How heterozygote advantage can promote the evolution of disassortative mating and shape the underlying genetic architecture ?

Ludovic Maisonneuve[1,*], Mathieu Chouteau[2,3], Mathieu Joron[2] and Violaine Llaurens[1]

1. Institut de Systematique, Evolution, Biodiversité (ISYEB), Museum National d'Histoire Naturelle, CNRS, Sorbonne-Université, EPHE, Université des Antilles, 45 rue Buffon, 75005 Paris, France;

2. Centre d'Ecologie Fonctionnelle et Evolutive, UMR 5175 CNRS-Université de Montpellier, École Pratique des Hautes Études, Université Paul Valéry, 34293 Montpellier 5, France;

3. Laboratoire Ecologie, Evolution, Interactions Des Systèmes Amazoniens (LEEISA), USR 3456, Université De Guyane, IFREMER, CNRS Guyane, 275 route de Montabo, 97334 Cayenne, French Guiana;

∗ Corresponding author: Ludovic Maisonneuve; e-mail: ludovic.maisonneuve@polytechnique.edu.

Manuscript elements: Table 1, figure 1, figure 2, figure 3, figure 4, figure 5, figure S1, figure S2, figure S3, figure S4,figure S5 , figure S6, figure S7. Figure 2 and figure S5 are to print in color.

Keywords: Heterogamy, Supergene, Frequency dependent selection, genetic load, mate preference, *Heliconius numata*.

Manuscript type: Article.

Prepared using the suggested LaTeX template for *Am. Nat.*

Abstract

Mate preferences exert strong selection on the evolution of trait variations. Here we aimed at understanding the evolution of disassortative mating based on the emblematic example of polymorphism in mimetic color patterns in the defended butterfly *Heliconius numata*. Positive frequency-dependent selection is exerted on wing color pattern by predators learning the association between warning coloration and chemical defenses, resulting in selection promoting mimicry toward local communities of defended species. In this well-characterized adaptive landscape, chromosomal inversions in the supergene controlling wing pattern variations have been reported. These inversions are often associated with deleterious mutations, inducing a heterozygote advantage at the supergene, favoring the evolution of disassortative mating based on wing color pattern. To explore the conditions underlying the emergence of disassortative mating, we modeled both the color pattern locus and a mate preference locus. We confirm that a heterozygote advantage favor the evolution of disassortative mating and show that disassortative mating is more likely to emerge if at least one supergene allele is free from any genetic load. Comparisons of hypothetical genetic architecture underlying mate choice behaviors show that rejection alleles linked to the supergene can be under positive selection and enable the emergence of disassortative mating behaviour.

Introduction

Mate preferences often play an important role in shaping traits diversity in natural populations. Disassortative mating, *i.e.* preferential crosses between individuals displaying a different phenotype, is known to promote the persistence of polymorphism in traits targeted by sexual selection. Obligate disassortative mating for sexes or mating types leads to the persistence of intermediate frequencies of sexes or mating types (Wright , 1939), and promotes polymorphism, with in some extreme cases, thousands of mating types being maintained, as in some Basidiomycete fungi for instance (Casselton , 2002). Few examples of disassortative mating have been documented for other traits, like the body chirality in *Amphridromus inversus* snails, where a greater fecundity was reported in inter-chiral matings, therefore promoting polymorphism within population (Schilthuizen et al. , 2007). Disassortative mating is frequently reported in traits where polymorphism is maintained because of natural selection : in the scale eating predator fish *Perissodus microlepis*, a dimorphism on the mouth-opening direction ('lefty' versus 'righty') is maintained within populations by negative frequency-dependent selection (Michio , 1993), due to prey behavior monitoring the most attacked side. A disassortative mating behavior based of the mouth-opening direction is also observed in this species (Hori , 1993). Disassortative mating based on odors has been reported in mice (Penn et al. , 1999) and humans (Wedekind et al. , 1995): odor profiles are indeed tightly linked to genotypes at the MHC loci controlling for variations in the immune response, known to be under strong balancing selection (Piertney et al. , 2006). The balancing selection in MHC partly stems from heterozygous advantage, whereby heterozygous genotypes might be able to recognize a large range of pathogens. Such heterozygote advantage may thus promotes the evolution of disassortative mating (Tregenza et al. , 2000). Extreme examples of heterozygotes advantage are observed for loci for which homozygotes have reduced survival. In the white-throated sparrow *Zonotrichia albicollis*, strong disassortative mating has been reported regarding the color of the head stripe. This plumage polymorphism is controlled by a supergene, with the allele controlling the white phenotype displaying a gene order inversion

(Tuttle et al. , 2016), thought to be homozygote lethal (Campagna , 2016). Deleterious mutations are likely to be associated with inversions (Kirkpatrick , 2010), because of the possibility of an initial capture of deleterious elements and the lack of DNA repair thereafter due to the suppression of genetic recombination. Inversion polymorphisms are thus prone to heterozygote advantage (Llaurens et al. , 2017), and could therefore favor the evolution of disassortative mating behavior since this behavior prevent the production of homozygote progenies with decreased fitness.

Nevertheless while the fitness advantage of disassortative mating is easily grasped, the genetic bases of disassortative mating preferences remains largely unknown. One exception is the self-incompatibility system in Brassicaceae where the *S*-locus controls for a specific rejection of incompatible pollen (Hiscock et al. , 2003). *S*-haplotypes contains tightly linked co-evolved *SCR* and *SRK* alleles, encoding for a protein of the pollen coat and a receptor kinase located in the pistil membrane respectively, leading to specific rejection due to specific receptor-ligand interactions. The existence of self-rejection mechanisms has been suggested to explain the disassortative mating behavior linked to odor in humans. Body odors are strongly influenced by genotypes at the immune genes *HLA* and rejection of potential partners has been shown to be linked to level of *HLA* similarity, rather than specific rejection of a given *HLA* genotype (Wedekind et al. , 1997). In the white-throated sparrow, disassortative mating stem from specific preferences for color plumage that differ between males and females, whereby tan-striped males are preferred by all females whereas white-striped females are preferred by all males (Houtman et al. , 1994). The various mechanisms involved in disassortative mating can be summarized under two main hypothetical genetic architecture : (1) Self-referencing, (i.e. when individual used its own signal to choose its mate) and (2) Preferences or rejection of a given phenotype (Kopp et al. , 2018). The locus controlling preference could be either be the same or different from the locus controlling cue variations, and in the latter case, the level of linkage disequilibrium between the two loci could have a strong impact on the evolution of disassortative mating. In models investigating the evolution of assortative mating on locally-adapted traits, theoretical simulations have demonstrated that assortative mating is favored when the preference and the cue locus are linked (Kopp

et al. , 2018).

Here we explore the evolutionary modalities leading to disassortative mating. We focus on the specific case of the butterfly species *Heliconius numata*, where high polymorphism in wing color pattern is maintained within population (Joron et al. , 1999) and strong disassortative mating based on these wing color patterns (Chouteau et al. , 2017). *H. numata* individuals are chemically-defended (Arias et al. , 2006), and their wing color patterns act as warning signals against predators. Predators indeed learn to associate warning patterns to chemical defense. This learning behavior generates a positive frequency-dependent selection (pFDS) on butterfly wing pattern (Chouteu et al. , 2016), because displaying a widely shared color pattern decreases the risk of encountering a naive predator (Sheratt , 2006). Positive FDS promote convergent evolution of wing color pattern among defended species (*i. e.* Müllerian mimicry, (Müller et al. , 1879)), forming so-called mimicry rings composed of individuals from different species sharing a common warning signal. Positive FDS usually leads to the fixation of a single color pattern within population (Merrill et al. , 2015). However, polymorphism of several wing color patterns is maintained within populations of *H. numata*, each pattern being a precise mimic of a different species of the distantly-related genus *Melinaea* living in sympatry. Polymorphism within *H. numata* populations may be favored by the fine-scale spatial heterogeneity of the distinct mimicry rings in which it take part of (Joron et al. , 1999). Disassortative mating based on wing color pattern, where *H. numata* females reject males displaying the same color pattern, might further promote this local polymorphism (Chouteau et al. , 2017). Nevertheless, why such disassortative mating has evolved is unclear, notably because of it tends to counteract the strong selection exerted against rare color pattern exerted by predators (Chouteu et al. , 2016). Variation in wing color patterns of *H. numata* is controlled by a single genomic region, called the supergene *P* (Joron et al. , 2006), which display chromosomal inversions (Joron et al. , 2011). Derived alleles carried by inversions are generally dominant over alleles carried by the ancestral gene order, resulting in strict dominance among alleles in sympatry (Arias et al. , 2014). These inversions drastically reduce recombination among mimetic alleles, and are likely to be linked with deleterious mu-

tations. Derived alleles may thus carry a stronger genetic load as compared to ancestral ones, because inversions might have initially captured deleterious elements. Such genetic load associated with mimetic alleles might generate heterozygote advantage at the color pattern locus and in turn promote the evolution of disassortative mate preferences based on color pattern cues. A lack oh dominant homozygotes individuals has indeed been reported in natural populations of *H. numata* has been reported, suggesting that heterozygote advantage might involved (Chouteau et al. , 2017). Wing colour patterns polymorphism and associated disassortative mating in *H. numata* is thus a relevant case-study, because the direction of natural selection can be inferred from the composition of local communities of defended butterfly species (Joron et al. , 1999), and the mating preferences are well-characterized (Chouteau et al. , 2017). By focusing on this well-documented example, we used a theoretical approach to provide general predictions on the evolution of disassortative mating in traits controlled by inversion polymorphism, and on expected genetic architecture underlying this behavior.

We specifically investigate whether genetic load associated with inversions may favor the evolution of mate preference and promote local polymorphism despite local directional selection generated by predators. We then explored two putative genetic architecture of mate preferences, (1) based on self referencing and (2) based on preference/trait matching rules, and tested their respective impact on the evolution of disassortative mating behavior.

