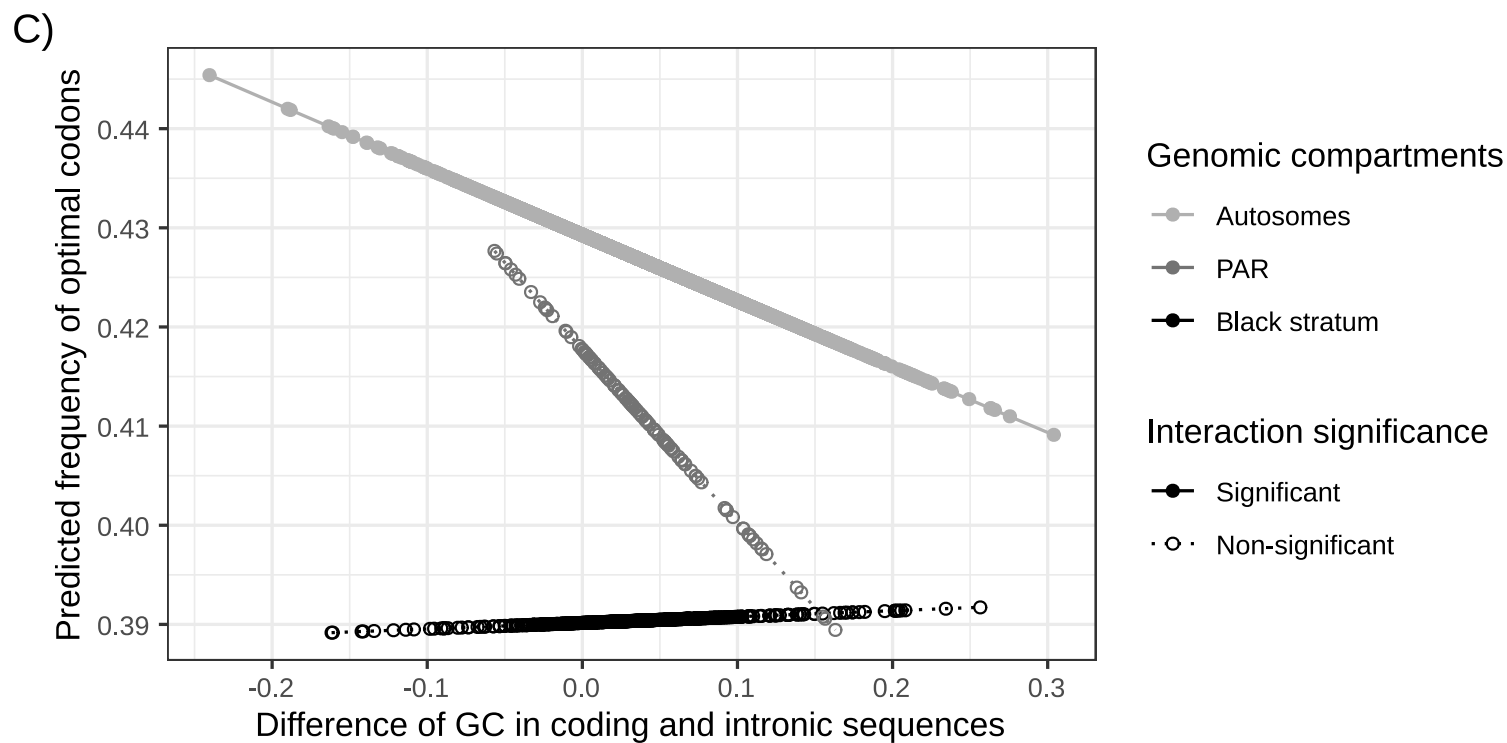
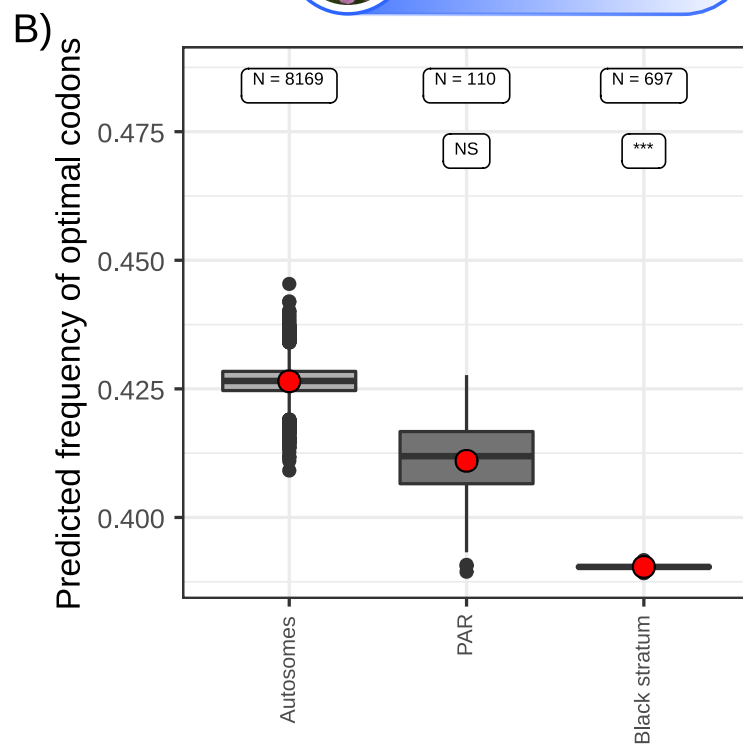
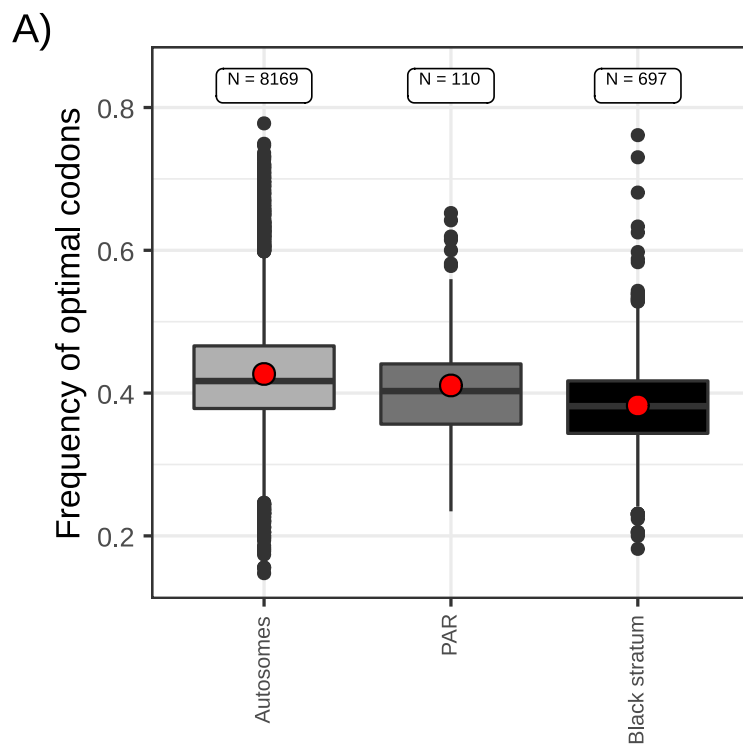


M. silenes-dioicae
on *Silene dioica*
a₁ genome





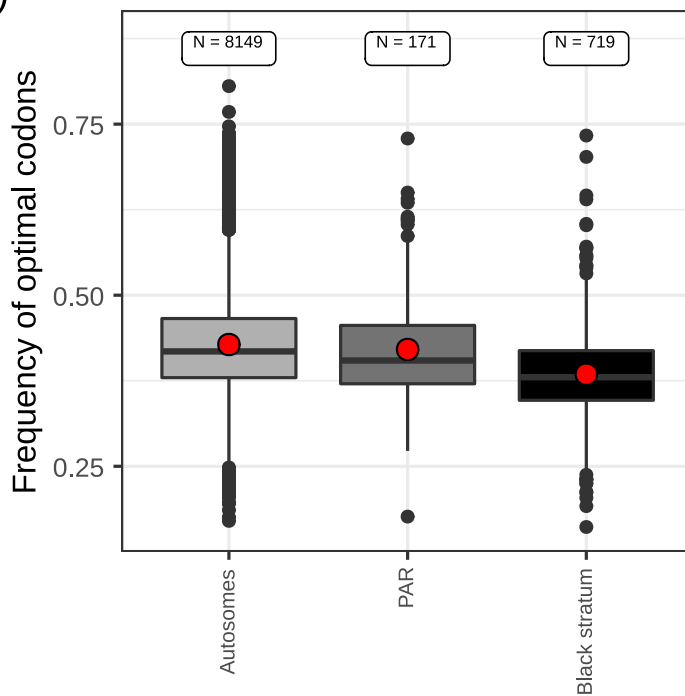
M. silenes-dioicae
on *Silene dioica*
a₂ genome



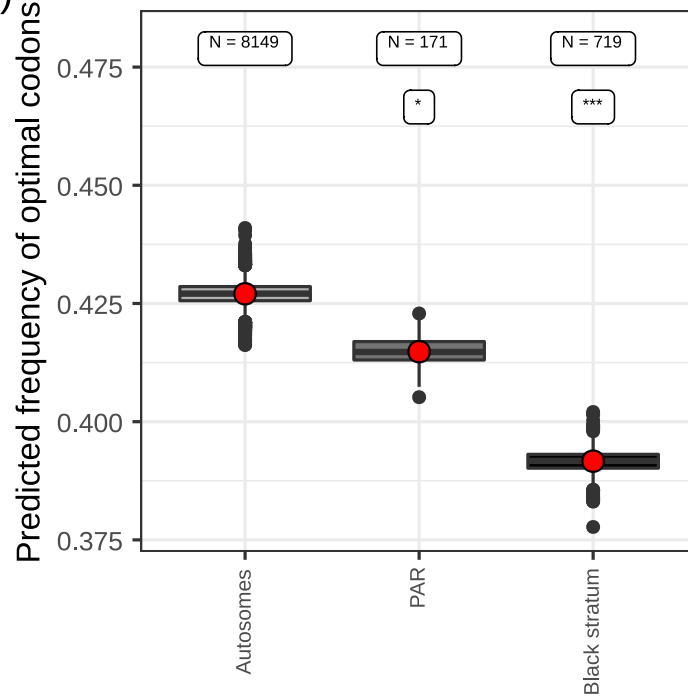


M. coronariae
on *Silene flos-cuculi*
 a_1 genome

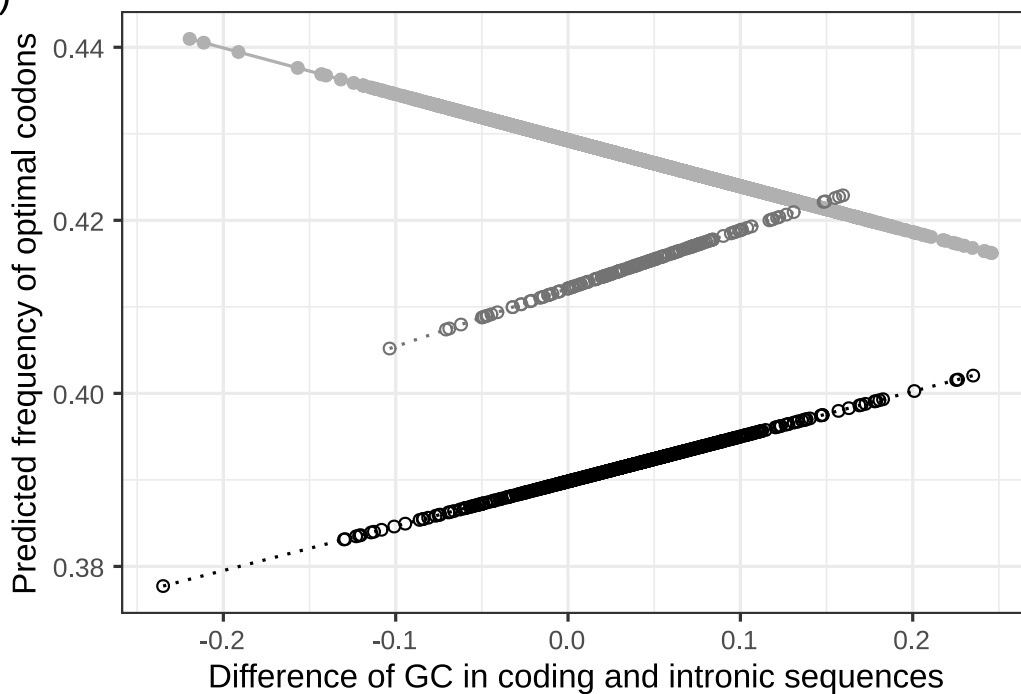
A)



B)



C)



Genomic compartments

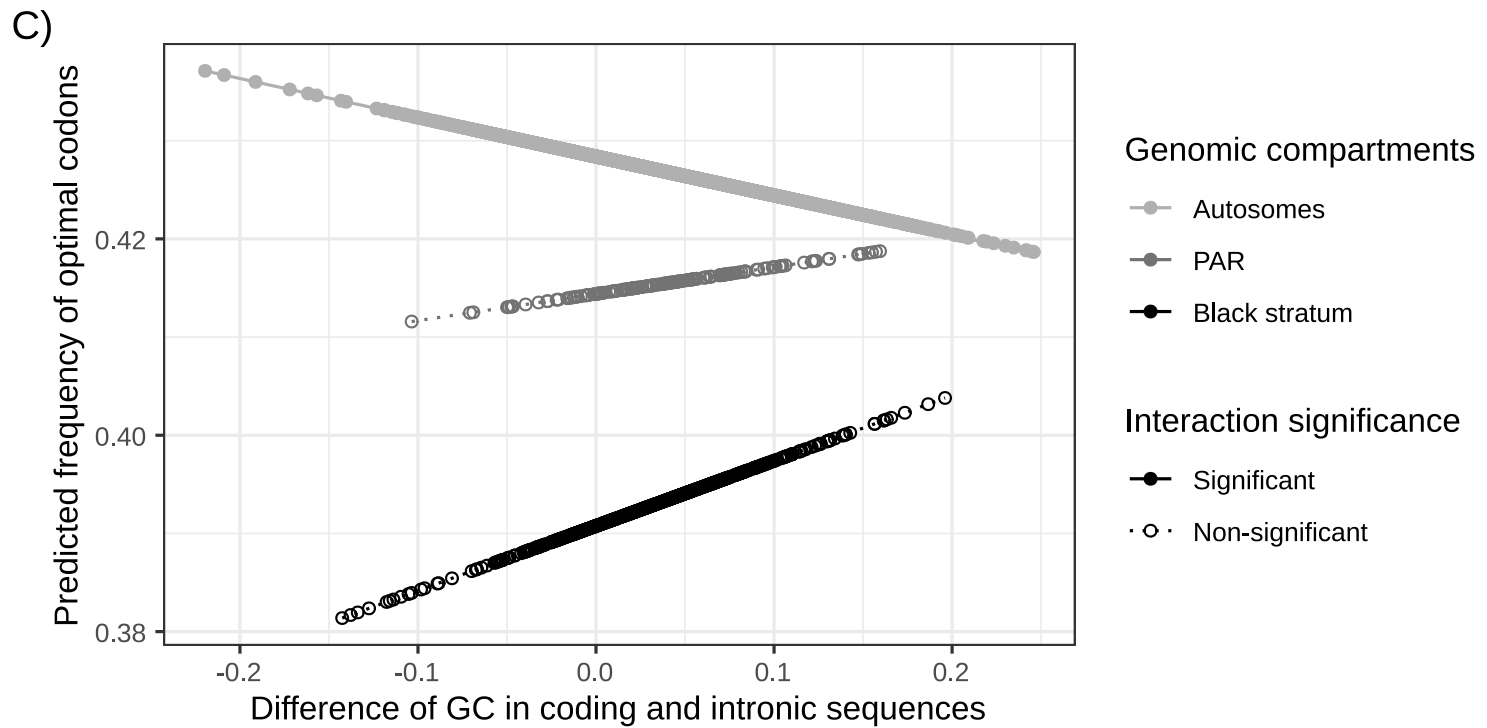
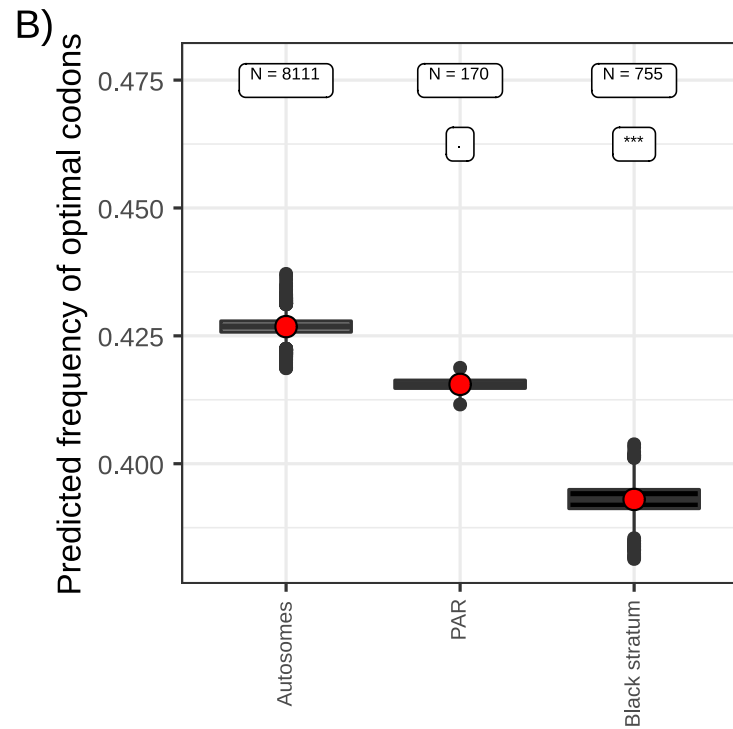
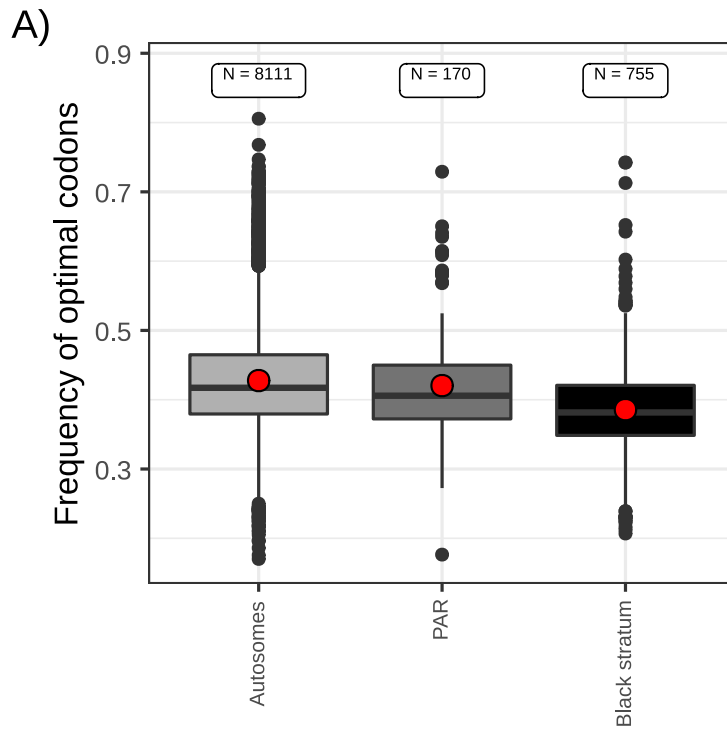
- Autosomes
- PAR
- Black stratum

Interaction significance

- Significant
- Non-significant



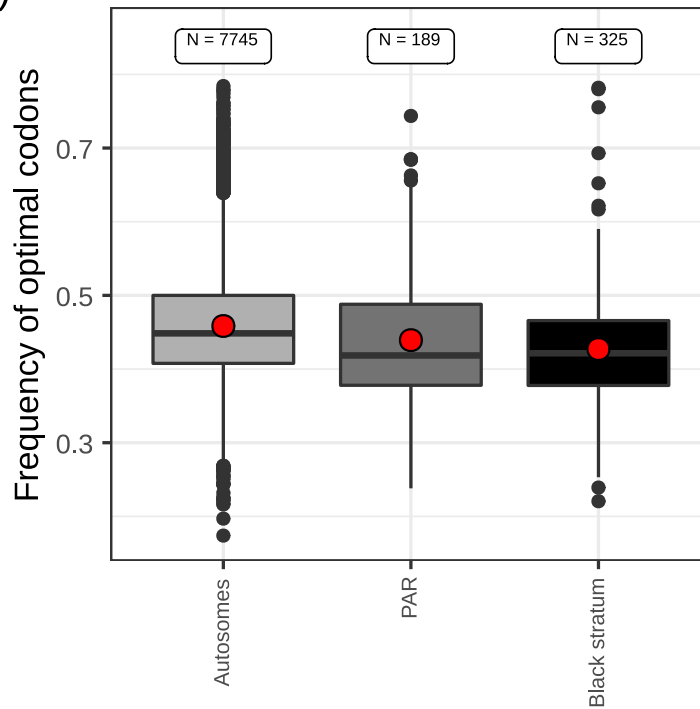
M. coronariae
on *Silene flos-cuculi*
a₂ genome



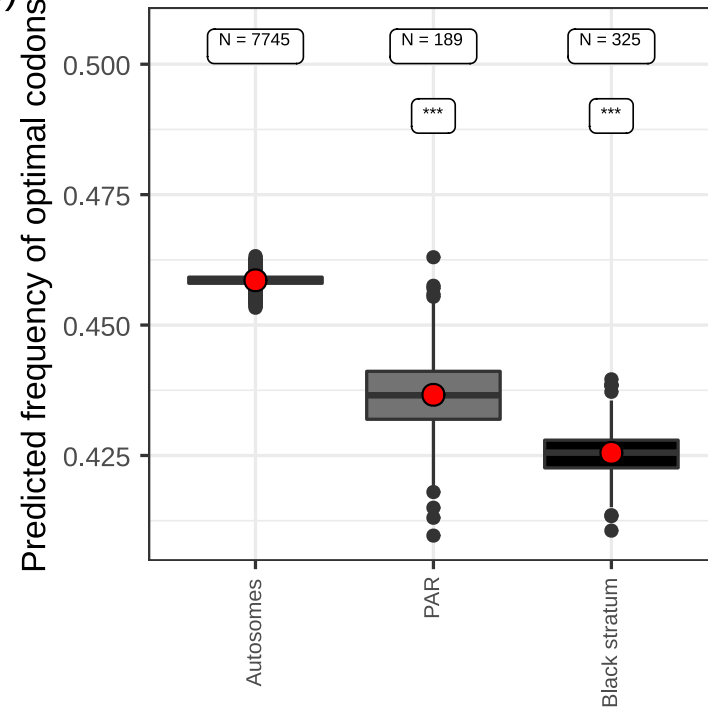


M. violaceum s.l.
on *Silene melanantha*
a₁ genome

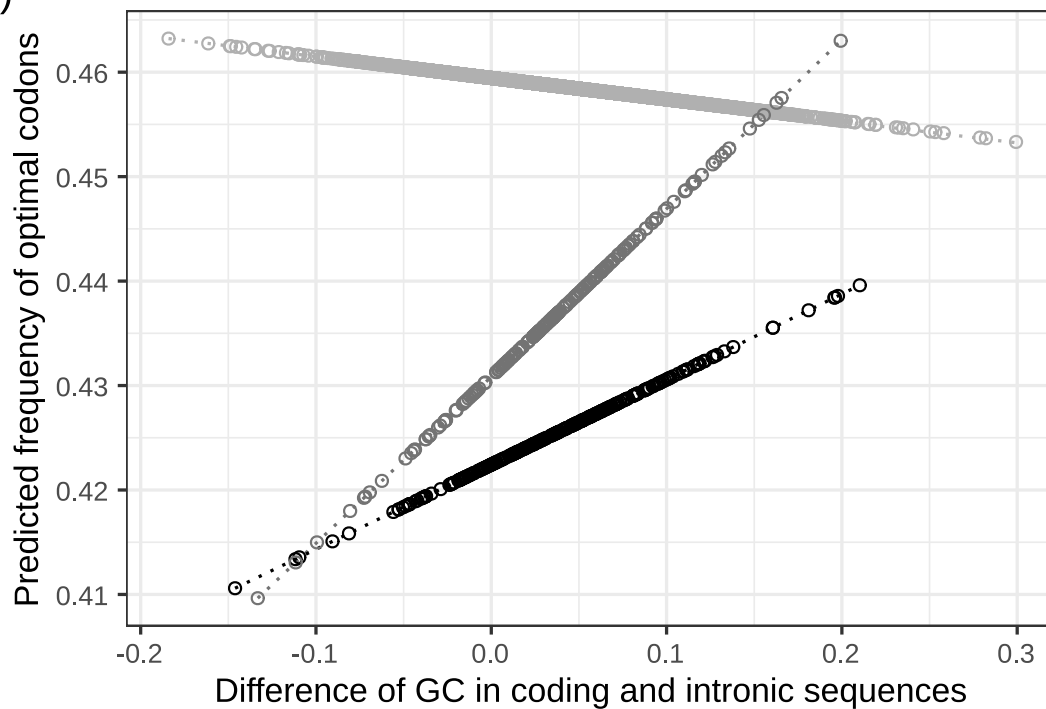
A)



B)



C)



Interaction significance

· ○ · Non-significant

Genomic compartments

—○— Autosomes

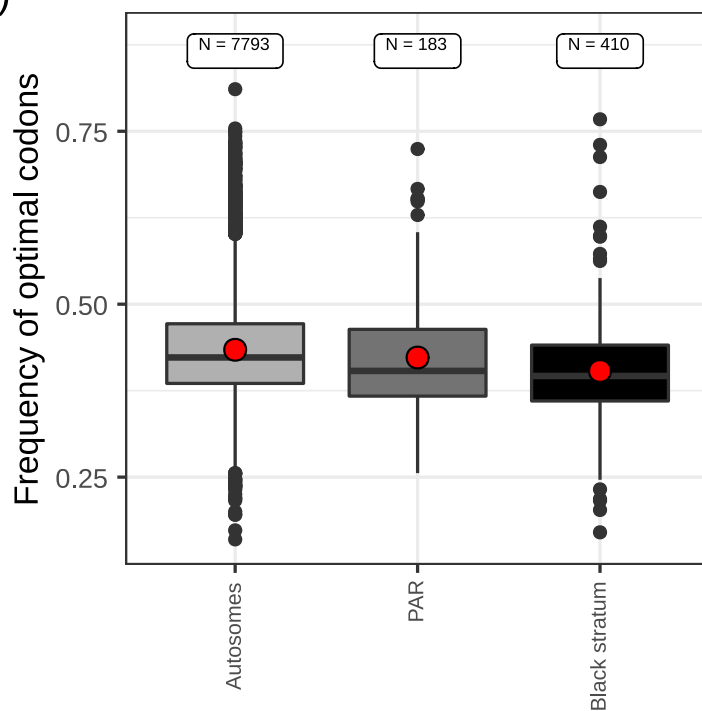
—●— PAR

—●— Black stratum

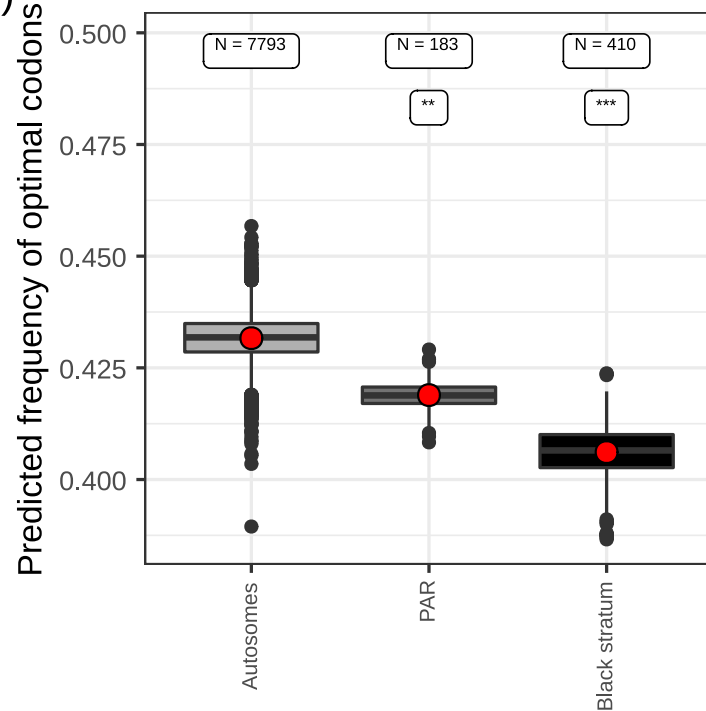


M. violaceum s.l.
on *Silene melanantha*
a₂ genome

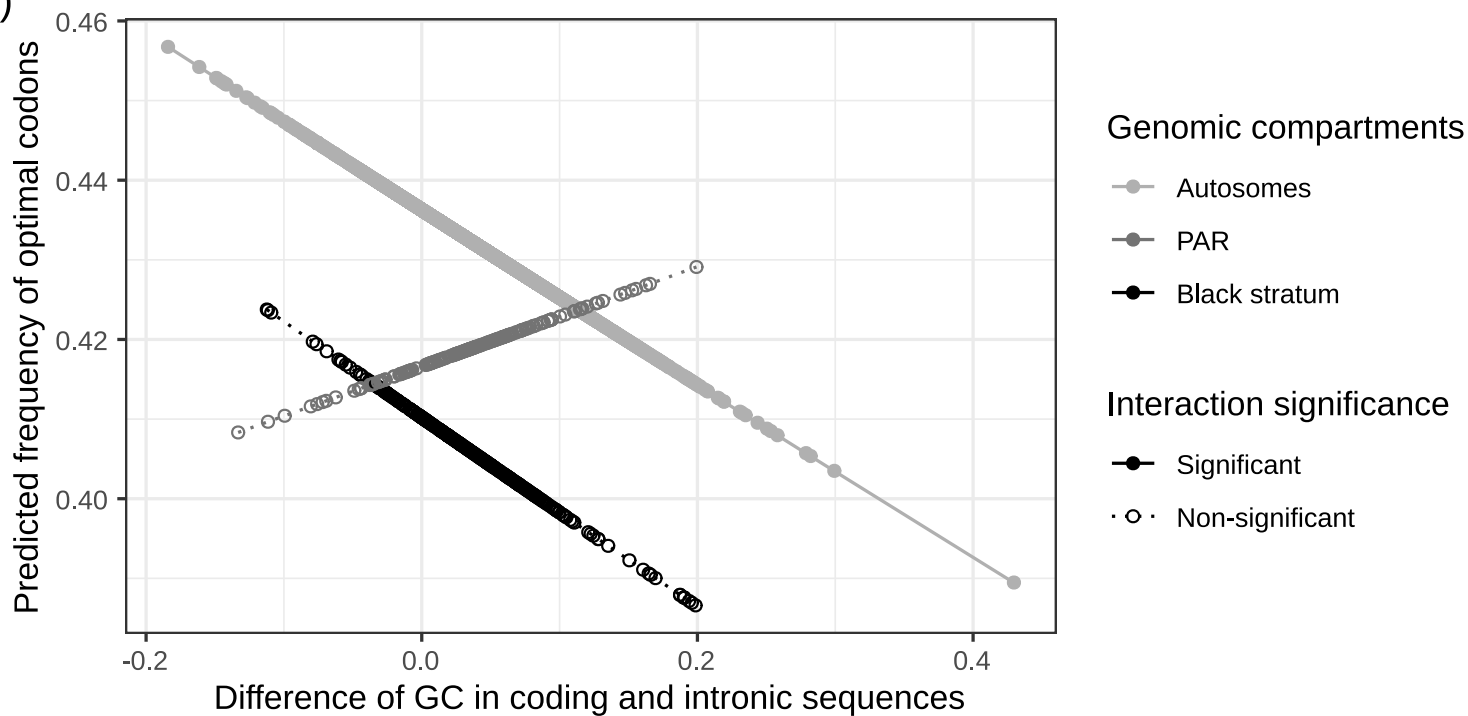
A)



