Tempo of degeneration across independently evolved non-recombining regions

Fantin Carpentier[1,2], Ricardo C. Rodríguez de la Vega[1], Michael H. Perlin[3], R. Margaret Wallen[3], Michael E. Hood#[4,] Tatiana Giraud#*[1]

# These authors co-supervised the study

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

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

[3] Department of Biology, Program on Disease Evolution, University of Louisville, Louisville, KY 40292, USA.

[4] Department of Biology, Amherst College, 01002-5000 Amherst, Massachusetts, United State of America.

***Corresponding author:** Tatiana Giraud

Laboratoire Ecologie, Systématique et Evolution, Bâtiment 360, Université Paris-Saclay, 91400 Orsay France

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

tatiana.giraud@universite-paris-saclay.fr

Running title:

Key words: synonymous codon usage, genomic degeneration, degeneration tempo, mating-type chromosomes, recombination suppression, frequency of optimal codon, fungi

Abstract

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

Introduction

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

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

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

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

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

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

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

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

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

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

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

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

Results

PR-HD linkage in three new *Microbotryum* genomes

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

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

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

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

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

Evolutionary strata

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

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

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

Absolute dating of recombination suppression events

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

Identification of optimal codons

To identify optimal codons in *Microbotryum* fungi, we investigated which codons were preferentially used in highly expressed autosomal genes in the two distantly related species, *M. lychnidis-dioicae* and *M. intermedium* (Figure 1). Selection for codon usage is expected to be strongest on most highly expressed genes and most effective on autosomes because recombination occurs. We had expression data for haploid cells cultivated clonally on nutritive medium (Table S4), for the two mating types separately. To study autosomal codon usage, we performed a within-group correspondence analysis (WCA; [70]), *i.e.*, a multivariate statistical method summarizing codon count data, normalizing the count of each codon by the average count of all the codons encoding for the same amino acid [71]. For the *M. intermedium* a_1 mating

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

Decrease in optimal codon odds in non-recombining regions

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

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

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

We found that the recombining regions located on mating-type chromosomes, i.e., PARs or on HD mating-type chromosome, had generally similar frequencies of optimal codons as autosomes (p-values > 0.05; Table S7a). The frequencies of optimal codons were, however, significantly lower in genes located in the recombining regions of the PR mating-type chromosome in *M. intermedium*, *M. lagerheimii* and *M. saponariae* and of the PAR of *M. violaceum* (*s.s.*), *M. v. scopulorum*, a_1 *M. coronarie* and a_2 *M. silenes-acaulis* (Table S7a). The frequencies of optimal codons in the young pink, white, light blue, brown evolutionary strata were not significantly different from autosomes in any species / mating type (p-value > 0.05; Table S7a).

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

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

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

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

Control for ancestral gene expression level

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

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

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

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

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

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

Discussion

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

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

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

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

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

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

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

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

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

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

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

Data access

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

Material and methods

Genome sequencing

We sequenced haploid genomes of opposite mating types for one diploid strain of each of the following species: *M. v. tatarinowii* parasitizing *S. tatarinowii* (#1400, GPS 39°59'45.4" 116°11'37.3", Xiangshan, Beijing, collected in September 2015), *M. v. melanantha* parasitizing *S. melanantha* (#1296; 27°01'50.6" 100°10'41.6" First Peak, Lijiang, date collected in July 2014), and *M. coronariae* parasitizing *L. flos cuculi* (#1247; GPS 55.7 -4.9, Great Cumbrae Island, UK, collected in 2000). Additionally, we sequenced the a_1 genome of *M. intermedium* parasitizing *Salvia pratensis* (#1389, GPS 44.3353 N 7.13637 E, Cuneo, Italy, collected in July 2011). Samples were collected before the publication of laws regarding the Nagoya protocol in the countries of collection. DNA extraction and sequencing based on Pacific Bioscience long-read sequencing was performed as described previously [60, 61]. Sequencing was carried out at UCSD IGM Genomics Center (San Diego, USA) with the P6 C4 chemistry. Coverage was between 24x and 35x for all genomes, except for *M. intermedium* for which it was 199x.

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

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

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

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

Gene models, orthologous group reconstruction, transposable elements, centromeres

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

Evolutionary scenarios of the mating-type chromosomes

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

RNAseq experiment and expression level quantification.

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

Agar (PDA) and grown for 48 hours at 22°C. Cells were scraped, ground in liquid nitrogen, and total RNA extracted following the manufacturer's protocol. For RNAseq analysis, equal amounts of total RNA individually isolated from a_1 and a_2 cultures were pooled. For RNA isolation from mated pairs, haploid a_1 and a_2 cultures were first grown separately in yeast extract peptone dextrose broth overnight. Cell density was measured spectrophotometrically. The concentration of each culture was adjusted to $O.D._{600}$ 1.0. Equal volumes of the two haploid cultures were mixed and plated on water agar. These plates were incubated for four days at 14°C, after which wet mounts were prepared from each mated plate to verify the presence of conjugation tubes, indicating active mating behaviour. Cells were scraped and total RNA harvested as previously described. After total RNA isolation, several quality control measures were taken. Concentration and purity were assessed using a NanoDrop 2000 spectrophotometer (Thermo Scientific); 260/280 and 260/230 ratios greater than 1.8 were considered satisfactory for RNAseq application. Additionally, cDNA was prepared and used as template for intron-spanning primers in PCR reactions to verify the lack of genomic DNA contamination. Bioanalyzer analysis was completed to detect intact 18S and 23S RNA as a measure of overall RNA quality. After passing all quality control measurements, RNA was sent for RNAseq analysis to CD Genomics (Shirley, New York).

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

Identifying optimal codons in *Microbotryum* fungi

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

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

Frequency of optimal codons and GC content statistics

For each coding sequence with base pair numbers multiple of three, we calculated the frequency of optimal codons (FOP), the GC content on the third position ($\%GC_3$) of each codon and the GC content of the introns of each gene ($\%GC_{introns}$). We removed from this gene set any gene overlapping with a TE. While parsing the coding sequence fasta files for the frequency of

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

Non-synonymous and synonymous divergence estimation

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

Absolute dating of recombination suppressions

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

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

caroliniana in the trees used for estimating the ages of the black stratum of *M. silenes-acaulis* (21 genes in common) and the pink stratum of *M. v. paradoxa* (10 genes in common). Absolute dates for the non-recombining regions in *M. lagerheimii* and *M. saponariae* were taken from a previous study [62]. Gene family alignments are available as supplementary material.

Statistics

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

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

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

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

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

Author contribution

Conception and design, TG, MEH, FC; Formal analysis, FC, RCRV; Interpretation, FC, TG, MEH, RCRV; Data acquisition, MHP, RMW; Original draft, FC, TG; Final draft, FC, TG, MEH, RCRV; Revision, FC, TG, RCRV, MEH, MHP, RMW. All authors read and approved the final version.

Competing interests

None of the authors have any competing interests.

Acknowledgements

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

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

Cited literature

1. Fisher RA: *The genetical theory of natural selection.* 1930.
2. Muller HJ: **Some genetic aspects of sex.** *The American Naturalist* 1932, **66:**118-138.
3. Bernstein H, Byers GS, Michod RE: **Evolution of sexual reproduction: importance of DNA repair, complementation, and variation.** *Am Nat* 1981, **117:**537-549.
4. Otto SP, Lenormand T: **Resolving the paradox of sex and recombination.** *Nature Reviews Genetics* 2002, **3:**252-261.
5. Hill WG, Robertson A: **The effect of linkage on limits to artificial selection.** *Genetics Research* 1966, **8:**269-294.
6. Maynard Smith J, Haigh J: **The hitch-hiking effect of a favourable gene.** *Genet Res* 1974, **23:**22-35.
7. Gutiérrez-Valencia J, Hughes PW, Berdan EL, Slotte T: **The genomic architecture and evolutionary fates of supergenes.** *Genome biology and evolution* 2021, **13**.
8. Schwander T, Libbrecht R, Keller L: **Supergenes and complex phenotypes.** *Current Biology* 2014, **24:**R288-R294.
9. Nadeau NJ: **Genes controlling mimetic colour pattern variation in butterflies.** *Current Opinion in Insect Science* 2016, **17:**24-31.
10. Saenko VS, Chouteau M, Prunier FP, Blugeon C, Joron M: **Unravelling the genes forming the wing pattern supergene in the polymorphic butterfly *Heliconius numata*.** *EvoDevo* 2019, **10:**16.
11. Wang J, Wurm Y, Nipitwattanaphon M, Riba-grognuz O, Huang Y-c, Shoemaker D, Keller L: **A Y-like social chromosome causes alternative colony organization in fire ants.** *Nature* 2013, **493:**664.
12. Charlesworth B, Charlesworth D: **A model for the evolution of dioecy and gynodioecy.** *The American Naturalist* 1978, **112:**975-997.
13. Hartmann FE, Duhamel M, Carpentier F, Hood ME, Foulongne-Oriol M, Silar P, Malagnac F, Grognet P, Giraud T: **Recombination suppression and evolutionary strata around mating-type loci in fungi: documenting patterns and understanding evolutionary and mechanistic causes.** *New Phytologist* 2021, **229:**2470-2491.
14. Ohno S: *Sex chromosomes and sex-linked genes.* Springer-V edn; 1967.
15. Westergaard M: **The mechanism of sex determination in dioecious flowering plants.** *Adv Genet* 1958, **9:**217-281.
16. Bachtrog D: **Y-chromosome evolution : emerging insights into processes of Y-chromosome degeneration.** *Nature Reviews Genetics* 2013, **14:**113-124.
17. Charlesworth B, Charlesworth D: **The degeneration of Y chromosomes.** *Philosophical Transactions of the Royal Society B-Biological Sciences* 2000, **355:**1563-1572.
18. Jay P, Chouteau M, Whibley A, Bastide H, Parrinello H, Llaurens V, Joron M: **Mutation load at a mimicry supergene sheds new light on the evolution of inversion polymorphisms.** *Nature 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, Aguileta 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, Zaghlul 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-Kawaguchl T, Minx PJ, Cordum HS, Hlllier LD, Brown LG, Repplng 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, Aguileta 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, Aguileta 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*.**
1040 *Molecular Biology and Evolution* 2011, **28:**771-780.
1041 79. Chibalina MV, Filatov DA: **Plant y chromosome degeneration is retarded by haploid purifying**
1042 **selection.** *Current Biology* 2011, **21:**1475-1479.
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: **OrthAgogue: 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.

100. Gladieux P, Vercken E, Fontaine MC, Hood ME, Jonot O, Couloux A, Giraud T: **Maintenance of fungal pathogen species that are specialized to different hosts: allopatric divergence and introgression through secondary contact.** *Molecular Biology and Evolution* 2011, **28:**459-471.
101. Ranwez V, Douzery EJP, Cambon C, Chantret N, Delsuc F: **MACSE v2: Toolkit for the Alignment of Coding Sequences Accounting for Frameshifts and Stop Codons.** *Molecular Biology and Evolution* 2018, **35:**2582-2584.
102. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R: **IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era.** *Molecular Biology and Evolution* 2020, **37:**1530-1534.

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

Figures Legends

Figure 1: Timetrees and rearrangement scenarios of recombination suppression events in the *Microbotryum* genus. A) Densitree plot of 1,000 random trees with the best posterior probabilities based on the concatenation of single-copy autosomal genes (green traces) and single-copy genes that are located in the non-recombining regions (black traces) of all species

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

Figure 2: Allelic divergence analysis between mating-type chromosomes for the newly sequenced species. For each species (*Microbotryum coronariae*, *M. v. melanantha*, *M. v. tatarinowii* and *M. lychnidis-dioicae)*, we show the synonymous divergence calculated between alternative mating types and plotted along the *M. intermedium* mating-type chromosome gene order. Zero d_S between mating-type chromosomes indicate recombination in these selfing species and the level of d_S between mating-type chromosomes increases with time since recombination suppression. Different evolutionary strata, i.e., different steps extending recombination suppression, are indicated by different colours. Divergence between the a_1 and

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

Figure 3: Analysis of the frequency of optimal codons in the a_1 genome of *Microbotryum silenes-dioicae*. A) Distribution of the frequency of optimal codons across genomic compartments. B) Distribution of the frequency of optimal codons predicted by the logistic regression, accounting for genomic compartment, difference of GC in coding and intronic sequences and their interactions as factor. In A) and B), boxplot grey shade refers to the genomic compartment. For each genomic compartment, the red dot indicates the mean frequency of optimal codons, the sample size (N) is labelled on top of the corresponding boxplot, as well as the significance level of the difference between a given genomic compartment and the autosomes (NS: non-significant; ***: p-value < 0.001) from the logistic regression. Note the difference in the ranges of Y axes for panels A and B. Similar analyses for all species studied here are presented in Figure S6 and S7 in the supplementary materials. C) Frequency of optimal codons predicted by the logistic regression plotted against the difference in GC content between coding and intronic sequences. From the logistic regression, significant interactions between a genomic compartment and the autosomes are indicated by solid line and filled dots, while non-significant interactions are indicated by dotted line and hollow dots. "Reference" refers to the reference considered in the logistic regression. D) Predicted effect of GC-biased gene conversion and selection on frequency of optimal codons. The figure shows the interpretation of the absence of interaction (2) or the presence of a positive (3) or negative interaction (4) between the difference in GC content between coding sequences and introns and the genomic compartments regarding the logistic regression performed in this study. The frequency of optimal codons in autosomes (1) decreases with the increase of difference in GC content and the frequency of optimal codons are lower in non-recombining regions (2, 3, 4).

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

Additional files:

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

2-All supplementary Figures as a single PDF file

Supplemental Figure legends

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

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

Figure S3: **Synteny and allelic divergence analysis between mating-type chromosomes of alternative mating types for the species *M. scabiosae* (A) and *Microbotryum v. caroliniana* (B).** For each species, we show synteny plots between a_1 and a_2 mating-type chromosomes (left top panel), synteny plots between a_1 and a_2 mating-type chromosomes of *M. intermedium* (right top panel), and the synonymous divergence calculated between alternative mating types and plotted along the *M. intermedium* mating-type chromosome gene order (bottom panel). Zero d_S between mating-type chromosomes indicate recombination in these selfing species and the level of d_S between mating-type chromosomes increases with time since recombination suppression. On the top left figure in B), the light blue evolutionary stratum is indicated by a light blue box on the outer track.

Figure S4: Allelic divergence analysis between mating-type chromosomes of alternative mating types for *Microbotryum v. paradoxa*. The synonymous divergence calculated between alternative mating types is plotted along the *M. intermedium* mating-type chromosome ancestral-like gene order. Zero d_S between mating-type chromosomes indicate recombination in these selfing species and the level of d_S between mating-type chromosomes increases with time since recombination suppression. Different evolutionary strata, i.e., different steps extending recombination suppression, are indicated by different colours. Divergence between the a_1 and a_2 pheromone receptor (PR) genes was too extensive, and d_S could not be calculated (depicted as "Un" for unalignable).

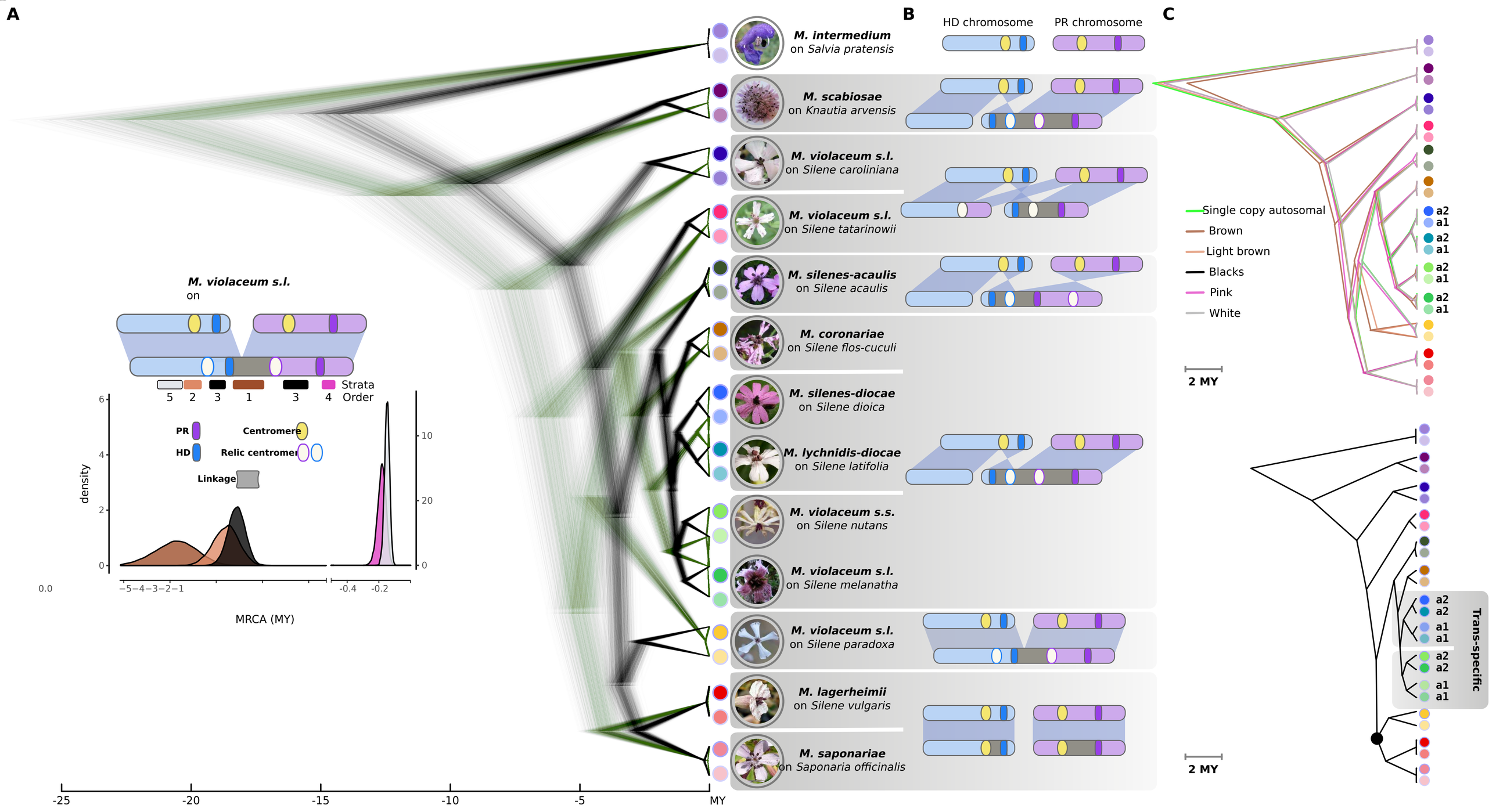
Figure S5: Within-correspondence analysis plots. A) WCA plots for the a_1 and a_2 genomes of *Microbotryum intermedium* (resp. left and right top panel) and a_1 and a_2 genomes of *M. lychnidis-dioicae* (resp. left and right bottom panel). B) Percentage of variance explained for each correspondence analysis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

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

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

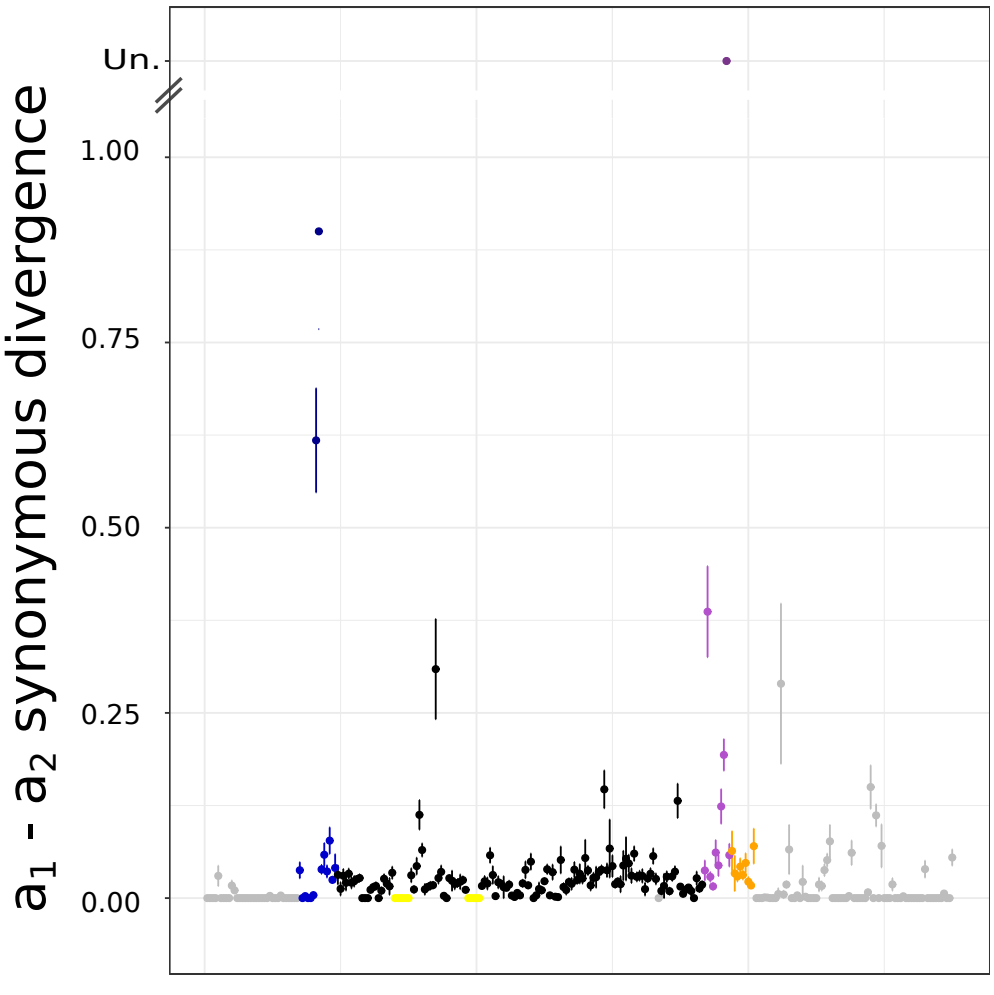
compartment and the autosomes (NS: non-significant; ".": p-value < 0.1; *: p-value < 0.05; **: p-value < 0.01; ***: p-value < 0.001). C) Frequency of optimal codons predicted by the logistic regression along the difference of GC content between coding and intronic sequences. Significant interactions between a genomic compartment and the autosomes are indicated by a solid line and a filled dot, while non-significant interactions are indicated by a dotted line and a hollow dot.

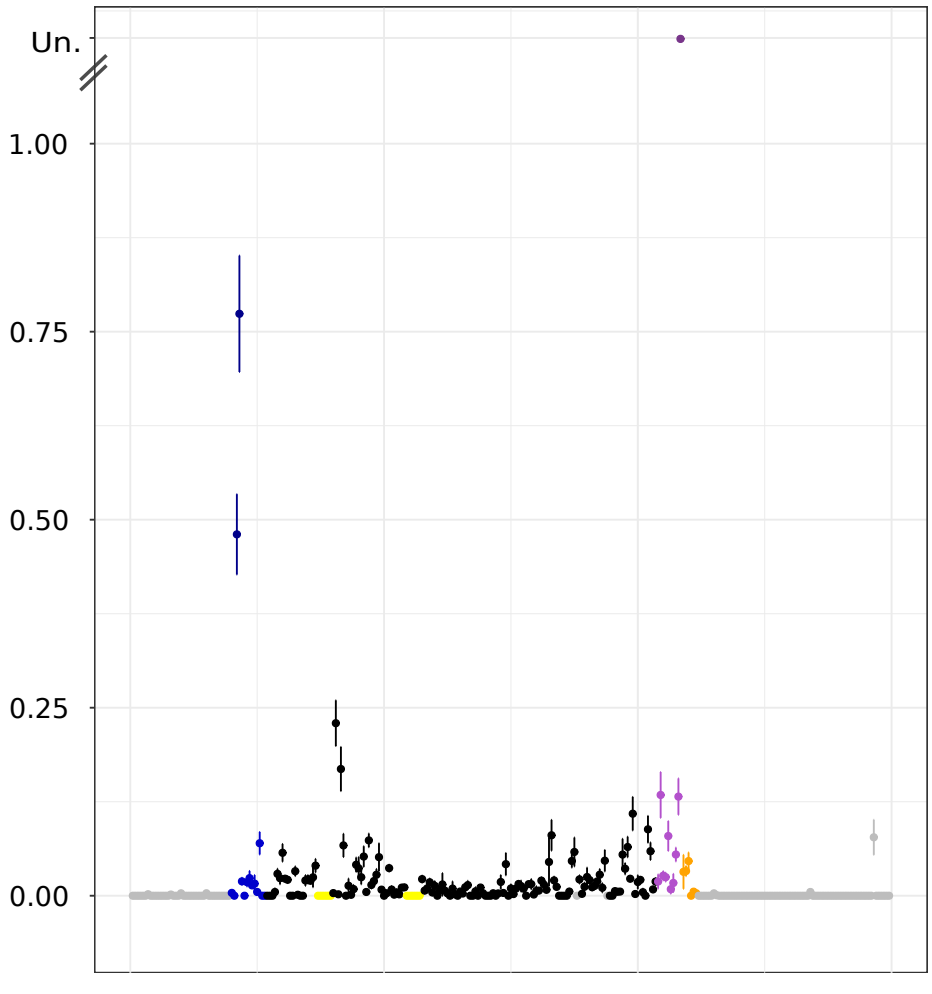
Figure S8: Analysis of the frequency of optimal codons in all *Microbotryum* genomes using all the genomic compartments. A) Distribution of the frequency of optimal codons across genomic compartments. B) Distribution of the predicted frequency of optimal codons predicted by the logistic regression. In A) and B), boxplot colours refer to the genomic compartment. For each genomic compartment, the red dot indicates the mean frequency of optimal codons, the sample size (N) is labelled on top of the corresponding boxplot, as well as the significance level of the difference between a given genomic compartment and the autosomes (NS: non-significant; ".": p-value < 0.1; *: p-value < 0.05; **: p-value < 0.01; ***: p-value < 0.001). C) Frequency of optimal codons predicted by the logistic regression along the difference of GC content between coding and intronic sequences. Significant interactions between a genomic compartment and the autosomes are indicated by a solid line and a filled dot, while non-significant interactions are indicated by a dotted line and a hollow dot.

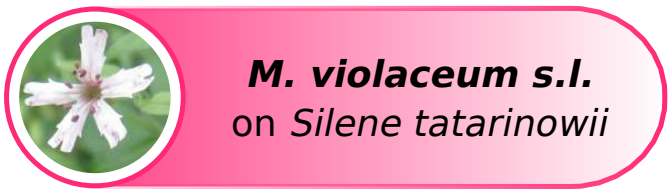

A
B
C
HD chromosome
PR chromosome
M. intermedium
on Salvia pratensis
M. scabiosae
on Knautia arvensis
M. violaceum s.l.
on Silene caroliniana
M. violaceum s.l.
on Silene tatarinowii
M. silenes-acaulis
on Silene acaulis
M. coronariae
on Silene flos-cuculi
M. silenes-diocae
on Silene dioica
M. lychnidis-diocae
on Silene latifolia
M. violaceum s.s.
on Silene nutans
M. violaceum s.l.
on Silene melanatha
M. violaceum s.l.
on Silene paradoxa
M. lagerheimii
on Silene vulgaris
M. saponariae
on Saponaria officinalis
-25
-20
-15
-10
-5
MY
M. violaceum s.l.
on
Strata
5 2 3 1 3 4
Order
PR
HD
Centromere
Relic centromere
Linkage
density
MRCA (MY)
Single copy autosomal
Brown
Light brown
Blacks
Pink
White
2 MY
a2
a1
Trans-specific

M. coronariae
on *Silene flos-cuculi*

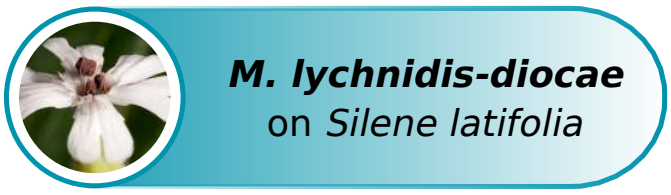
M. violaceum s.l.
on *Silene melanatha*



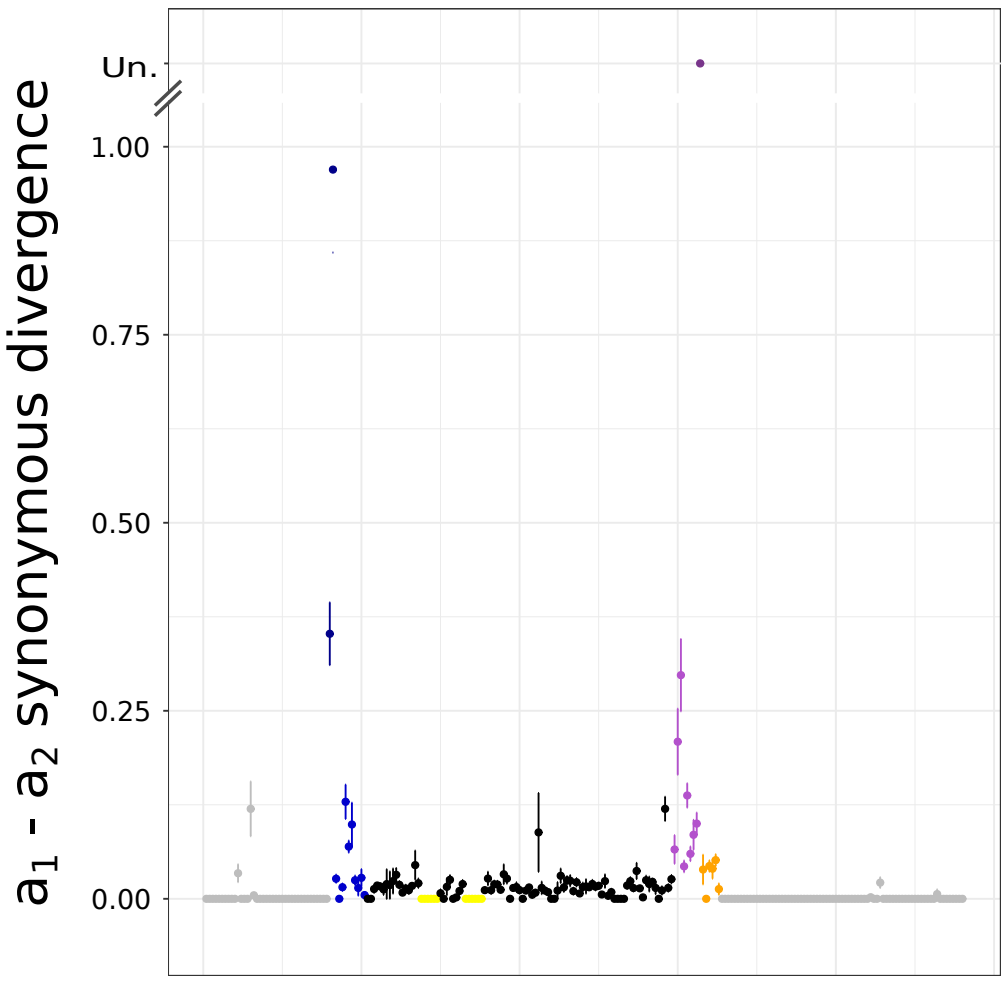


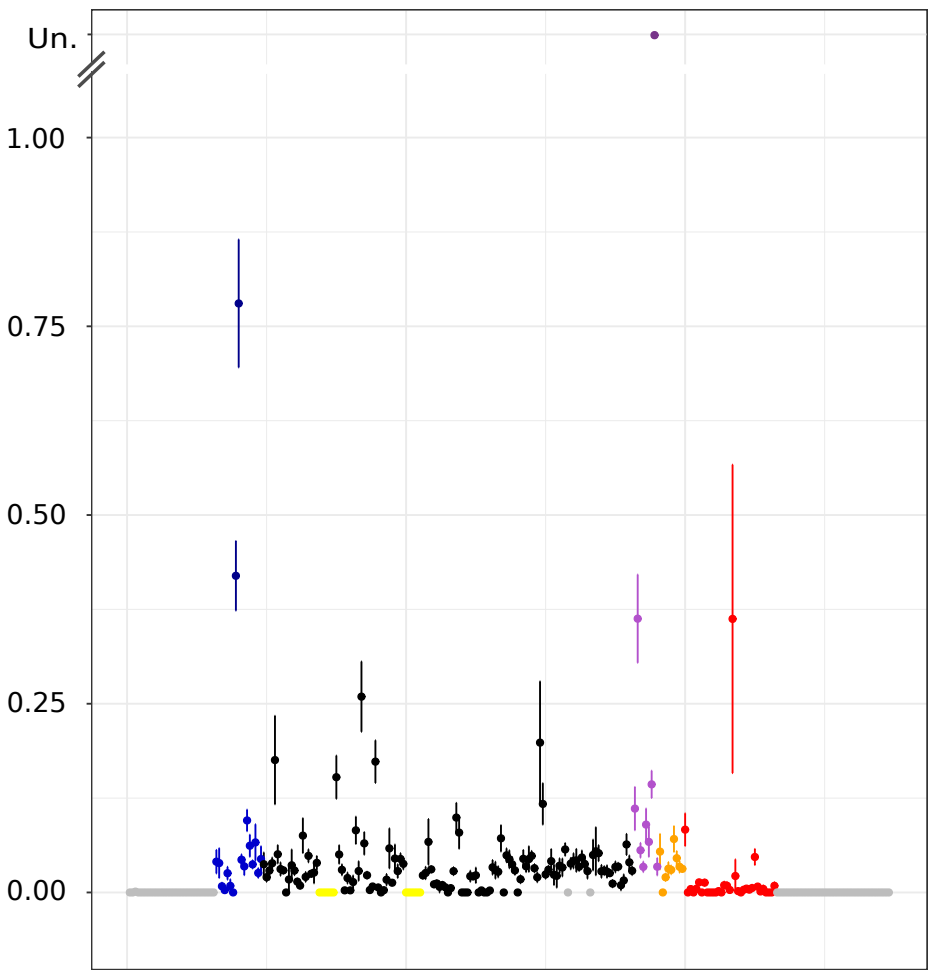


M. violaceum s.l.
on *Silene tatarinowii*

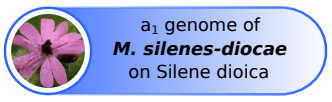


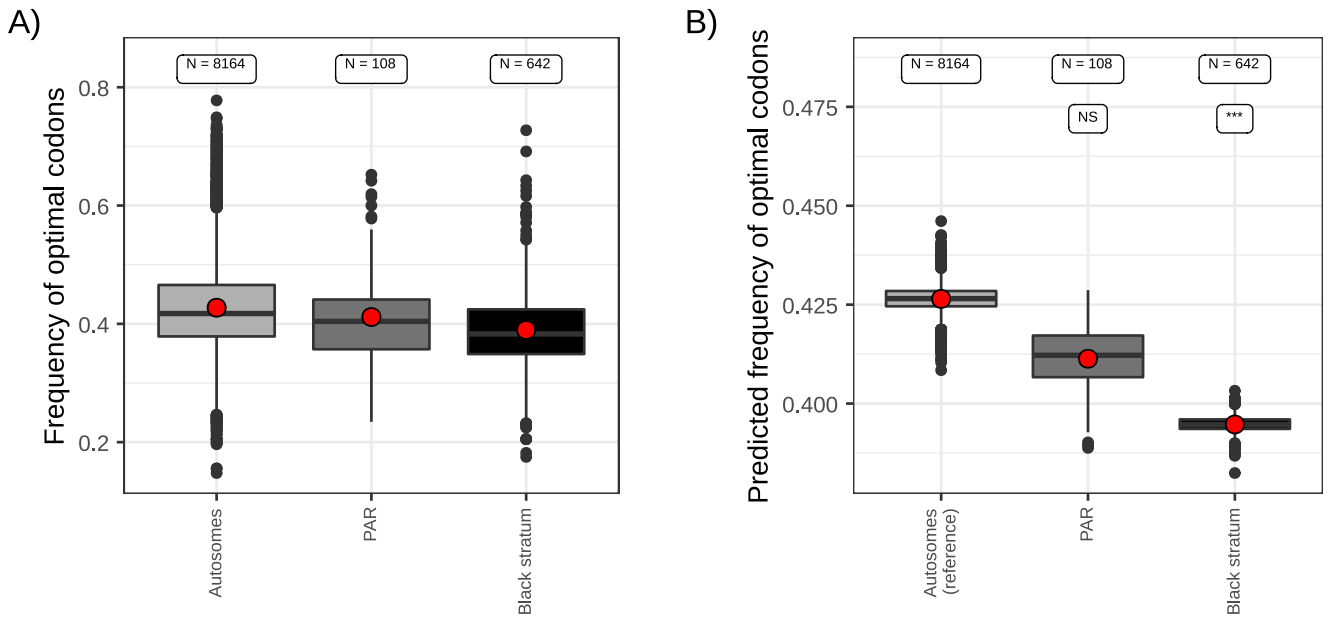
M. lychnidis-diocae
on *Silene latifolia*

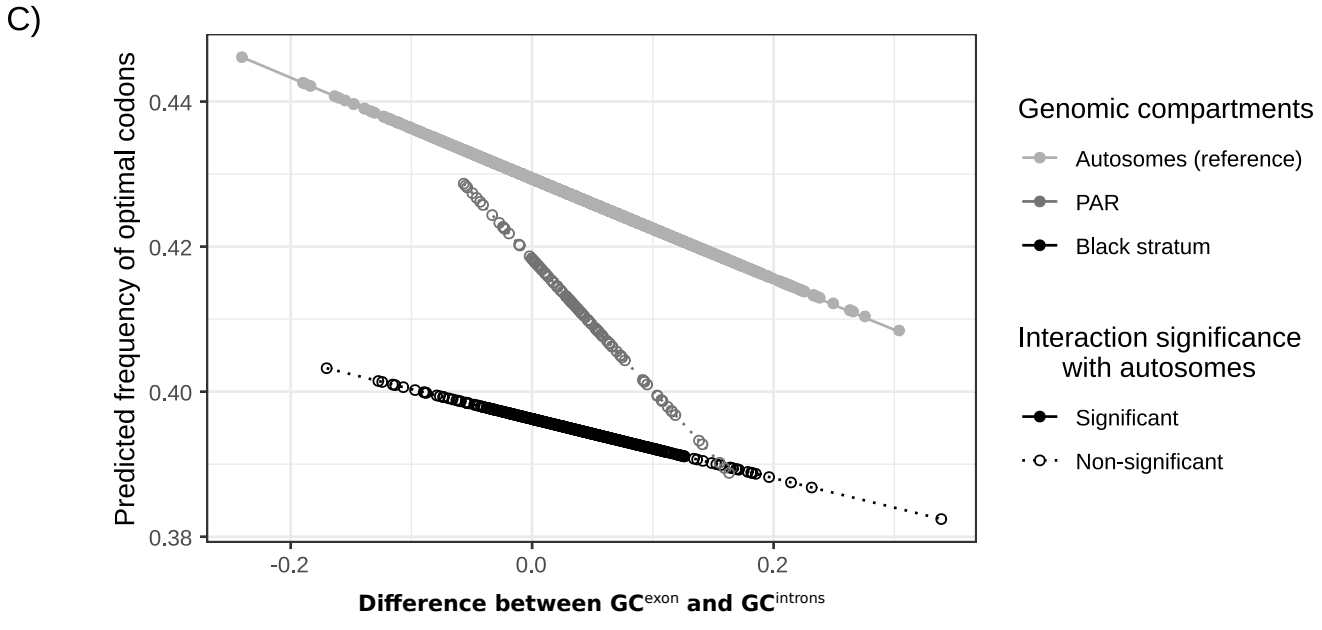


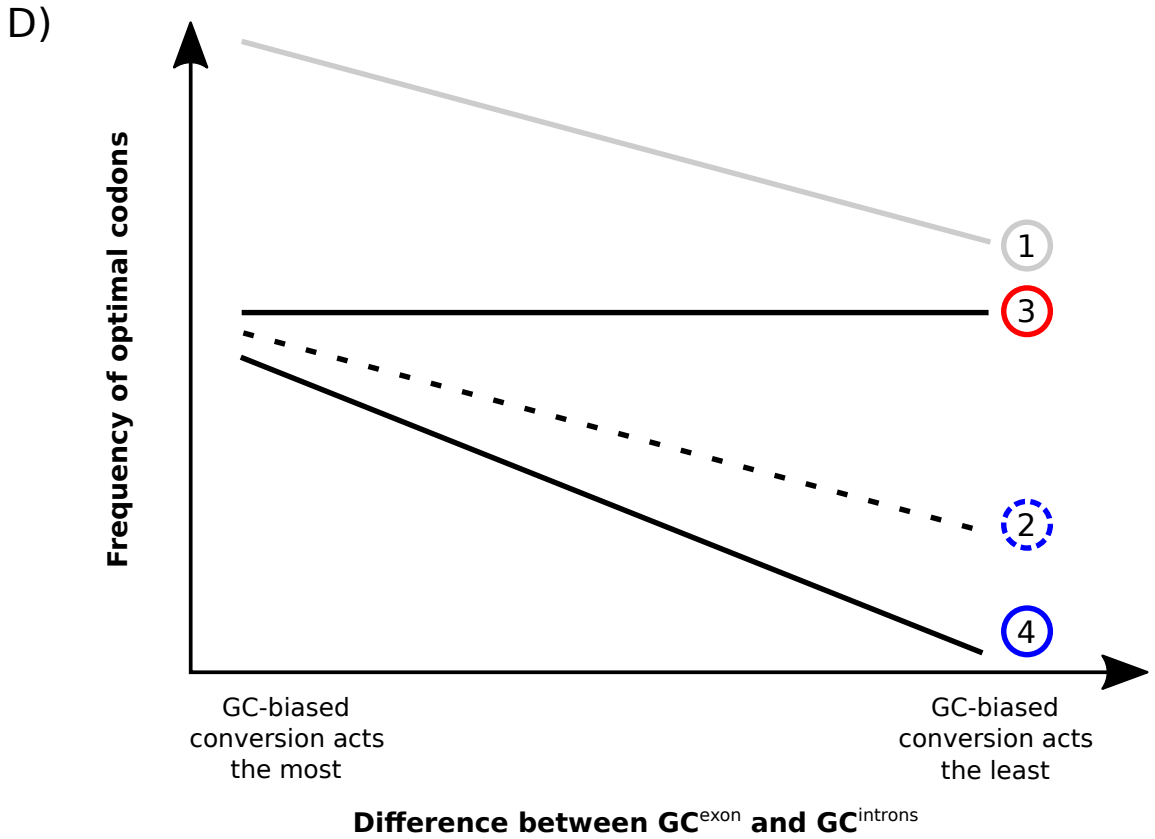


Gene order along the mating-type chromosomes of *M. intermedium*









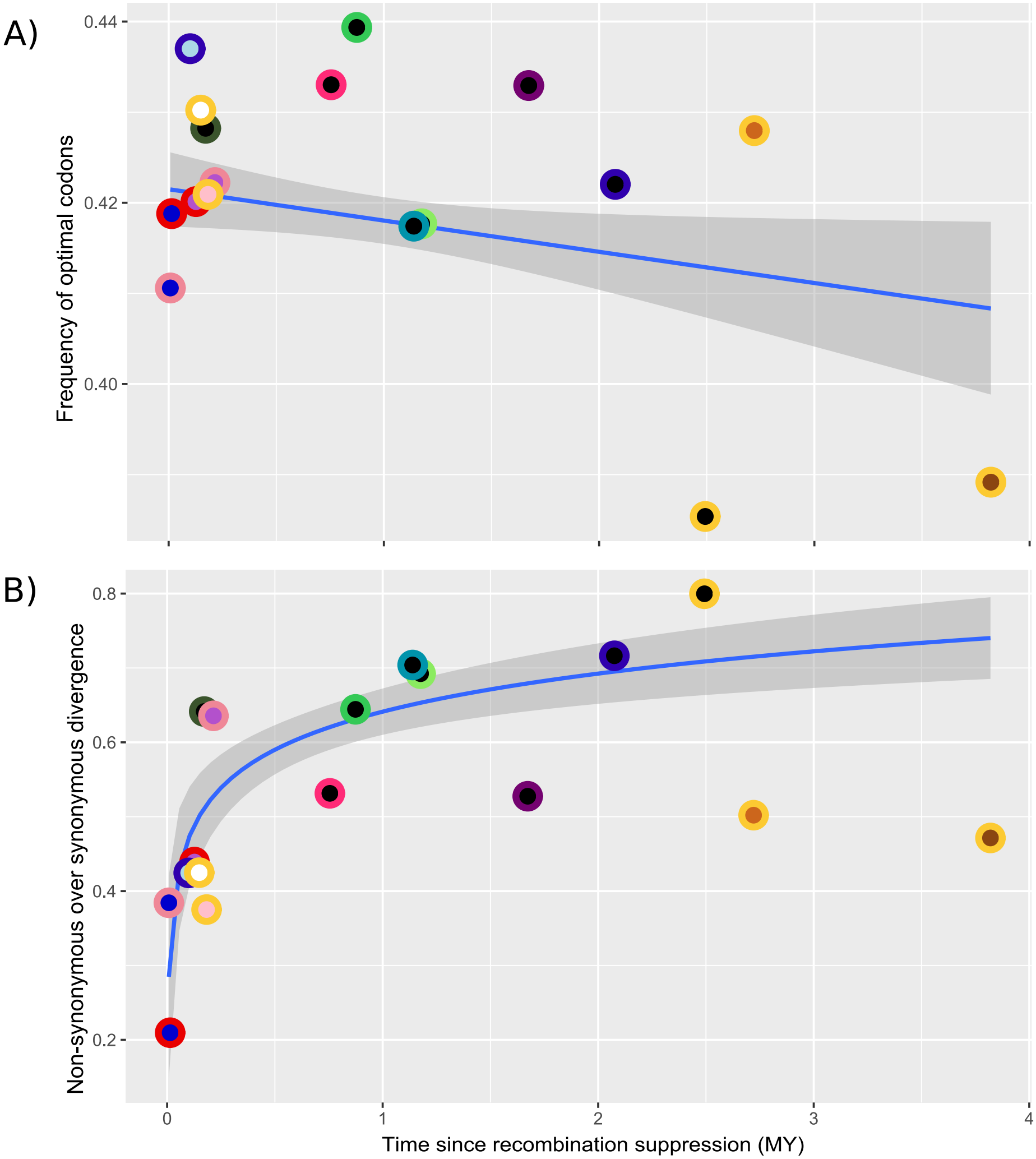
1 **In autosomes**, the reduction of GC-biased gene conversion results in a decrease in frequency of optimal codons.

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.



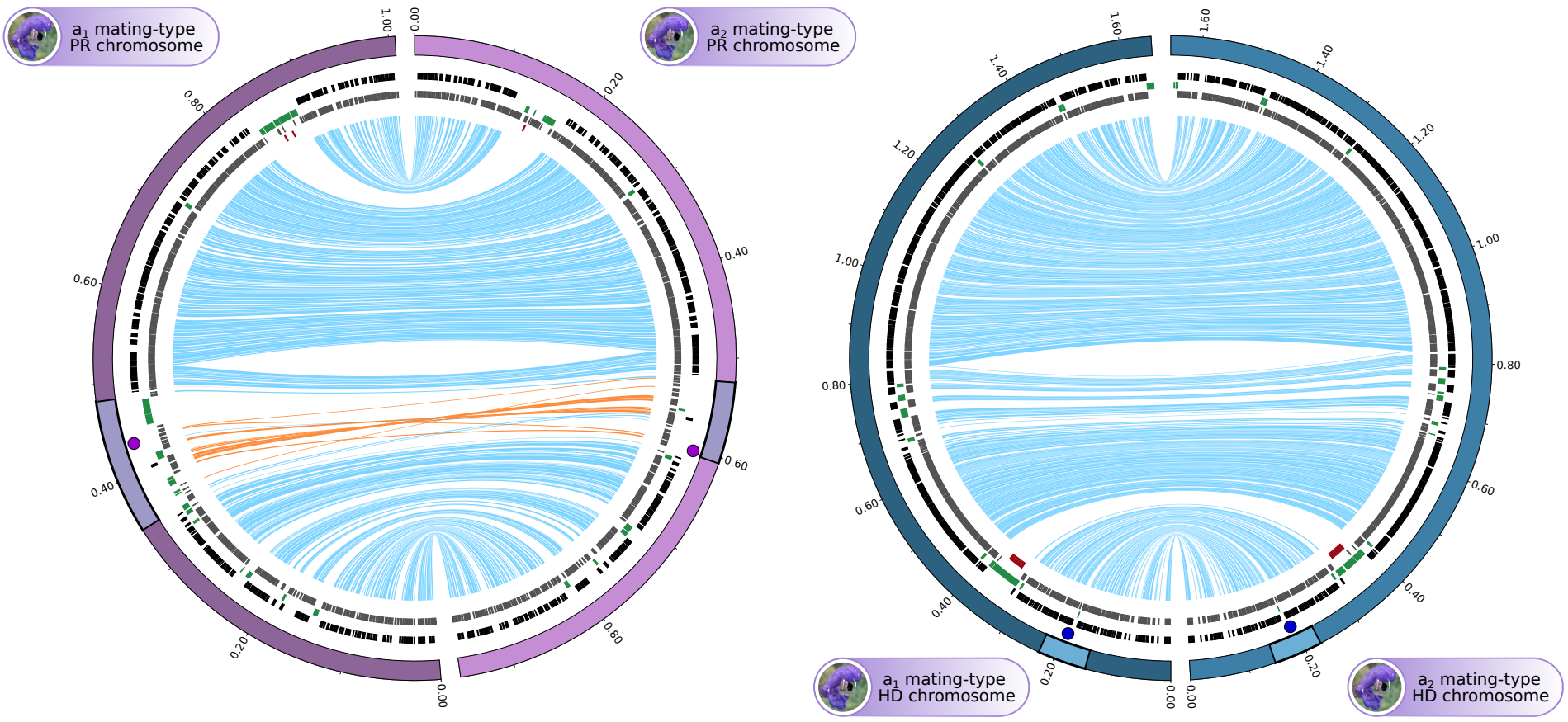
Legend

- 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*



Key to tracks

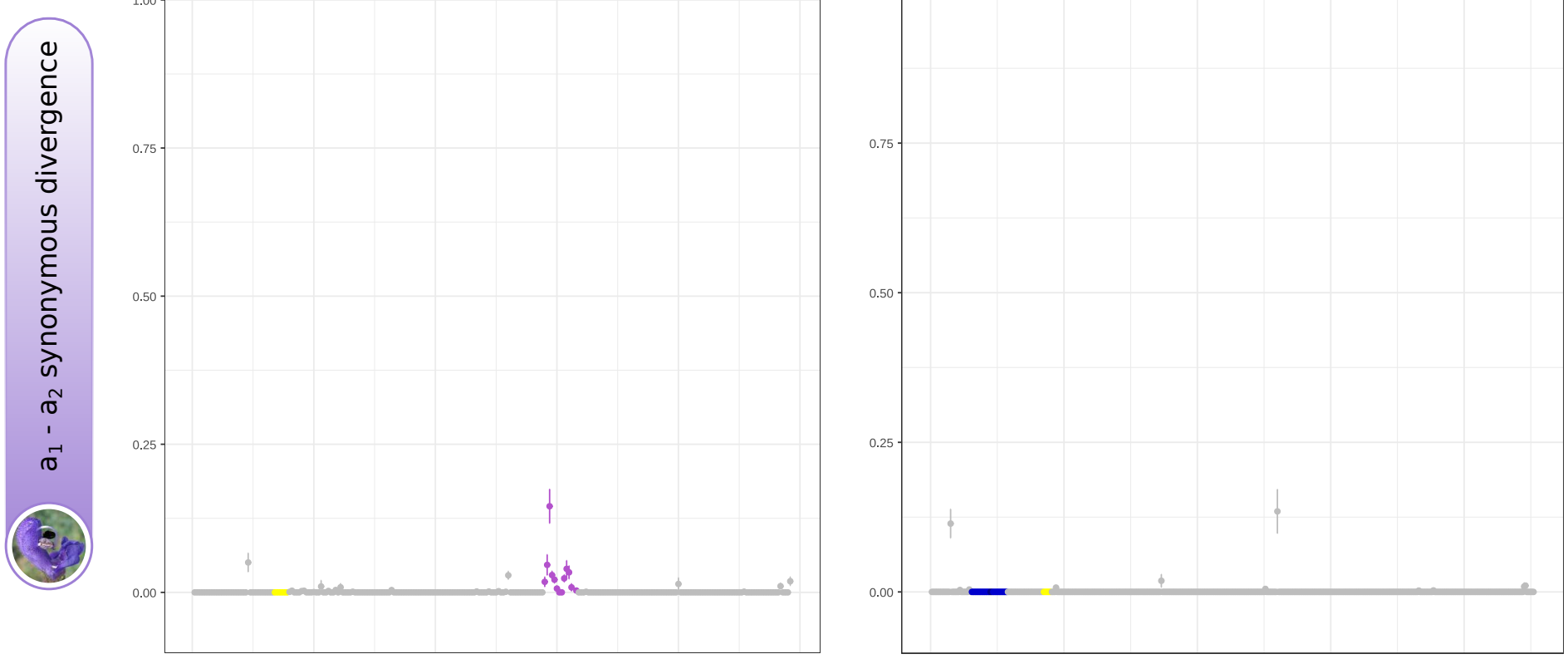
- Genes with $d_S = 0$
- Transposable elements
- Genes
- Centromeric repeats

Key to links

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

Evolutionary strata

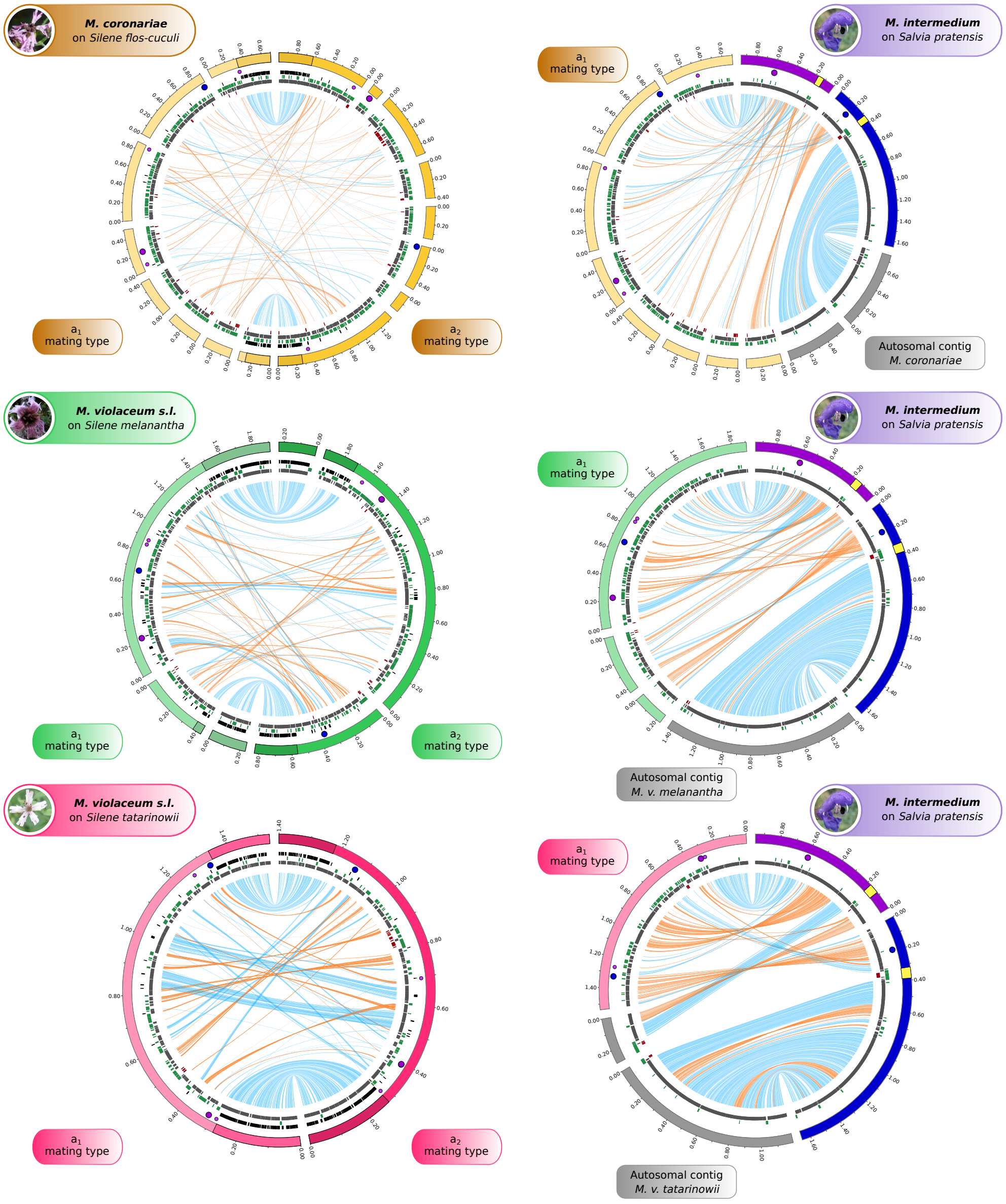
- Purple stratum
- Blue stratum



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

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

Figure S2



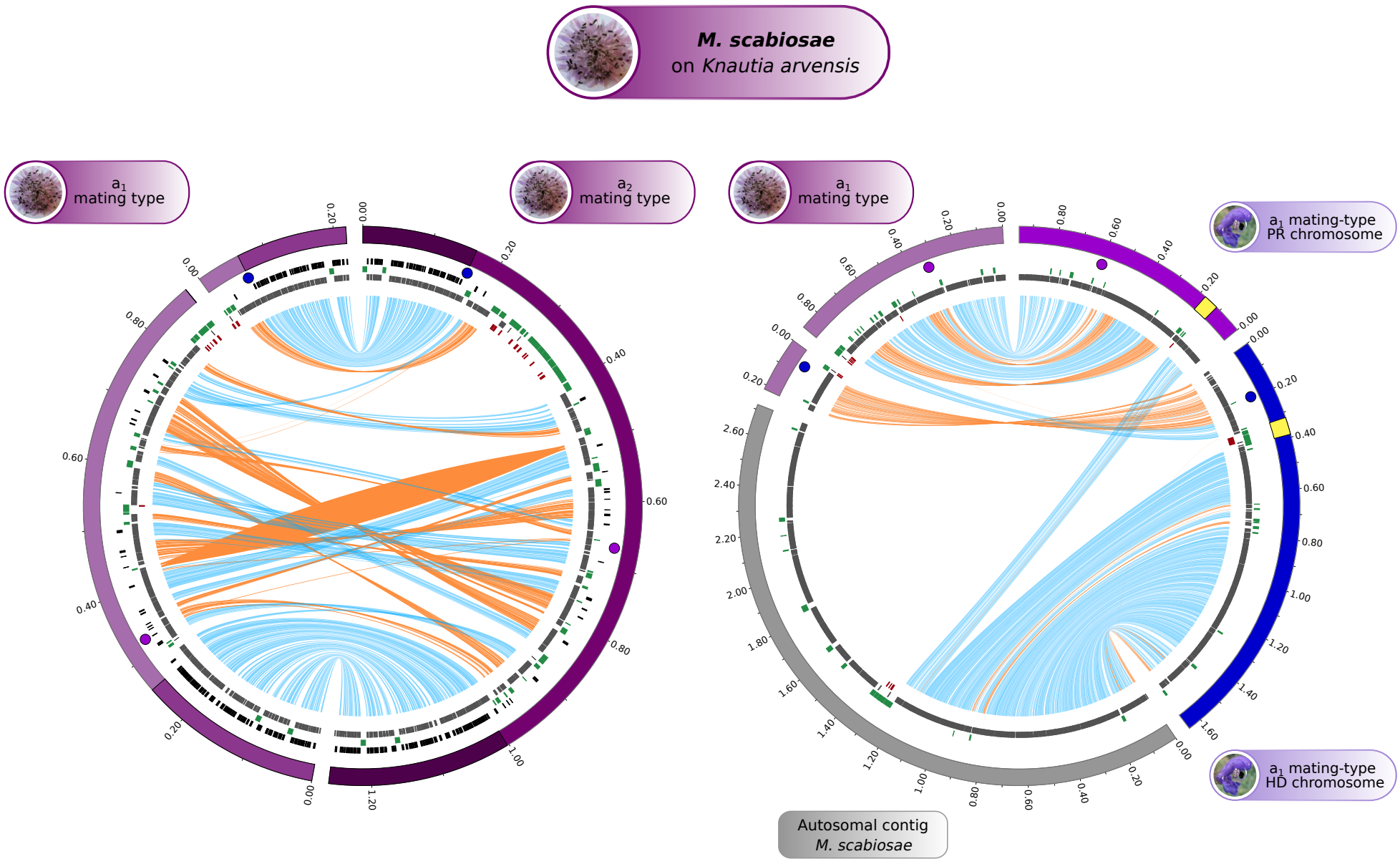
Key to tracks

- 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



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

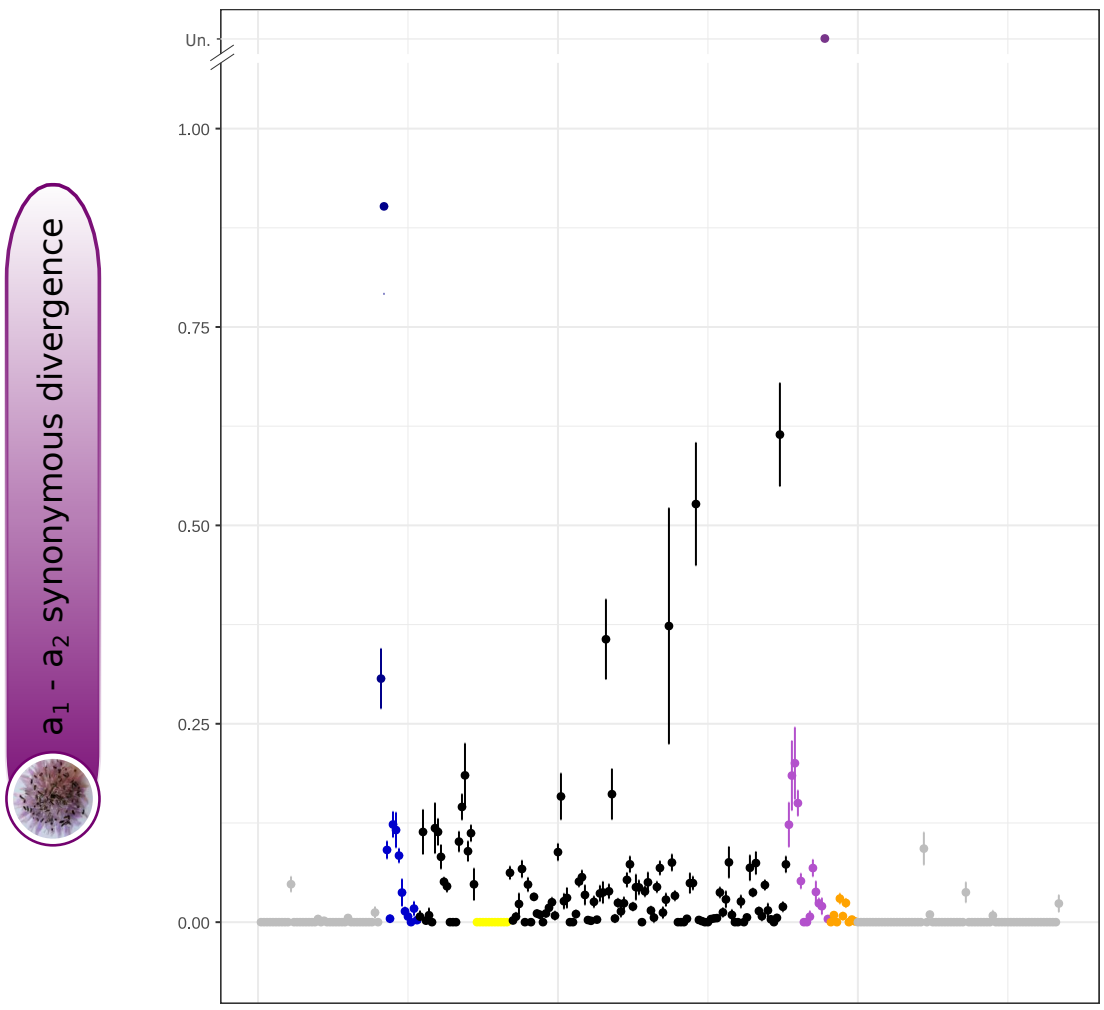
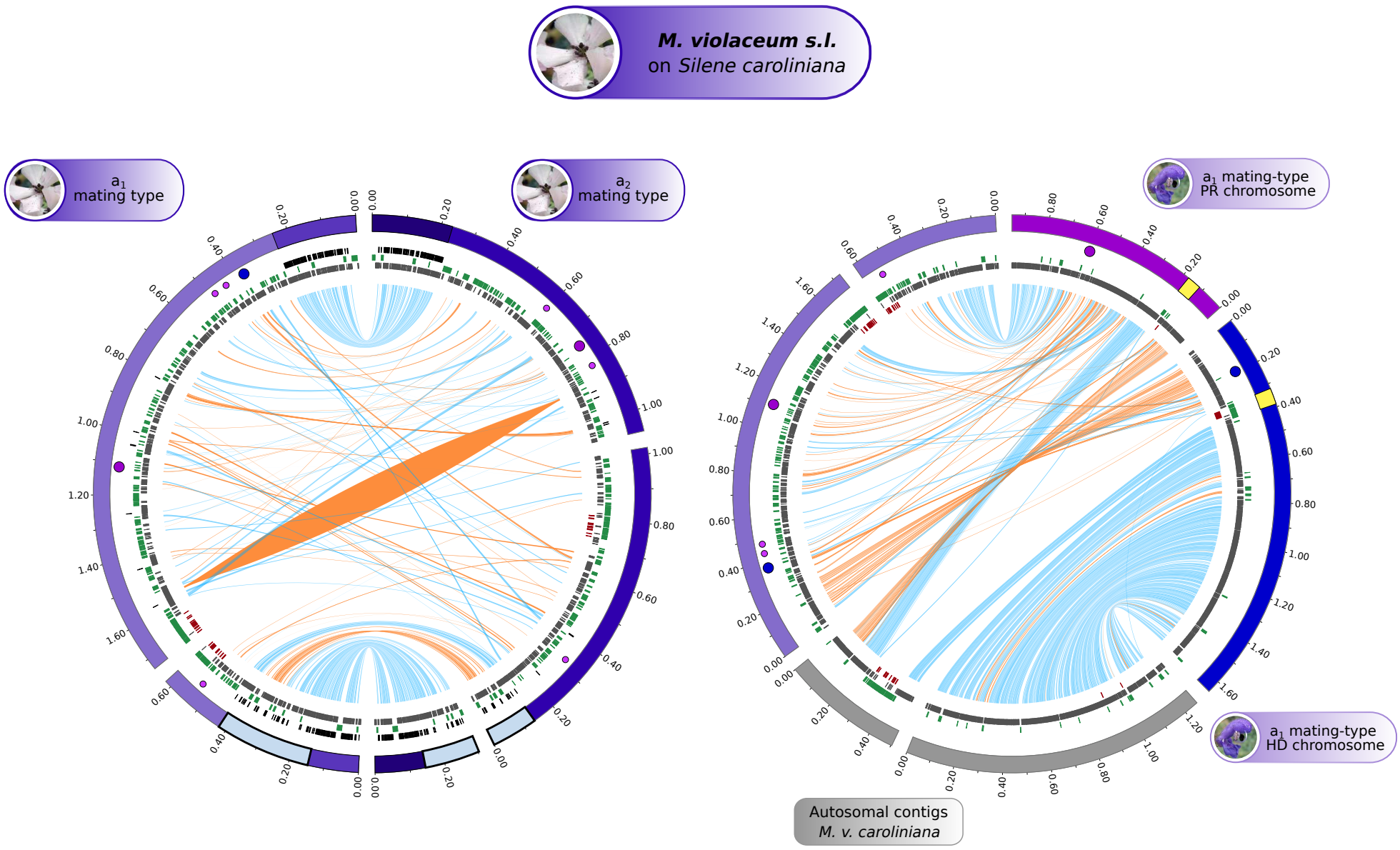


Figure S3b



Key to tracks

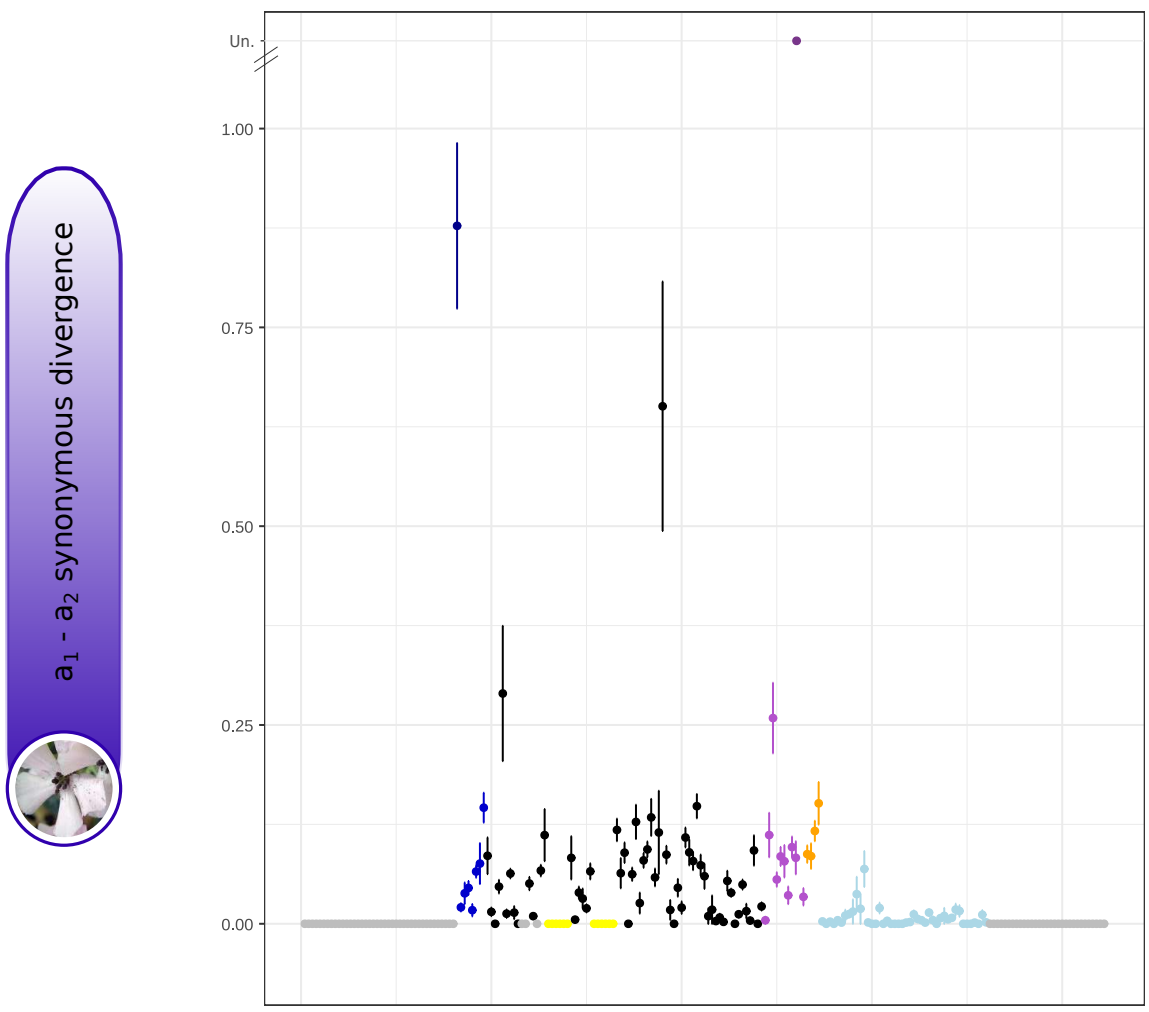
- Genes with $d_S = 0$
- Transposable elements
- Genes
- Centromeric repeats