Methods

Based on the model of (Joron et al. , 2005) extended to diploid populations (Llaurens et al. , 2013), we described a two populations model with a locus P controlling mimetic color pattern under positive FDS and spatial variations in mimetic communities. We explicitly modeled the genetic architecture controlling mate preference by a locus M assuming either (1) preference based on the phenotype of the chooser or (2) preference for a given phenotype. We also assumed different levels of genetic load associated with the mimetic alleles.

We tracked the evolution of allele frequencies at the locus P controlling variations in wing color pattern and at the locus locus M controlling mate preference. A recombination rate ρ between these two locus was assumed.

Mimetic color patterns

At locus P, three alleles were assumed to segregate, namely alleles a, b and c, encoding for phenotypes [A], [B] and [C] respectively. We assumed strict dominance among the three alleles with $a > b > c$ in agreement with the strict dominance observed among supergene P alleles within natural populations of *H. numata* (Arias et al. , 2014). The three color pattern phenotypes were assumed to be perceived as strictly different by both mating partners and predators. The resemblance Res[i][j] between pairs of individuals exhibiting phenotype [i] and [j] respectively was thus set to 1 for identical phenotypes and to 0 for dissimilar one. The resemblance matrix among the three phenotypes is :

$$Res = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Spatial variation in mimetic communities

We assumed two populations of an unpalatable species involved in Müllerian mimicry with other chemically-defended species. The environment differed in communities of local species involved in mimicry (*i.e.* mimicry rings). We considered two patches occupied by different mimetic communities: population 1 was located in a patch where the local community (*i.e.* other aposematic species, not including *H. numata*) mostly displayed phenotype [A], and population 2 in a patch where mimetic communities mostly displayed phenotype [B]. This spatial heterogeneity was represented by the parameter $\sigma \in [0, 1]$ simulating the relative proportion of phenotypes [A] and [B] in mimicry rings of patch 1 and 2 respectively, so that the higher is σ, the more the two

communities differed. This spatial heterogeneity σ plays a central role on the predation suffered by the different phenotypes in the two patches (see Predation section below). The focal mimetic species was polymorphic for those two phenotypes, corresponding to the locally advantageous phenotypes [A] or [B].

Positive frequency-dependent predation

Every individual of the focal (polymorphic) species suffered a predation risk modulated by its resemblance to the local mimetic community of butterflies. We assumed a symmetrical condition where the mortality coefficient was $d(1-\sigma)$ for phenotypes matching the local mimicry ring (*i.e.* diminishing predation exerted on genotypes displaying phenotype [A] in population 1 and genotypes displaying [B] in population 2) and $d(1+\sigma)$ otherwise (*i.e.* increasing predation exerted on genotypes displaying phenotype [B] or [C] in population 1 and on genotypes displaying phenotype [A] or [C] in population 2), where d represented the baseline predation risk and σ the spatial heterogeneity of mimicry communities in patch 1 and 2.

Predation exerted on a given phenotype depended on its match to the local mimicry environment, but also on its own abundance in the patch. Number-dependent predator avoidance in the focal species was assumed to depend on its unpalatability coefficient (λ) and the density of each phenotype, so that protection gained by resemblance among phenotypes was greater for higher values of the unpalatability coefficient λ. This results in the following change in number of each genotype i in population pop due to predation.

$$i = (p1, p2, m1, m2) \tag{1}$$

$$\Delta P^t_{i,pop} = -\frac{d}{1+\lambda(\sum_j Res_{[i],[j]})N^t_{j,pop}}[(1+\sigma)(1-Res_{[i],[pop]}) + (1-\sigma)Res_{[i],[pop]}]N^t_{i,pop} \tag{2}$$

with $N_{i,pop}$ representing the total number of individuals with genotype i in population k, $Res_{[i],[pop]}$ representing the resemblance of the phenotype expressed by genotype i to the local

mimetic community. This resemblance is equal, in population 1, to the resemblance of the genotype i to the wild-type homozygote aa (displaying phenotype [A]), and to homozygote bb in population 2 (displaying phenotype [B]). Individuals displaying phenotype [C] were non-mimetic in both populations, and therefore suffer from high predation risk.

Migration

The change in the number of individuals with genotype i in population pop due to migration between populations pop and pop' was given by:

$$\Delta M^{t}_{i,pop} = mig(N_{i,pop'} - N_{i,pop}) \tag{3}$$

with mig is the migration coefficient $mig \in [0, 1]$).

Mate preferences

Two hypothetical mate preference mechanisms were investigated. Under the self-referencing hypothesis (hyp 1), two alleles were assumed at loci M, coding for (i) random mating (r) and (ii) preferential mating behaviors (either assortative sim or disassortative dis) respectively (see fig. S1 for more details). Hyp 1 allows to test whether preferential mating (either assortative disassortative) can evolve in a population where mating occur at random initially. Preferential mating alleles were assumed as dominant over the random mating allele. Note that under (hyp. 1), mating behavior is based on a self-referencing, and thus crucially depends on the color pattern of the individual expressing the preference.

An alternative model of mechanisms of mate preference was investigated, assuming a specific recognition of color patterns, acting as mating cue (hyp.2). Under hyp.2, four alleles segregated at locus M: allele Mr, coding for an absence of color pattern recognition (leading to random mating behavior), and Ma, Mb and Mc coding for recognition of color pattern phenotype [A], [B] and [C]. The 'no preference' allele Mr was recessive over all the preference alleles Ma, Mb

and *Mc*, and preference alleles were co-dominant, so that that heterozygotes can recognize two different alleles. Then, the recognition enabled by preference alleles *Ma*, *Mb* and *Mc* triggered attraction (hyp.2.a) or repulsion (hyp.2.b) toward recognized color pattern, leading to assortative or disassortative mating behavior depending on the genotype at locus *M* (see figure S2 and S3 for more details).

We expect the evolution of disassortative mating to be favored when preference alleles (*Ma*, *Mb* and *Mc*) generate rejection behavior (hyp.2.b) rather than attraction (hyp.2.a). Disassortative mating of females indeed implies the avoidance of males displaying their color pattern. Such behavior can simply emerge from an haplotype combining allele *a* at color pattern locus *P* and M_a at preference locus *M* assuming the genetic architecture triggering rejection (hyp 2.b). Assuming a genetic architecture generating attraction (hyp 2.a) however, disassortative mating only emerge when females displaying the color pattern phenotype [A] (*i.e.* with genotypes *aa*, *ab* or *ac*) carry the heterozygous genotype M_bM_c at the preference locus *M*, preventing a complete fixation of this behavior.

To characterize female mating preferences generated by the different genotypes at locus *M*, we distinguished two main behaviors emerging under hyp.2 (fig. S2 and S3 for attraction (hyp.2.a) and rejection (hyp.2.b) hypotheses respectively):

- Self-acceptance : females mate with males displaying their own color pattern phenotype.
- Self-avoidance : females do not mate with males displaying their own color pattren phenotype.

Reproduction

We also assumed a balanced sex-ratio, a carrying capacity *K* and a growth rate *r* all equal in both populations, and $N^t_{tot,pop}$ the total density of individuals in population *pop* at time *t*. Sexual reproduction was computed explicitly. Assuming Mendelian segregation and a recombination at rate ρ between both locus, the change in the number of individuals with genotype *i* in population

pop due to reproduction was then described as follows:

The frequency of genotype i in population pop (defined as $f^t_{i,pop}$) was first computed. The change in frequency $F^t_{i,pop}$ then considered the frequencies of each genotype in the population, mendelian segregation and the mate preferences computed in equation 1.4. Consistent with mate choices observed in *H.numata* (Chouteau et al. , 2017), we assumed that females were the choosy sex and that males courted all female whatever her morph.

The preference matrix $Pref$ was initially set as $Pref_{i,[j]} = 1$ when female i accept male j as mating partner and $Pref_{i,[j]} = 0$ otherwise.

We define the fertility of the individual i as below

$$f_i = Pref_{i,A}P_A + Pref_{i,B}P_B + Pref_{i,C}P_C \tag{4}$$

Where P_i refer to the proportion of the morph i in the population.

Let $cost$ be the parameters representing the cost of the choosiness, as described in (Kirkpatrick et al. , 2004), where this cost is linked to mate availability: when $cost = 0$, females have access to a large number of potential mates, so that their fertility is not limited when they become choosy ("Animal" model), whereas when when $cost = 1$, females have access to a limited number of potential mates, so that their fertility tend to decrease when they become choosy ("Plant" model). This cost of choosiness is known to limit the evolution of mating preferences (Otto et al. , 2008).

$$F^{t+1}_{i,pop} = \sum_{j,k} coef(j,k,æ)\frac{1 - cost + cost f_j}{f_j} Pref_{j[k]} \frac{f^t_{j,pop}}{2}\frac{f^t_{k,pop}}{2} \tag{5}$$

Where $coef$ controls the mendelian segregation of alleles during reproduction between an individual of genotype j and an individual of genotype k, therefore depends on the recombination rate ρ between the color pattern locus P and the preference locus M.

We normalized this matrix as $\forall i \sum_j f^{t+1}_{j,pop} = 1$

$$f^{t+1}_{i,pop} = \frac{F^{t+1}_{i,pop}}{\sum_j F^{t+1}_{j,pop}} \tag{6}$$

Overall, the change in the number of genotype i in population pop is given by:

$$\Delta R^t_{i,pop} = r(1 - \frac{N^t_{tot,pop}}{K})N^t_{i,pop}f^{t+1}_{i,pop} \tag{7}$$

Survival

We assumed a mortality of larvae δ, independent from color pattern genotype. We assumed that a recessive genetic load δ_i associated with the different alleles at locus P with $i \in 1, 2, 3$, influencing on the survival probabilities of homozygote at the supergene P.

$$\delta_i = \begin{cases} \delta_1 & \text{if } (m1, m2) = (a, a) \\ \delta_2 & \text{if } (m1, m2) = (b, b) \\ \delta_3 & \text{if } (m1, m2) = (c, c) \\ 0 & \text{else} \end{cases} \tag{8}$$

$$\Delta S^t_{i,pop} = -(1 - (1 - \delta)(1 - \delta_i))N^t_{i,pop} \tag{9}$$

Tracking the evolution of the population using numerical analyses

Overall, the change in the number of genotype i in the population pop is given by:

$$\Delta N^t_{i,pop} = \Delta P^t_{i,pop} + \Delta R^t_{i,pop} + \Delta M^t_{i,pop} + \Delta S^t_{i,pop} \tag{10}$$

All parameters and range values used in the different simulations are summarized in table 1 below. Simulations were performed using Python v.3.