B)



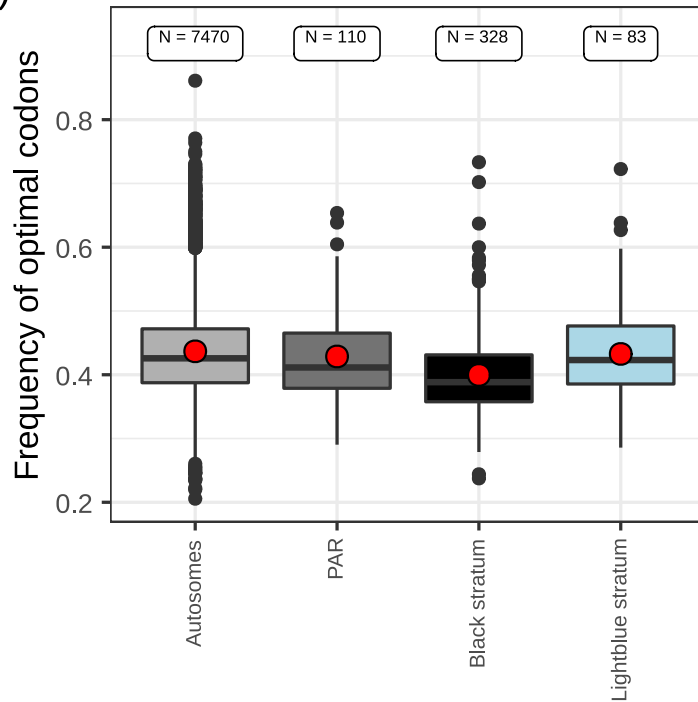
C)



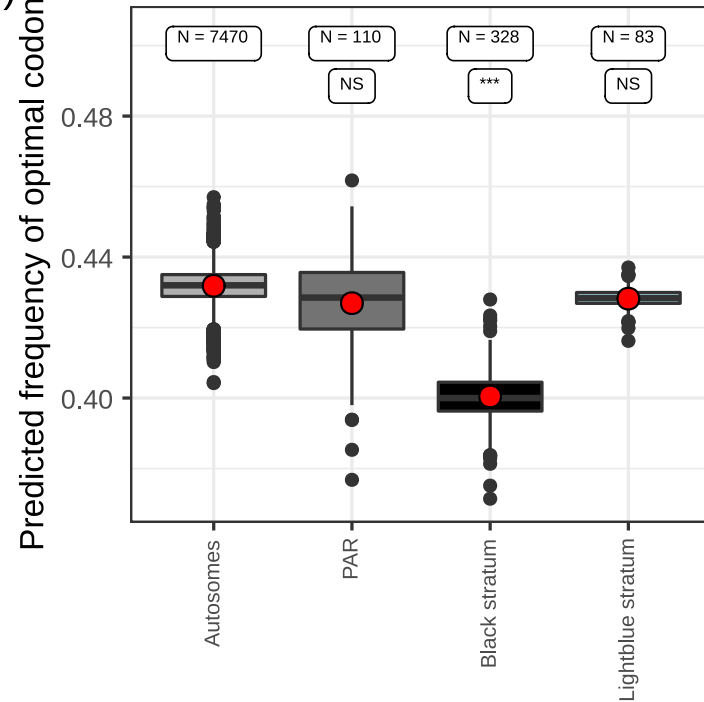


M. violaceum s.l.
on *Silene caroliniana*
a₁ genome

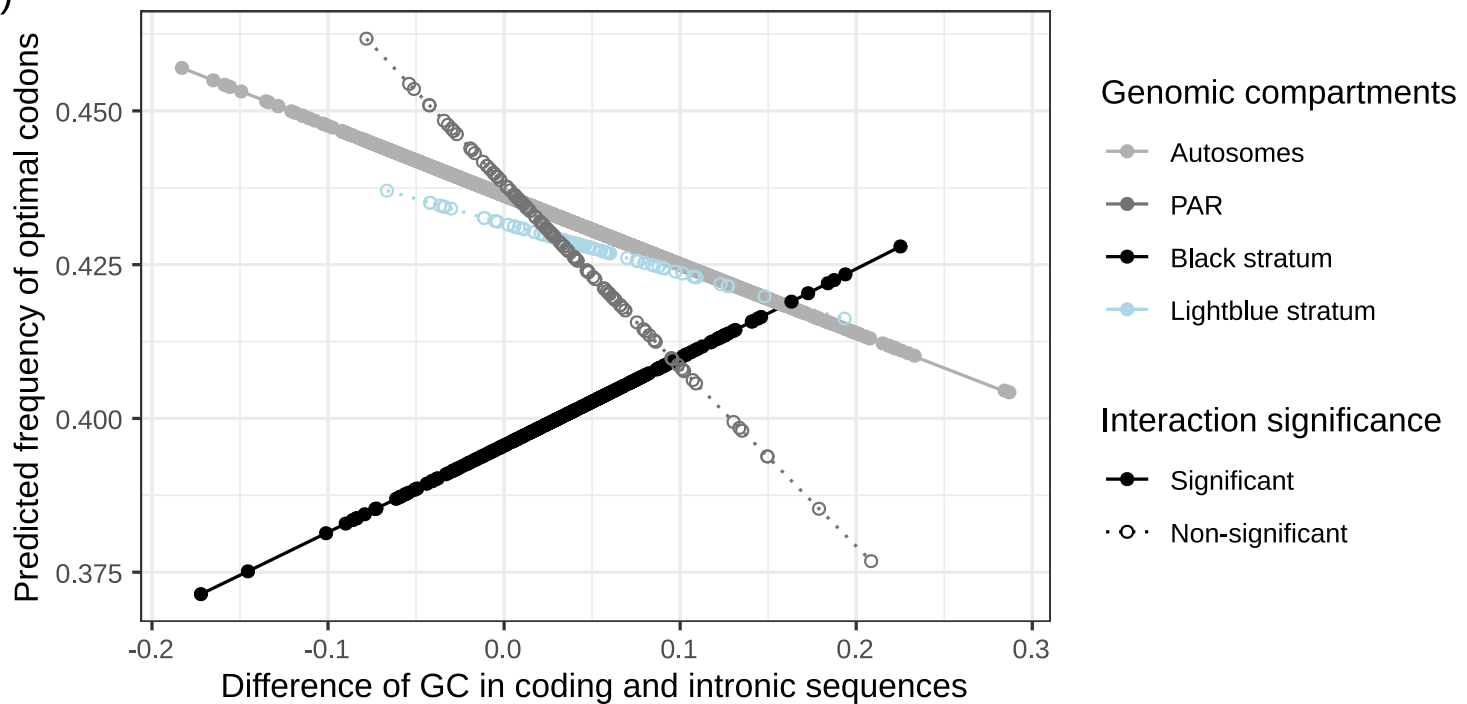
A)



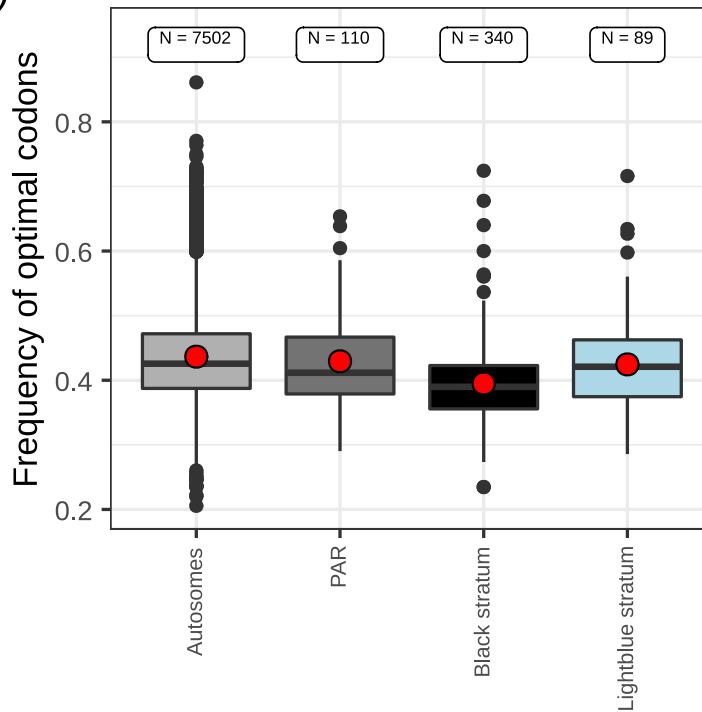
B)



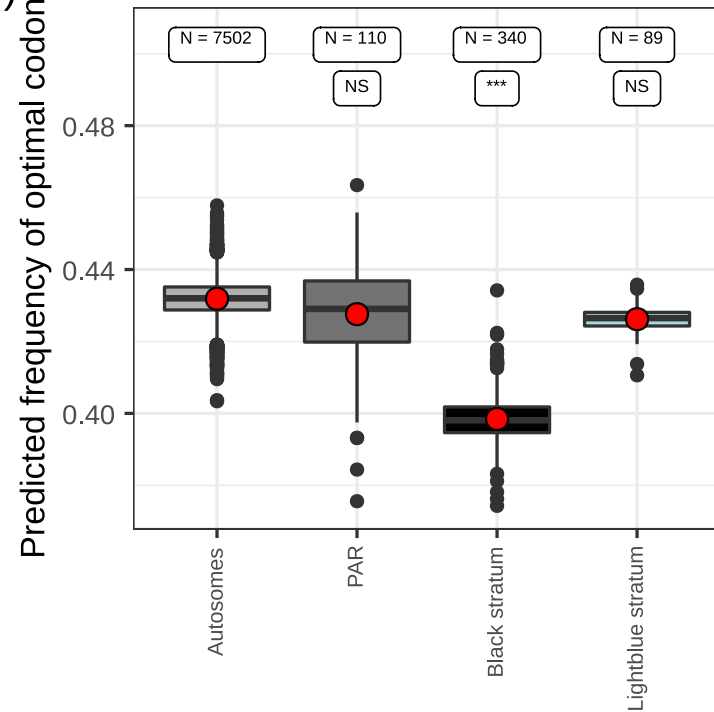
C)



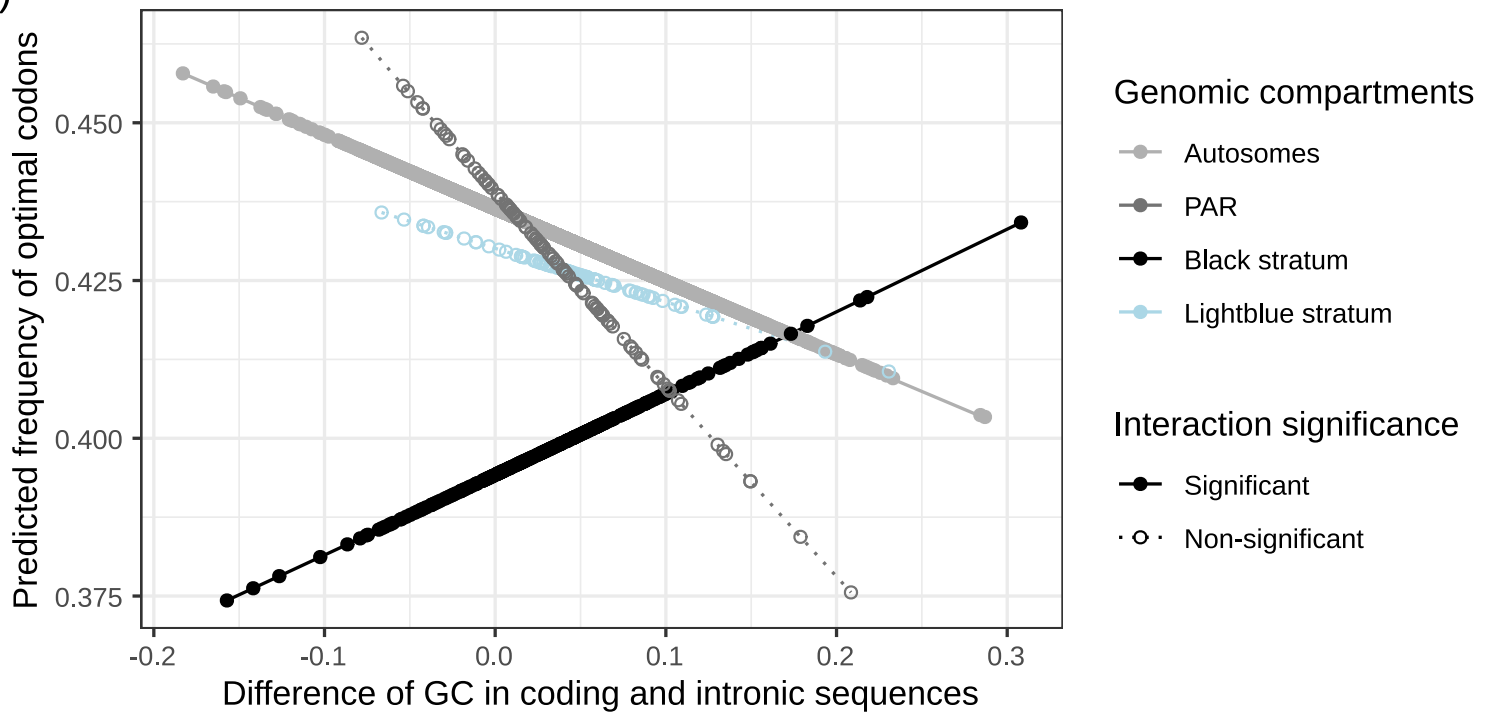
A)



B)

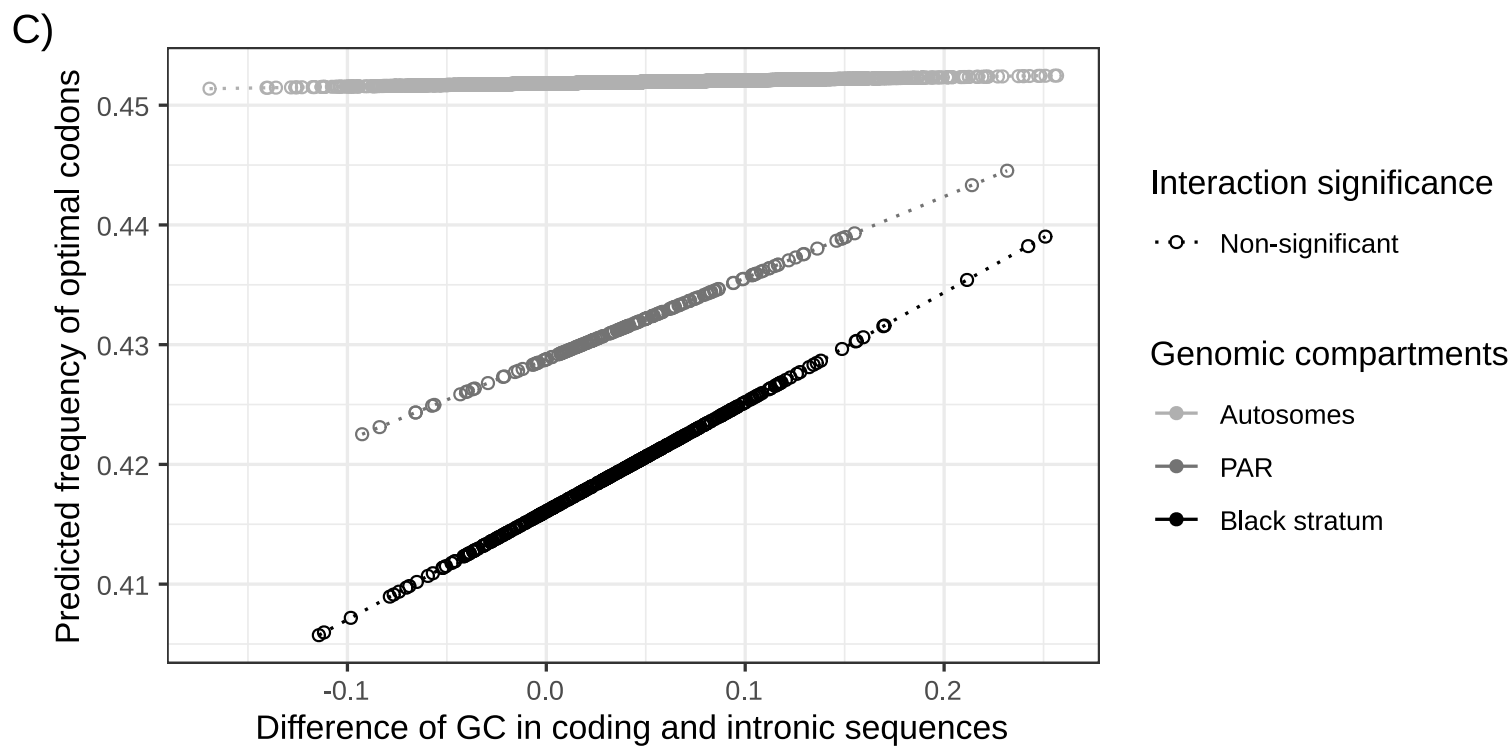
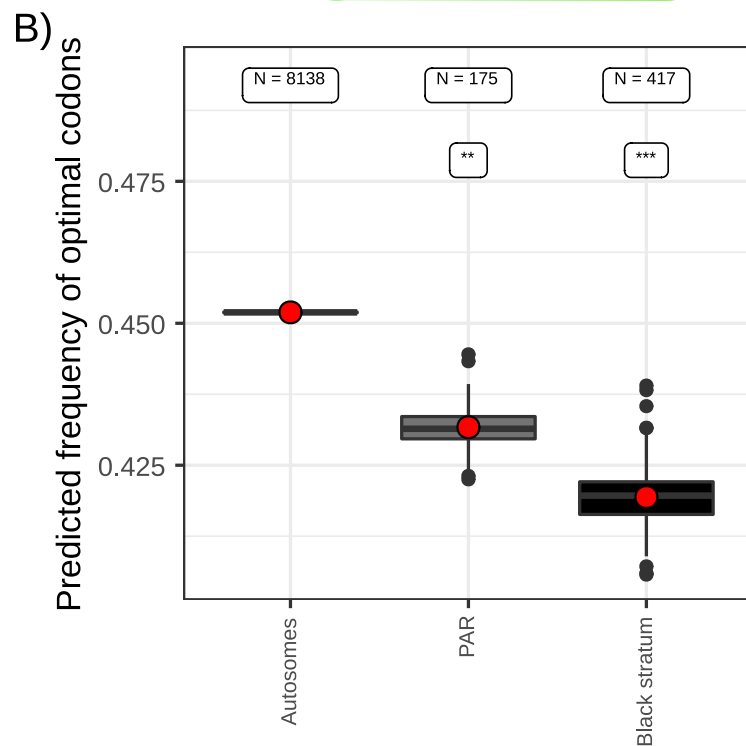
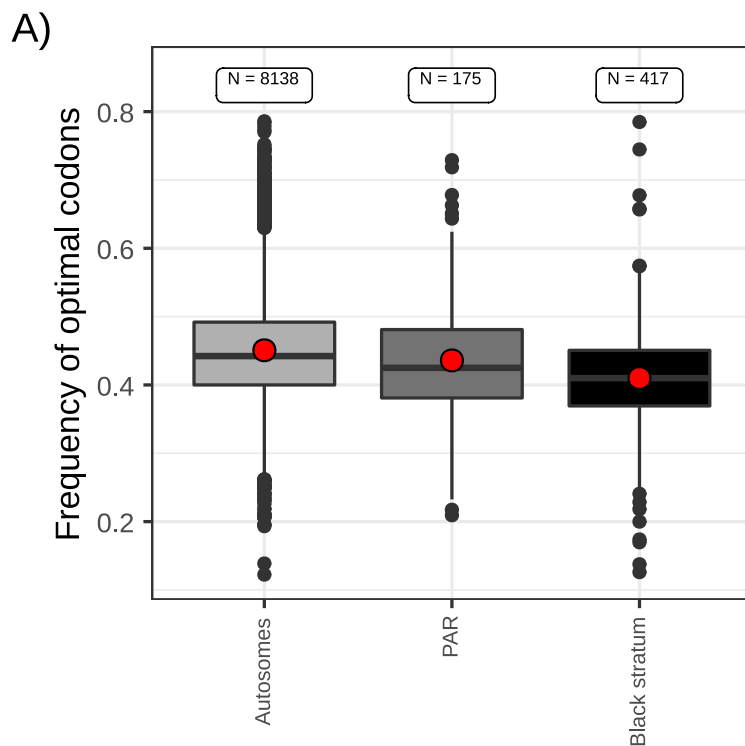


C)

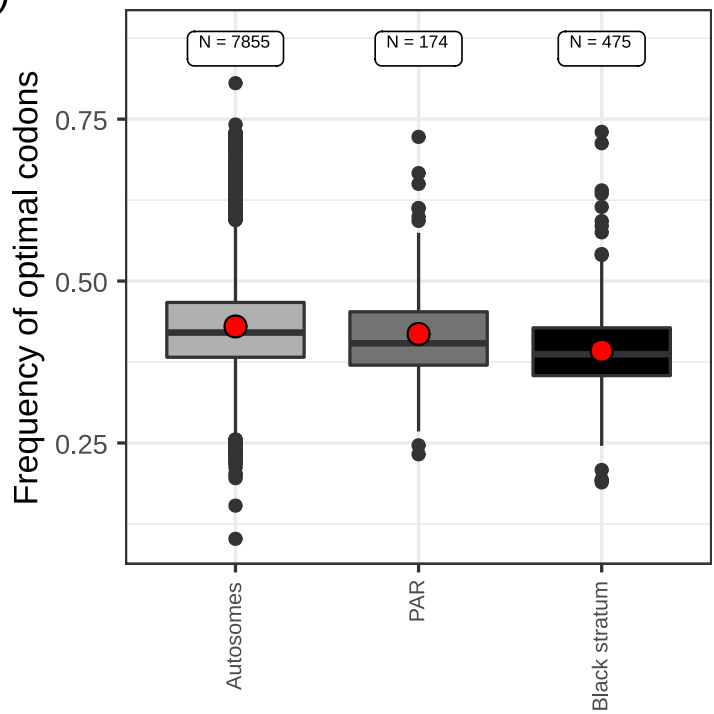




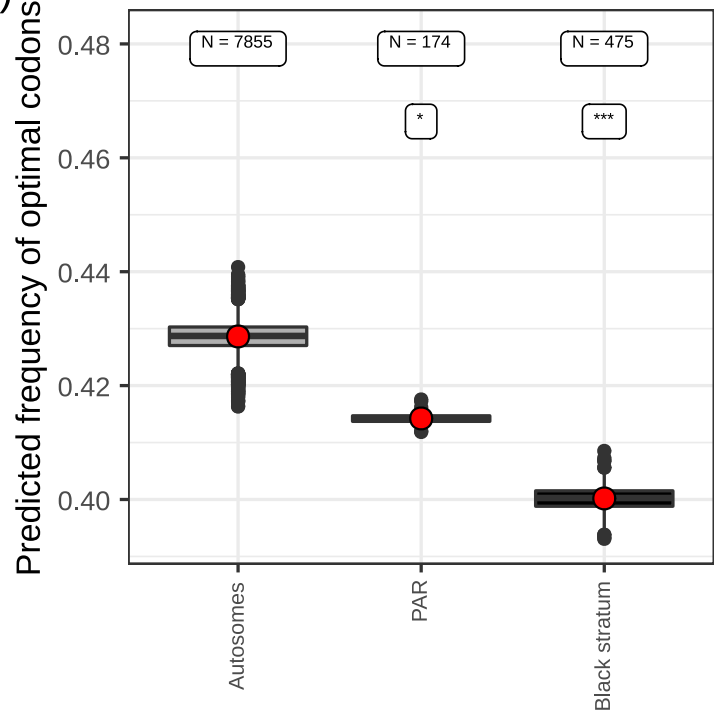
M. violaceum s.s.
on *Silene nutans*
a₁ genome



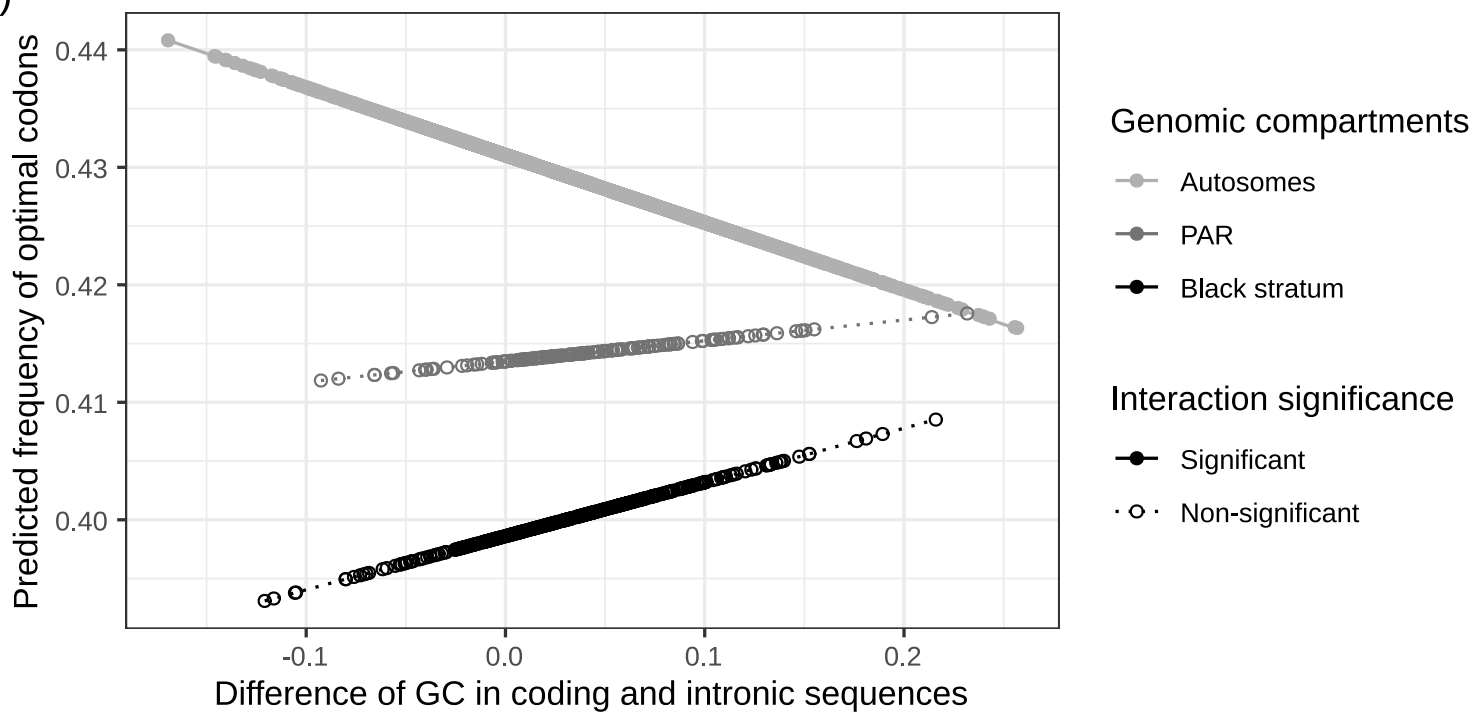
A)



B)



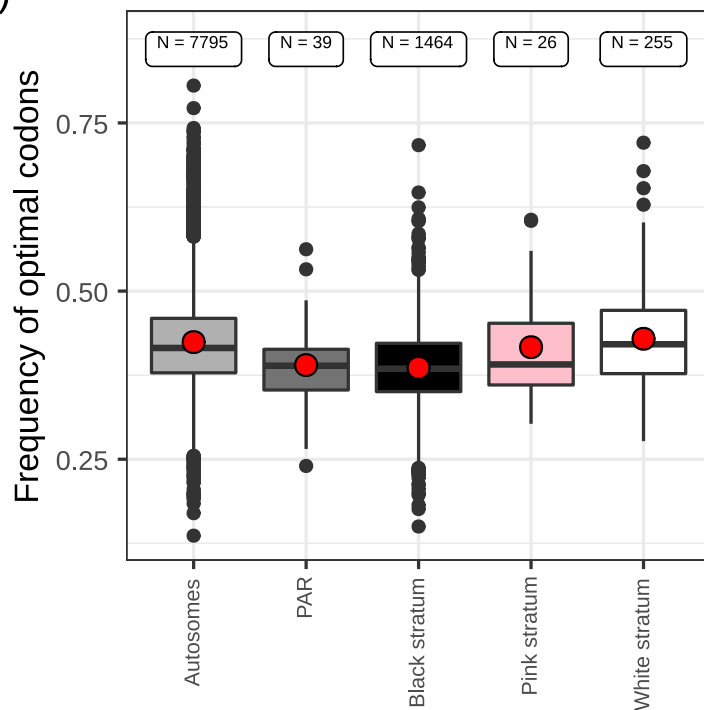
C)