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*

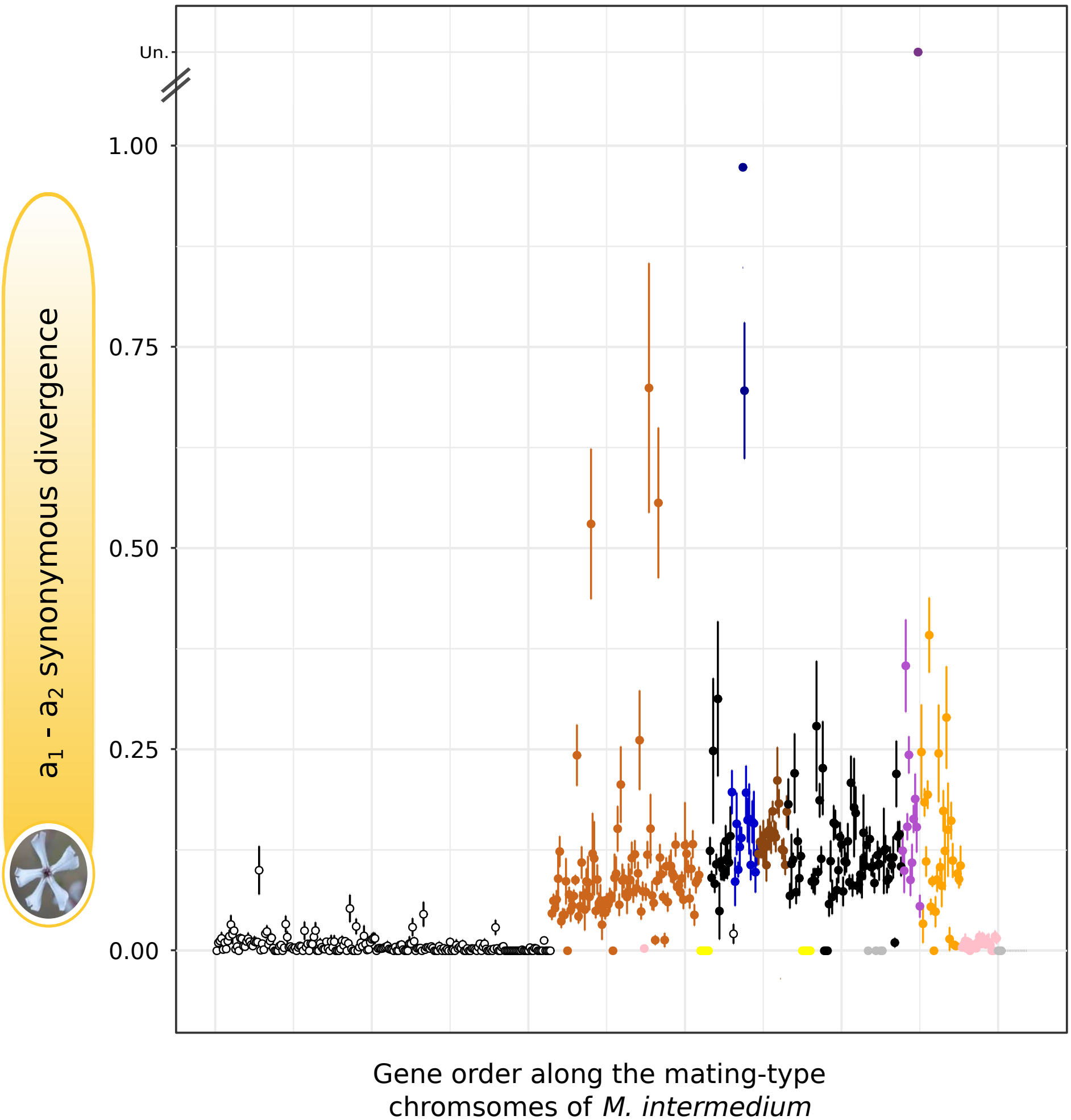


Figure S5a

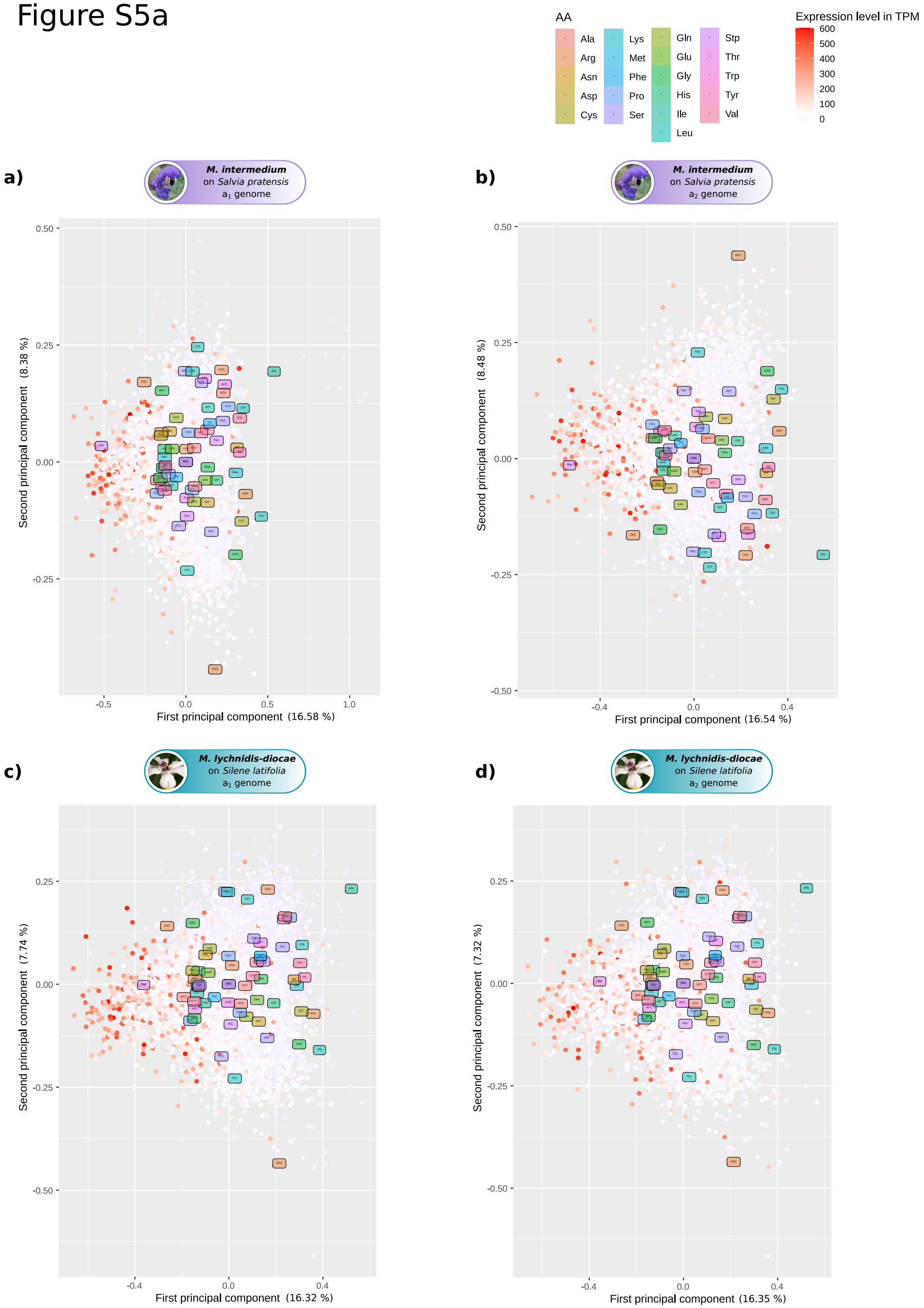

AA
Ala
Arg
Asn
Asp
Cys
Lys
Met
Phe
Pro
Ser
Gln
Glu
Gly
His
Ile
Leu
Stp
Thr
Trp
Tyr
Val
Expression level in TPM
600
500
400
300
200
100
0
a)
M. intermedium
on Salvia pratensis
a1 genome
b)
M. intermedium
on Salvia pratensis
a2 genome
c)
M. lychnidis-diocae
on Silene latifolia
a1 genome
d)
M. lychnidis-diocae
on Silene latifolia
a2 genome
Second principal component (8.38 %)
First principal component (16.58 %)
Second principal component (8.48 %)
First principal component (16.54 %)
Second principal component (7.74 %)
First principal component (16.32 %)
Second principal component (7.32 %)
First principal component (16.35 %)

Figure S5b

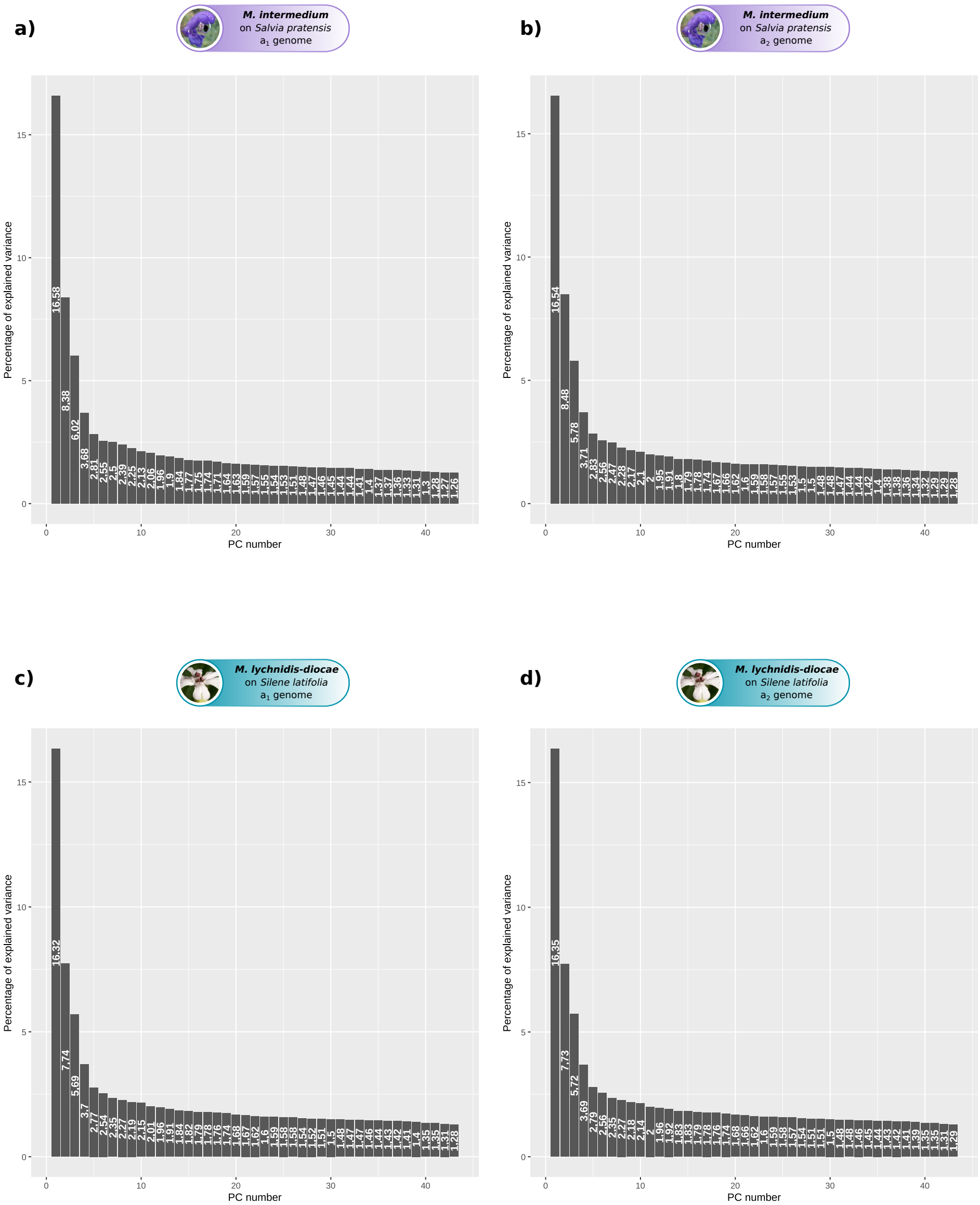
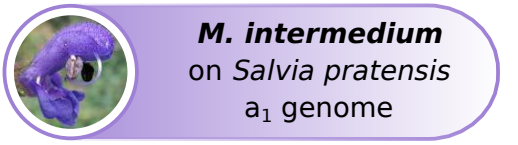
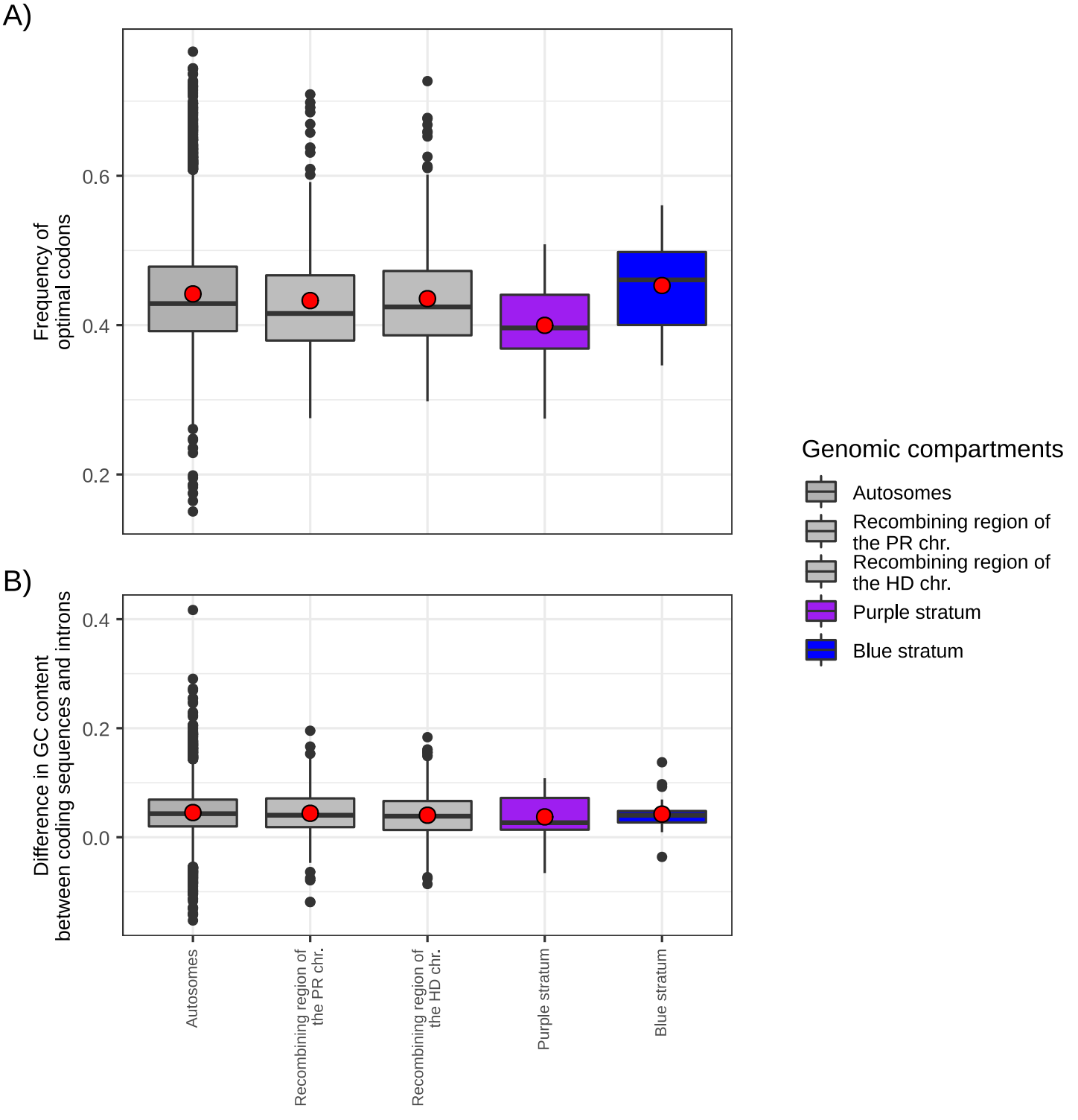
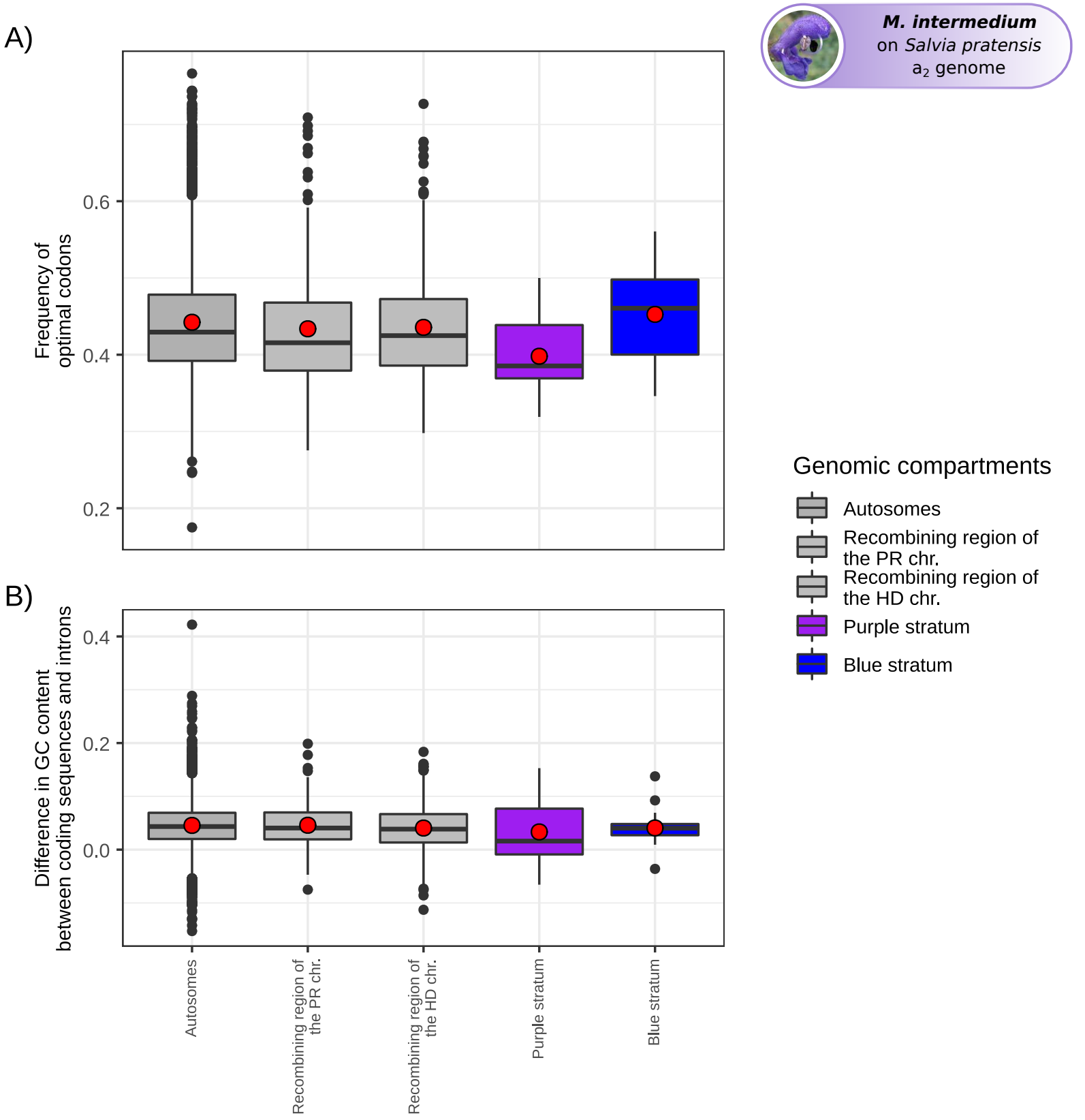
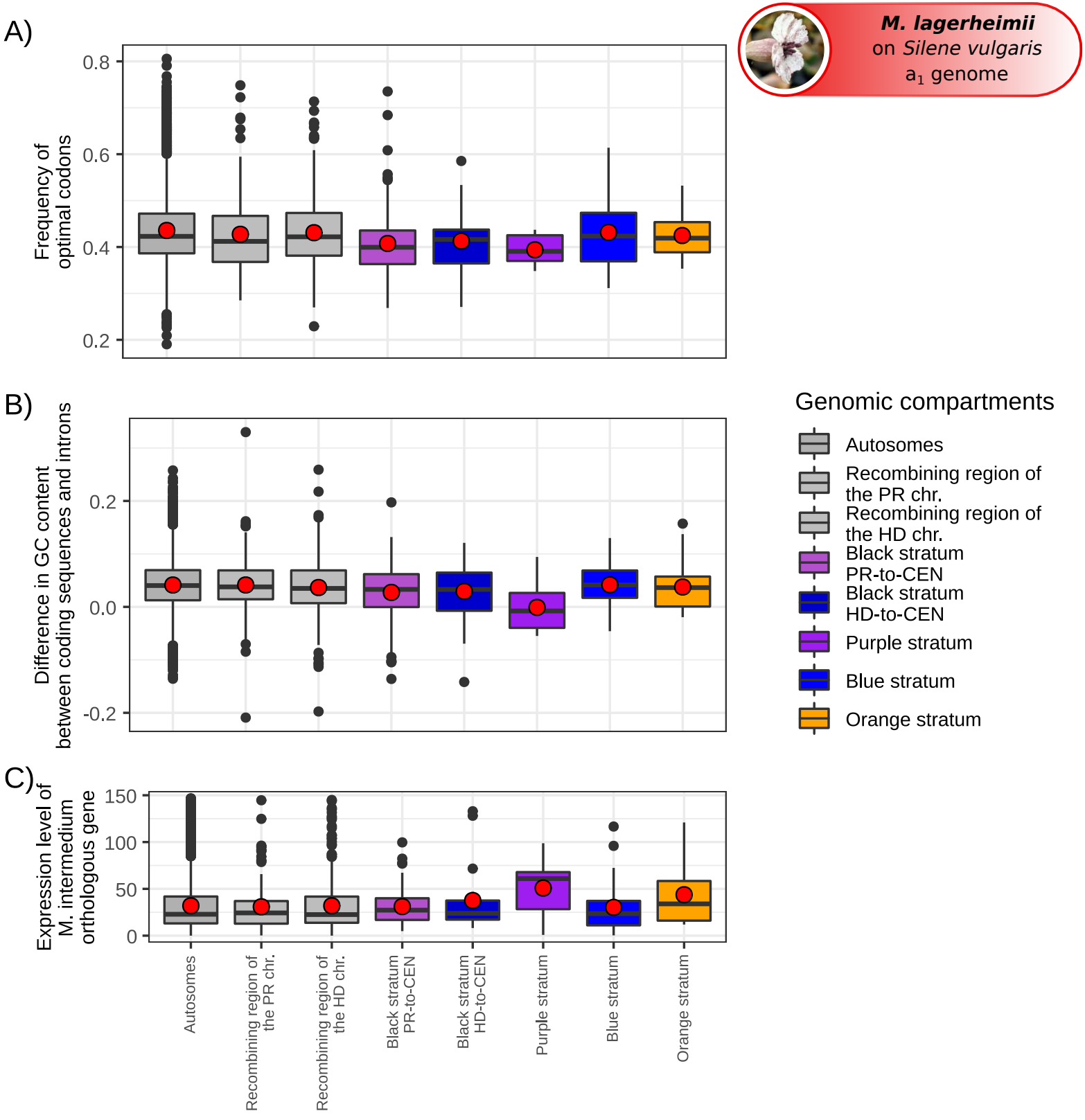

a)
M. intermedium
on Salvia pratensis
a1 genome
Percentage of explained variance
PC number
16.58 8.38 6.02 3.68 2.81 2.55 2.5 2.39 2.25 2.13 2.06 1.96 1.9 1.84 1.77 1.75 1.74 1.71 1.64 1.63 1.59 1.57 1.55 1.54 1.53 1.51 1.48 1.47 1.46 1.45 1.44 1.44 1.41 1.4 1.37 1.37 1.36 1.33 1.31 1.3 1.28 1.27 1.26
b)
M. intermedium
on Salvia pratensis
a2 genome
Percentage of explained variance
PC number
16.54 8.48 5.78 3.71 2.83 2.56 2.47 2.28 2.17 2.1 2 1.95 1.91 1.8 1.79 1.78 1.74 1.67 1.66 1.62 1.6 1.59 1.58 1.57 1.55 1.53 1.5 1.5 1.48 1.48 1.47 1.44 1.44 1.42 1.4 1.38 1.38 1.36 1.34 1.32 1.29 1.29 1.28
c)
M. lychnidis-diocae
on Silene latifolia
a1 genome
Percentage of explained variance
PC number
16.32 7.74 5.69 3.7 2.77 2.54 2.35 2.27 2.19 2.15 2.01 1.96 1.91 1.84 1.82 1.79 1.78 1.76 1.74 1.68 1.67 1.62 1.6 1.59 1.58 1.58 1.54 1.52 1.51 1.5 1.48 1.47 1.47 1.46 1.44 1.43 1.42 1.41 1.4 1.35 1.35 1.31 1.28
d)
M. lychnidis-diocae
on Silene latifolia
a2 genome
Percentage of explained variance
PC number
16.35 7.73 5.72 3.69 2.79 2.56 2.35 2.27 2.18 2.14 2 1.96 1.92 1.83 1.83 1.79 1.78 1.76 1.74 1.68 1.66 1.62 1.6 1.59 1.58 1.57 1.54 1.51 1.51 1.5 1.48 1.48 1.46 1.45 1.44 1.44 1.43 1.42 1.41 1.39 1.35 1.35 1.31 1.29

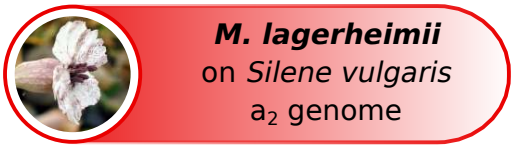
Figure S6


M. intermedium
on *Salvia pratensis*
a_1 genome

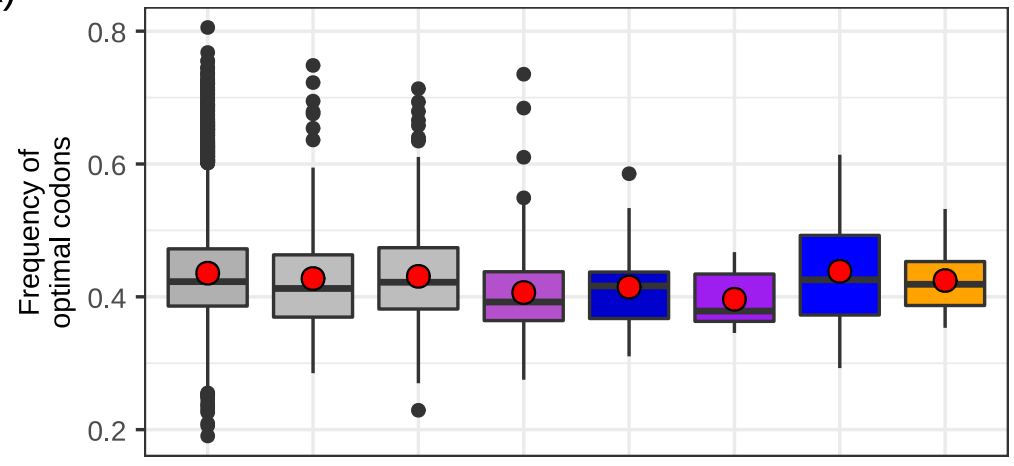

A)
Frequency of optimal codons
0.6
0.4
0.2
B)
Difference in GC content between coding sequences and introns
0.4
0.2
0.0
Autosomes
Recombining region of the PR chr.
Recombining region of the HD chr.
Purple stratum
Blue stratum
Genomic compartments
Autosomes
Recombining region of the PR chr.
Recombining region of the HD chr.
Purple stratum
Blue stratum


M. intermedium
on Salvia pratensis
a2 genome
A)
Frequency of optimal codons
0.6
0.4
0.2
B)
Difference in GC content between coding sequences and introns
0.4
0.2
0.0
Autosomes
Recombining region of the PR chr.
Recombining region of the HD chr.
Purple stratum
Blue stratum
Genomic compartments
Autosomes
Recombining region of the PR chr.
Recombining region of the HD chr.
Purple stratum
Blue stratum

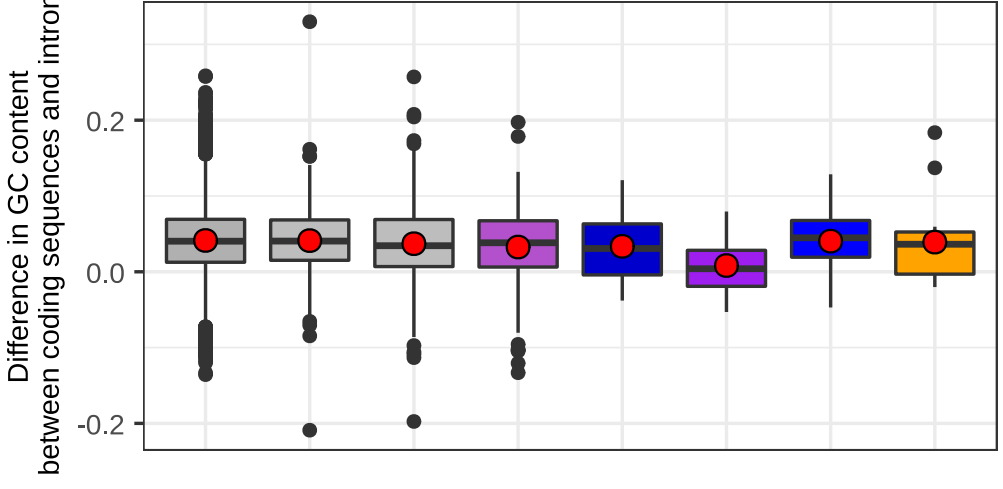

M. lagerheimii
on Silene vulgaris
a1 genome
A)
Frequency of optimal codons
0.8
0.6
0.4
0.2
B)
Difference in GC content between coding sequences and introns
0.2
0.0
-0.2
C)
Expression level of M. intermedium orthologous gene
150
100
50
0
Autosomes
Recombining region of the PR chr.
Recombining region of the HD chr.
Black stratum PR-to-CEN
Black stratum HD-to-CEN
Purple stratum
Blue stratum
Orange stratum
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



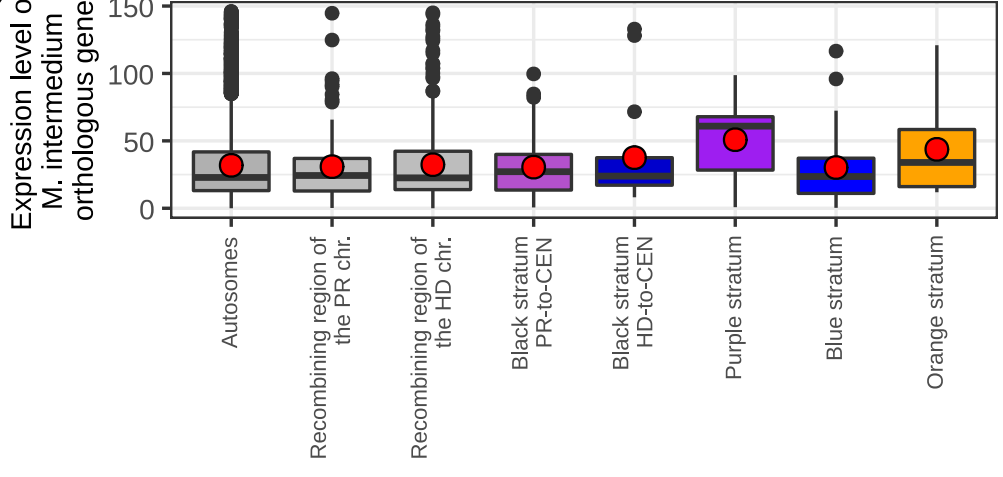
A)



B)

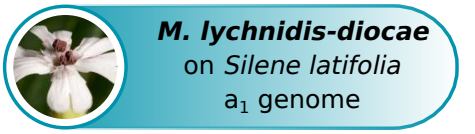


C)

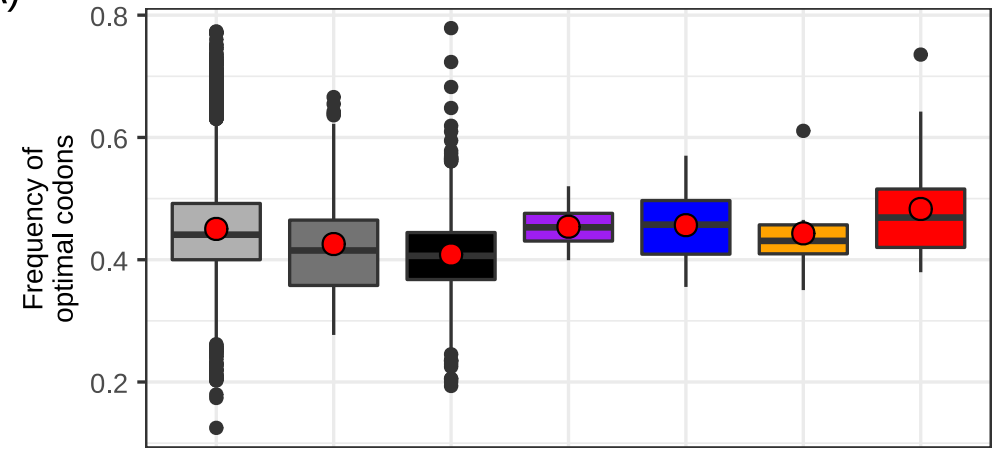


Genomic compartments

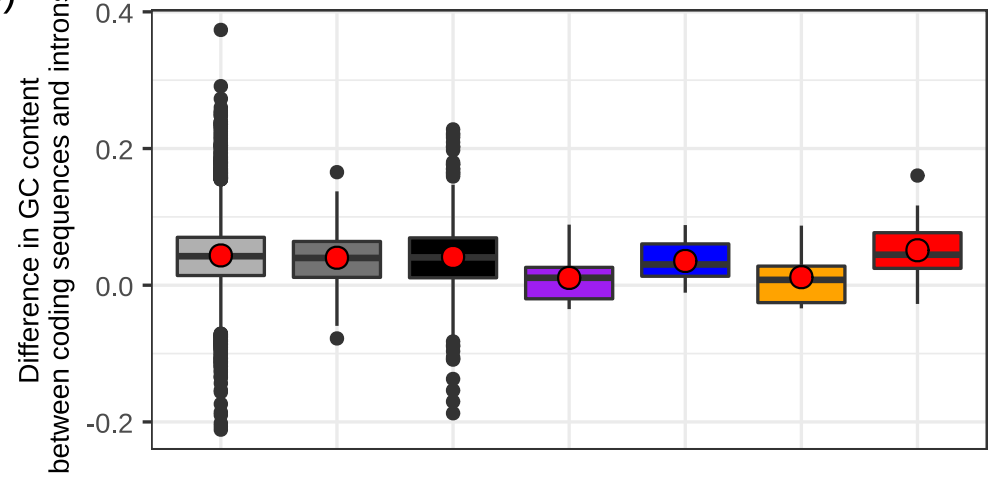
- Autosomes
- Recombining region of the PR chr.
- Recombining region of the HD chr.
- Black stratum PR-to-CEN
- Black stratum HD-to-CEN
- Purple stratum
- Blue stratum
- Orange stratum



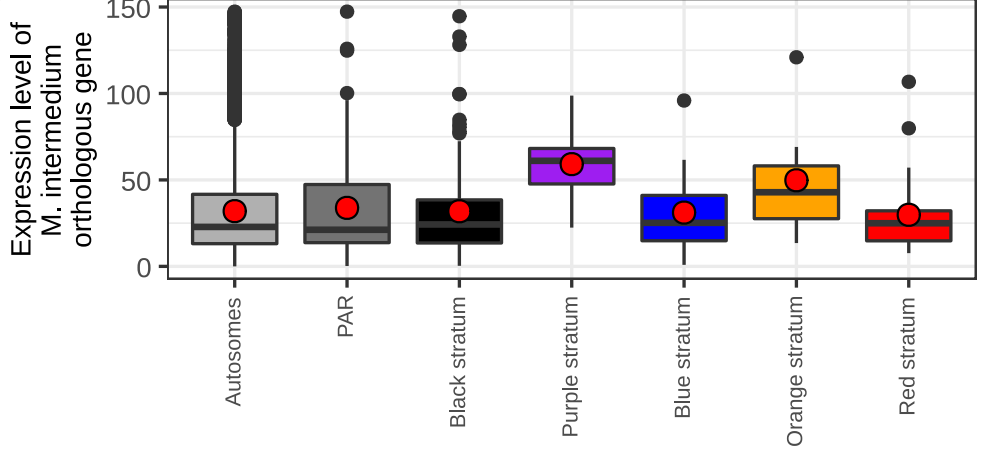
A)



B)

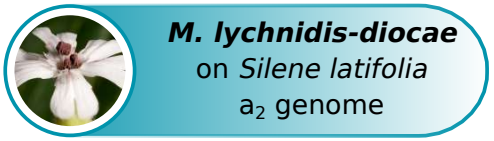


C)

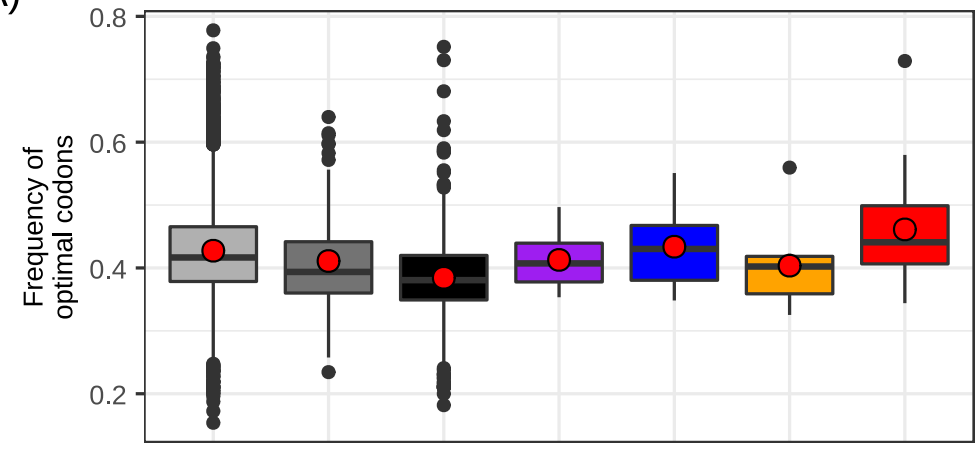


Genomic compartments

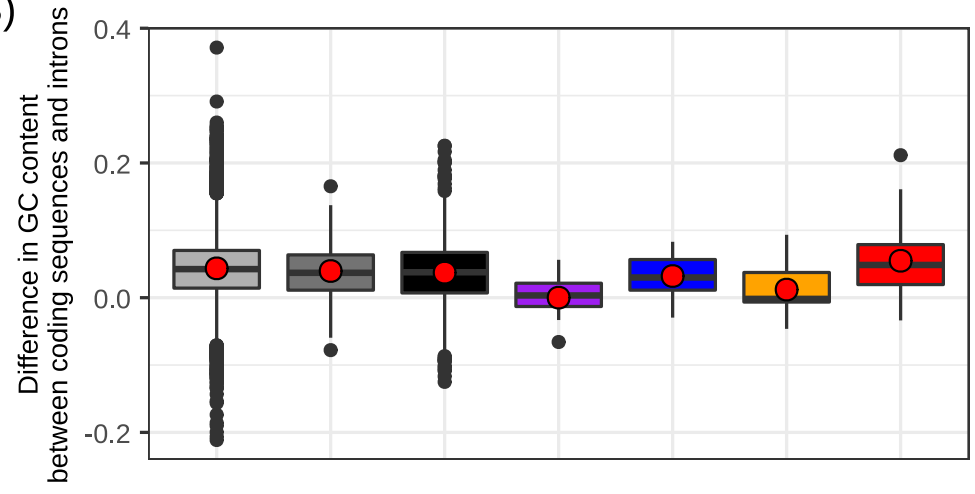
- Autosomes
- PAR
- Black stratum
- Purple stratum
- Blue stratum
- Orange stratum
- Red stratum



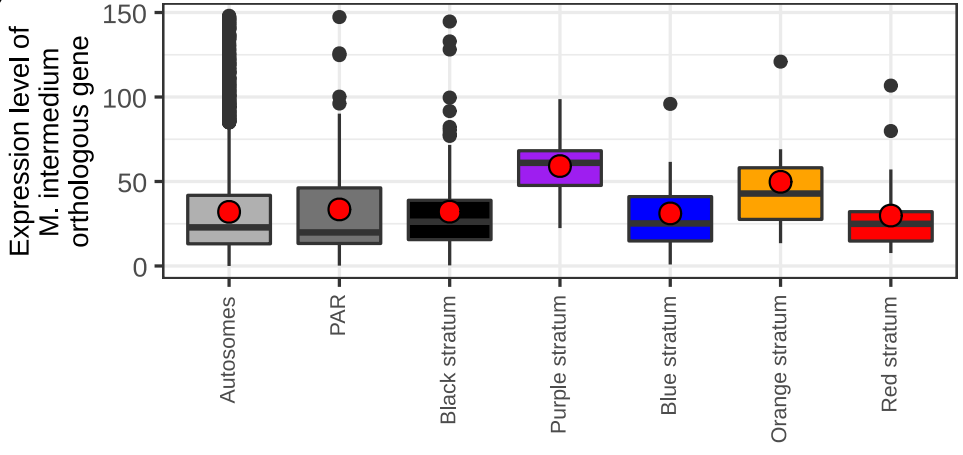
A)



B)

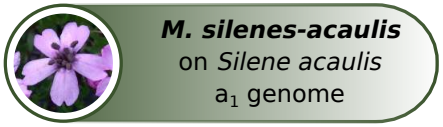


C)



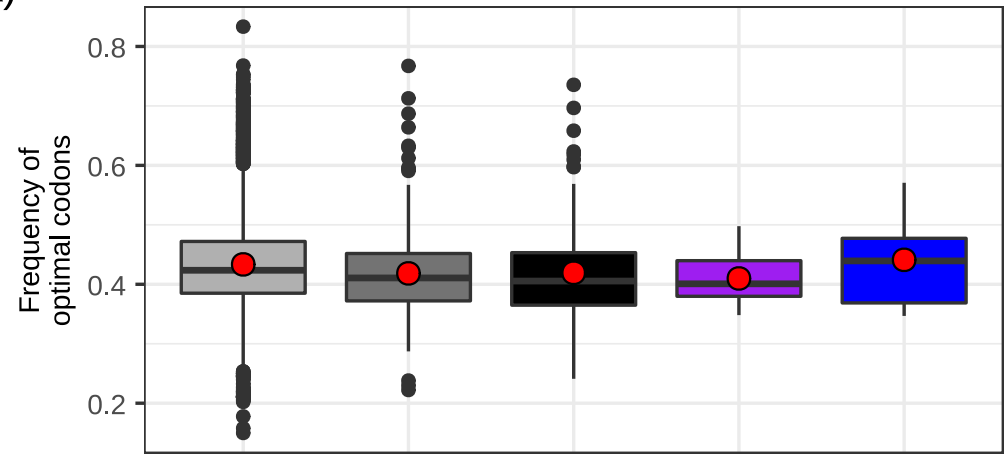
Genomic compartments

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

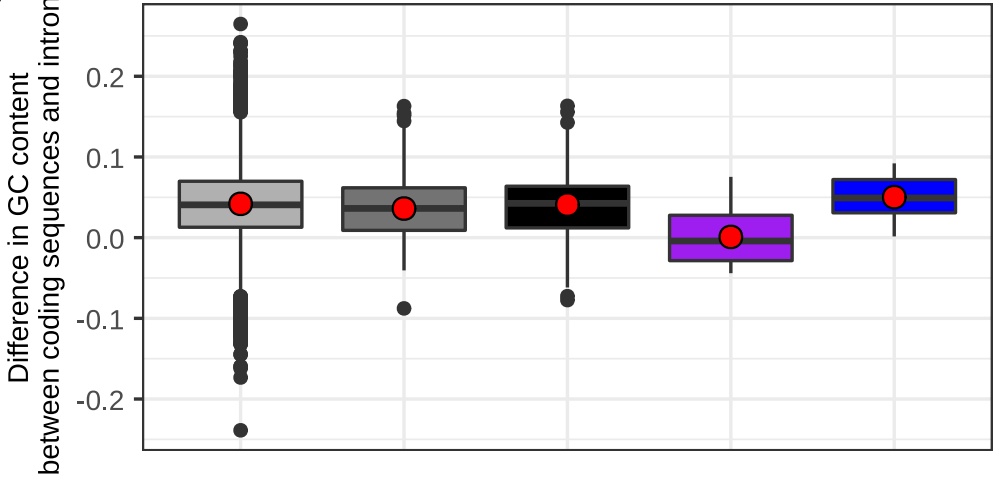


M. silenes-acaulis
on *Silene acaulis*
a_1 genome

A)



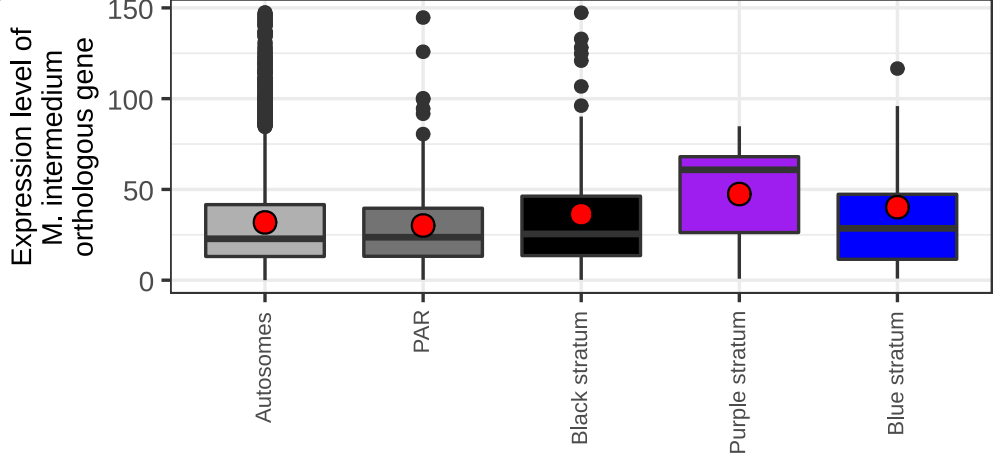
B)

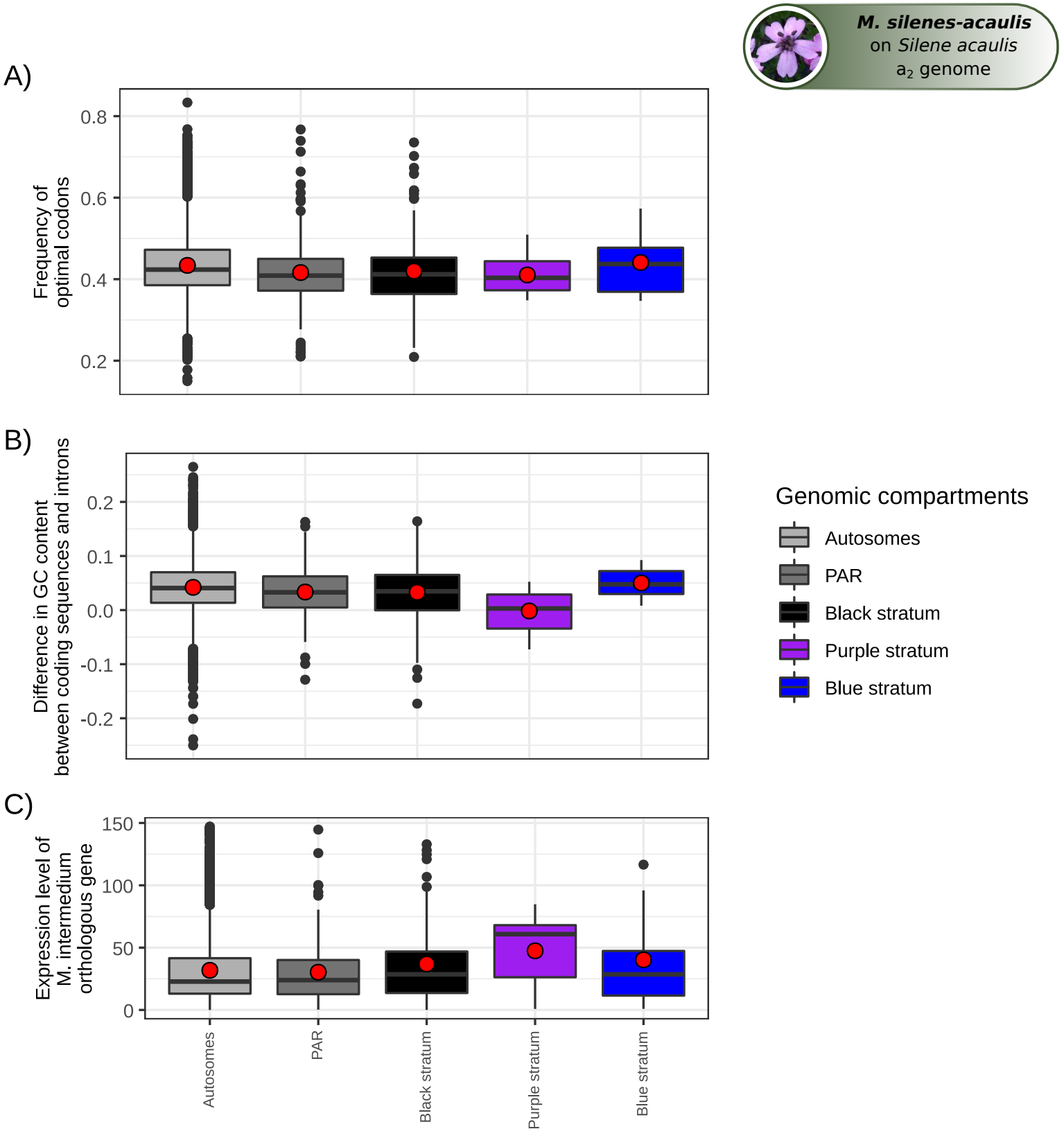


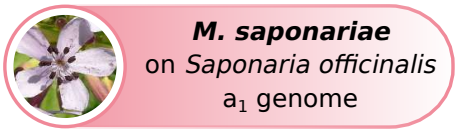
Genomic compartments

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

C)

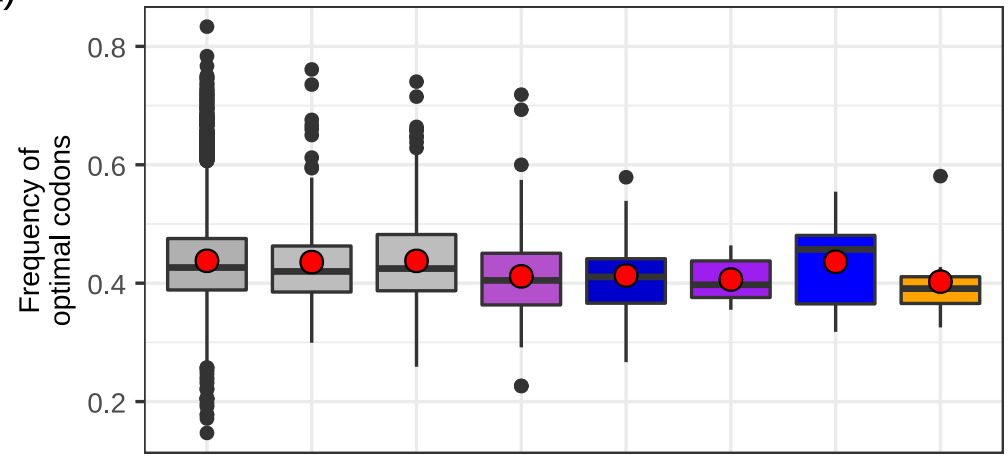



M. silenes-acaulis
on Silene acaulis
a2 genome
A)
Frequency of optimal codons
0.8
0.6
0.4
0.2
B)
Difference in GC content between coding sequences and introns
0.2
0.1
0.0
-0.1
-0.2
C)
Expression level of M. intermedium orthologous gene
150
100
50
0
Autosomes
PAR
Black stratum
Purple stratum
Blue stratum
Genomic compartments
Autosomes
PAR
Black stratum
Purple stratum
Blue stratum

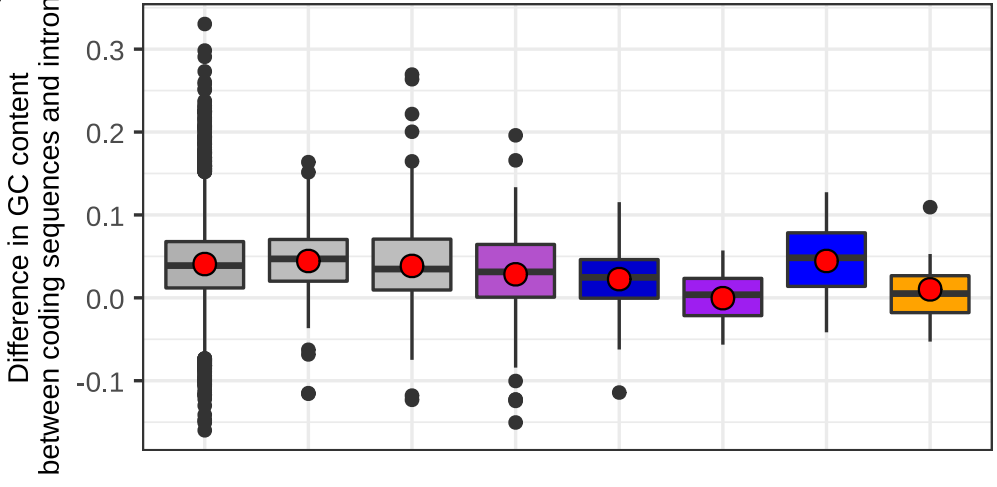


M. saponariae
on *Saponaria officinalis*
a_1 genome

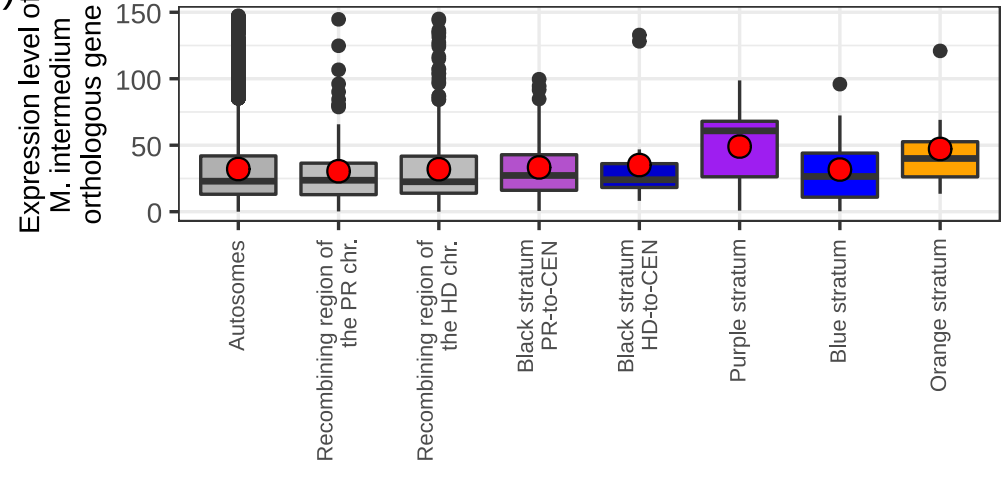
A)



B)

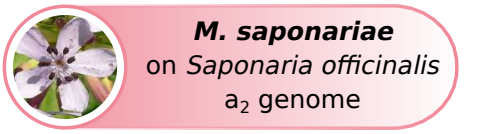


C)

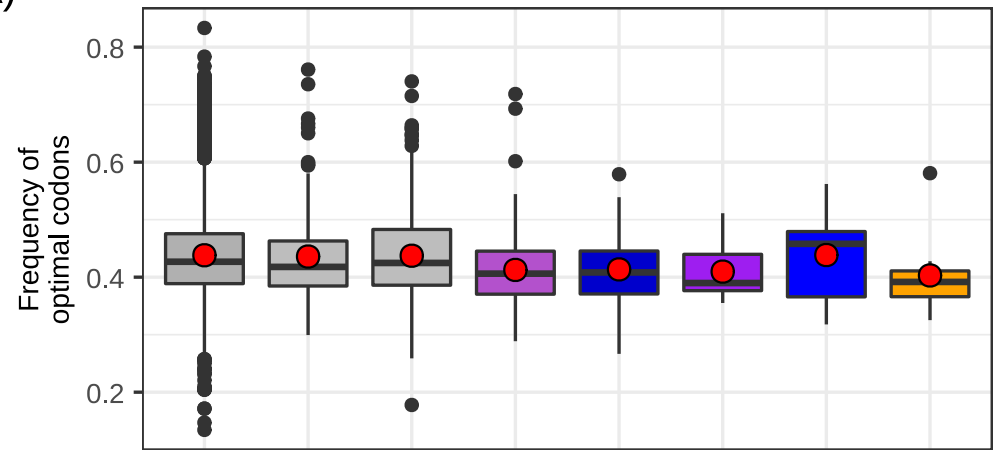


Genomic compartments

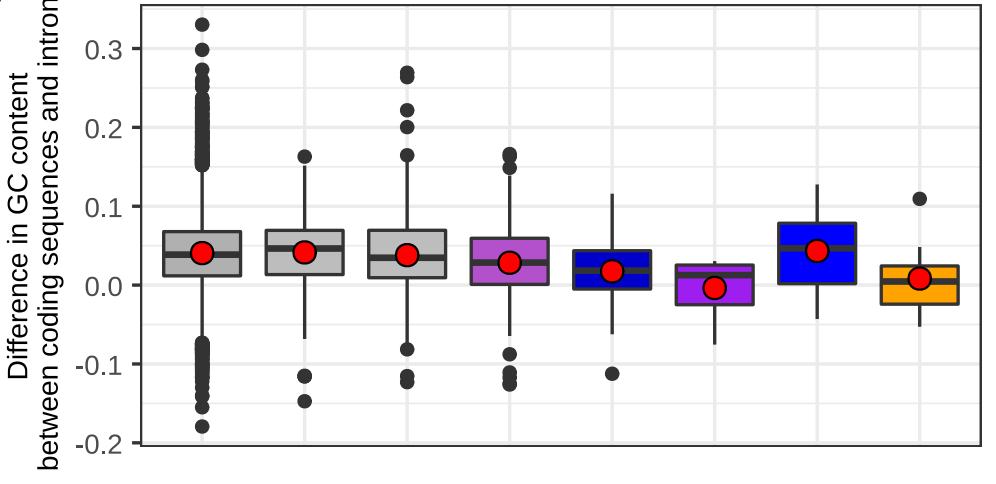
- Autosomes
- Recombining region of the PR chr.
- Recombining region of the HD chr.
- Black stratum PR-to-CEN
- Black stratum HD-to-CEN
- Purple stratum
- Blue stratum
- Orange stratum



A)



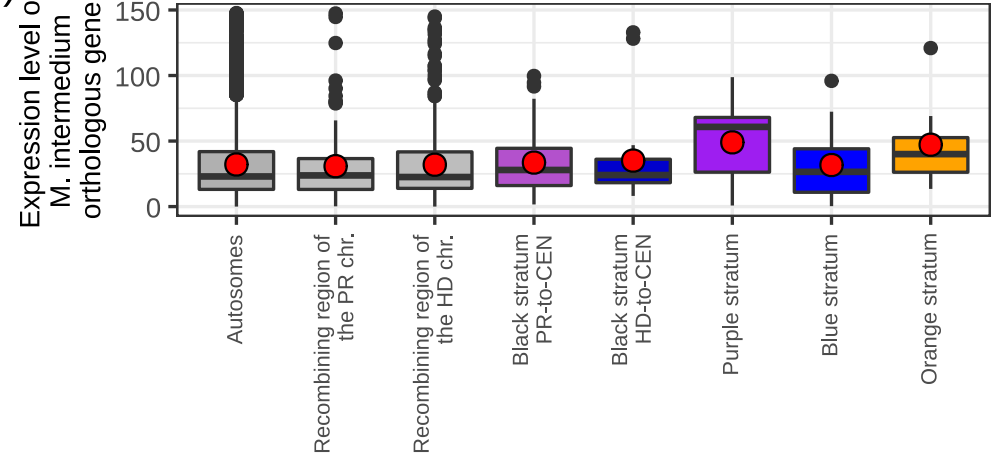
B)

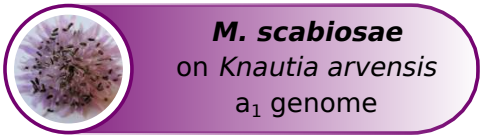


Genomic compartments

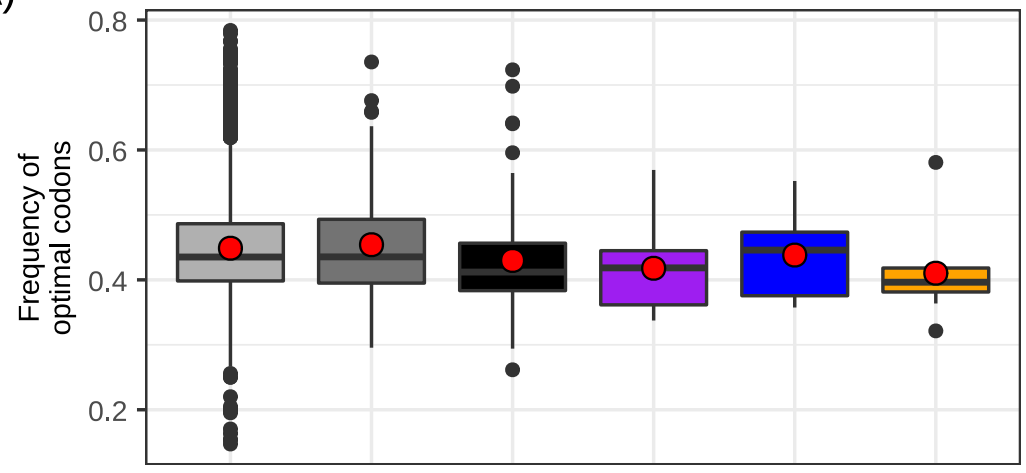
- Autosomes
- Recombining region of the PR chr.
- Recombining region of the HD chr.
- Black stratum PR-to-CEN
- Black stratum HD-to-CEN
- Purple stratum
- Blue stratum
- Orange stratum

C)

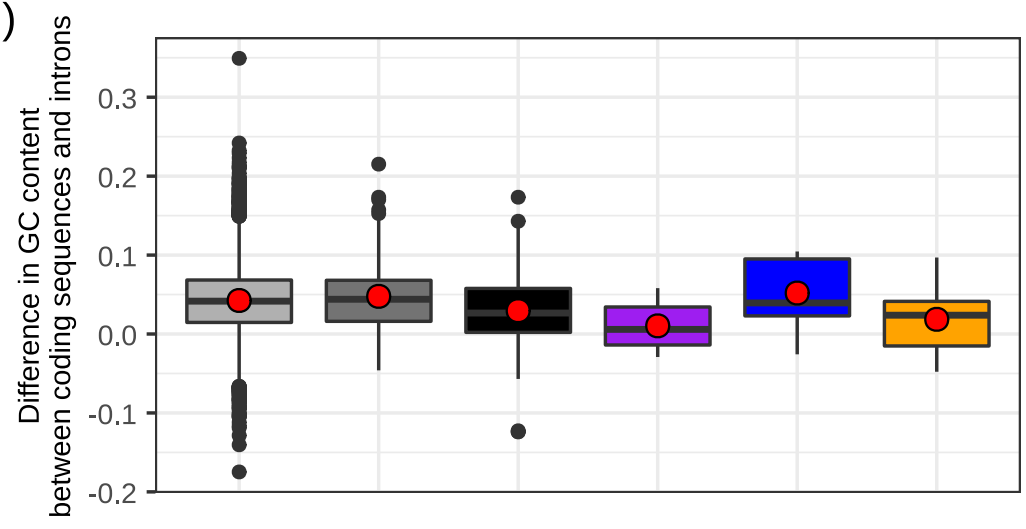




A)



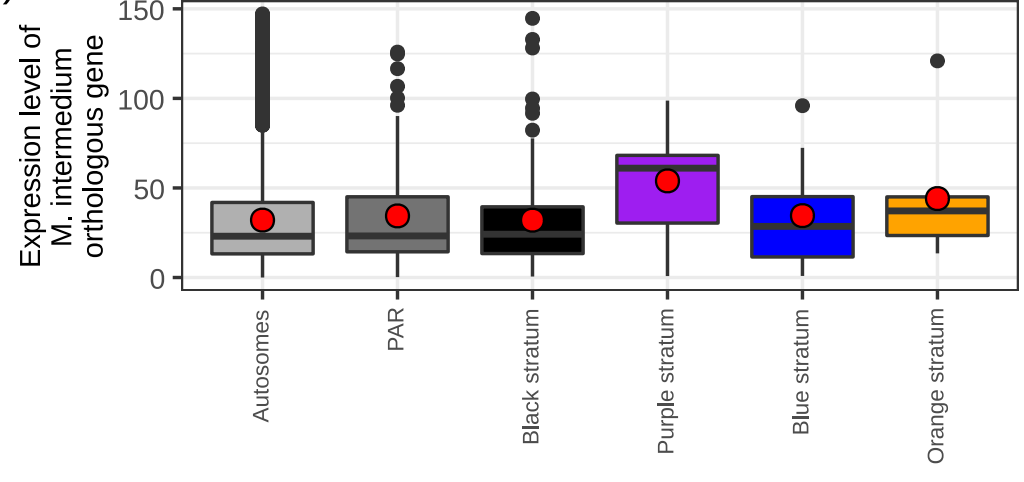
B)

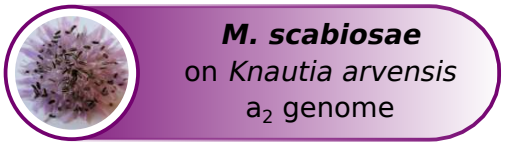


Genomic compartments

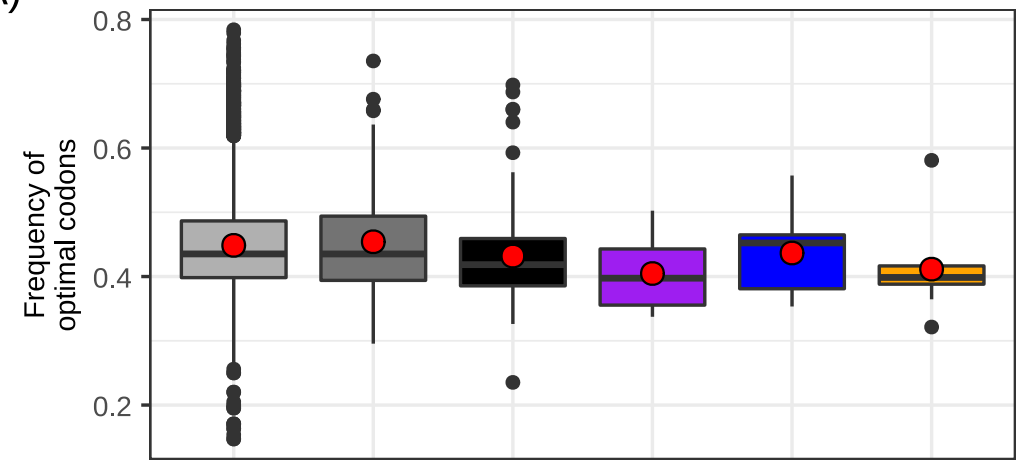
- Autosomes
- PAR
- Black stratum
- Purple stratum
- Blue stratum
- Orange stratum

C)

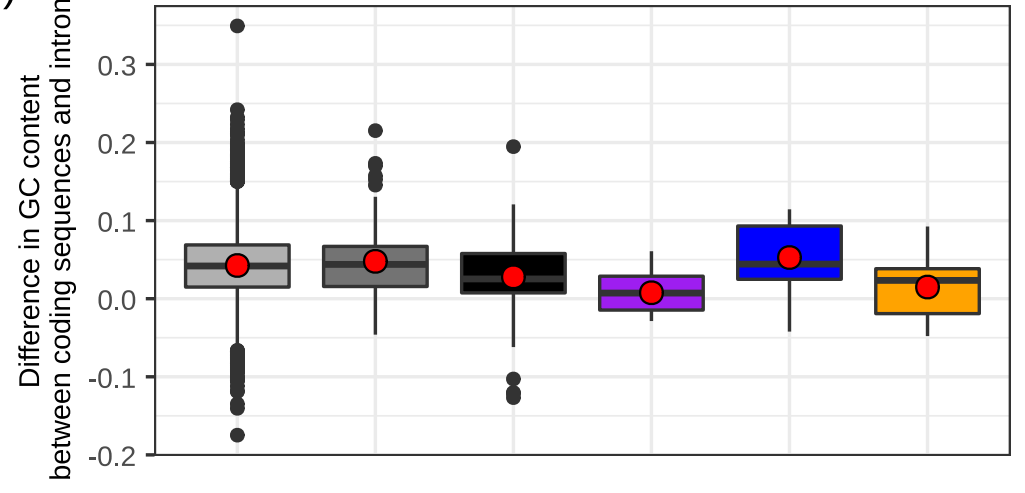


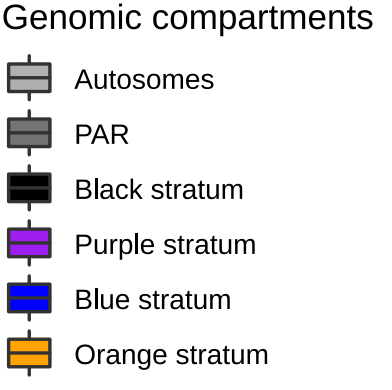

M. scabiosae
on Knautia arvensis
a2 genome

A)