The complexity of this two-locus diploid model prevents comprehensive exploration with analytical methods. The model was thus studied using deterministic simulations, to provide general predictions, neglecting the effect of stochastic processes, such as drift. Our predictions might thus be relevant for species with large effective population size, such as *H. numata*. We

Abbreviation	Name	Parameter range
N_i^0	Initial size of the population i	100
d	Predation strength	[0,1]
σ	Spatial heterogeneity of local mimicry ring	0.5
λ	Unpalatability coefficient	0.0002
mig	Migration rate	[0,1]
ρ	Recombination rate	[0, 0.5]
r	Growth rate	2
K	Carrying capacity	2000
δ	Baseline death rate	[0, 1]
δ_i	Genetic load linked to allele i	[0, 1]
$cost$	cost of choosiness	[0, 1]
Δf^t	Temporal variations	

Table 1: **Description of parameters used in the model and range explored in simulations.**

used discrete time simulations where all events (reproduction, predation and migration) occur simultaneously, therefore relevantly stimulating a natural population with overlapping generations.

In our simulations, the growth rate r was set to 2, the carrying capacity K was assumed to be equal to 2000 per population. Initial population sizes $N^0_{tot,1}$ and $N^0_{tot,2}$ were 100 individuals. The three alleles at the locus P controlling color pattern variations were introduced in proportion $\frac{1}{3}$ in each population. We set the toxicity parameter λ to 0.0002, and the spatial heterogeneity of mimetic communities σ to 0.5. These parameter values are selected as conditions where wing color pattern polymorphism could be maintained without any genetic load or disassortative mating behavior, based on a previous study (Llaurens et al. , 2013).

Results

Effect of mate choice on polymorphism

As already highlighted in the literature (Llaurens et al. , 2013), assuming random mating, polymorphism can be maintained through equilibrium between spatially heterogeneous selection and migration, a mechanism notably addressed in situations where mimetic communities of defended species are spatial heterogenic (Joron et al. , 2005). Without migration, allele a and b become fixed in population 1 and 2 respectively, because of their provided mimicry advantage within their local communities. Polymorphism with persistence of alleles a and b within both patchs can be maintain with an intermediate migration rate but the non-mimetic allele c is always lost in both populations (fig.1 A).

To test the effect of mate choice on the selection/migration equilibrium previously described, simulations were carried out introducing mate choice (random, assortative and disassortative) and computing the evolution of frequencies at the color pattern locus during 10 000 generations for different migration rates *mig*. Assuming assortative mating *via* self-referencing (hyp. 1) leads to fixation of allele a in both patches for migration rate tested, because allele a is the

most frequently expressed due to its dominance over the other alleles (fig.1 B). In contrast, disassortative mating maintains higher polymorphism, with both the two mimetic alleles *a* and *b*, and the non-mimetic one *c* persisting, whatever the strength of the migration. The non-mimetic phenotype [C] is rarely expressed because it is carried by the recessive allele *c* persisting at low frequency, but it still benefits from high reproductive success when expressed because of disassortative mating. The strong negative FDS generated by disassortative mating acting on color pattern alleles counteracts the positive FDS due to predator behavior acting on the same trait. Mate choice thus strongly impact the dynamics of alleles within and between patches.

Linked genetic load favor the persistence of non-mimetic allele

In the following simulations, migration parameter *mig* was thus set to 0.1, allowing a persistence of polymorphism of alleles *a* and *b* at the color pattern locus *P*, under random mating assumption. We then investigated the role of genetic load associated with the different color pattern alleles on the polymorphism at wing color pattern locus. This allows to infer the effect of heterozygote advantage generated by genetic load on polymorphism, independently from the evolution of mating preferences. We assumed that the genetic load associated with the most dominant allele *a* and the intermediate allele *b* had similar strength, *i.e.* $\delta_1 = \delta_2$, assumed higher than the genetic load associated with the recessive allele *c*, namely δ_3. Overall, we observed that polymorphism of phenotypes [A] and [B] is maintained in both populations when there is a genetic load associated with the non mimetic allele *c* only ($\delta_1 = \delta_2 = 0$ and $\delta_3 > 0$) or when this load is higher that the one associated with alleles *a* and *b* (Supp. table S4). The non-mimetic allele *c* is then maintained with the other alleles *a* and *b* within both populations, when (i) all three alleles carry a genetic load with similar strength, *i.e* $\delta_1 = \delta_2 = \delta_3 > 0$ or (ii) when the allele *c* is the only one not carrying a genetic load ($\delta_1 = \delta_2 > 0$ and $\delta_3 = 0$) (Supp. table S4). The heterozygote advantage generated by genetic load linked to the alleles at the locus *P* thus allows the persistence of balanced polymorphism and more specifically favors the maintenance of the non mimetic allele *c* within both populations.

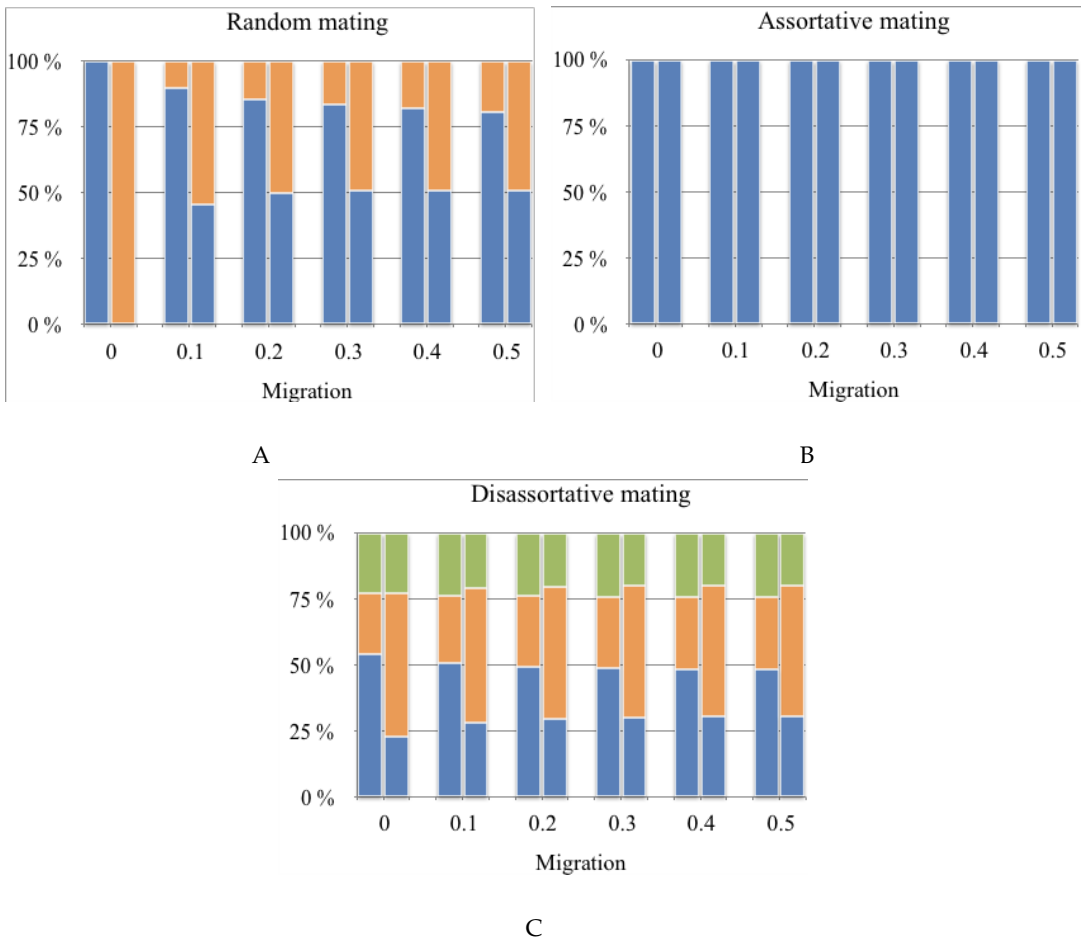


Figure 1: **Impact of mating preferences on color pattern diversity in both populations**. The equilibrium frequencies of color pattern phenotypes in population 1 and 2 for different migration rates *mig* are computed assuming different mating behaviors, *i.e.* assortative *(a)*, random *(b)* or disassortative *(c)*. Joint bars the frequencies of color pattern phenotypes [A], [B] and [C] (as blue, orange and green areas respectively) in population 1 and 2 (on the left and right side respectively, within each level of migration). The three alleles at the locus P controlling color pattern variations were introduced in proportion $\frac{1}{3}$ in each population and the genetic architecture to describe the locus M corresponded to self-referencing (hyp. 1). Simulations were run assuming $r = 2$, $K = 2000$, $N^0_{tot,1} = N^0_{tot,2} = 100$, $\lambda = 0.0002$, $\sigma = 0.5$, $d = 0$, $\rho = 0$, $cost = 0.1$, $\delta_1 = \delta_2 = \delta_3 = 0$ and $\delta = 0$.

Evolution of disassortative mating

Because heterozygote advantage at color pattern locus P enhances the evolution of disassortative mating preferences at the locus M, we first investigate the effect of genetic load on the evolution of

disassortative behavior, by testing the invasion of disassortative mutant inducing self-avoidance (hyp. 1) in a population initially performing random mating. Figure 2 shows that the genetic load associated with alleles a and b ($\delta_1 = \delta_2$) has a strong positive impact on the emergence of disassortative mating. The genetic load associated with the recessive allele c (δ_3) has a slighter positive effect on the evolution of disassortative mating. Simulations assuming different costs of choosiness (*cost*) show a similar effect of associated genetic loads, although increasing this cost slows down the invasion of the choosy disassortative mating mutant *dis* (see Sup. fig. S5). Overall, this confirms that linked genetic load is favoring the evolution of disassortative mating behavior in both populations.