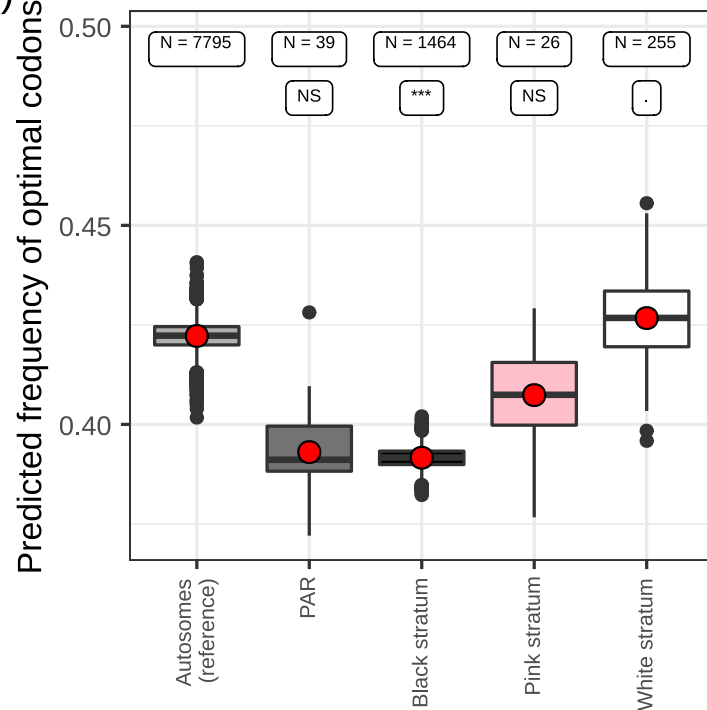


M. violaceum s.l.
on *Silene paradoxa*
a₁ genome

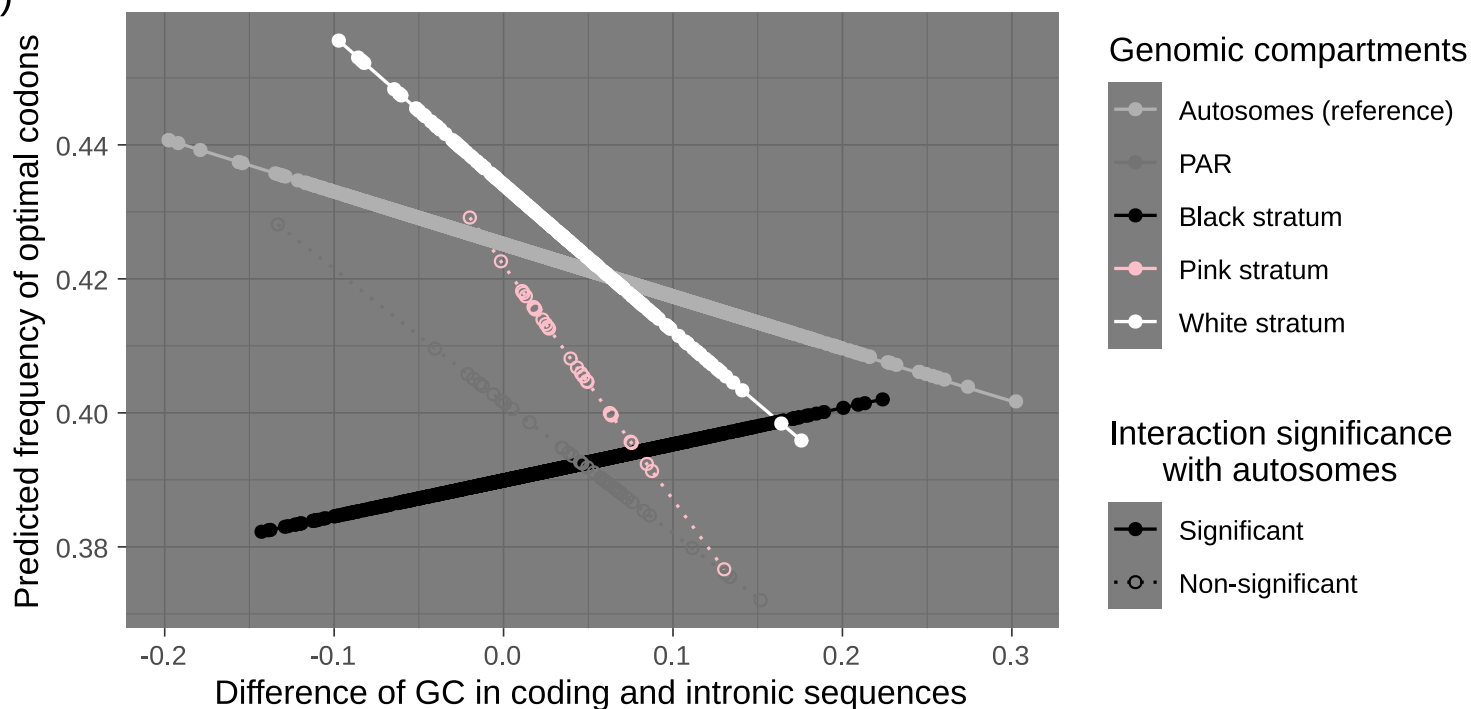
A)

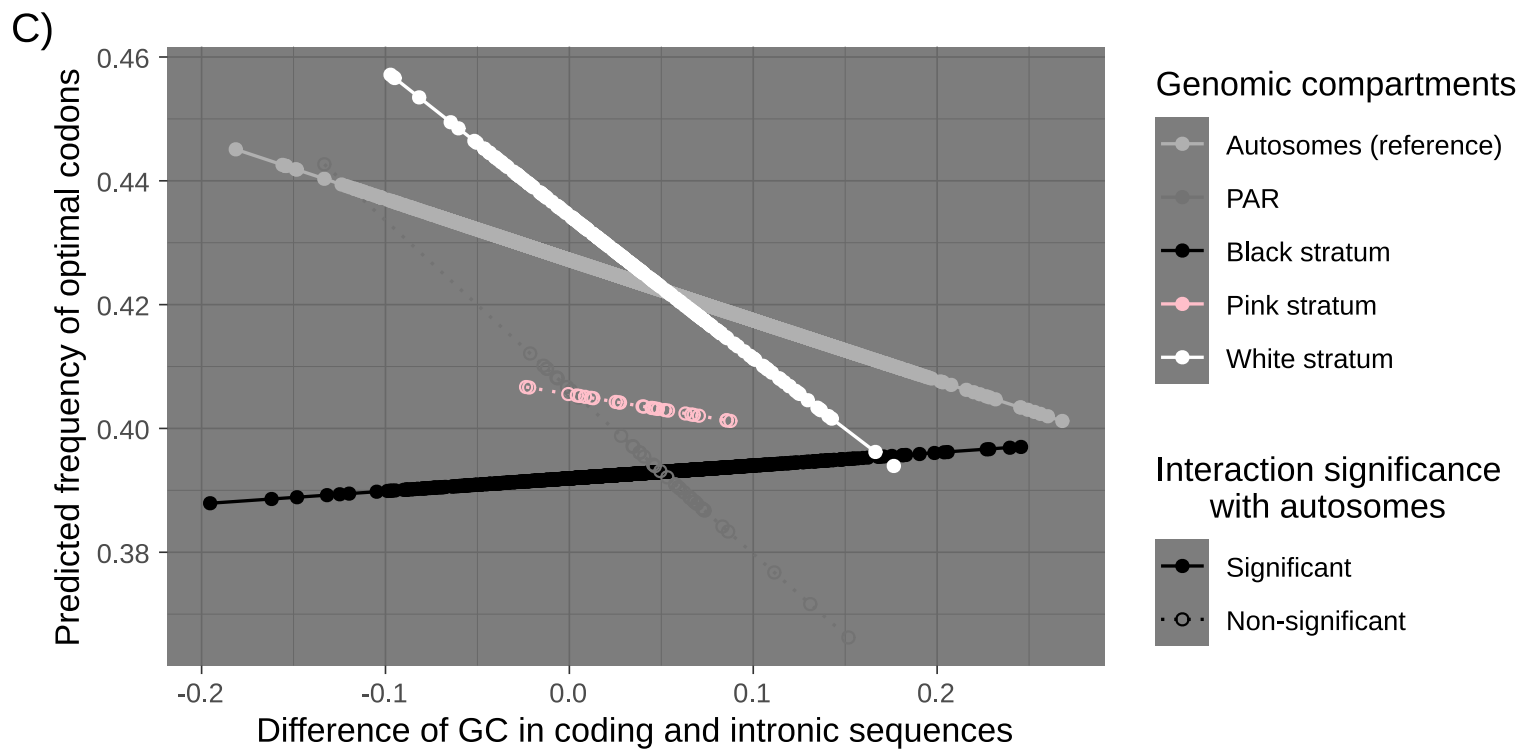
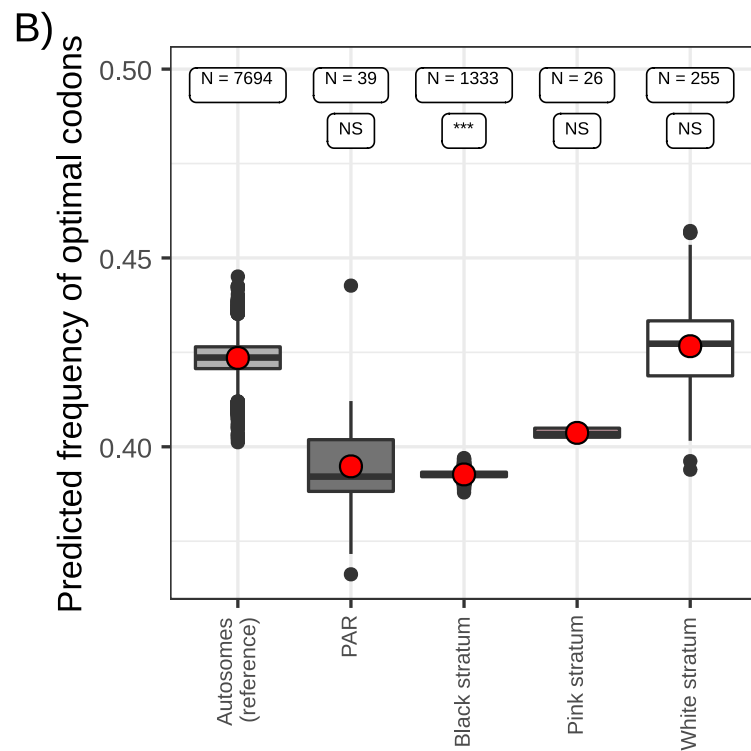
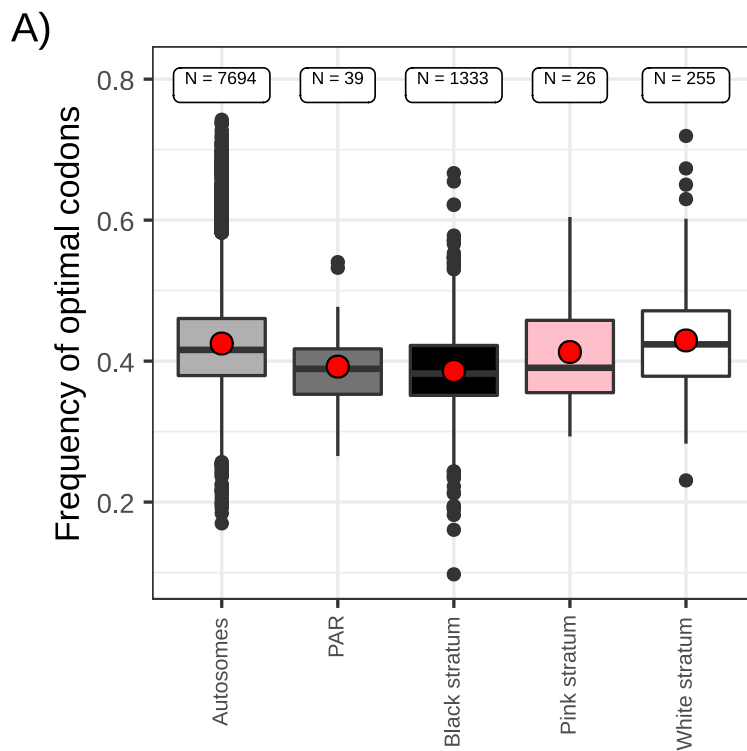


B)



C)

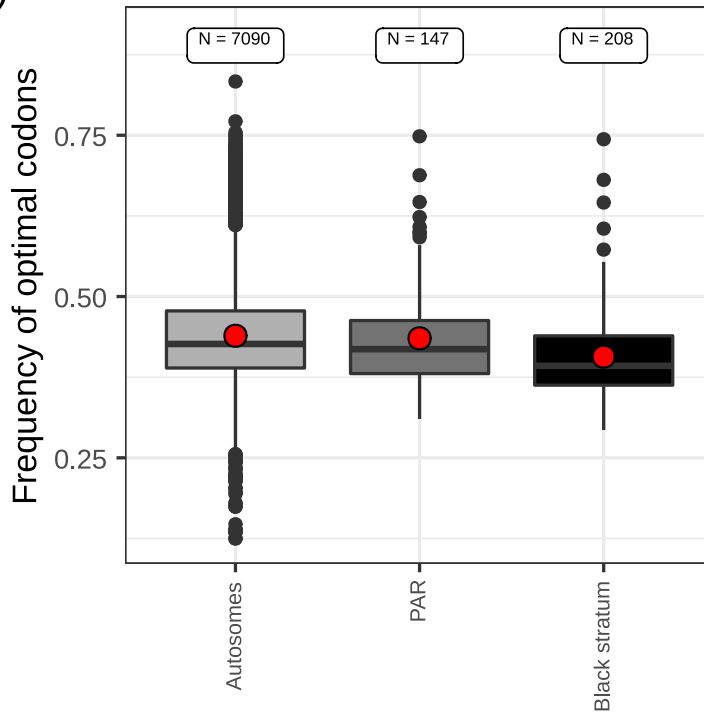




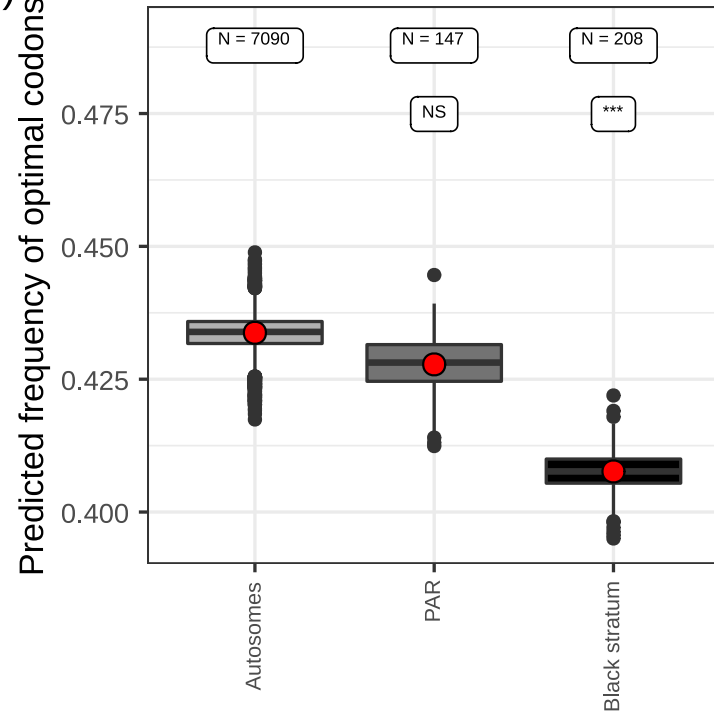


M. violaceum s.l.
on *Silene tatarinowii*
a₂ genome

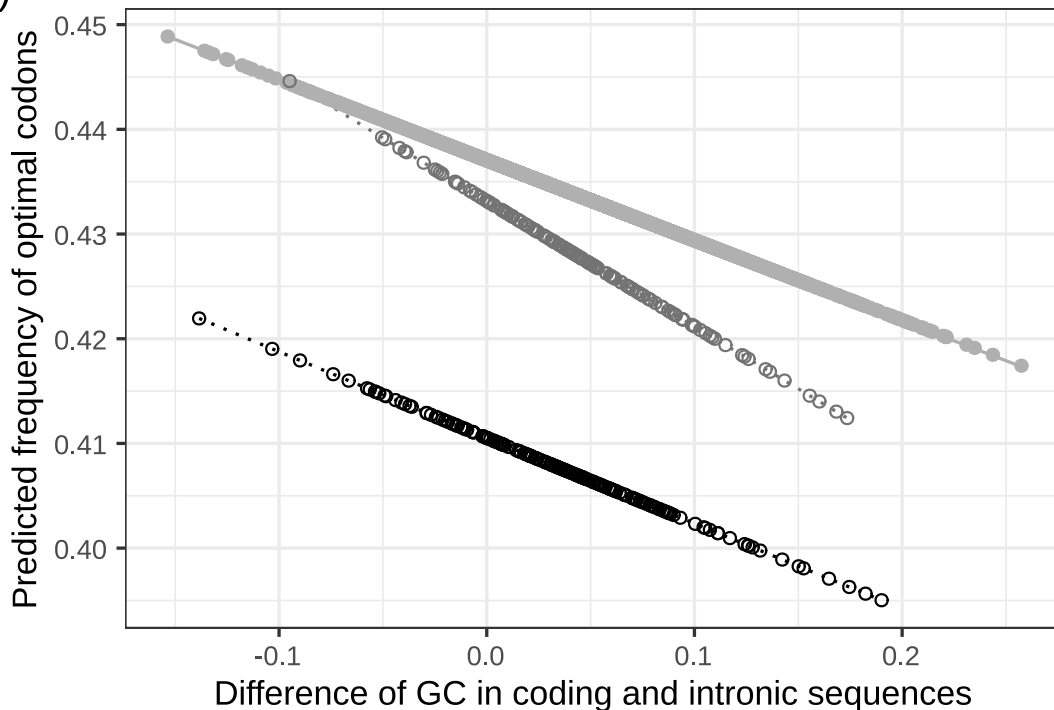
A)



B)



C)



Genomic compartments

- Autosomes
- PAR
- Black stratum

Interaction significance

- Significant
- Non-significant

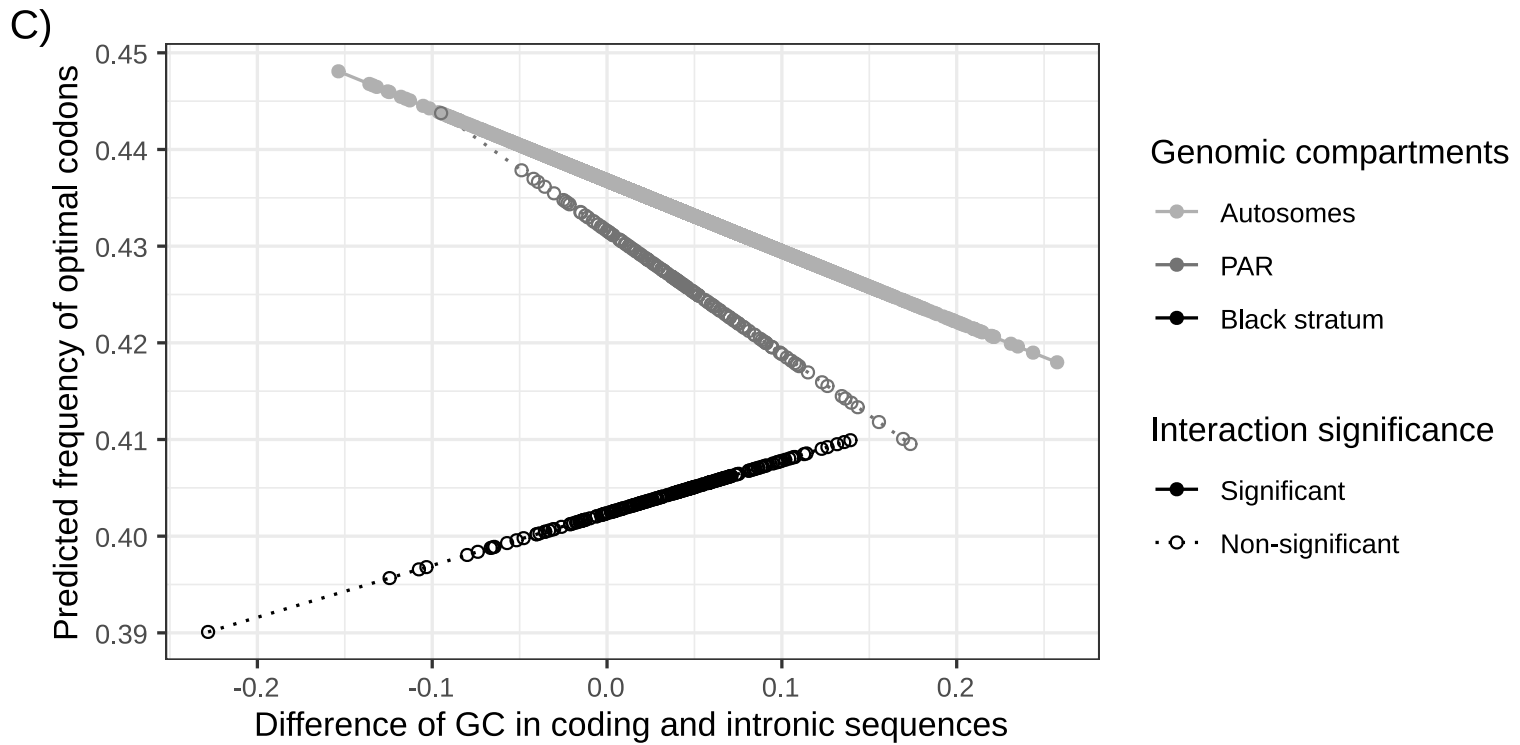
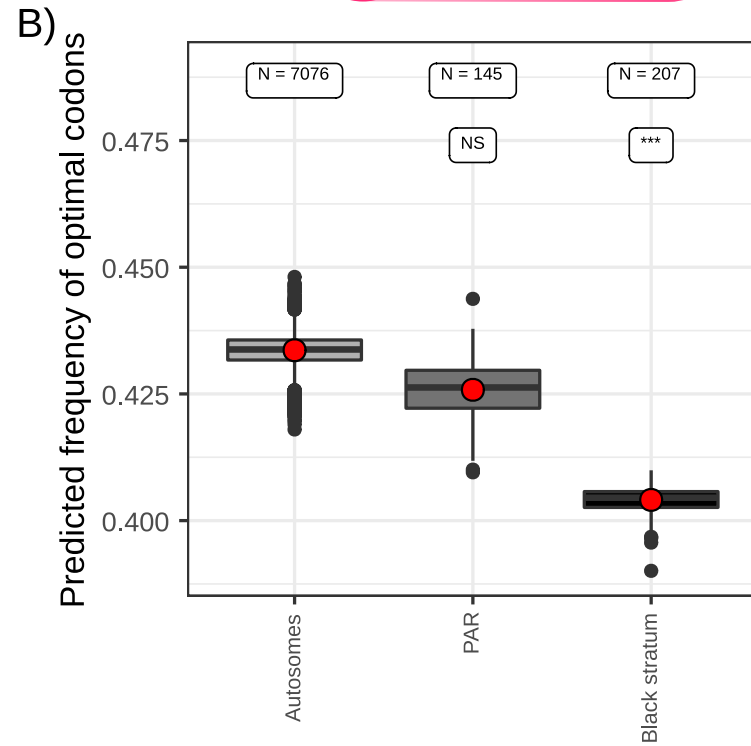
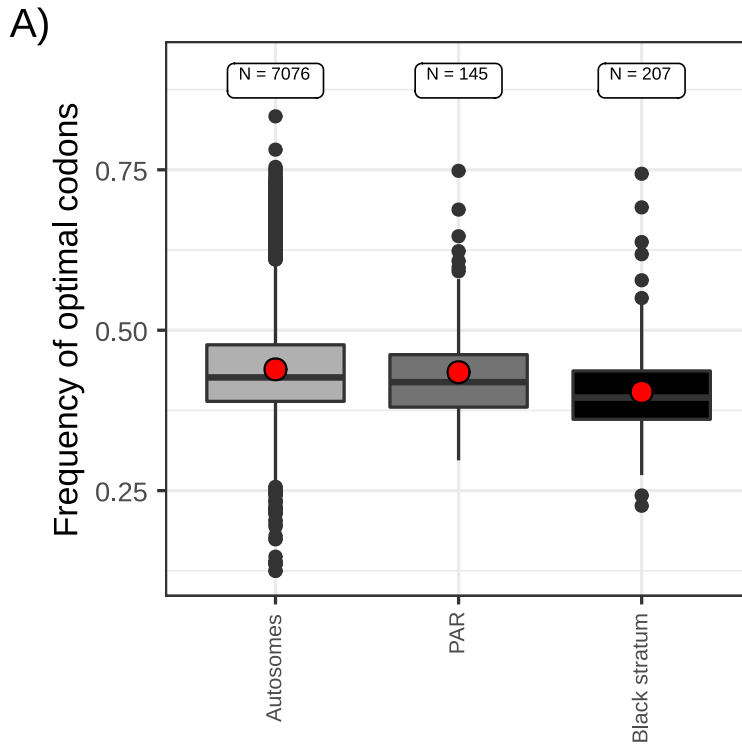
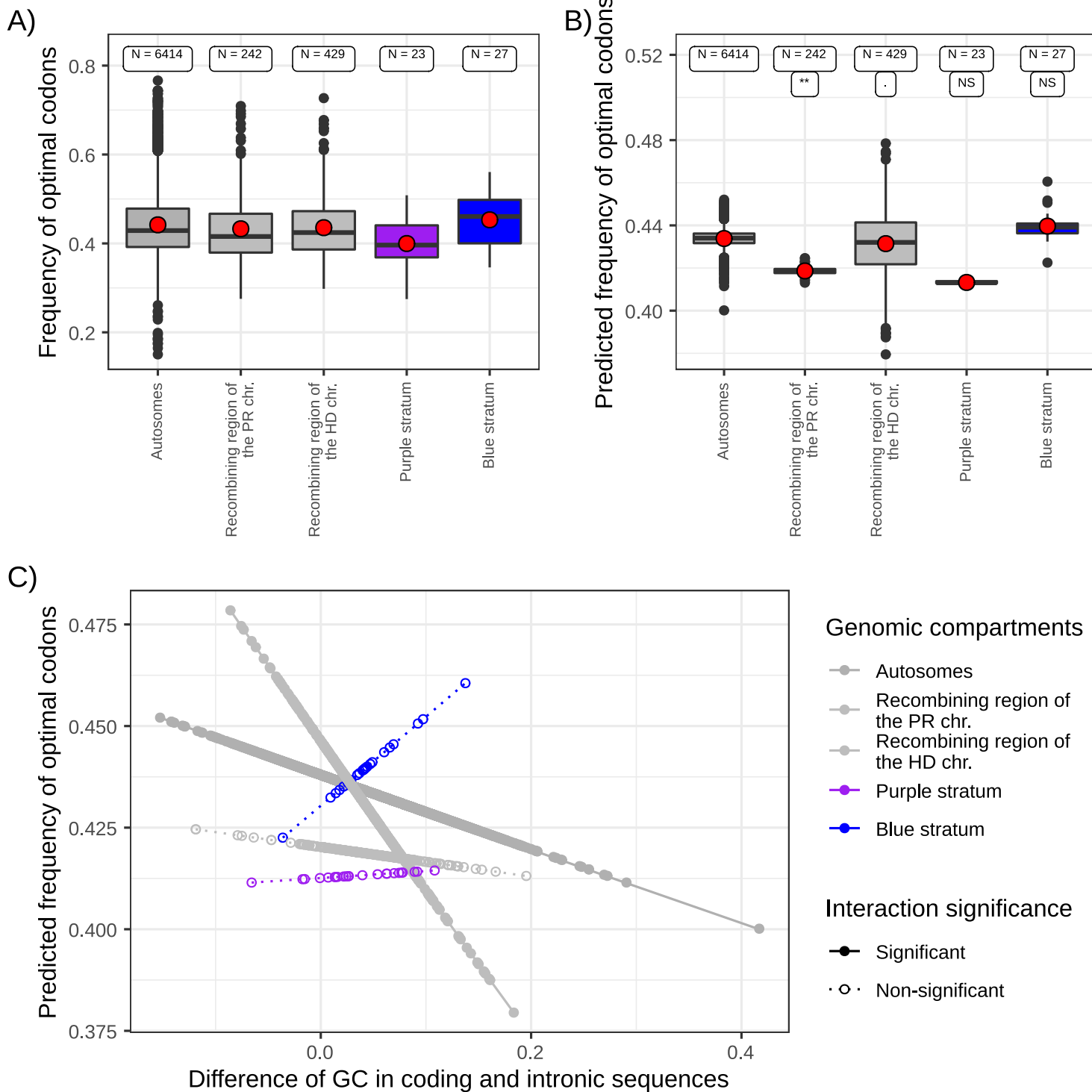
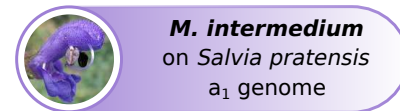