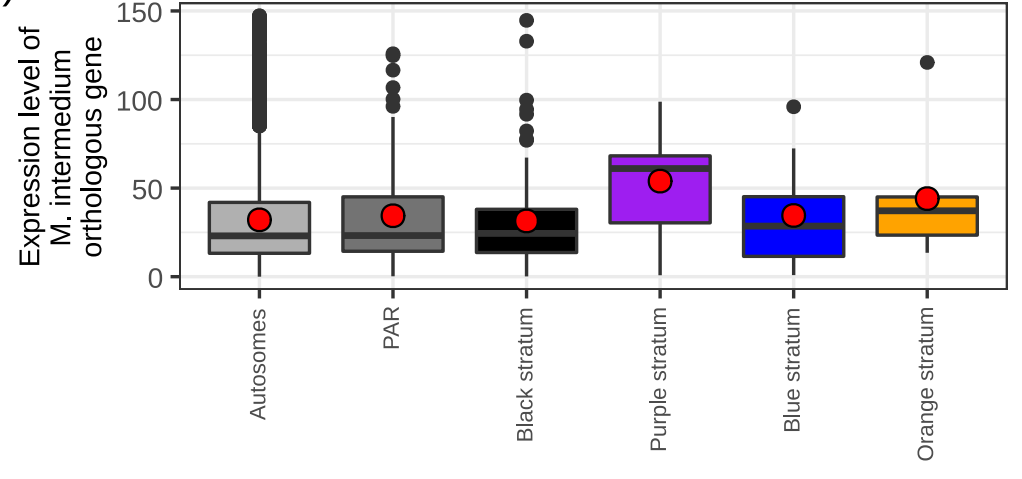

Frequency of optimal codons
0.8
0.6
0.4
0.2

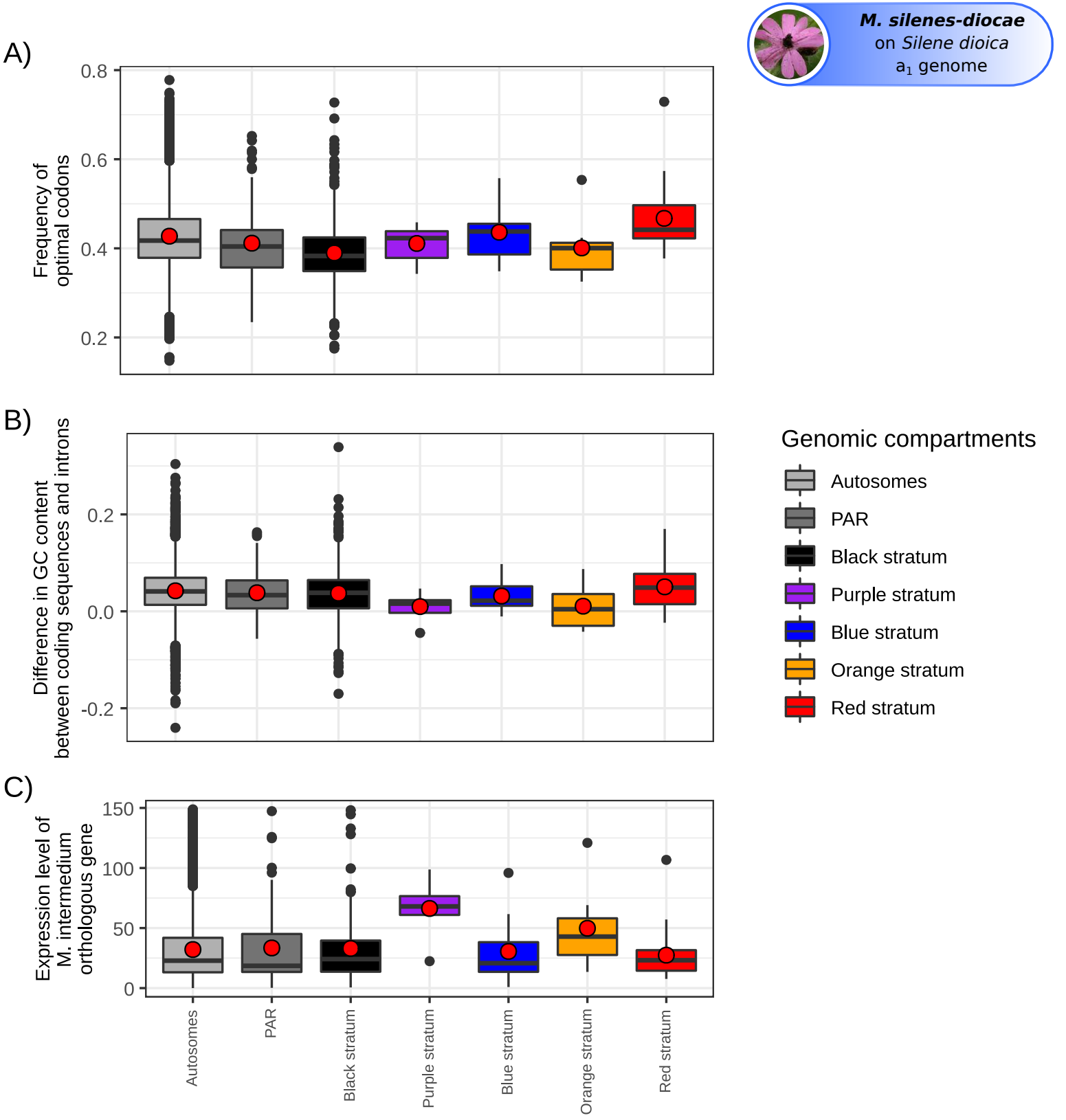
B)


Difference in GC content between coding sequences and introns
0.3
0.2
0.1
0.0
-0.1
-0.2


Genomic compartments
Autosomes
PAR
Black stratum
Purple stratum
Blue stratum
Orange stratum

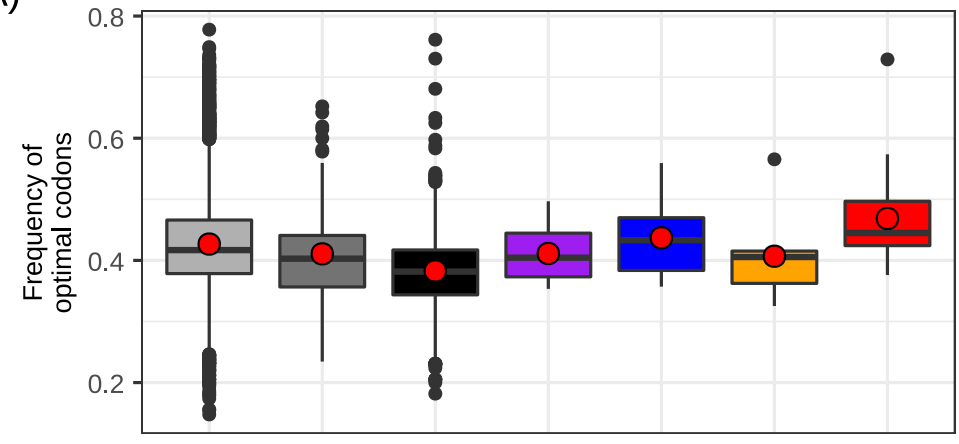
C)


Expression level of M. intermedium orthologous gene
150
100
50
0
Autosomes
PAR
Black stratum
Purple stratum
Blue stratum
Orange stratum

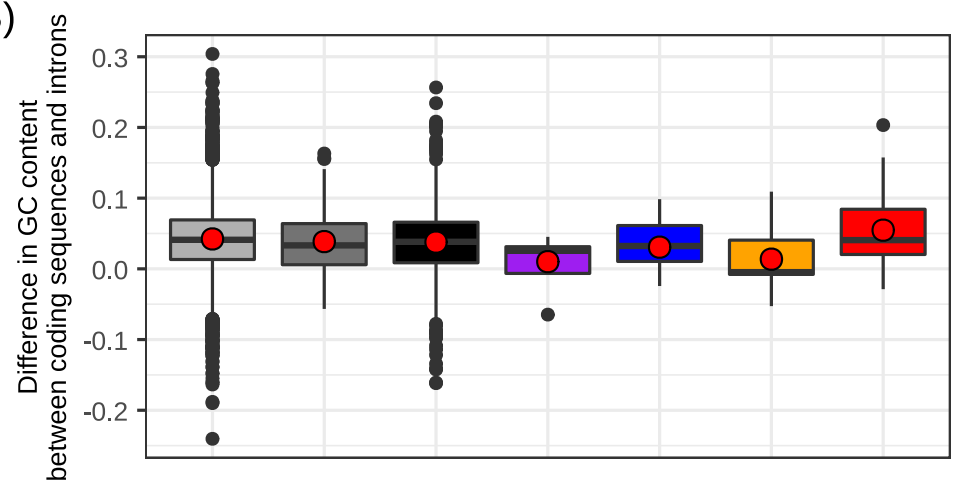

M. silenes-diocae
on Silene dioica
a1 genome
A)
Frequency of optimal codons
0.8
0.6
0.4
0.2
B)
Difference in GC content between coding sequences and introns
0.2
0.0
-0.2
C)
Expression level of M. intermedium orthologous gene
150
100
50
0
Autosomes
PAR
Black stratum
Purple stratum
Blue stratum
Orange stratum
Red stratum
Genomic compartments
Autosomes
PAR
Black stratum
Purple stratum
Blue stratum
Orange stratum
Red stratum

M. silenes-diocae
on *Silene dioica*
a_2 genome

A)



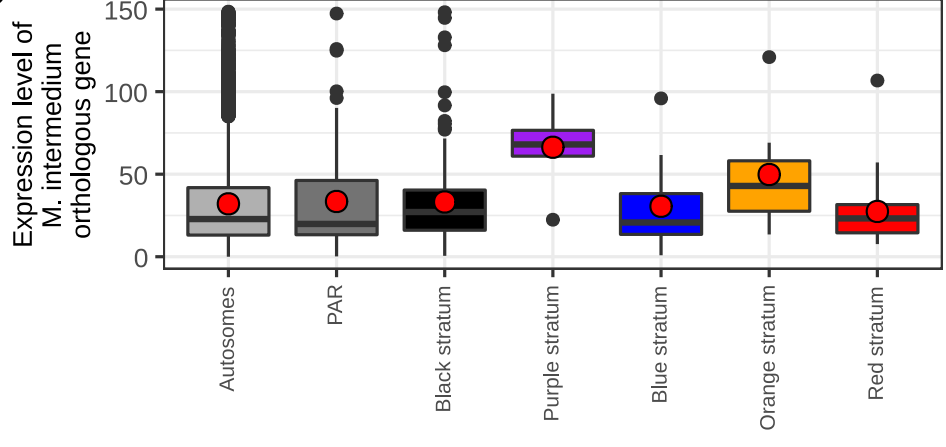
B)

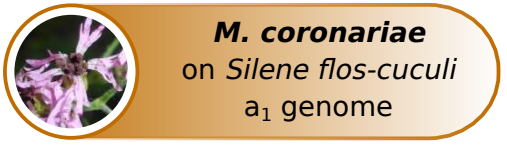


Genomic compartments

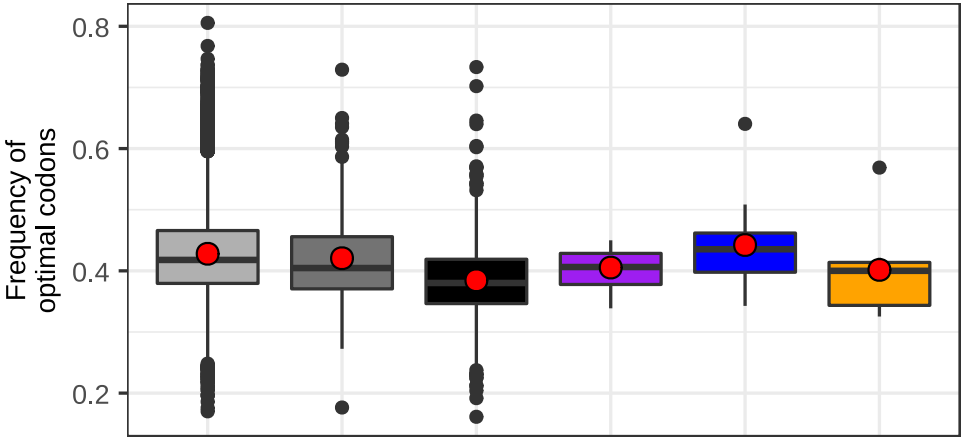
- Autosomes
- PAR
- Black stratum
- Purple stratum
- Blue stratum
- Orange stratum
- Red stratum

C)

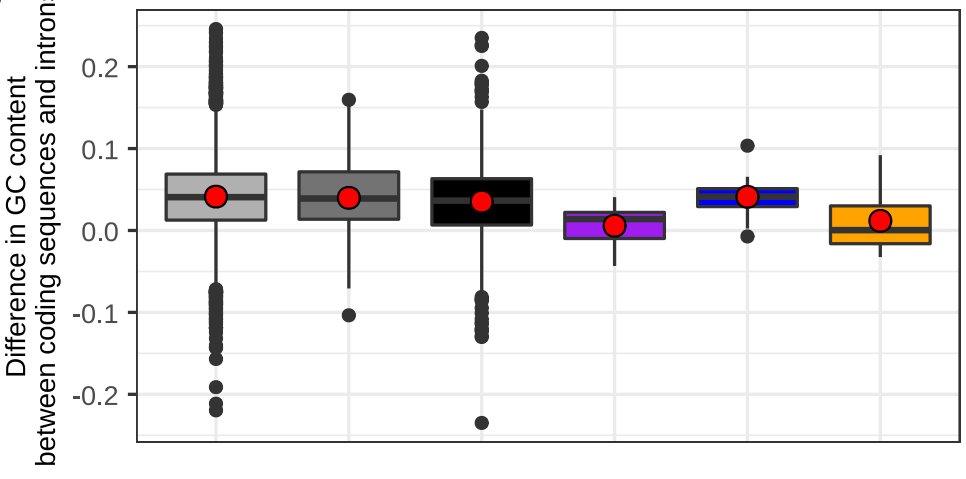


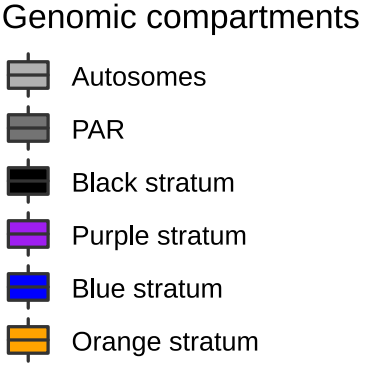

M. coronariae
on Silene flos-cuculi
a1 genome

A)

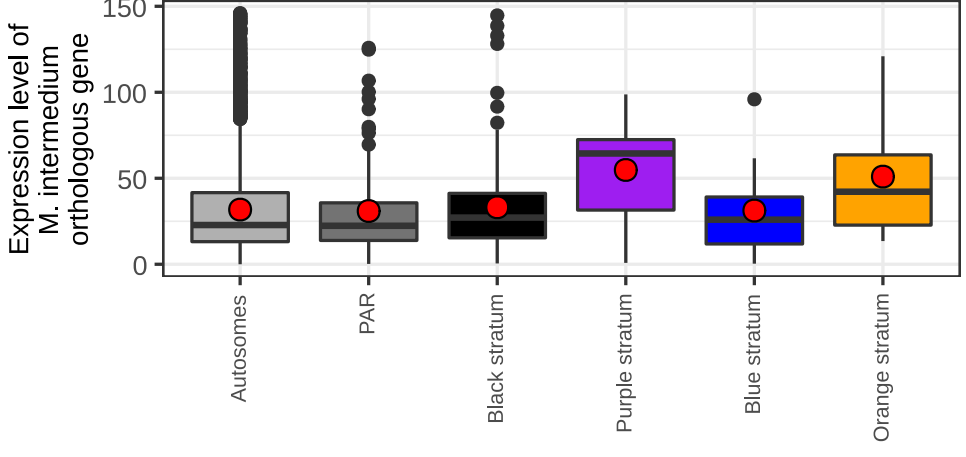

Frequency of optimal codons
0.8
0.6
0.4
0.2

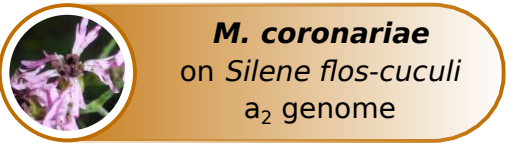
B)


Difference in GC content between coding sequences and introns
0.2
0.1
0.0
-0.1
-0.2

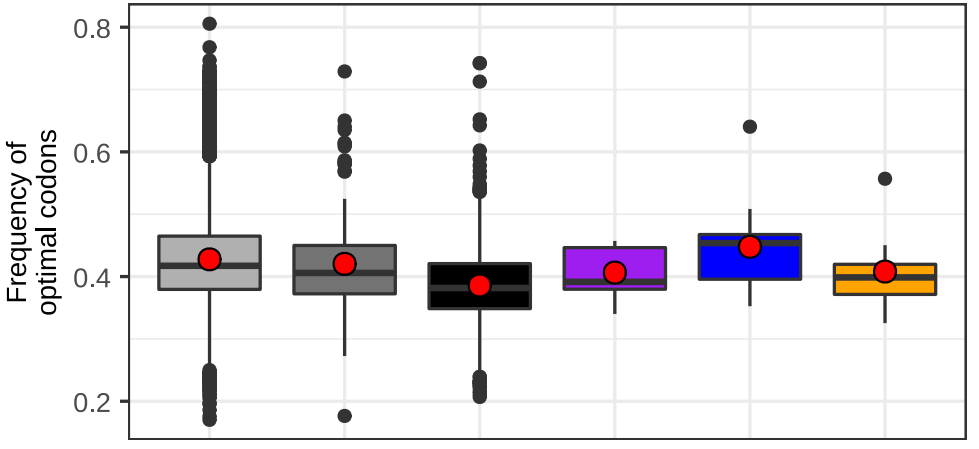

Genomic compartments
Autosomes
PAR
Black stratum
Purple stratum
Blue stratum
Orange stratum

C)

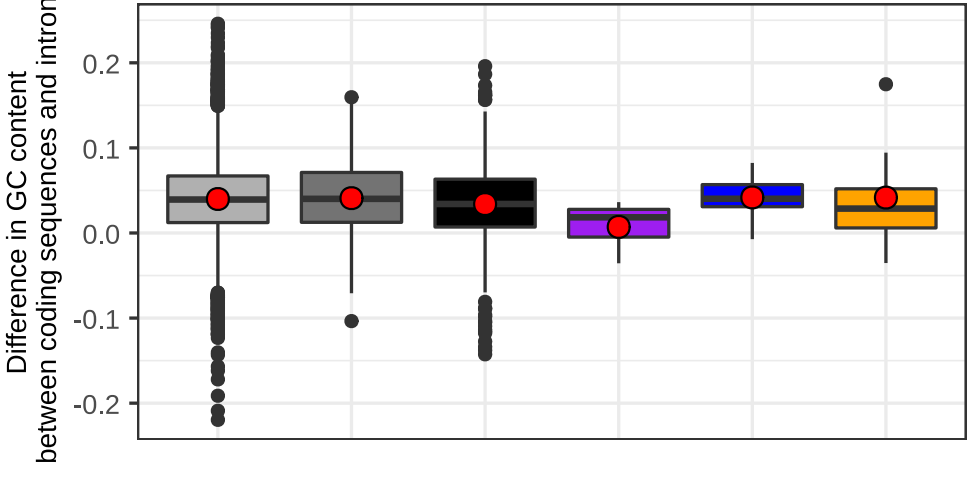

Expression level of M. intermedium orthologous gene
150
100
50
0
Autosomes
PAR
Black stratum
Purple stratum
Blue stratum
Orange stratum



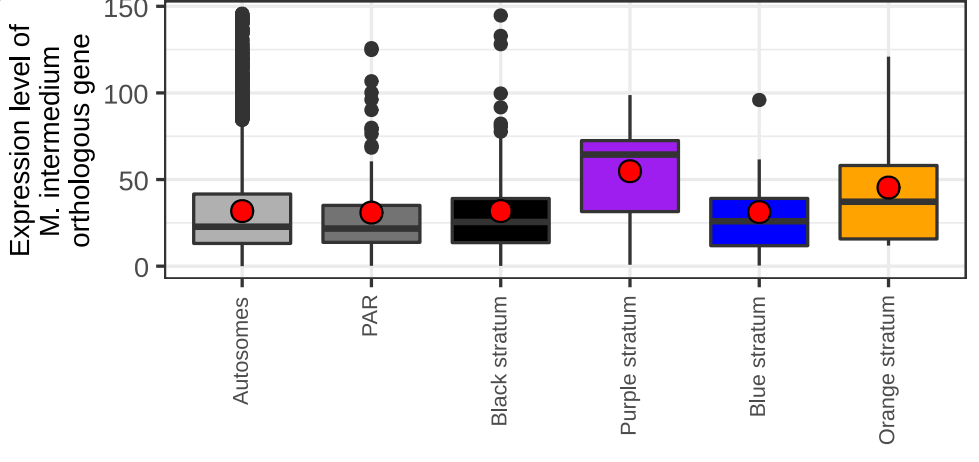
A)



B)

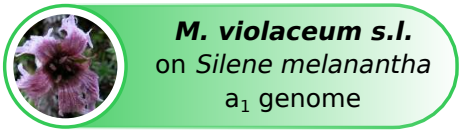


C)

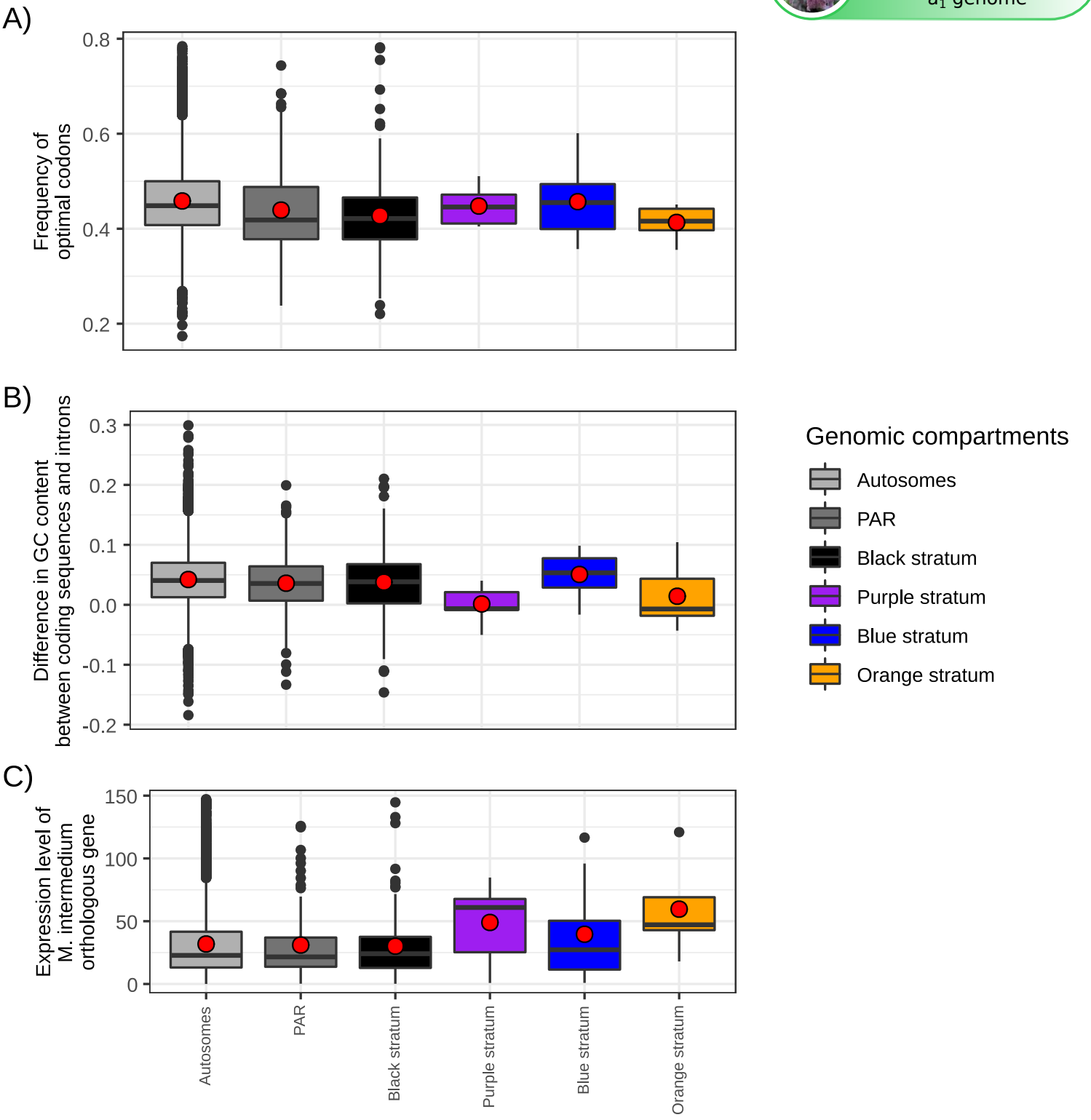


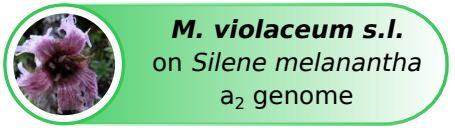
Genomic compartments

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

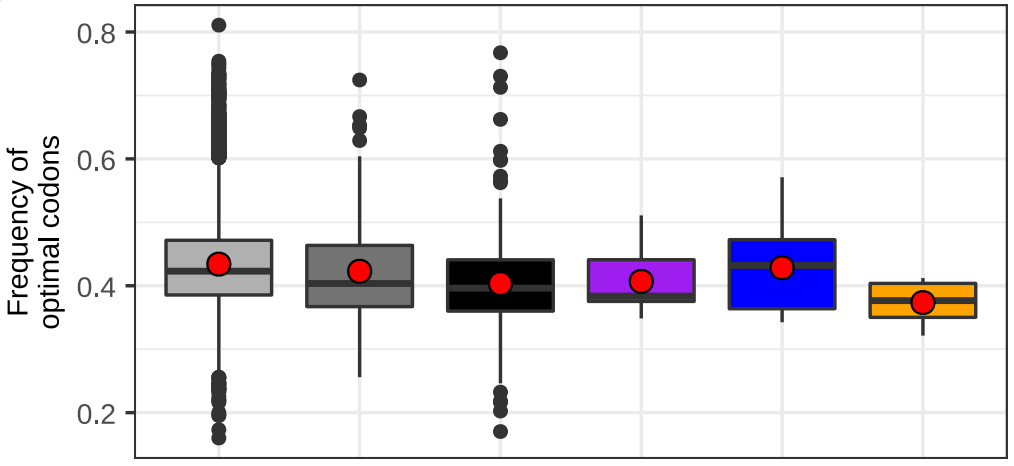


M. violaceum s.l.
on *Silene melanantha*
a_1 genome

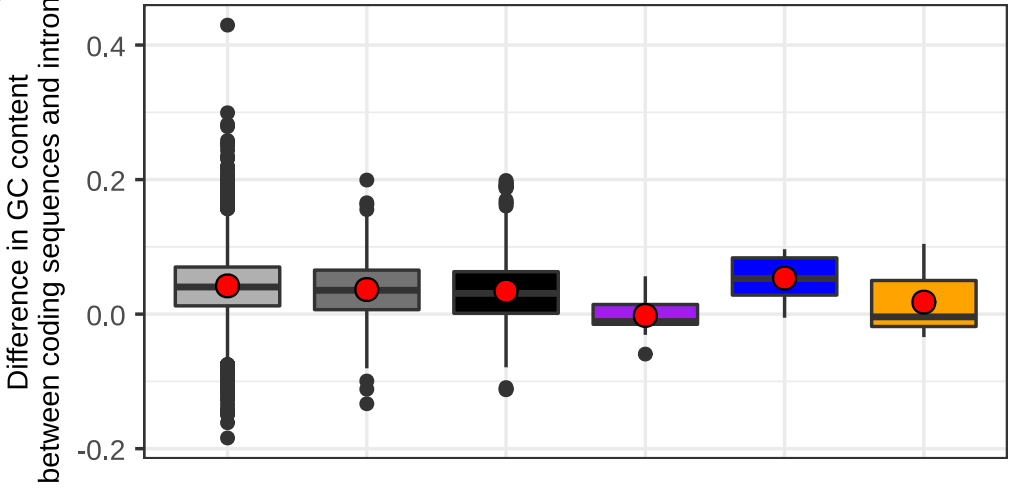




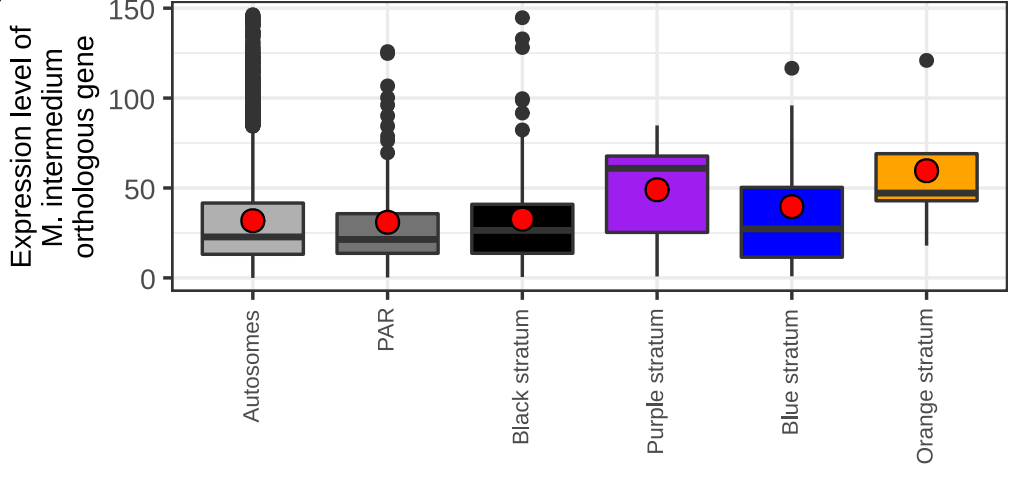
A)



B)

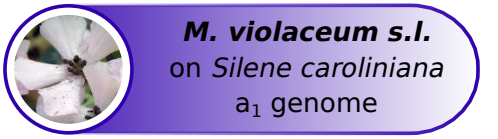


C)



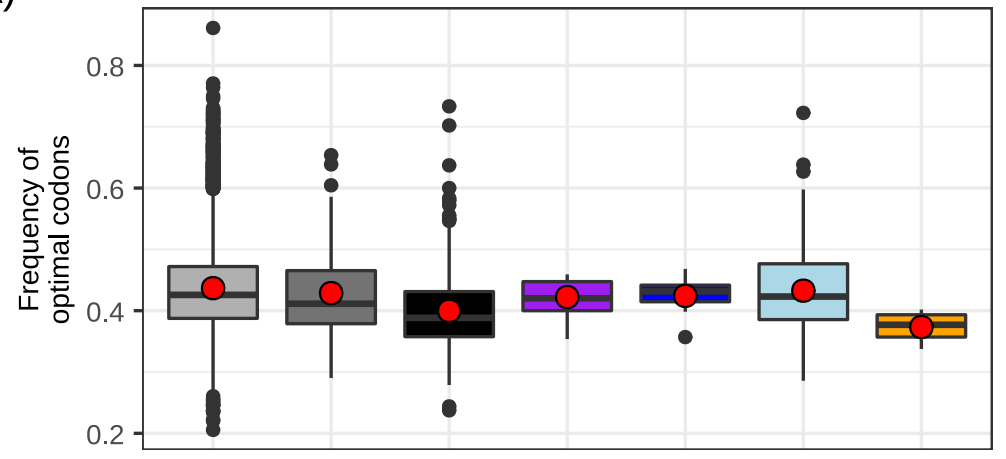
Genomic compartments

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

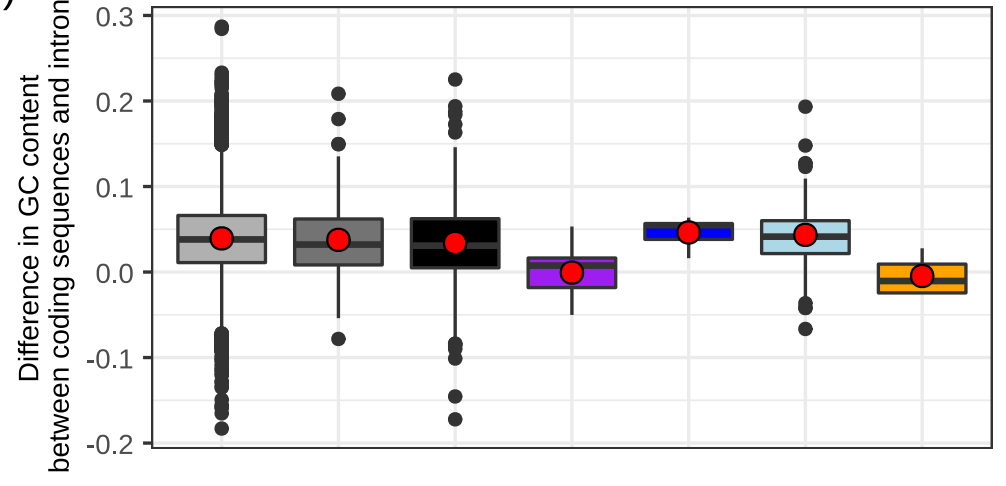


M. violaceum s.l.
on *Silene caroliniana*
a_1 genome

A)



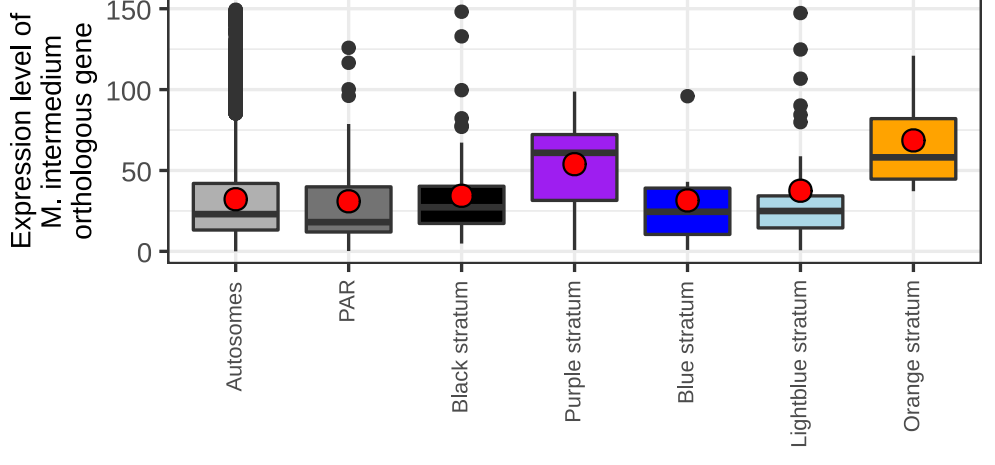
B)

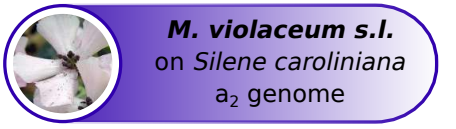


Genomic compartments

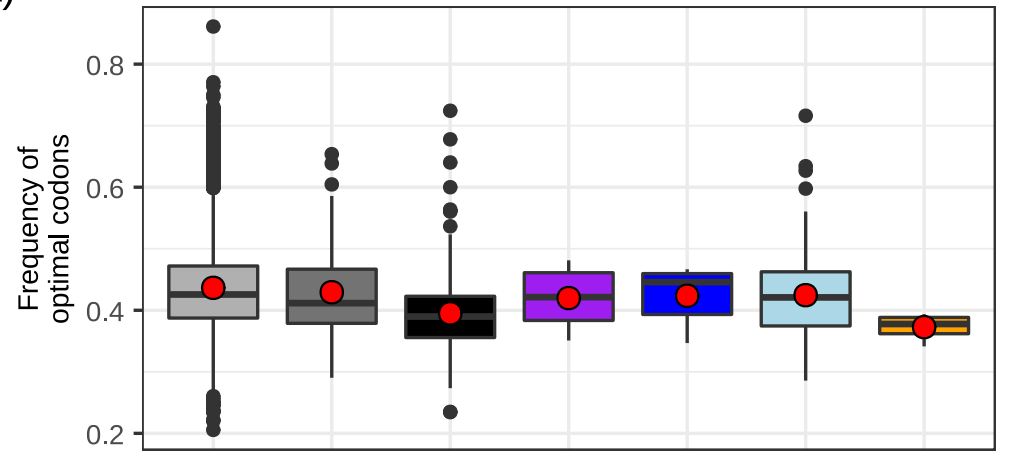
- Autosomes
- PAR
- Black stratum
- Purple stratum
- Blue stratum
- Lightblue stratum
- Orange stratum

C)

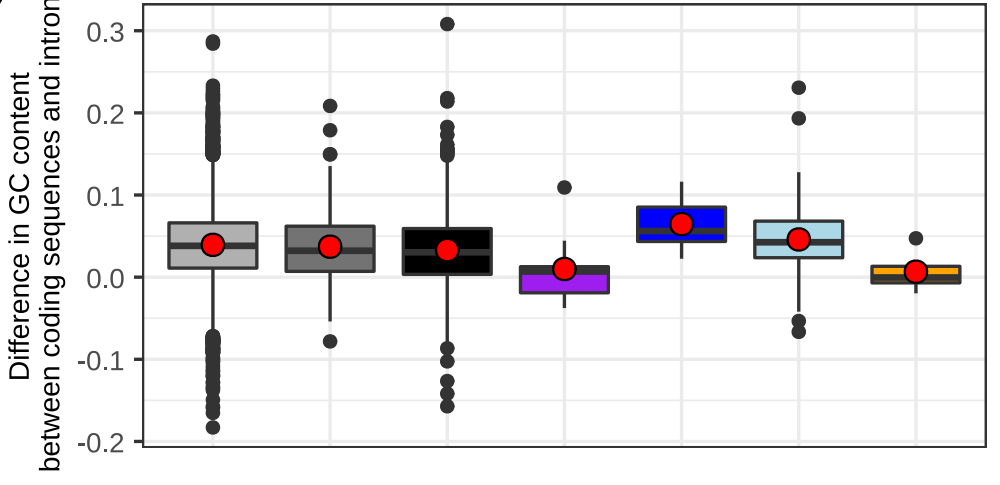




A)



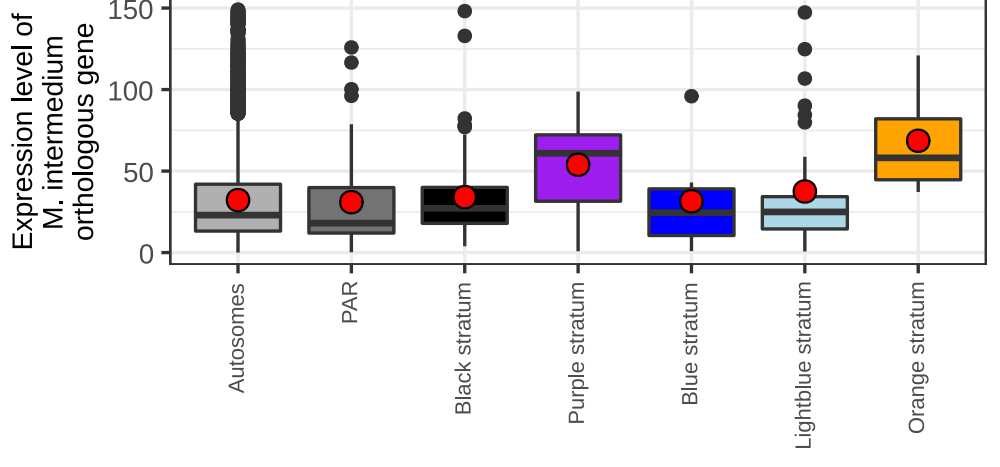
B)

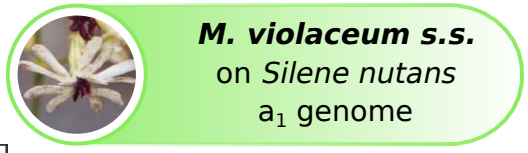


Genomic compartments

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

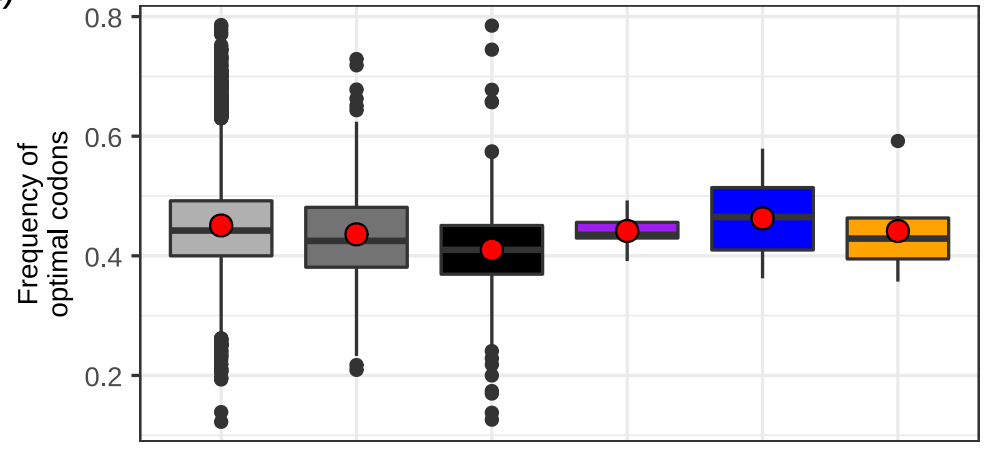
C)



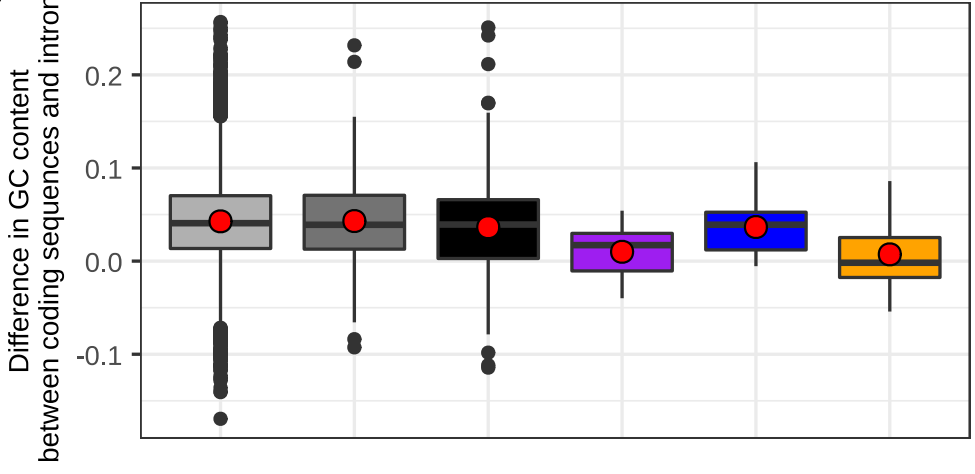


M. violaceum s.s.
on *Silene nutans*
a_1 genome

A)



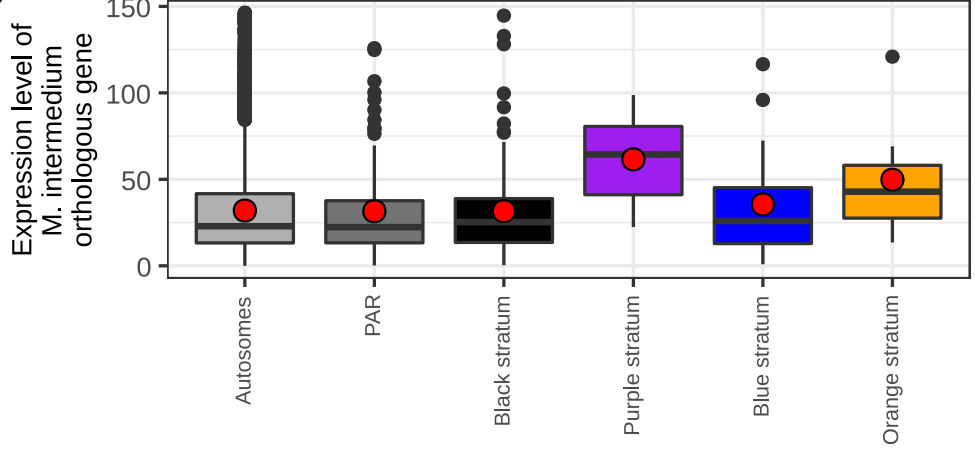
B)

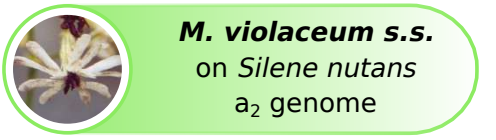


Genomic compartments

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

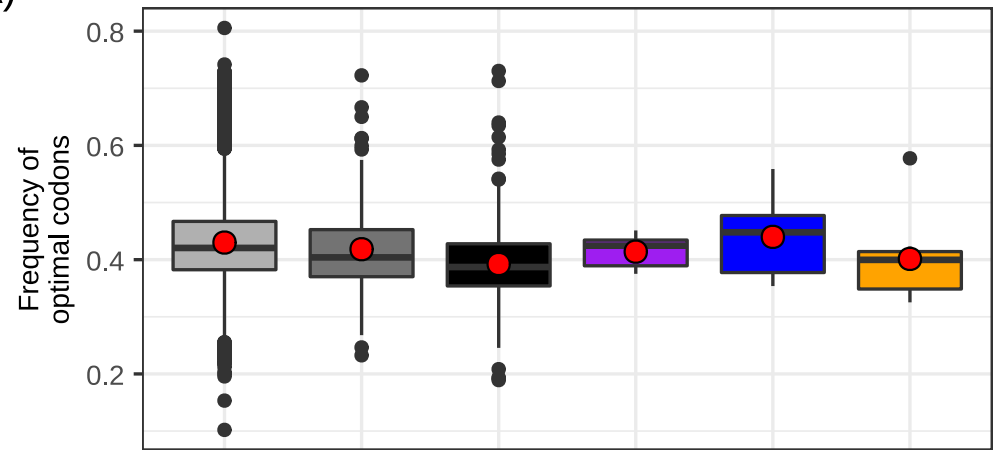
C)



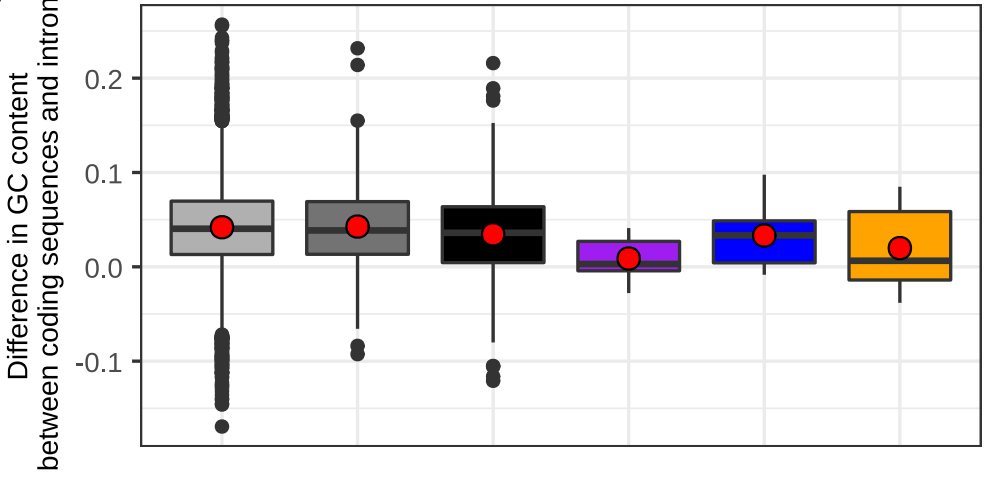


M. violaceum s.s.
on *Silene nutans*
a_2 genome

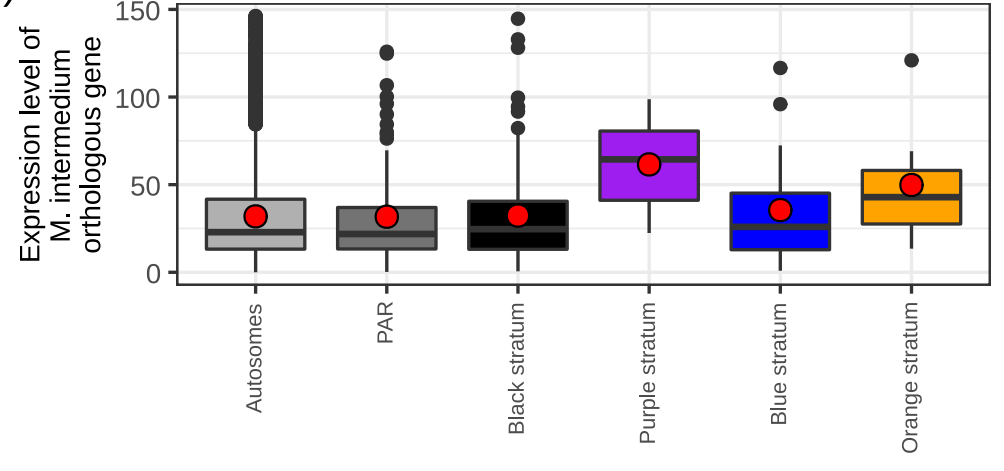
A)



B)

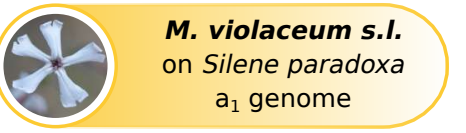


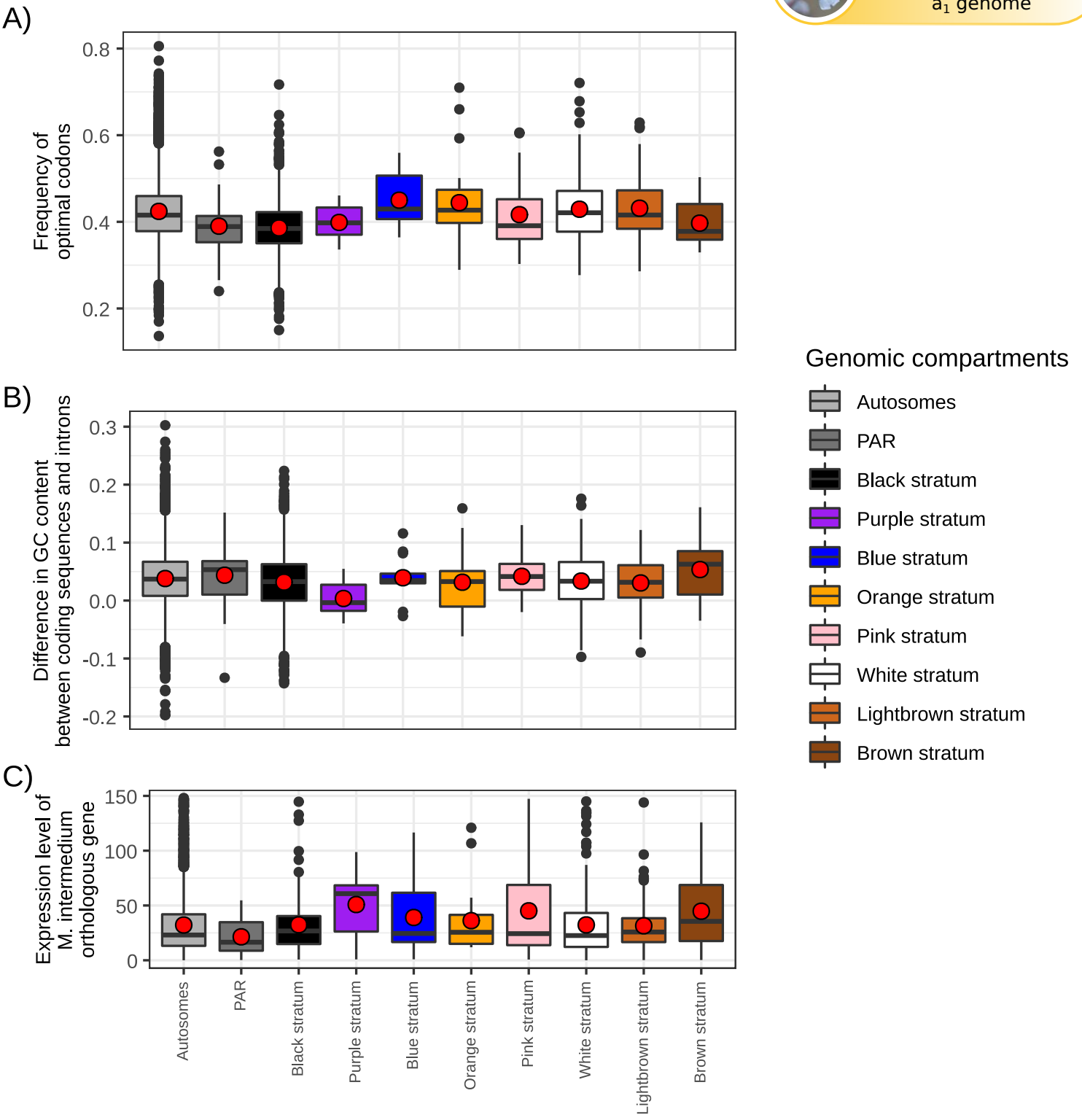
C)

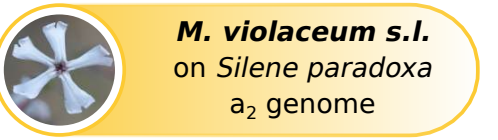


Genomic compartments

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

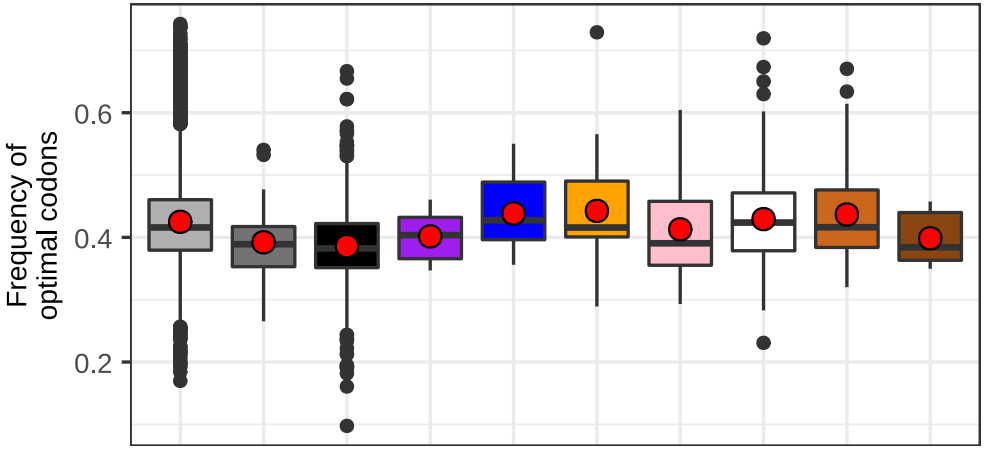

M. violaceum s.l.
on Silene paradoxa
a1 genome


A)
Frequency of optimal codons
0.8
0.6
0.4
0.2
B)
Difference in GC content between coding sequences and introns
0.3
0.2
0.1
0.0
-0.1
-0.2
C)
Expression level of M. intermedium orthologous gene
150
100
50
0
Autosomes
PAR
Black stratum
Purple stratum
Blue stratum
Orange stratum
Pink stratum
White stratum
Lightbrown stratum
Brown stratum
Genomic compartments
Autosomes
PAR
Black stratum
Purple stratum
Blue stratum
Orange stratum
Pink stratum
White stratum
Lightbrown stratum
Brown stratum

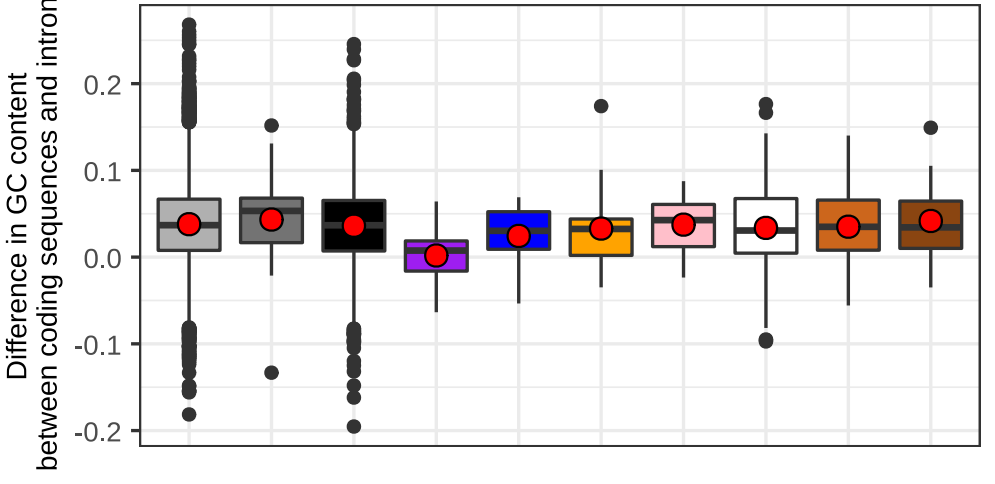


M. violaceum s.l.
on *Silene paradoxa*
a_2 genome

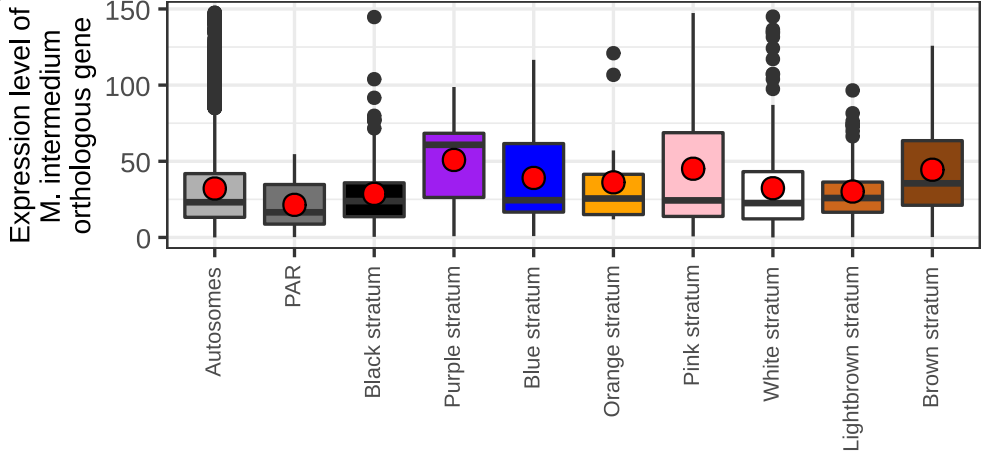
A)



B)

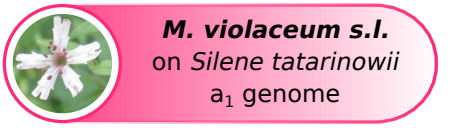


C)



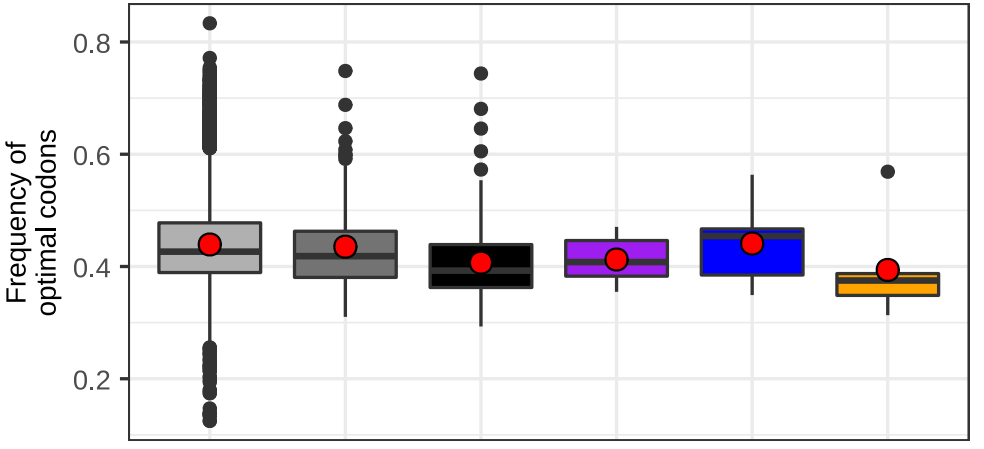
Genomic compartments

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

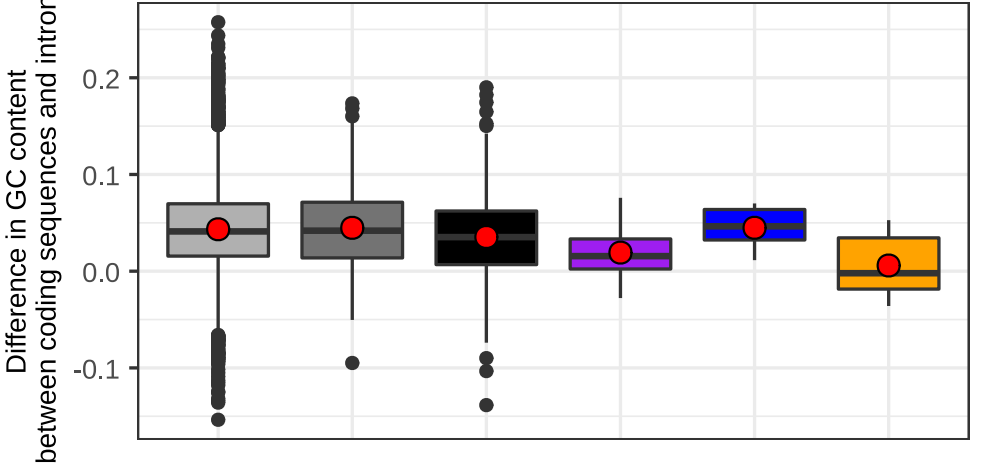


M. violaceum s.l.
on *Silene tatarinowii*
a_1 genome

A)



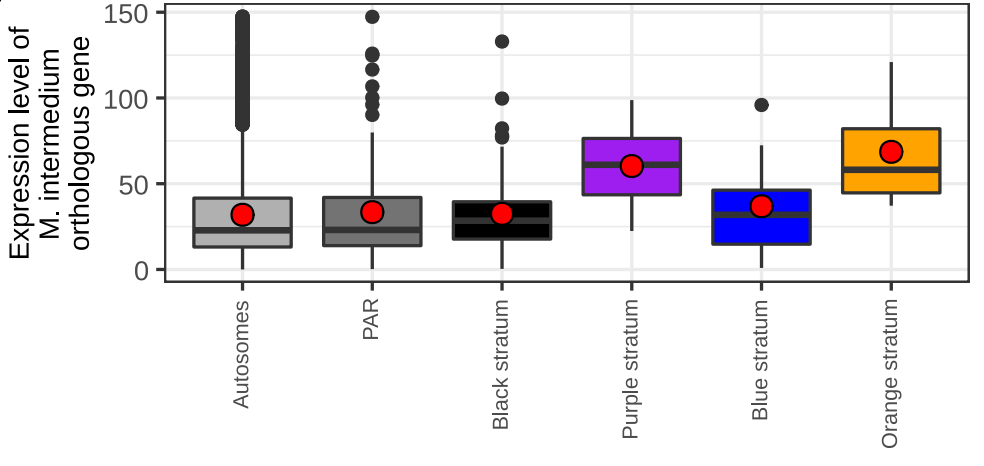
B)

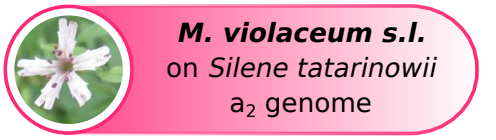


Genomic compartments

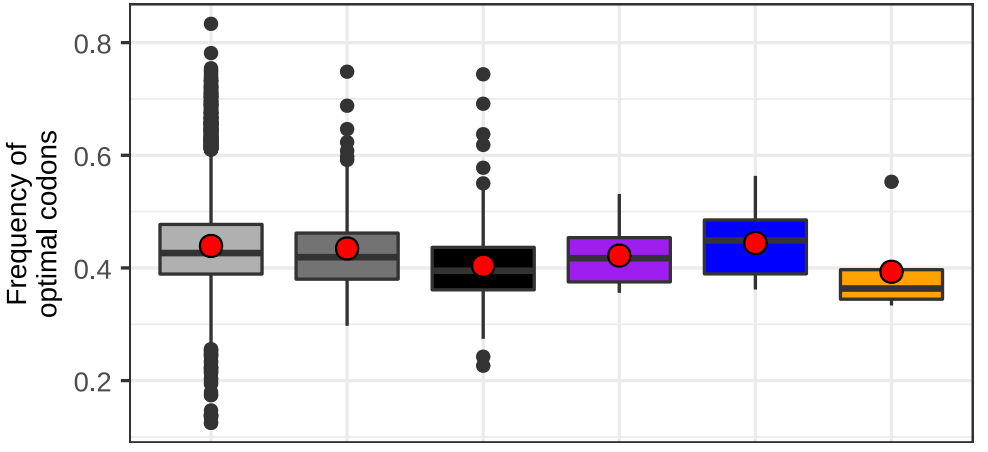
- Autosomes
- PAR
- Black stratum
- Purple stratum
- Blue stratum
- Orange stratum

C)

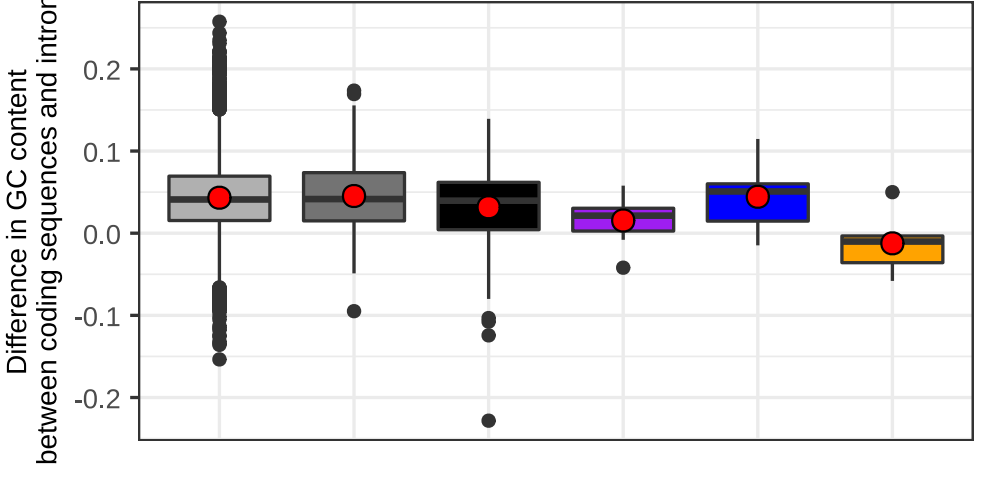




A)



B)



Genomic compartments

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

C)

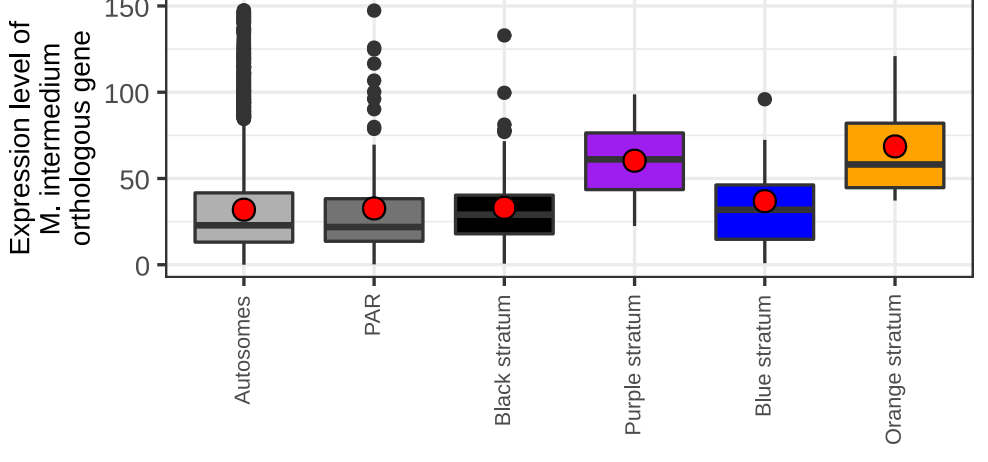
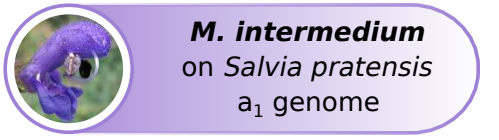
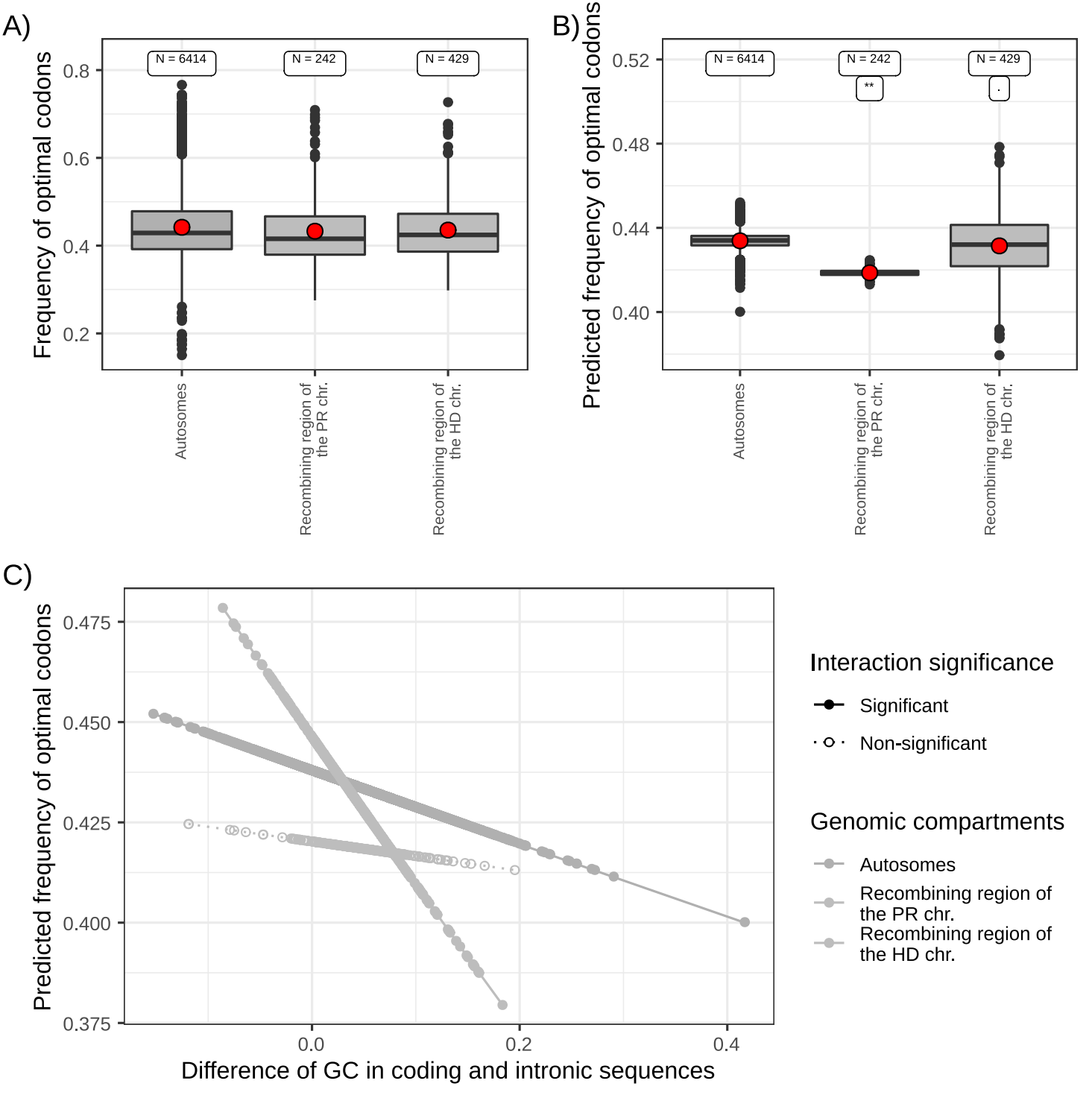


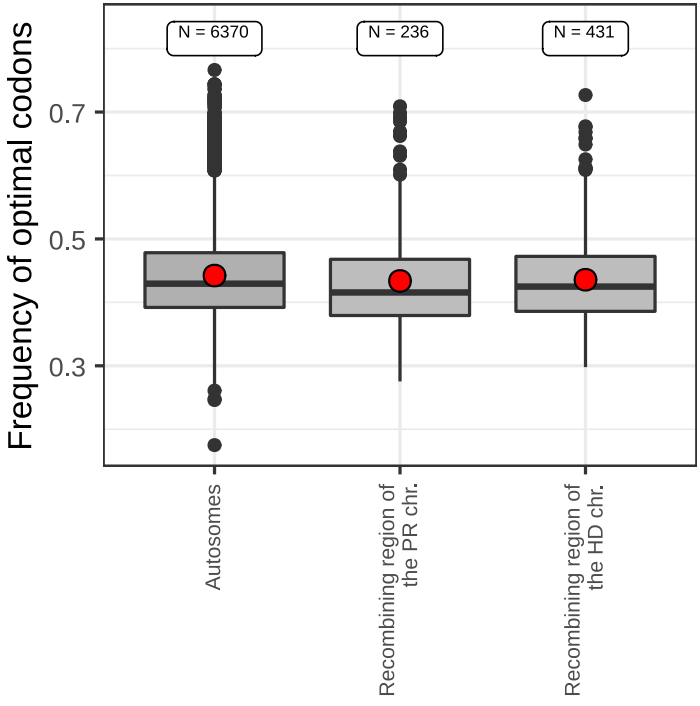
Figure S7


M. intermedium
on Salvia pratensis
a1 genome

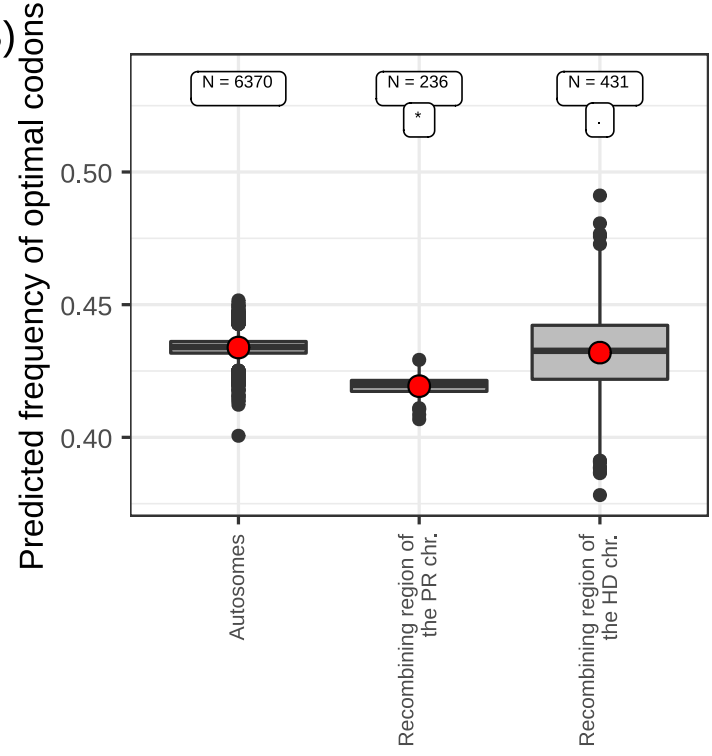

A)
Frequency of optimal codons
0.8
0.6
0.4
0.2
N = 6414
N = 242
N = 429
Autosomes
Recombining region of the PR chr.
Recombining region of the HD chr.
B)
Predicted frequency of optimal codons
0.52
0.48
0.44
0.40
N = 6414
N = 242
**
N = 429
.
Autosomes
Recombining region of the PR chr.
Recombining region of the HD chr.
C)
Predicted frequency of optimal codons
0.475
0.450
0.425
0.400
0.375
0.0
0.2
0.4
Difference of GC in coding and intronic sequences
Interaction significance
Significant
Non-significant
Genomic compartments
Autosomes
Recombining region of the PR chr.
Recombining region of the HD chr.