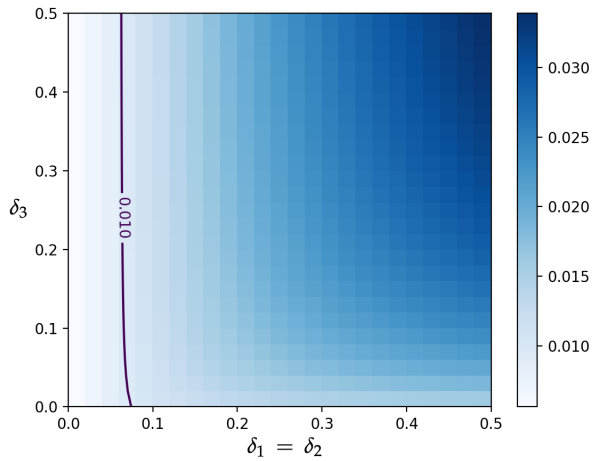
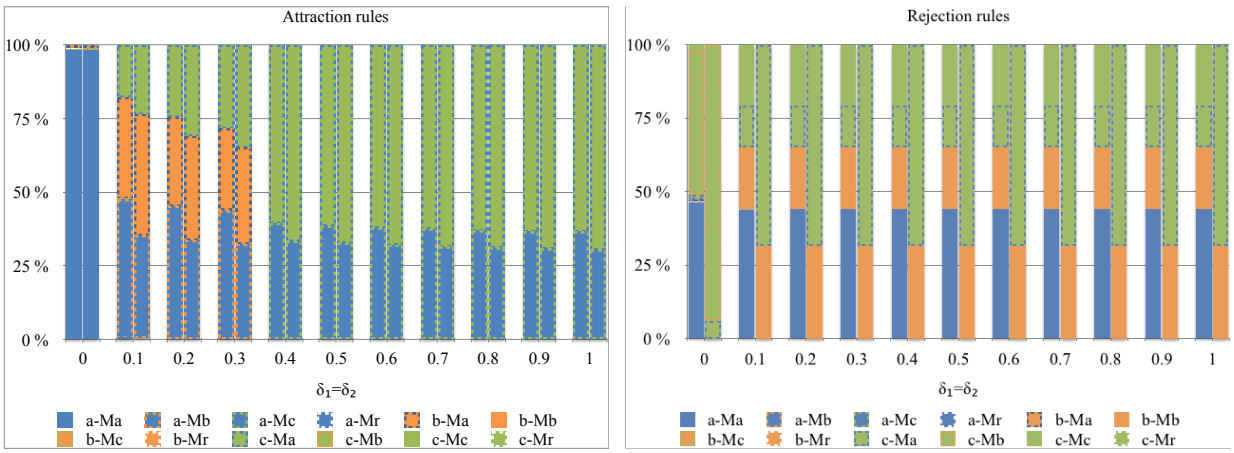


Figure 2: **Influence of a linked genetic load on the evolution of disassortative mating, assuming self-referencing (Hyp.1).** The invasion of a mutant with disassortative mating preferences is shown as a function of the strength of the genetic load associated with dominant alleles a and b assumed equal ($\delta_1 = \delta_2$) (x axis) and with the recessive allele c (δ_3) (y axis). The shade of blue indicates the frequency of the mutant with disassortative mating preferences dis, inducing self-avoidance based on phenotype (hyp. 1), after 100 generations. The three alleles at the locus controlling colour pattern variations were introduced in even proportion (1/3) in each population, and the initial frequency of the mutant was 0.01, shown by the vertical purple line marking the limit of invasion by the mutant. Simulations were run assuming Simulations were run assuming $r = 2$, $K = 2000$, $N^0_{tot,1} = N^0_{tot,2} = 100$, $\lambda = 0.0002$, $\sigma = 0.5$, $d = 0.1$, $\rho = 0$, $mig = 0.1$, $\delta = 0.1$ and $cost = 0.1$

Which genetic architecture of mating preferences favors the evolution of disassortative mating behavior ?

To study the evolution of mating behavior assuming different genetic architecture of mate preferences, we investigated the invasion of mate preference alleles M_r, M_a, M_b and M_c controlling random mating, recognition of phenotype A, B and C respectively. We compared simulations where the recognition alleles triggered either attraction or rejection of the recognized color pattern phenotype (fig.3 A and fig.3 B respectively) .



A B

Figure 3: **Influence of a genetic load on haplotype diversity, assuming attraction alleles at the preference locus (hyp. 2a) and (hyp. 2b)**. The proportion of haplotypes obtained after 2000 generations in both populations under hyp.2a for different values of δ_1 and $\delta 2$. The three alleles at the locus P controlling color pattern variations were introduced in even proportion $\frac{1}{3}$ in each population and the genetic architecture to describe the locus M corresponded to specific recognition of colour pattern alleles inducing attraction (hyp.2a). Simulations were run assuming $r = 2$, $K = 2000$, $N^0_{tot,1} = N^0_{tot,2} = 100$, $\lambda = 0.0002$, $\sigma = 0.5$, $d = 0.1$, $\rho = 0$, $mig = 0.1$, $\delta_3 = 0$, $\delta = 0.1$ and $cost = 0.1$.

When preference alleles lead to attraction behavior (Hyp.2a), we observed high equilibrium frequencies of haplotypes $a - M_b$ and $b - M_a$ in both populations when δ_2 and δ_3 differed from 0 (fig.3 A). These two haplotypes benefit from both mimicry and limited expression of genetic load. The haplotype $c - M_a$ is maintained because of the benefit associated with the choice of the

most frequent mimetic haplotypes a, and the limited expression of the phenotype C, enabled by recessiveness. Nevertheless, the haplotype $b - M_a$ is lost as the genetic load increases and cannot be compensated by the beneficial effect of mimicry. As a consequence, the mimetic phenotype B is not maintained in populations where genetic load is high, and the dominant phenotype A becomes predominantly expressed in both populations.

When preference alleles lead to rejection behavior (Hyp.2b) and a genetic load is associated with the mimetic alleles a and b at locus P, these alleles become associated with the corresponding rejection alleles at locus M (*i.e.* $a - M_a$ and $b - M_b$ have an intermediate frequency in both populations) (fig.3 B). The non mimetic allele c is either associated with a self-avoiding allele $c - M_c$, or an allele rejecting the dominant allele a ($c - M_a$). The three alleles (a, b and c) persist as a polymorphism for all positive values of genetic load (fig.3 B), in contrast with the previously-described case where preference alleles lead to attraction (hyp. 2a) where the mimetic allele b is lost when genetic load is high (fig. 3 A). Although the equilibrium haplotype frequencies is similar for all positive values of genetics load assuming preference allele coding for rejection (Hyp.2b), the strength of genetic load still impact the temporal dynamics of haplotypes, the equilibrium being reached earlier as the genetic load increases (see sup. fig S6).

We then investigated how the evolution of haplotype frequencies translates into individual behavior: assuming that preference alleles lead to attraction (hyp.2a) the majority of individuals display assortative preferences, even when genetic load is high (figure 4A). In contrast, when preference alleles lead to rejection (hyp.2b), most individuals have a disassortative mating behavior (figure 4B). This highlights that the genetic architecture of mate preference plays a key role in the evolution of individuals mating behaviors.

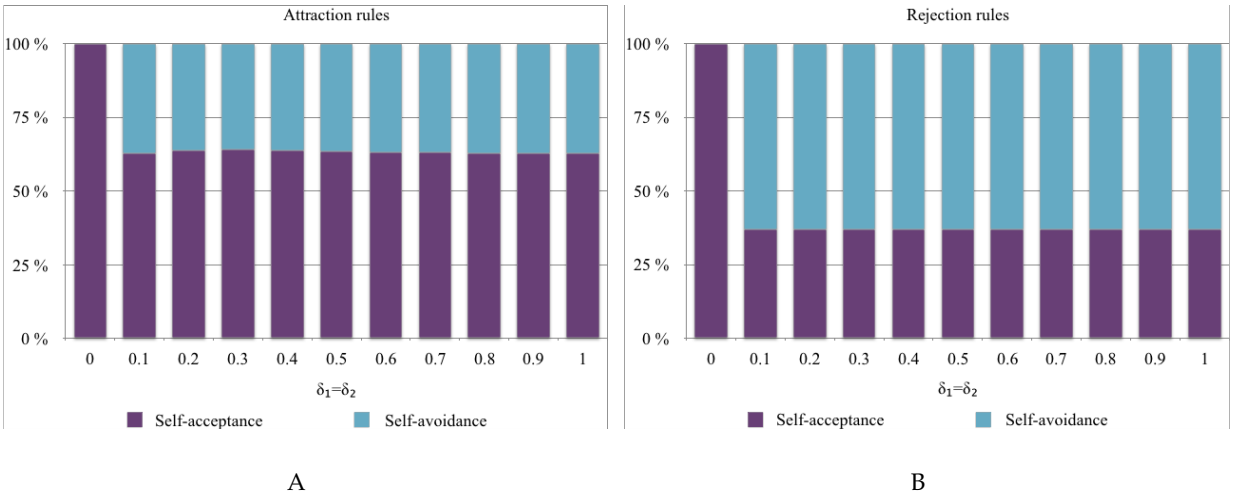


Figure 4: **Influence of a genetic load on the distribution of mating behaviour observed at the population level, assuming attraction (a) or rejection (b) alleles at the preference locus**. The proportion of individual displaying self-acceptance (in purple) and self-avoidance (in blue) behaviors depends on the level of genetic load of δ_1 and $\delta 2$, assuming that preference alleles either lead to *(a)* attraction (hyp.2a), or *(b)* rejection (hyp.2b). The three alleles at the locus P controlling color pattern variations were introduced in even proportion $\frac{1}{3}$ in each population. Simulations were run assuming $r = 2$, $K = 2000$, $N^0_{tot,1} = N^0_{tot,2} = 100$, $\lambda = 0.0002$, $\sigma = 0.5$, $d = 0.1$, $\rho = 0$, $mig = 0.1$, $\delta_3 = 0$, $\delta = 0.1$ and $cost = 0.1$.