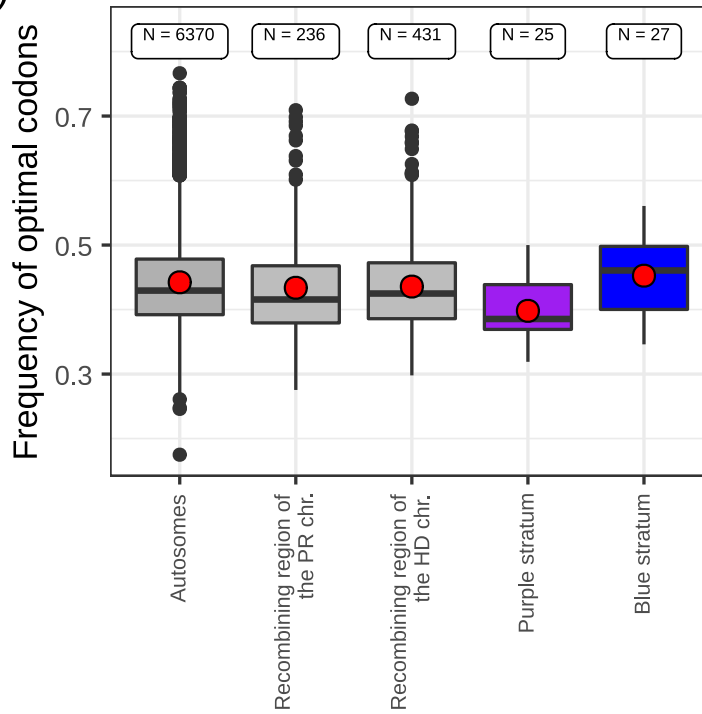
Figure S8



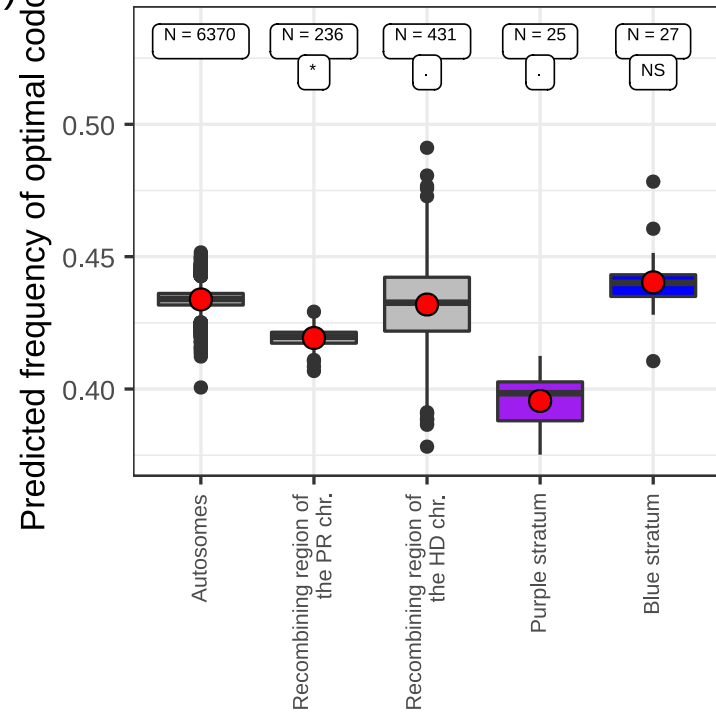


M. intermedium
on *Salvia pratensis*
a₂ genome

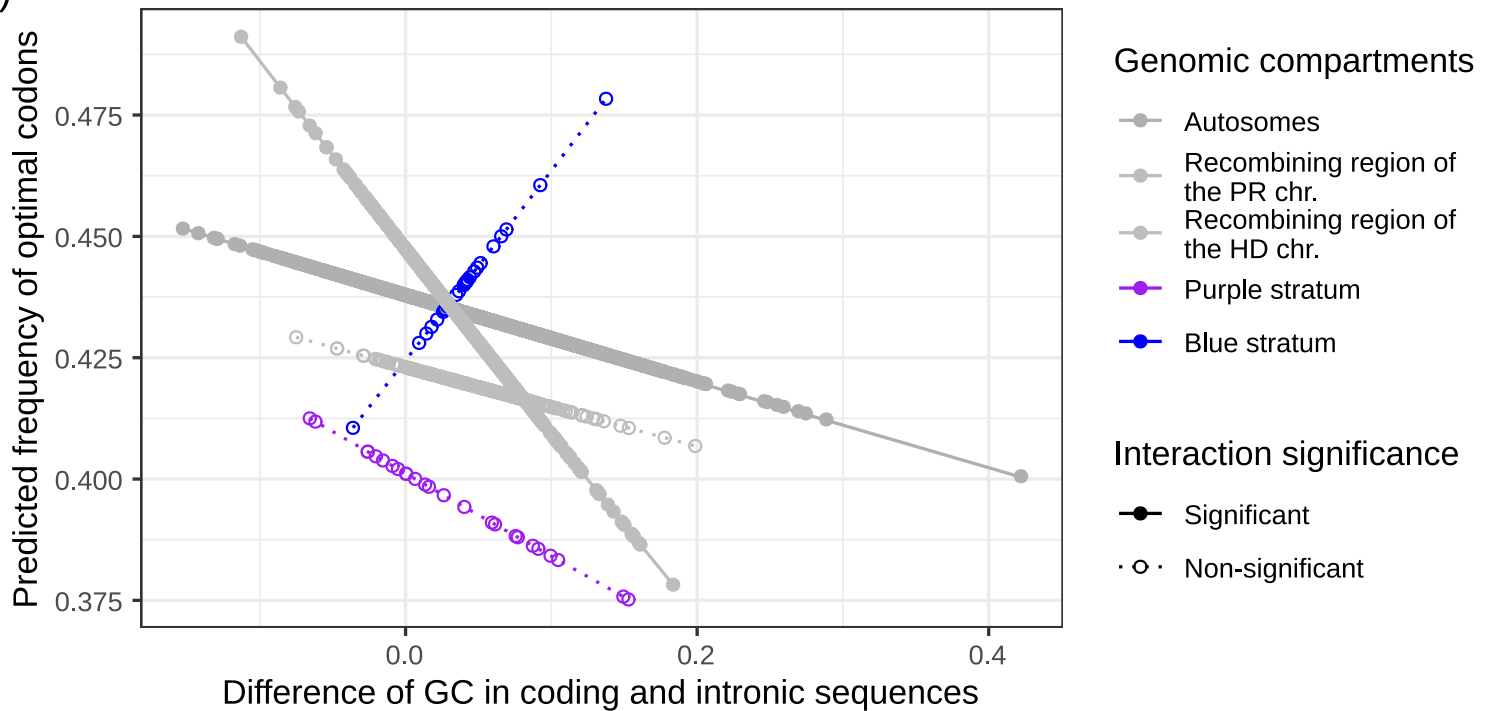
A)



B)



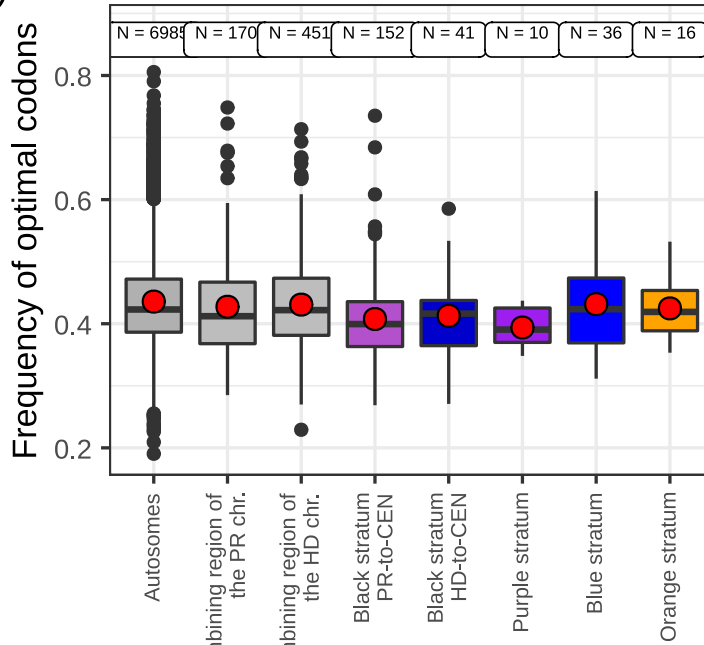
C)



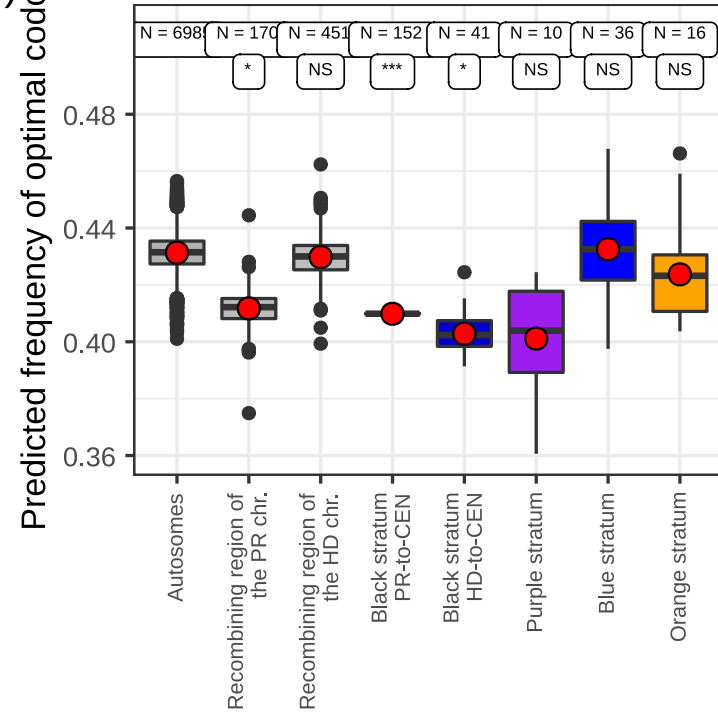


M. lagerheimii
on *Silene vulgaris*
a₁ genome

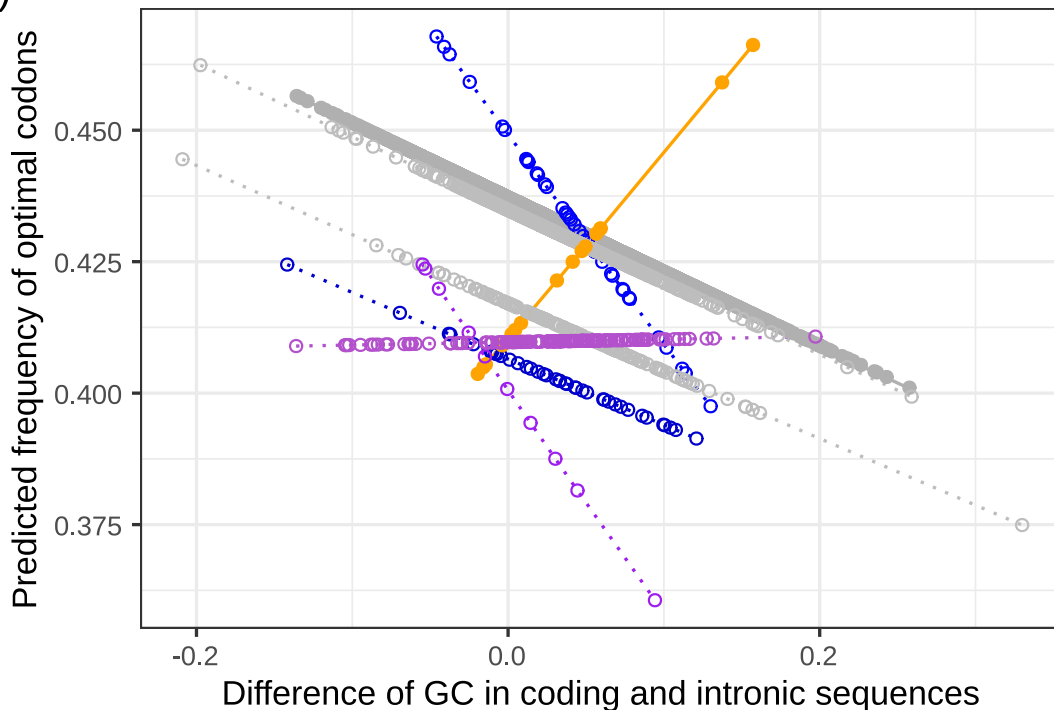
A)



B)



C)

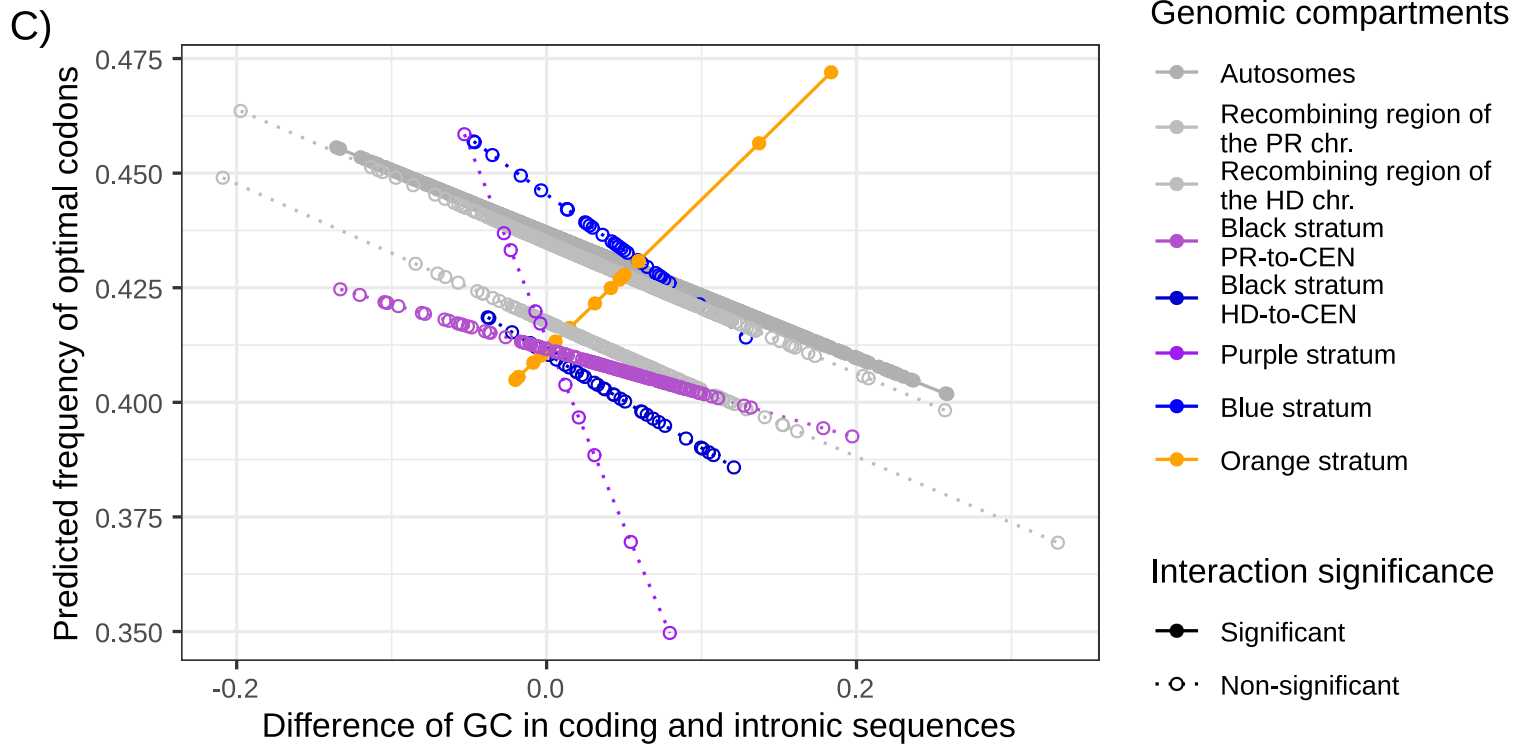
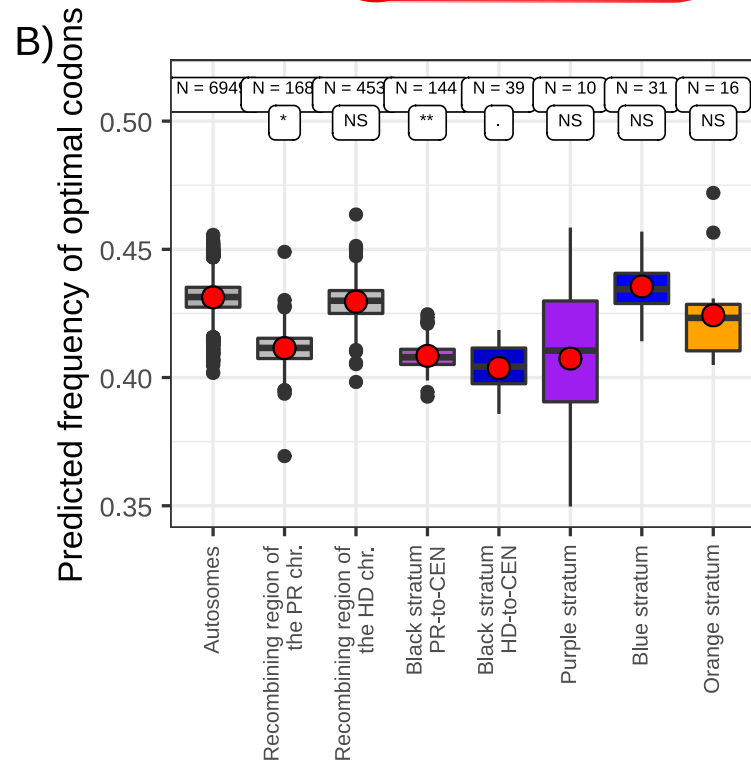
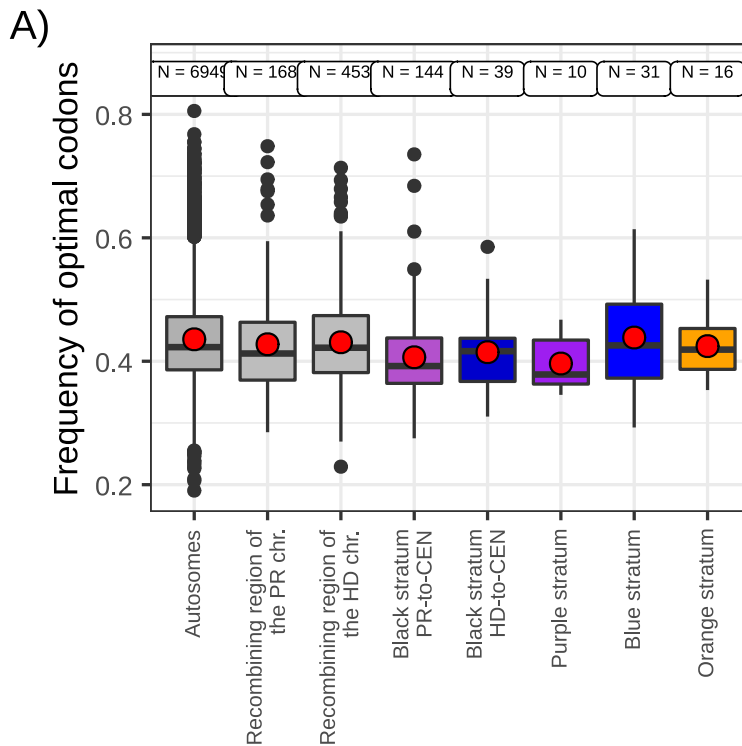


Genomic compartments

- Autosomes
- Recombining region of the PR chr.
- Recombining region of the HD chr.
- Black stratum PR-to-CEN
- Black stratum HD-to-CEN
- Purple stratum
- Blue stratum
- Orange stratum

Interaction significance

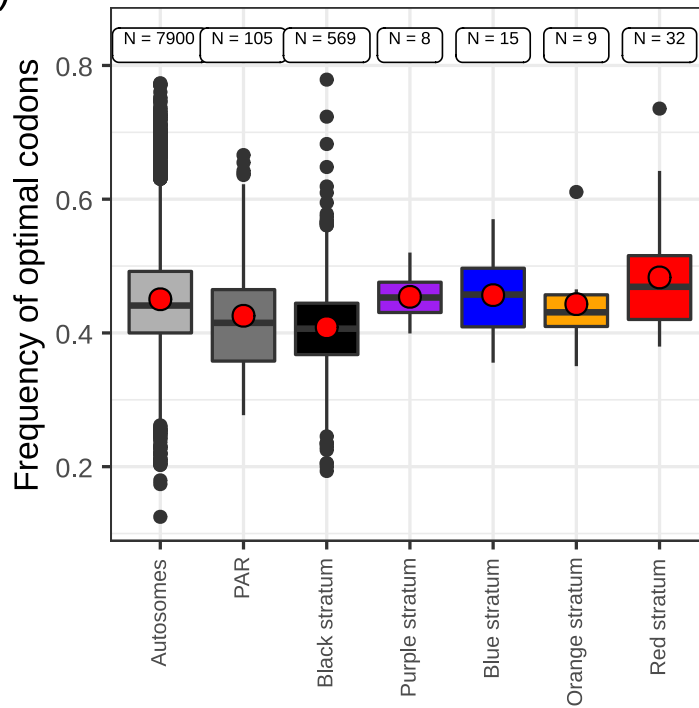
- Significant
- Non-significant



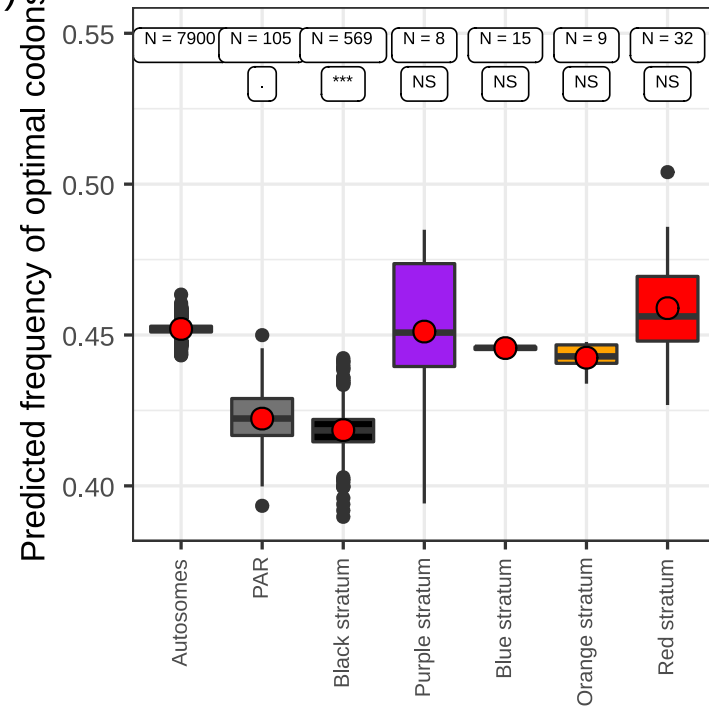


M. lychnidis-diocae
on *Silene latifolia*
a₁ genome

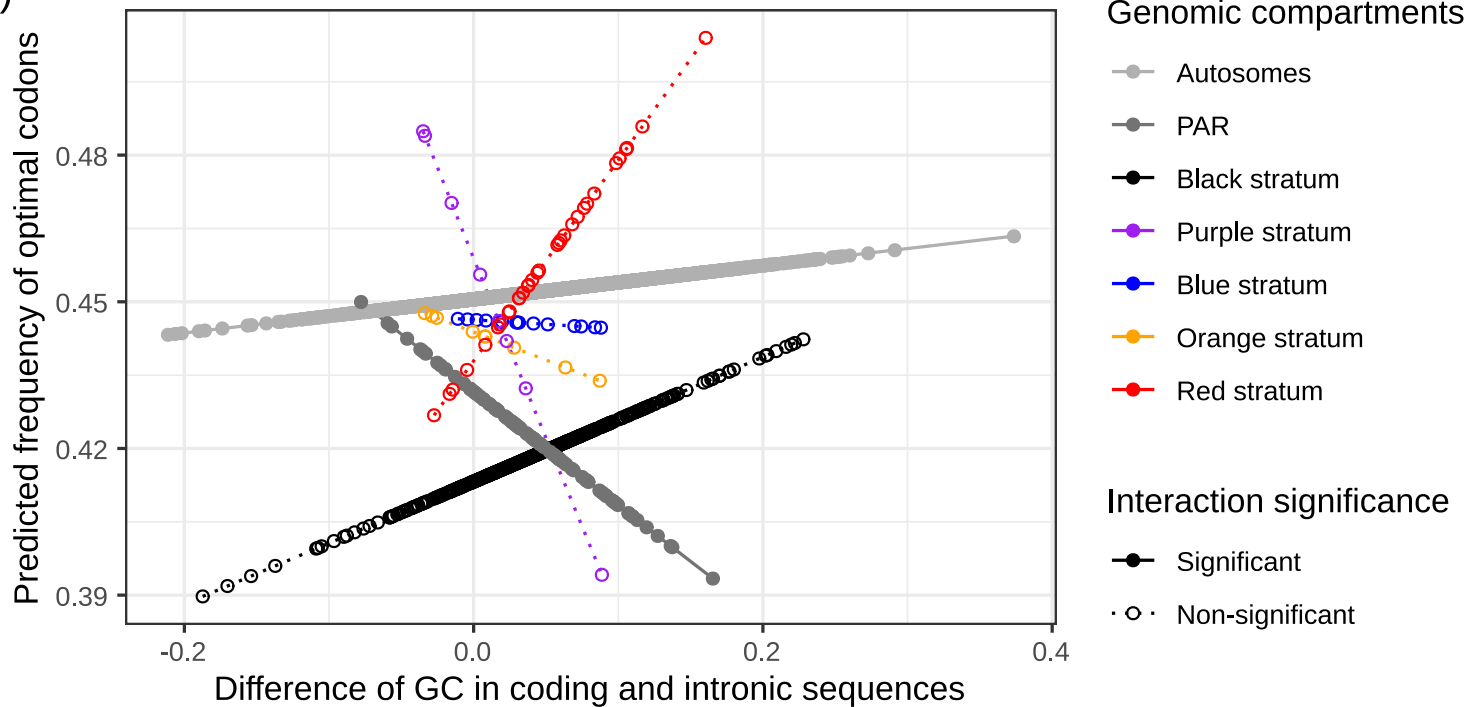
A)



B)

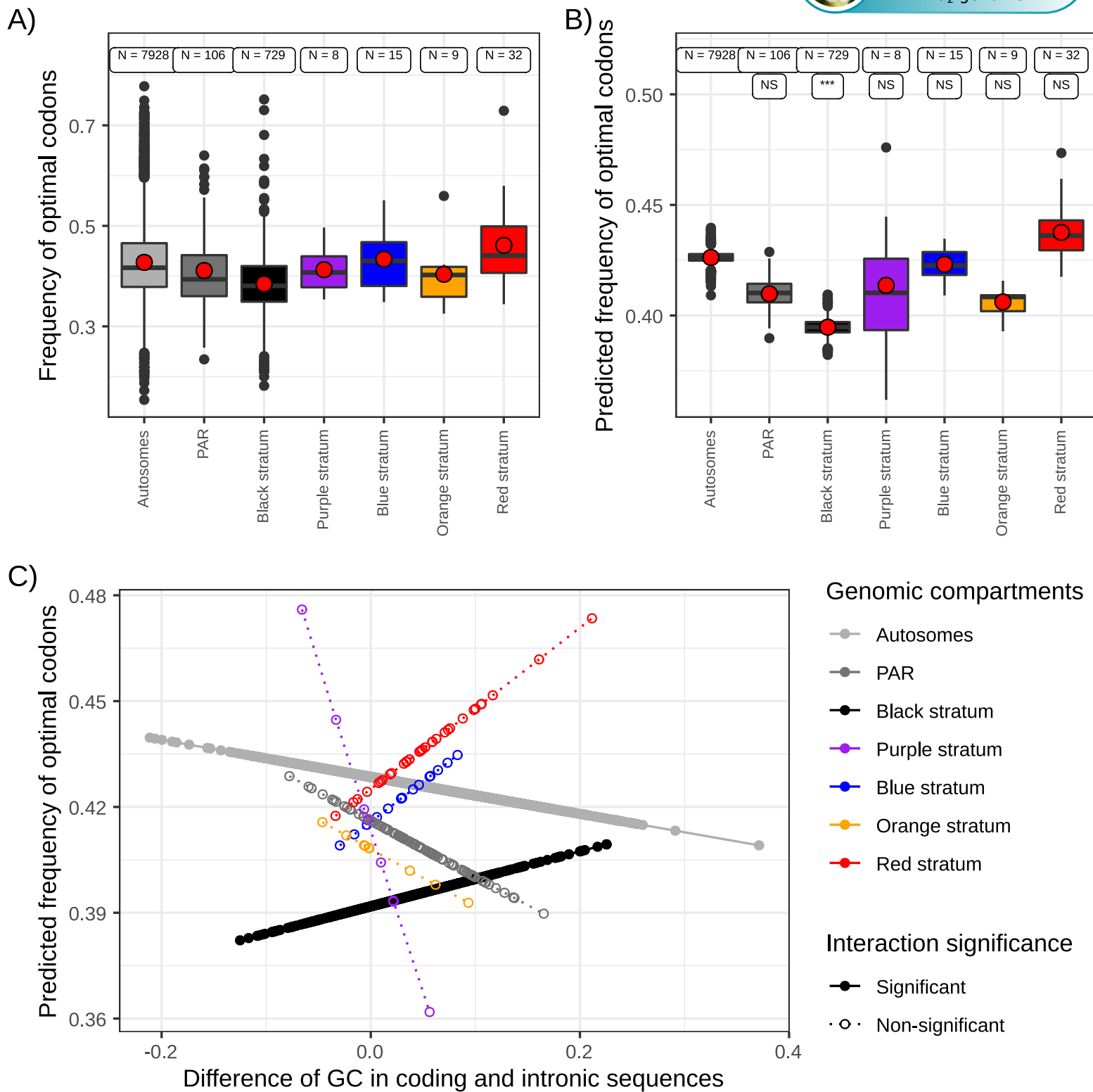


C)