M. intermedium
on Salvia pratensis
a2 genome

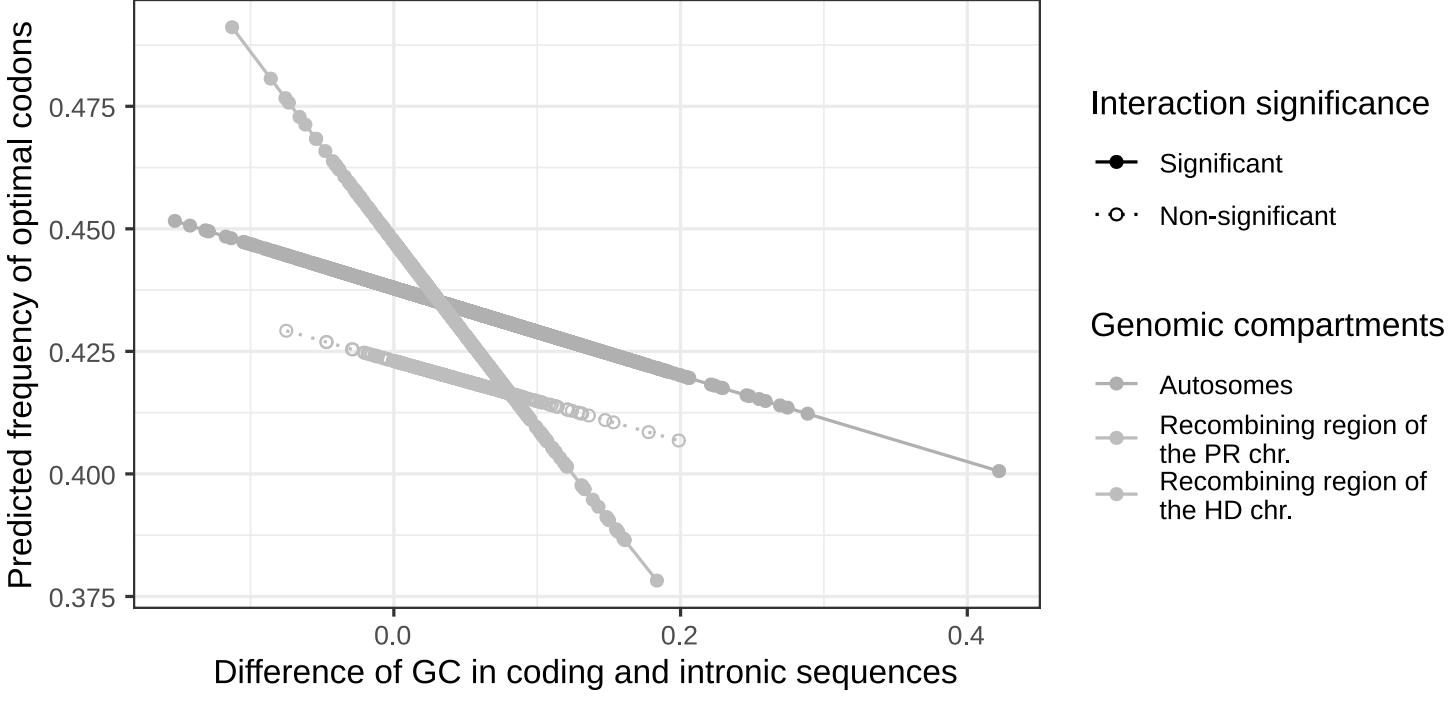
A)

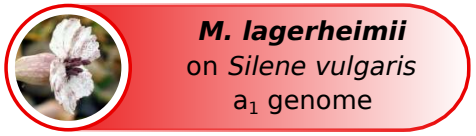

Frequency of optimal codons
0.7
0.5
0.3
N = 6370
N = 236
N = 431
Autosomes
Recombining region of the PR chr.
Recombining region of the HD chr.

B)

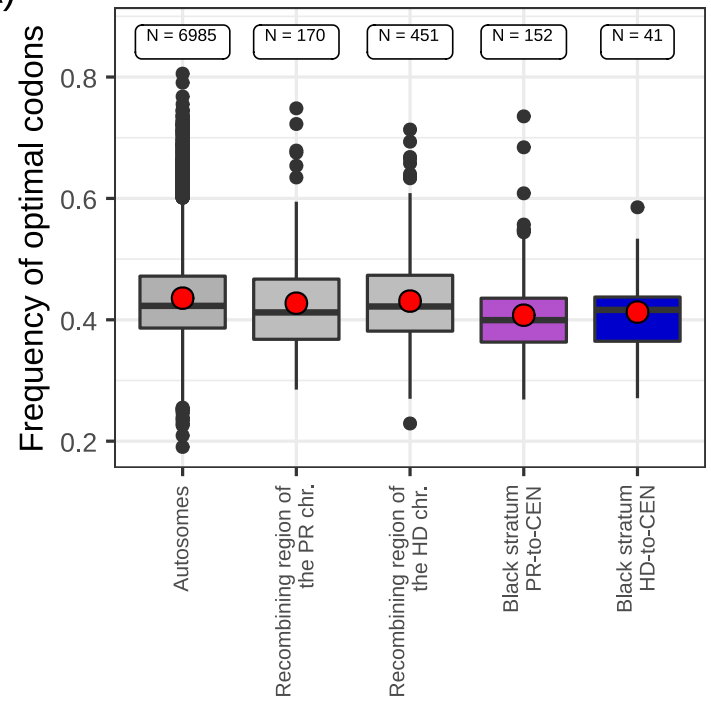

Predicted frequency of optimal codons
0.50
0.45
0.40
N = 6370
N = 236
N = 431
*
.
Autosomes
Recombining region of the PR chr.
Recombining region of the HD chr.

C)

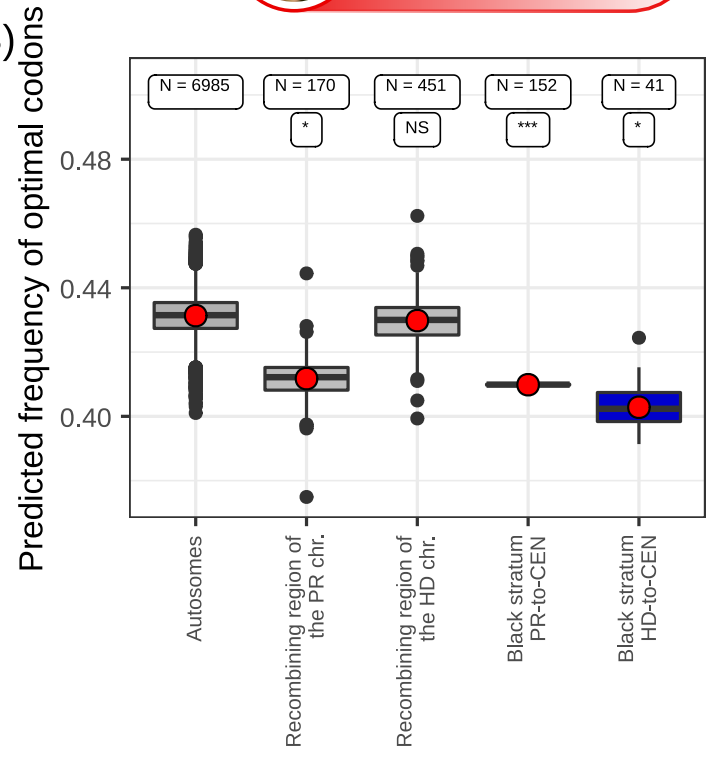

Predicted frequency of optimal codons
0.475
0.450
0.425
0.400
0.375
0.0
0.2
0.4
Difference of GC in coding and intronic sequences
Interaction significance
Significant
Non-significant
Genomic compartments
Autosomes
Recombining region of the PR chr.
Recombining region of the HD chr.


M. lagerheimii
on Silene vulgaris
a1 genome

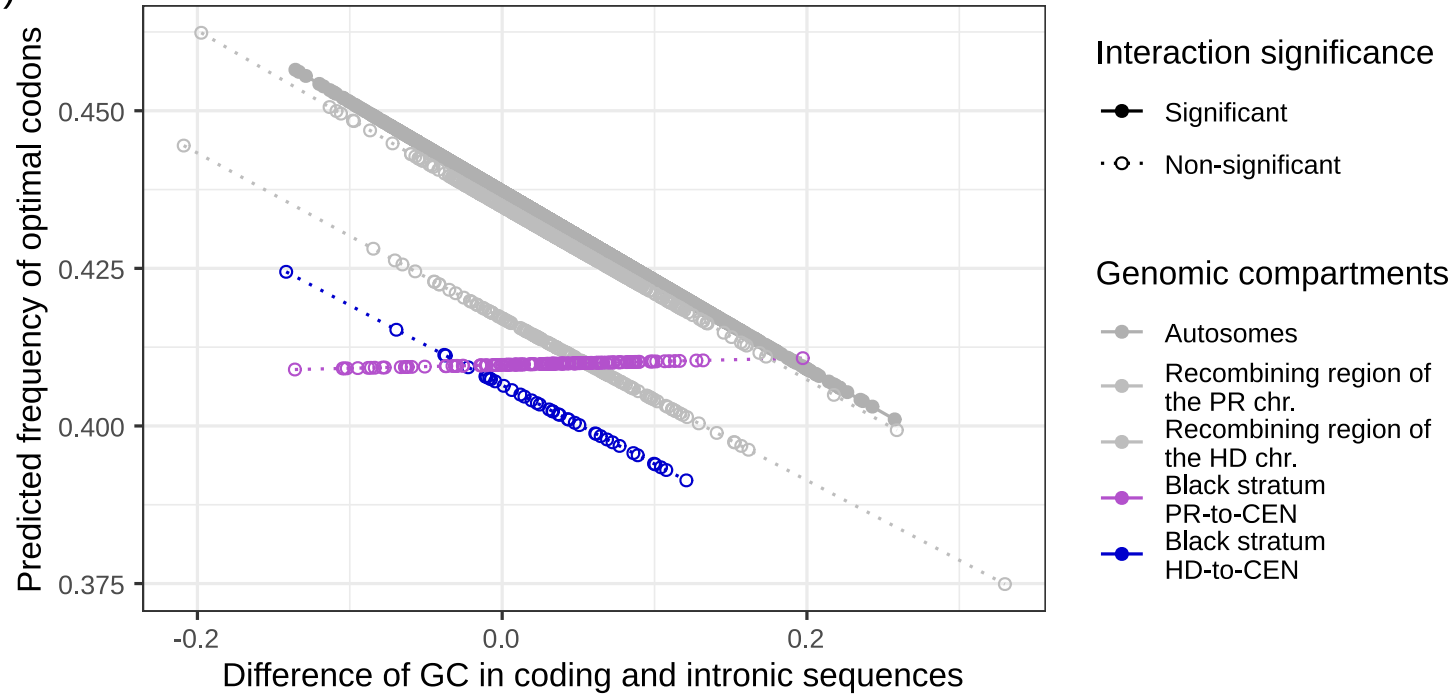
A)

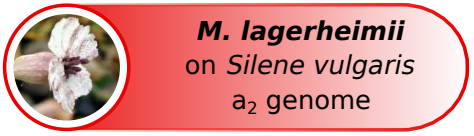

Frequency of optimal codons
0.8
0.6
0.4
0.2
N = 6985
N = 170
N = 451
N = 152
N = 41
Autosomes
Recombining region of the PR chr.
Recombining region of the HD chr.
Black stratum PR-to-CEN
Black stratum HD-to-CEN

B)

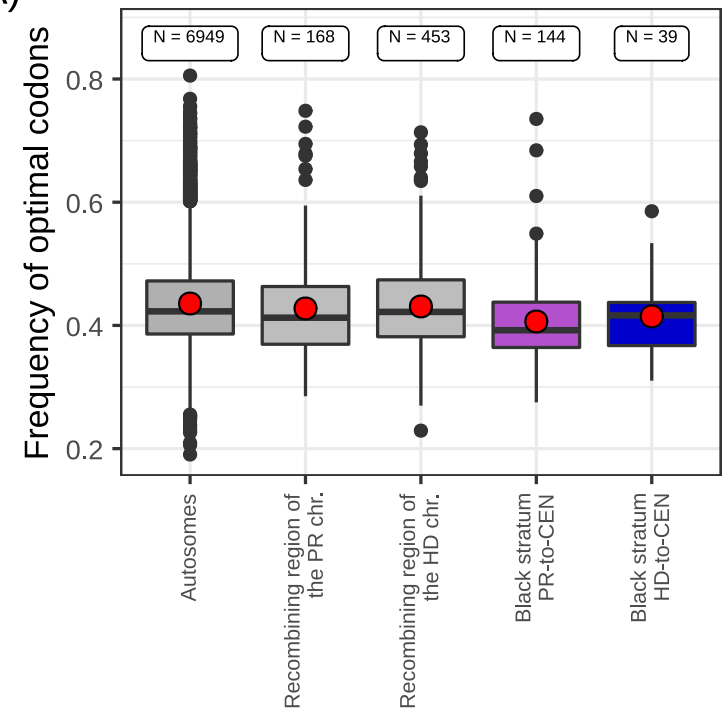

Predicted frequency of optimal codons
0.48
0.44
0.40
N = 6985
N = 170
N = 451
N = 152
N = 41
*
NS
***
*
Autosomes
Recombining region of the PR chr.
Recombining region of the HD chr.
Black stratum PR-to-CEN
Black stratum HD-to-CEN

C)

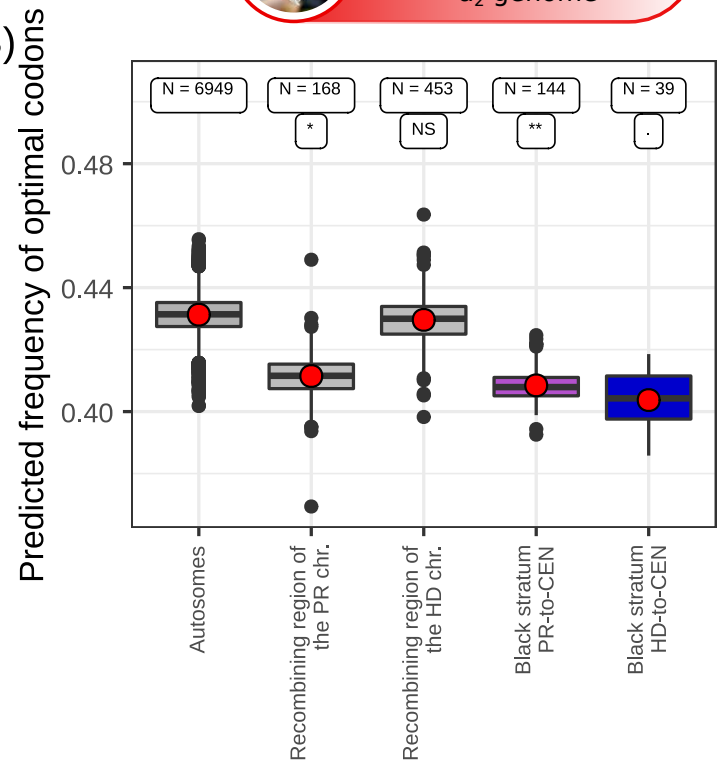

Predicted frequency of optimal codons
0.450
0.425
0.400
0.375
-0.2
0.0
0.2
Difference of GC in coding and intronic sequences
Interaction significance
Significant
Non-significant
Genomic compartments
Autosomes
Recombining region of the PR chr.
Recombining region of the HD chr.
Black stratum PR-to-CEN
Black stratum HD-to-CEN


M. lagerheimii
on Silene vulgaris
a2 genome

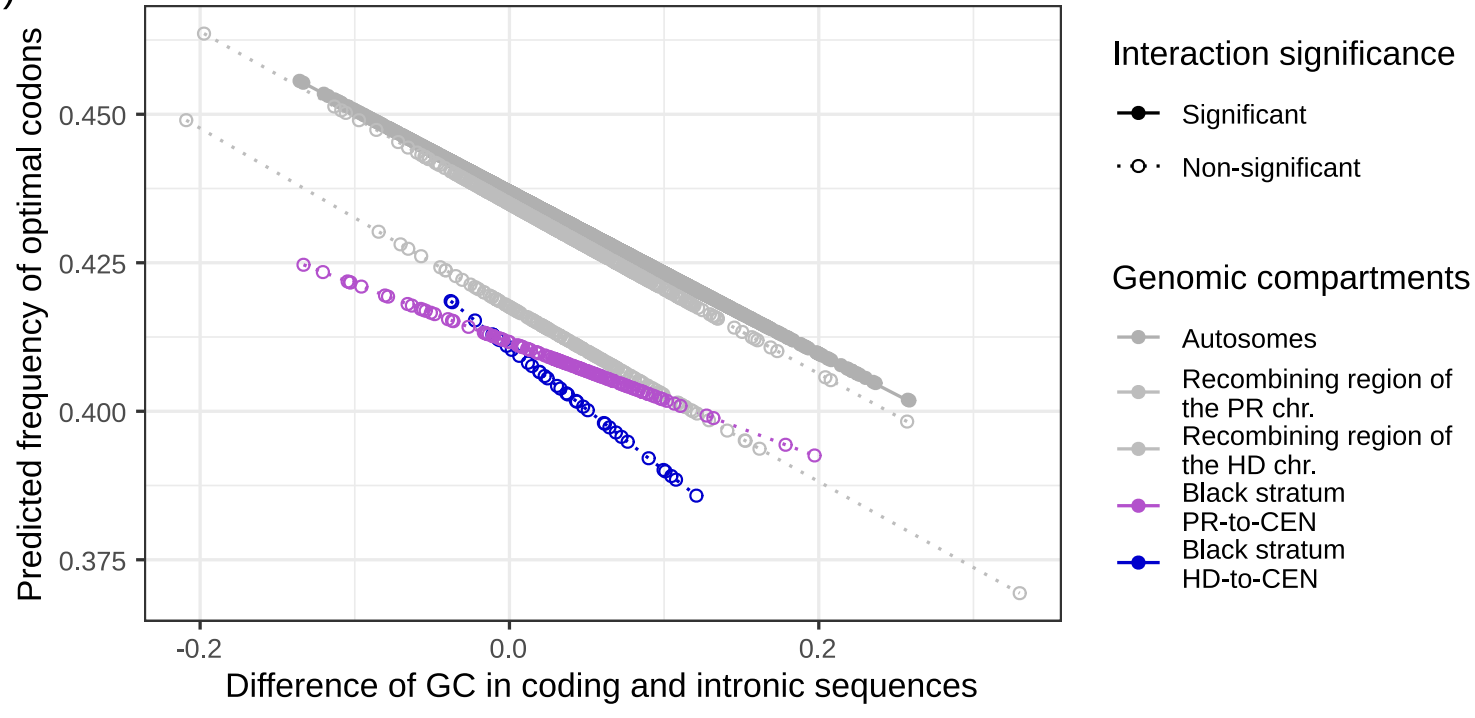
A)

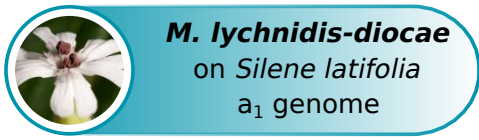

Frequency of optimal codons
0.8
0.6
0.4
0.2
N = 6949
N = 168
N = 453
N = 144
N = 39
Autosomes
Recombining region of the PR chr.
Recombining region of the HD chr.
Black stratum PR-to-CEN
Black stratum HD-to-CEN

B)

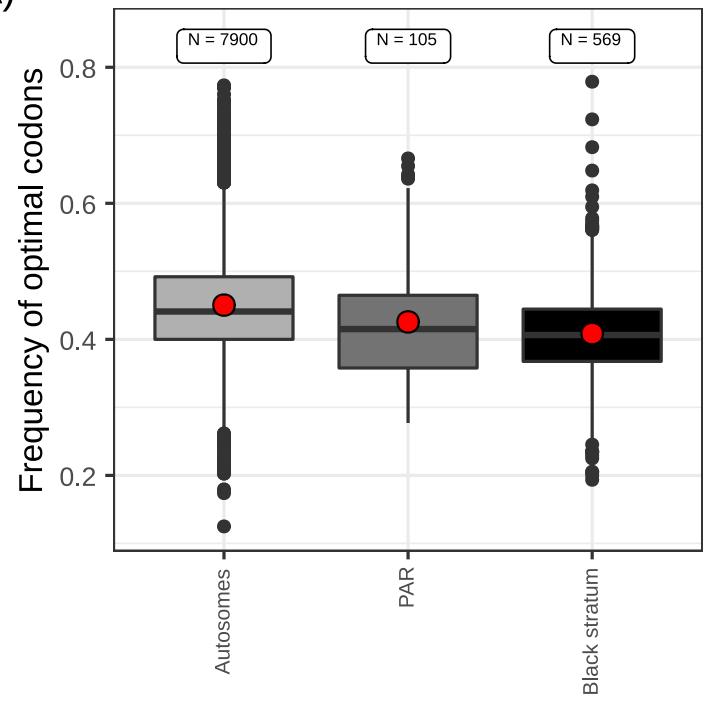

Predicted frequency of optimal codons
0.48
0.44
0.40
N = 6949
N = 168
N = 453
N = 144
N = 39
*
NS
**
.
Autosomes
Recombining region of the PR chr.
Recombining region of the HD chr.
Black stratum PR-to-CEN
Black stratum HD-to-CEN

C)

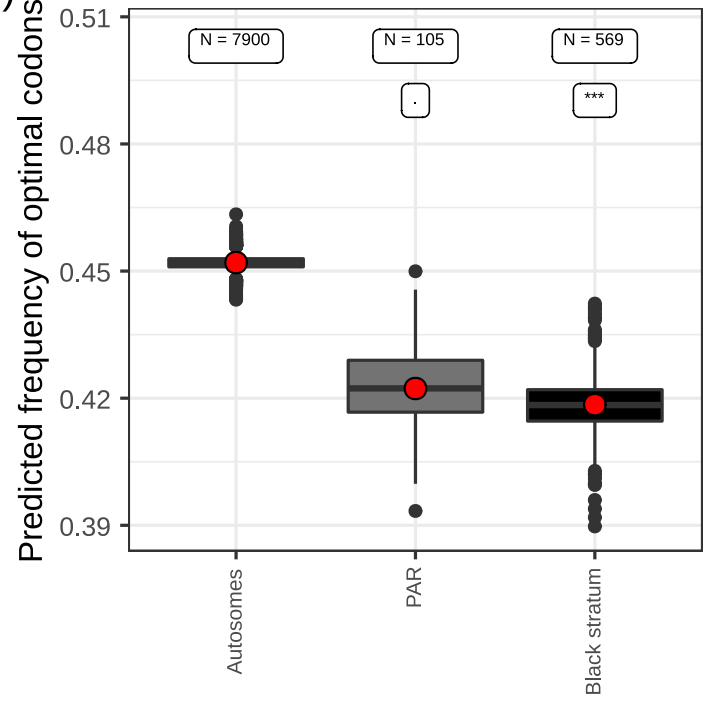

Predicted frequency of optimal codons
0.450
0.425
0.400
0.375
-0.2
0.0
0.2
Difference of GC in coding and intronic sequences
Interaction significance
Significant
Non-significant
Genomic compartments
Autosomes
Recombining region of the PR chr.
Recombining region of the HD chr.
Black stratum PR-to-CEN
Black stratum HD-to-CEN


M. lychnidis-diocae
on Silene latifolia
a1 genome

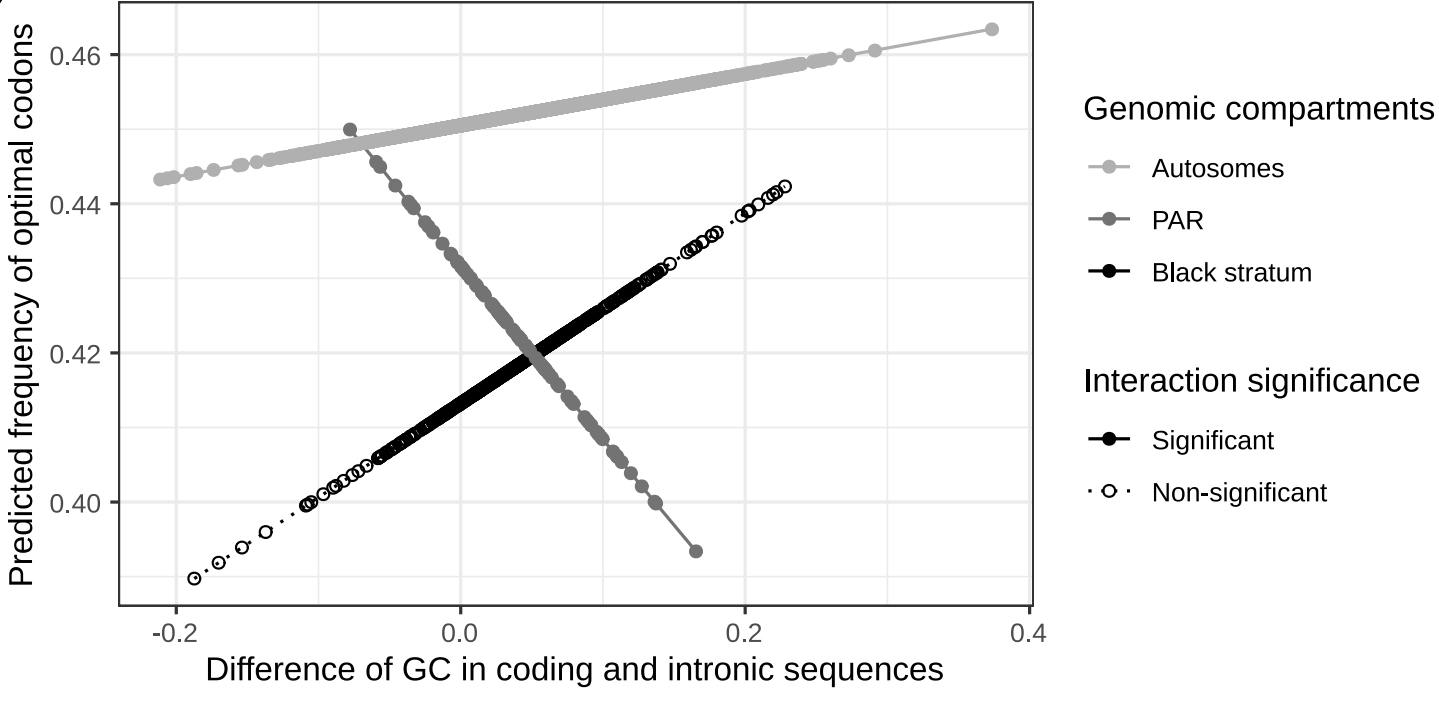
A)

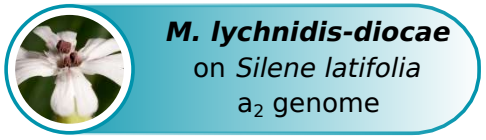

N = 7900
N = 105
N = 569
Frequency of optimal codons
0.8
0.6
0.4
0.2
Autosomes
PAR
Black stratum

B)

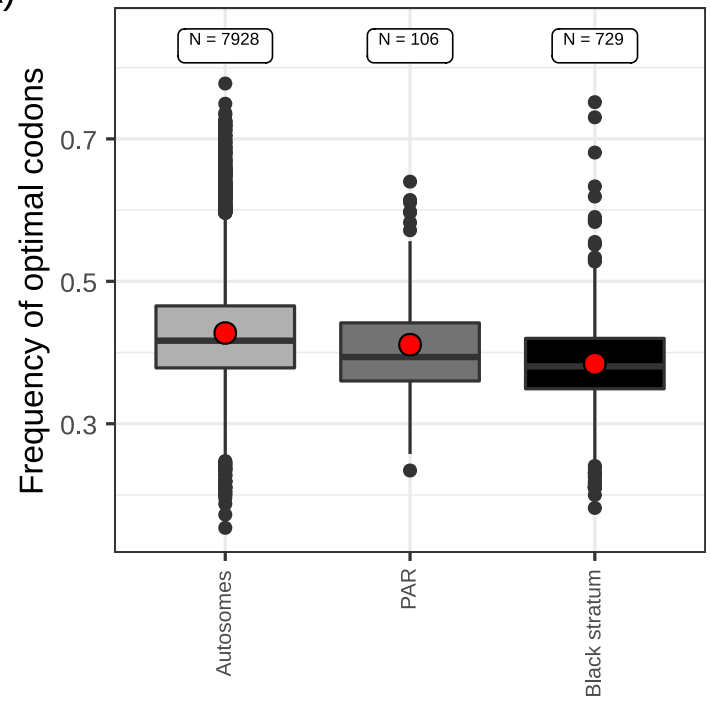

N = 7900
N = 105
N = 569
.
***
Predicted frequency of optimal codons
0.51
0.48
0.45
0.42
0.39
Autosomes
PAR
Black stratum

C)

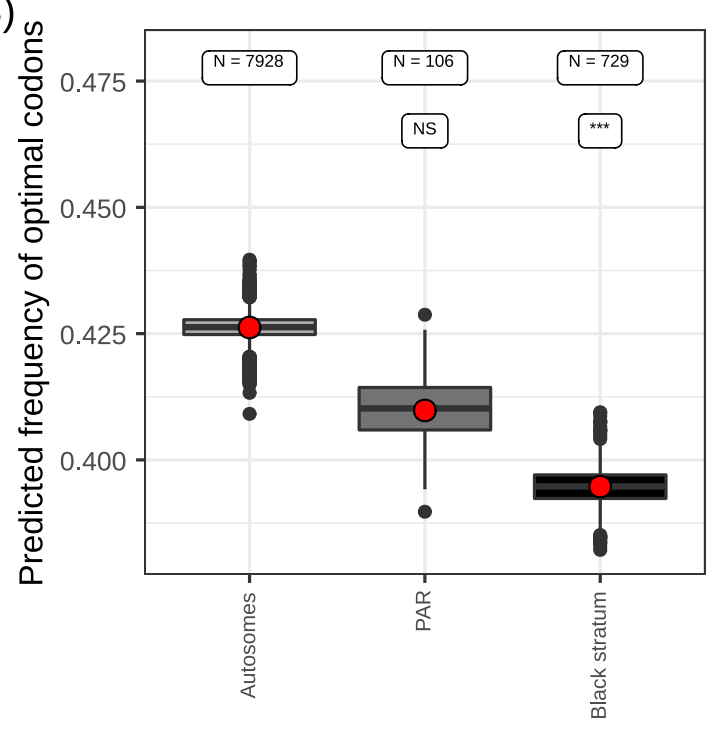

Predicted frequency of optimal codons
0.46
0.44
0.42
0.40
-0.2
0.0
0.2
0.4
Difference of GC in coding and intronic sequences
Genomic compartments
Autosomes
PAR
Black stratum
Interaction significance
Significant
Non-significant


M. lychnidis-diocae
on Silene latifolia
a2 genome

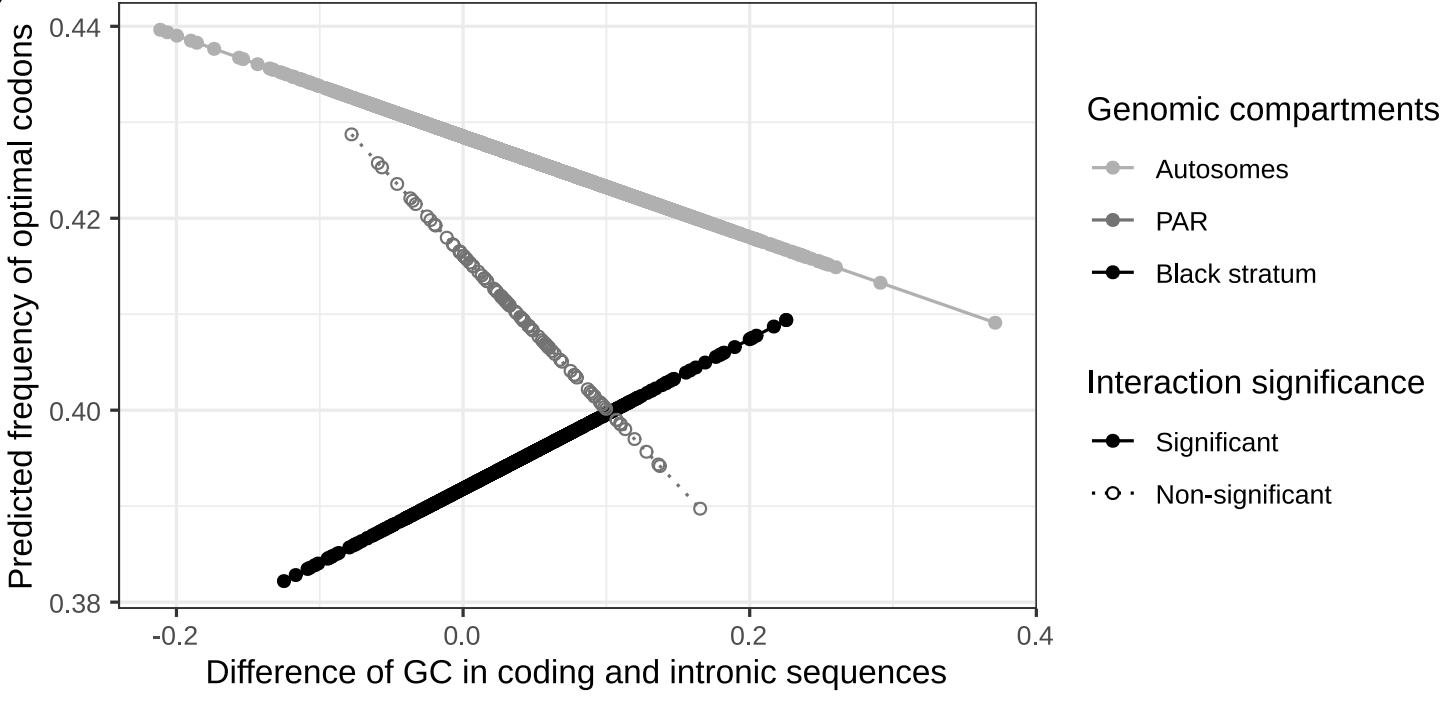
A)

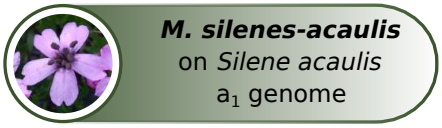

N = 7928
N = 106
N = 729
Frequency of optimal codons
0.7
0.5
0.3
Autosomes
PAR
Black stratum

B)

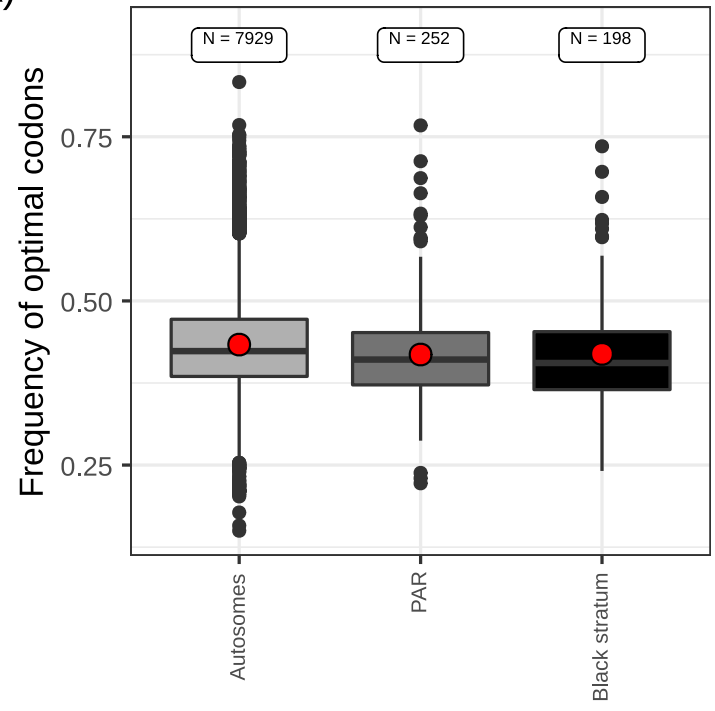

N = 7928
N = 106
N = 729
NS
***
Predicted frequency of optimal codons
0.475
0.450
0.425
0.400
Autosomes
PAR
Black stratum

C)

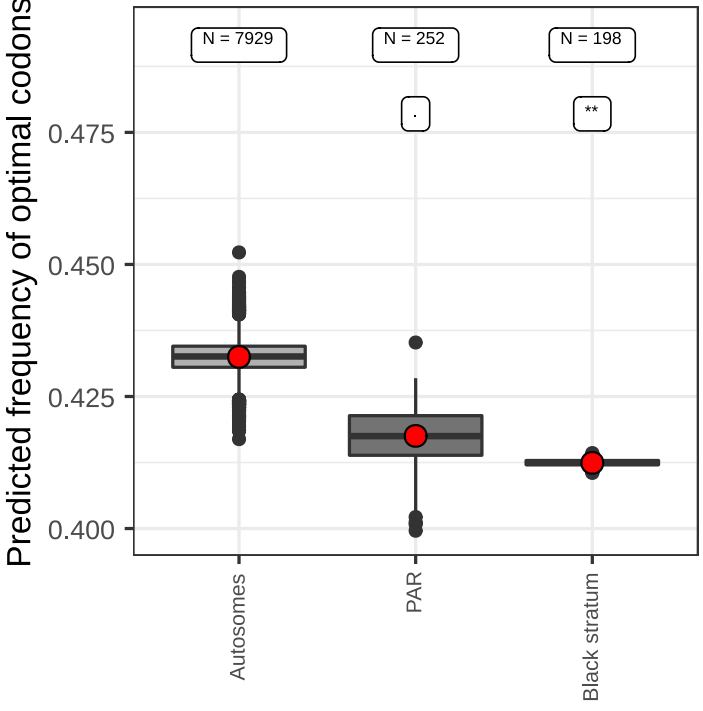

Predicted frequency of optimal codons
0.44
0.42
0.40
0.38
-0.2
0.0
0.2
0.4
Difference of GC in coding and intronic sequences
Genomic compartments
Autosomes
PAR
Black stratum
Interaction significance
Significant
Non-significant


M. silenes-acaulis
on Silene acaulis
a1 genome

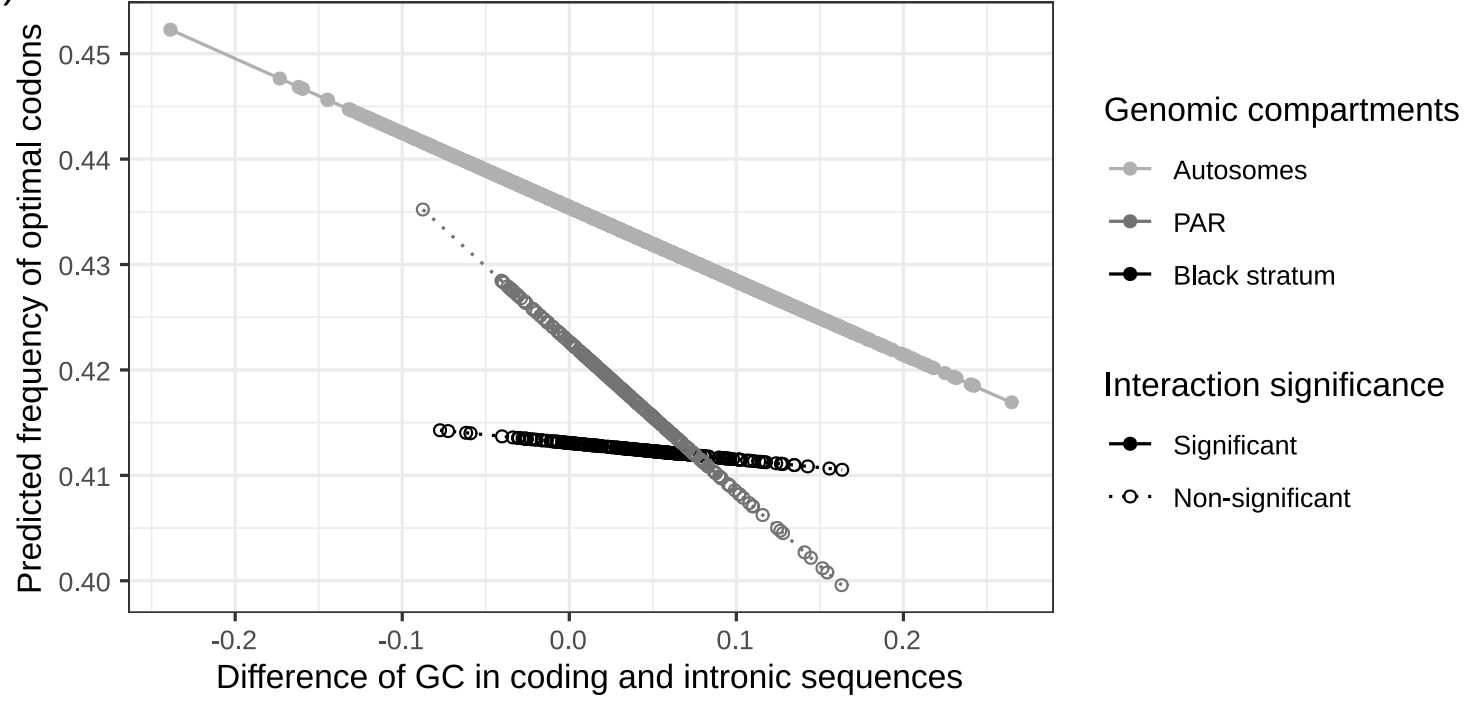
A)

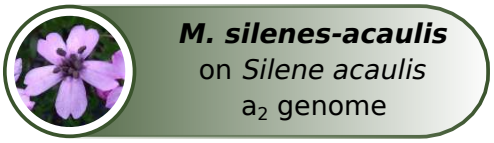

N = 7929
N = 252
N = 198
Frequency of optimal codons
0.75
0.50
0.25
Autosomes
PAR
Black stratum

B)

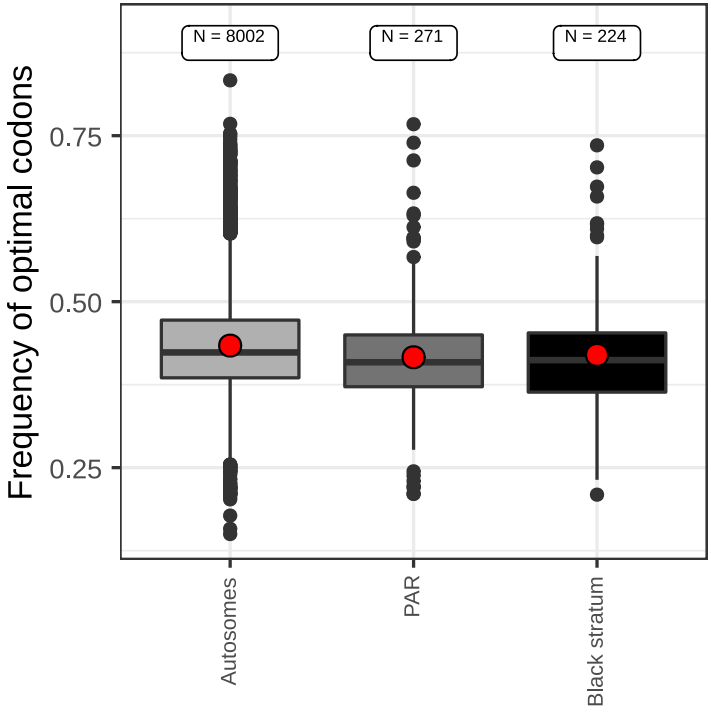

N = 7929
N = 252
N = 198
.
**
Predicted frequency of optimal codons
0.475
0.450
0.425
0.400
Autosomes
PAR
Black stratum

C)

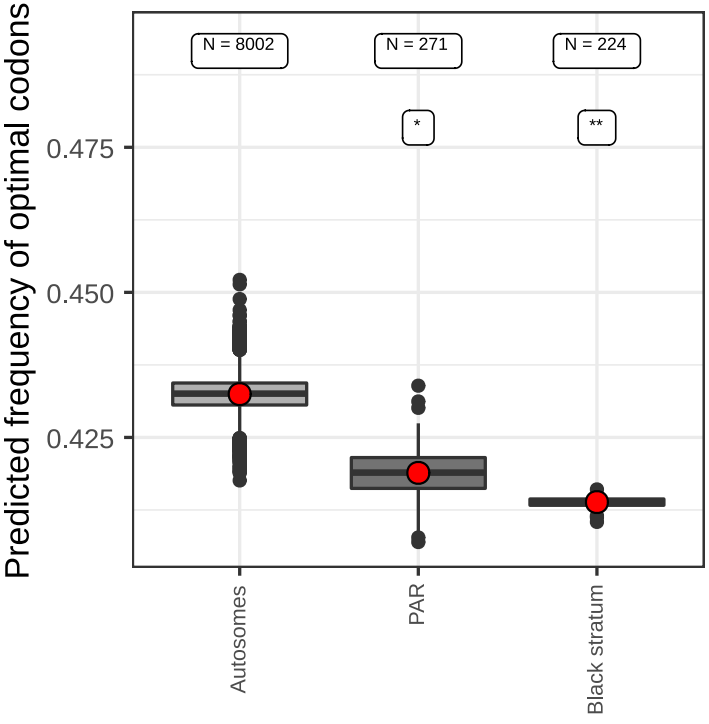

Predicted frequency of optimal codons
0.45
0.44
0.43
0.42
0.41
0.40
-0.2
-0.1
0.0
0.1
0.2
Difference of GC in coding and intronic sequences
Genomic compartments
Autosomes
PAR
Black stratum
Interaction significance
Significant
Non-significant


M. silenes-acaulis
on Silene acaulis
a2 genome

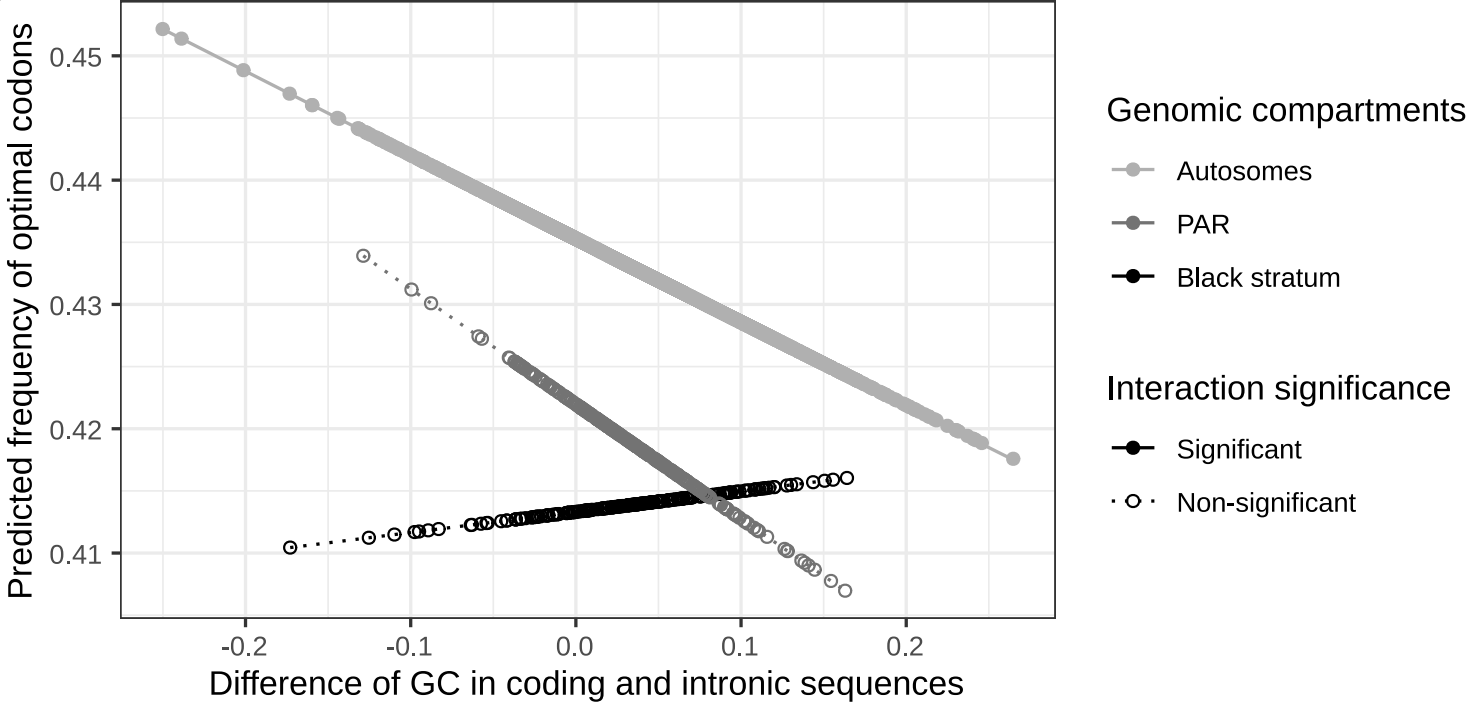
A)

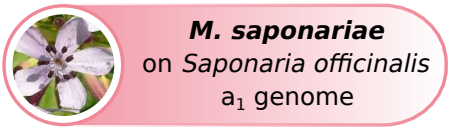

N = 8002
N = 271
N = 224
Frequency of optimal codons
0.75
0.50
0.25
Autosomes
PAR
Black stratum

B)


N = 8002
N = 271
N = 224
*
**
Predicted frequency of optimal codons
0.475
0.450
0.425
Autosomes
PAR
Black stratum

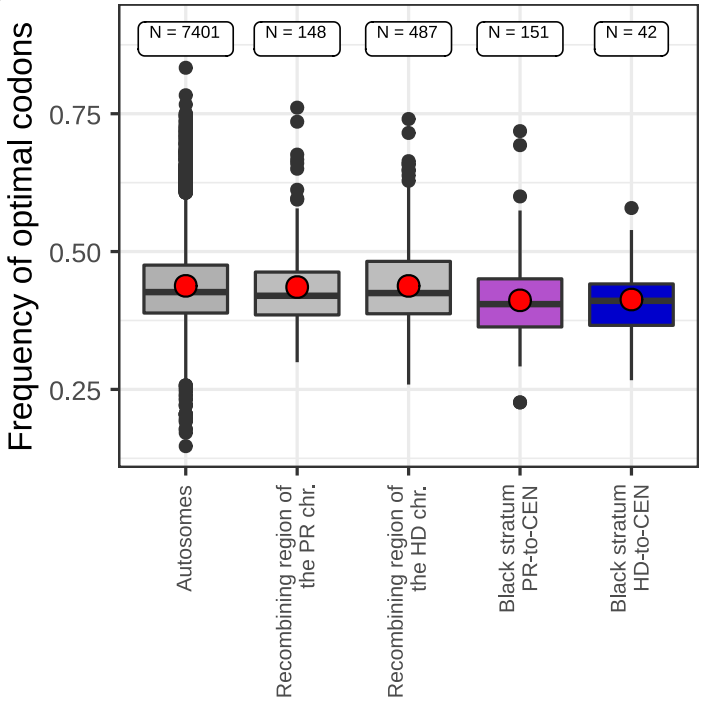
C)


Predicted frequency of optimal codons
0.45
0.44
0.43
0.42
0.41
-0.2
-0.1
0.0
0.1
0.2
Difference of GC in coding and intronic sequences
Genomic compartments
Autosomes
PAR
Black stratum
Interaction significance
Significant
Non-significant

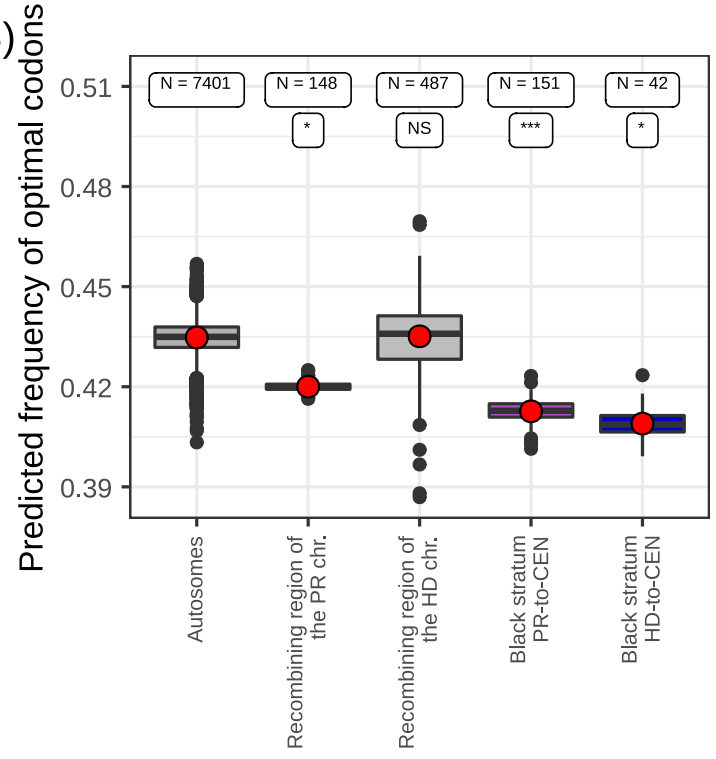


M. saponariae
on *Saponaria officinalis*
a_1 genome

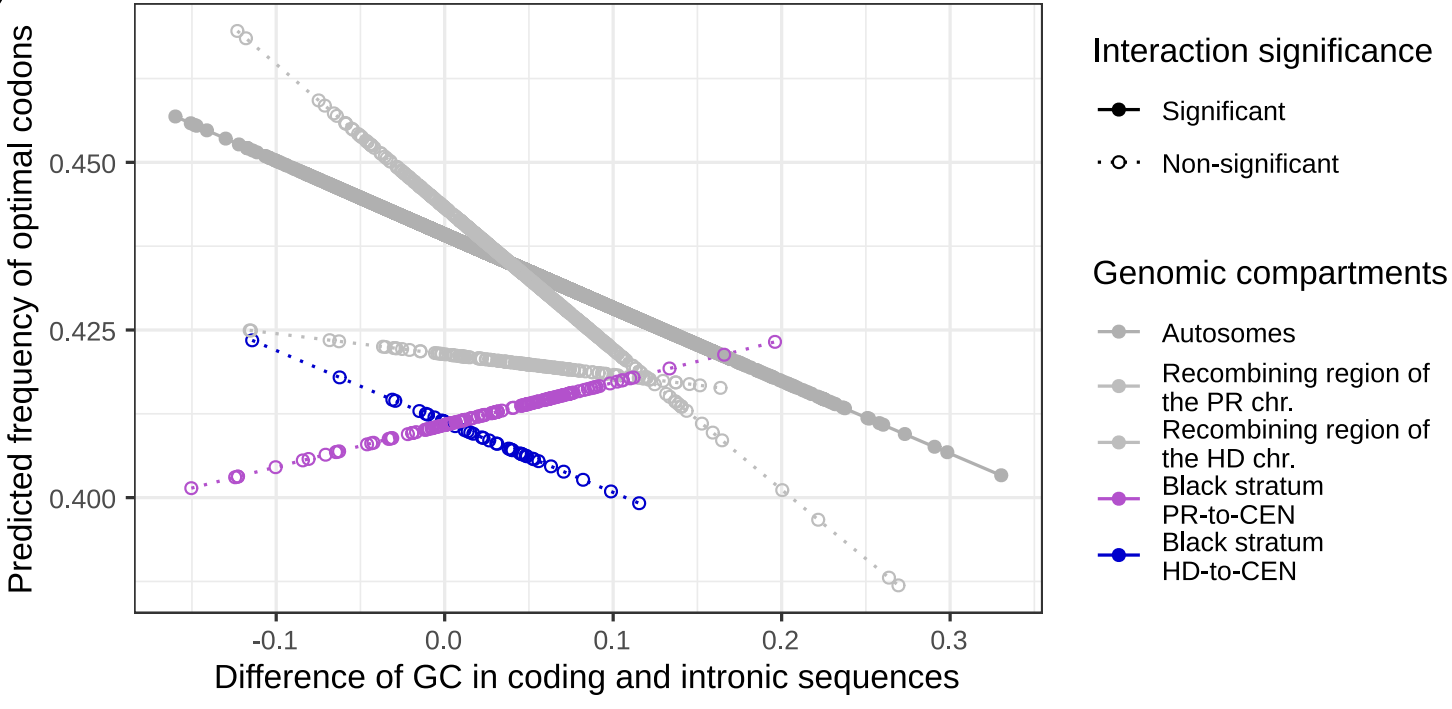
A)

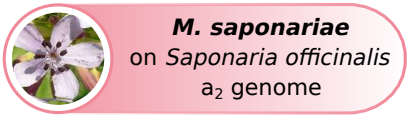


B)

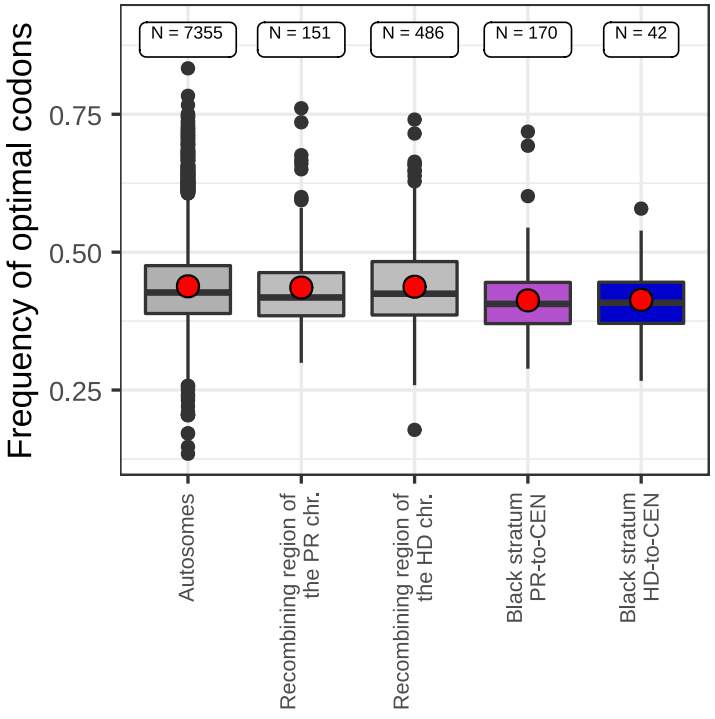


C)

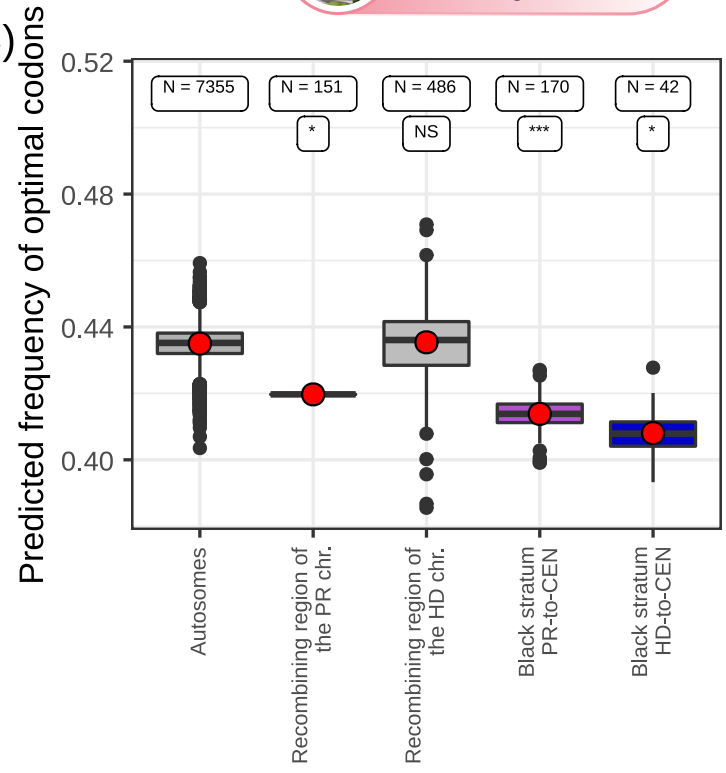



M. saponariae
on Saponaria officinalis
a2 genome

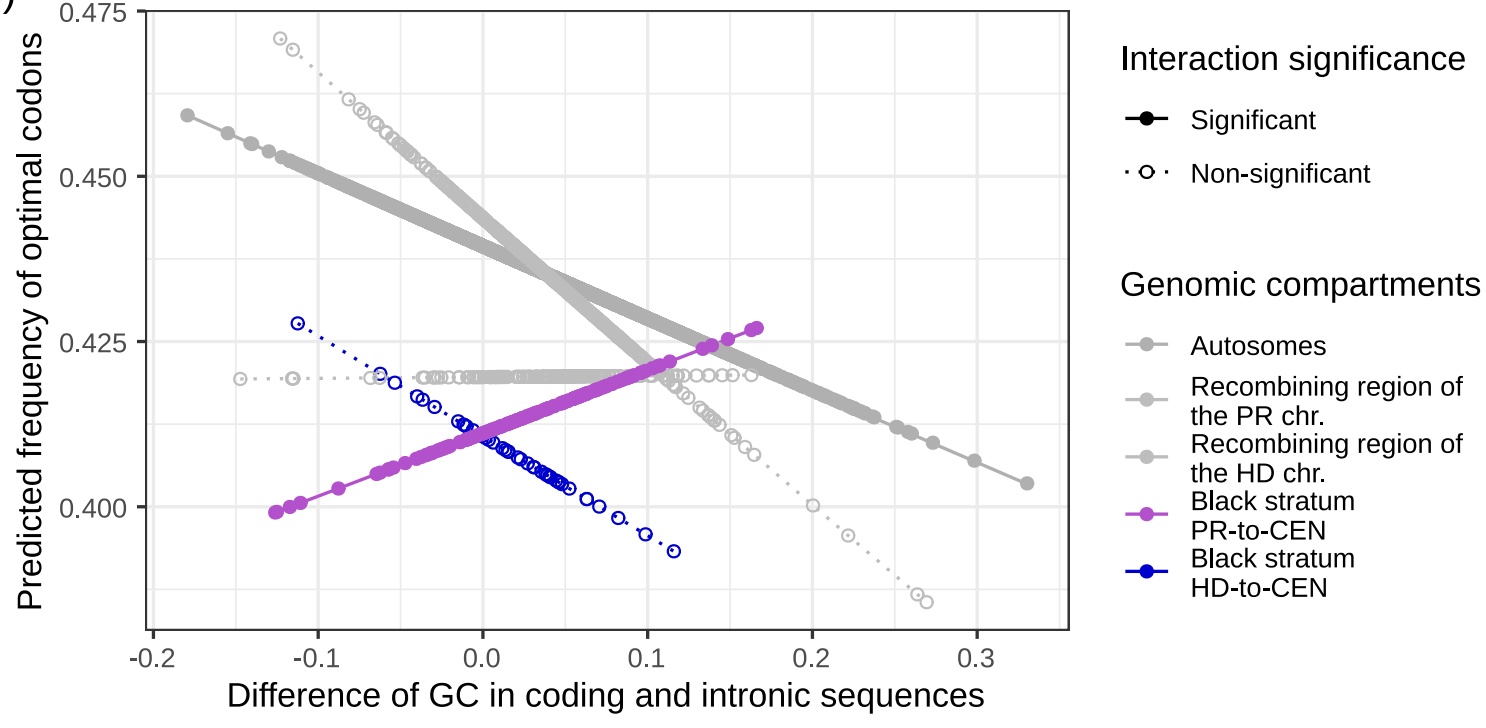
A)

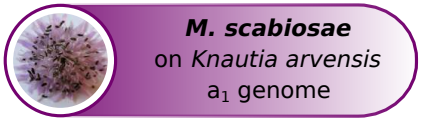

Frequency of optimal codons
0.75
0.50
0.25
N = 7355
N = 151
N = 486
N = 170
N = 42
Autosomes
Recombining region of the PR chr.
Recombining region of the HD chr.
Black stratum PR-to-CEN
Black stratum HD-to-CEN

B)