Impact of genetic linkage between locus P and M on the evolution of disassortative mating

We observed that the genetic load associated with the two most dominant alleles at the color pattern locus P impact the evolution of mate choice, when the color pattern locus P and the preference locus M are fully linked. We then tested the effect of recombination between alleles at the two loci on the evolution of mate choice by performing simulations with different values of the recombination rate ρ. Assuming self-referencing (hyp.1), increasing the level of recombination rate further promotes the invasion of the disassortative allele *dis* (see Sup. fig S7). Under hyp. 1, the mating preference depends on the phenotype displayed by the individual, so that the allele *dis* always translates into a disassortative behavior, irrespective the linkage disequilibrium between the preference locus and the color pattern locus. Recombination thus only results in a

more rapid fixation of the disassortative mating allele *dis*, advantaged by the associated genetic load.

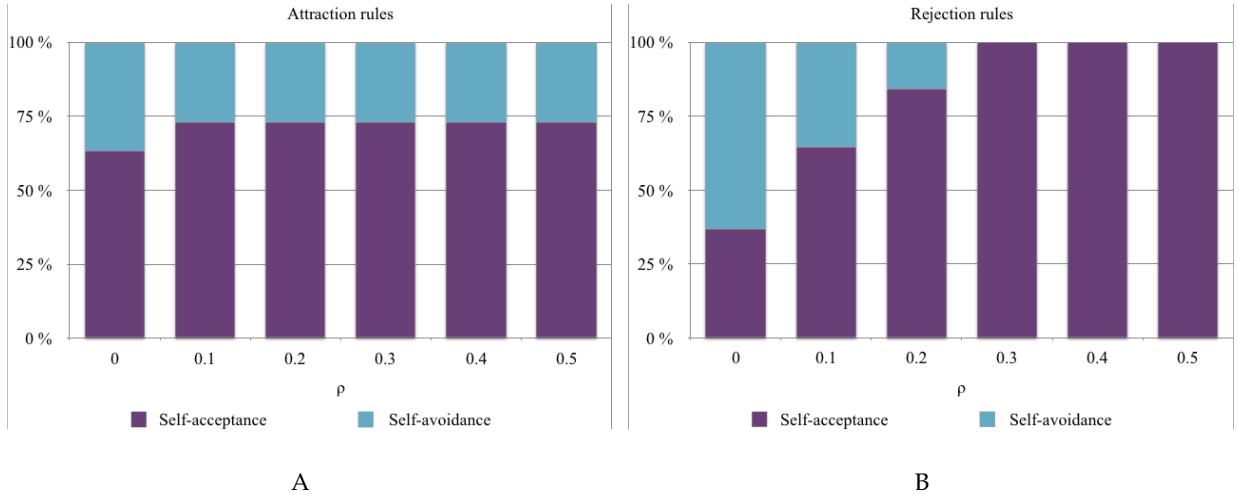


Figure 5: **Influence of recombination between colour pattern and preference alleles on the distribution of mating behaviours at the population level, assuming *(a)* attraction or *(b)* rejection alleles at the preference locus**. The proportion of individuals displaying self-acceptance (in purple) and self-avoidance (in blue) depends on the recombination rate ρ between the preference locus M and the color pattern locus P, assuming that preference alleles either lead to *(a)* attraction (hyp.2a), or *(b)* rejection (hyp.2b). The three alleles at the locus P controlling color pattern variations were introduced in proportion $\frac{1}{3}$ in each population. Simulations were run assuming $r = 2$, $K = 2000$, $N^0_{tot,1} = N^0_{tot,2} = 100$, $\lambda = 0.0002$, $\sigma = 0.5$, $d = 0.1$, $\rho = 0$, $mig = 0.1$, $\delta_3 = 0$, $\delta = 0.1$ and $cost = 0$.

Contrastingly, when assuming specific preference for a given color pattern allele (hyp.2), mating behavior depends on the genotype at the preference locus M independently from the phenotype of the choosing individuals, so that we expected a stronger effect of recombination rate on mate choice evolution assuming this genetic architecture. Figure 5 indeed confirmed that, by decoupling preference alleles from color patterns alleles, recombination between locus P and M leads to a decrease in the proportion of individuals performing self-avoidance. The evolution of disassortative mating behaviors is even more impaired assuming that preference alleles generate rejection (hyp. 2a), where disassortative mating behavior completely disappears when recombination is free (*i.e.* when $\rho = 0.5$). Under attraction rule (hyp.2.a), for each color pattern allele, two third of the possible haplotypes lead to self-avoidance (for instance $a - Mb$

and $a - Mc$ for color pattern allele a). In contrast, under rejection rule (hyp.2.b), only one out the three possible haplotypes leads to self-avoidance (for instance $a - Ma$ for color pattern allele a). Moreover, the allele encoding for the rejection of a given color pattern (*e.g.* Ma) rarely co-occur the rejected allele (*e.g.* a) within genotype because mate choice limits crosses between an individual carrying a rejection allele on one hand and an individual displaying the rejected allele on the other hand. This rare co-occurrence between rejecting and rejected alleles further limits the creation of disassortative haplotypes (*e.g.* $a - Ma$) by recombination when assuming rejection rule (Hyp.2b). Overall, genetic architectures enabling recombination between color pattern locus and preference locus thus limit the evolution of disassortative mating, when assuming a preference locus with allele encoding for a specific recognition of mating cues (*i.e.* under hyp.2).

Discussion

The dark side of adaptive alleles: how linked genetic load leads to balanced polymorphism

Our results show that polymorphism of haplotypes can be maintained within locus either via heterozygote advantage generated by linked recessive deleterious mutations, or via disassortative mating. Both mechanisms enable a recessive non-mimetic haplotype to be maintained within populations. Many polymorphic traits involve inversion polymorphisms, and ancestral haplotypes are generally recessive over derived ones (see (Llaurens et al. , 2017)). Following a mechanism known as Haldane's sieve(Haldane et al. , 1956), new adaptive haplotypes initially occurring as heterozygotes with the ancestral haplotype are indeed more likely to be favored by natural selection when dominant, because recessive haplotypes, unexpressed in heterozygotes, are likely to be lost by drift. Ancestral haplotypes are often free from associated deleterious mutations because of purging enabled by recombination. They are then maintained after the invasion of a new inverted haplotype, because of heterozygote advantage generated by deleterious

mutations captured by the inversions. Many rearranged haplotypes are indeed associated with serious fitness reduction at homozygotes state, such as in the derived haplotype of the supergene controlling alternative social organisation in the fire ant Solenopsis (Wang et al. , 2013), or that of the supergene controlling plumage and behavioural variation in the white-throated sparrow (Horton et al. , 2013). Structural heterozygotes may also have intermediate fitness, as in the supergene controlling plumage and behavioural variations in the ruff in which the rearranged haplotypes are homozygote lethal and also suffer from some significant survival cost as heterozygotes (Küpper et al. , 2016). Inversions may thus be under strong positive selection because they capture an adaptive allele or lock together a combination of adaptive alleles, but they are also likely to carry along deleterious mutations that are protected from purge by the limited recombination with standard haplotypes. Adaptive haplotypes carried by inversion may therefore fail to become fixed within populations, resulting in persistent polymorphism. As highlighted in our results, such heterozygote advantage can then promote the evolution of disassortative mating.

Evolution of disassortative mating results from interactions between dominance and deleterious mutations

The evolution of disassortative mating depends on the level of genetic load and its association with the different haplotypes. Here we show that a genetic load associated with dominant haplotypes play a more significant role than when associated with the most recessive haplotype. In the polymorphic butterfly *Heliconius numata*, the most dominant haplotype bicoloratus is associated with a significant deficit of homozygotes in natural populations, reflecting that the high frequency of the corresponding phenotype is mostly composed of heterozygotes (Chouteau et al. , 2017). This suggests that a genetic load might be associated with this dominant haplotype at the supergene P (Arias et al. , 2014). Similarly, in the self-incompatibility locus of the Brassicaceae, a significant lack of dominant homozygotes in controlled crosses have revealed that dominant haplotypes carried a higher genetic load than recessive hapotypes (Llaurens et al. , 2009).

Disassortative mating thus bring fitness advantage by reducing the number of homozygote offspring suffering from the expression of deleterious mutataions. Once disassortative mating is established within populations, as in self-incompatible plants, recessive deleterious mutations associated with the dominant haplotype become sheltered because the formation of dominant homozygotes is strongly reduced,limiting the opportunities for purging deleterious mutations via recombination (Llaurens et al. , 2009). Disassortative mating might therefore reinforce heterozygote advantage, through an interaction between two aspects of the genetic architecture, dominance and recombination, affecting respectively positive selection on beneficial haplotypes (Haldane's sieve) and the purge of associated deleterious mutations (limited recombination e.g. via inversions) (Haldane et al. , 1956). Here, based on a model of mimicry supergene, we assumed that mate choice relied on wing pattern phenotype, and that heterozygotes had an identical phenotype to dominant homozygotes, as observed in controlled crosses (Arias et al. , 2014). We assumed that mate choice relied on wing pattern as a cue, in accordance with behavioural test (Chouteau et al. , 2017), but other cues, like pheromones, might also be associated with the supergene, possibly enabling discrimination between homozygotes and heterozygotes. Depending on the cues and the dominance relationships among haplotypes, different mate choice behaviours may evolve, that might modulate the evolution of polymorphism within population. In the ant Formica seylisi, shich displays polymorphic social organisation controlled by chromosomal inversions (Purcell et al. , 2014), asymmetric assortative mating was reported, where wild-caught queens with the homozygote genotype m/m were never mated with males carrying the alternative haplotype p (Avril et al. , 2018). Our model thus stresses the need to document dominance relationships precisely across haplotypes segregating at polymorphic loci, as well as mate choice behaviour and cues, to understand the evolution forces involved in disassortative mating evolution.