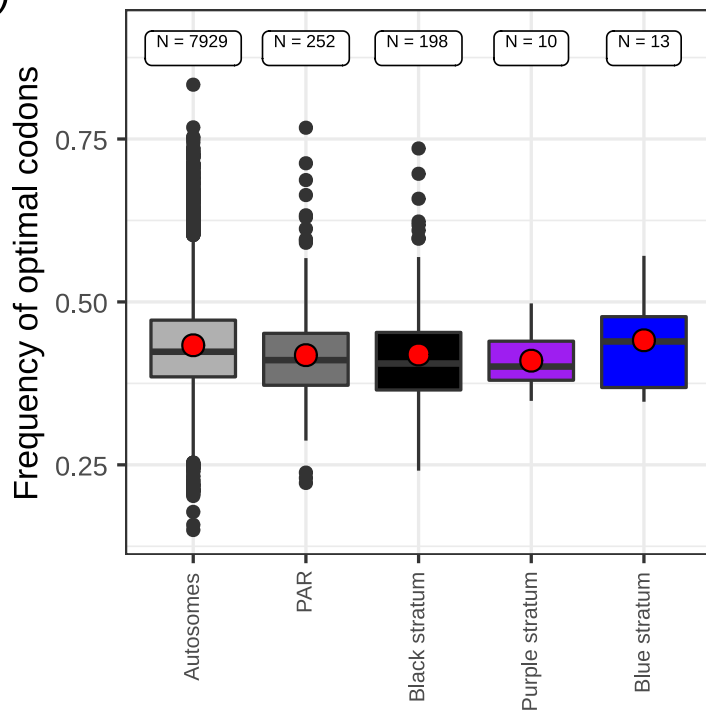
M. lychnidis-dioicae
on *Silene latifolia*
a₂ genome



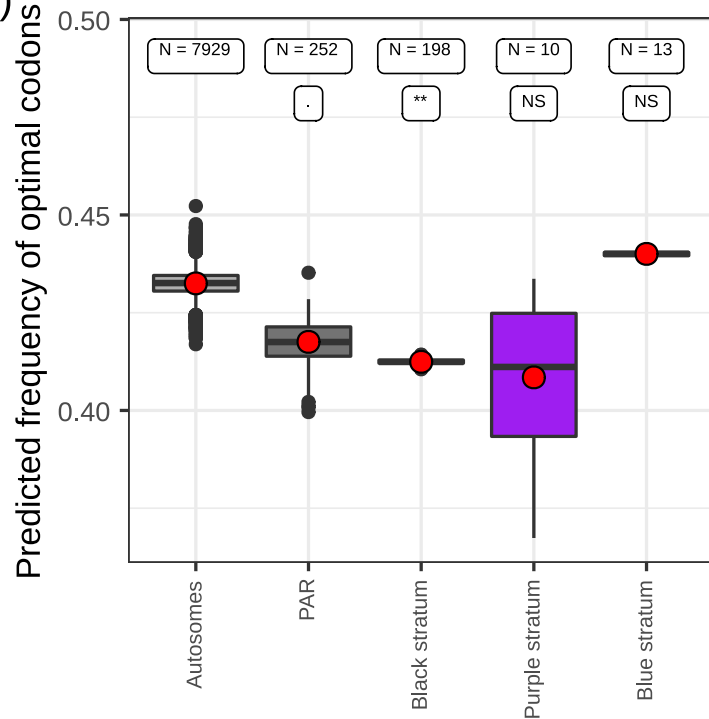


M. silenes-acaulis
on *Silene acaulis*
a₁ genome

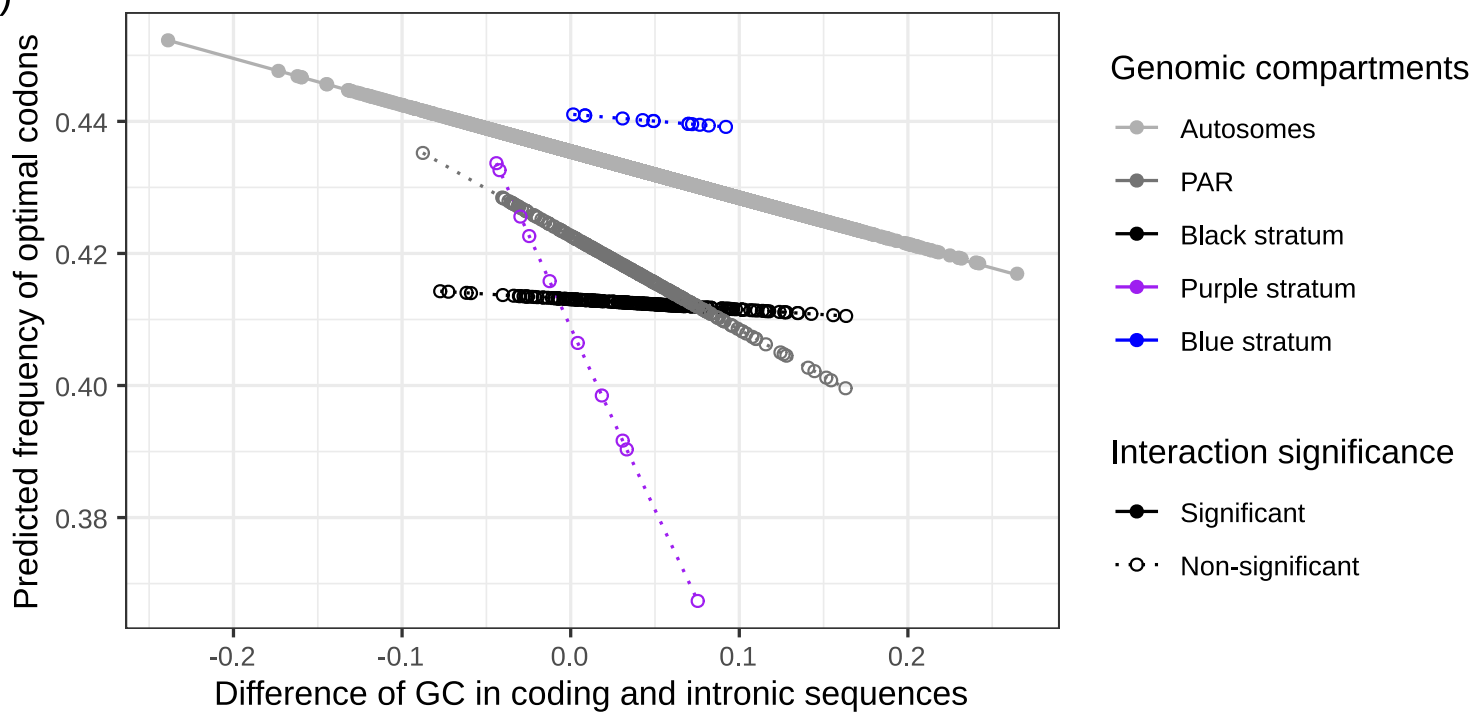
A)

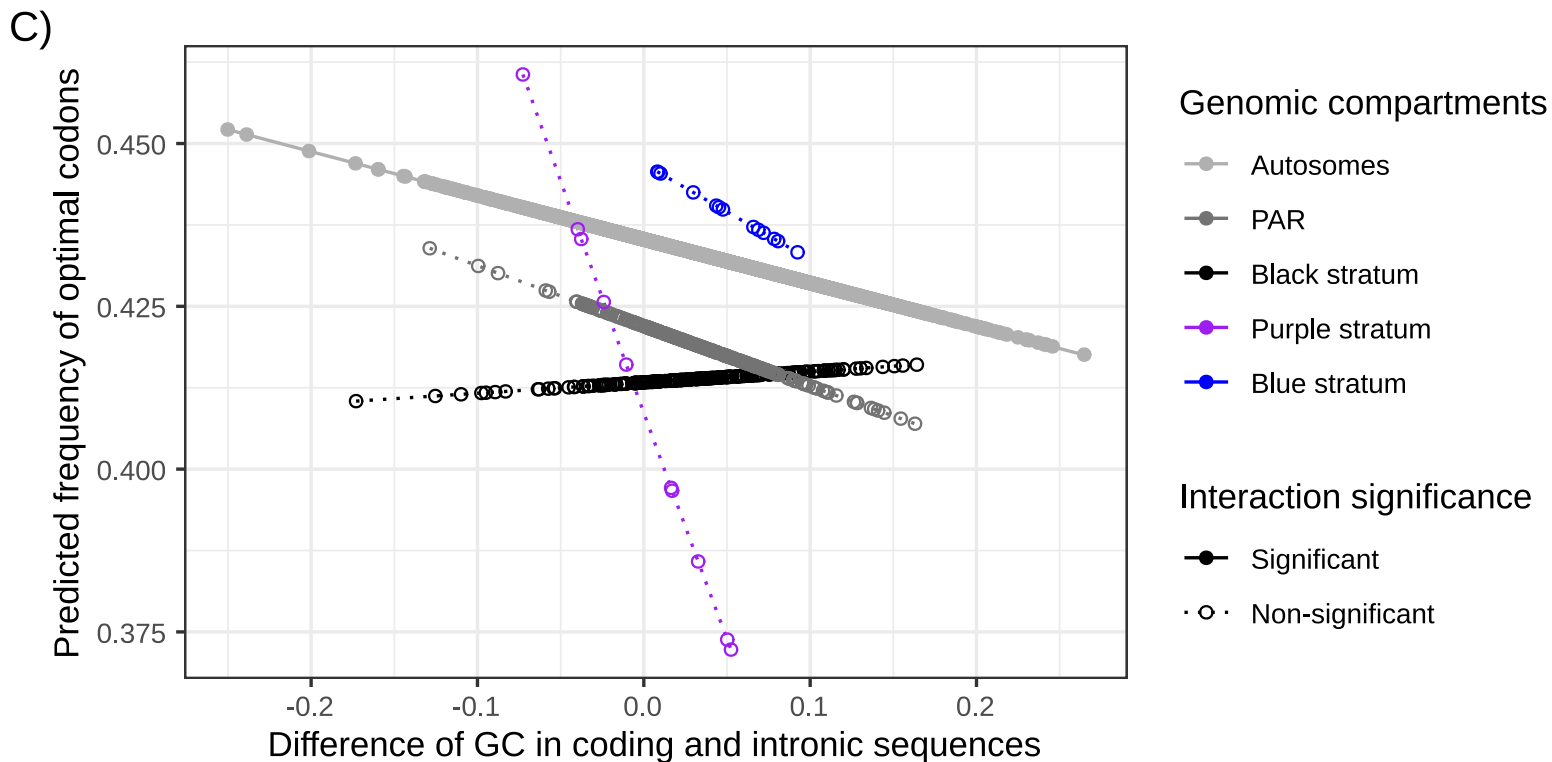
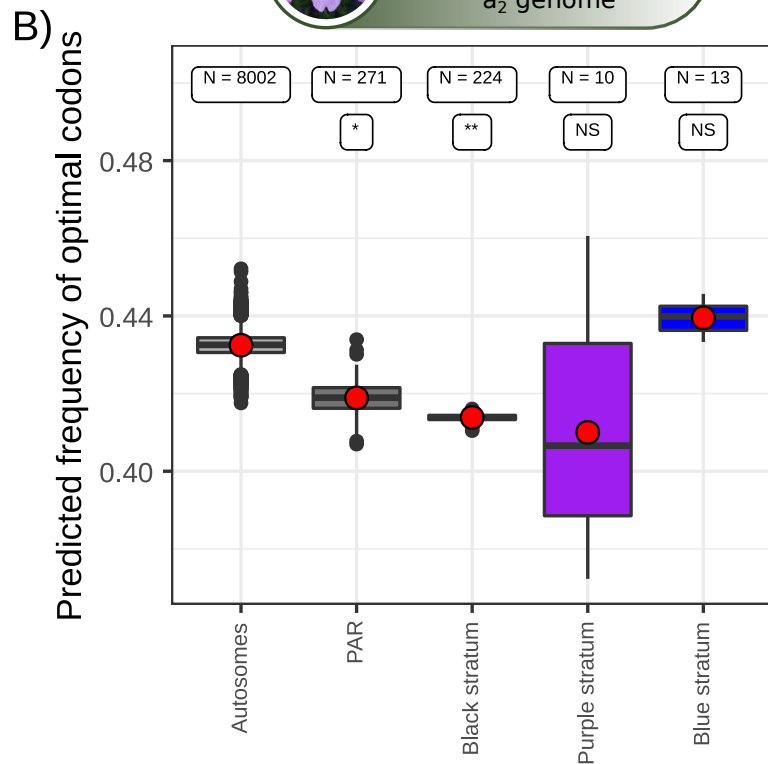
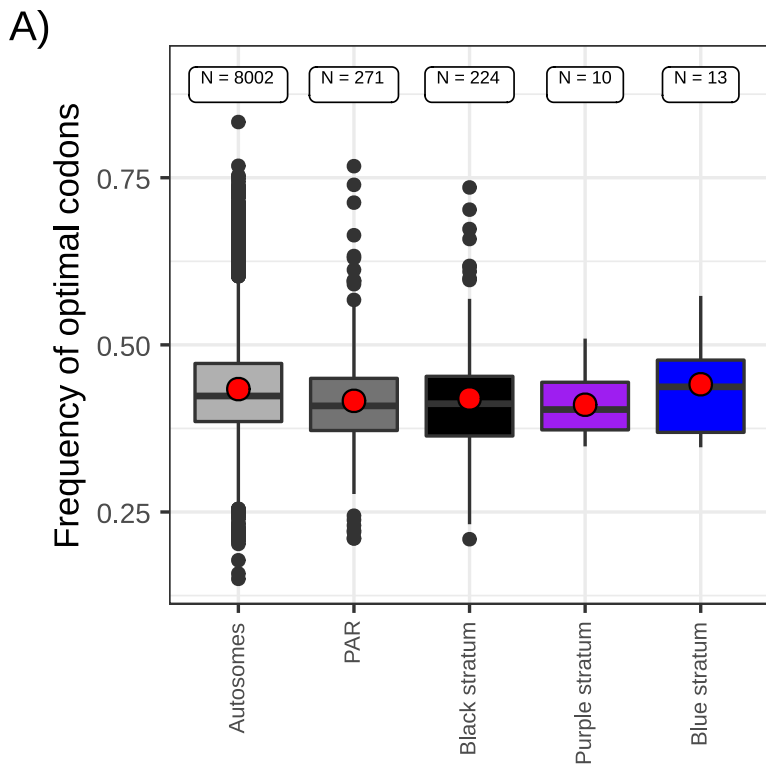
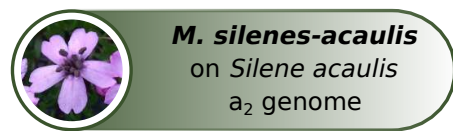


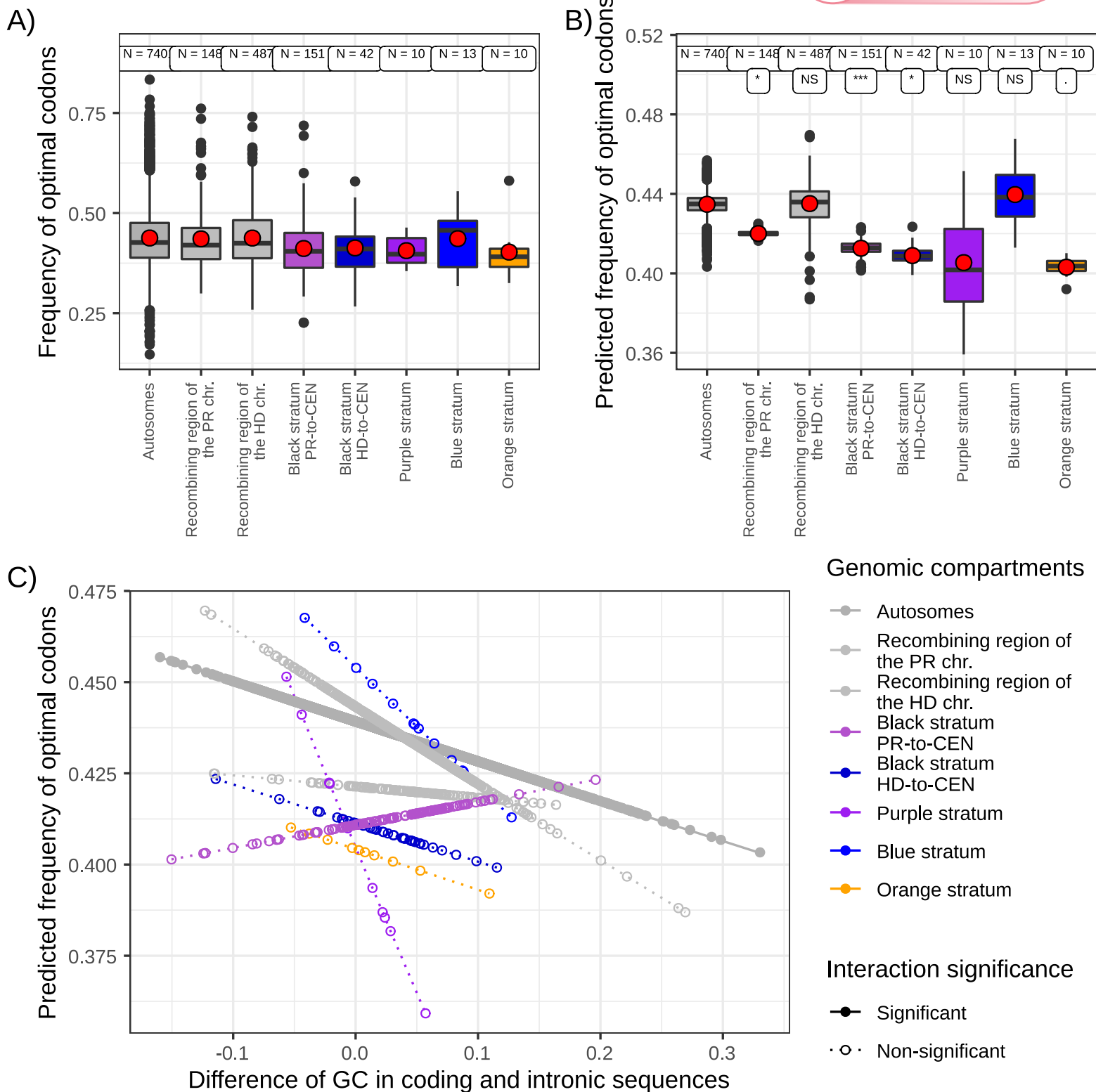
B)



C)

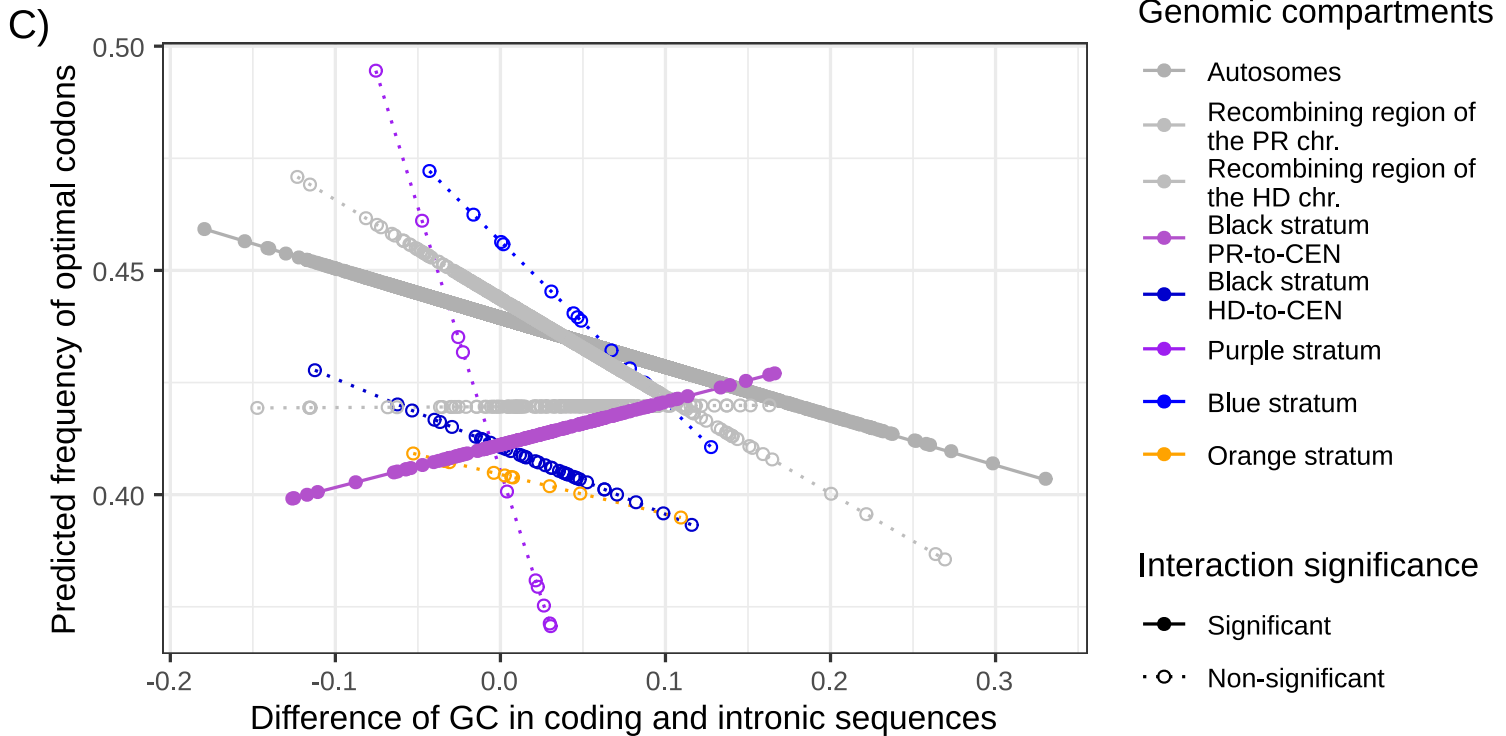
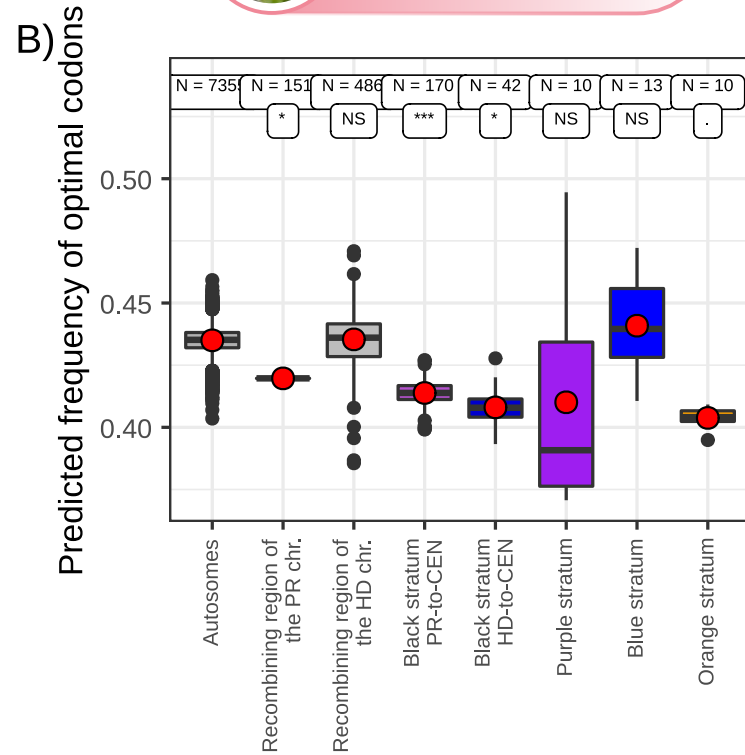
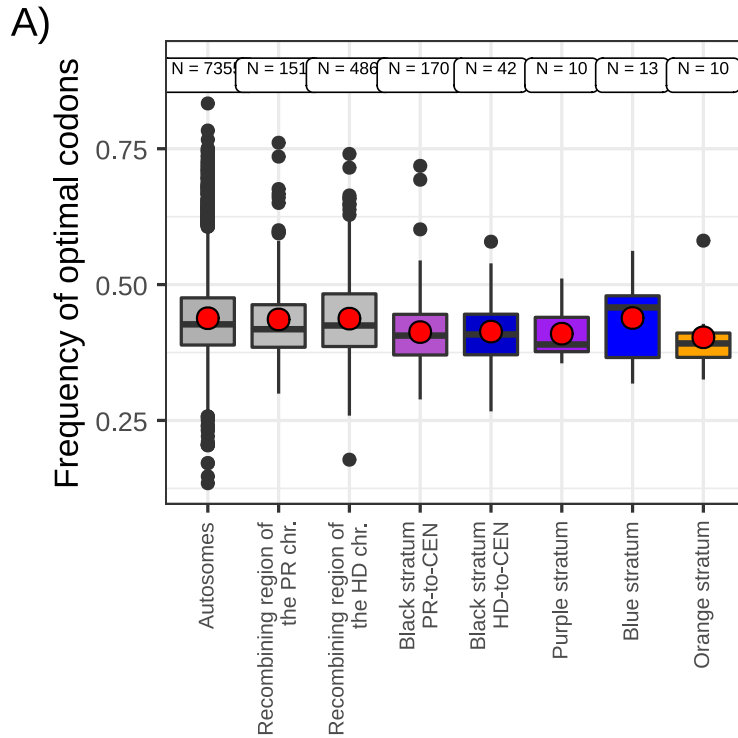








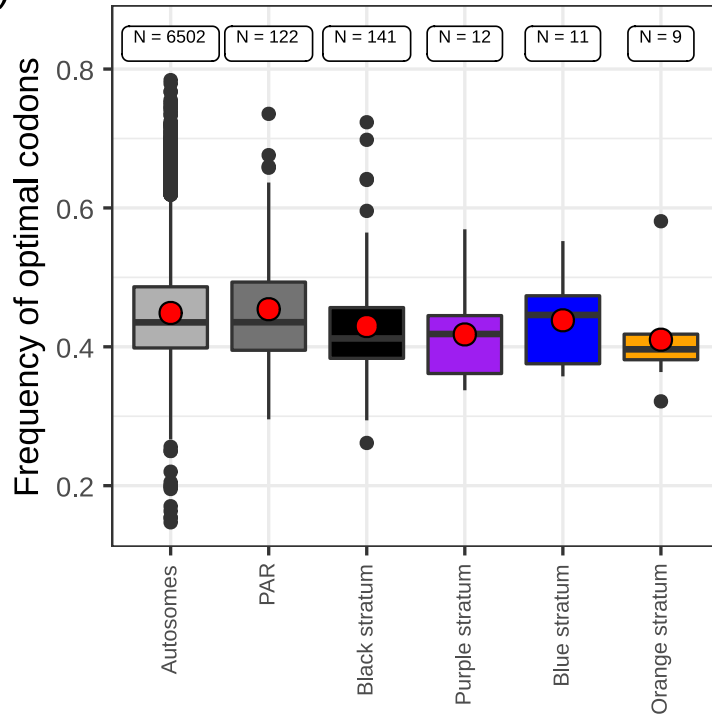
M. saponariae
on *Saponaria officinalis*
a₂ genome



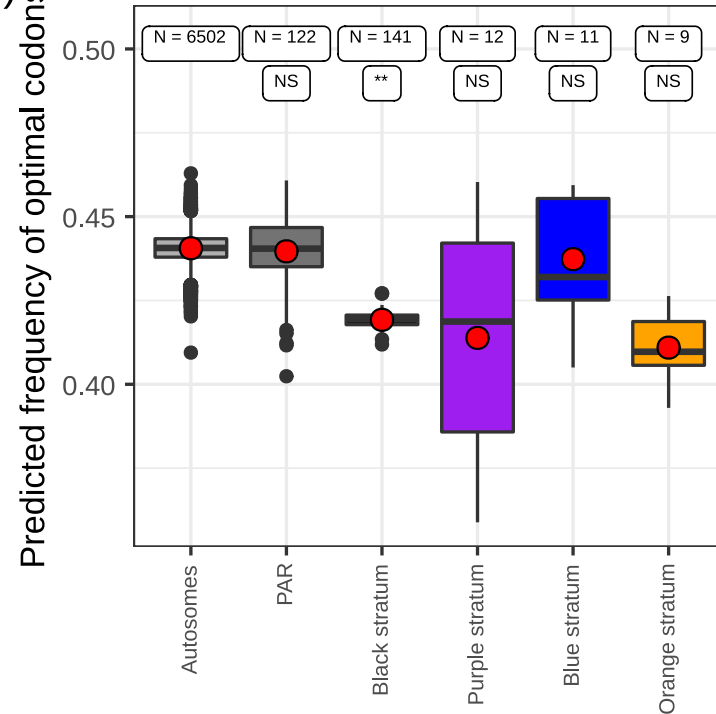


M. scabiosae
on *Knautia arvensis*
a₁ genome

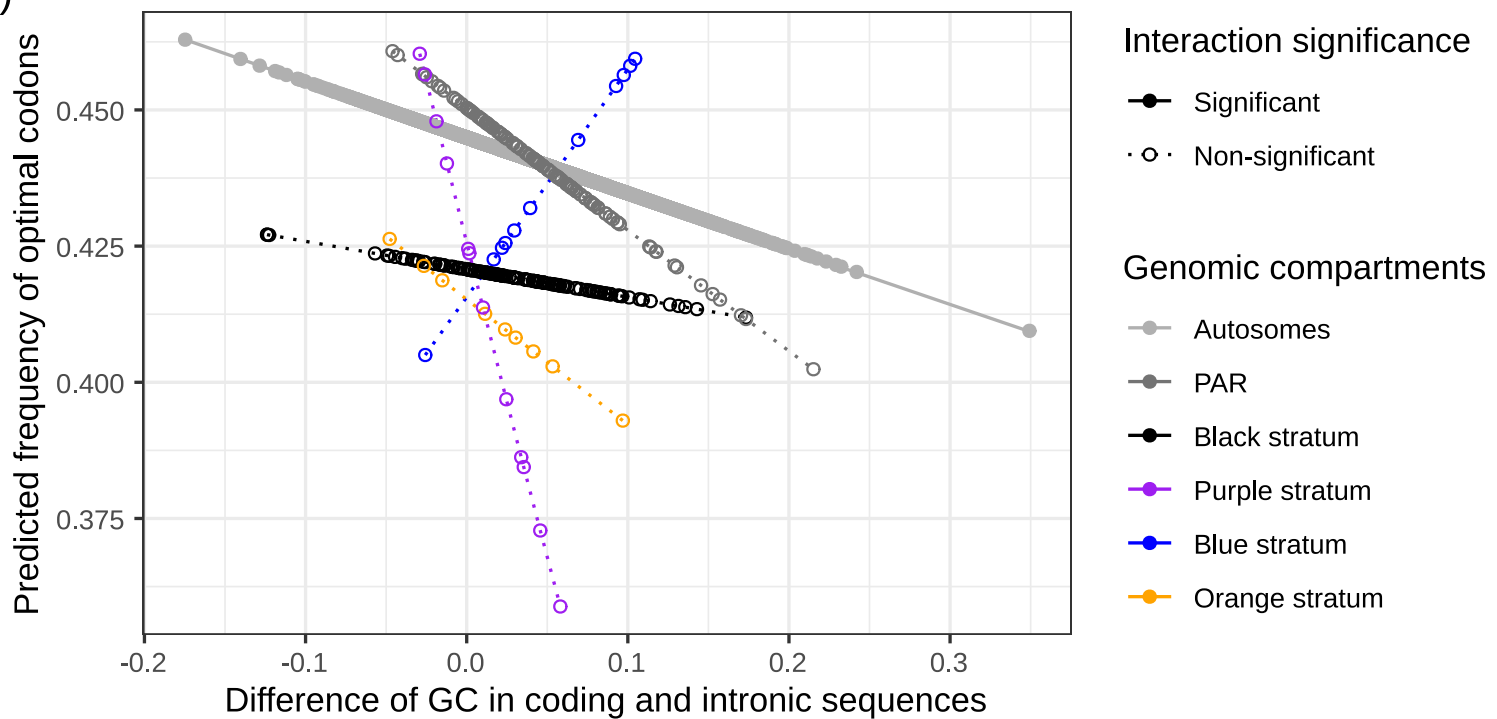
A)