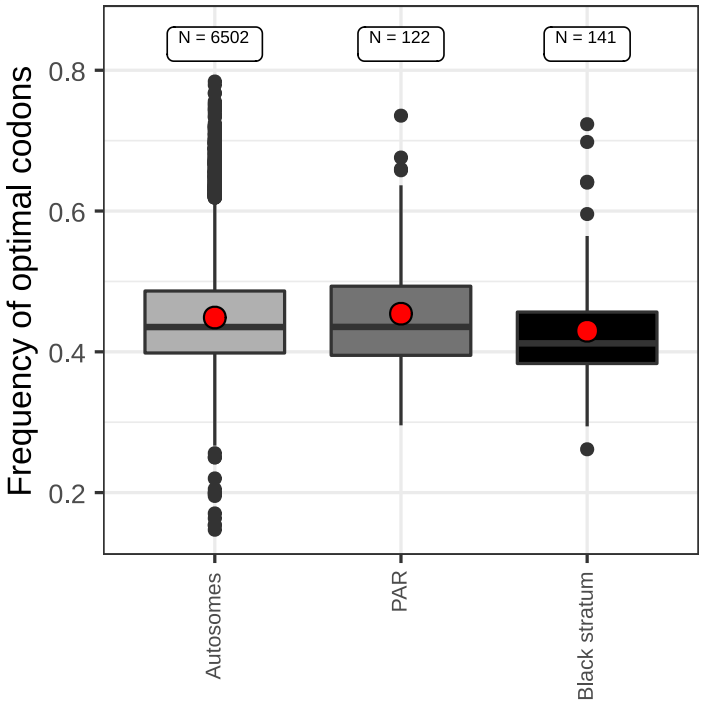

Predicted frequency of optimal codons
0.52
0.48
0.44
0.40
N = 7355
N = 151
N = 486
N = 170
N = 42
*
NS
***
*
Autosomes
Recombining region of the PR chr.
Recombining region of the HD chr.
Black stratum PR-to-CEN
Black stratum HD-to-CEN

C)

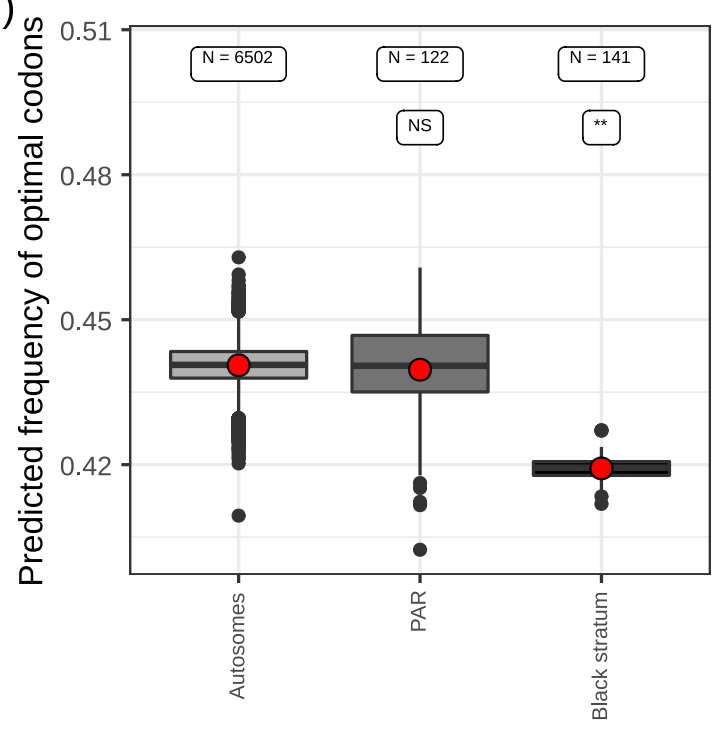

Predicted frequency of optimal codons
0.475
0.450
0.425
0.400
-0.2
-0.1
0.0
0.1
0.2
0.3
Difference of GC in coding and intronic sequences
Interaction significance
Significant
Non-significant
Genomic compartments
Autosomes
Recombining region of the PR chr.
Recombining region of the HD chr.
Black stratum PR-to-CEN
Black stratum HD-to-CEN


M. scabiosae
on Knautia arvensis
a1 genome

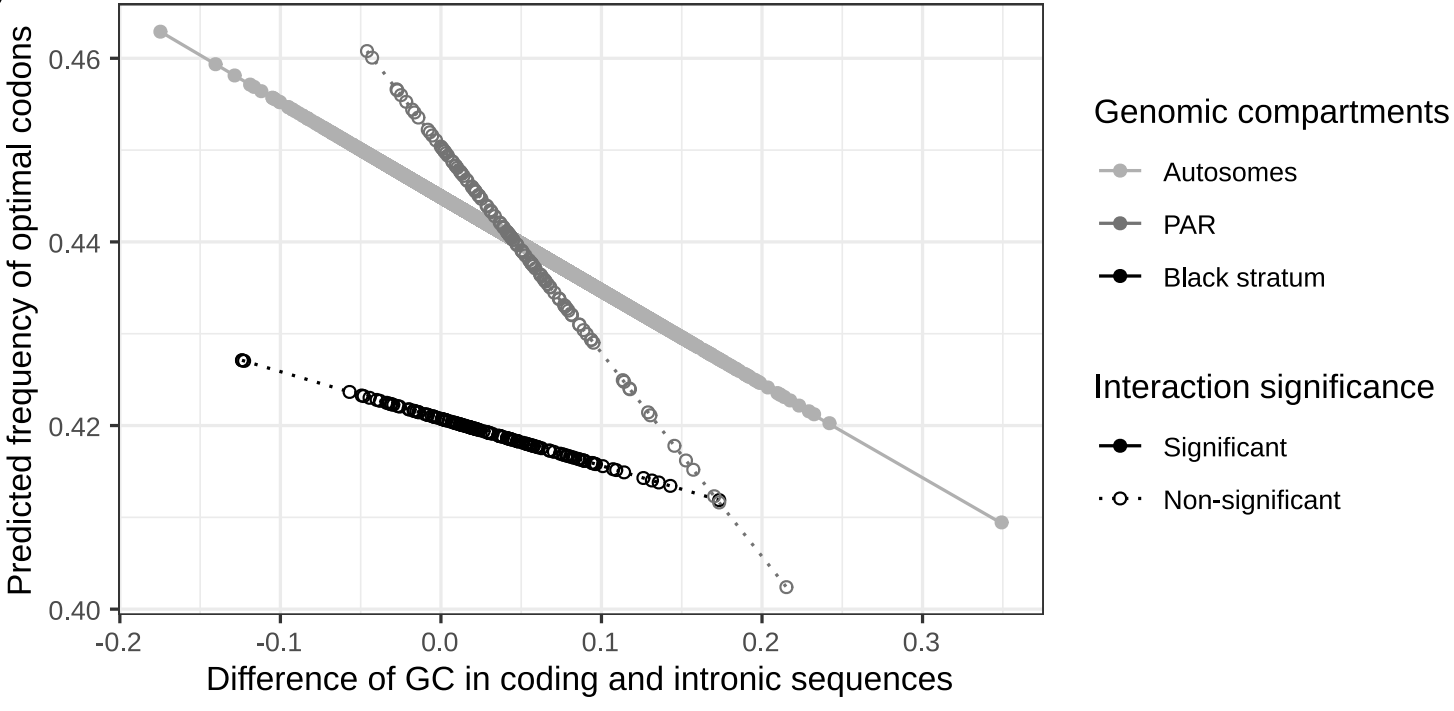
A)

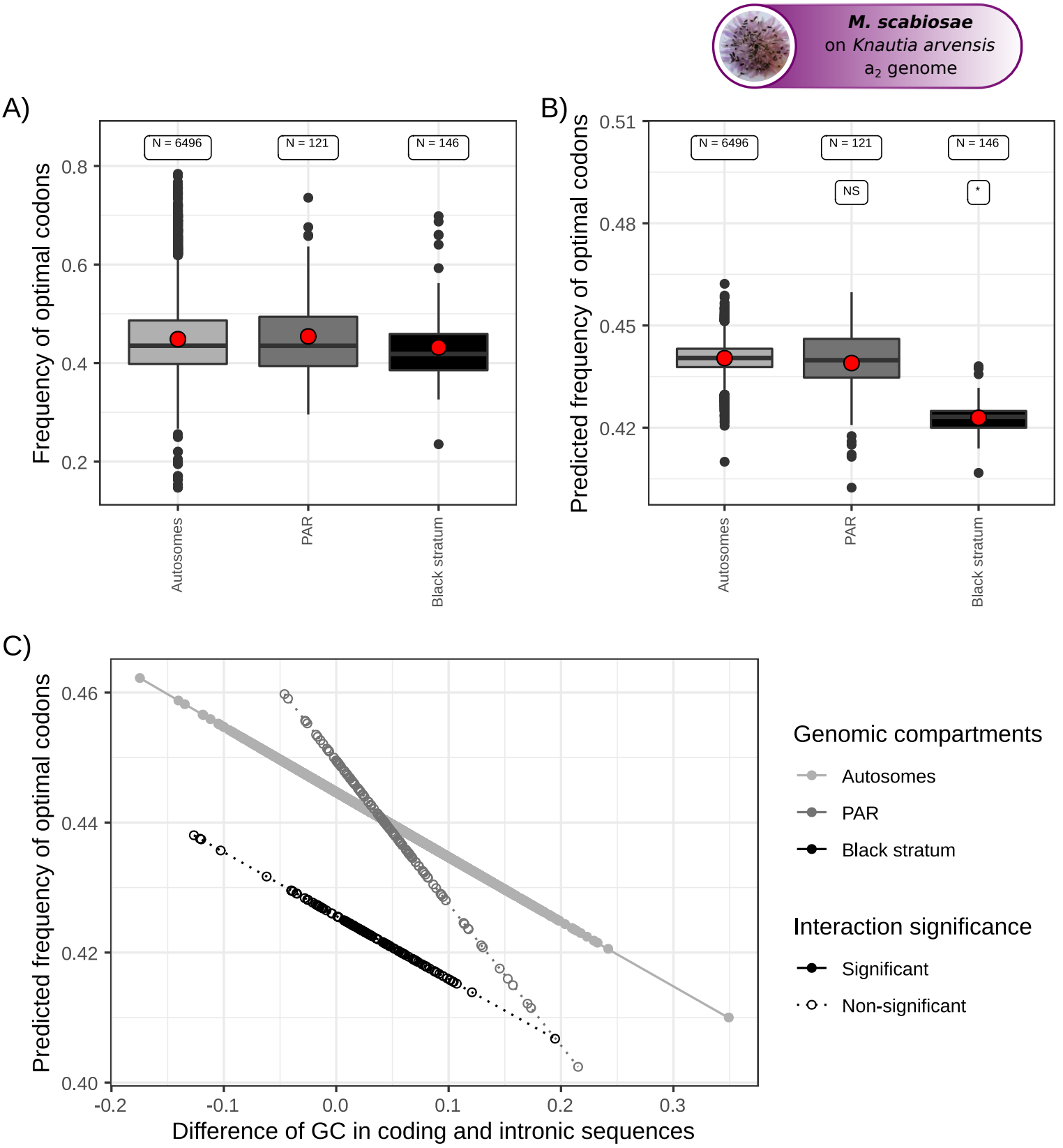

N = 6502
N = 122
N = 141
Frequency of optimal codons
0.8
0.6
0.4
0.2
Autosomes
PAR
Black stratum

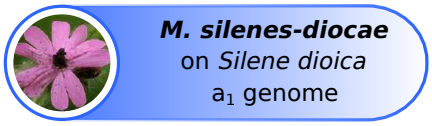
B)


N = 6502
N = 122
N = 141
NS
**
Predicted frequency of optimal codons
0.51
0.48
0.45
0.42
Autosomes
PAR
Black stratum

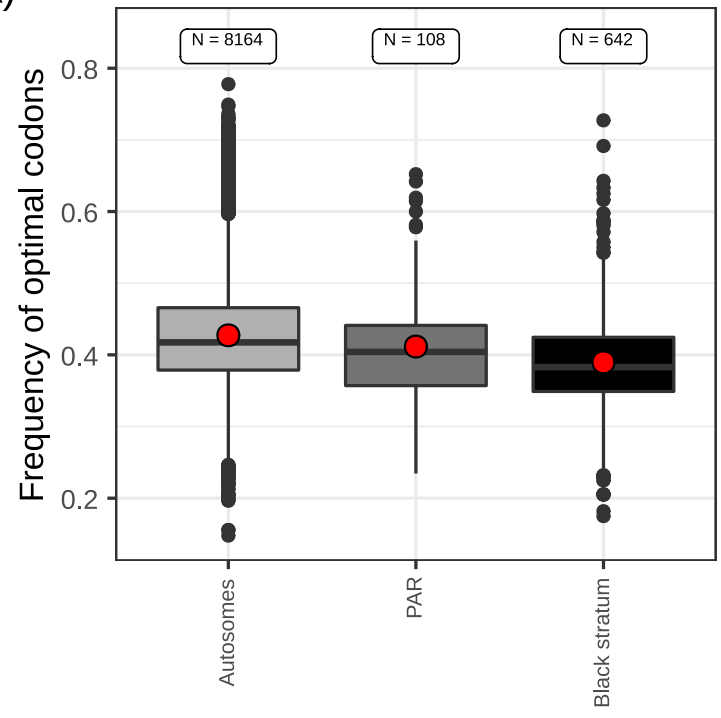
C)


Predicted frequency of optimal codons
0.46
0.44
0.42
0.40
-0.2
-0.1
0.0
0.1
0.2
0.3
Difference of GC in coding and intronic sequences
Genomic compartments
Autosomes
PAR
Black stratum
Interaction significance
Significant
Non-significant

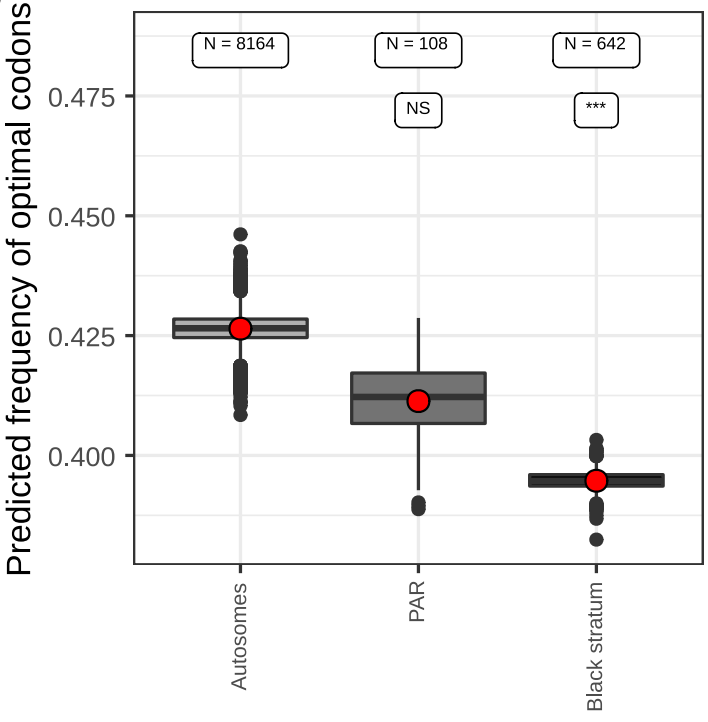

M. scabiosae
on Knautia arvensis
a2 genome
A)
Frequency of optimal codons
N = 6496
N = 121
N = 146
Autosomes
PAR
Black stratum
B)
Predicted frequency of optimal codons
NS
*
C)
Difference of GC in coding and intronic sequences
Genomic compartments
Autosomes
PAR
Black stratum
Interaction significance
Significant
Non-significant


M. silenes-diocae
on Silene dioica
a1 genome

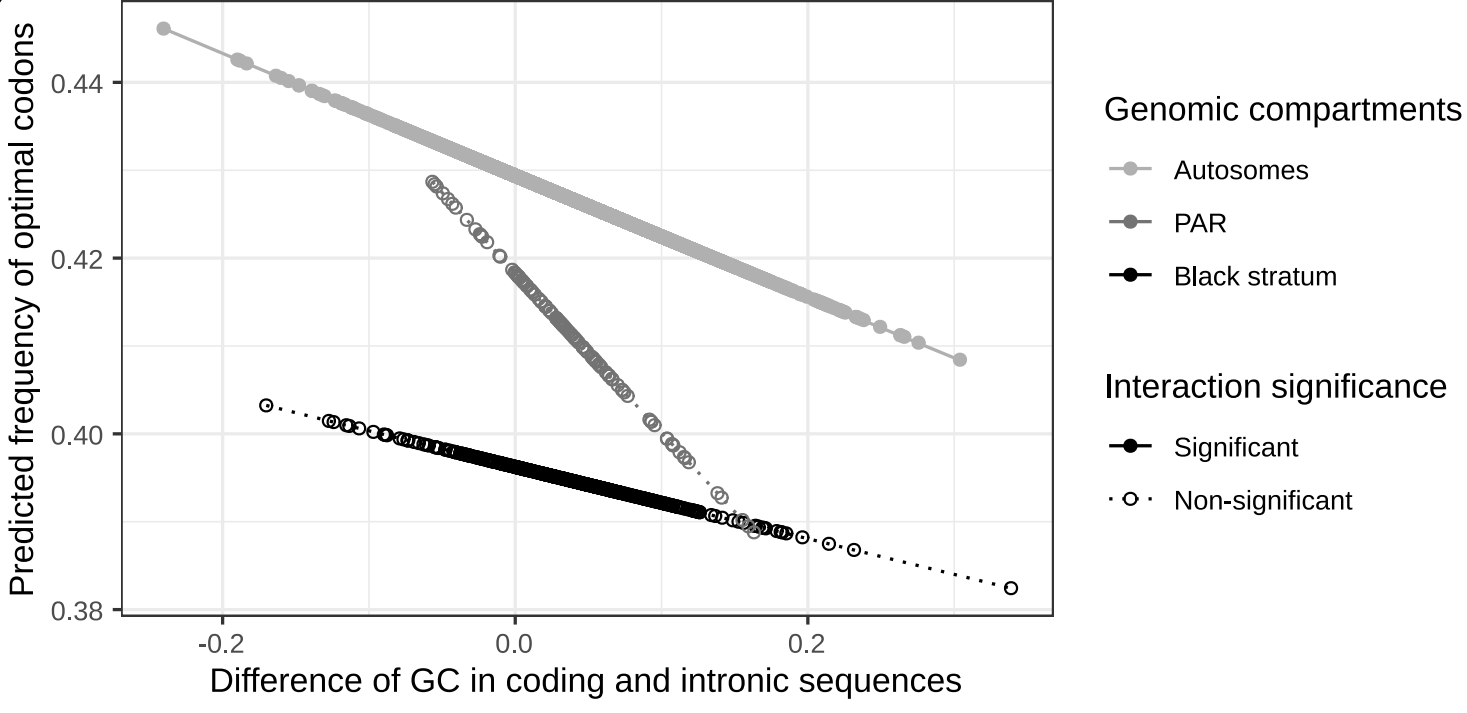
A)

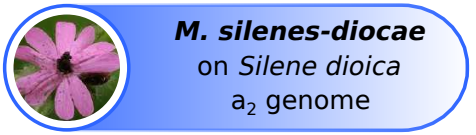

N = 8164
N = 108
N = 642
Frequency of optimal codons
0.8
0.6
0.4
0.2
Autosomes
PAR
Black stratum

B)

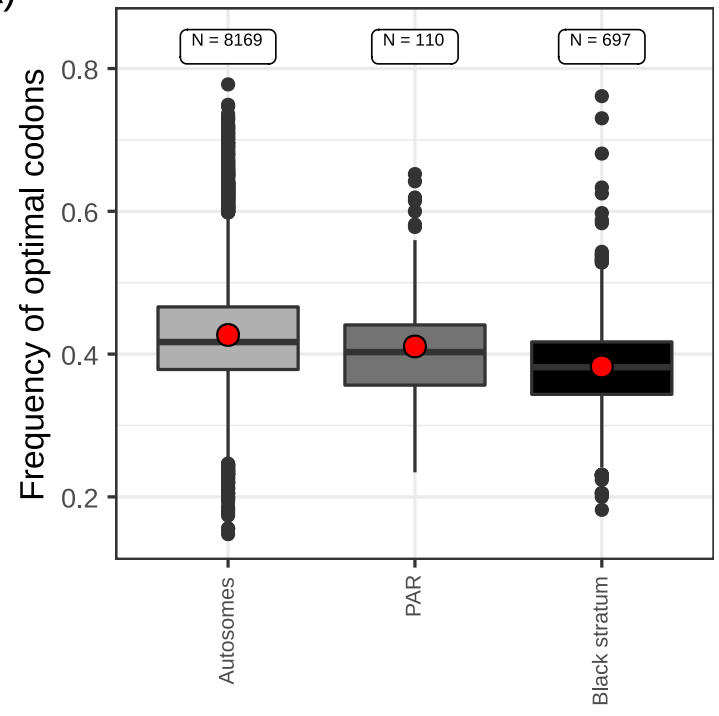

N = 8164
N = 108
N = 642
NS
***
Predicted frequency of optimal codons
0.475
0.450
0.425
0.400
Autosomes
PAR
Black stratum

C)

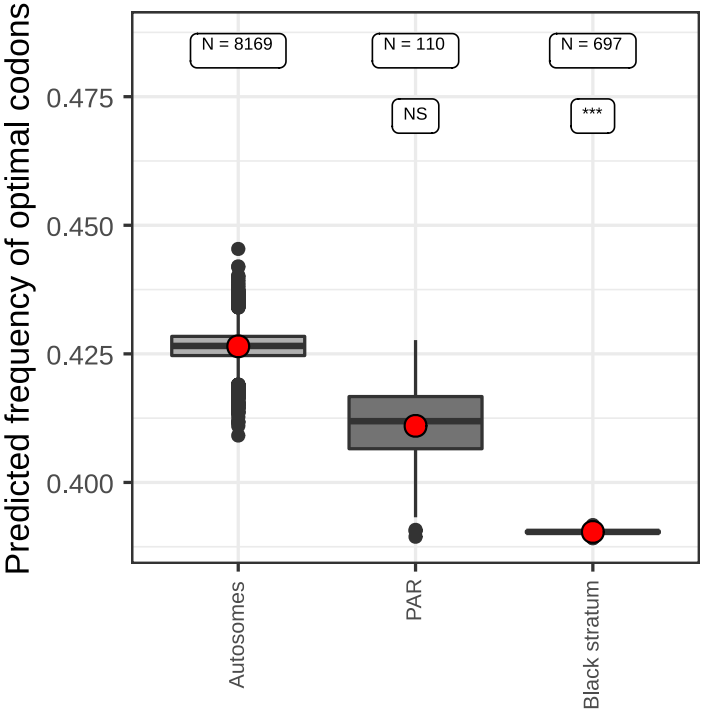

Predicted frequency of optimal codons
0.44
0.42
0.40
0.38
-0.2
0.0
0.2
Difference of GC in coding and intronic sequences
Genomic compartments
Autosomes
PAR
Black stratum
Interaction significance
Significant
Non-significant


M. silenes-diocae
on Silene dioica
a2 genome

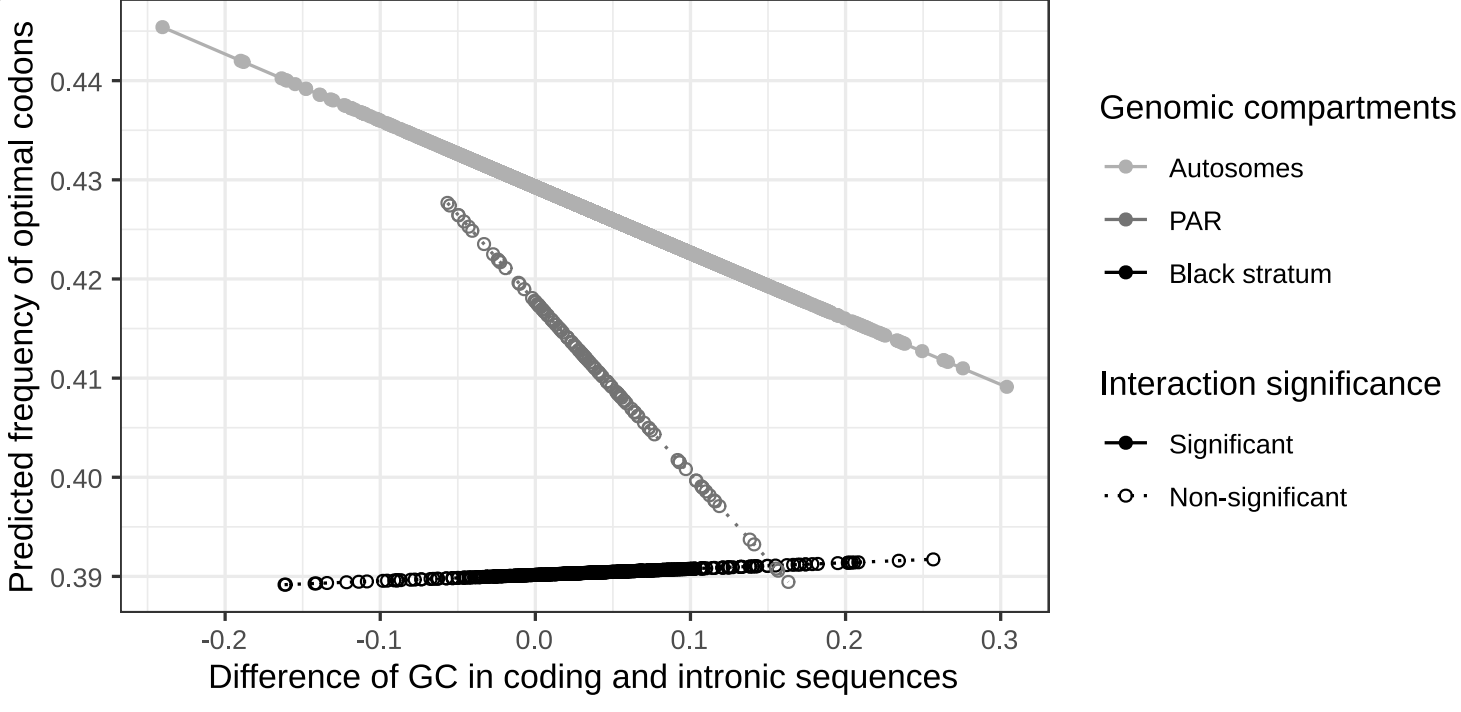
A)

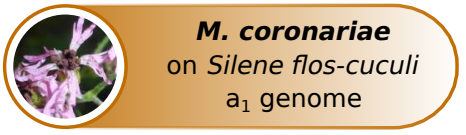

N = 8169
N = 110
N = 697
Frequency of optimal codons
0.8
0.6
0.4
0.2
Autosomes
PAR
Black stratum

B)

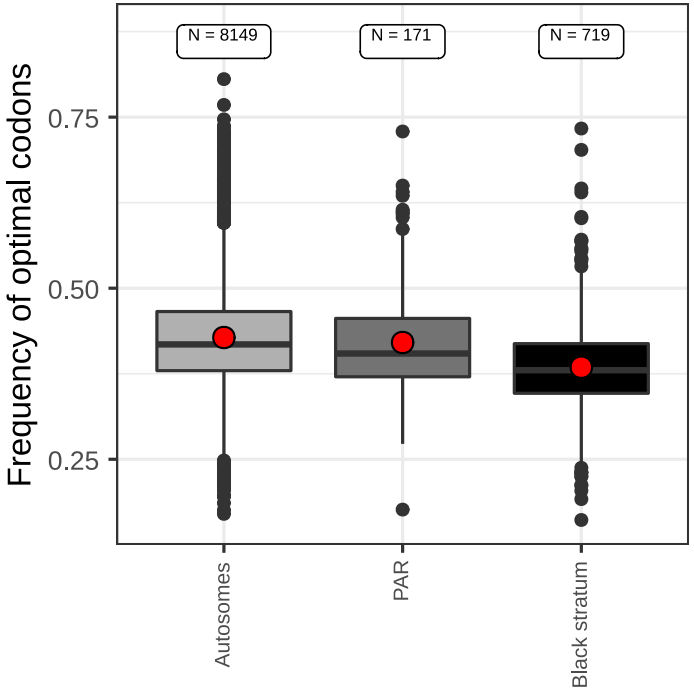

N = 8169
N = 110
N = 697
NS
***
Predicted frequency of optimal codons
0.475
0.450
0.425
0.400
Autosomes
PAR
Black stratum

C)

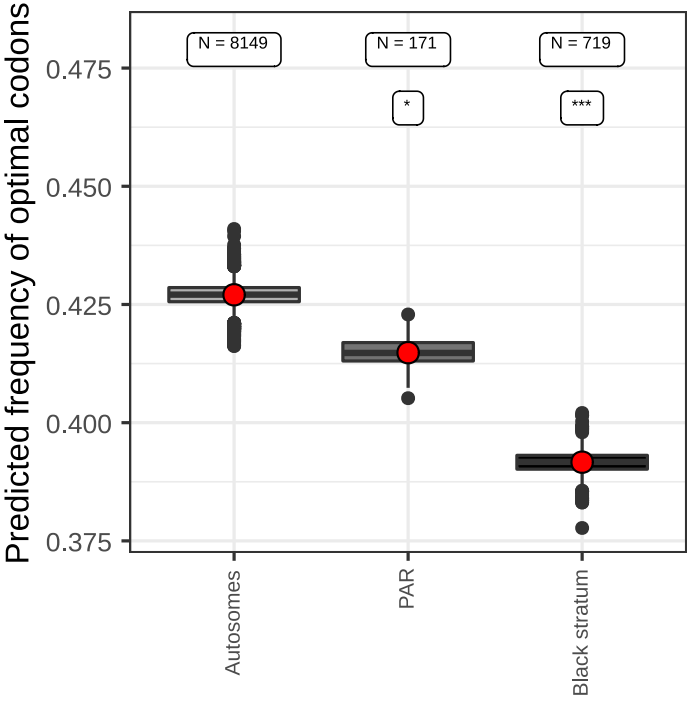

Predicted frequency of optimal codons
0.44
0.43
0.42
0.41
0.40
0.39
-0.2
-0.1
0.0
0.1
0.2
0.3
Difference of GC in coding and intronic sequences
Genomic compartments
Autosomes
PAR
Black stratum
Interaction significance
Significant
Non-significant


M. coronariae
on Silene flos-cuculi
a1 genome

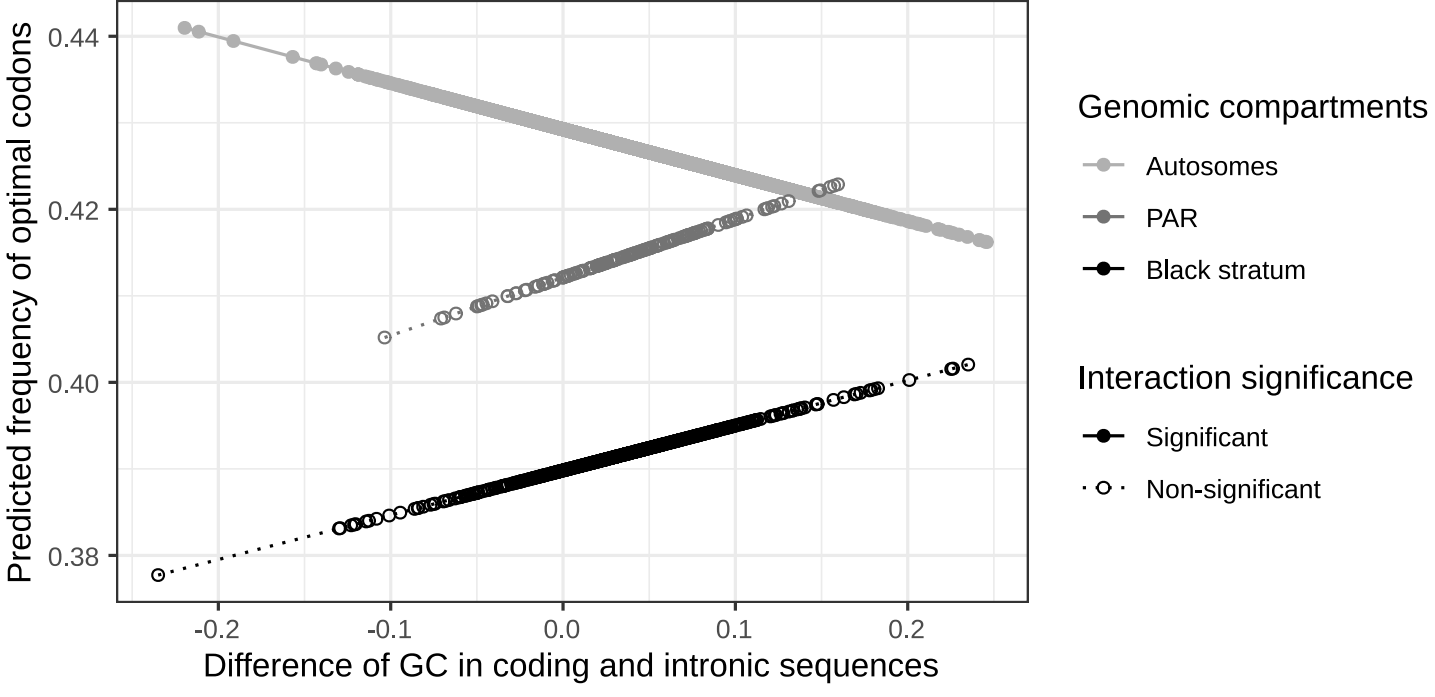
A)

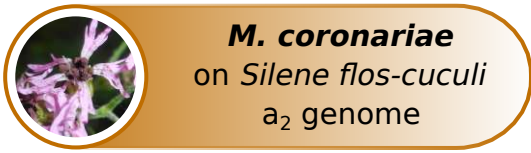

N = 8149
N = 171
N = 719
Frequency of optimal codons
0.75
0.50
0.25
Autosomes
PAR
Black stratum

B)

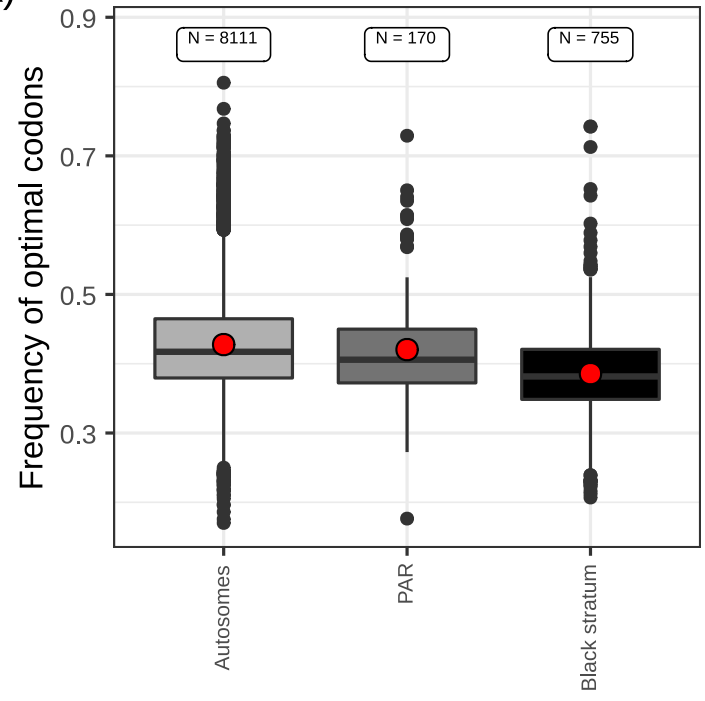

N = 8149
N = 171
N = 719
*
***
Predicted frequency of optimal codons
0.475
0.450
0.425
0.400
0.375
Autosomes
PAR
Black stratum

C)

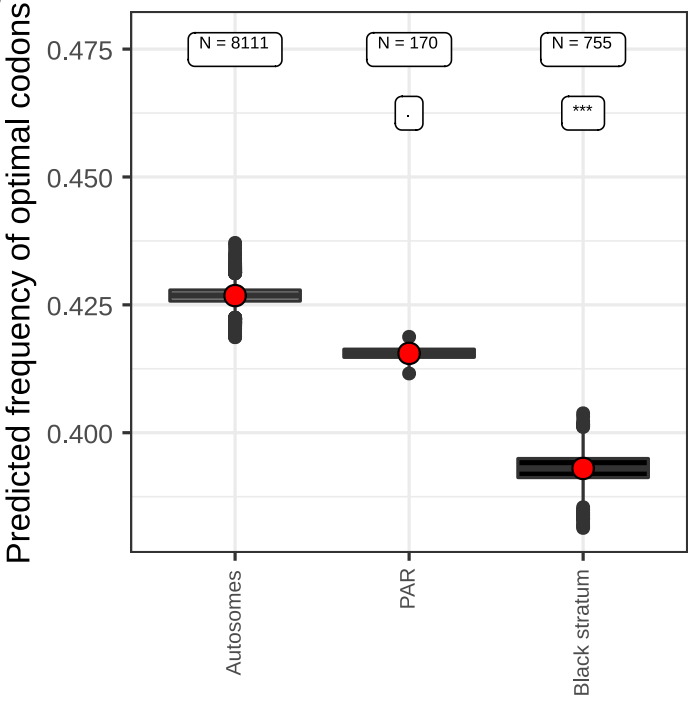

Predicted frequency of optimal codons
0.44
0.42
0.40
0.38
-0.2
-0.1
0.0
0.1
0.2
Difference of GC in coding and intronic sequences
Genomic compartments
Autosomes
PAR
Black stratum
Interaction significance
Significant
Non-significant



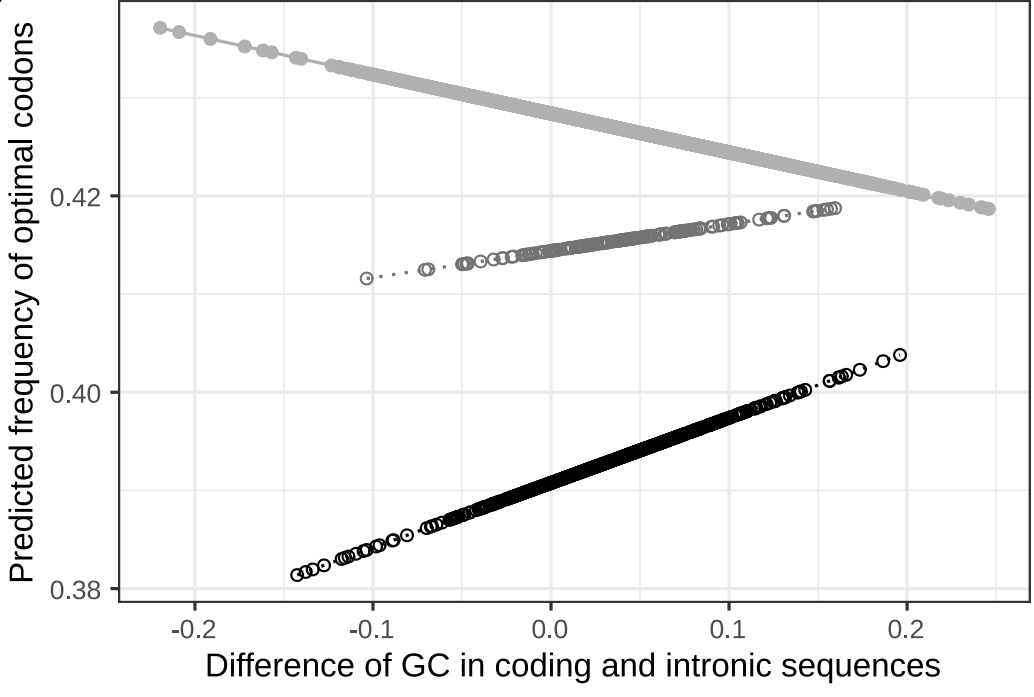
A)



B)



C)

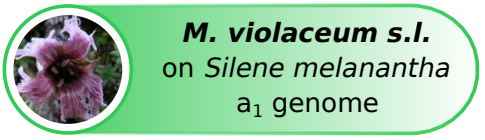


Genomic compartments

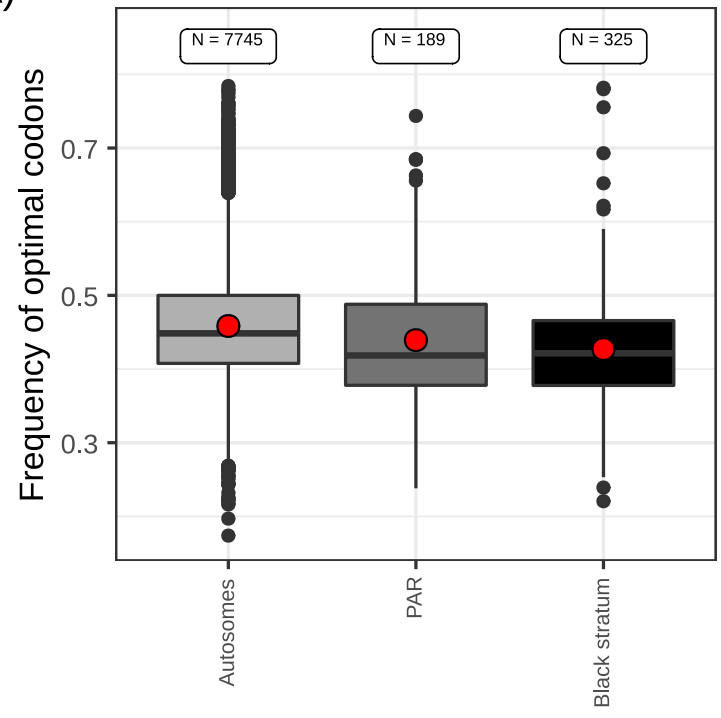
- Autosomes
- PAR
- Black stratum

Interaction significance

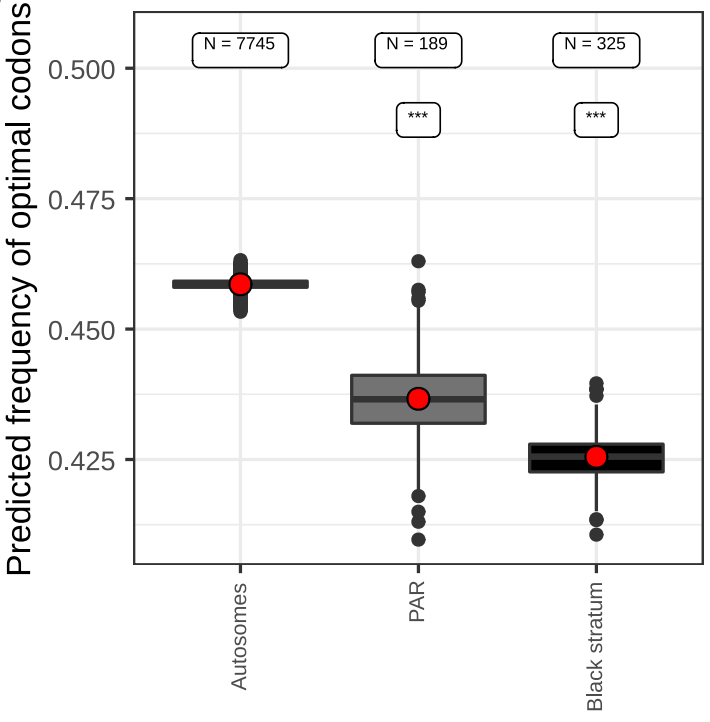
- Significant
- Non-significant


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

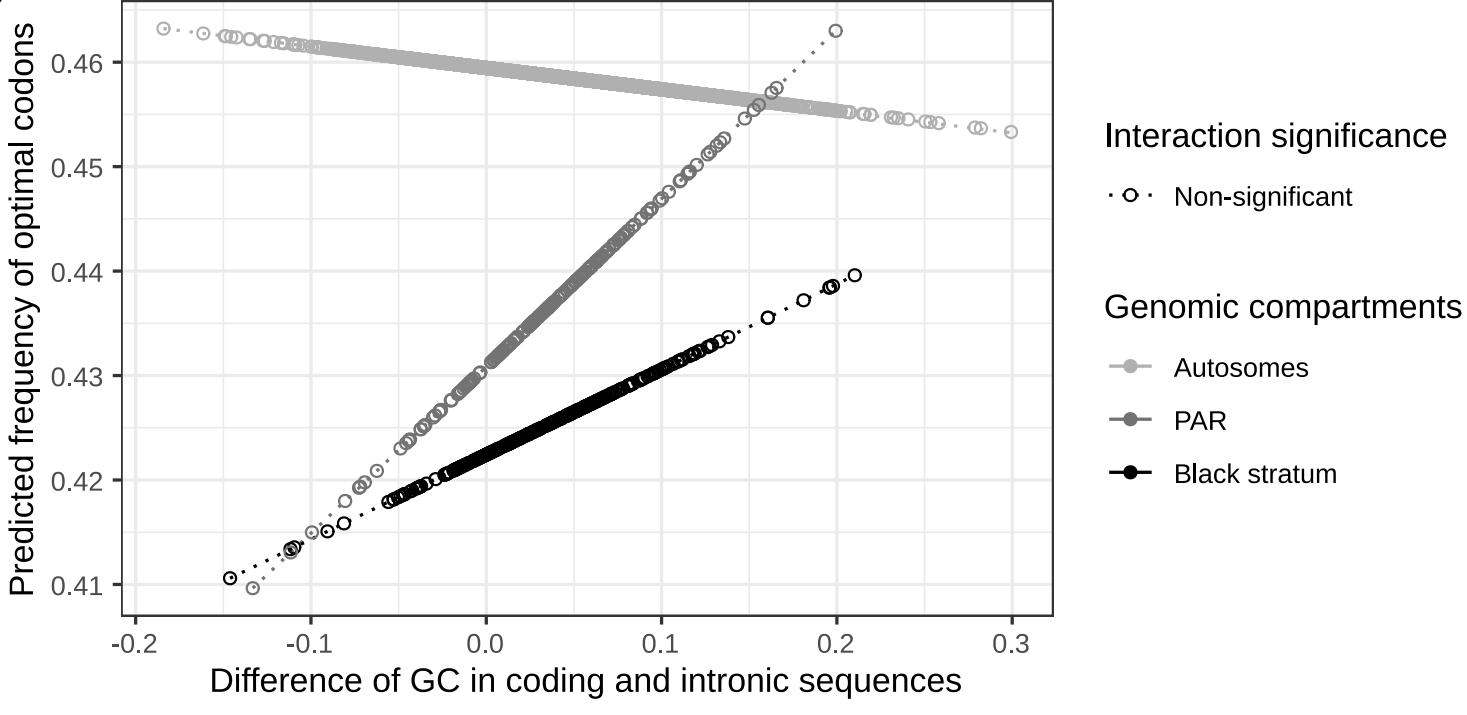
A)

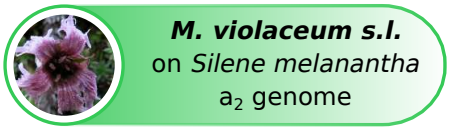

N = 7745
N = 189
N = 325
Frequency of optimal codons
0.7
0.5
0.3
Autosomes
PAR
Black stratum

B)


N = 7745
N = 189
N = 325
***
***
Predicted frequency of optimal codons
0.500
0.475
0.450
0.425
Autosomes
PAR
Black stratum

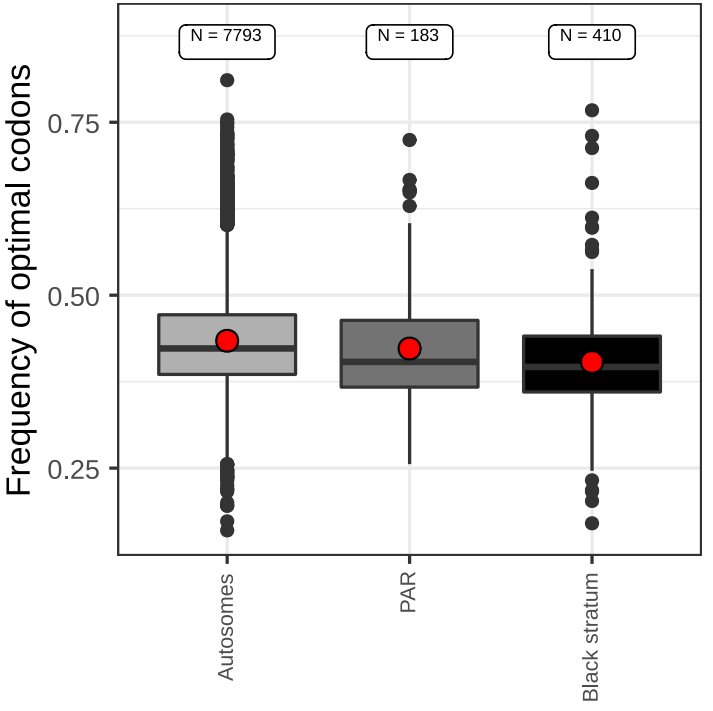
C)


Predicted frequency of optimal codons
0.46
0.45
0.44
0.43
0.42
0.41
-0.2
-0.1
0.0
0.1
0.2
0.3
Difference of GC in coding and intronic sequences
Interaction significance
Non-significant
Genomic compartments
Autosomes
PAR
Black stratum

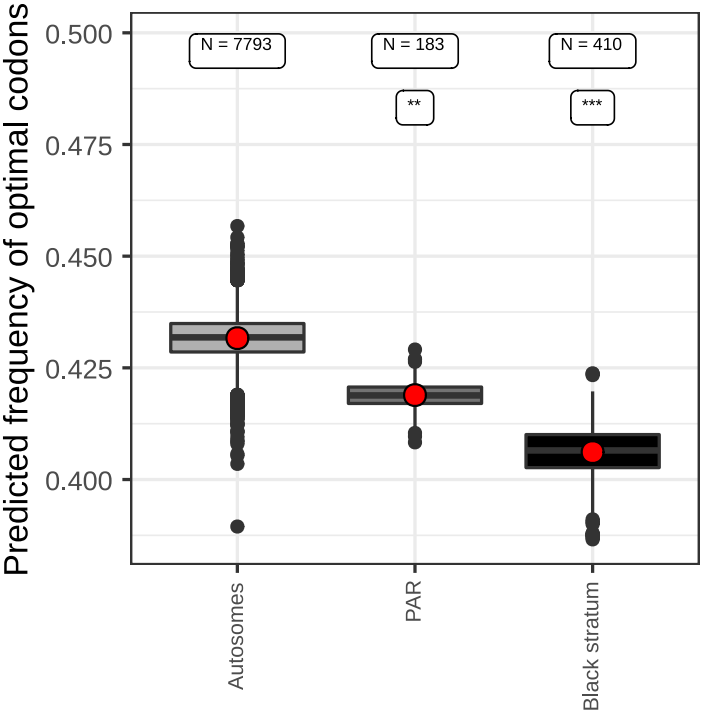


M. violaceum s.l.
on *Silene melanantha*
a_2 genome

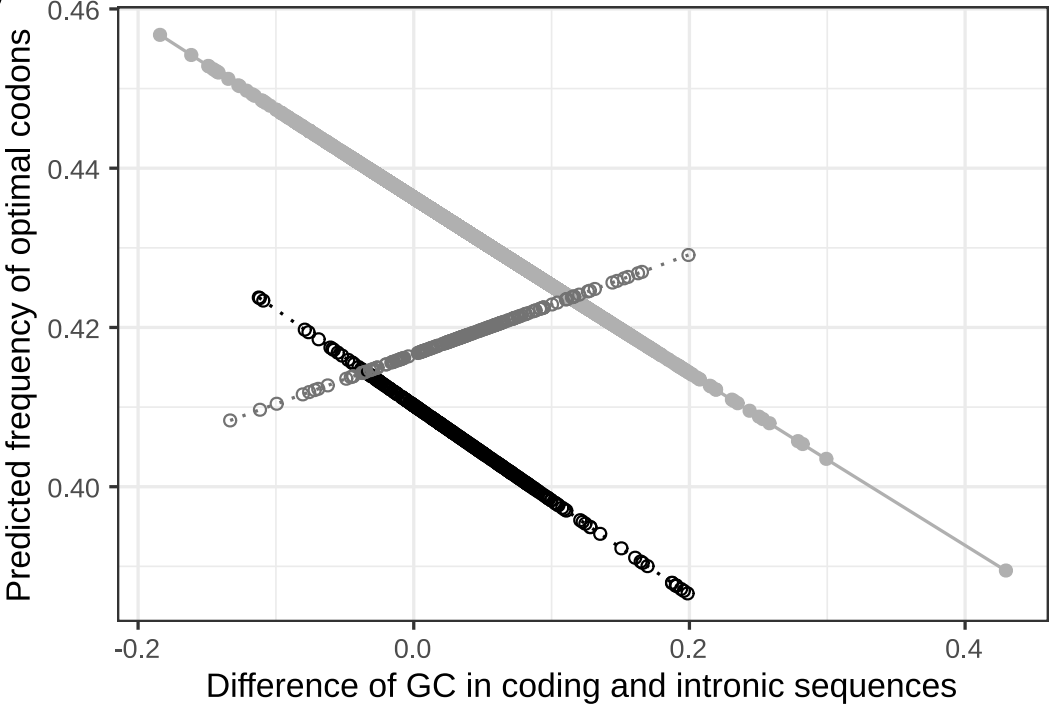
A)



B)



C)

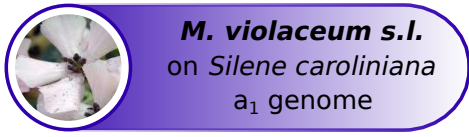


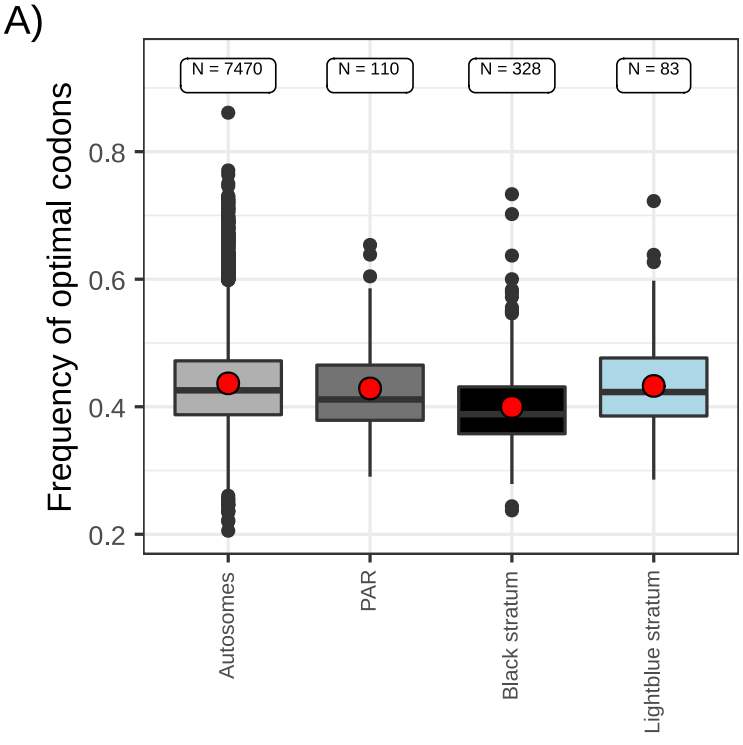
Genomic compartments

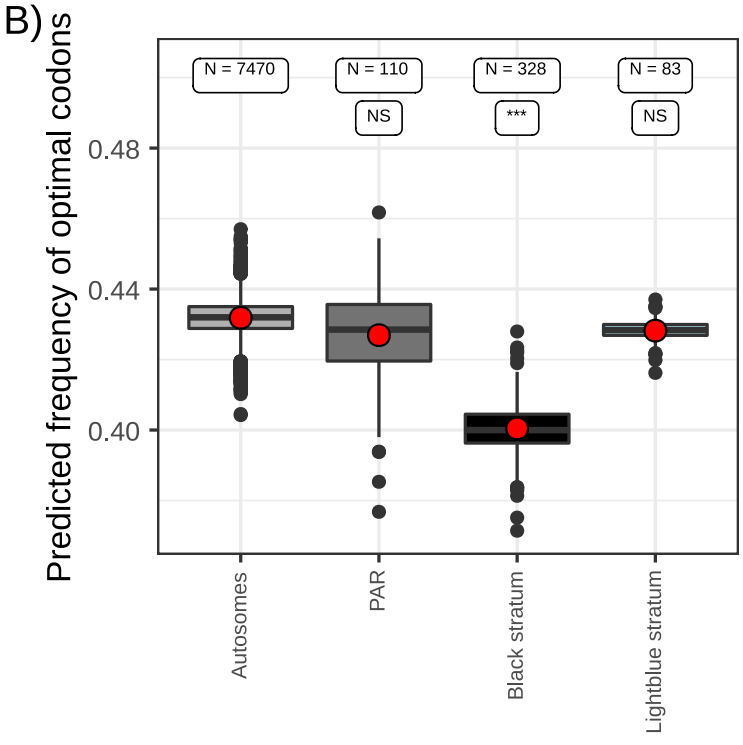
- Autosomes
- PAR
- Black stratum

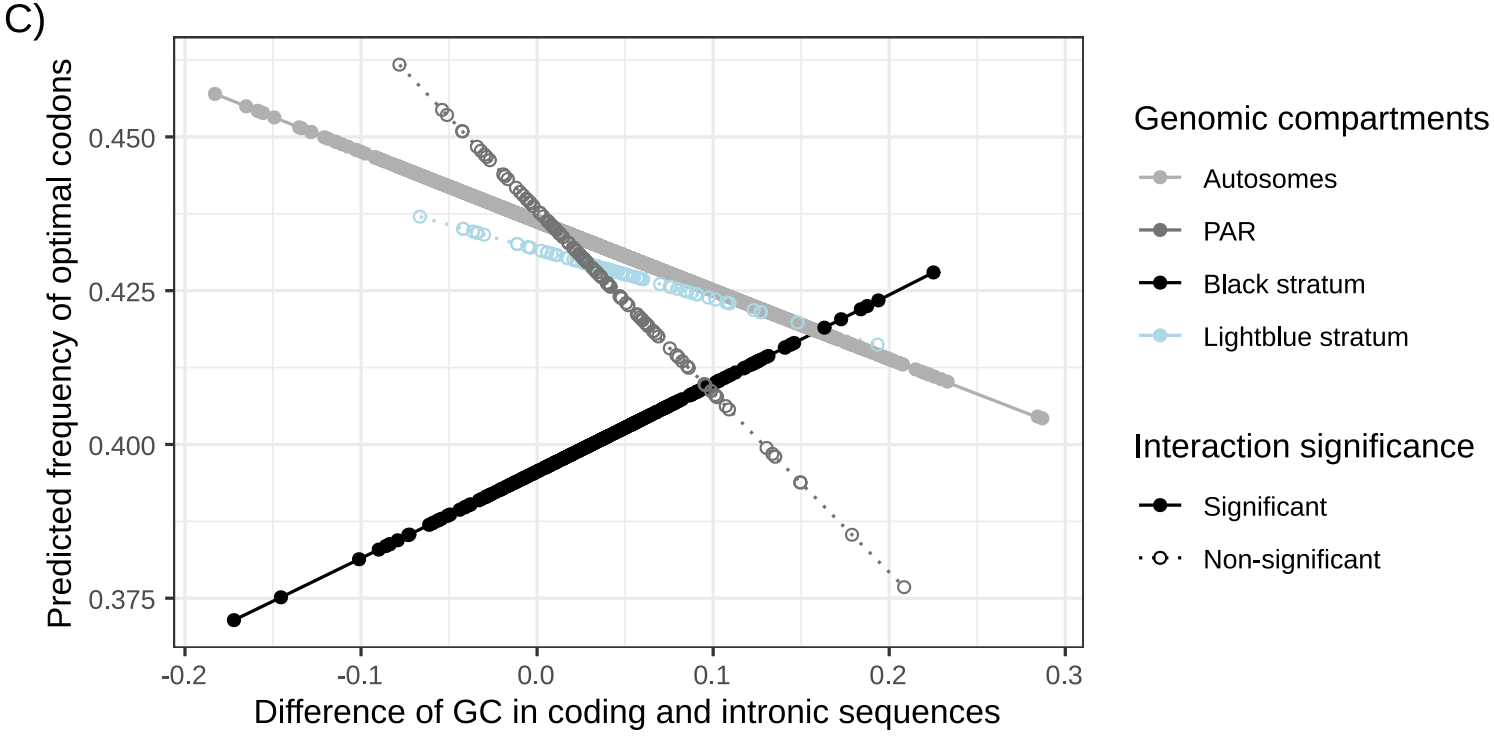
Interaction significance

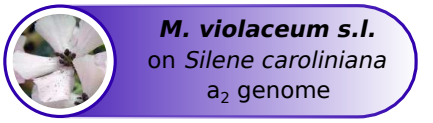
- Significant
- Non-significant


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

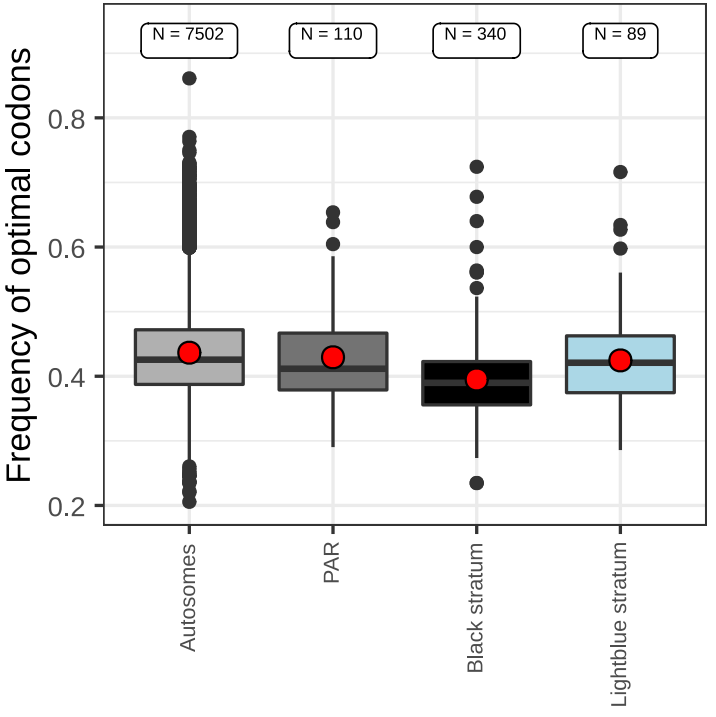

A)
Frequency of optimal codons
0.8
0.6
0.4
0.2
N = 7470
N = 110
N = 328
N = 83
Autosomes
PAR
Black stratum
Lightblue stratum


B)
Predicted frequency of optimal codons
0.48
0.44
0.40
N = 7470
N = 110
NS
N = 328
***
N = 83
NS
Autosomes
PAR
Black stratum
Lightblue stratum

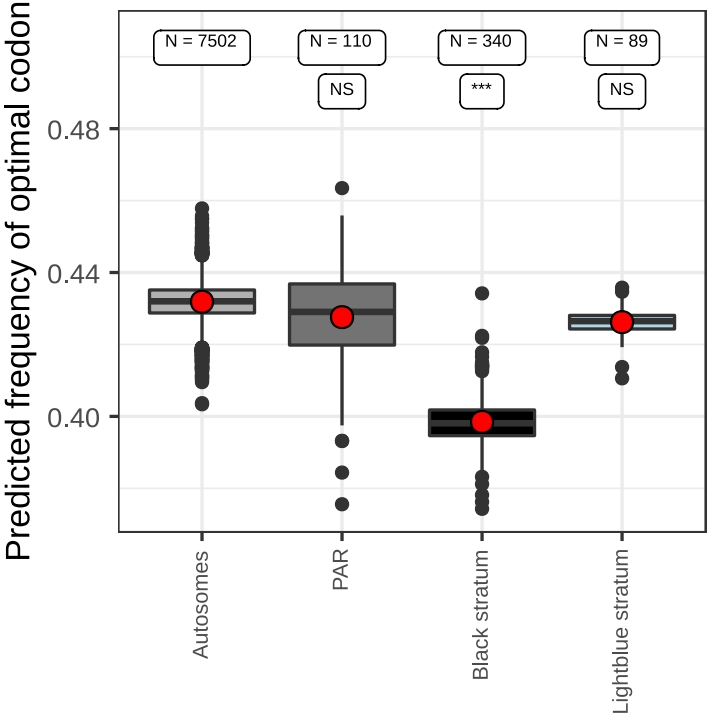

C)
Predicted frequency of optimal codons
0.450
0.425
0.400
0.375
-0.2
-0.1
0.0
0.1
0.2
0.3
Difference of GC in coding and intronic sequences
Genomic compartments
Autosomes
PAR
Black stratum
Lightblue stratum
Interaction significance
Significant
Non-significant



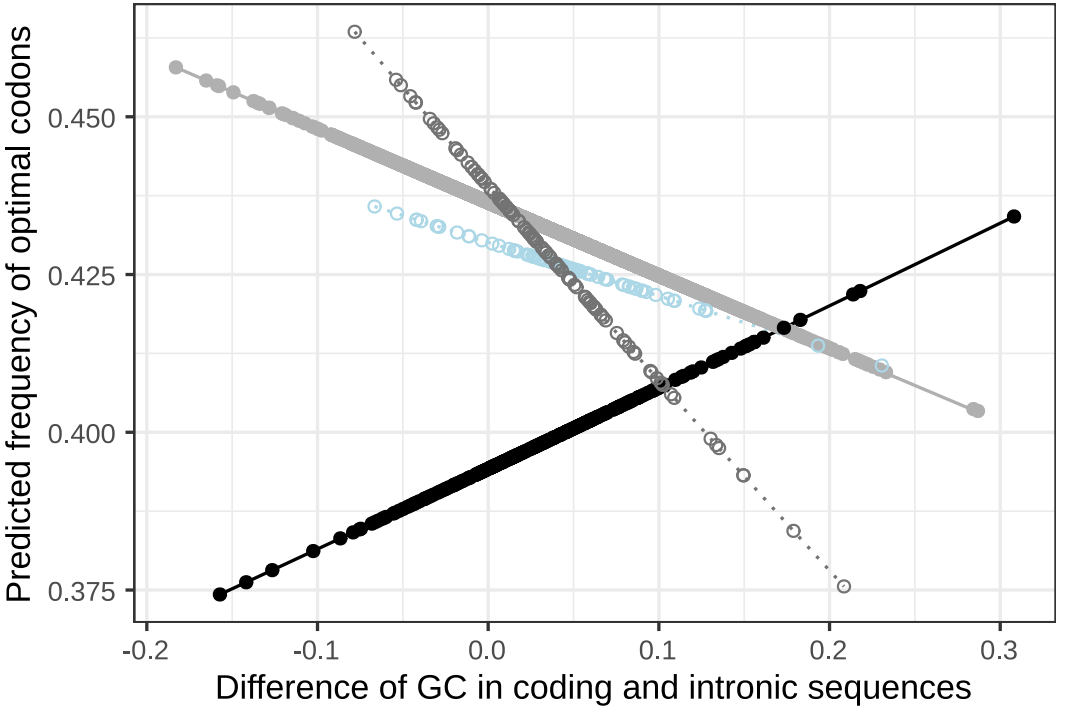
A)



B)



C)

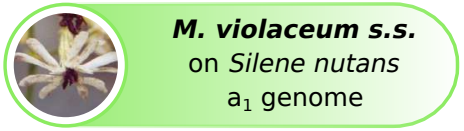


Genomic compartments

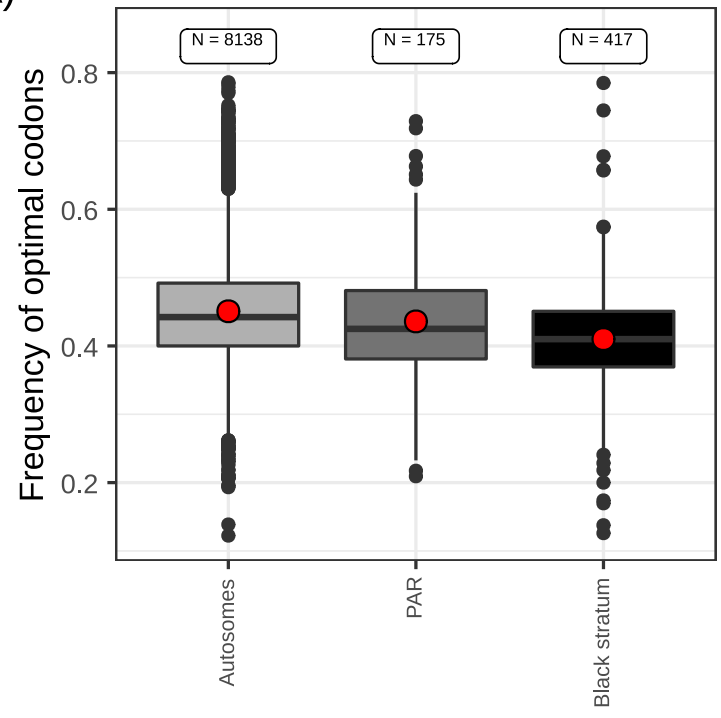
- Autosomes
- PAR
- Black stratum
- Lightblue stratum