Genetic architecture of disassortative mating: predictions brought by modelling

Our model shows that disassortative mating is more likely to emerge when genetic architecture induces self-referencing. Nevertheless, the few documented loci involved in mate preference in *Heliconius* butterflies seem to be associated with attraction towards a specific cue. Wing-pattern-based assortative mating is reported in the vast majority of *Heliconius* species (Merrill et al. , 2015) and a few preference loci have been identified. The locus controlling preference for yellow in *H.cydno* maps close to the locus *K* (Kronforst et al. , 2006), controlling white/yellow switch in this species. More recently, (Merrill et al. , 2019) identify a major QTL associated with preference towards the red color in crosses between*H.melpomene* displaying red and black wing color pattern and *H. melpomene* displaying black and white color patterns. Moreover, this QTL is located close to the gene *optix* involved in the variation of red patterning in *H. melpomene*. The evolution of assortative mating in *Heliconius* butterflies thus seems to rely on alleles encoding for attraction towards specific cues in linkage with loci involved in the variation of these cues. Nevertheless, empirical data are still scarce and QTL for preference behavior have also been found unlinked to known wing-patterning loci (Merrill et al. , 2019).

Here we explored the evolution of disassortative through the emergence of rejection or attraction haplotypes and show that disassortative mating is more likely to emerge assuming rejection haplotypes linked to the locus controlling color pattern variation. Nevertheless, even when such haplotypes invade the population, the resulting behaviors were only partially disassortative, with significant variations in preferences across individuals. These results might nevertheless fit the empirical data of tetrad experiments carried out in *H. numata*: when two males of each phenotype were put with two females of each phenotype, only 75 percent of crosses happens between individuals with different phenotypes (Chouteau et al. , 2017). Our theoretical predictions point at the need to precisely characterize the genetic basis of mate preference, and the linkage desequilibrium with the locus controlling variations in the mating cues.

Conclusions

By focusing on a well-documented case-study of disassortative mating on cues subjected to natural selection, we show that balancing selection promoting (i) local polymorphism and (ii) leading to heterozygote advantage because of the genetic architecture involving inversions, is likely to favor the evolution of disassortative mating behavior. The genetic basis of such behavior is predicted to involve haplotypes triggering rejection toward specific cues. Such rejection loci promotes the disassortative mating when they are in tight linkage with the locus controlling mating cue variations.

Acknowledgments

The authors would like to thank Charline Smadi and Emmanuelle Porcher for feedbacks on the modeling approach developed here. We also thank Thomas Aubier and Richard Merrill and the whole *Heliconius* group for stimulating discussion on mate choice evolution in our favorite butterflies. This work was supported by the Emergence program from Paris City Council to VL and the ANR grant SUPERGENE to MJ.

Literature Cited

Cook, O. E., G. H. Collaborator, and A. Q. Expert. 2015. Data from: Template and guidelines for using LaTeX in *The American Naturalist*. American Naturalist, Dryad Digital Repository, http://dx.doi.org/10.5061/dryad.XYZAB.

Davis, E. B., K. A. Brakora, and A. H. Lee. 2011. Evolution of ruminant headgear: a review. Proceedings of the Royal Society B 278:2857–2865.

Inglis, R. F., P. G. Roberts, A. Gardner, and A. Buckling. 2011. Spite and the scale of competition in *Pseudomonas aeruginosa*. American Naturalist 178:276–285.

Lemodèle, P.-Q., A. B. Kapitelschreiber, and C. D. E. Exemplar. 2007. An exemplary instance of chapters in books. Pages 231–245 *in* J.-P. Écrivain and M. A. Természettudós, eds. Inspiring Instances of Brilliant Writing. Truth Pudding Press, Fond du Lac, WI.

Xiao, X., D. J. McGlinn, and E. P. White. 2015. A strong test of the maximum entropy theory of ecology. American Naturalist 185:E705–E80.

Sherratt, Thomas. 2008. The evolution of Müllerian mimicry Naturwissenschaften

Thomas N. Sherratt The optimal sampling strategy for unfamiliar prey Evolution

Richard M. Merrill et al. Disruptive ecological selection on a mating cue The Royal Society

Mathieu Chouteau, Violaine Llaurens, Florence Piron-Prunier, and Mathieu Joron. 2017. Polymorphism at a mimicry supergene maintained by opposing frequency-dependent selection pressures PNAS

Joron, Mathieu and Papa, Riccardo and Beltrán, Margarita and Chamberlain, Nicola and Mavárez, Jesús and Baxter, Simon and Abanto, Moisés and Bermingham, Eldredge and Humphray, Sean J and Rogers, Jane and others. 2006. A conserved supergene locus controls colour pattern diversity in Heliconius butterflies PLoS biology

Joron, Mathieu and Frezal, Lise and Jones, Robert T and Chamberlain, Nicola L and Lee, Siu F and Haag, Christoph R and Whibley, Annabel and Becuwe, Michel and Baxter, Simon W and Ferguson, Laura and others Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry Nature

Yuexin Jiang, Daniel I. Bolnick and Mark Kirkpatrick Assortative Mating in Animals The American Naturalist

Joseph Felsenstein Continuous-genotype models and assortative mating Theoretical Population Biology

Alexey S. Kondrashov, Max Shpak On the origin of species by means of assortative mating The Royal Society

Michael Lynch, Bruce Walsh Genetics and Analysis of Quantitative Traits

Anne Pusey, Marisa Wolf Inbreeding avoidance in animals

Kopp, Michael and Servedio, Maria R and Mendelson, Tamra C and Safran, Rebecca J and Rodríguez, Rafael L and Hauber, Mark E and Scordato, Elizabeth C and Symes, Laurel B and Balakrishnan, Christopher N and Zonana, David M and others Mechanisms of assortative mating in speciation with gene flow: connecting theory and empirical research The American Naturalist

Tuttle, Elaina M and Bergland, Alan O and Korody, Marisa L and Brewer, Michael S and Newhouse, Daniel J and Minx, Patrick and Stager, Maria and Betuel, Adam and Cheviron, Zachary A and Warren, Wesley C and others Divergence and functional degradation of a sex chromosome-like supergene Current Biology

Tetsumi Takahashi, Michio Hori Evidence of disassortative mating in a Tanganyikan cichlid fish and its role in the maintenance of intrapopulation dimorphism The Royal Society

Michio Hori Frequency-Dependent Natural Selection in the Handedness of Scale-Eating Cichlid Fish Science

Tregenza, T and Wedell, N Genetic compatibility, mate choice and patterns of parentage: invited review

Leonardo Campagna Supergenes: The Genomic Architecture of a Bird with Four Sexes Current Biology

Joron, Mathieu and Wynne, Ian R and Lamas, Gerardo and Mallet, James Variable selection and the coexistence of multiple mimetic forms of the butterfly Heliconius numata Evolutionary Ecology

Arias, Mónica and Meichanetzoglou, Aimilia and Elias, Marianne and Rosser, Neil and De-Silva, Donna Lisa and Nay, Bastien and Llaurens, Violaine Variation in cyanogenic compounds concentration within a Heliconius butterfly community: does mimicry explain everything? BMC evolutionary biology

Le Poul, Yann and Whibley, Annabel and Chouteau, Mathieu and Prunier, Florence and Llaurens, Violaine and Joron, Mathieu Evolution of dominance mechanisms at a butterfly mimicry supergene Nature Communications

Thomas N. Sherratt Spatial mosaic formation through frequency-dependent selection in Müllerian mimicry complexes Journal of Theoretical Biology

Merrill, Richard M and Dasmahapatra, Kanchon K and Davey, JW and Dell'Aglio, DD and Hanly, JJ and Huber, B and Jiggins, Chris D and Joron, M and Kozak, KM and Llaurens, V and others The diversification of Heliconius butterflies: what have we learned in 150 years? Journal of Evolutionary Biology

Chouteau, Mathieu and Arias, Mónica and Joron, Mathieu Warning signals are under positive frequency-dependent selection in nature Proceedings of the national Academy of Sciences

Joron, Mathieu and Iwasa, Yoh The evolution of a Müllerian mimic in a spatially distributed community Journal of Theoretical Biology

Emmanuelle Porcher, Russell Lande, and C. Fenster Loss of gametophytic self-incompatibility with evolution of inbreeding depression Evolution

Llaurens, Violaine and Billiard, Sylvain and Joron, Mathieu The effect of dominance on polymorphism in Müllerian mimicry Journal of theoretical biology

Richard M. Merrill, Pasi Rastas, Maria C. Melo, Sarah Barker, JohnDavey, W. Owen McMillan3 and Chris D. Jiggins Genetic dissection of assortative mating behavior

Kirkpatrick, Mark How and why chromosome inversions evolve PLoS biology

Sergey Gavrilets Models of Speciation: Where Are We Now? Journal of Heredity

Clement Küpper, Michael Stocks, Judith E Risse, Natalie dos Remedios, Lindsay L Farrell, Susan B McRae, Tawna C Morgan, Natalia Karlionova, Pavel Pinchuk, Yvonne I Verkuil, Alexander S Kitaysky, John C Wingfield, Theunis Piersma, Kai Zeng, Jon Slate1, Mark Blaxter, David B Lank & Terry Burke A supergene determines highly divergent male reproductive morphs in the ruff Nature genetics

Hori, Michio Frequency-dependent natural selection in the handedness of scale-eating cichlid fish Science

Schilthuizen, Menno and Craze, Paul Graham and Cabanban, Annadel Sarmiento and Davison, A and Stone, J and Gittenberger, E and Scott, BJ Sexual selection maintains whole-body chiral dimorphism in snails Journal of evolutionary biology

Penn, Dustin J and Potts, Wayne K The evolution of mating preferences and major histocompatibility complex genes The American Naturalist

Wedekind, Claus and Seebeck, Thomas and Bettens, Florence and Paepke, Alexander J MHC-dependent mate preferences in humans Proceedings of the Royal Society of London. Series B: Biological Sciences

Piertney, SB and Oliver, MK The evolutionary ecology of the major histocompatibility complex Heredity

Hiscock, Simon J and McInnis, Stephanie M Pollen recognition and rejection during the sporophytic self-incompatibility response: Brassica and beyond Trends in plant science