B)



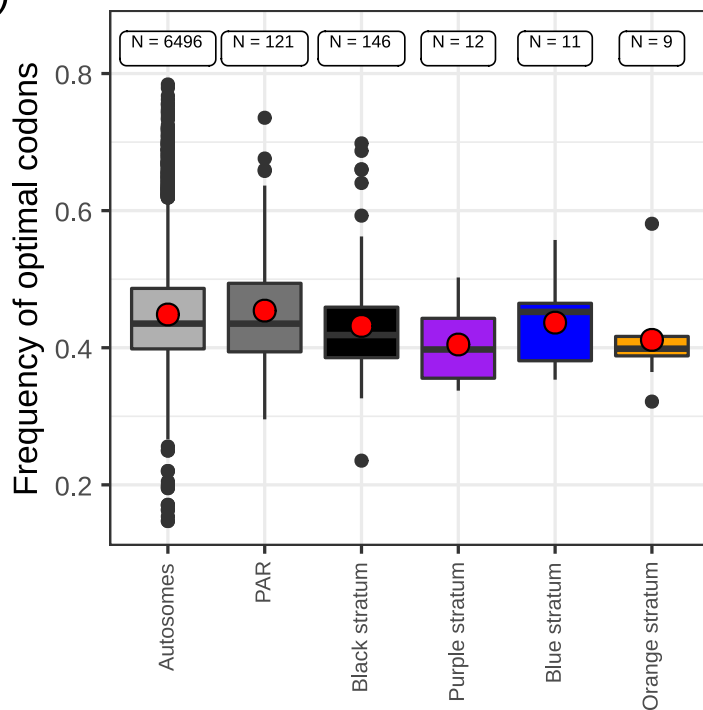
C)



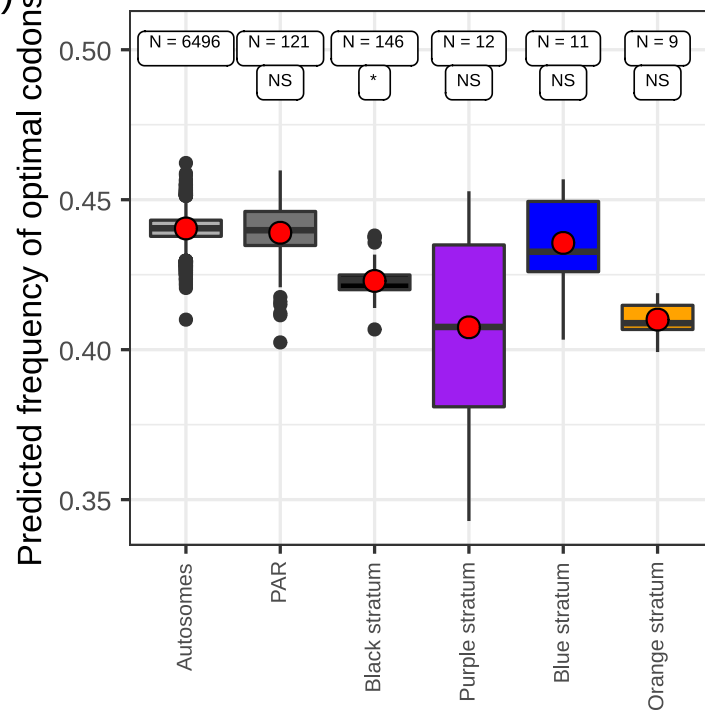


M. scabiosae
on *Knautia arvensis*
a₂ genome

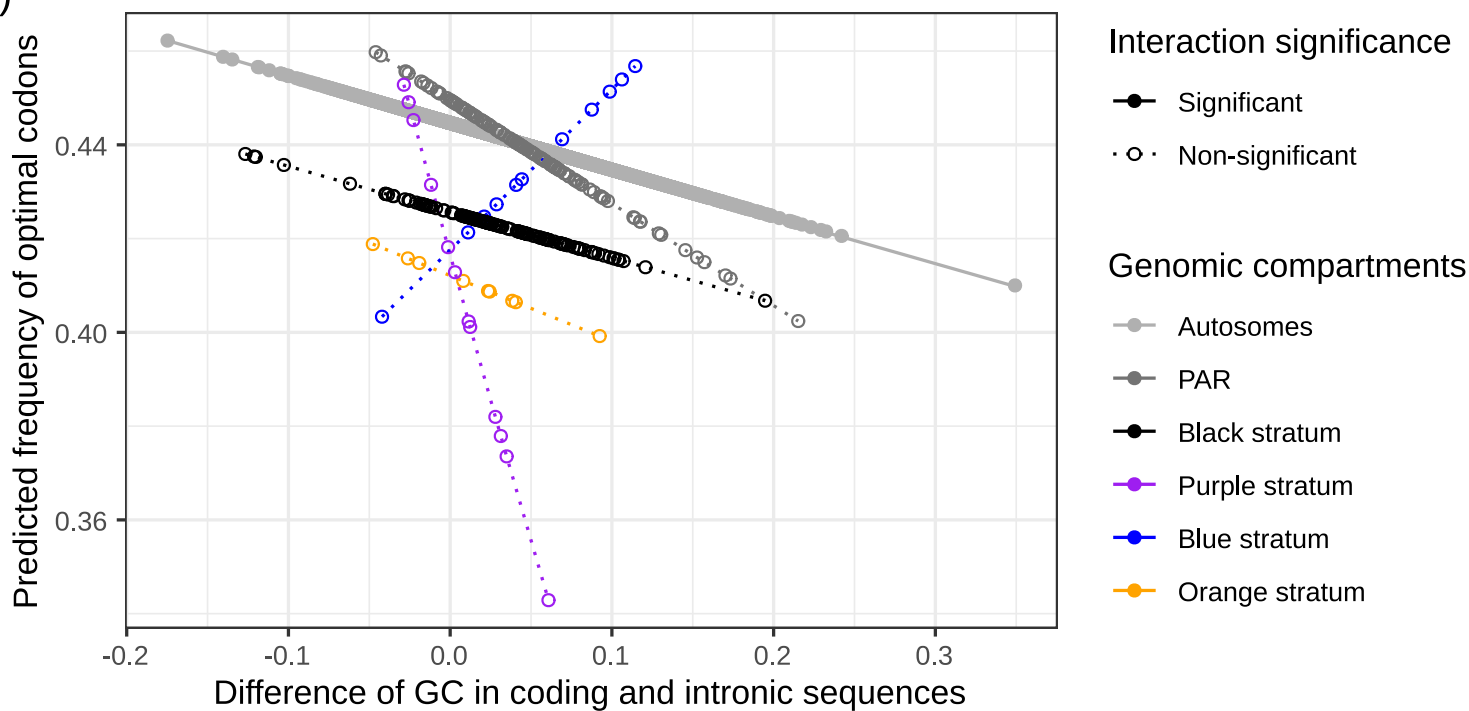
A)



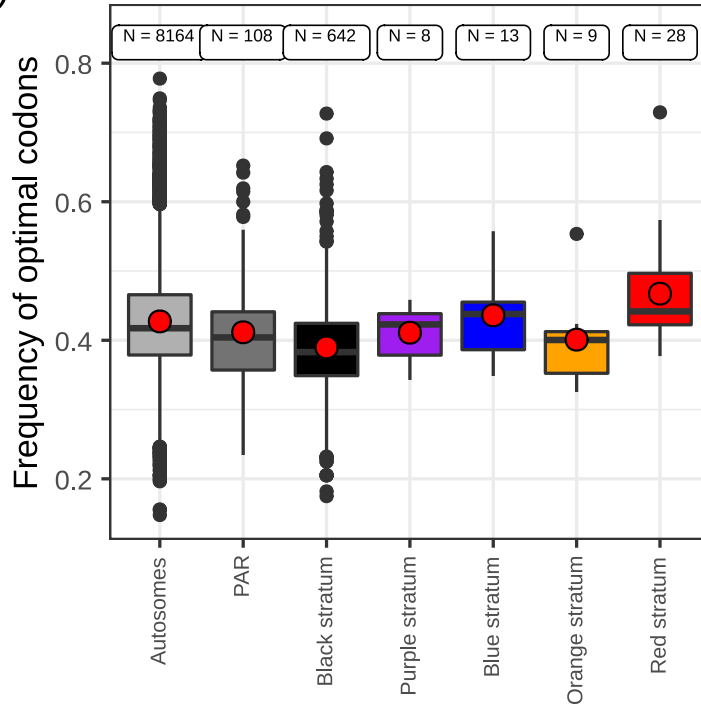
B)



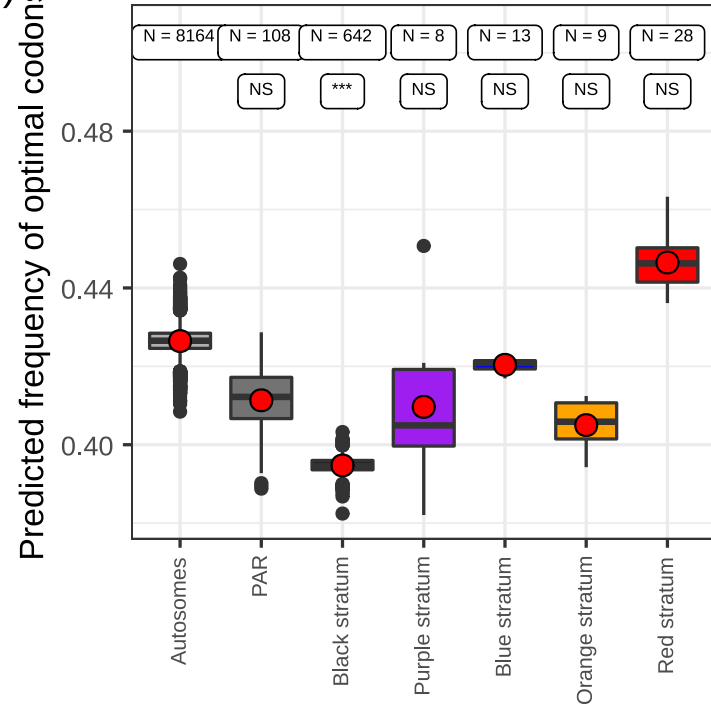
C)



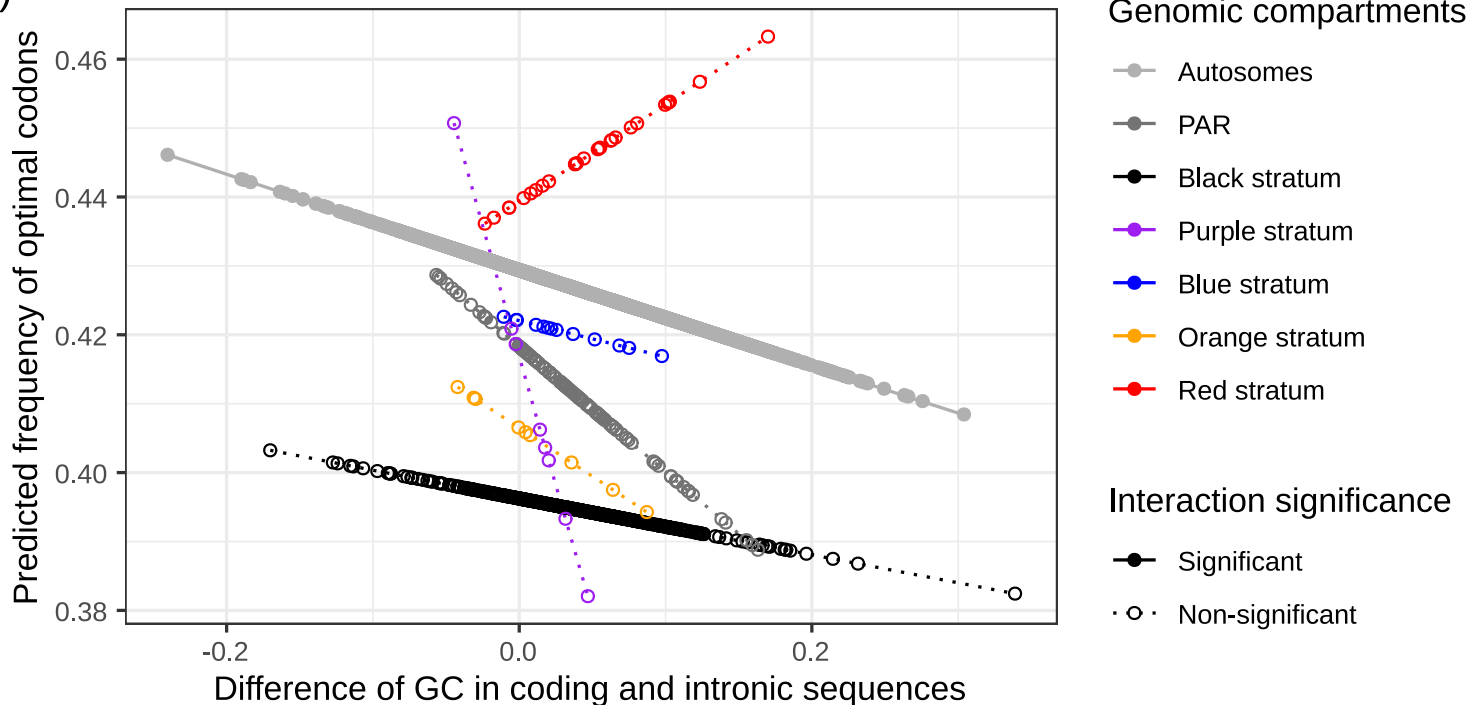
A)



B)



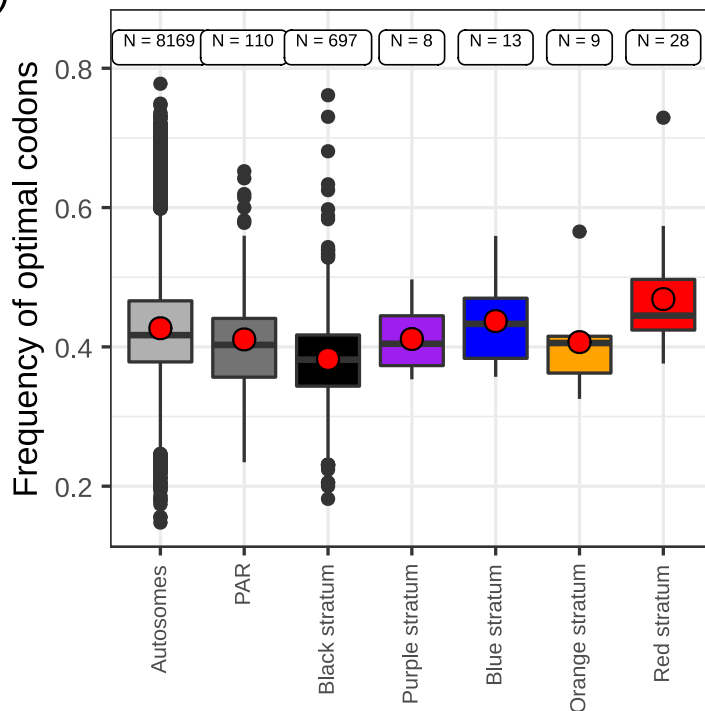
C)



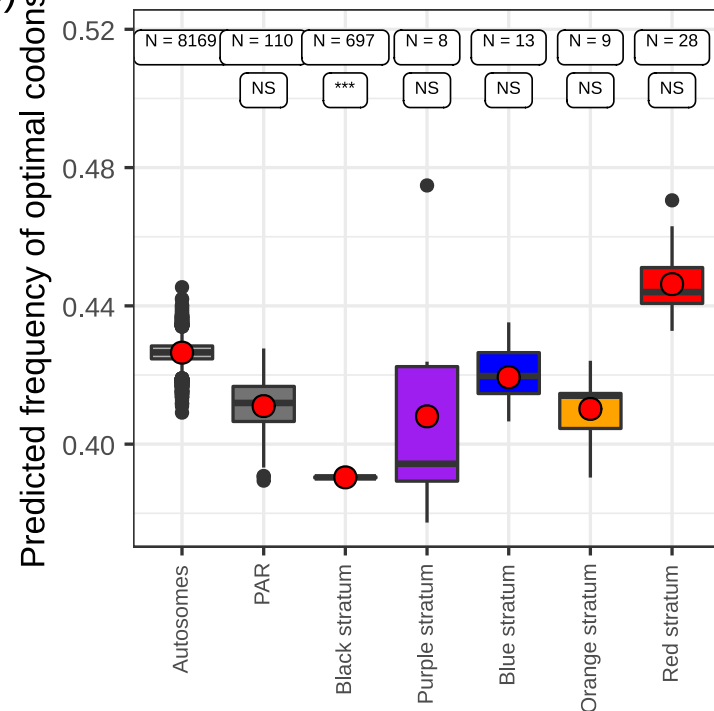


M. silenes-dioicae
on *Silene dioica*
a₂ genome

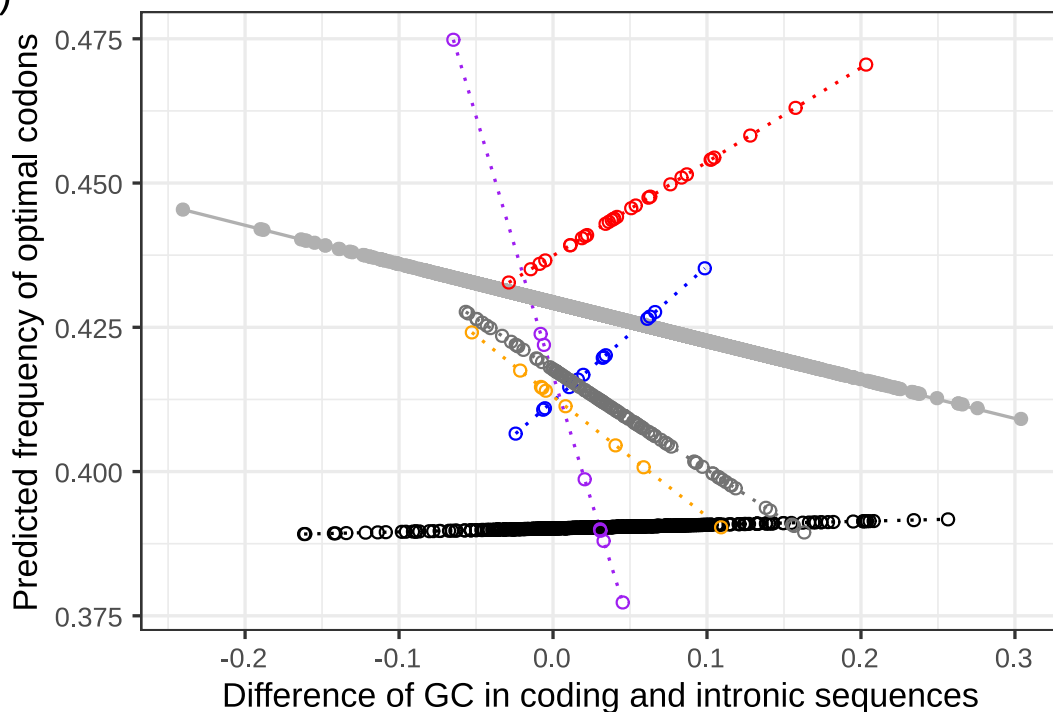
A)



B)



C)



Genomic compartments

- Autosomes
- PAR
- Black stratum
- Purple stratum
- Blue stratum
- Orange stratum
- Red stratum

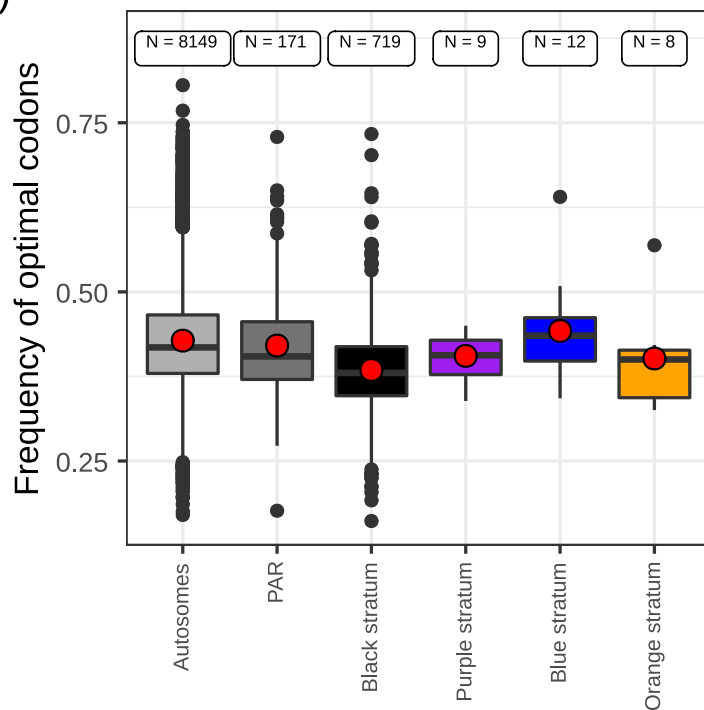
Interaction significance

- Significant
- Non-significant

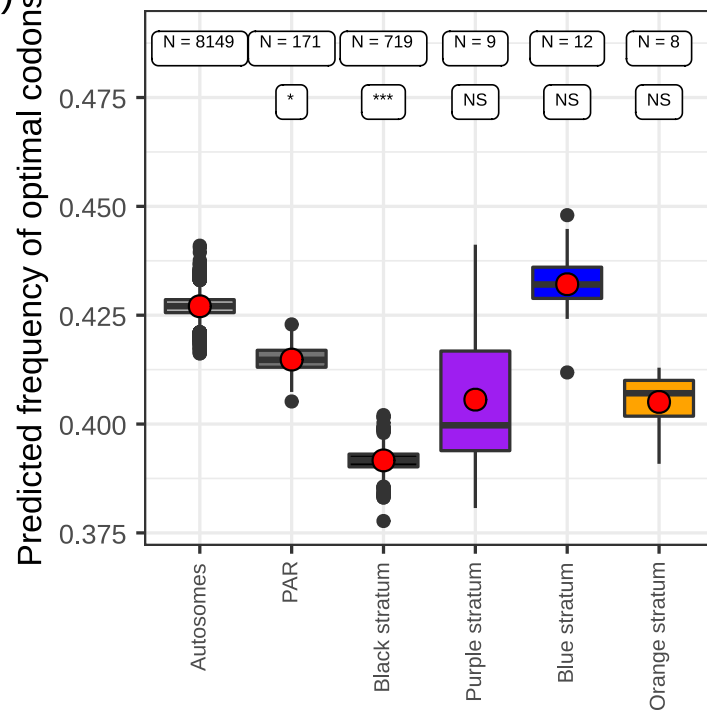


M. coronariae
on *Silene flos-cuculi*
a₁ genome

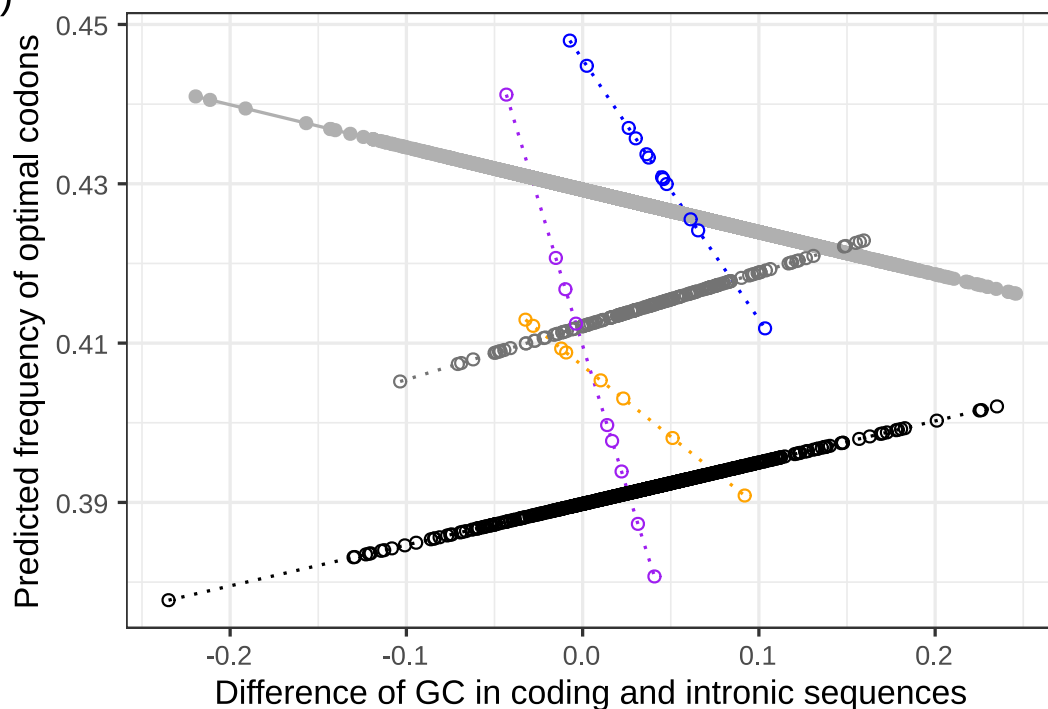
A)



B)



C)