Interaction significance

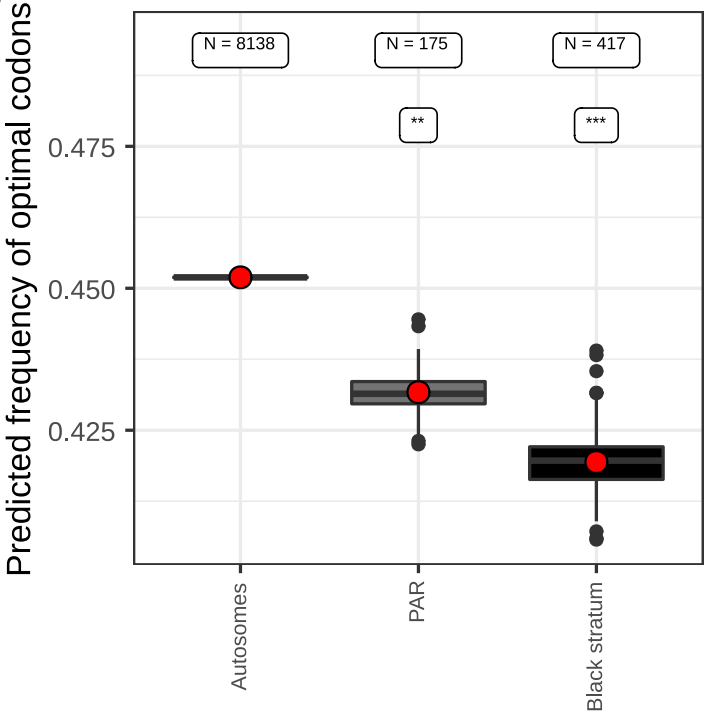
- Significant
- Non-significant


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

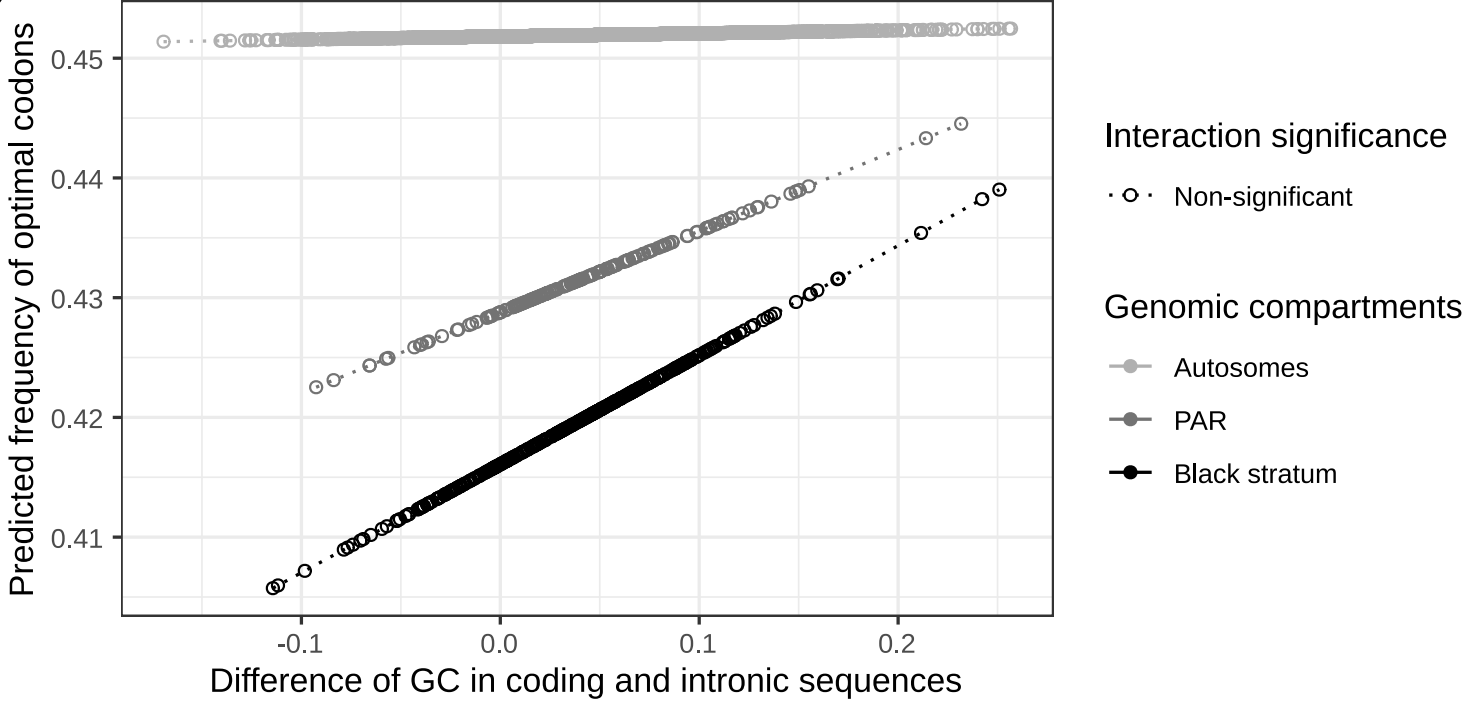
A)

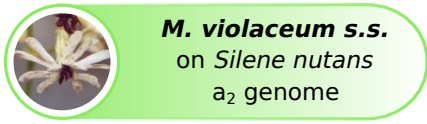

N = 8138
N = 175
N = 417
Frequency of optimal codons
0.8
0.6
0.4
0.2
Autosomes
PAR
Black stratum

B)

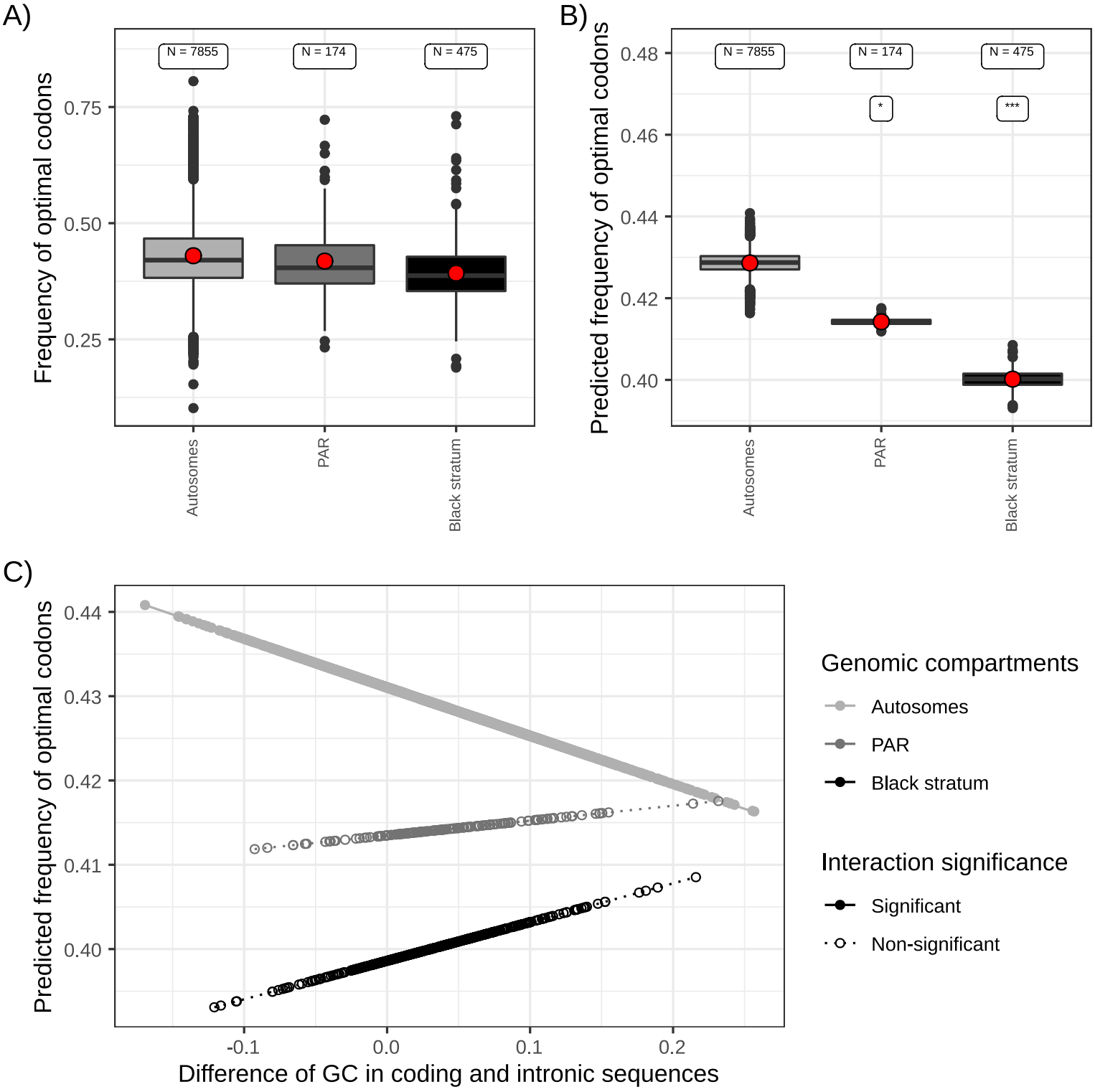

N = 8138
N = 175
N = 417
**
***
Predicted frequency of optimal codons
0.475
0.450
0.425
Autosomes
PAR
Black stratum

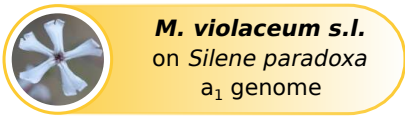
C)

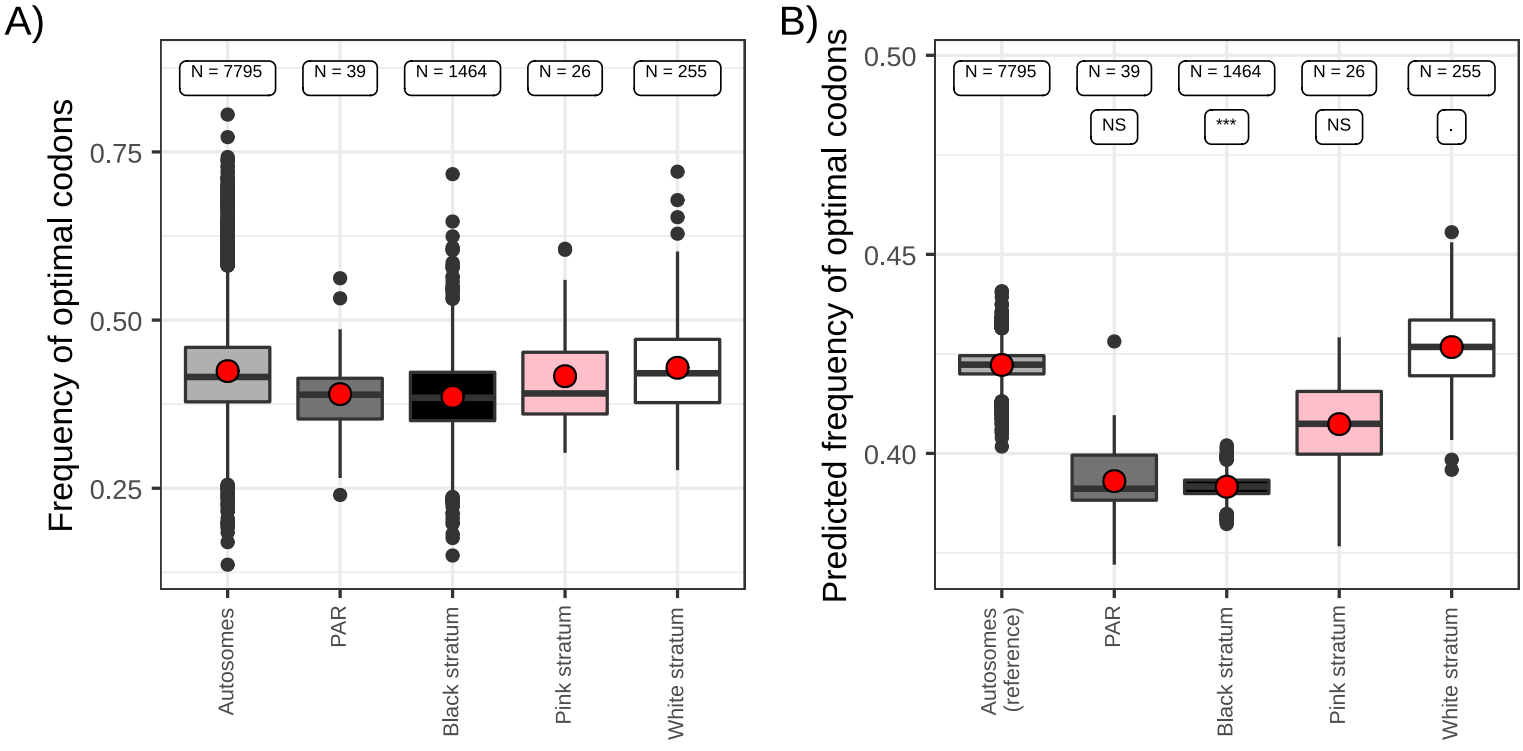

Predicted frequency of optimal codons
0.45
0.44
0.43
0.42
0.41
-0.1
0.0
0.1
0.2
Difference of GC in coding and intronic sequences
Interaction significance
Non-significant
Genomic compartments
Autosomes
PAR
Black stratum

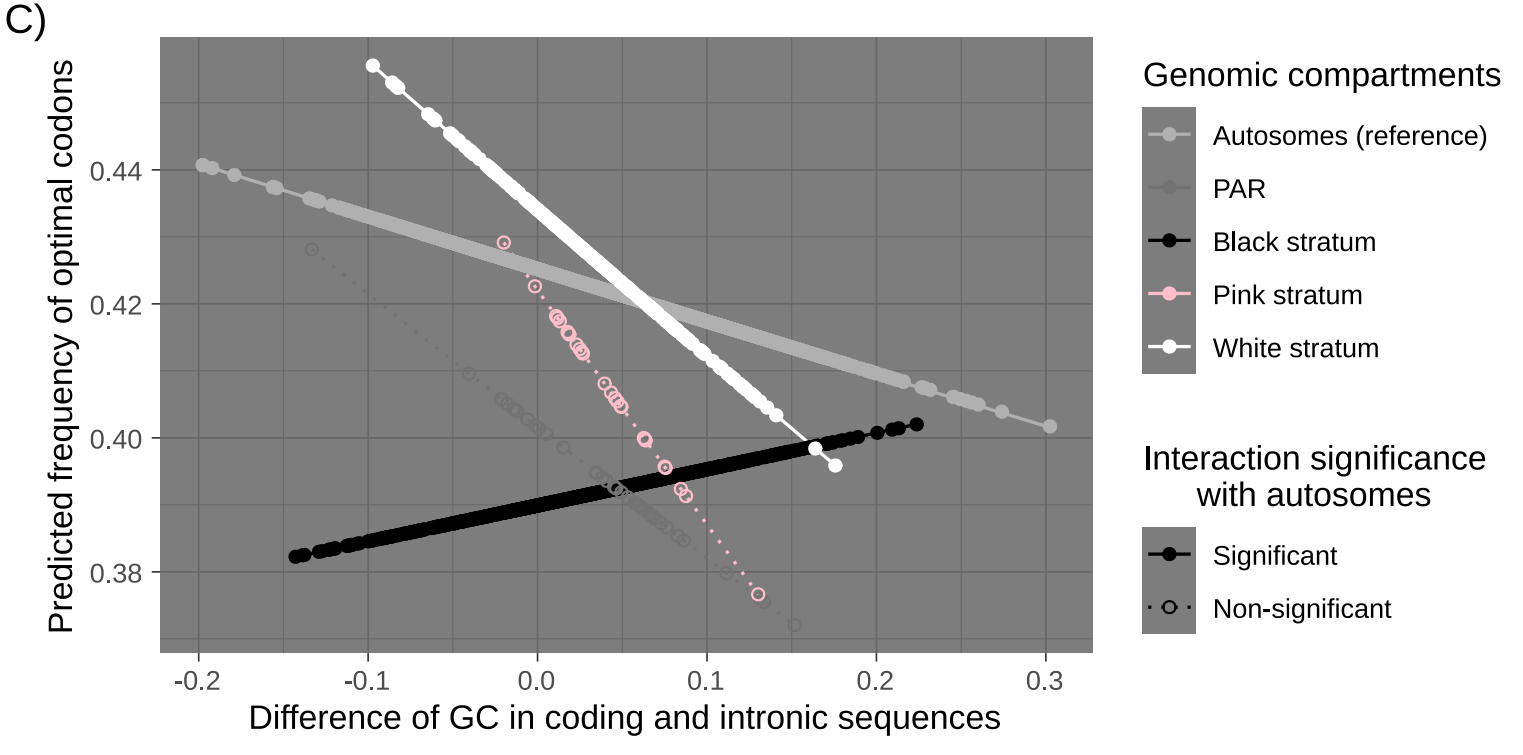


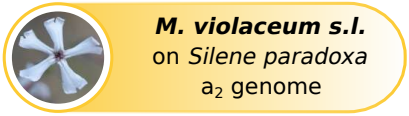
M. violaceum s.s.
on *Silene nutans*
a_2 genome



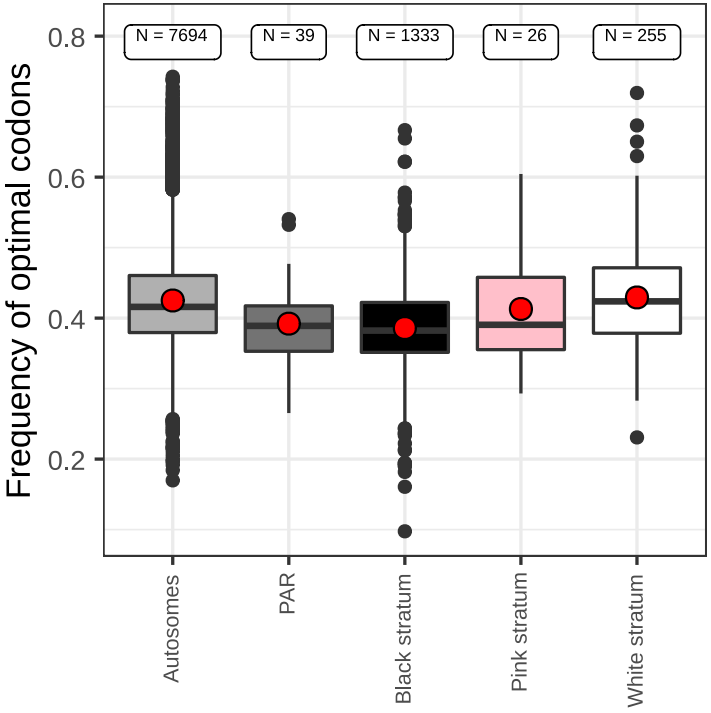

M. violaceum s.l.
on Silene paradoxa
a1 genome


A)
Frequency of optimal codons
0.75
0.50
0.25
N = 7795
N = 39
N = 1464
N = 26
N = 255
Autosomes
PAR
Black stratum
Pink stratum
White stratum
B)
Predicted frequency of optimal codons
0.50
0.45
0.40
N = 7795
N = 39
N = 1464
N = 26
N = 255
NS
***
NS
.
Autosomes (reference)
PAR
Black stratum
Pink stratum
White stratum

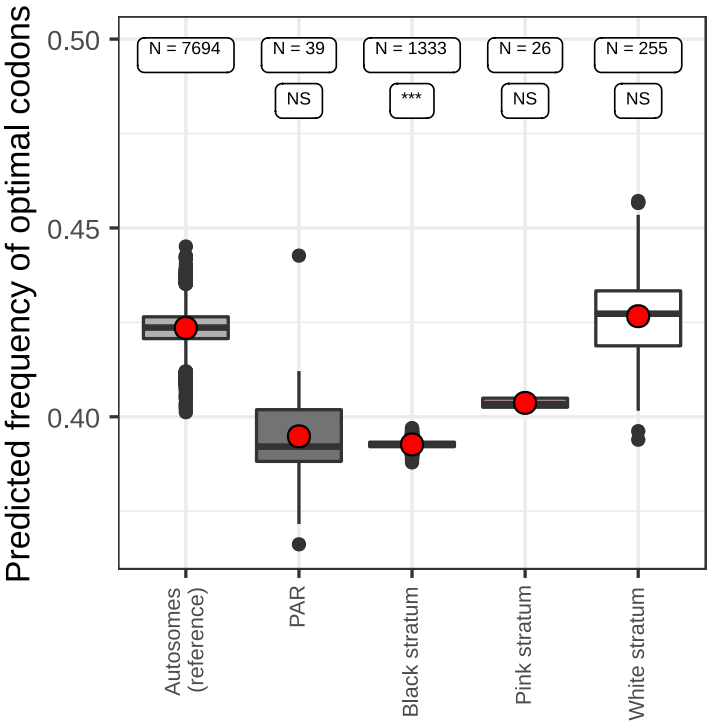

C)
Predicted frequency of optimal codons
0.44
0.42
0.40
0.38
-0.2
-0.1
0.0
0.1
0.2
0.3
Difference of GC in coding and intronic sequences
Genomic compartments
Autosomes (reference)
PAR
Black stratum
Pink stratum
White stratum
Interaction significance with autosomes
Significant
Non-significant


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

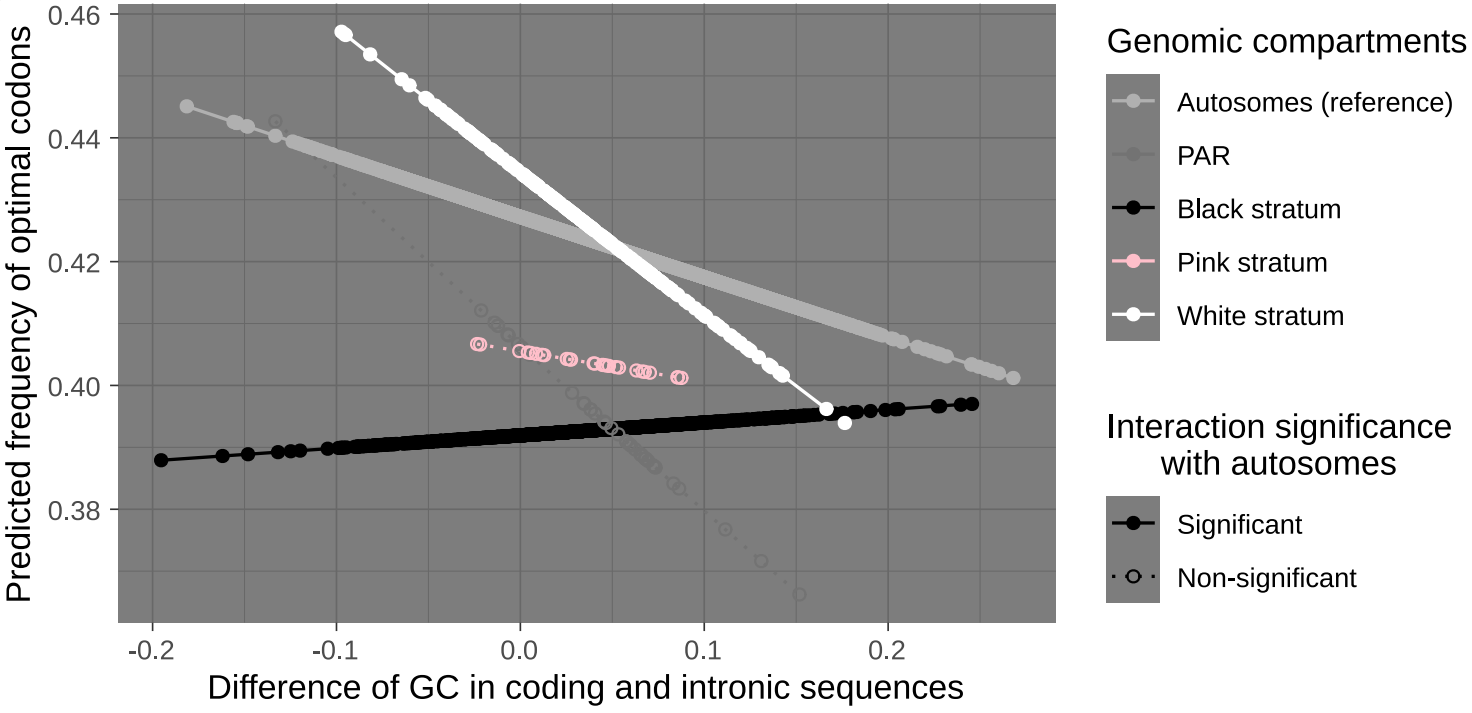
A)

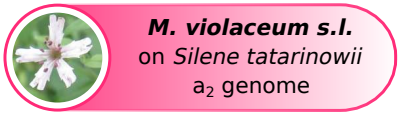

Frequency of optimal codons
0.8
0.6
0.4
0.2
N = 7694
N = 39
N = 1333
N = 26
N = 255
Autosomes
PAR
Black stratum
Pink stratum
White stratum

B)

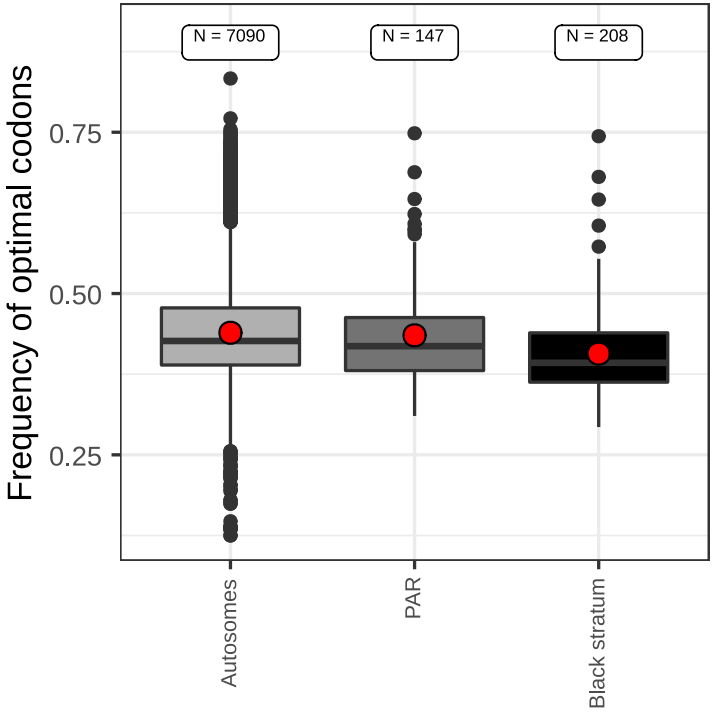

Predicted frequency of optimal codons
0.50
0.45
0.40
N = 7694
N = 39
N = 1333
N = 26
N = 255
NS
***
NS
NS
Autosomes (reference)
PAR
Black stratum
Pink stratum
White stratum

C)

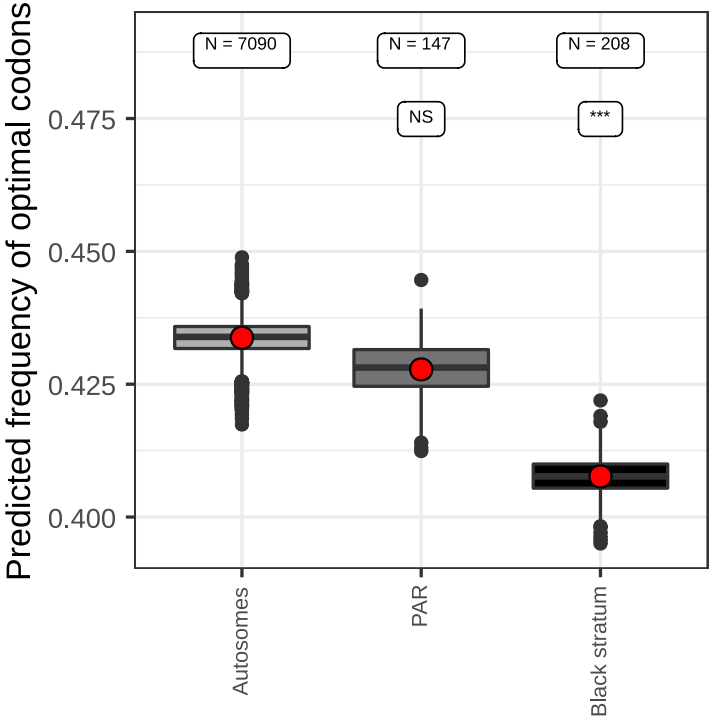

Predicted frequency of optimal codons
0.46
0.44
0.42
0.40
0.38
-0.2
-0.1
0.0
0.1
0.2
Difference of GC in coding and intronic sequences
Genomic compartments
Autosomes (reference)
PAR
Black stratum
Pink stratum
White stratum
Interaction significance with autosomes
Significant
Non-significant


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

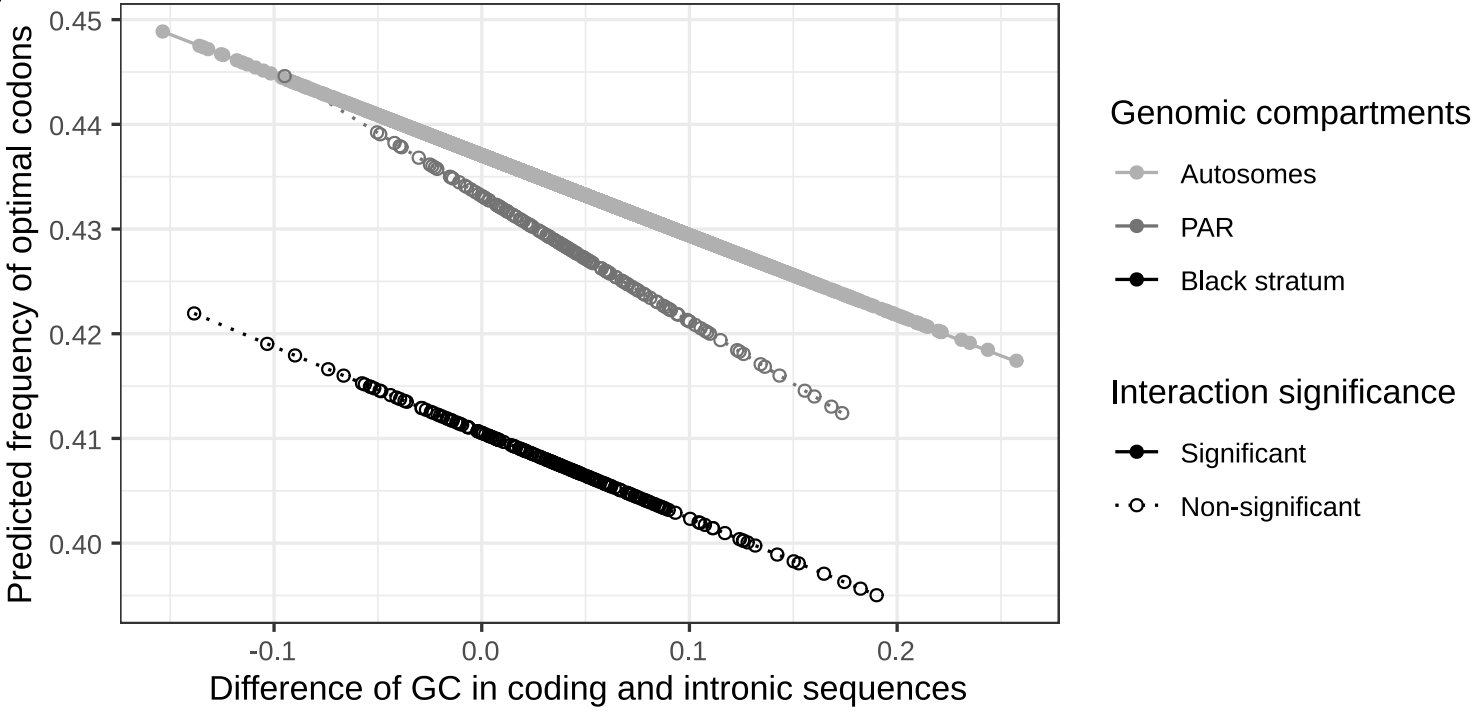
A)

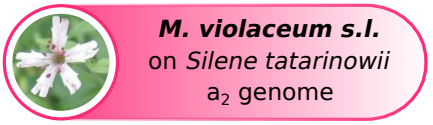

N = 7090
N = 147
N = 208
Frequency of optimal codons
0.75
0.50
0.25
Autosomes
PAR
Black stratum

B)


N = 7090
N = 147
N = 208
NS
***
Predicted frequency of optimal codons
0.475
0.450
0.425
0.400
Autosomes
PAR
Black stratum

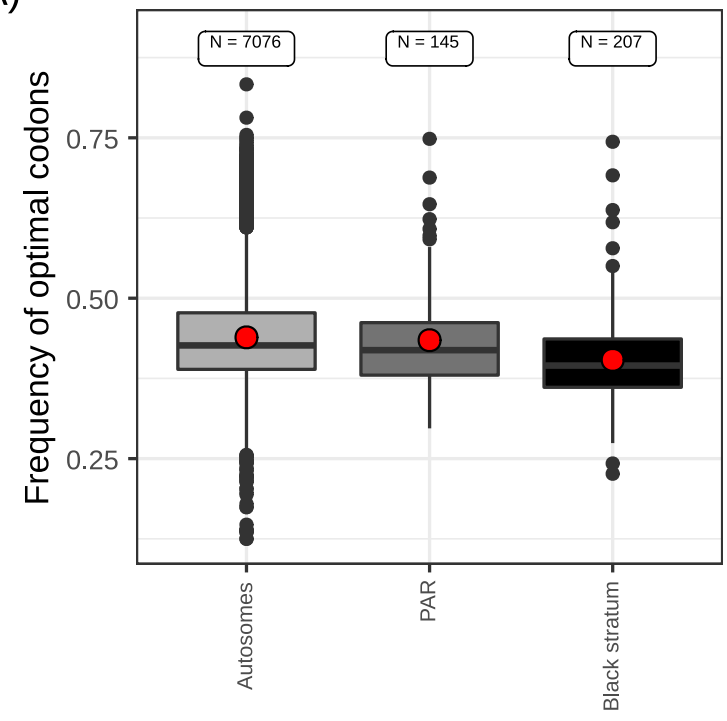
C)


Predicted frequency of optimal codons
0.45
0.44
0.43
0.42
0.41
0.40
-0.1
0.0
0.1
0.2
Difference of GC in coding and intronic sequences
Genomic compartments
Autosomes
PAR
Black stratum
Interaction significance
Significant
Non-significant

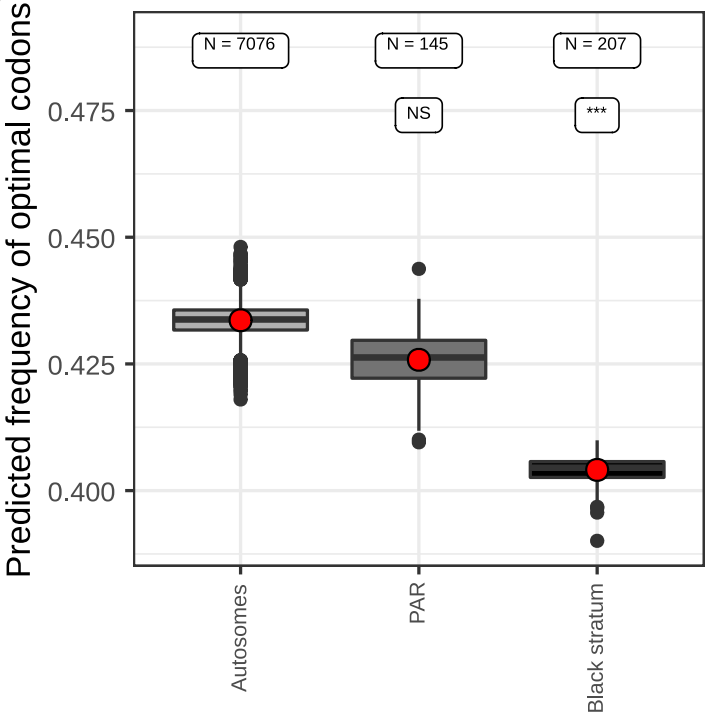


M. violaceum s.l.
on *Silene tatarinowii*
a_2 genome

A)



B)



C)

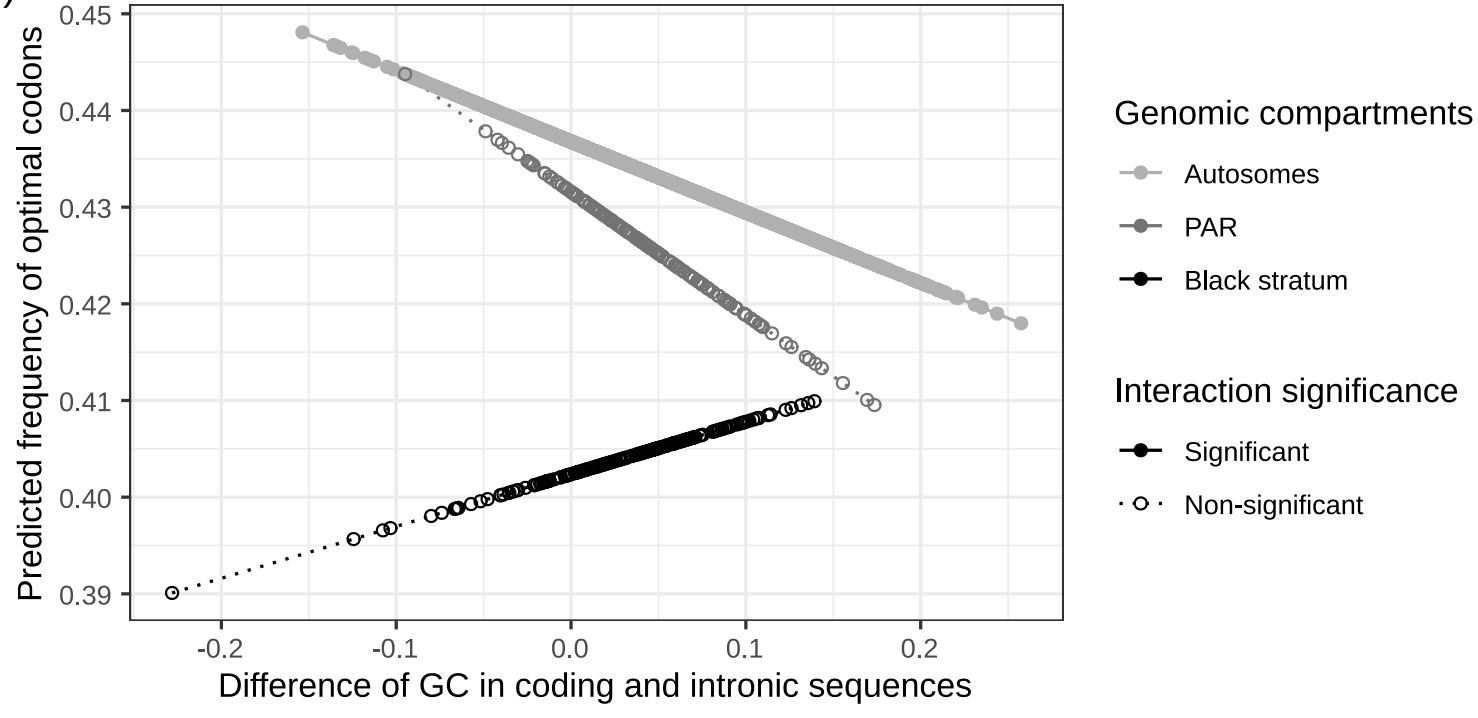
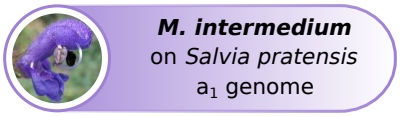
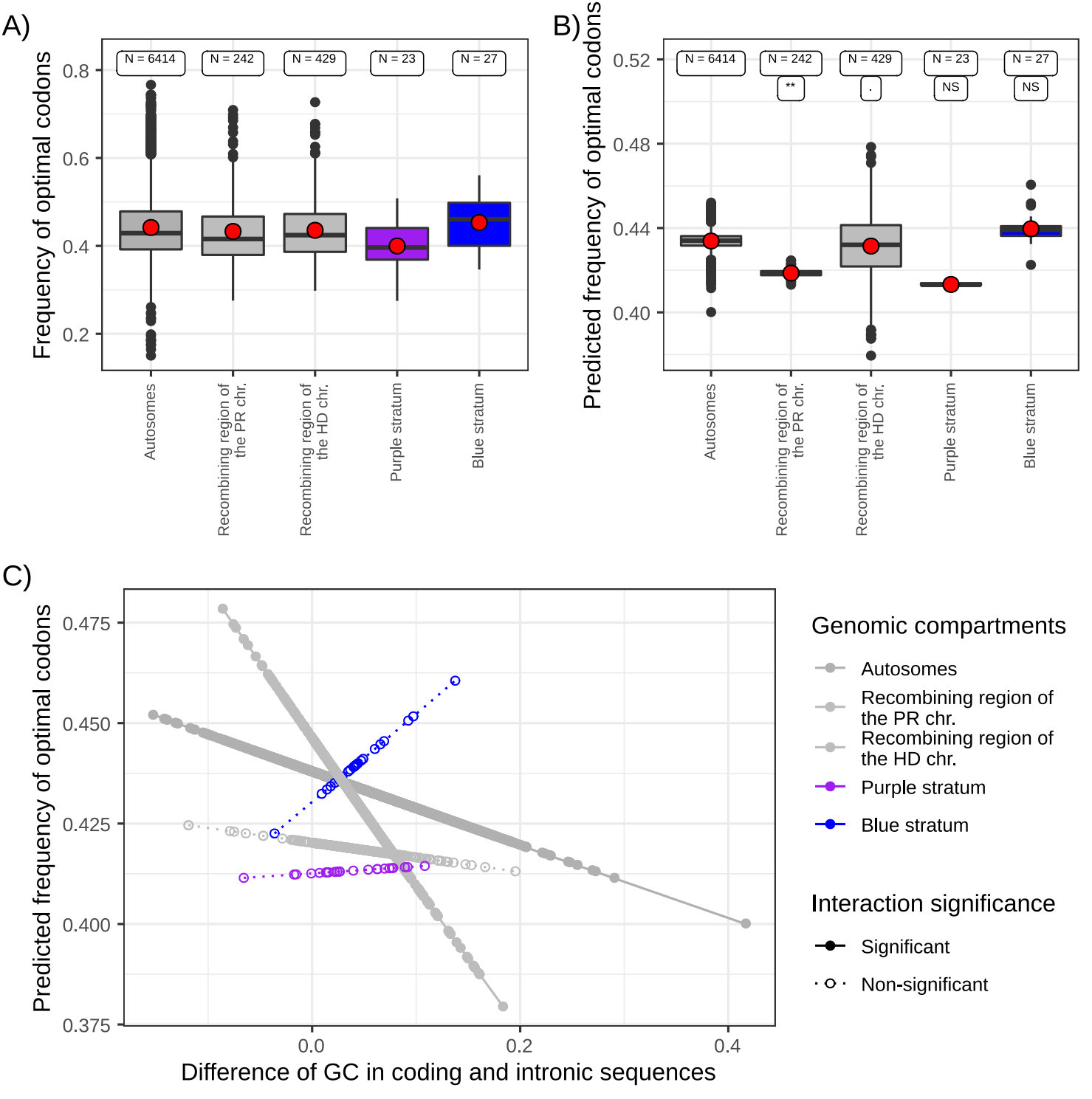
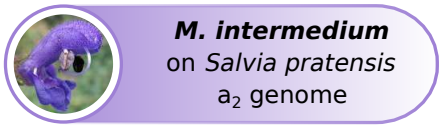


Figure S8

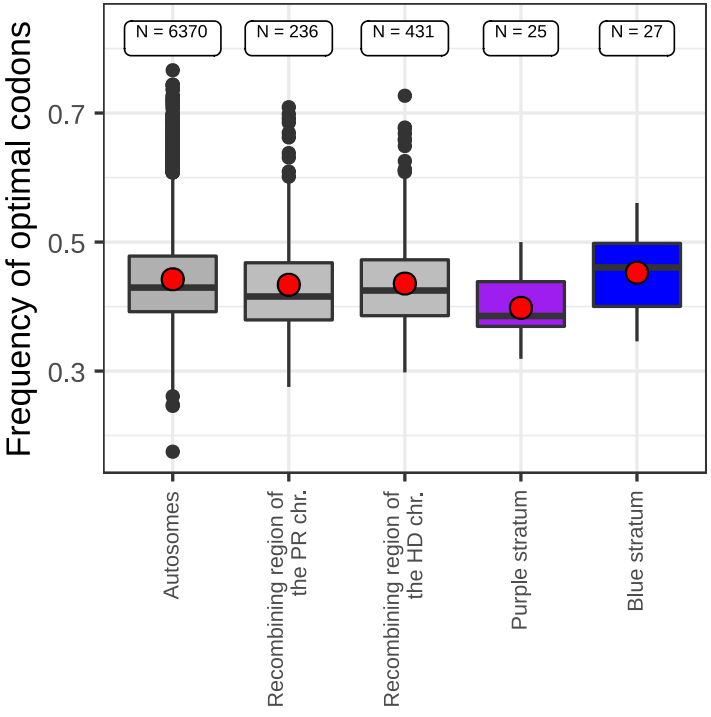

M. intermedium
on Salvia pratensis
a1 genome


A)
Frequency of optimal codons
0.8
0.6
0.4
0.2
N = 6414
N = 242
N = 429
N = 23
N = 27
Autosomes
Recombining region of the PR chr.
Recombining region of the HD chr.
Purple stratum
Blue stratum
B)
Predicted frequency of optimal codons
0.52
0.48
0.44
0.40
N = 6414
N = 242
N = 429
N = 23
N = 27
**
.
NS
NS
Autosomes
Recombining region of the PR chr.
Recombining region of the HD chr.
Purple stratum
Blue stratum
C)
Predicted frequency of optimal codons
0.475
0.450
0.425
0.400
0.375
0.0
0.2
0.4
Difference of GC in coding and intronic sequences
Genomic compartments
Autosomes
Recombining region of the PR chr.
Recombining region of the HD chr.
Purple stratum
Blue stratum
Interaction significance
Significant
Non-significant

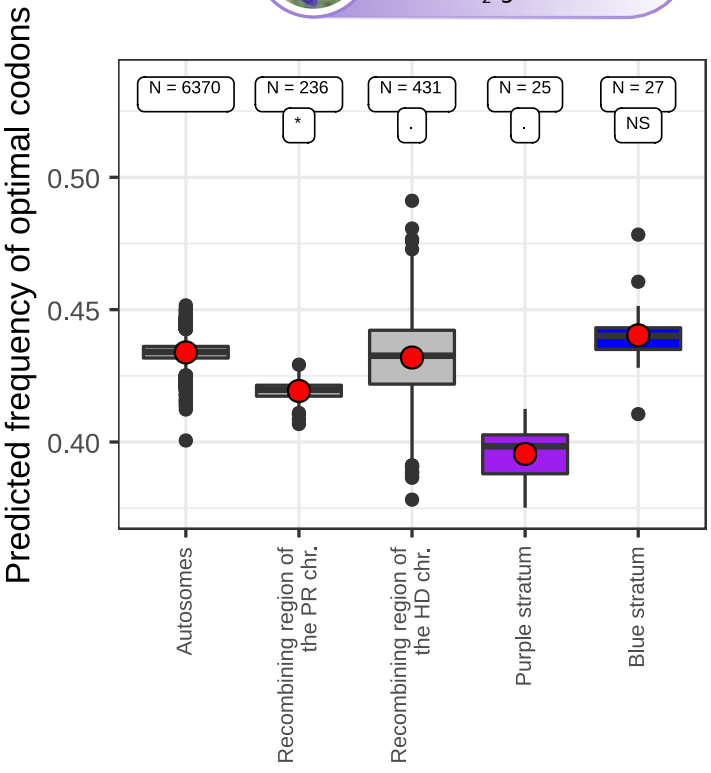


M. intermedium
on *Salvia pratensis*
a_2 genome

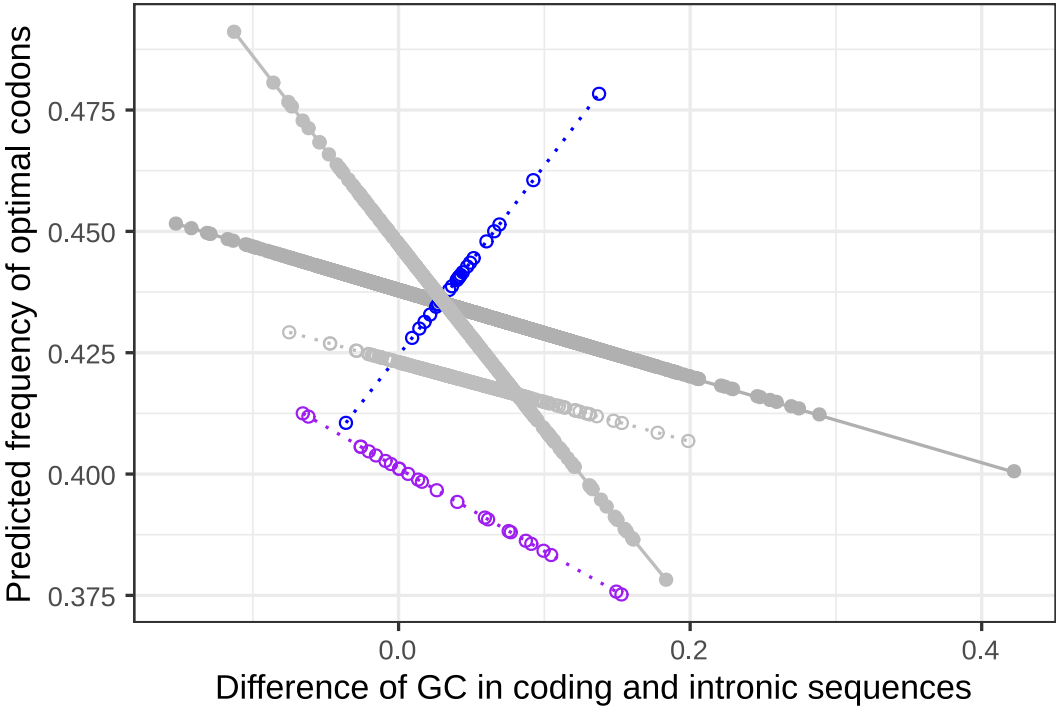
A)



B)



C)

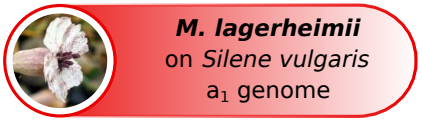


Genomic compartments

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

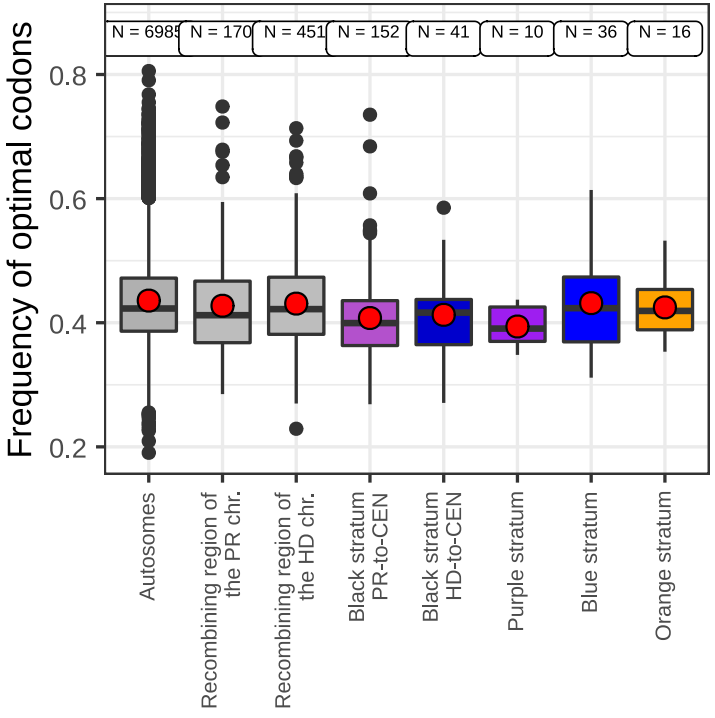
Interaction significance

- Significant
- Non-significant

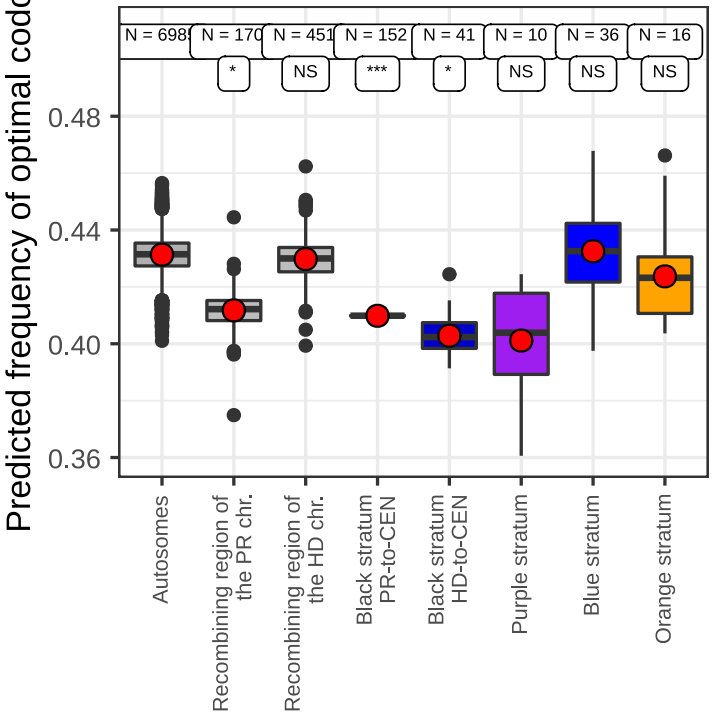


M. lagerheimii
on *Silene vulgaris*
a_1 genome

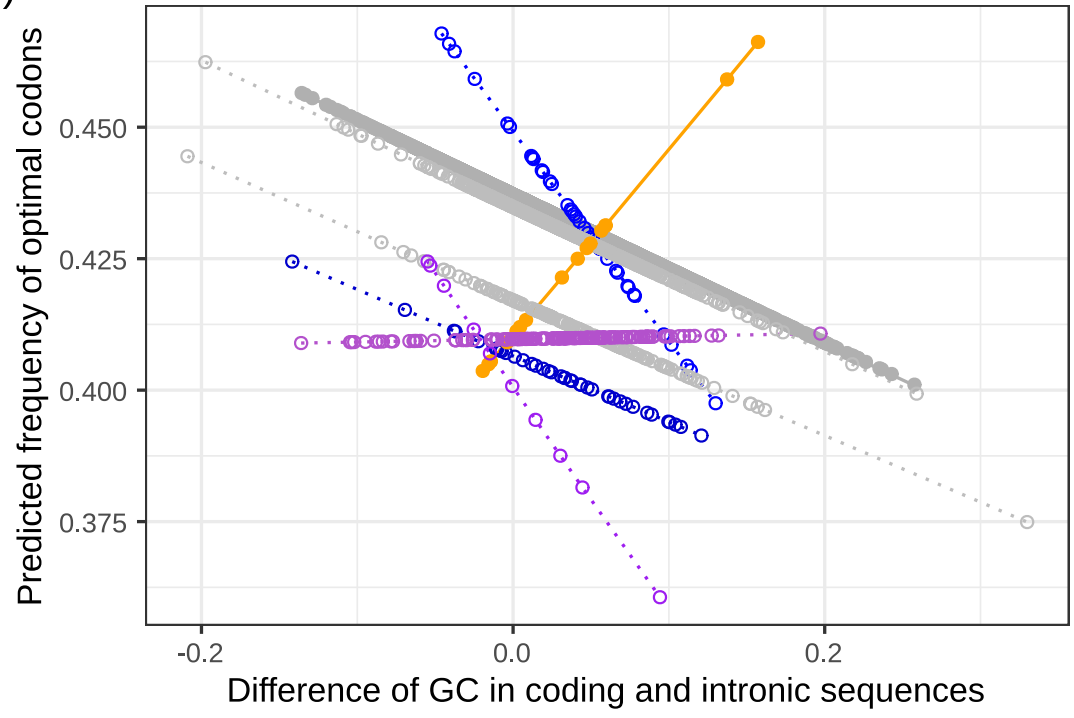
A)



B)



C)

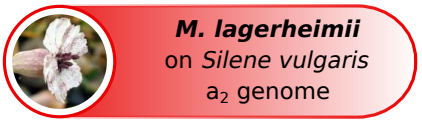


Genomic compartments

- 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

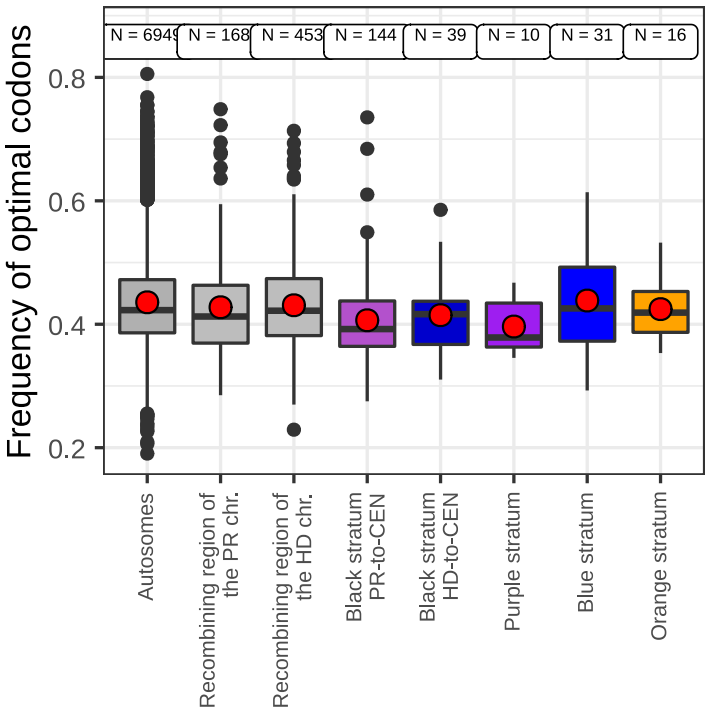
Interaction significance

- Significant
- Non-significant

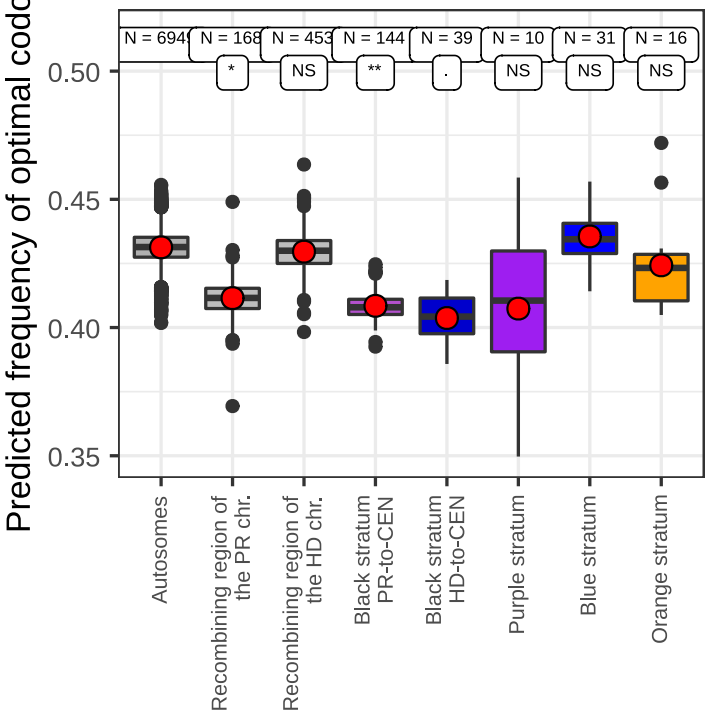


M. lagerheimii
on *Silene vulgaris*
a_2 genome

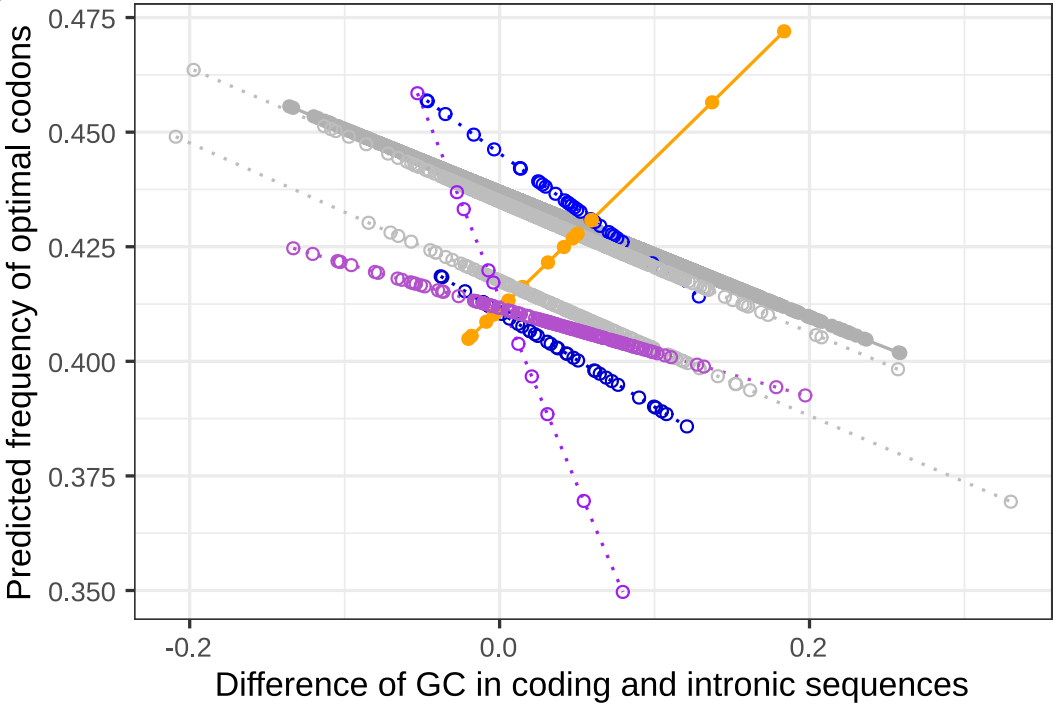
A)



B)



C)

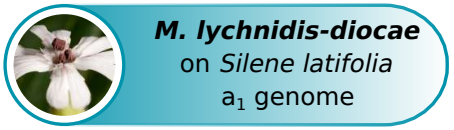


Genomic compartments

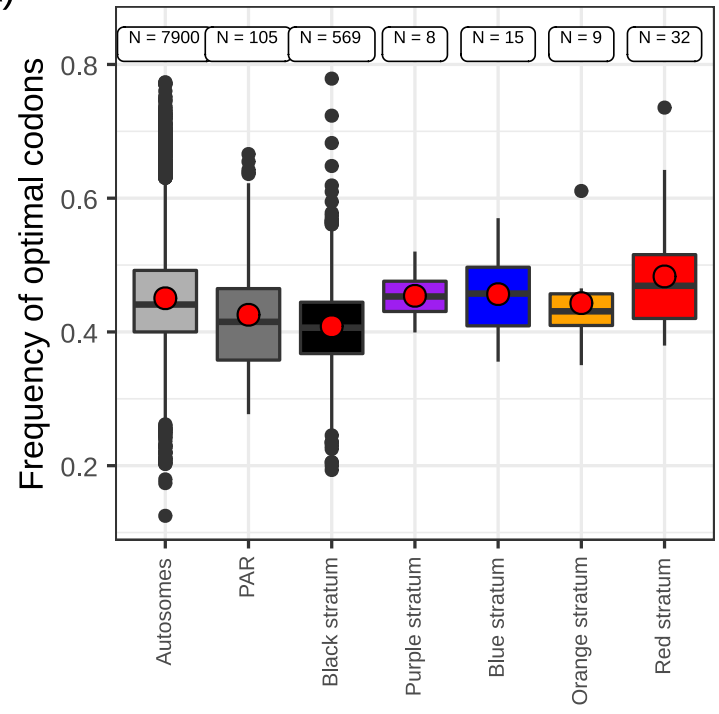
- Autosomes
- Recombining region of the PR chr.
- Recombining region of the HD chr.
- Black stratum PR-to-CEN
- Black stratum HD-to-CEN
- Purple stratum
- Blue stratum
- Orange stratum

Interaction significance

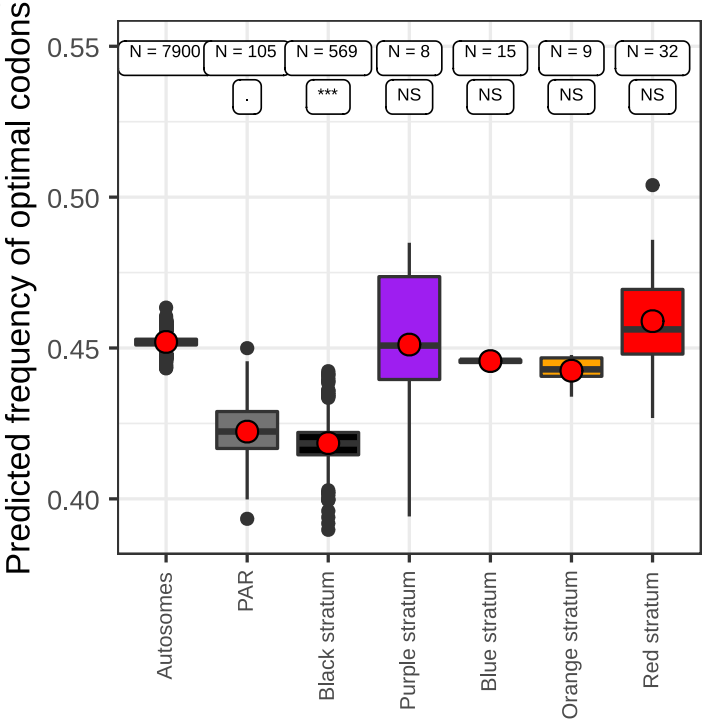
- Significant
- Non-significant


M. lychnidis-diocae
on Silene latifolia
a1 genome

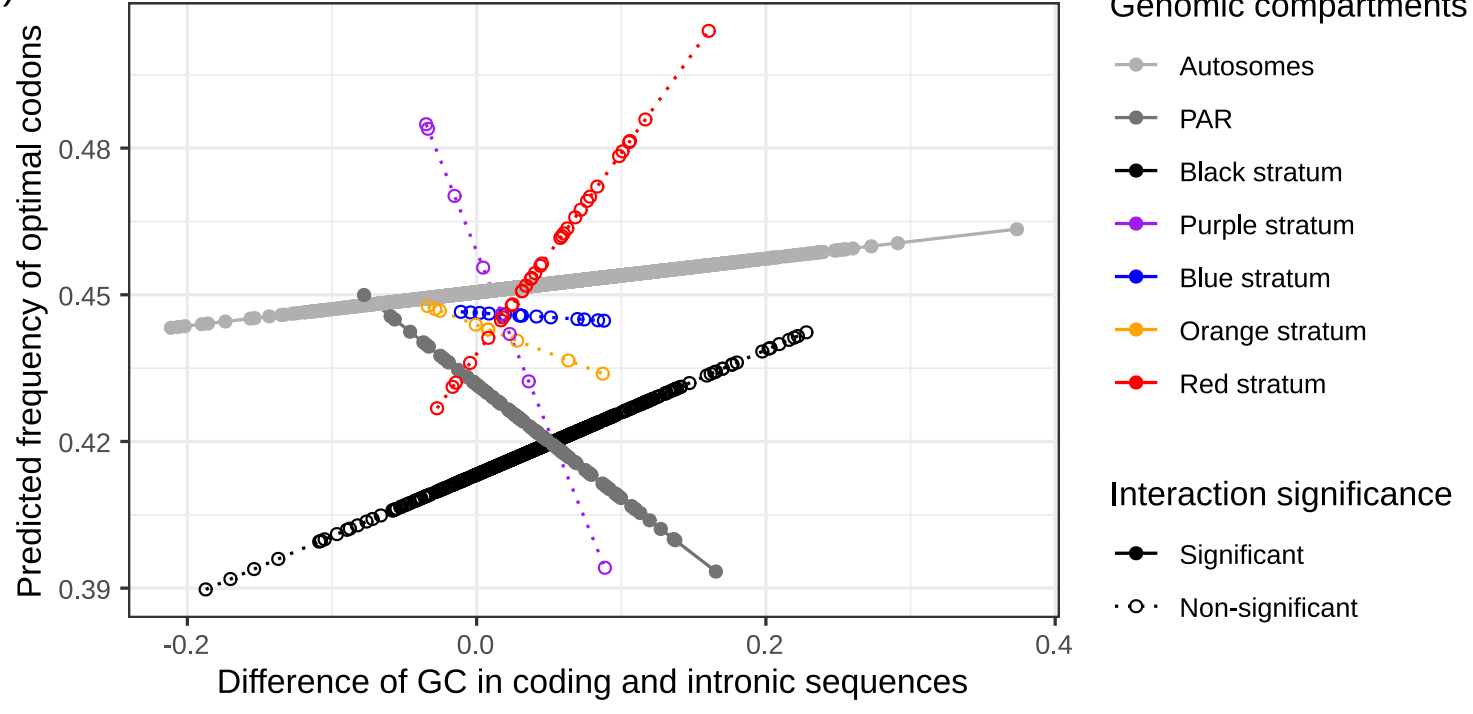
A)

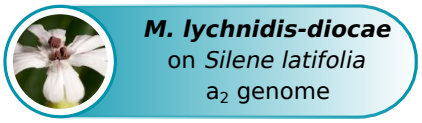

N = 7900
N = 105
N = 569
N = 8
N = 15
N = 9
N = 32
Frequency of optimal codons
0.8
0.6
0.4
0.2
Autosomes
PAR
Black stratum
Purple stratum
Blue stratum
Orange stratum
Red stratum

B)

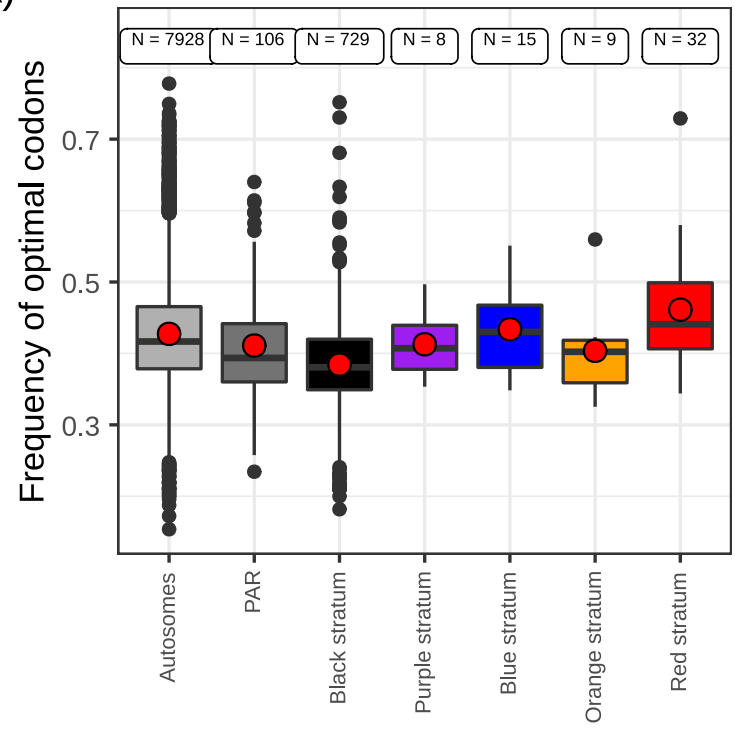

N = 7900
N = 105
N = 569
N = 8
N = 15
N = 9
N = 32
.
***
NS
NS
NS
NS
Predicted frequency of optimal codons
0.55
0.50
0.45
0.40
Autosomes
PAR
Black stratum
Purple stratum
Blue stratum
Orange stratum
Red stratum

C)