Wedekind, Claus and Füri, Sandra", journal="Proceedings of the Royal Society of London. Series B: Biological Sciences Body odour preferences in men and women: do they aim for specific MHC combinations or simply heterozygosity? The Royal Society"

Kirkpatrick, Mark and Nuismer, Scott L Sexual selection can constrain sympatric speciation Proceedings of the Royal Society of London. Series B: Biological Sciences

Houtman, Anne M and Falls, J Bruce Negative assortative mating in the white-throated sparrow, Zonotrichia albicollis: the role of mate choice and intra-sexual competition Animal Behaviour

Otto, Sarah P and Servedio, Maria R and Nuismer, Scott L Frequency-dependent selection and the evolution of assortative mating Genetics

Casselton, LA Mate recognition in fungi Heredity

Wright, Sewall The distribution of self-sterility alleles in populations Genetics

Wang, John and Wurm, Yannick and Nipitwattanaphon, Mingkwan and Riba-Grognuz, Oksana and Huang, Yu-Ching and Shoemaker, DeWayne and Keller, Laurent A Y-like social chromosome causes alternative colony organization in fire ants Nature

Horton, Brent M and Hu, Yuchen and Martin, Christa L and Bunke, Brian P and Matthews, Beth S and Moore, Ignacio T and Thomas, James W and Maney, Donna L Behavioral characteriza-

tion of a white-throated sparrow homozygous for the ZAL2 m chromosomal rearrangement Behavior genetics

Küpper, Clemens and Stocks, Michael and Risse, Judith E and dos Remedios, Natalie and Farrell, Lindsay L and McRae, Susan B and Morgan, Tawna C and Karlionova, Natalia and Pinchuk, Pavel and Verkuil, Yvonne I and others A supergene determines highly divergent male reproductive morphs in the ruff Nature Genetics

Llaurens, Violaine and Gonthier, Lucy and Billiard, Sylvain The sheltered genetic load linked to the S locus in plants: new insights from theoretical and empirical approaches in sporophytic self-incompatibility Genetics

Haldane, John Burdon Sanderson The theory of selection for melanism in Lepidoptera Proceedings of the Royal Society of London. Series B-Biological Sciences

Chouteau, Mathieu and Llaurens, Violaine and Piron-Prunier, Florence and Joron, Mathieu Polymorphism at a mimicry supergene maintained by opposing frequency-dependent selection pressures Proceedings of the National Academy of Sciences

Avril, Amaury and Purcell, Jessica and Brelsford, Alan and Chapuisat, Michel Asymmetric assortative mating and queen polyandry are linked to a supergene controlling ant social organization Molecular ecology

Purcell, Jessica and Brelsford, Alan and Wurm, Yannick and Perrin, Nicolas and Chapuisat, Michel Convergent genetic architecture underlies social organization in ants Current Biology

Merrill, Richard M and Rastas, Pasi and Martin, Simon H and Melo, Maria C and Barker, Sarah and Davey, John and McMillan, W Owen and Jiggins, Chris D Genetic dissection of assortative mating behavior PLoS biology

Westerman, Erica L and VanKuren, Nicholas W and Massardo, Darli and Tenger-Trolander, Ayşe and Zhang, Wei and Hill, Ryan I and Perry, Michael and Bayala, Erick and Barr, Kenneth

and Chamberlain, Nicola and others Aristaless controls butterfly wing color variation used in mimicry and mate choice Current Biology

Kronforst, Marcus R and Young, Laura G and Kapan, Durrell D and McNeely, Camille and O'Neill, Rachel J and Gilbert, Lawrence E Linkage of butterfly mate preference and wing color preference cue at the genomic location of wingless Proceedings of the National Academy of Sciences

Llaurens, Violaine and Whibley, Annabel and Joron, Mathieu Genetic architecture and balancing selection: the life and death of differentiated variants Molecular ecology

Müller, F Ituna and Thyridia; a remarkable case of mimicry in butterflies (transl. by Ralph Meldola from the original German article in Kosmos, May 1879, 100) Transactions of the Entomological Society of London

Worley, Kirsty and Collet, Julie and Spurgin, Lewis G and Cornwallis, Charlie and Pizzari, Tommaso and Richardson, David S MHC heterozygosity and survival in red junglefowl Molecular Ecology

Supplementary Material

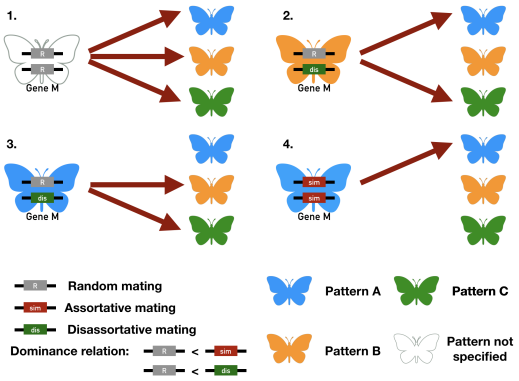


Figure S1: **Mate preferences expressed by the different genotypes at locus *M*, assuming self-referencing (Hyp.1)**. 1. Butterflies carrying two *r* alleles mate at random, independently from either their own color pattern or the color pattern displayed by mating partners. 2-3. Butterflies carrying a *dis* allele display disassortative mating behavior, and mate preferentially with individuals whose color pattern differ from their own. 4. Butterflies carrying a *sim* allele display an assortative mating behavior and thus preferentially mate with individuals displaying the same color pattern. Cases 1 and 4 therefore lead to *self-acceptance*, while cases 2 and 3 lead to *self-avoidance*.

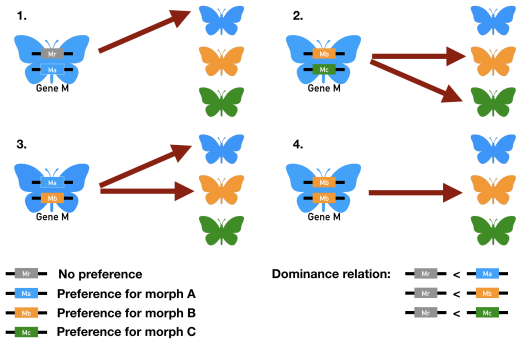


Figure S2: **Mate preferences expressed by the different genotypes at locus *M* assuming preference allele encoding for attraction of specific color patterns (hyp.2.a)**. 1. A butterfly displaying phenotype [A] (in blue) carried one allele coding for specific attraction toward partner displaying phenotype [A] (in blue) and the allele coding for random mating at the locus M controlling the mate choice. This butterfly will mate preferentially with individuals displaying phenotype [A], resulting in assortative mating. 2. A butterfly displaying phenotype [A] (in blue) carries one allele coding for specific attraction toward partner displaying phenotype [B] (in orange) and one allele coding for specific attraction toward partner displaying phenotype [C] (in green). This individual will preferentially mate with individuals displaying phenotype [B] and [C], resulting in disassortative mating 3. A butterfly displaying phenotype [A] (in blue) carries one allele coding for specific attraction toward partner displaying phenotype [A] (in blue) and one allele coding for specific attraction toward partner displaying phenotype [B] (in orange). This individual will preferentially mate with individuals displaying phenotype [A] and [B] 4. A butterfly displaying phenotype [A] (in blue) carries two alleles coding for specific attraction toward partner displaying phenotype [B] (in orange). This individual will preferentially mate with individuals displaying phenotype [B], resulting in disassortative mating. Cases 1 and 3 therefore lead to *self-acceptance*, while cases 2 and 4 lead to *self-avoidance*.

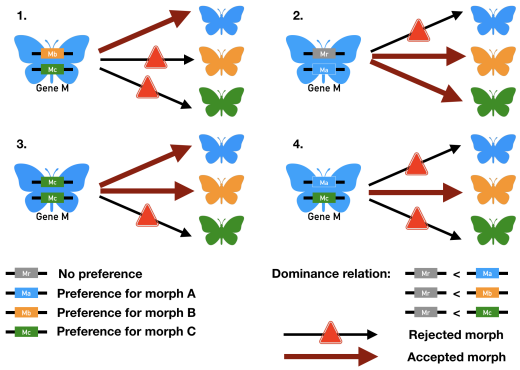


Figure S3: **Mate preferences expressed by the different genotypes at locus *M* preference allele encoding for rejection of specific color patterns (hyp.2.a)**. 1. A butterfly displaying phenotype [A] (in blue) carried one allele coding for specific rejection toward partner displaying phenotype [B] (in orange) and one allele one allele coding for specific rejection toward partner displaying phenotype [C] (in orange). This butterfly will mate preferentially with individuals displaying phenotype [A], resulting in assortative mating. 2. A butterfly displaying phenotype [A] (in blue) carried one allele coding for specific rejection toward partner displaying phenotype [A] (in orange) and one allele coding for random mating (in grey). This butterfly will mate preferentially with individuals displaying phenotypes [B] and [C], resulting in disassortative mating. 3. A butterfly displaying phenotype [A] (in blue) carried two alleles coding for specific rejection toward partners displaying phenotype [C] (in green). This butterfly will mate preferentially with individuals displaying phenotypes [A] and [B]. 4. A butterfly displaying phenotype [A] (in blue) carried one allele coding for specific rejection toward partner displaying phenotype [A] (in blue) and one allele coding for specific rejection toward partner displaying phenotype [C] (in green). This butterfly will mate preferentially with individuals displaying phenotype [B] resulting in disassortative mating. Cases 1 and 3 therefore lead to *self-acceptance*, while cases 2 and 4 lead to *self-avoidance*.