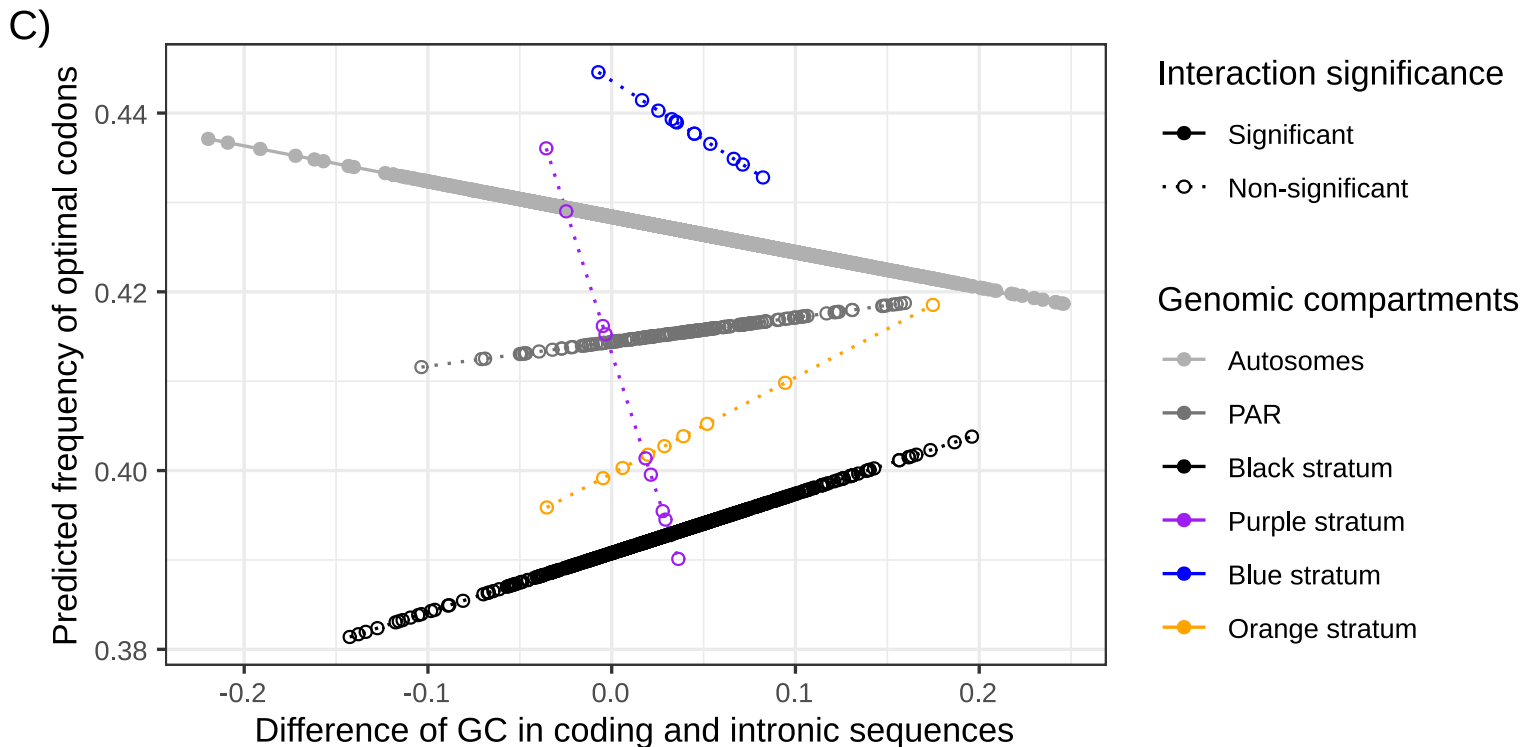
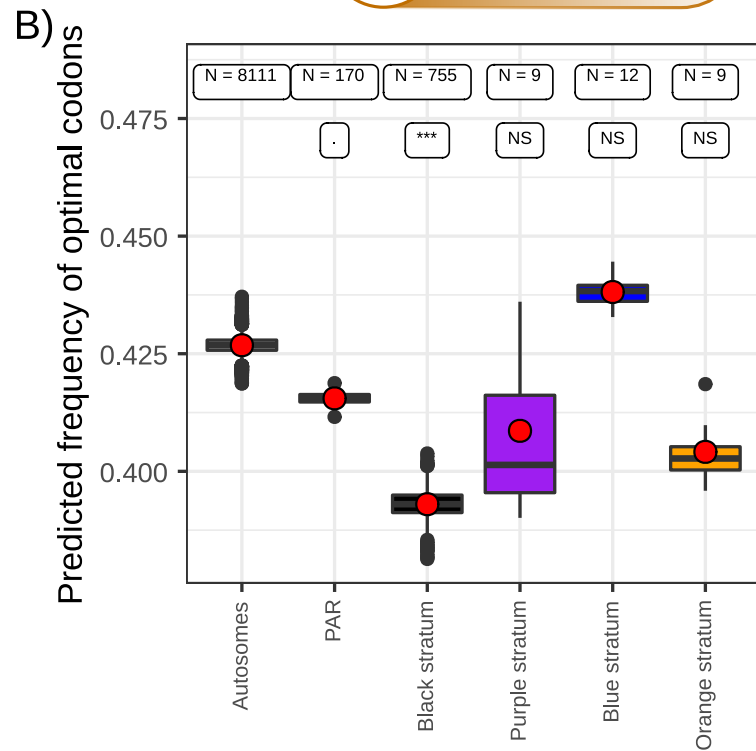
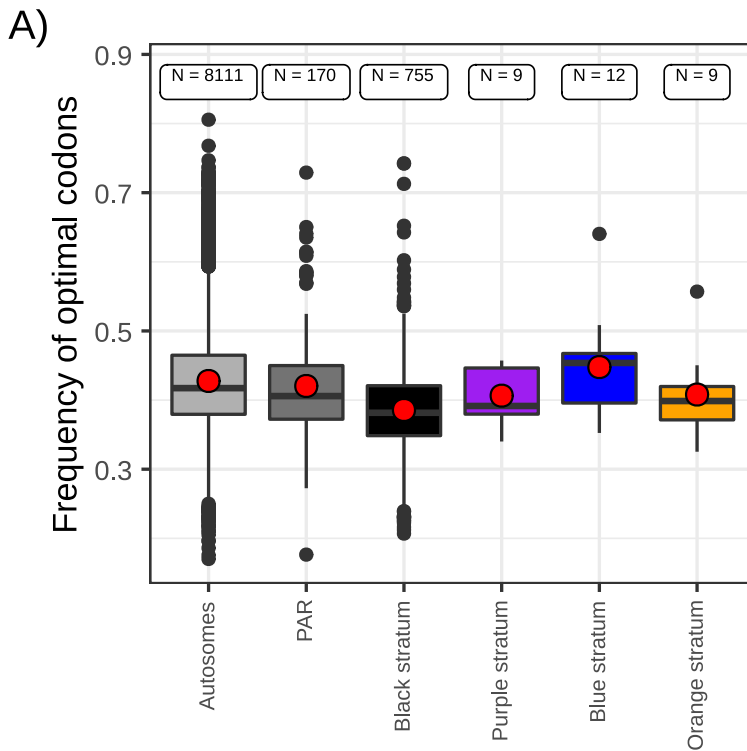


Interaction significance

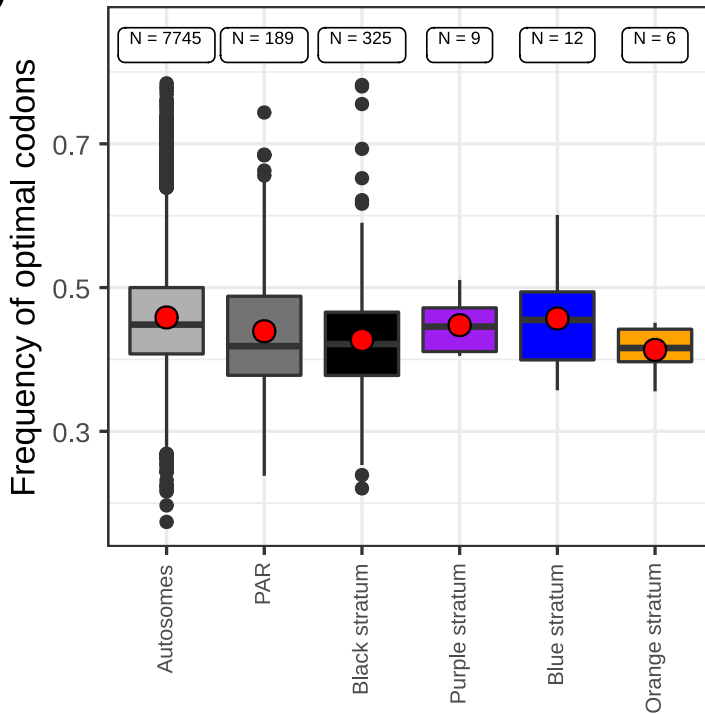
- Significant
- Non-significant

Genomic compartments

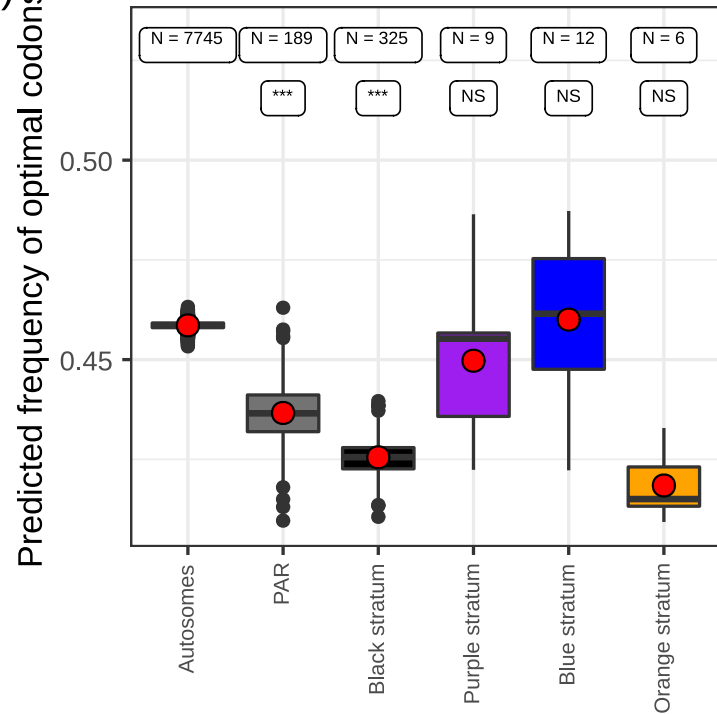
- Autosomes
- PAR
- Black stratum
- Purple stratum
- Blue stratum
- Orange stratum



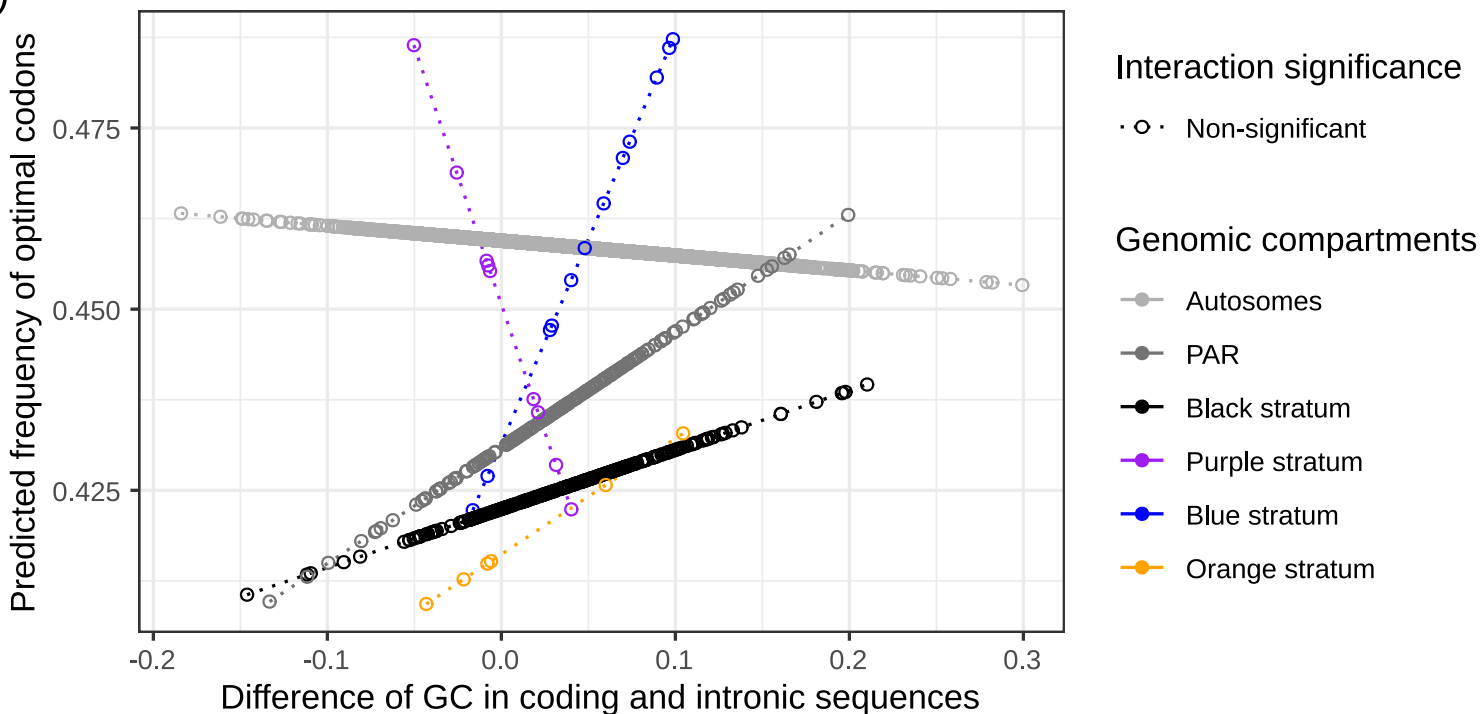
A)



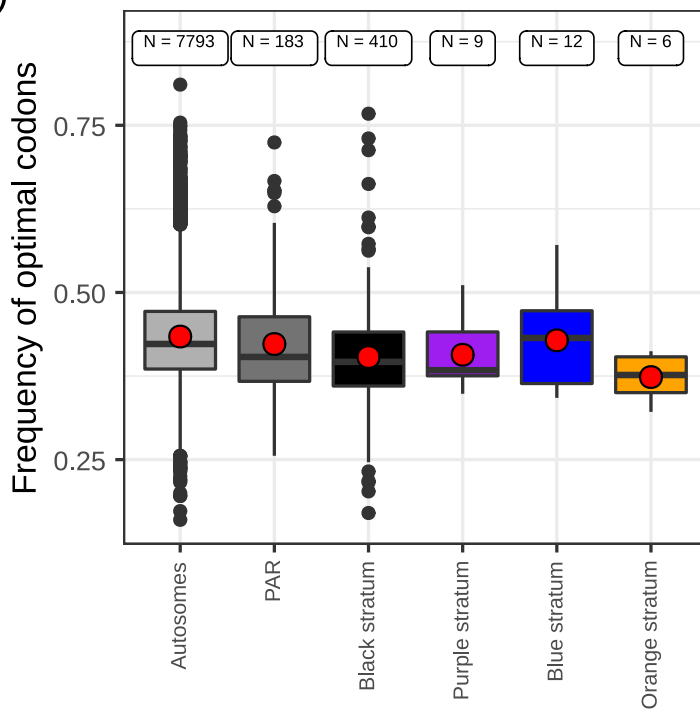
B)



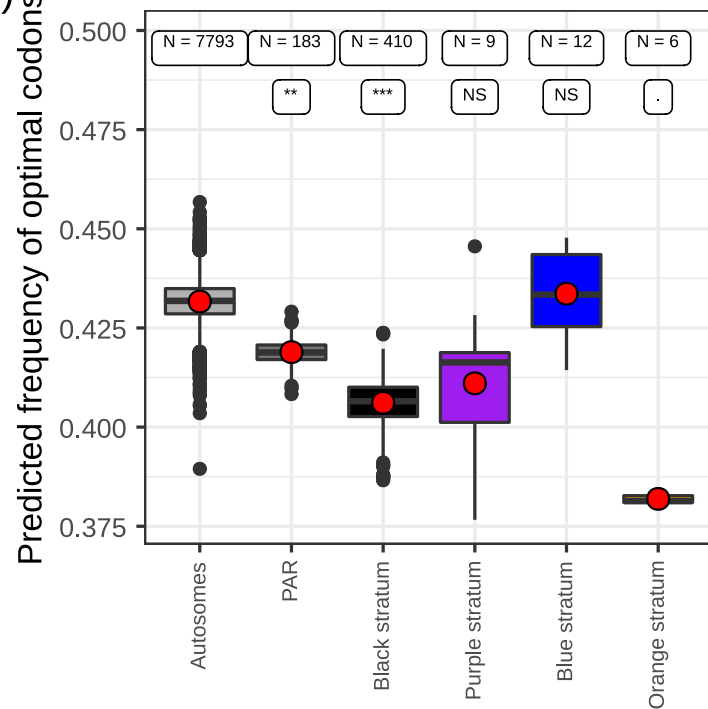
C)



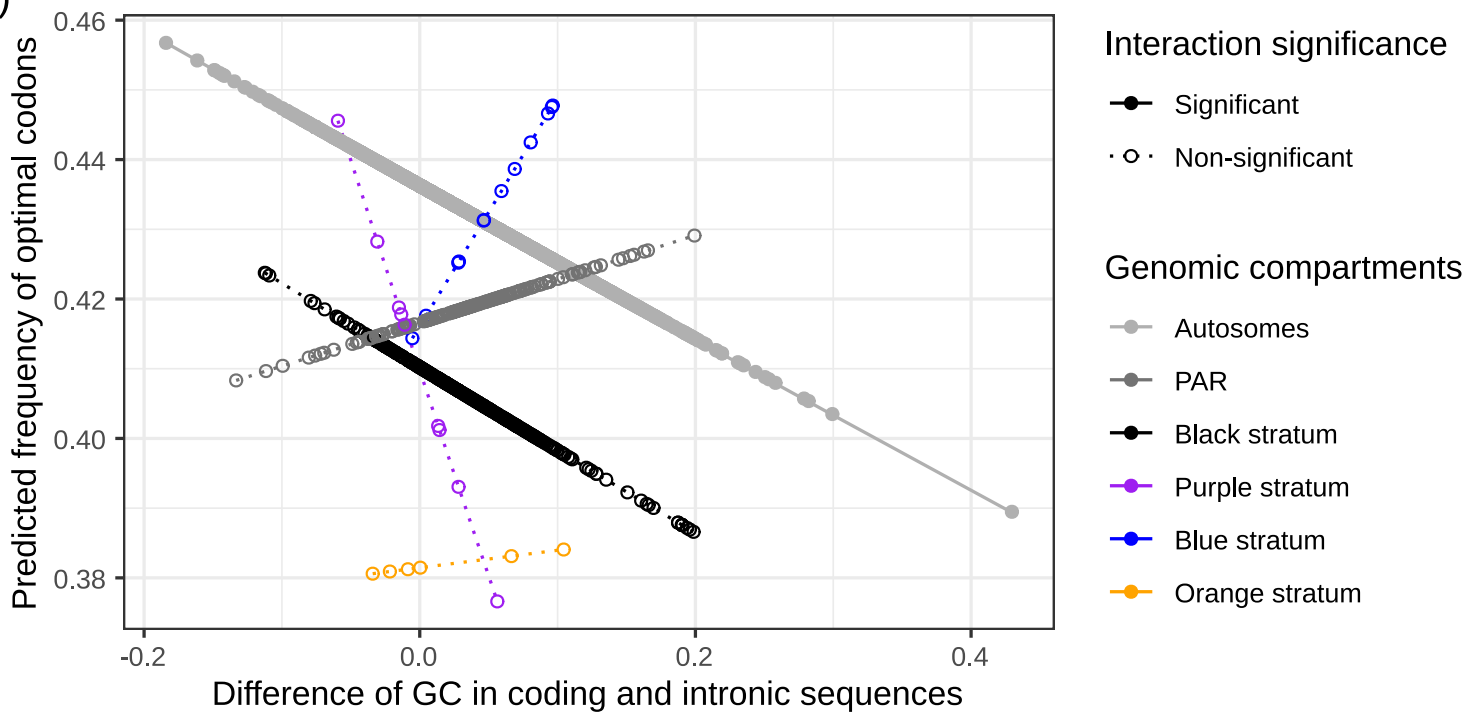
A)

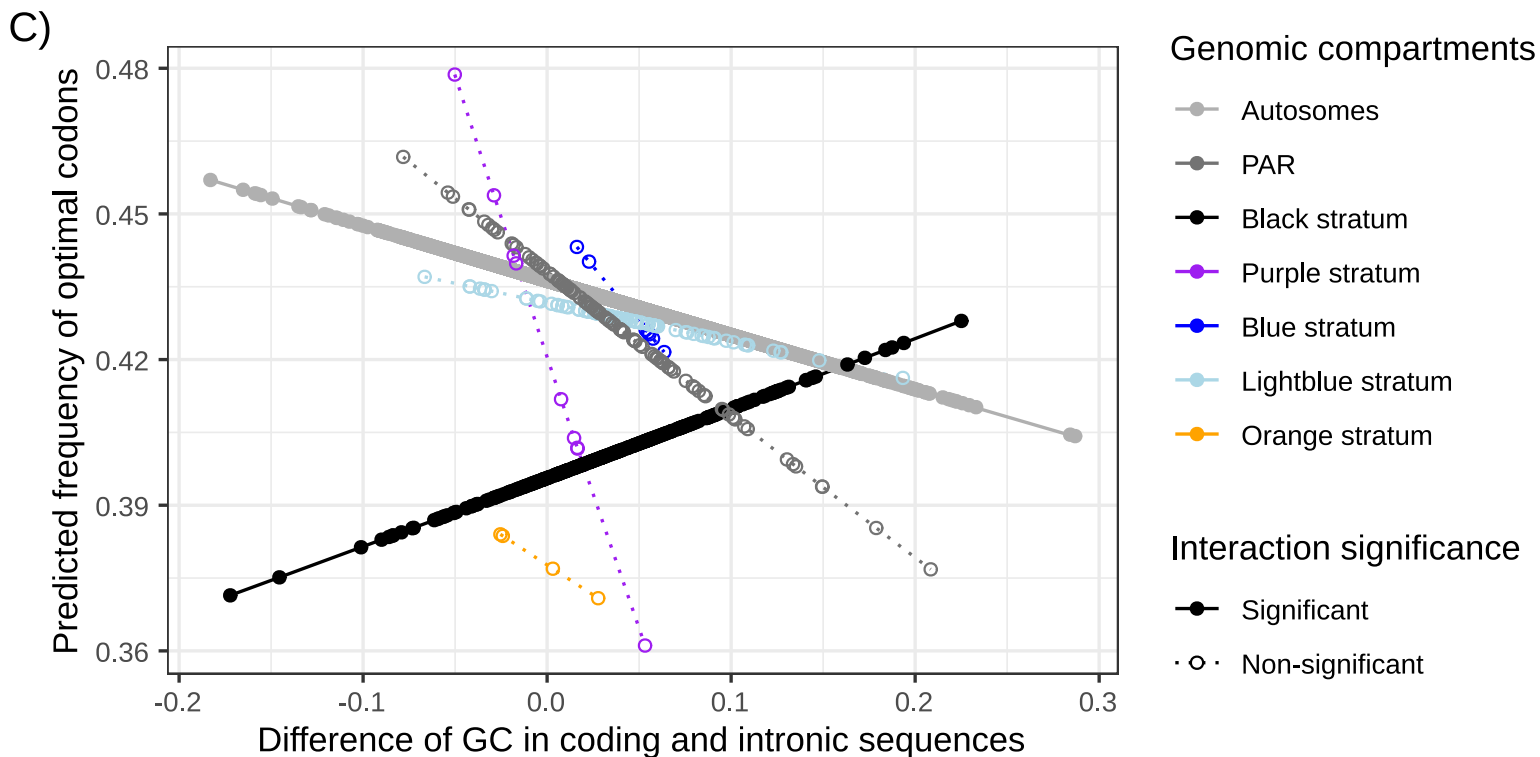
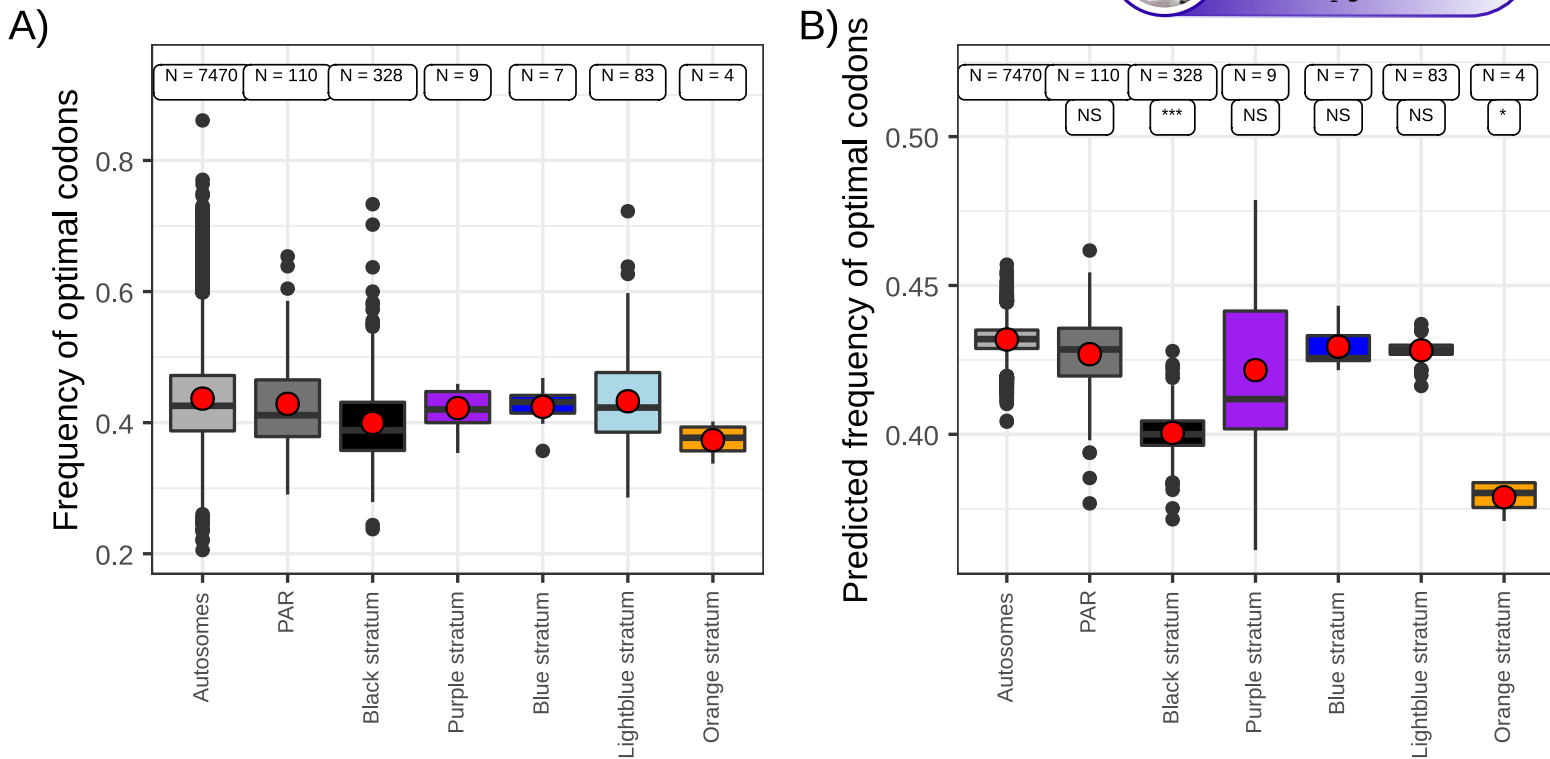


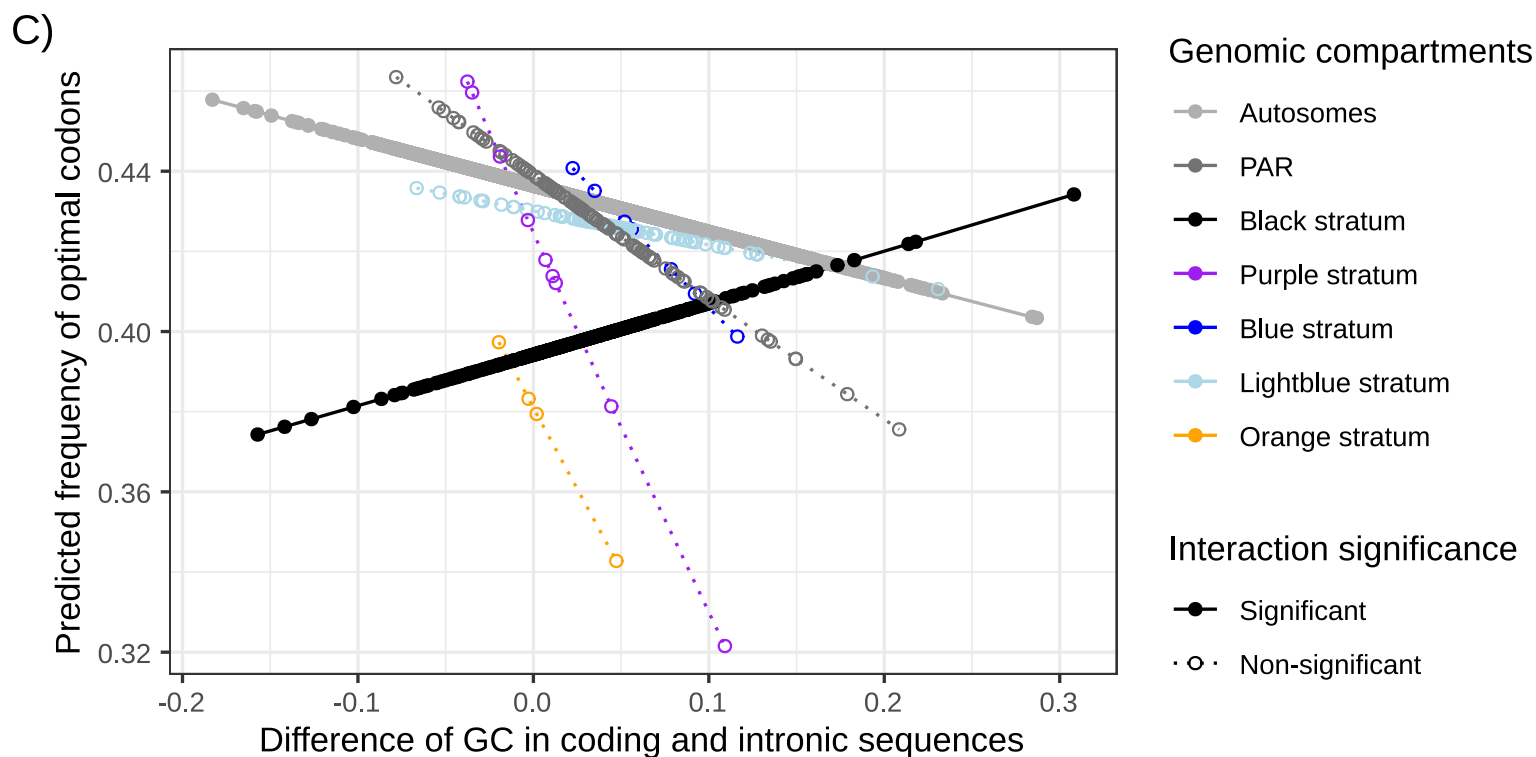
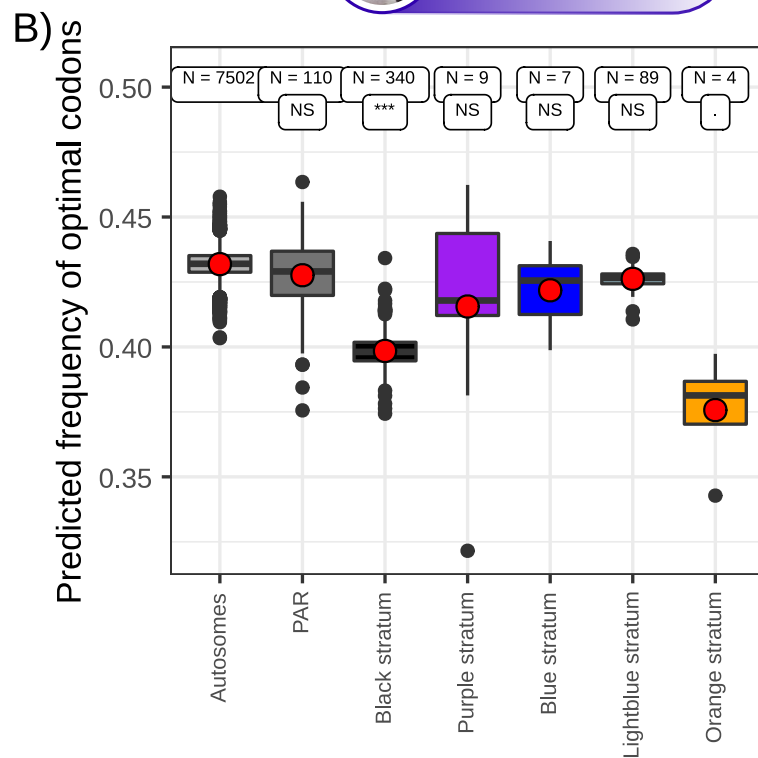
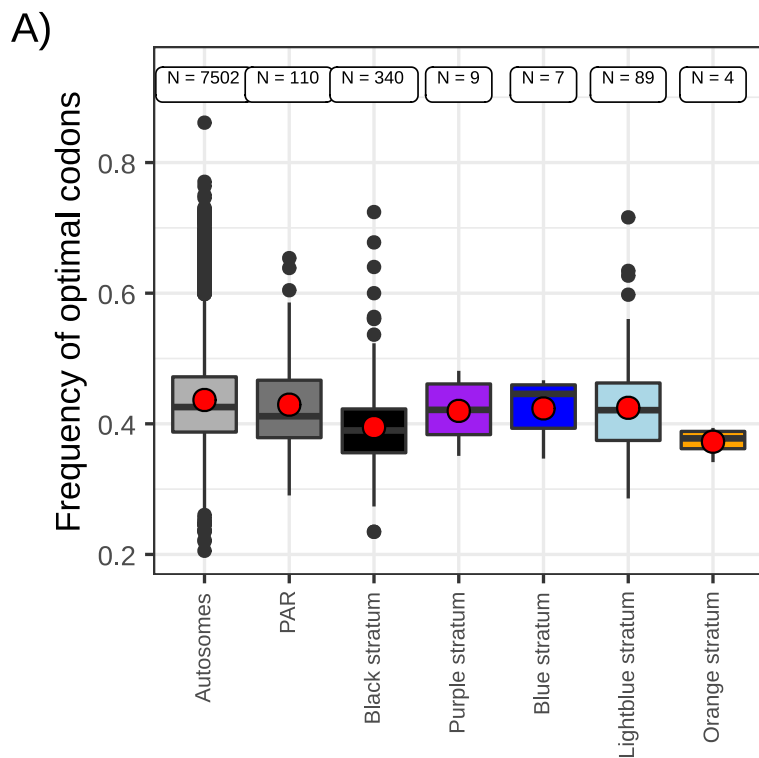
B)