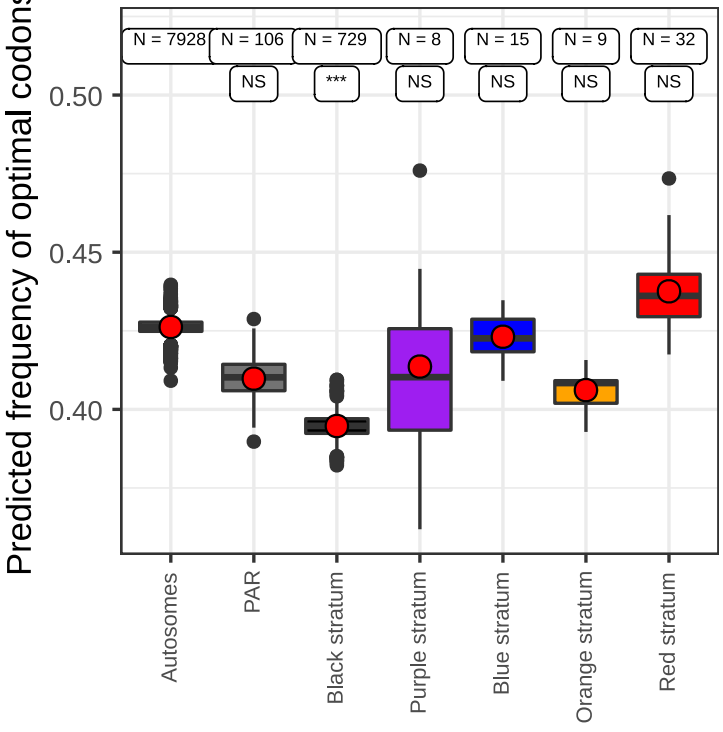

Predicted frequency of optimal codons
0.48
0.45
0.42
0.39
-0.2
0.0
0.2
0.4
Difference of GC in coding and intronic sequences
Genomic compartments
Autosomes
PAR
Black stratum
Purple stratum
Blue stratum
Orange stratum
Red stratum
Interaction significance
Significant
Non-significant


M. lychnidis-diocae
on Silene latifolia
a2 genome

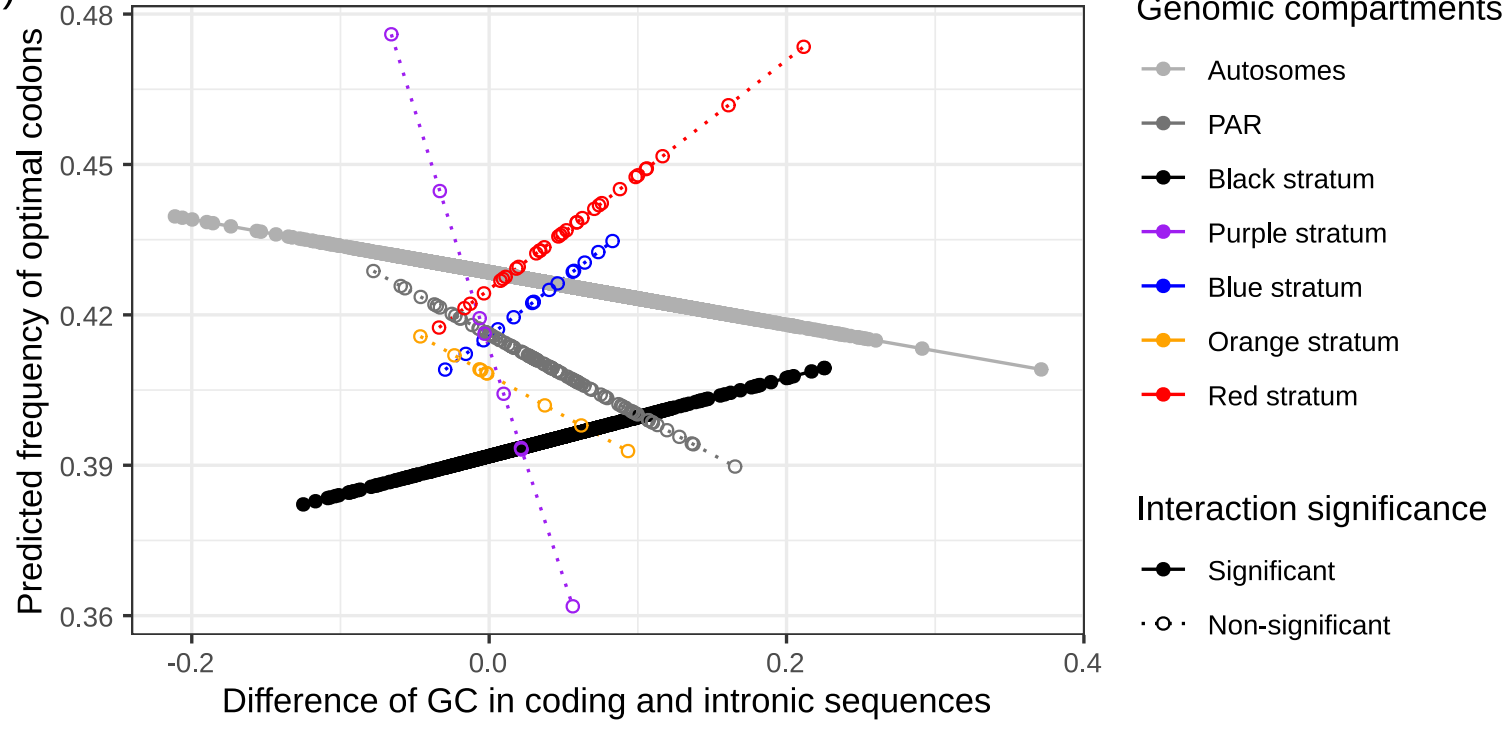
A)

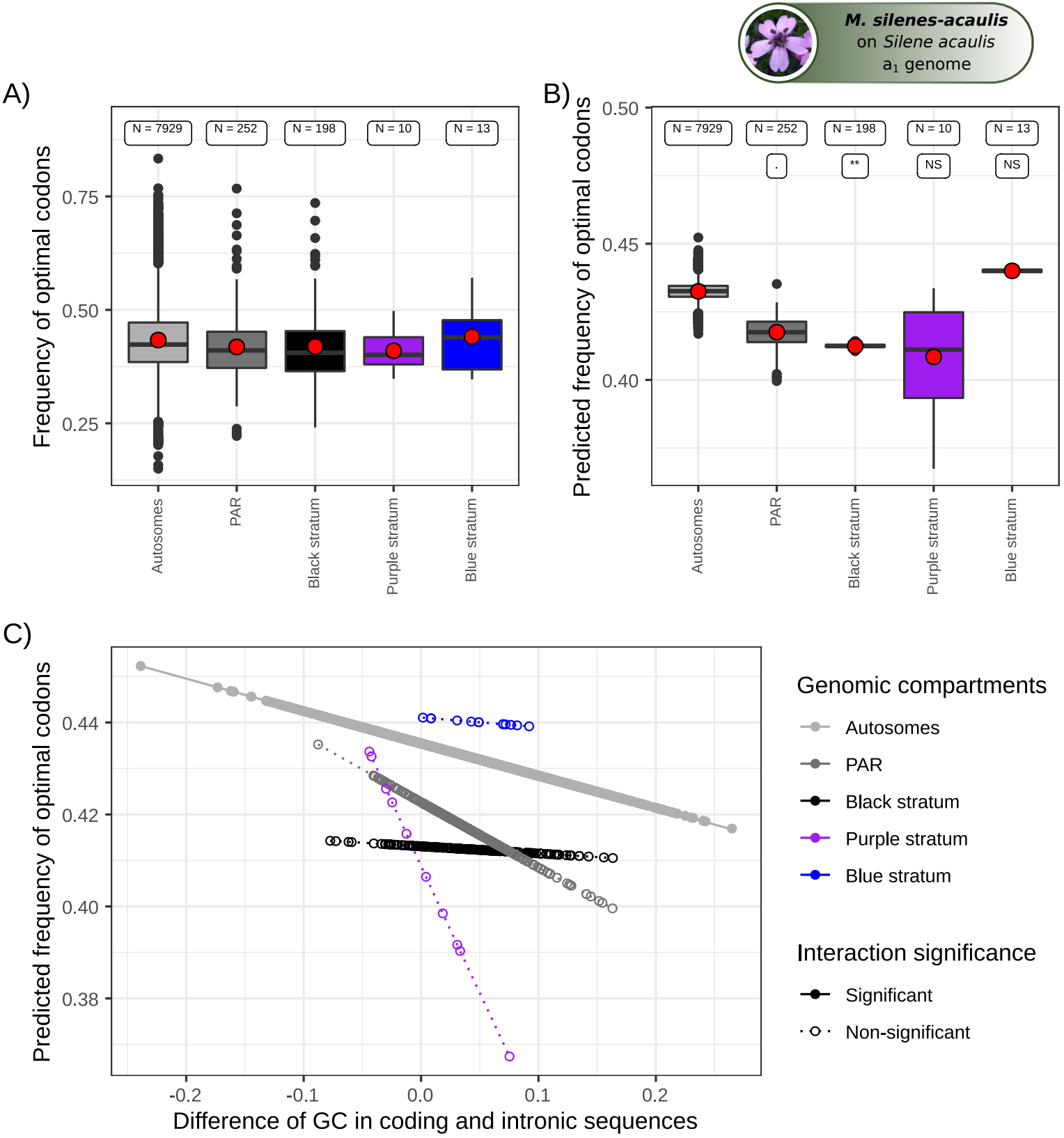

N = 7928
N = 106
N = 729
N = 8
N = 15
N = 9
N = 32
Frequency of optimal codons
0.7
0.5
0.3
Autosomes
PAR
Black stratum
Purple stratum
Blue stratum
Orange stratum
Red stratum

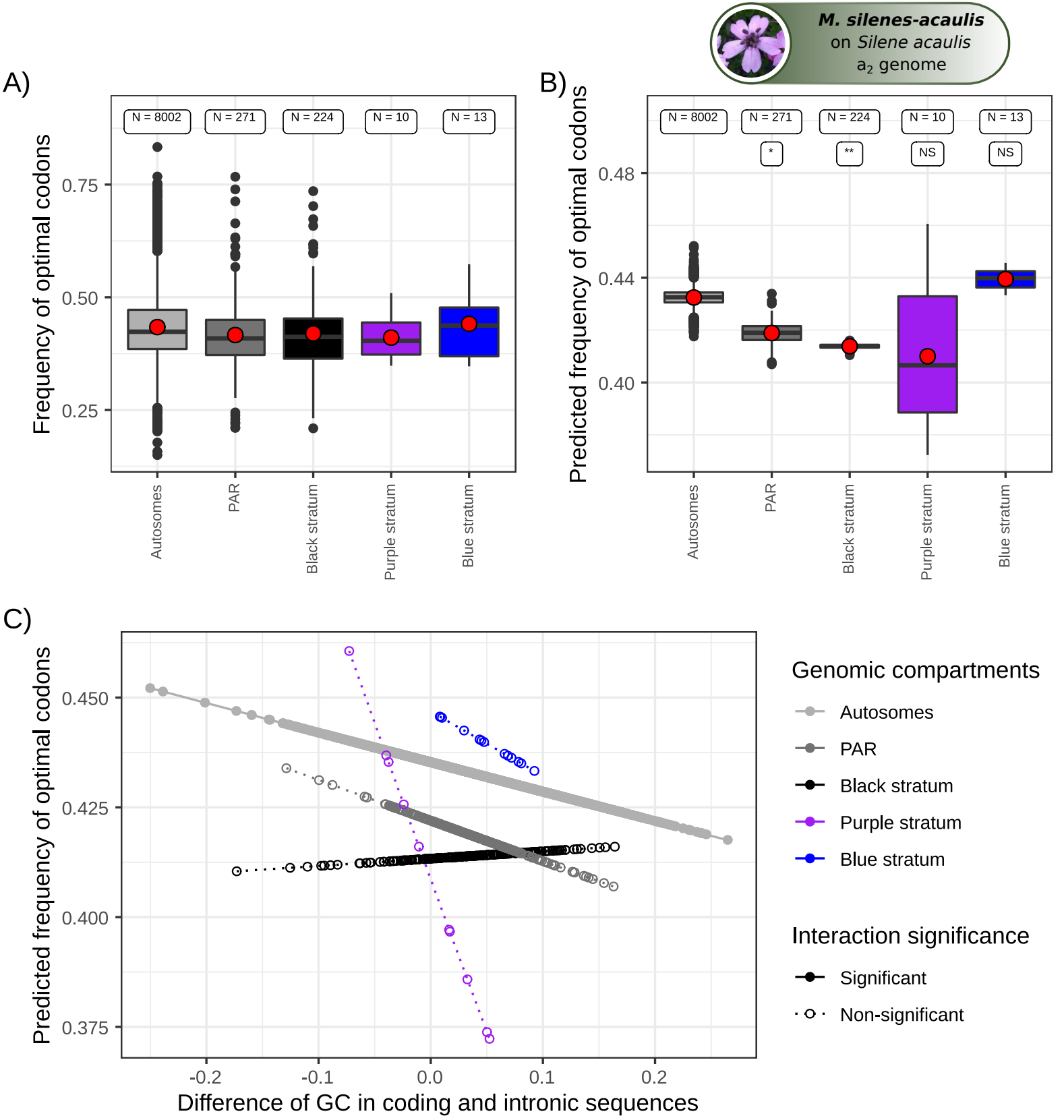
B)

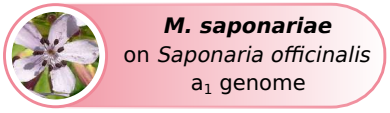

N = 7928
N = 106
N = 729
N = 8
N = 15
N = 9
N = 32
NS
***
NS
NS
NS
NS
Predicted frequency of optimal codons
0.50
0.45
0.40
Autosomes
PAR
Black stratum
Purple stratum
Blue stratum
Orange stratum
Red stratum

C)


Predicted frequency of optimal codons
0.48
0.45
0.42
0.39
0.36
-0.2
0.0
0.2
0.4
Difference of GC in coding and intronic sequences
Genomic compartments
Autosomes
PAR
Black stratum
Purple stratum
Blue stratum
Orange stratum
Red stratum
Interaction significance
Significant
Non-significant

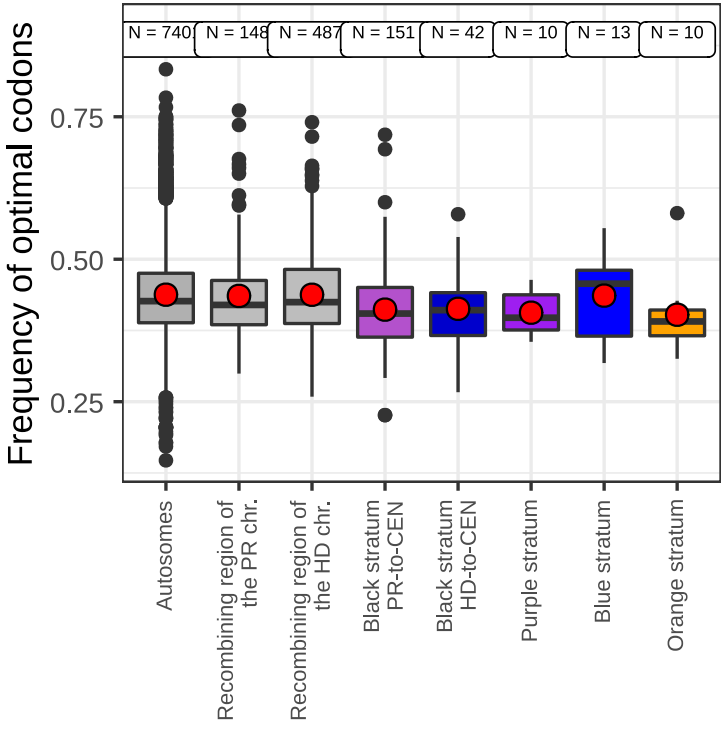

M. silenes-acaulis
on Silene acaulis
a1 genome
A)
Frequency of optimal codons
0.75
0.50
0.25
N = 7929
N = 252
N = 198
N = 10
N = 13
Autosomes
PAR
Black stratum
Purple stratum
Blue stratum
B)
Predicted frequency of optimal codons
0.50
0.45
0.40
N = 7929
N = 252
N = 198
N = 10
N = 13
.
**
NS
NS
Autosomes
PAR
Black stratum
Purple stratum
Blue stratum
C)
Predicted frequency of optimal codons
0.44
0.42
0.40
0.38
-0.2
-0.1
0.0
0.1
0.2
Difference of GC in coding and intronic sequences
Genomic compartments
Autosomes
PAR
Black stratum
Purple stratum
Blue stratum
Interaction significance
Significant
Non-significant


M. silenes-acaulis
on Silene acaulis
a2 genome
A)
Frequency of optimal codons
0.75
0.50
0.25
N = 8002
N = 271
N = 224
N = 10
N = 13
Autosomes
PAR
Black stratum
Purple stratum
Blue stratum
B)
Predicted frequency of optimal codons
0.48
0.44
0.40
*
**
NS
NS
C)
Predicted frequency of optimal codons
0.450
0.425
0.400
0.375
-0.2
-0.1
0.0
0.1
0.2
Difference of GC in coding and intronic sequences
Genomic compartments
Autosomes
PAR
Black stratum
Purple stratum
Blue stratum
Interaction significance
Significant
Non-significant

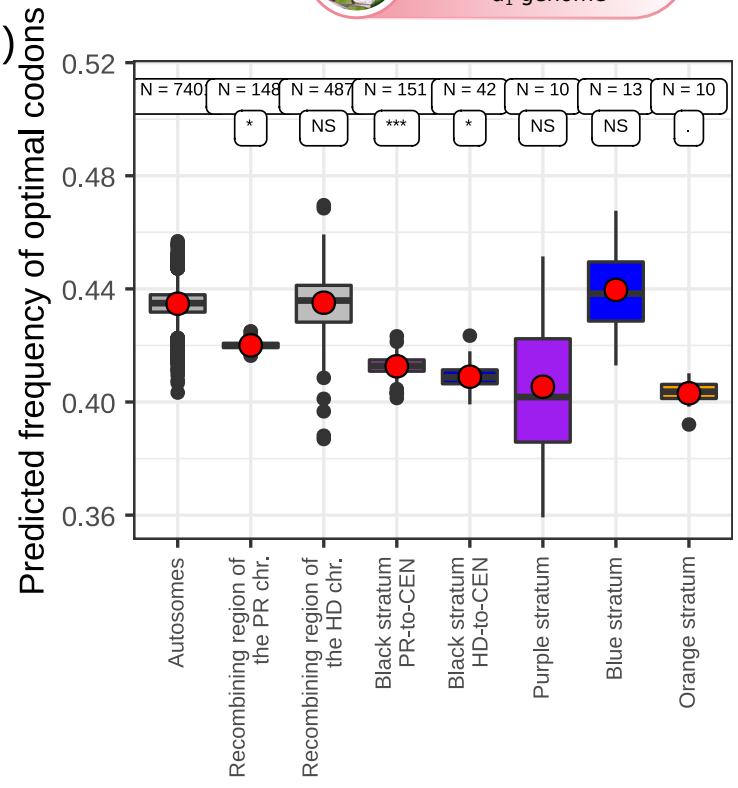


M. saponariae
on *Saponaria officinalis*
a_1 genome

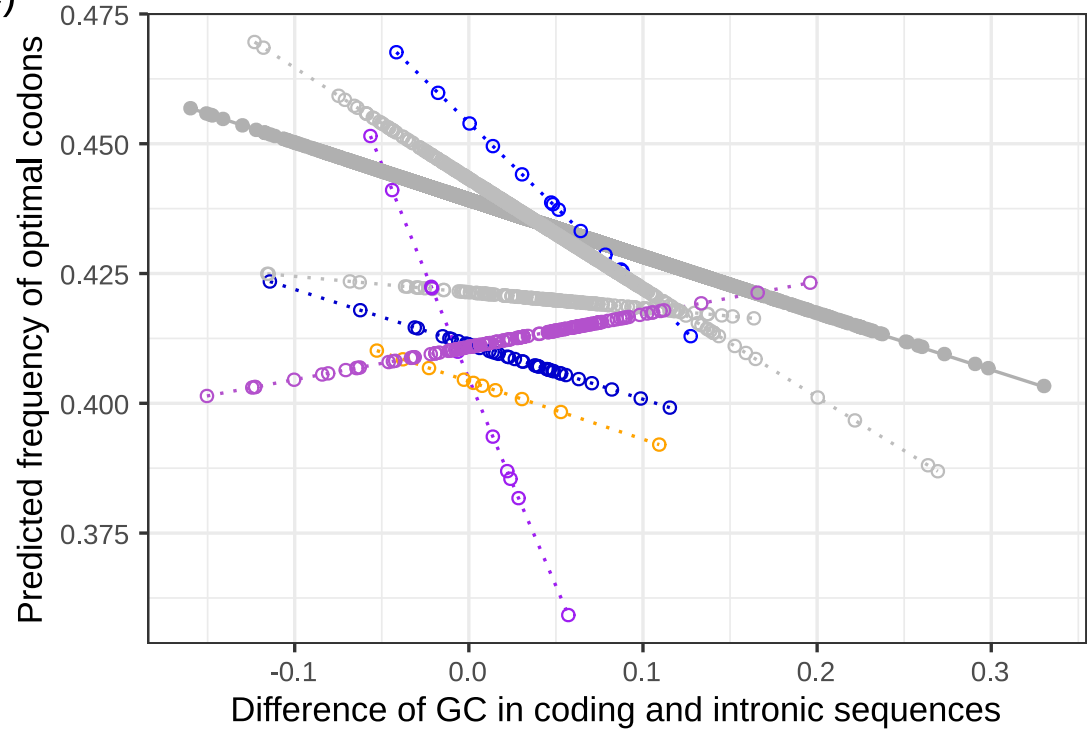
A)



B)



C)

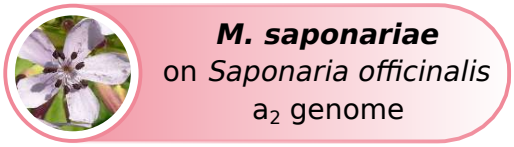


Genomic compartments

- 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

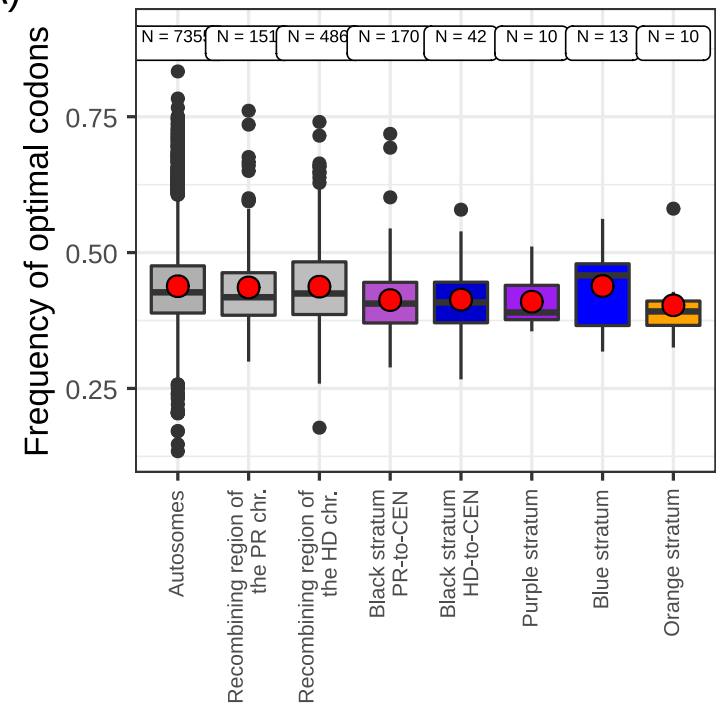
Interaction significance

- Significant
- Non-significant

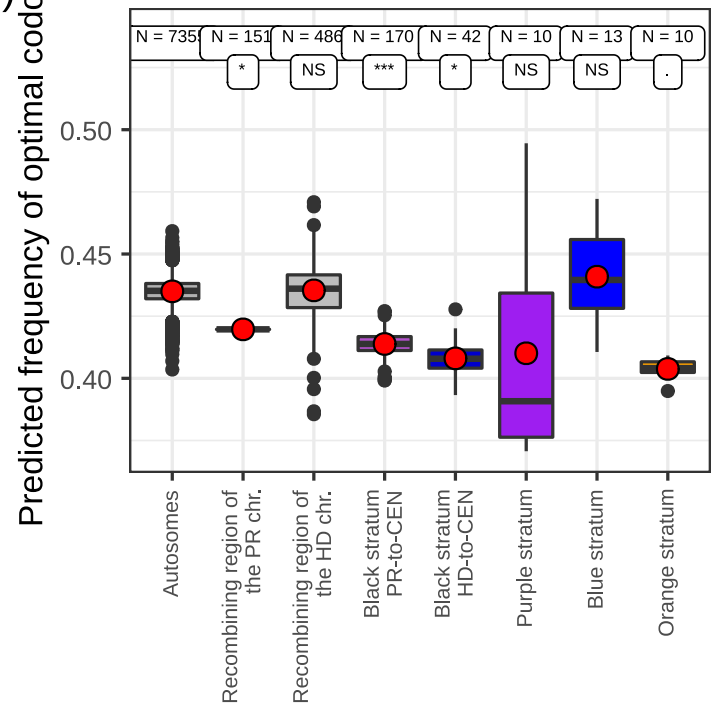


M. saponariae
on *Saponaria officinalis*
a_2 genome

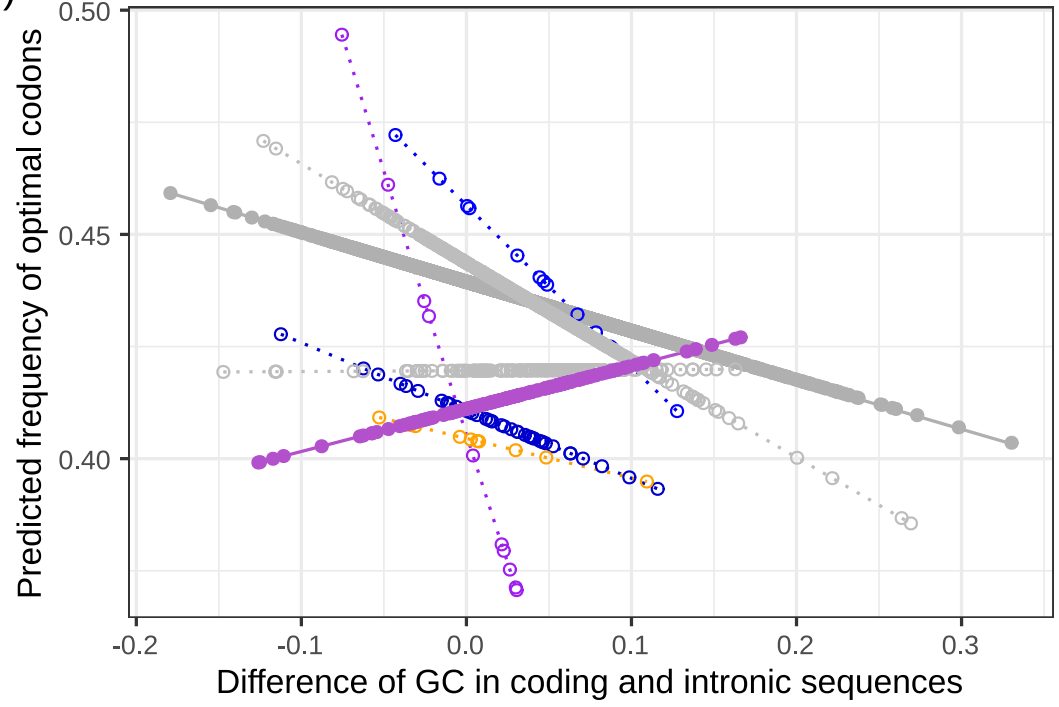
A)



B)



C)

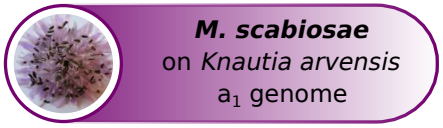


Genomic compartments

- 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

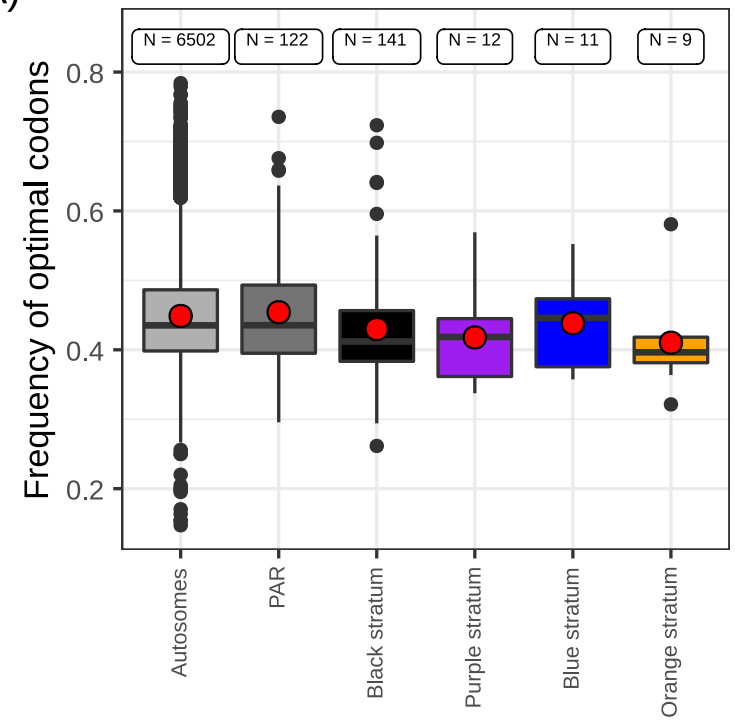
Interaction significance

- Significant
- Non-significant

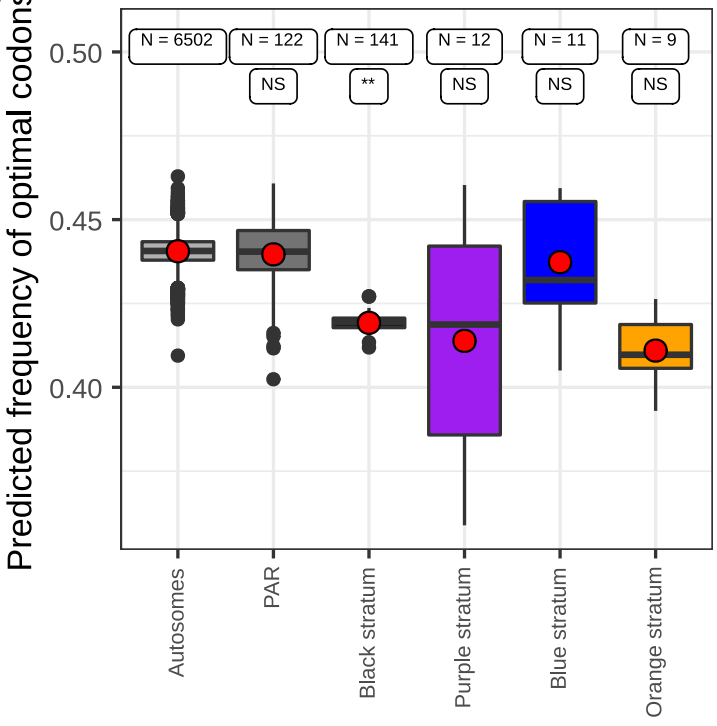


M. scabiosae
on *Knautia arvensis*
a_1 genome

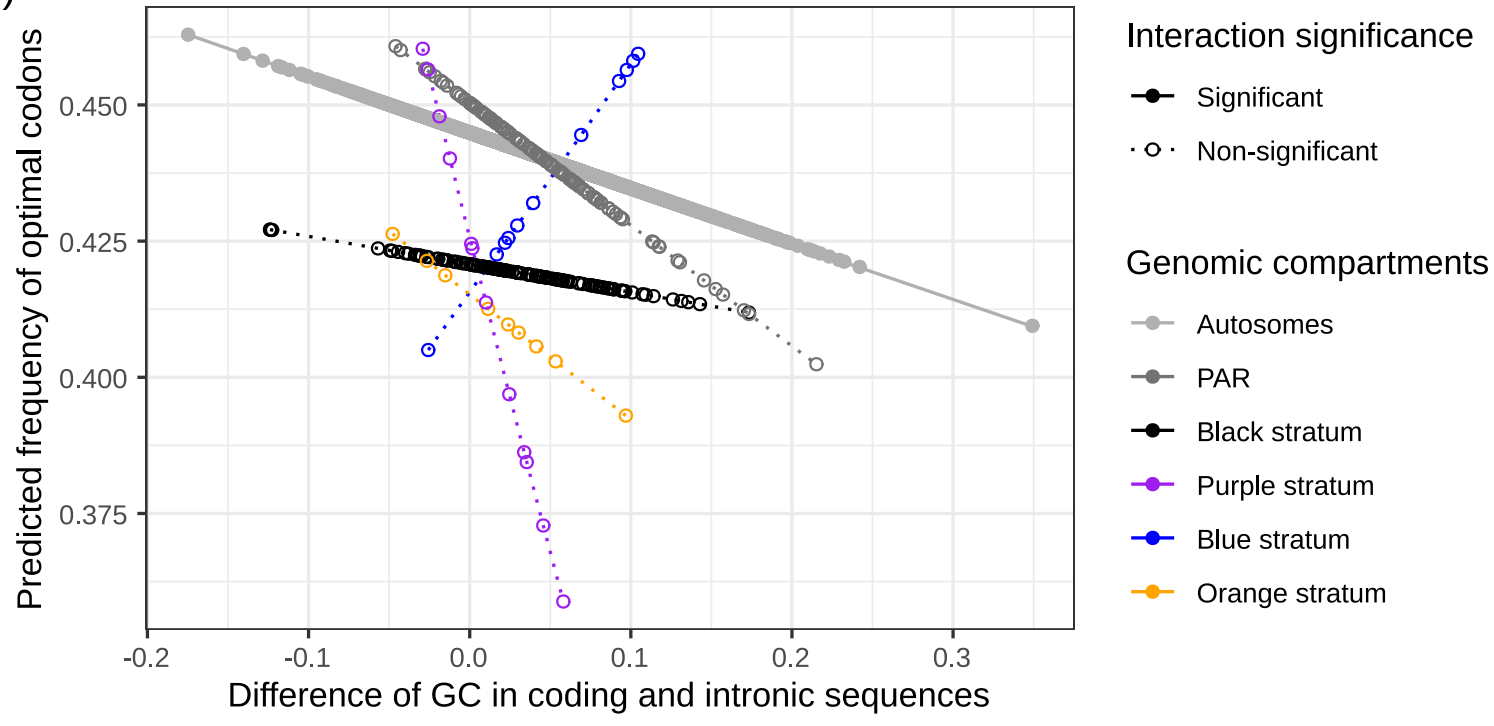
A)

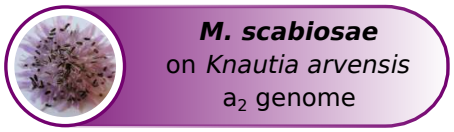


B)

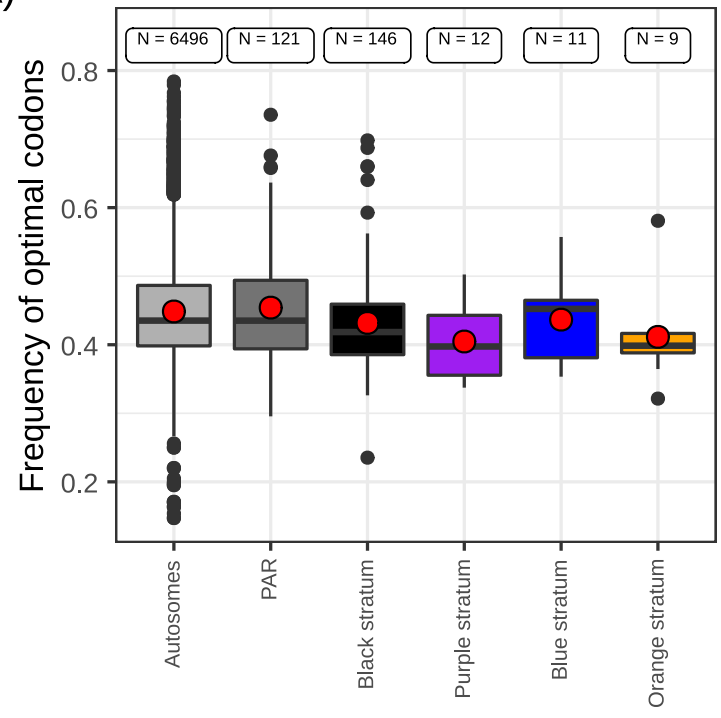


C)

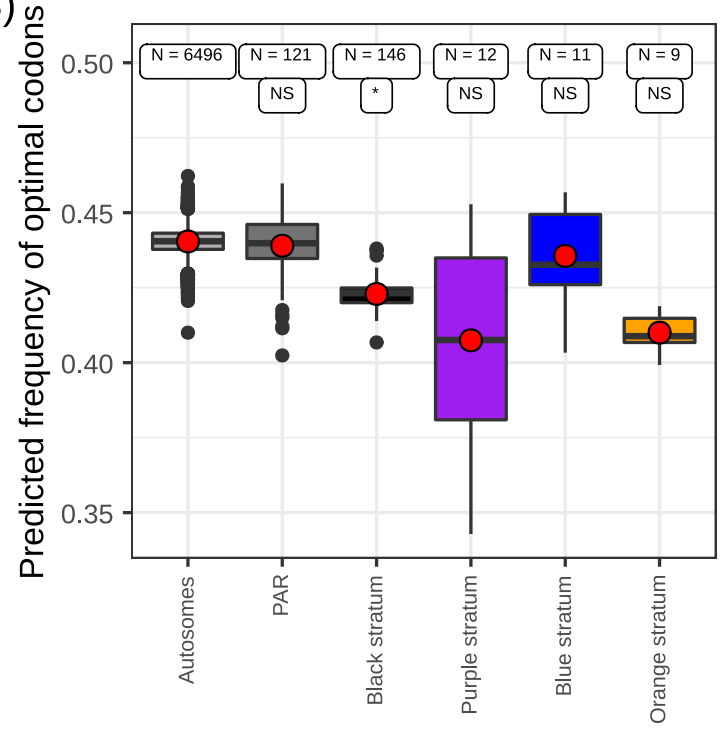



M. scabiosae
on Knautia arvensis
a2 genome

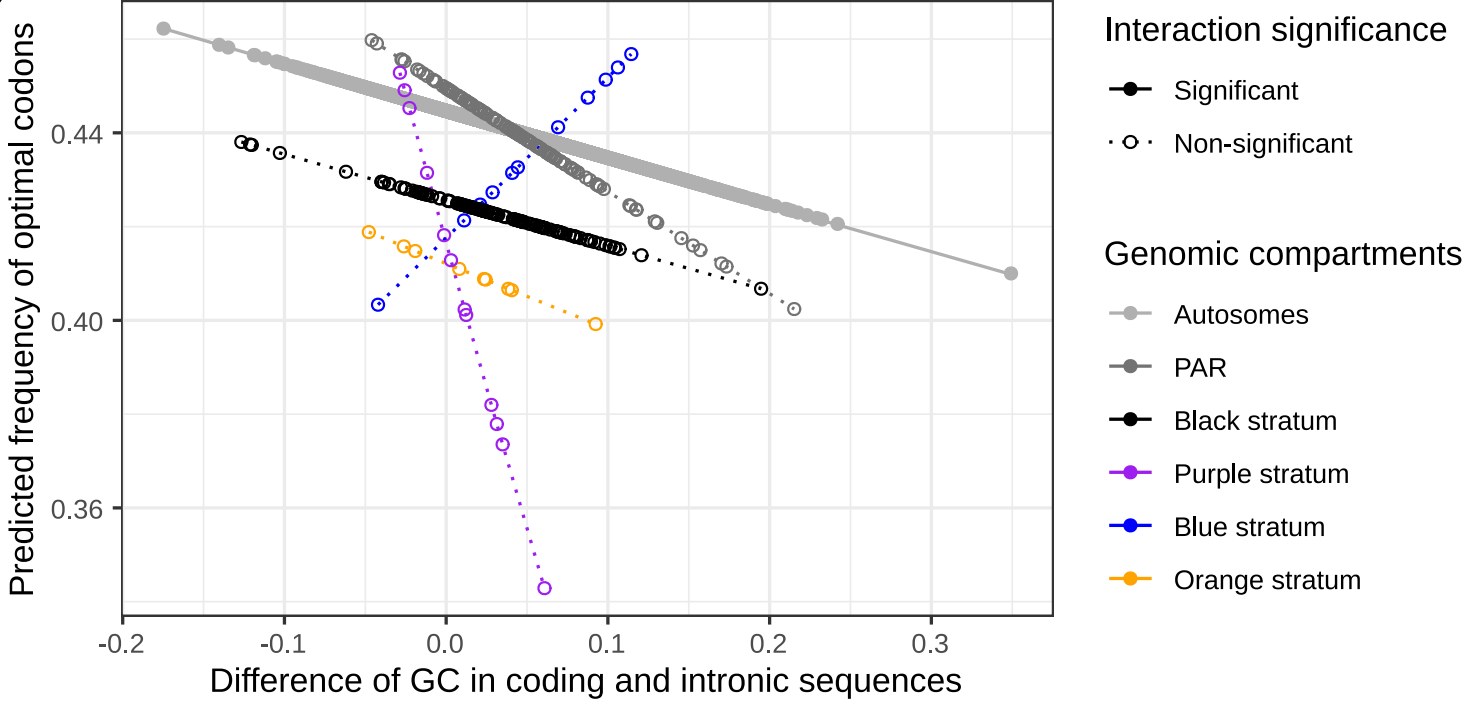
A)

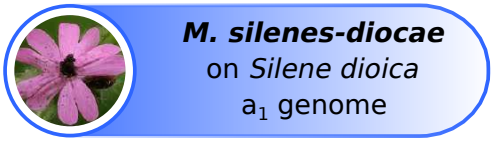

N = 6496
N = 121
N = 146
N = 12
N = 11
N = 9
Frequency of optimal codons
0.8
0.6
0.4
0.2
Autosomes
PAR
Black stratum
Purple stratum
Blue stratum
Orange stratum

B)


N = 6496
N = 121
NS
N = 146
*
N = 12
NS
N = 11
NS
N = 9
NS
Predicted frequency of optimal codons
0.50
0.45
0.40
0.35
Autosomes
PAR
Black stratum
Purple stratum
Blue stratum
Orange stratum

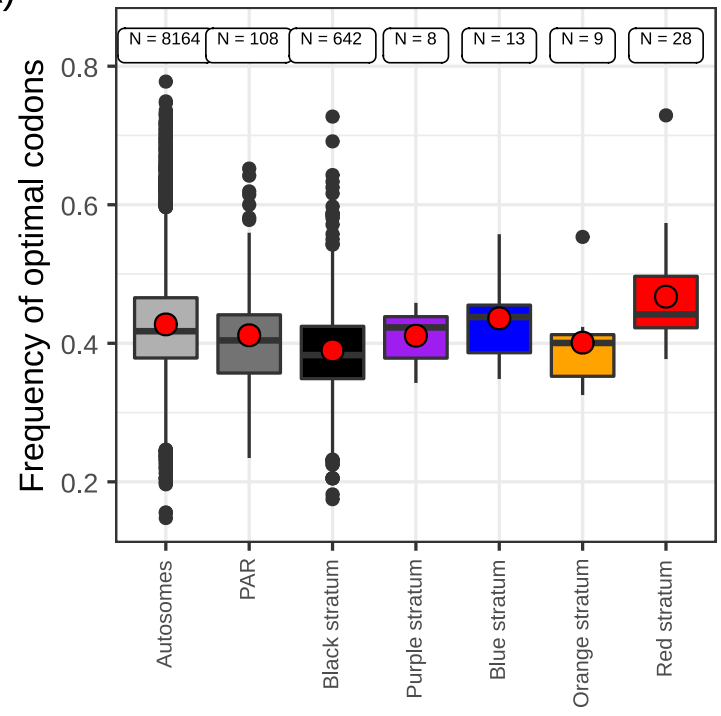
C)


Predicted frequency of optimal codons
0.44
0.40
0.36
-0.2
-0.1
0.0
0.1
0.2
0.3
Difference of GC in coding and intronic sequences
Interaction significance
Significant
Non-significant
Genomic compartments
Autosomes
PAR
Black stratum
Purple stratum
Blue stratum
Orange stratum

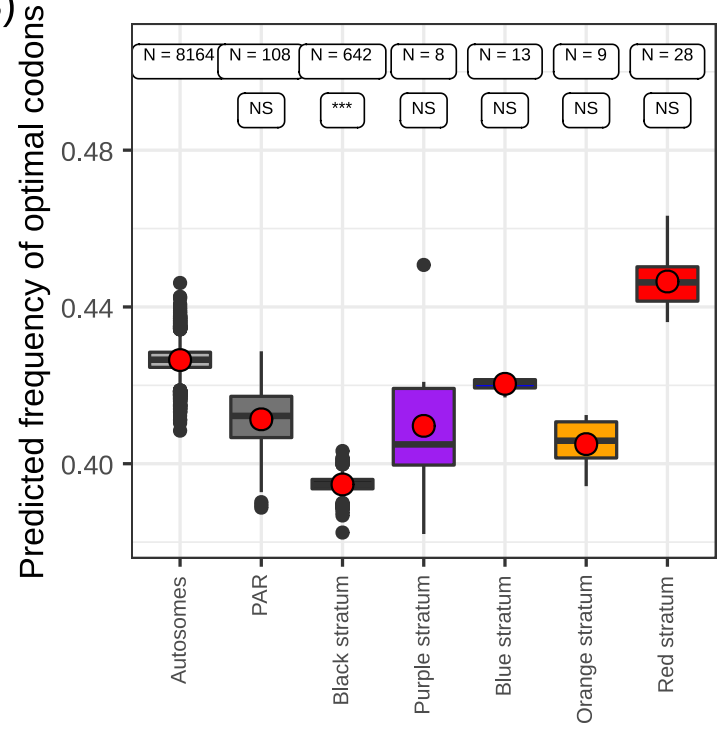


M. silenes-diocae
on *Silene dioica*
a_1 genome

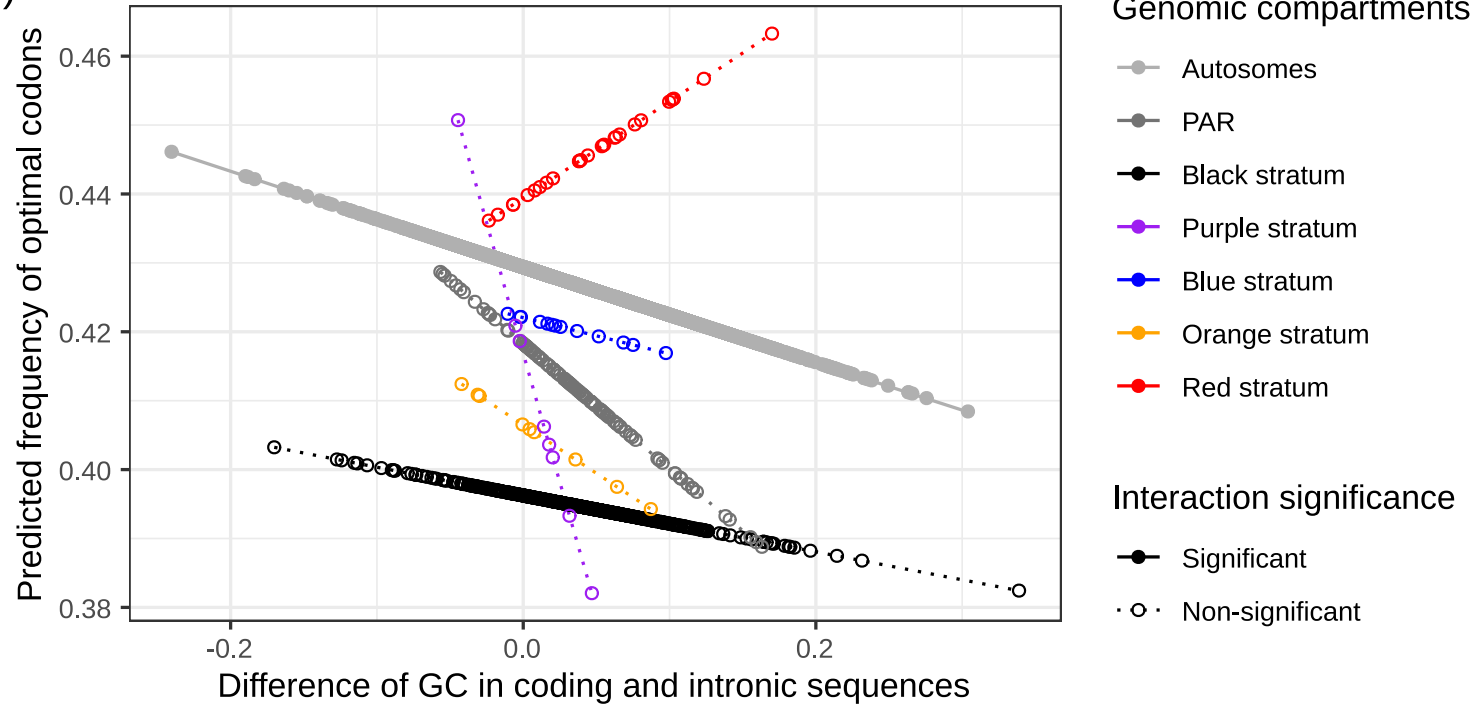
A)

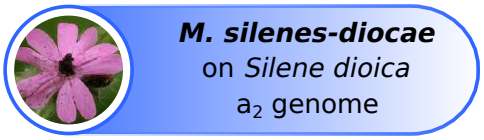


B)

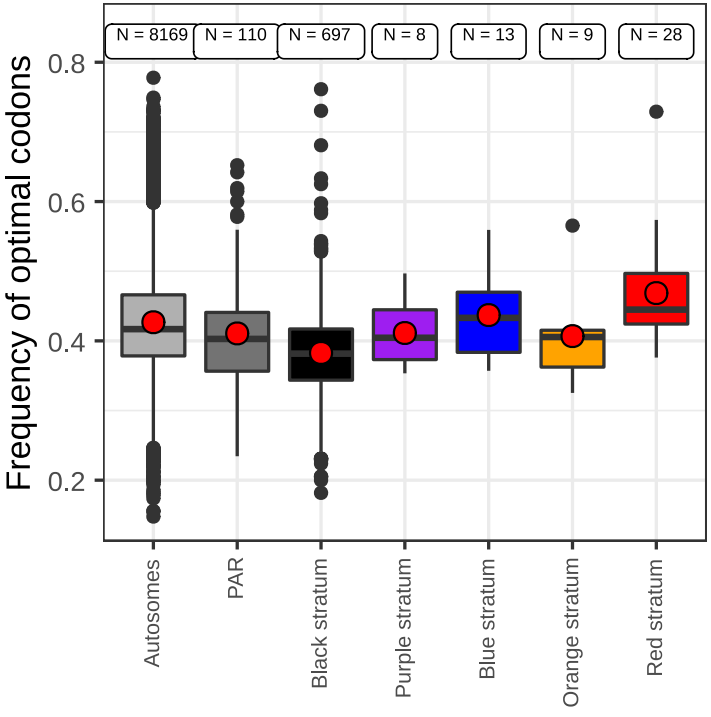


C)

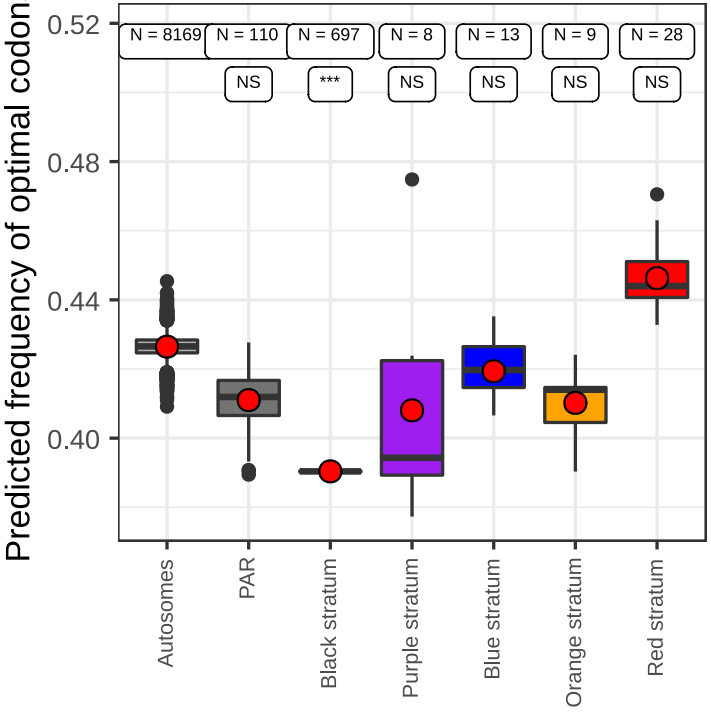




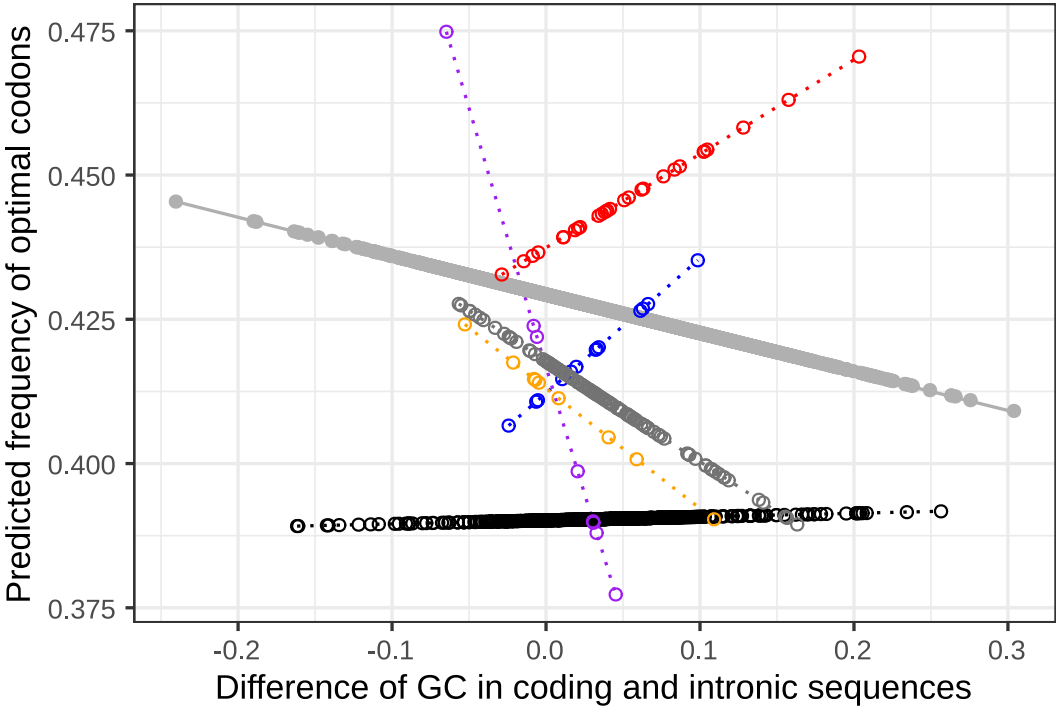
A)



B)



C)

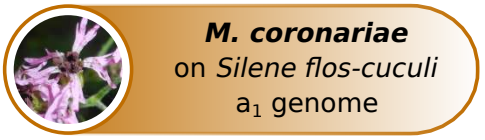


Genomic compartments

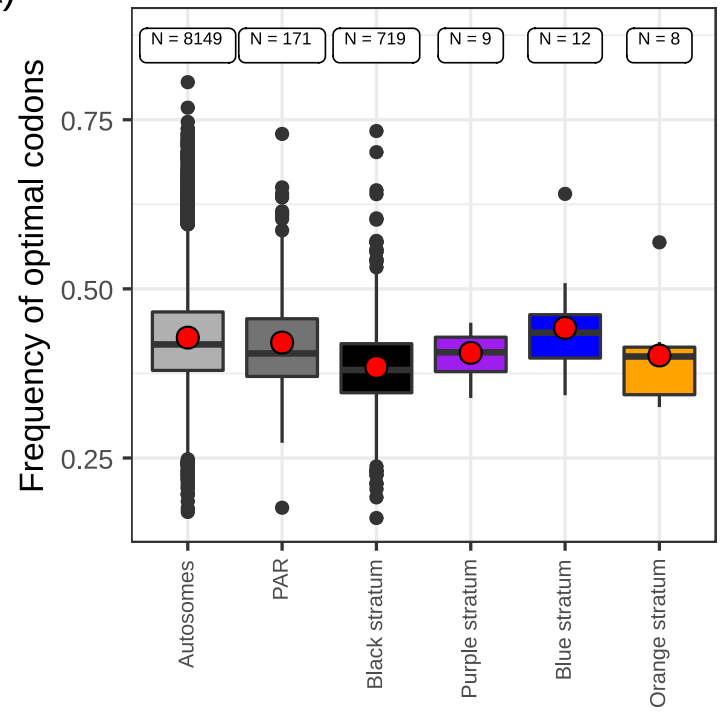
- Autosomes
- PAR
- Black stratum
- Purple stratum
- Blue stratum
- Orange stratum
- Red stratum

Interaction significance

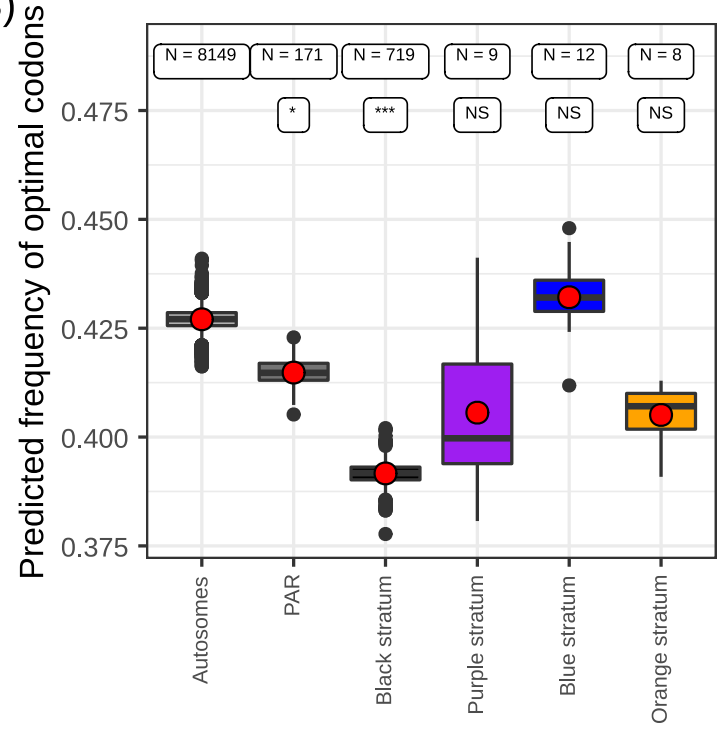
- Significant
- Non-significant


M. coronariae
on Silene flos-cuculi
a1 genome

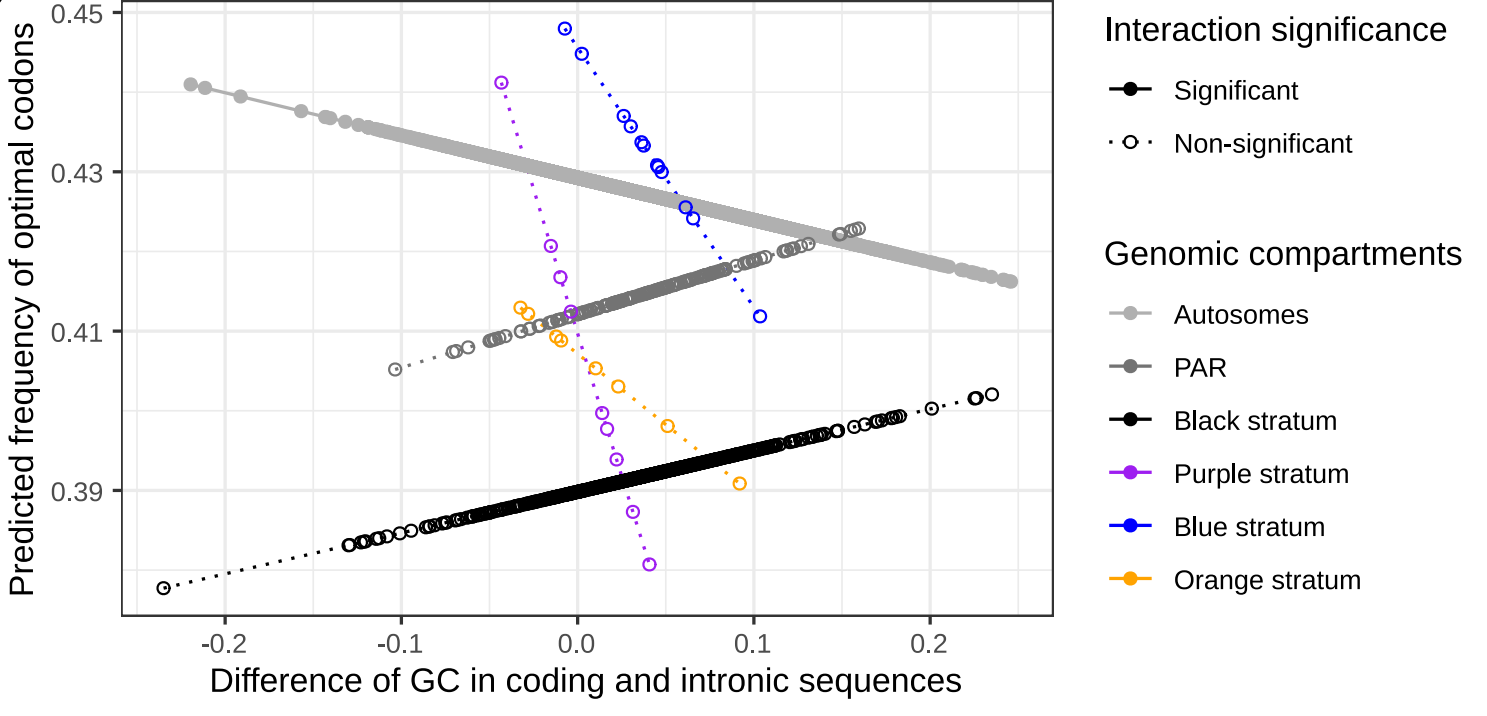
A)

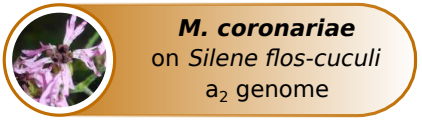

N = 8149
N = 171
N = 719
N = 9
N = 12
N = 8
Frequency of optimal codons
0.75
0.50
0.25
Autosomes
PAR
Black stratum
Purple stratum
Blue stratum
Orange stratum

B)

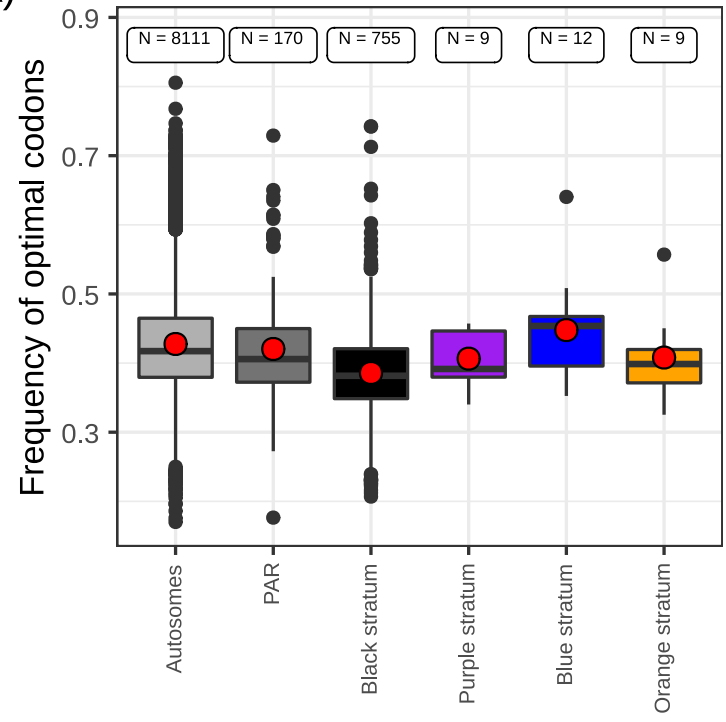

N = 8149
N = 171
N = 719
N = 9
N = 12
N = 8
*
***
NS
NS
NS
Predicted frequency of optimal codons
0.475
0.450
0.425
0.400
0.375
Autosomes
PAR
Black stratum
Purple stratum
Blue stratum
Orange stratum

C)

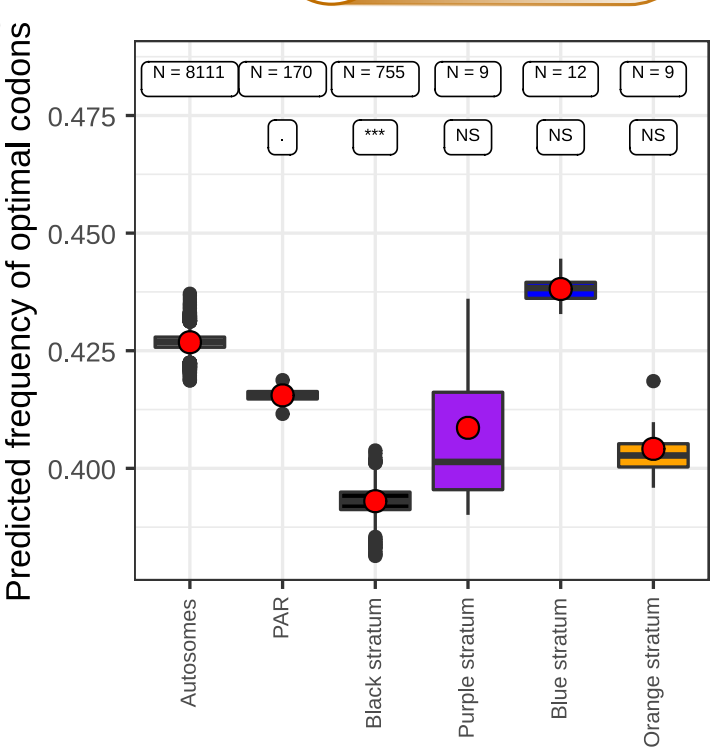

Predicted frequency of optimal codons
0.45
0.43
0.41
0.39
-0.2
-0.1
0.0
0.1
0.2
Difference of GC in coding and intronic sequences
Interaction significance
Significant
Non-significant
Genomic compartments
Autosomes
PAR
Black stratum
Purple stratum
Blue stratum
Orange stratum



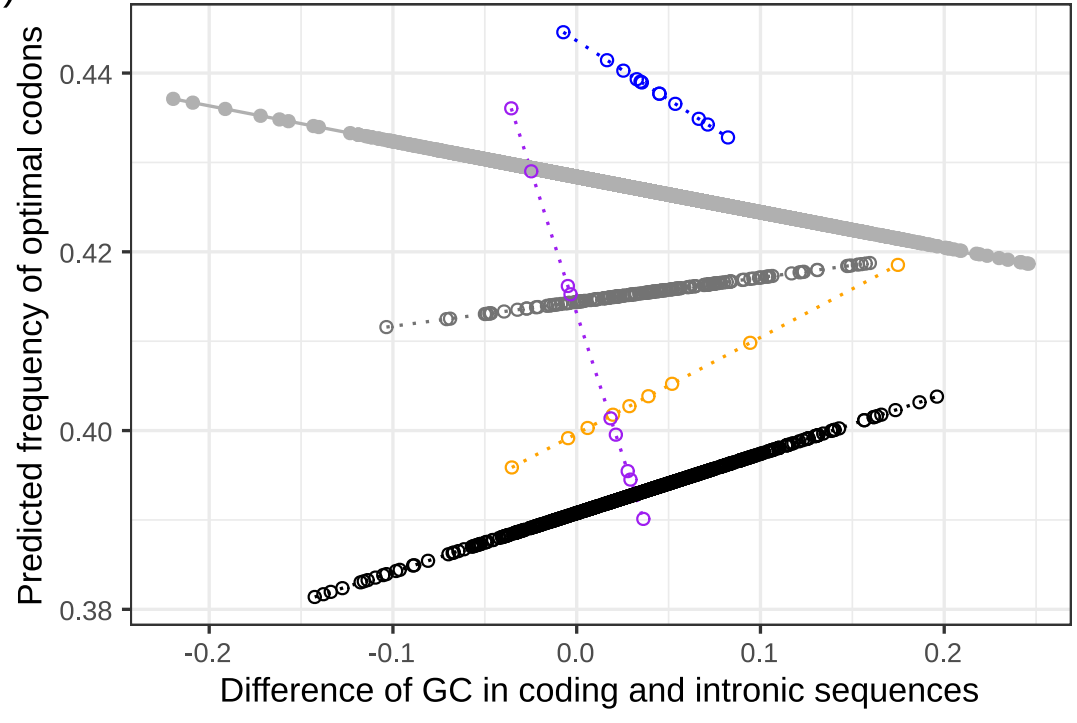
A)



B)



C)

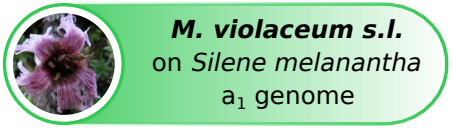


Interaction significance

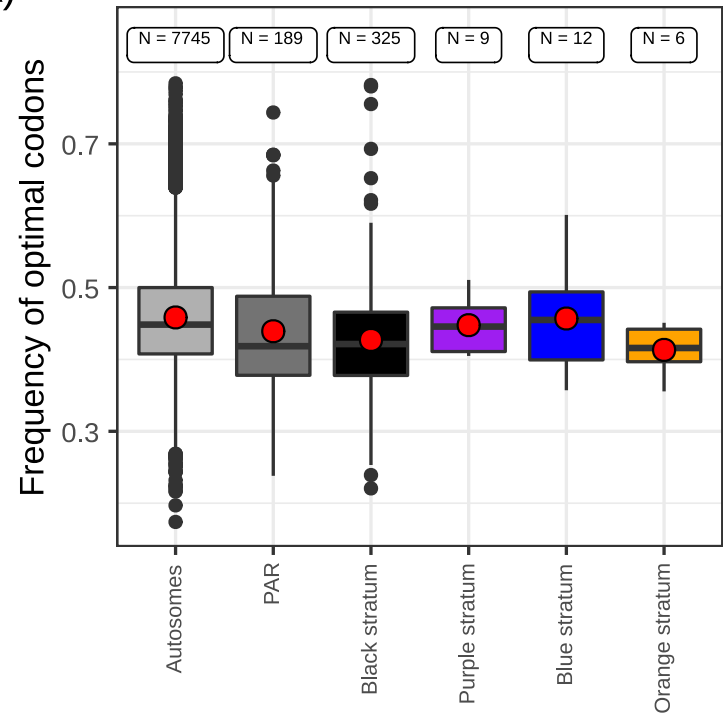
- Significant
- Non-significant

Genomic compartments

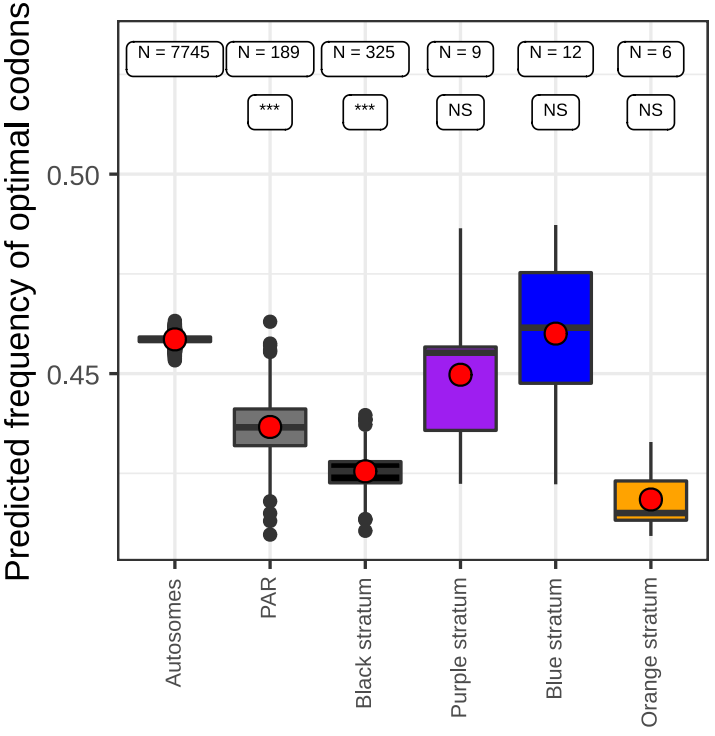
- Autosomes
- PAR
- Black stratum
- Purple stratum
- Blue stratum
- Orange stratum