$\delta_1 = \delta_2$	δ_3	Population 1			Population 2		
		Proportion of morph A	Proportion of morph B	Proportion of morph C	Proportion of morph A	Proportion of morph B	Proportion of morph C
0,00	0,00	90 %	10 %	0 %	46 %	54 %	0 %
0,00	0,25	90 %	10 %	0 %	46 %	54 %	0 %
0,00	0,50	90 %	10 %	0 %	46 %	54 %	0 %
0,00	1,00	90 %	10 %	0 %	46 %	54 %	0 %
0,25	0,00	63 %	8 %	28 %	22 %	53 %	25 %
0,25	0,25	79 %	18 %	3 %	35 %	59 %	6 %
0,25	0,50	80 %	18 %	2 %	38 %	58 %	4 %
0,25	1,00	82 %	18 %	1 %	41 %	57 %	2 %
0,50	0,00	56 %	7 %	37 %	19 %	51 %	31 %
0,50	0,25	76 %	19 %	5 %	32 %	59 %	9 %
0,50	0,50	78 %	19 %	3 %	36 %	59 %	5 %
0,50	1,00	80 %	19 %	1 %	39 %	58 %	3 %
1,00	0,00	51 %	5 %	43 %	17 %	48 %	35 %
1,00	0,25	74 %	19 %	7 %	31 %	58 %	11 %
1,00	0,50	77 %	19 %	4 %	35 %	59 %	6 %
1,00	1,00	79 %	19 %	2 %	38 %	58 %	3 %

Figure S4: **Impact of linked genetic load on color pattern polymorphism, assuming random mating.** The proportion of phenotypes [A], [B] and [C] in the population 1 and 2 after 1000 generations depend on the different values of genetic load associated with the recessive allele c (δ_1), intermediate-dominant allele b (δ_2) and dominant allele c (δ_3). Simulation were run assuming $r = 2$, $K = 2000$, $N^0_{tot,1} = N^0_{tot,2} = 100$, $\lambda = 0.0002$, $\sigma = 0.5$, $d = 0.1$, $\rho = 0$, $mig = 0.1$, $\delta = 0.1$ and $cost = 0.1$.

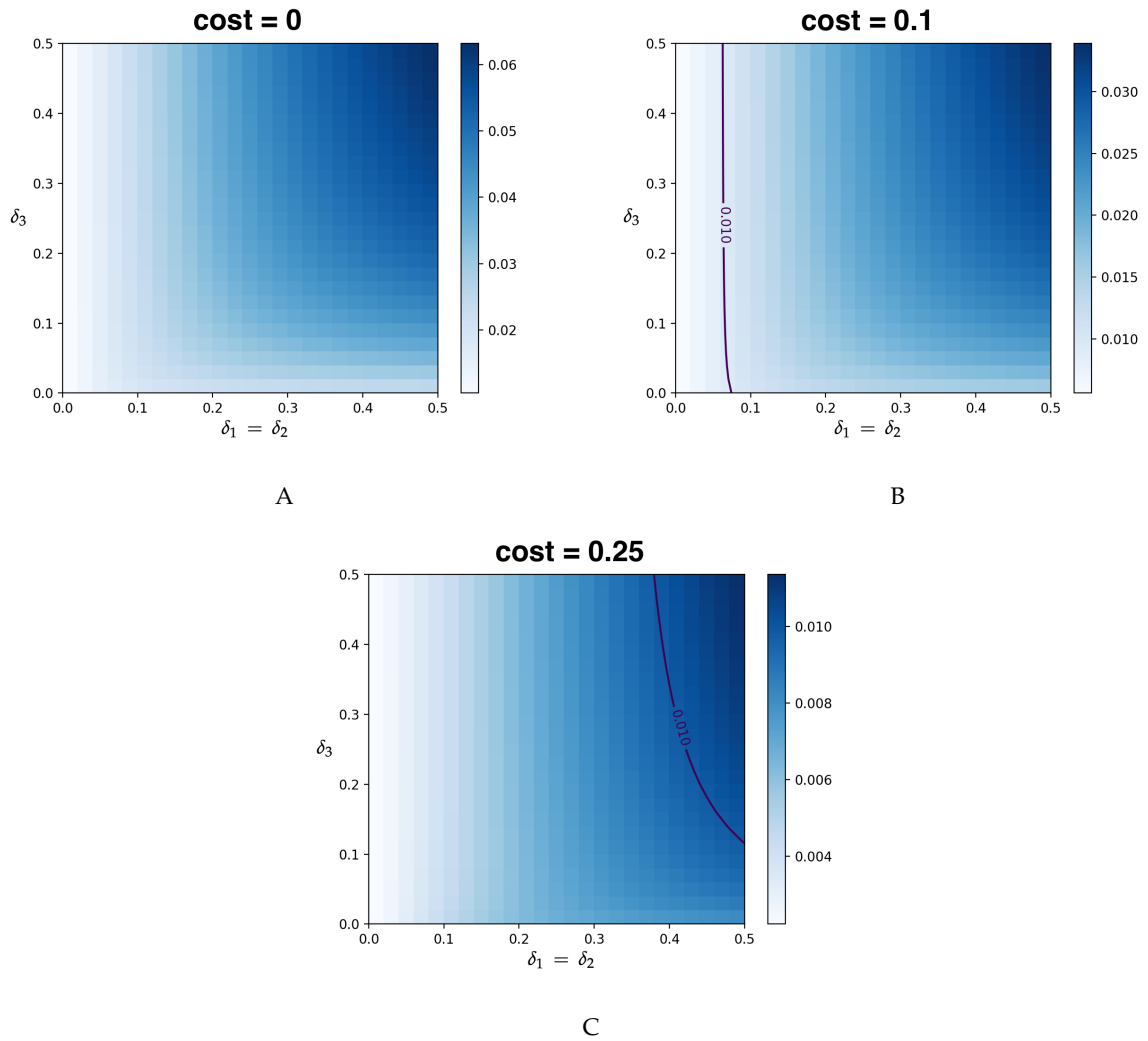


Figure S5: **Effect of the cost of choosiness *cost* on the invasion of the disassortative mutant *dis*, under the self-referencing hypothesis (Hyp.1)**. Simulations were run assuming either *(a)* no cost of choosiness $cost = 0$, *(b)* low cost of choosiness $cost = 0.1$ or *(c)* elevated cost of choosiness $cost = 0.25$. The invasion of the disassortative mutant *dis* always depends on the strength of genetic load associated with the dominant alleles a and b ($\delta_1 = \delta_2$) on the x-axis and to the recessive allele c, δ_3, on the y-axis. Level of blue indicates the frequency of the disassortative mutant *dis*, inducing self-avoidance based on phenotype (hyp. 1), after 100 generations.The three alleles at the locus P controlling color pattern variations were introduced in proportion $\frac{1}{3}$ in each population, and the initial frequency of the mutant was 0.01, shown by the vertical purple line, marking the limit of invasion by the mutant. Simulation were run assuming $r = 2$, $K = 2000$, $N^0_{tot,1} = N^0_{tot,2} = 100$, $\lambda = 0.0002$, $\sigma = 0.5$, $d = 0$, $mig = 0.1$ and $\rho = 0$.

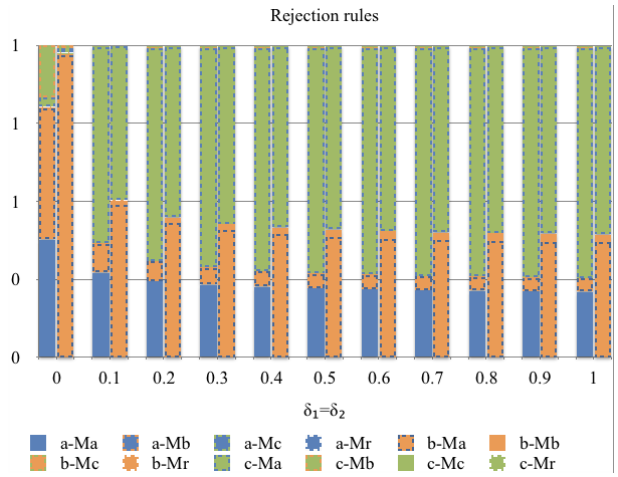


Figure S6: **Impact of the genetic load on haplotype diversity, assuming rejection alleles at the preference locus (Hyp. 2b), during the emergence of preference alleles**. The proportion of haplotypes in both populations 200 generations after the introduction of the mutant, for different values of δ_1 and δ_2. The three alleles at the locus P controlling color pattern variations were introduced in proportion $\frac{1}{3}$ in each population and the genetic architecture to describe the locus M corresponded to specific recognition of P-alleles inducing attraction (hyp.2b). Simulations were run assuming $r = 2$, $K = 2000$, $N^0_{tot,1} = N^0_{tot,2} = 100$, $\lambda = 0.0002$, $\sigma = 0.5$, $d = 0.1$, $\rho = 0$, $mig = 0.1$, $\delta_3 = 0$, $\delta = 0.1$ and $cost = 0.1$.

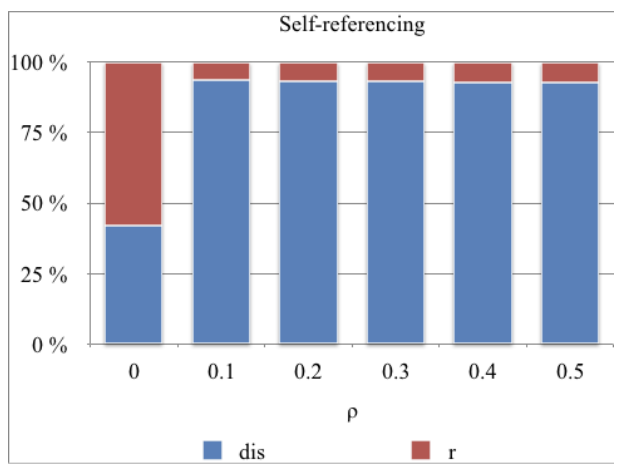


Figure S7: **Impact of recombination between color pattern (locus P) and preference alleles (locus M) on mating behavior, assuming self-referencing preference alleles (Hyp.1).** The proportion of *dis* and *r* alleles in both populations for different values of recombination rate ρ. The three alleles at the locus P controlling color pattern variations were introduced in proportion $\frac{1}{3}$ in each population and the genetic architecture to describe the locus M corresponded to self-referencing (hyp.1). Simulations were run assuming $r = 2$, $K = 2000$, $N^0_{tot,1} = N^0_{tot,2} = 100$, $\lambda = 0.0002$, $\sigma = 0.5$, $d = 0.1$, $mig = 0.1$, $\delta_1 = 0.5$, $\delta_2 = 0.5$ $\delta_3 = 0$, $\delta = 0.1$ and $cost = 0.1$.