C)

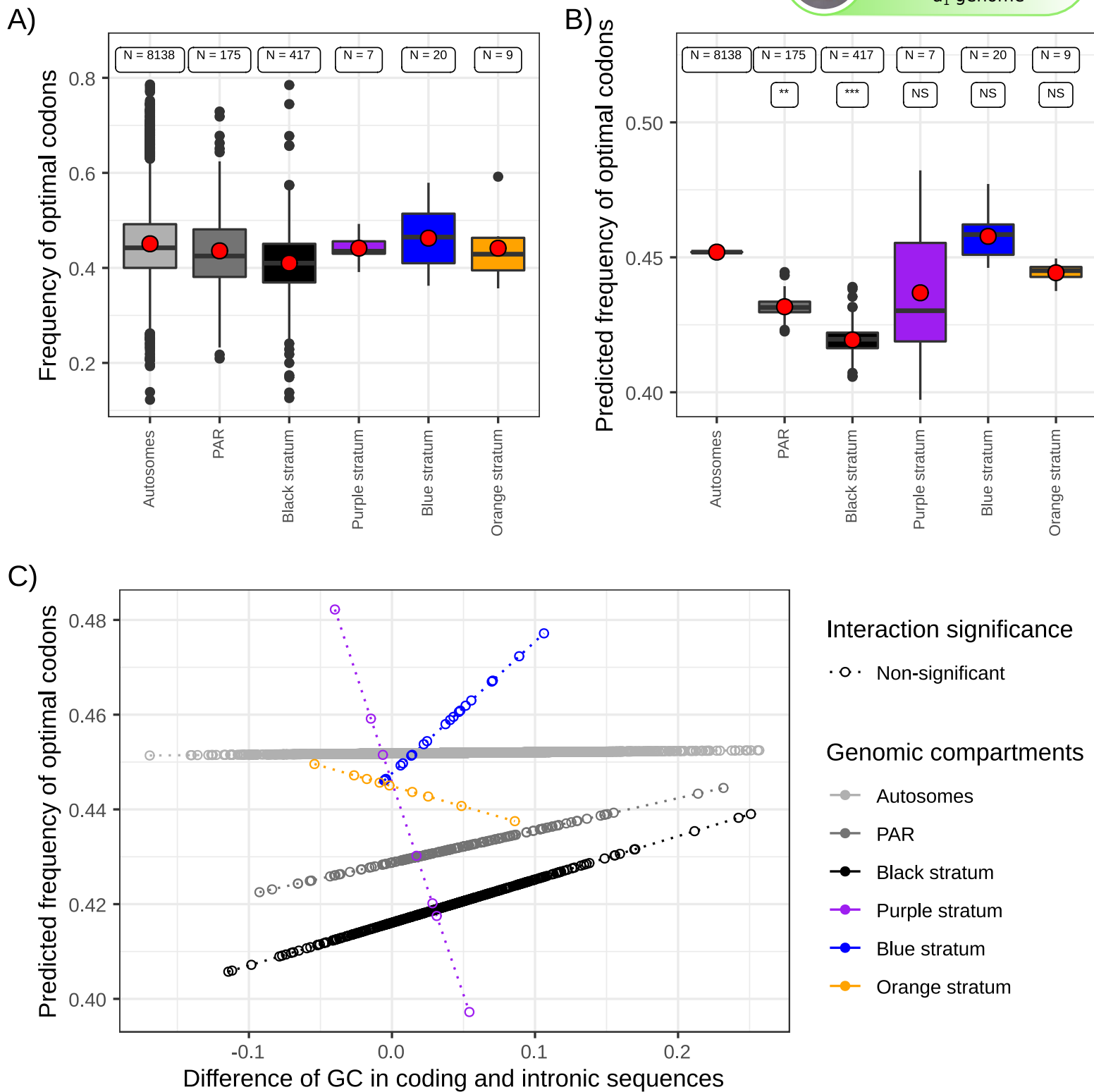






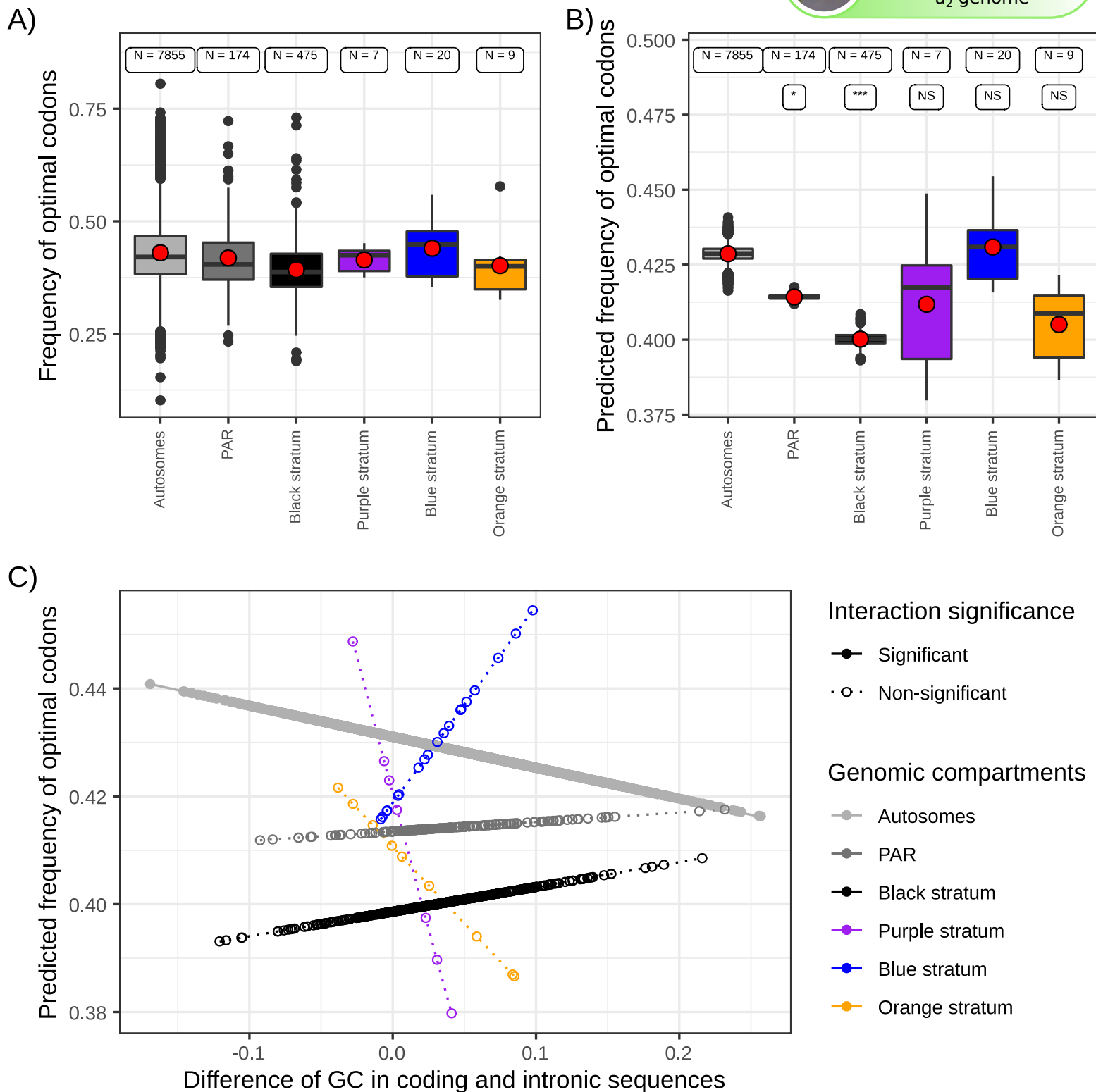


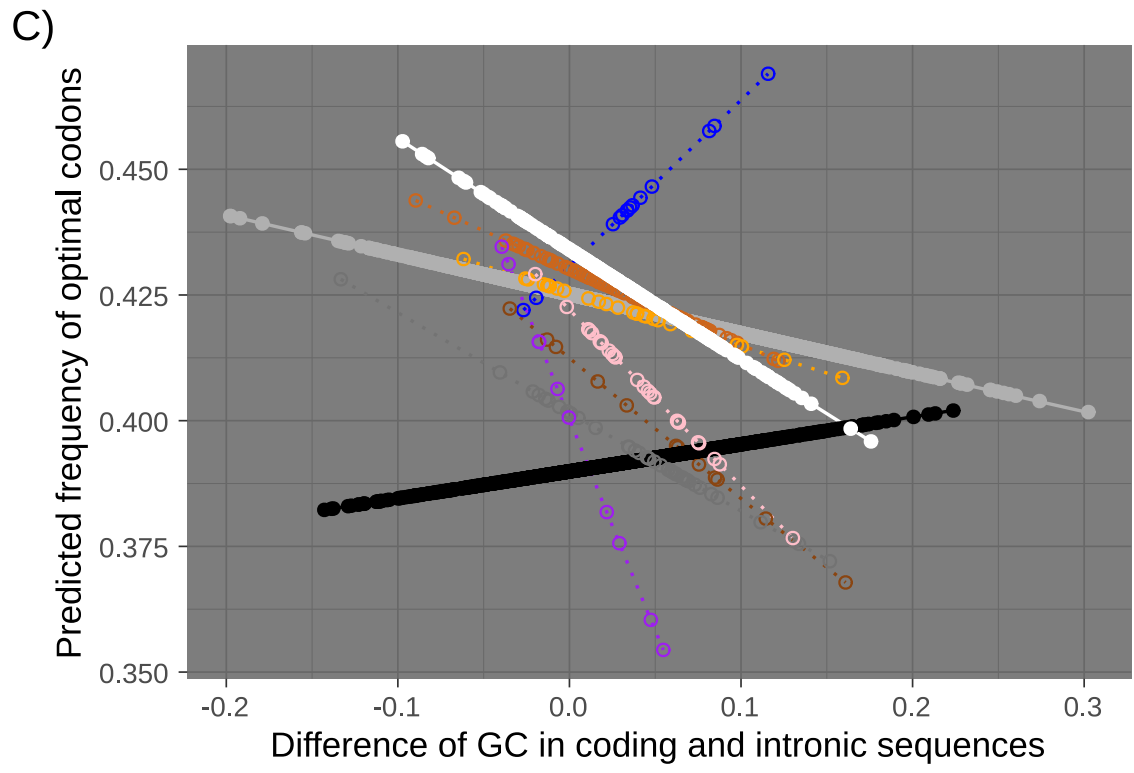
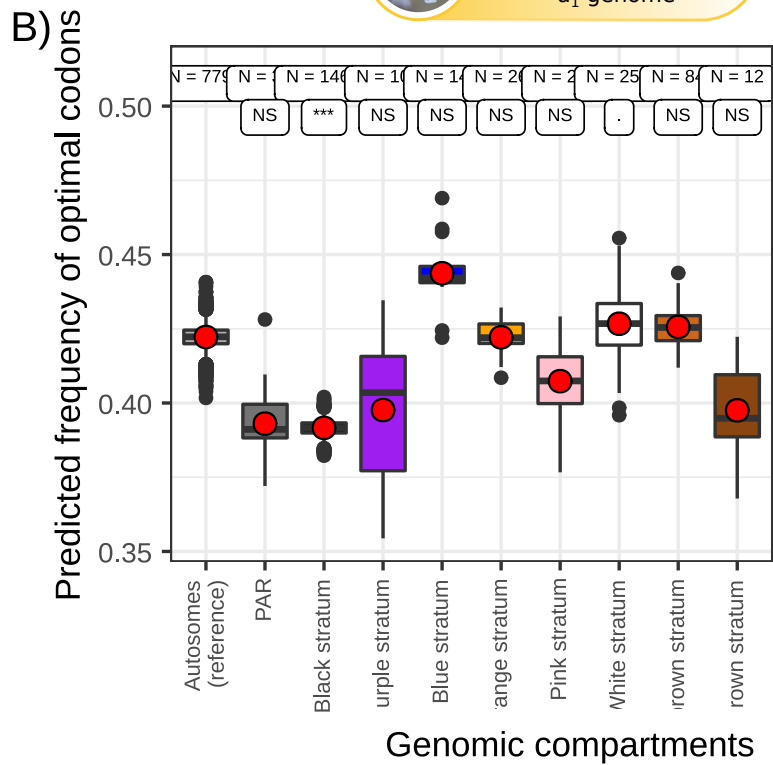
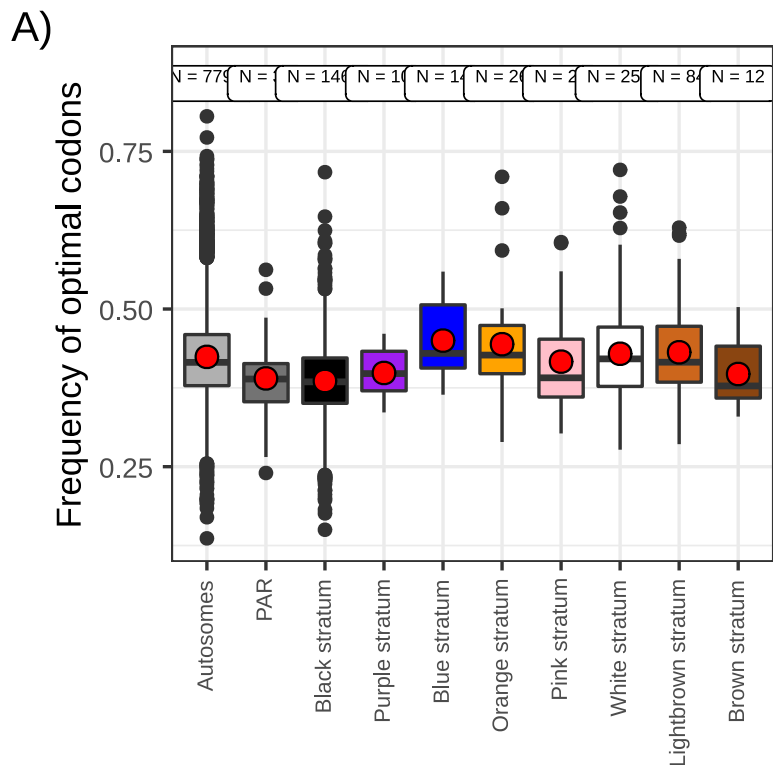
M. violaceum s.s.
on *Silene nutans*
a₁ genome





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



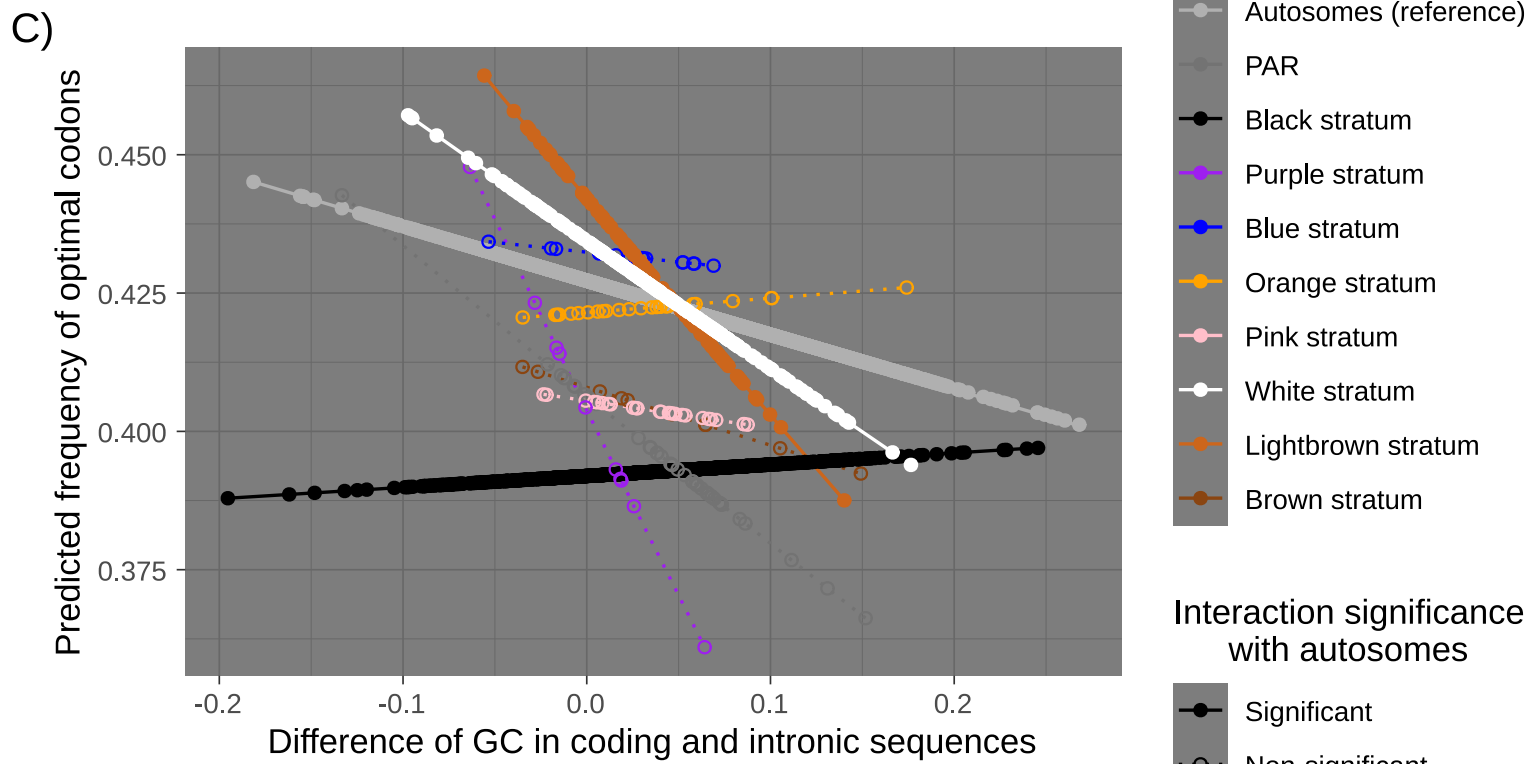
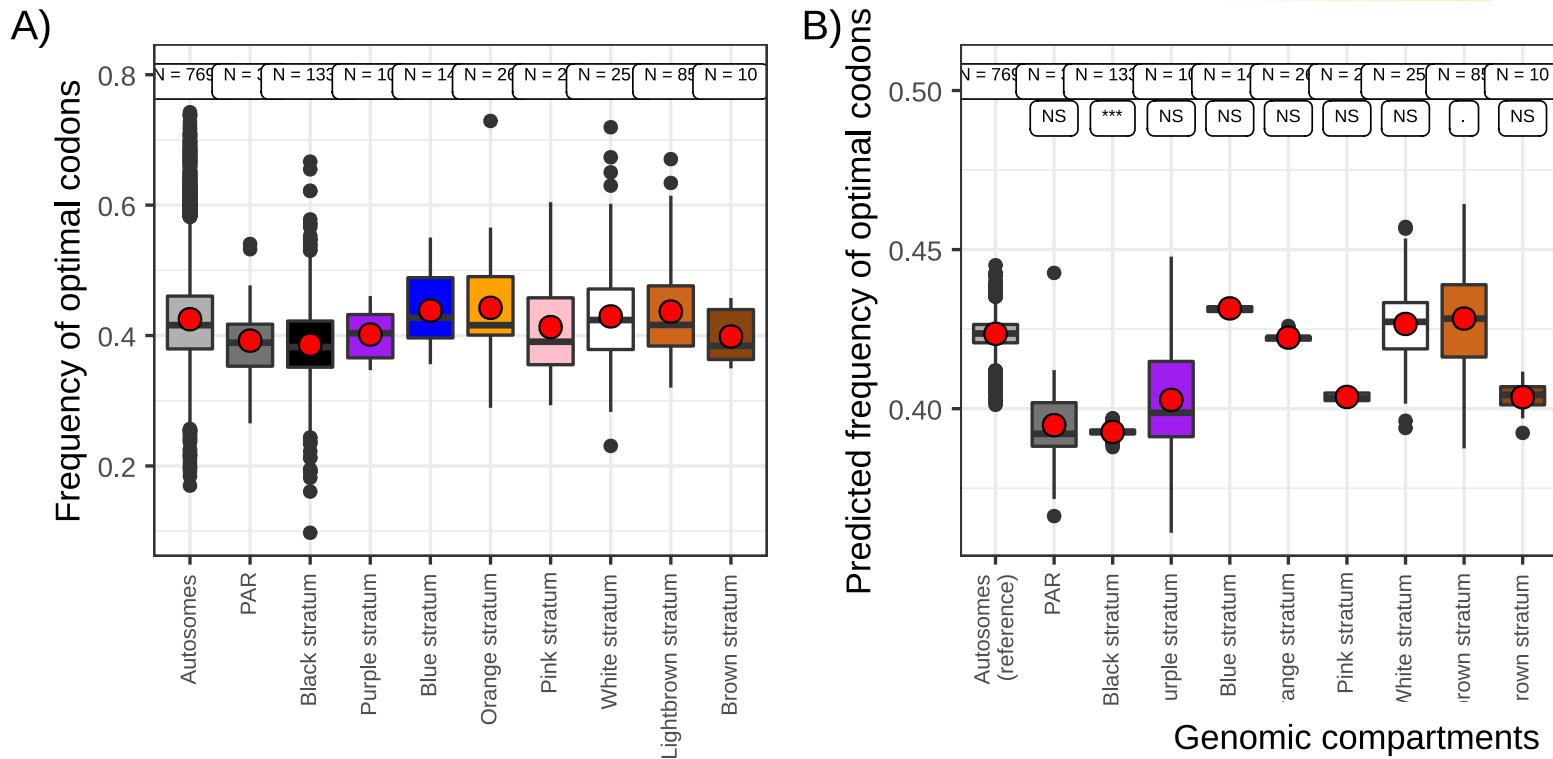


Genomic compartments

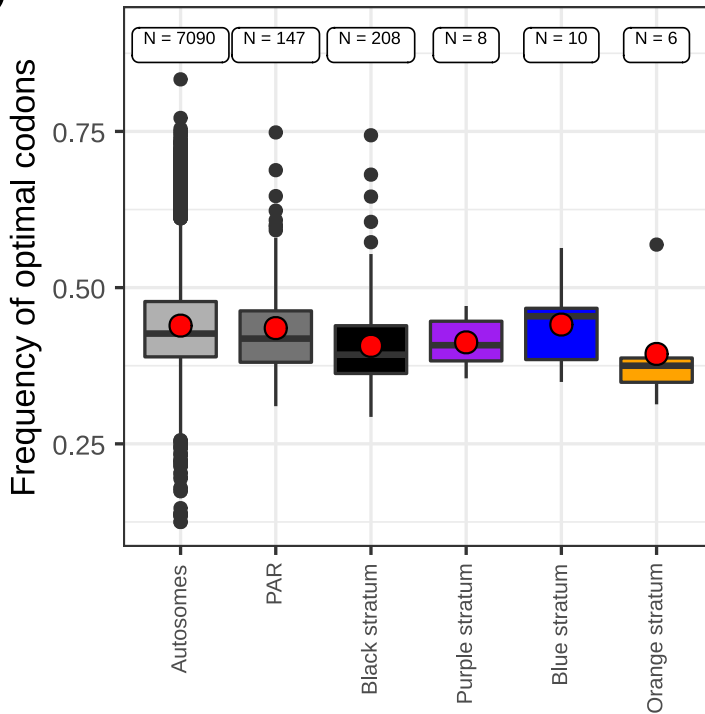
- Autosomes (reference)
- PAR
- Black stratum
- Purple stratum
- Blue stratum
- Orange stratum
- Pink stratum
- White stratum
- Lightbrown stratum
- Brown stratum

Interaction significance with autosomes

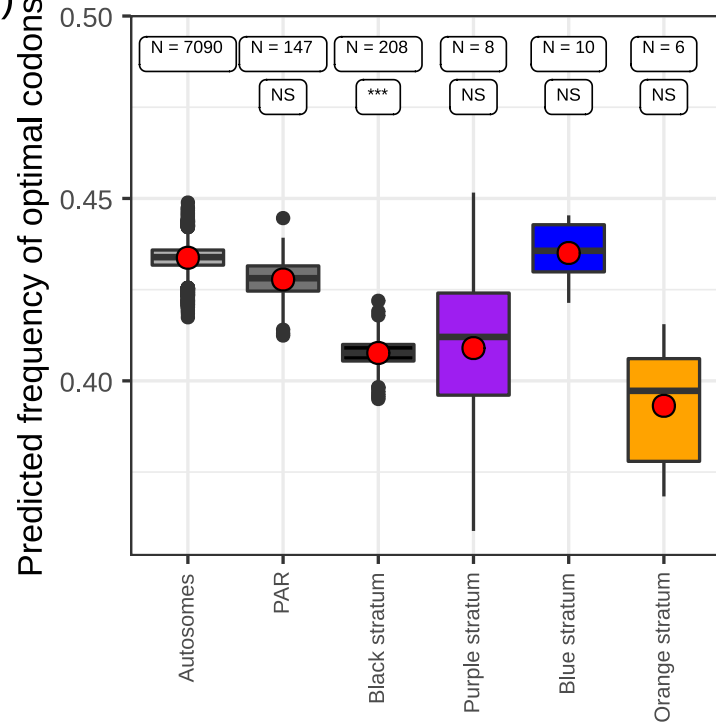
- Significant
- Non-significant



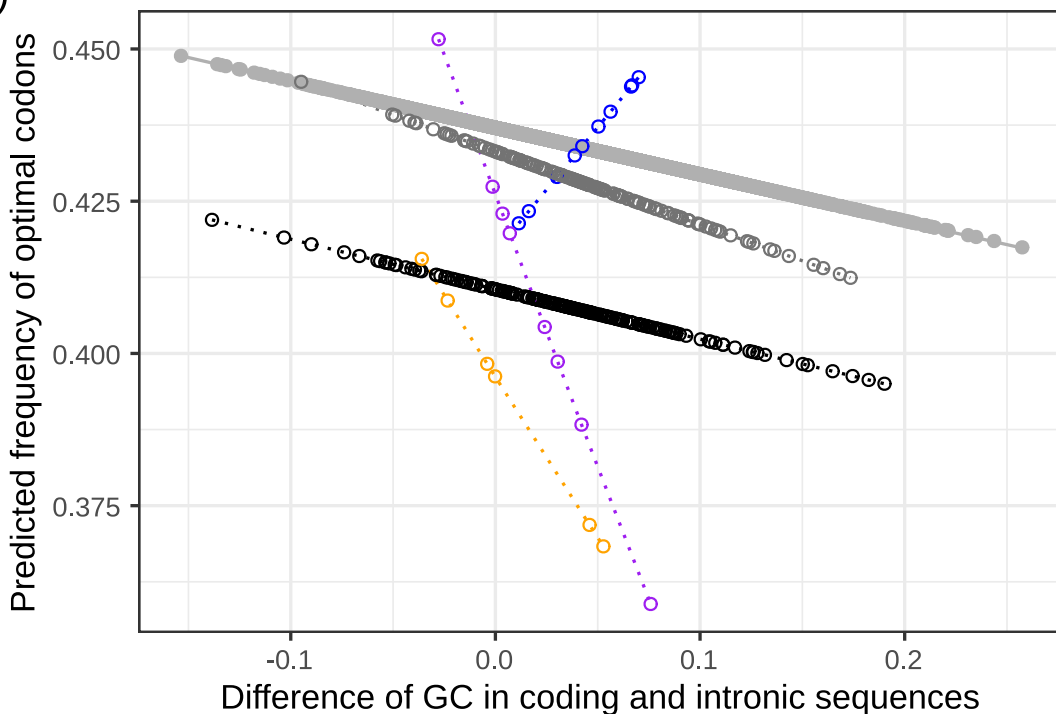
A)



B)



C)



Interaction significance

- Significant
- Non-significant

Genomic compartments

- Autosomes
- PAR
- Black stratum
- Purple stratum
- Blue stratum
- Orange stratum