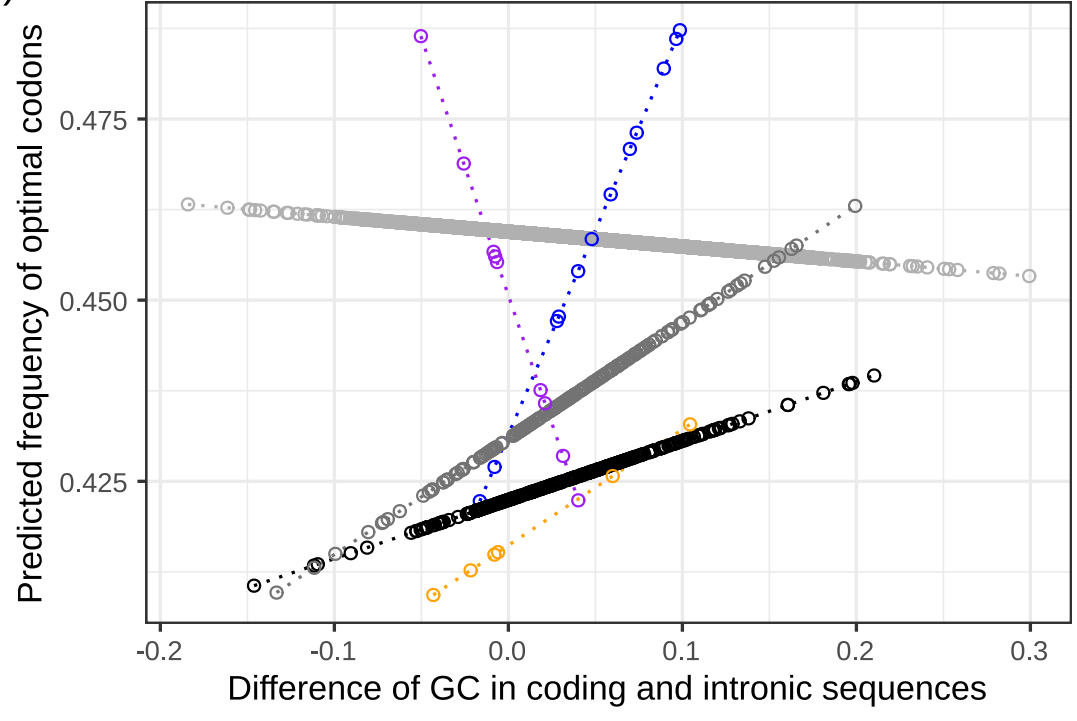
A)



B)



C)



Interaction significance

Non-significant

Genomic compartments

Autosomes

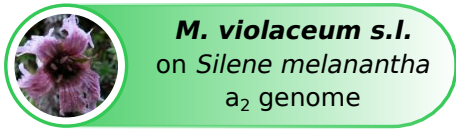
PAR

Black stratum

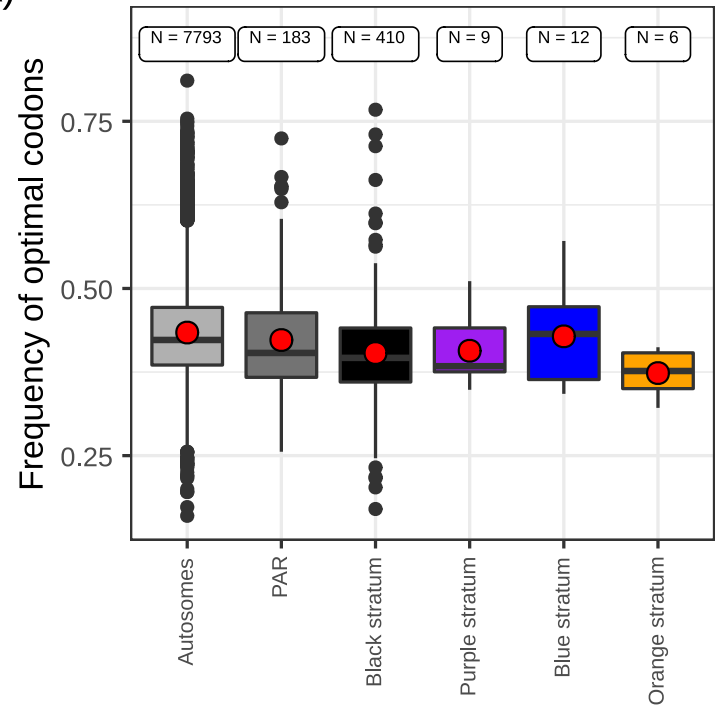
Purple stratum

Blue stratum

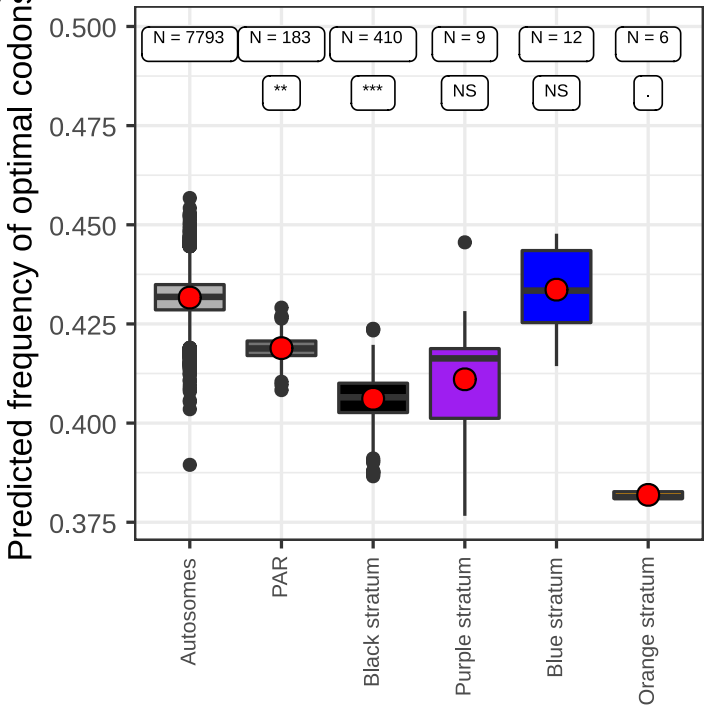
Orange stratum


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

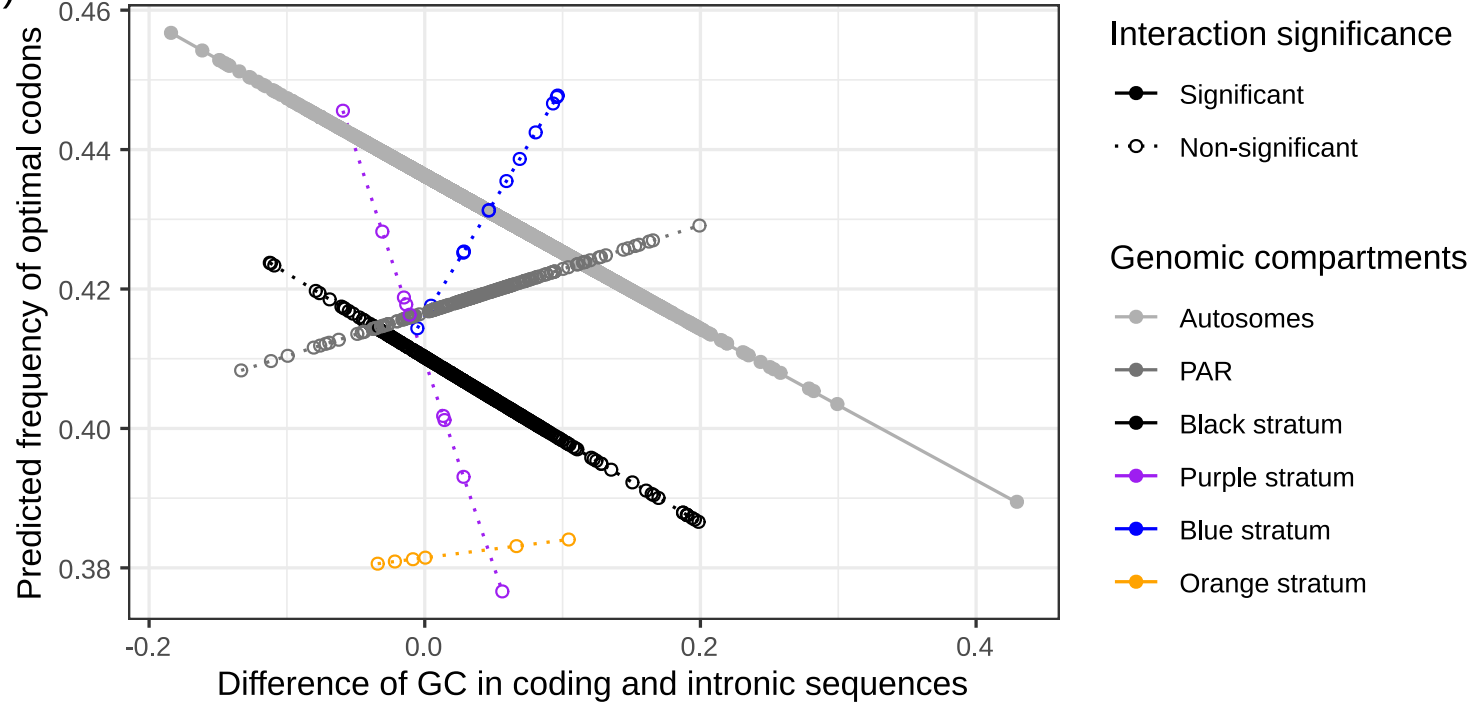
A)

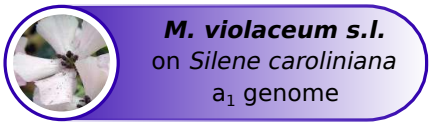

N = 7793
N = 183
N = 410
N = 9
N = 12
N = 6
Frequency of optimal codons
0.75
0.50
0.25
Autosomes
PAR
Black stratum
Purple stratum
Blue stratum
Orange stratum

B)


N = 7793
N = 183
N = 410
N = 9
N = 12
N = 6
**
***
NS
NS
.
Predicted frequency of optimal codons
0.500
0.475
0.450
0.425
0.400
0.375
Autosomes
PAR
Black stratum
Purple stratum
Blue stratum
Orange stratum

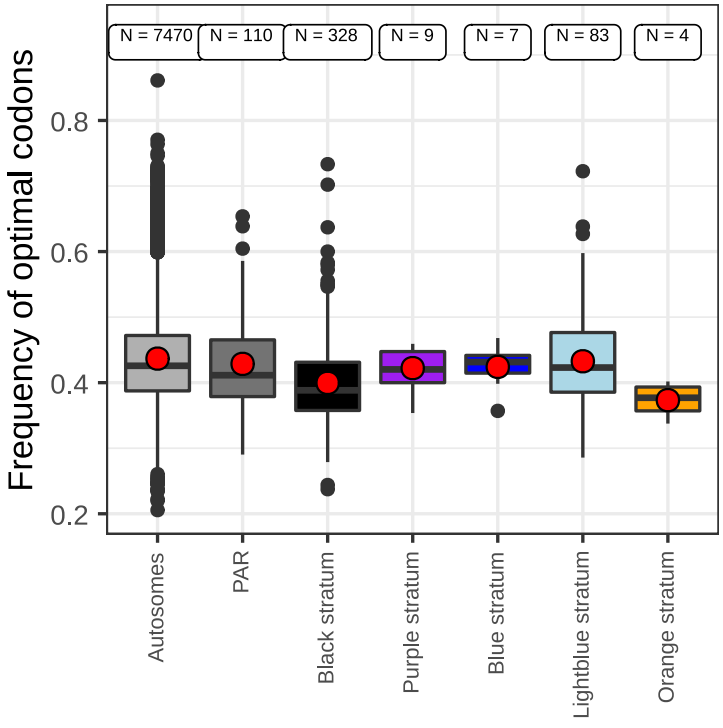
C)


Predicted frequency of optimal codons
0.46
0.44
0.42
0.40
0.38
-0.2
0.0
0.2
0.4
Difference of GC in coding and intronic sequences
Interaction significance
Significant
Non-significant
Genomic compartments
Autosomes
PAR
Black stratum
Purple stratum
Blue stratum
Orange stratum

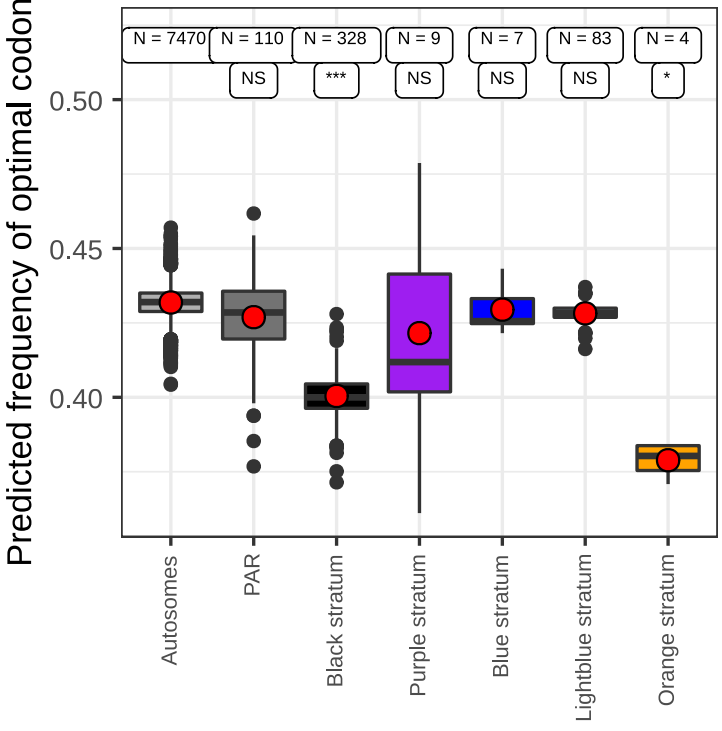


M. violaceum s.l.
on *Silene caroliniana*
a_1 genome

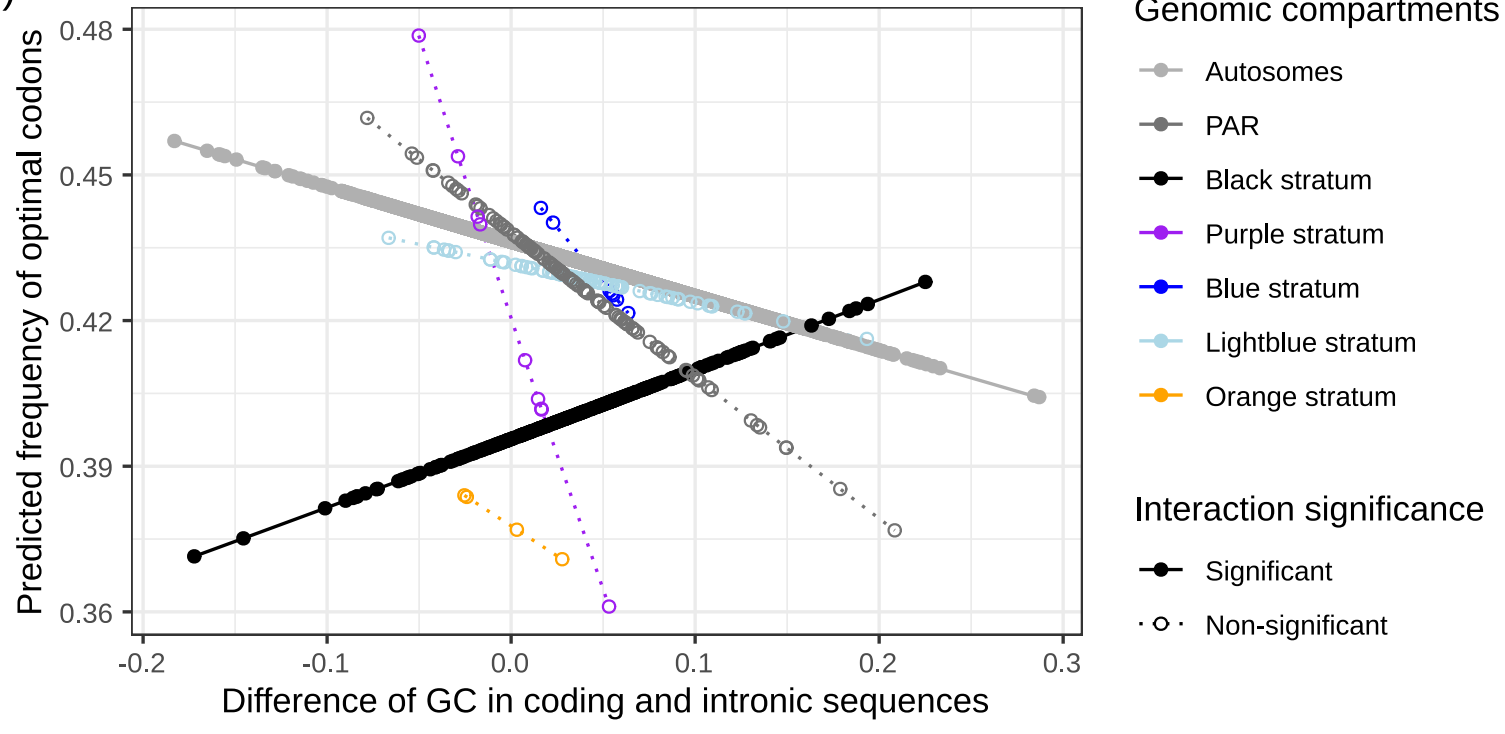
A)

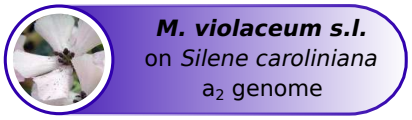


B)

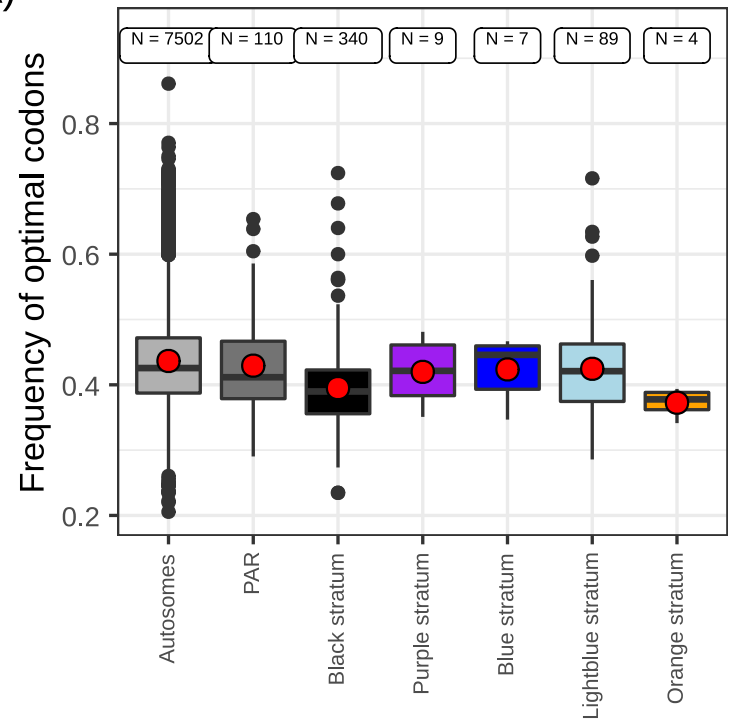


C)

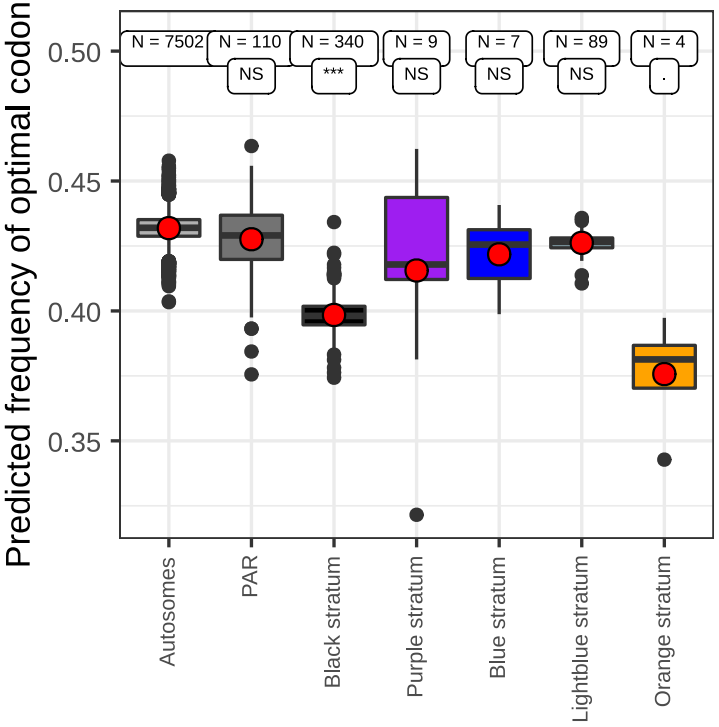



M. violaceum s.l.
on Silene caroliniana
a2 genome

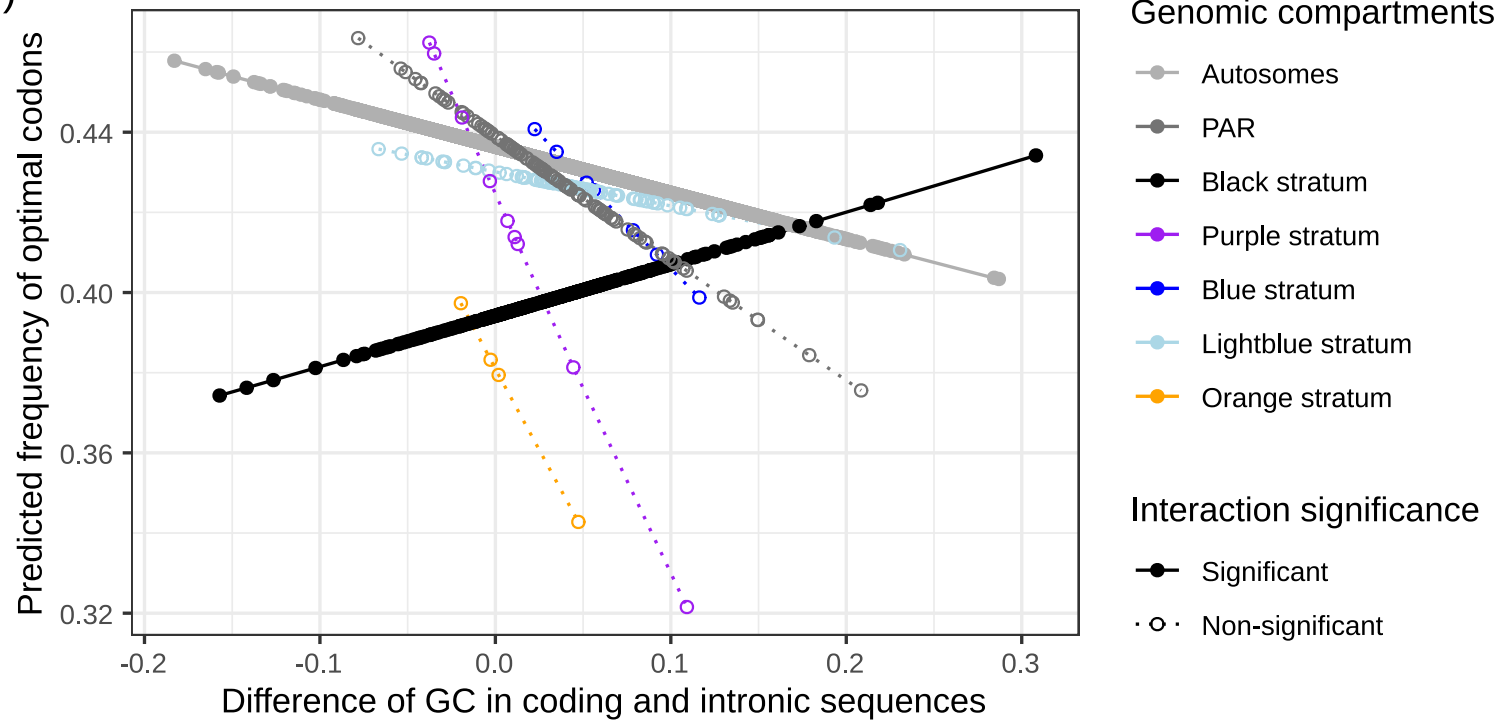
A)

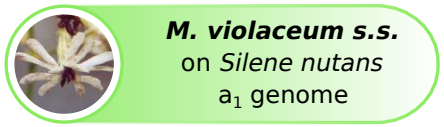

N = 7502
N = 110
N = 340
N = 9
N = 7
N = 89
N = 4
Frequency of optimal codons
0.8
0.6
0.4
0.2
Autosomes
PAR
Black stratum
Purple stratum
Blue stratum
Lightblue stratum
Orange stratum

B)

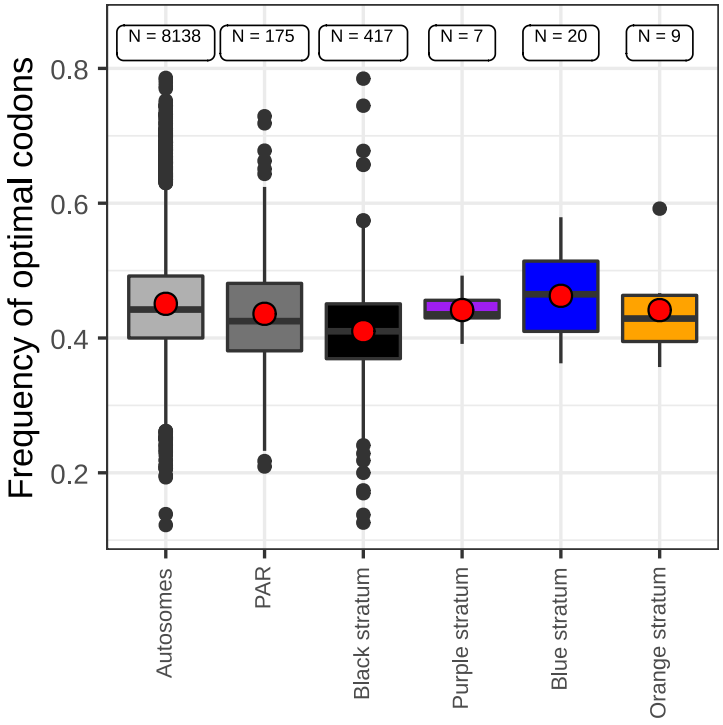

N = 7502
N = 110
N = 340
N = 9
N = 7
N = 89
N = 4
NS
***
NS
NS
NS
.
Predicted frequency of optimal codons
0.50
0.45
0.40
0.35
Autosomes
PAR
Black stratum
Purple stratum
Blue stratum
Lightblue stratum
Orange stratum

C)

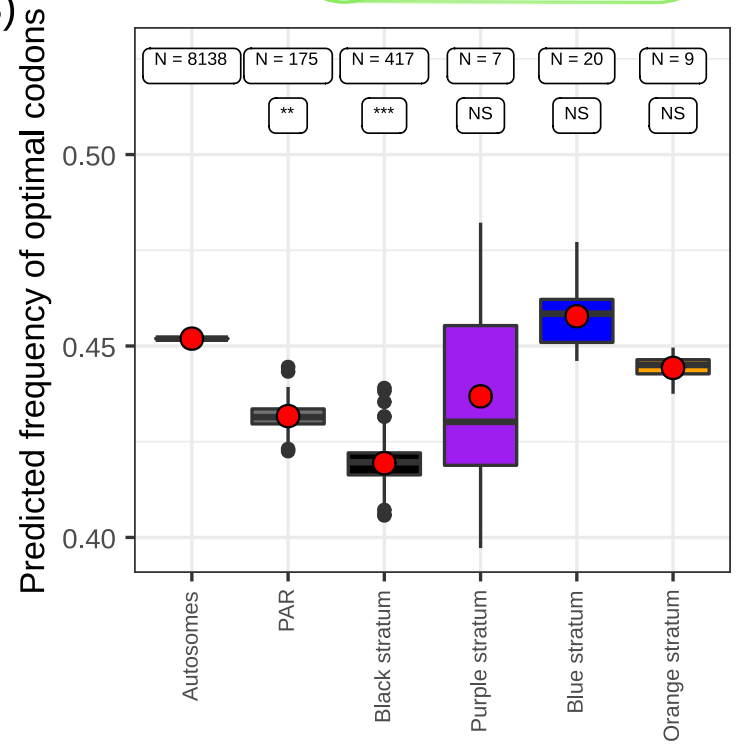

Predicted frequency of optimal codons
0.44
0.40
0.36
0.32
-0.2
-0.1
0.0
0.1
0.2
0.3
Difference of GC in coding and intronic sequences
Genomic compartments
Autosomes
PAR
Black stratum
Purple stratum
Blue stratum
Lightblue stratum
Orange stratum
Interaction significance
Significant
Non-significant


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

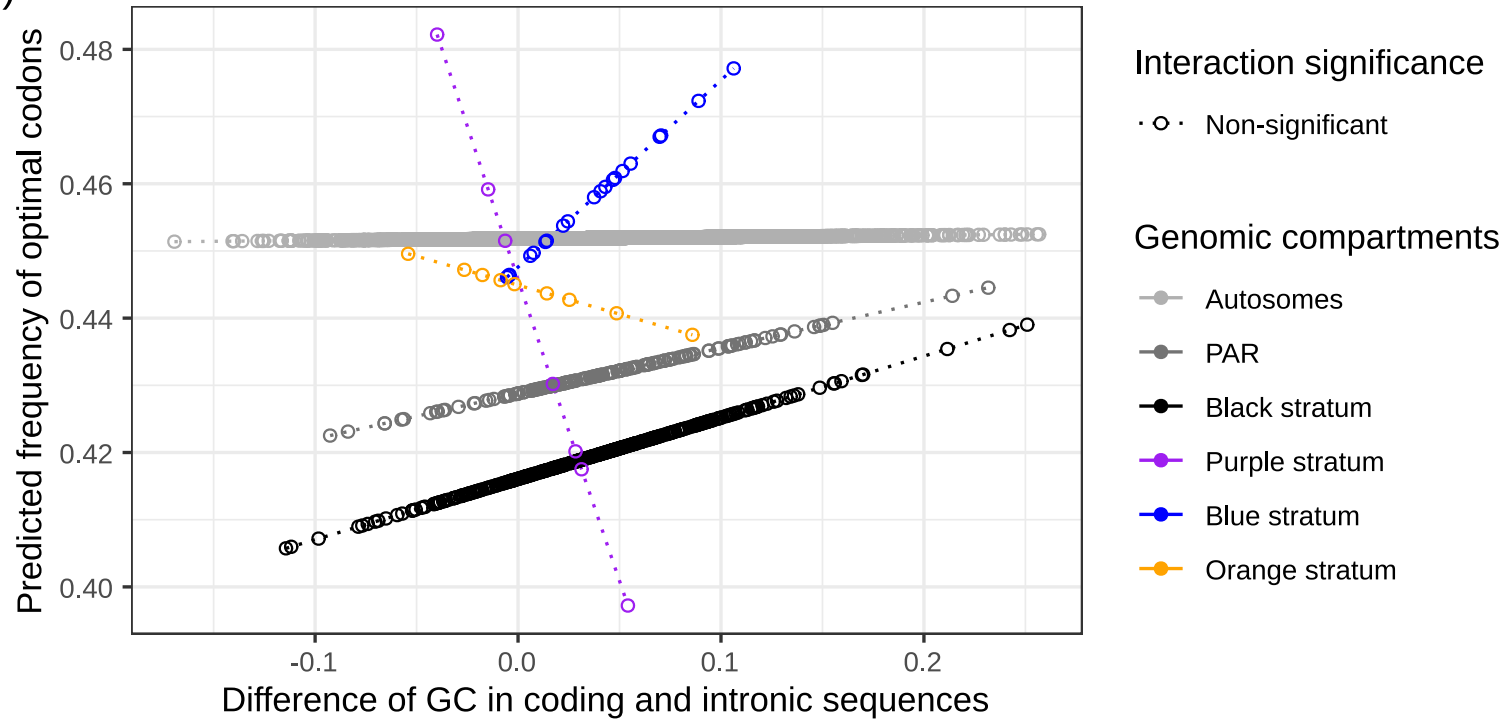
A)

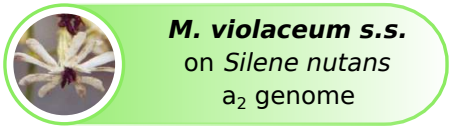

N = 8138
N = 175
N = 417
N = 7
N = 20
N = 9
Frequency of optimal codons
0.8
0.6
0.4
0.2
Autosomes
PAR
Black stratum
Purple stratum
Blue stratum
Orange stratum

B)

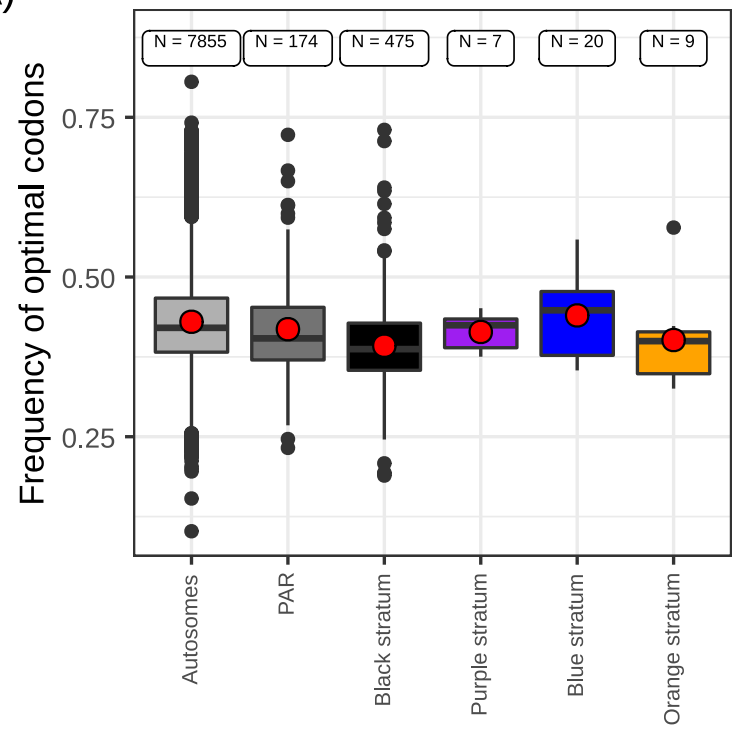

N = 8138
N = 175
N = 417
N = 7
N = 20
N = 9
**
***
NS
NS
NS
Predicted frequency of optimal codons
0.50
0.45
0.40
Autosomes
PAR
Black stratum
Purple stratum
Blue stratum
Orange stratum

C)

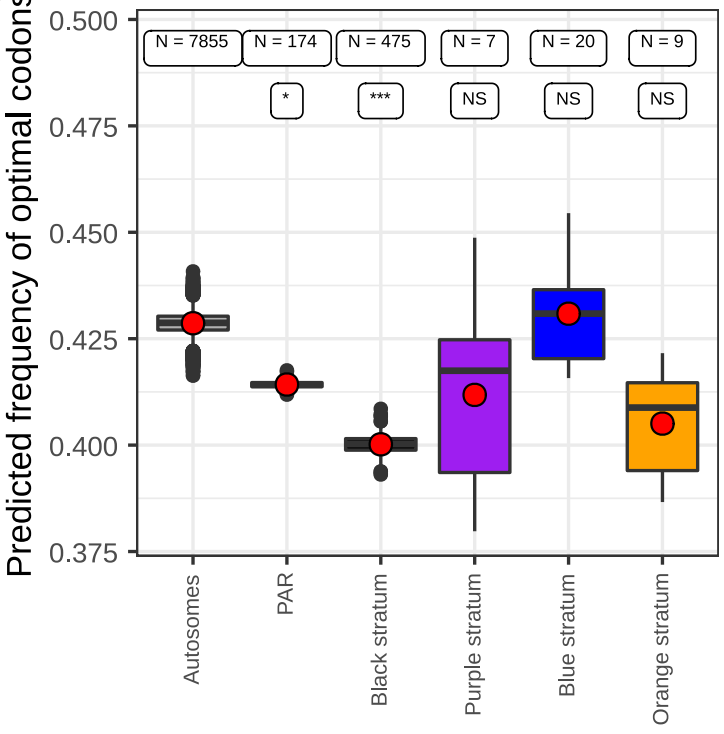

Predicted frequency of optimal codons
0.48
0.46
0.44
0.42
0.40
-0.1
0.0
0.1
0.2
Difference of GC in coding and intronic sequences
Interaction significance
Non-significant
Genomic compartments
Autosomes
PAR
Black stratum
Purple stratum
Blue stratum
Orange stratum


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

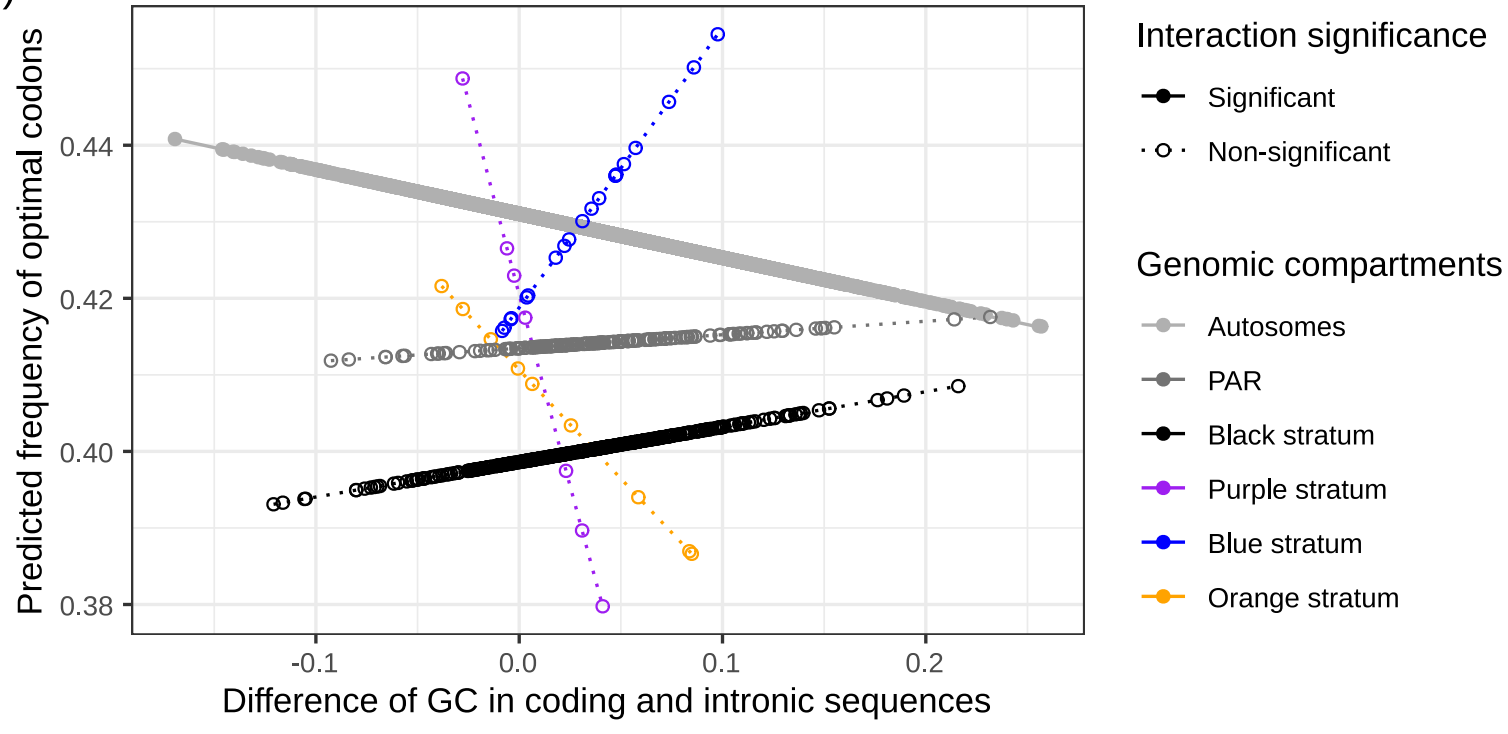
A)

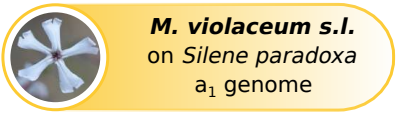

N = 7855
N = 174
N = 475
N = 7
N = 20
N = 9
Frequency of optimal codons
0.75
0.50
0.25
Autosomes
PAR
Black stratum
Purple stratum
Blue stratum
Orange stratum

B)

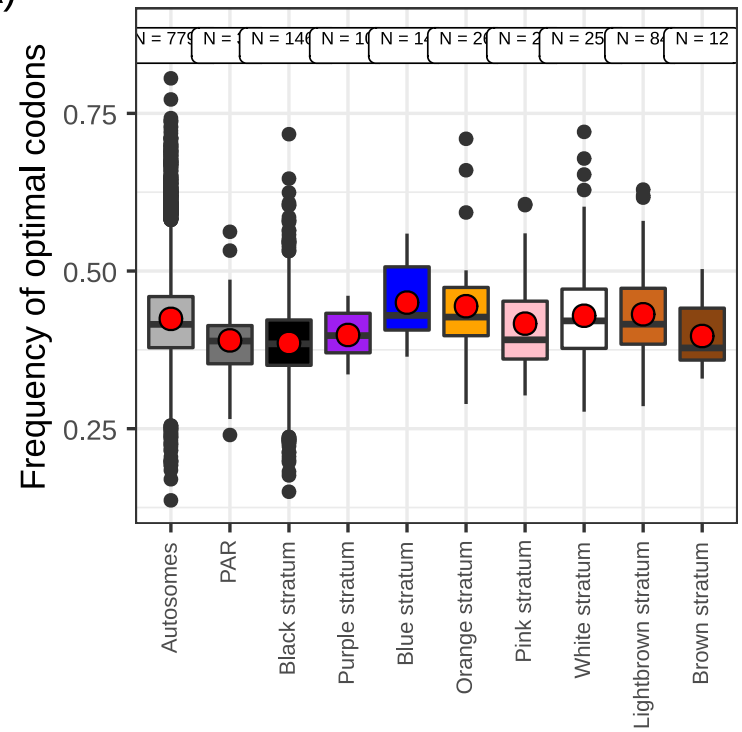

N = 7855
N = 174
N = 475
N = 7
N = 20
N = 9
*
***
NS
NS
NS
Predicted frequency of optimal codons
0.500
0.475
0.450
0.425
0.400
0.375
Autosomes
PAR
Black stratum
Purple stratum
Blue stratum
Orange stratum

C)

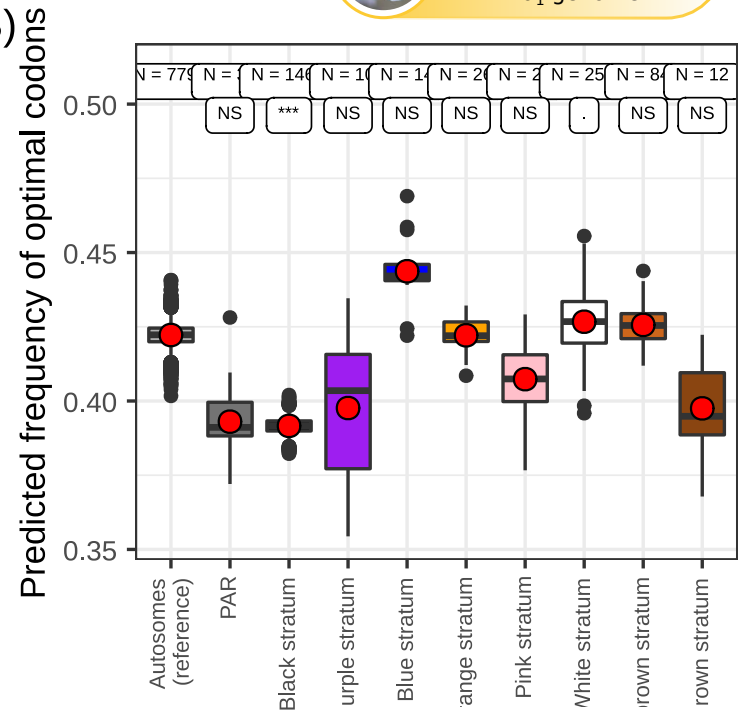

Predicted frequency of optimal codons
0.44
0.42
0.40
0.38
-0.1
0.0
0.1
0.2
Difference of GC in coding and intronic sequences
Interaction significance
Significant
Non-significant
Genomic compartments
Autosomes
PAR
Black stratum
Purple stratum
Blue stratum
Orange stratum



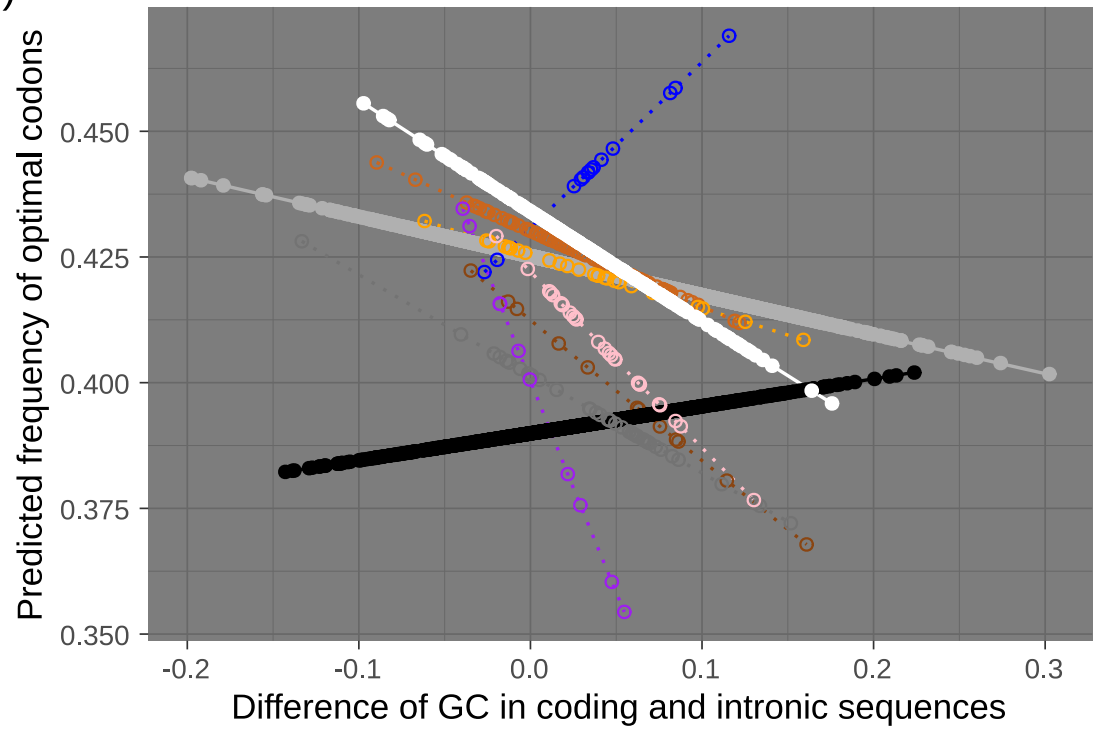
A)



B)



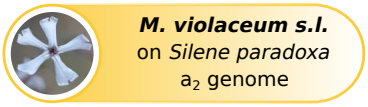
C)



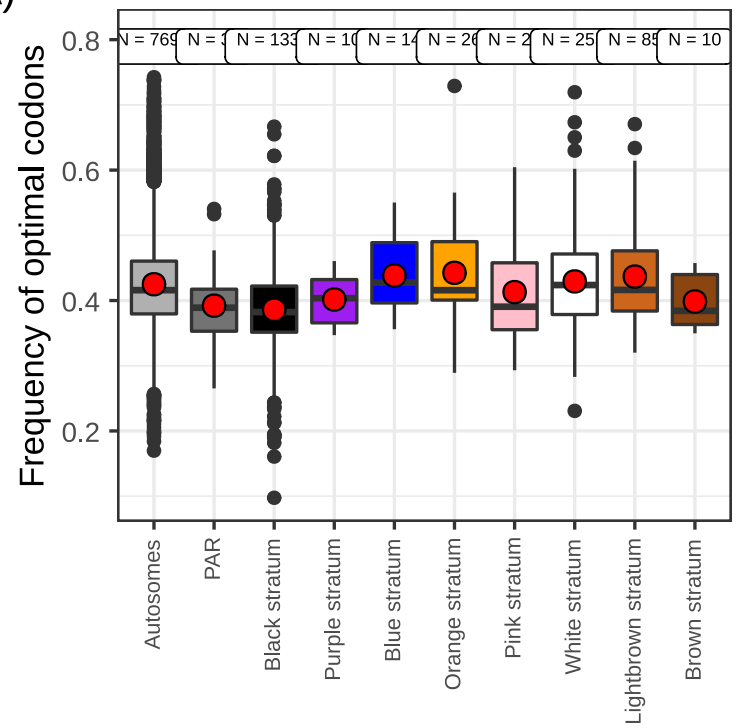
Autosomes (reference)
PAR
Black stratum
Purple stratum
Blue stratum
Orange stratum
Pink stratum
White stratum
Lightbrown stratum
Brown stratum

Interaction significance with autosomes

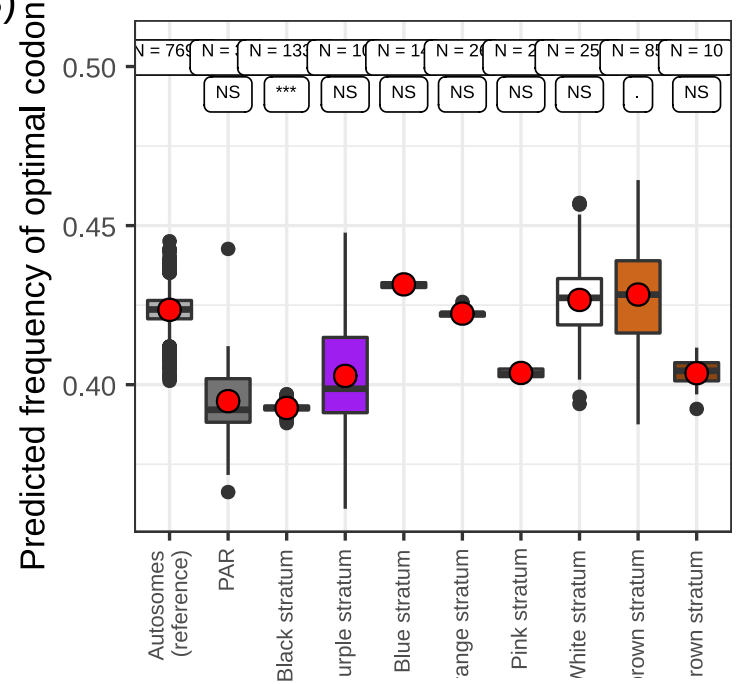
Significant
Non-significant



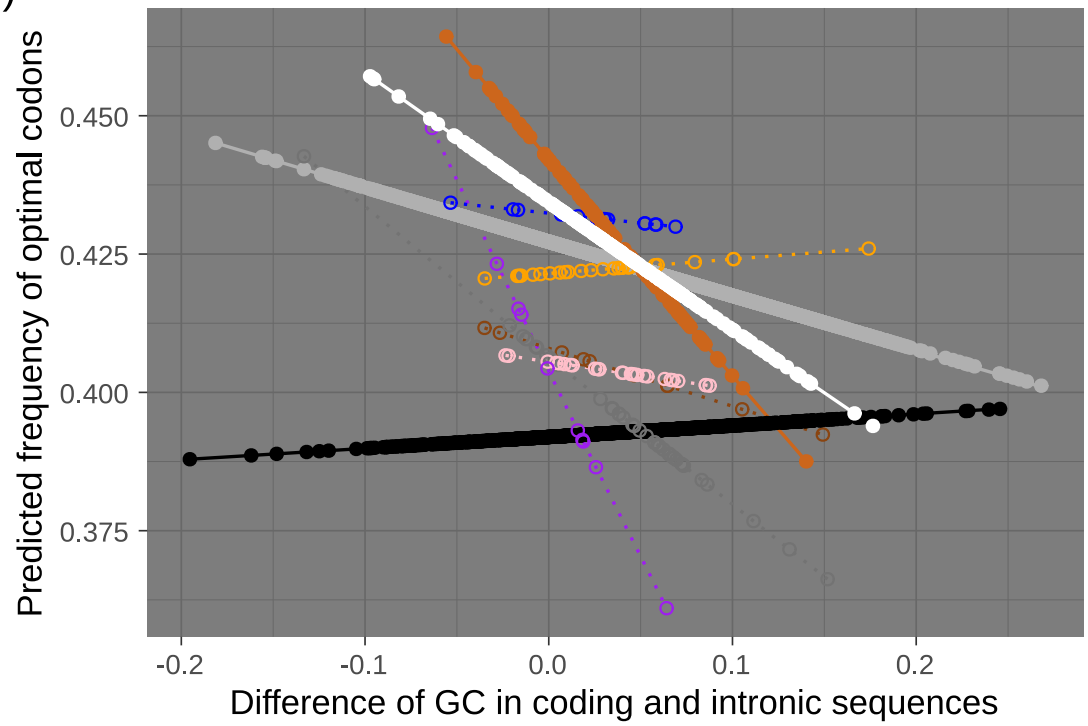
A)



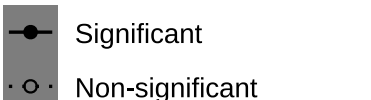
B)

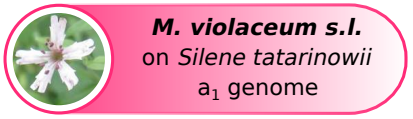


C)

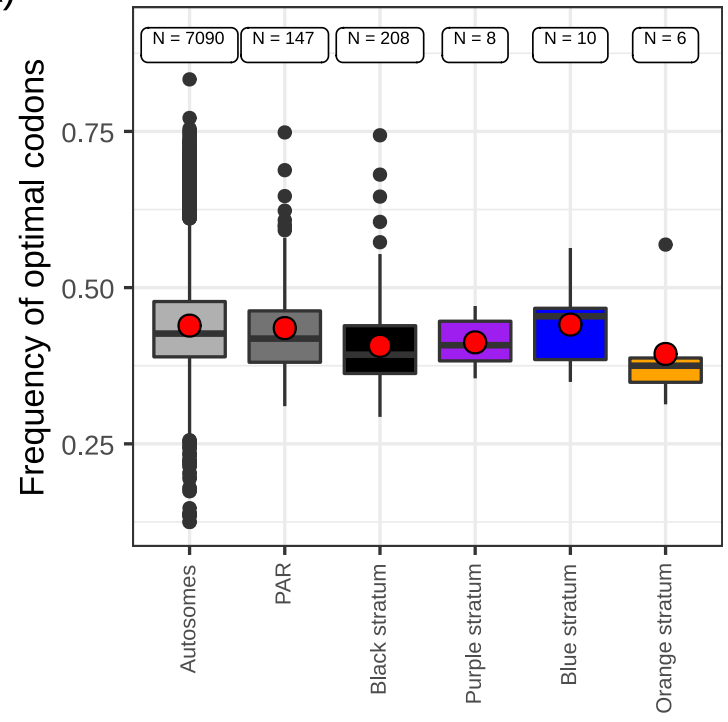


Interaction significance with autosomes

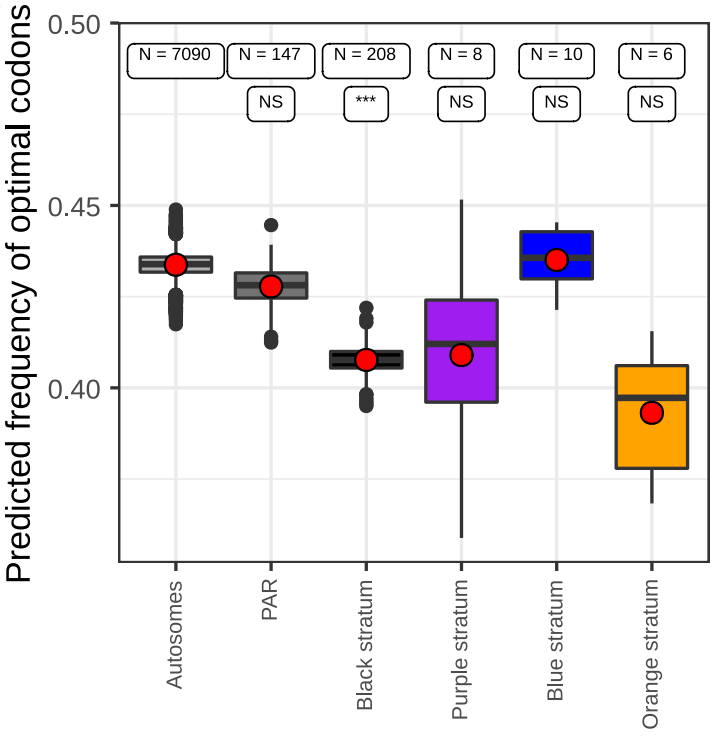



M. violaceum s.l.
on Silene tatarinowii
a1 genome

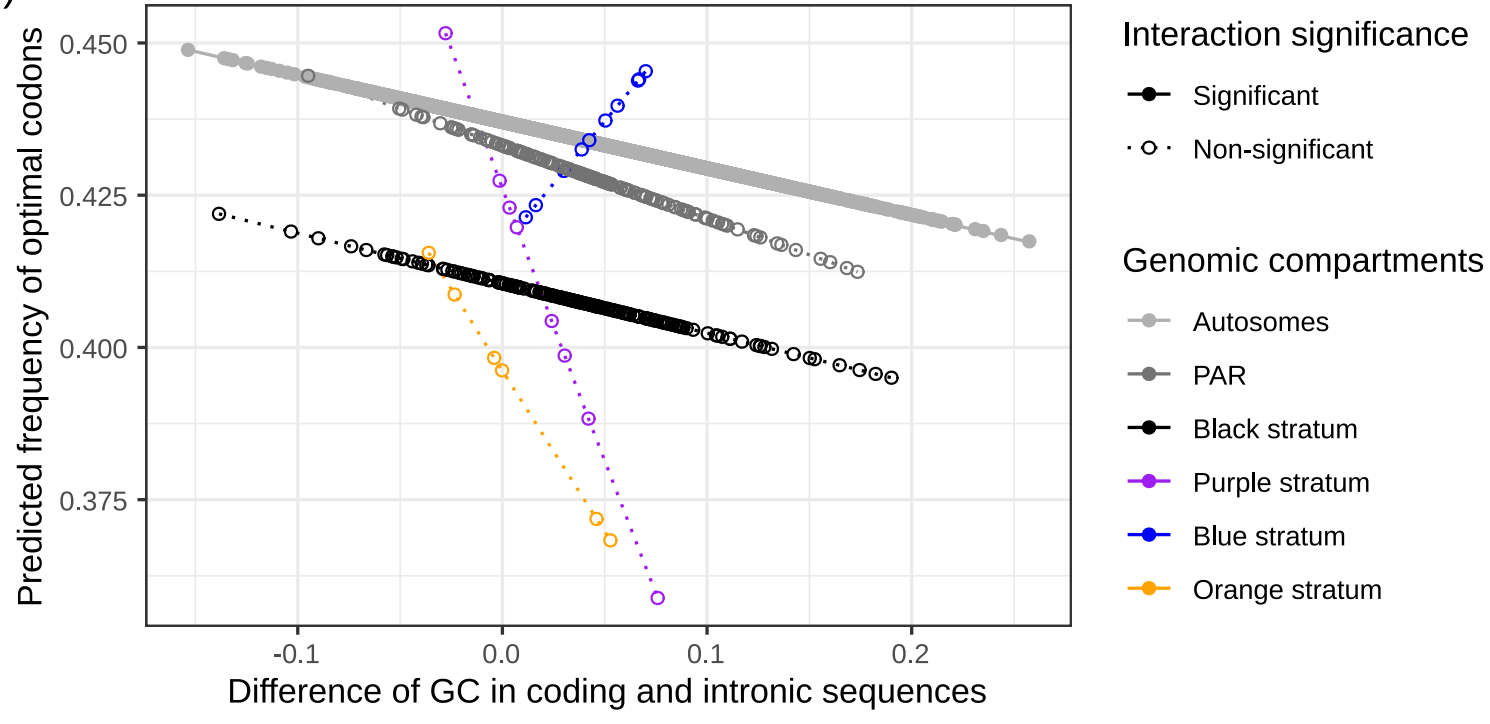
A)

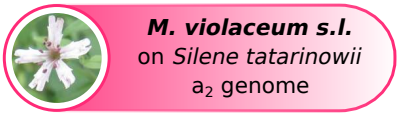

Frequency of optimal codons
0.75
0.50
0.25
N = 7090
N = 147
N = 208
N = 8
N = 10
N = 6
Autosomes
PAR
Black stratum
Purple stratum
Blue stratum
Orange stratum

B)

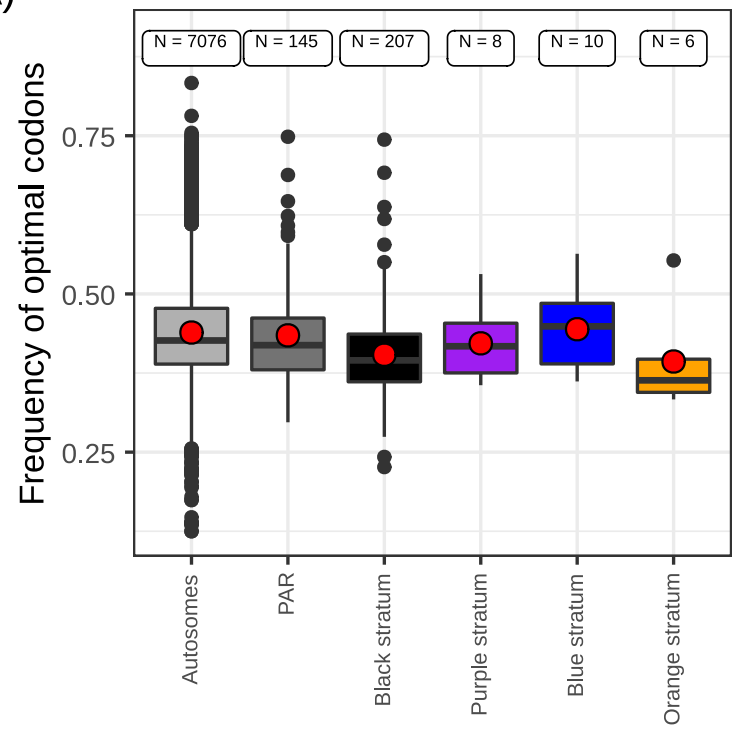

Predicted frequency of optimal codons
0.50
0.45
0.40
N = 7090
N = 147
N = 208
N = 8
N = 10
N = 6
NS
***
NS
NS
NS
Autosomes
PAR
Black stratum
Purple stratum
Blue stratum
Orange stratum

C)

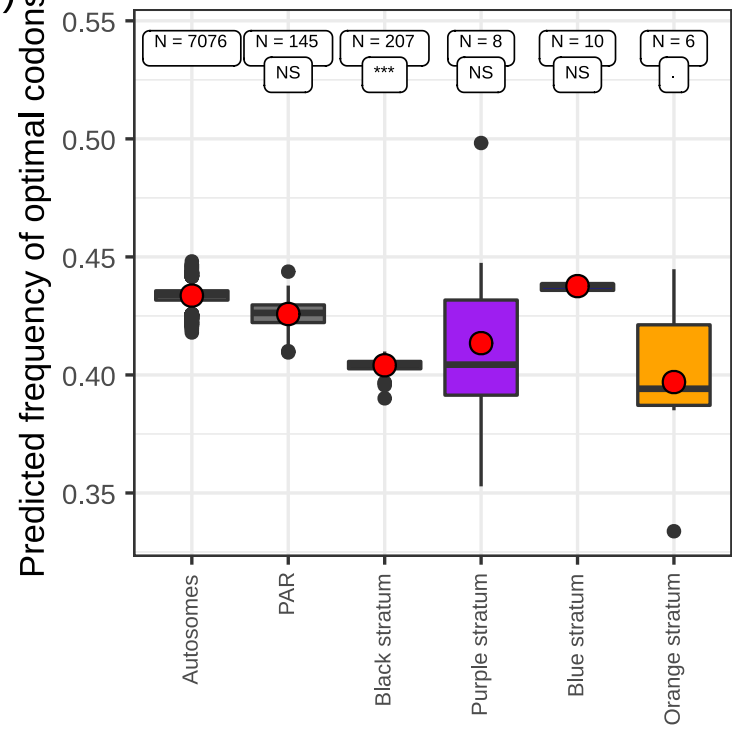

Predicted frequency of optimal codons
0.450
0.425
0.400
0.375
-0.1
0.0
0.1
0.2
Difference of GC in coding and intronic sequences
Interaction significance
Significant
Non-significant
Genomic compartments
Autosomes
PAR
Black stratum
Purple stratum
Blue stratum
Orange stratum


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

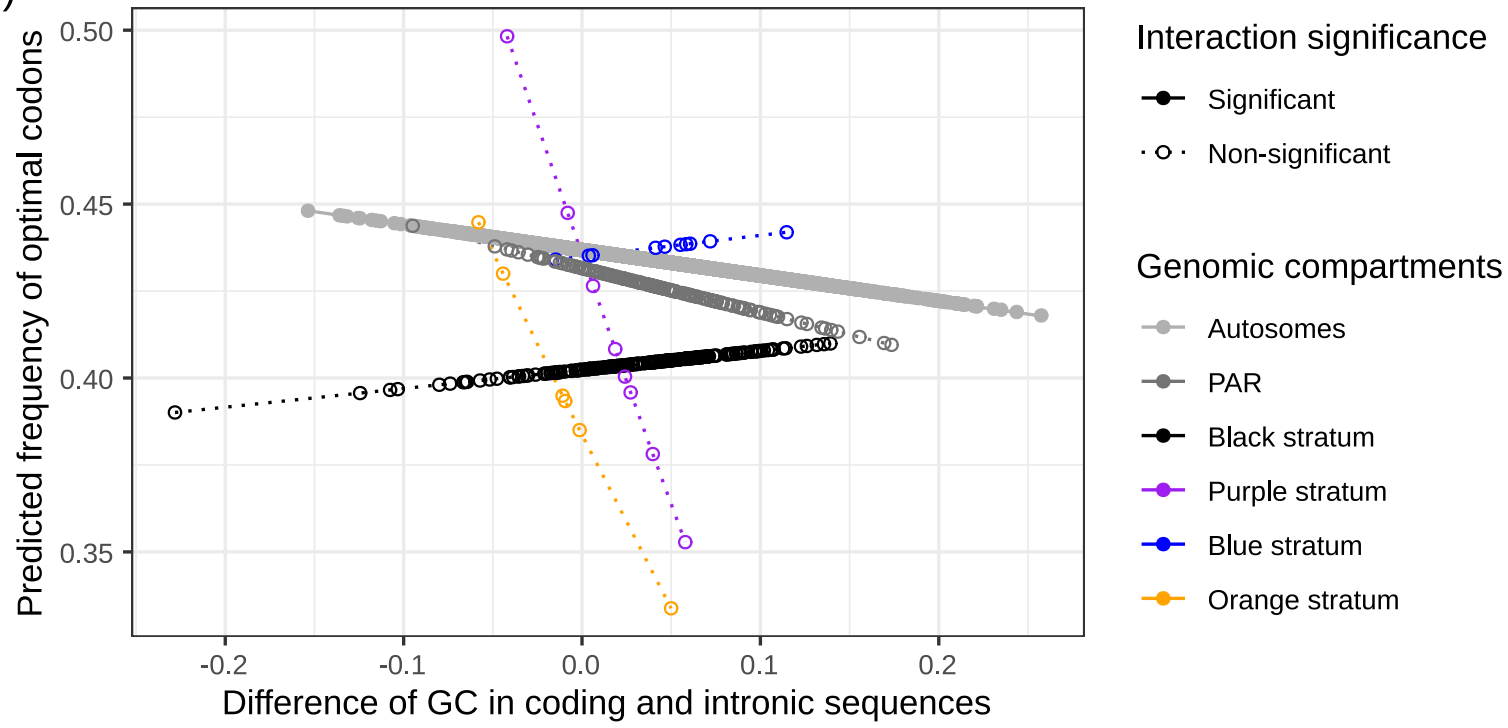
A)


N = 7076
N = 145
N = 207
N = 8
N = 10
N = 6
Frequency of optimal codons
0.75
0.50
0.25
Autosomes
PAR
Black stratum
Purple stratum
Blue stratum
Orange stratum

B)


N = 7076
N = 145
NS
N = 207
***
N = 8
NS
N = 10
NS
N = 6
.
Predicted frequency of optimal codons
0.55
0.50
0.45
0.40
0.35
Autosomes
PAR
Black stratum
Purple stratum
Blue stratum
Orange stratum

C)


Predicted frequency of optimal codons
0.50
0.45
0.40
0.35
-0.2
-0.1
0.0
0.1
0.2
Difference of GC in coding and intronic sequences
Interaction significance
Significant
Non-significant
Genomic compartments
Autosomes
PAR
Black stratum
Purple stratum
Blue stratum
Orange stratum