Evolution and genetic architecture of disassortative mating at a locus under heterozygote advantage

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

1. Institut de Systematique, Evolution, Biodiversité (UMR7205), Museum National d'Histoire Naturelle, CNRS, Sorbonne-Université, EPHE, Université des Antilles, CP50, 57 rue Cuvier, 75005 Paris, France;

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

3. (CEFE), Univ Montpellier, CNRS, EPHE, IRD, Univ Paul Valéry Montpellier 3, Montpellier, France;

∗ Corresponding author: Ludovic Maisonneuve; e-mail: ludovic.maisonneuve.2015@polytechnique.org.

Manuscript elements: Table 1, Table 2, figure 1, figure 2, figure 3, figure 4, figure 5, figure S5, figure S5, figure S7,figure S8 , figure S9, figure S10, figure S11.

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

Manuscript type: Article.

Abstract

The evolution of mate preferences may depend on natural selection acting on the mating cues and on the underlying genetic architecture. While the evolution of assortative mating with respect to locally adapted traits has been well-characterized, the evolution of disassortative mating is poorly characterized. Here we aim at understanding the evolution of disassortative mating for traits under strong balancing selection, by focusing on polymorphic mimicry as an illustrative example. Positive frequency-dependent selection exerted by predators generates local selection on wing patterns acting against rare variants and promoting local monomorphism. This acts across species boundaries, favouring Mullerian mimicry among defended species. In this well-characterized adaptive landscape, polymorphic mimicry is rare but is observed in a butterfly species, associated with polymorphic chromosomal inversions. Because inversions are often associated with recessive deleterious mutations, we hypothesize they may induce heterozygote advantage at the color pattern locus, putatively favoring the evolution of disassortative mating. To explore the conditions underlying the emergence of disassortative mating, we modeled both a trait locus (colour pattern for instance), subject to mutational load, and a preference locus. We confirm that heterozygote advantage favors the evolution of disassortative mating and show that disassortative mating is more likely to emerge if at least one allele at the trait locus is free from any recessive deleterious mutations. We modelled different possible genetic architectures underlying mate choice behaviour, such as self referencing alleles, or specific preference or rejection alleles. Our results showed that self referencing or rejection alleles linked to the color pattern locus can be under positive selection and enable the emergence of disassortative mating. However rejection alleles allow the emergence of disassortative mating only when the color pattern and preference loci are tightly linked. Our results therefore provide relevant predictions on both the selection regimes and the genetic architecture favoring the emergence of disassortative mating and a theoretical framework in which to interprete empirical data on mate preferences in wild populations.

Introduction

Mate preferences often play an important role in shaping trait diversity in natural populations, but the mechanisms responsible for their emergence often remain to be characterized. While the evolution of assortative mating on locally adapted traits is relatively well understood (Otto et al., 2008; de Cara et al., 2008; Thibert-Plante and Gavrilets, 2013), the selective forces involved in the evolution of disassortative mating are still largely unknown. Disassortative mating, *i.e.* the preferential mating between individuals displaying different phenotypes, is a rare form of mate preference (Jiang et al., 2013). In populations where individuals tend to mate with phenotypically distinct partners, individuals with a rare phenotype have a larger number of available mates, resulting in a higher reproductive success. By generating negative frequency-dependent selection on mating cues, disassortative mating is often regarded as a process generating and/or maintaining polymorphism within populations. Obligate disassortative mating leads to the persistence of intermediate frequencies of sexes or mating types (Wright, 1939), and promotes polymorphism (e.g. the extreme case of some Basidiomycete fungi where thousands of mating types are maintained (Casselton, 2002)). Disassortative mating can be based on different traits. Disassortative mating based on odors is known to operate in mice (Penn and Potts, 1999) and humans (Wedekind et al., 1995). Odor profiles are associated with genotype at the MHC loci affecting the immune response, known to be under strong balancing selection (Piertney and Oliver, 2006). Balancing selection on MHC alleles partly stems from heterozygous advantage, whereby heterozygous genotypes might confer an ability to recognize a larger range of pathogens. Such heterozygote advantage may promote the evolution of disassortative mating (Tregenza and Wedell, 2000). Extreme examples of heterozygote advantage are observed for loci with reduced homozygote survival. In the seaweed fly *Coelopa frigida* heterozygotes ($\alpha\beta$) at the locus Adh have a higher fitness than homozygotes ($\alpha\alpha$ or $\beta\beta$) (Butlin et al., 1984; Mérot et al., 2019) and females prefer males with a genotype that differs from their own (Day and Butlin, 1987). In the white-throated sparrow *Zonotrichia albicollis*, strong disassortative mating is known to operate with respect to

the color of the head stripe and associated with chromosomal dimorphism (Throneycroft, 1975). This plumage dimorphism is associated with a spectacular chromosomal polymorphism (Tuttle et al., 2016), with a complete lack of homozygous individuals for the rearranged chromosome (Horton et al., 2013).

While the fitness advantage of disassortative mating targeting loci with overdominance seems straightforward, the genetic basis of disassortative preferences remains largely unknown. One exception is the self-incompatibility system in *Brassicaceae* where the S-locus determines a specific rejection of incompatible pollens (Hiscock and McInnis, 2003). S-haplotypes contain 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, preventing fertilization from self-incompatible pollen due to specific receptor-ligand interactions. Self-rejection has also been proposed as an explanation for the disassortative mating associated with 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 related to the level of HLA similarity, rather than to a particular HLA genotype (Wedekind and Füri, 1997). In the white-throated sparrow, disassortative mating results from specific preferences for color plumage that differ between males and females; *tan*-striped males are preferred by all females while *white*-striped females are preferred by all males (Houtman and Falls, 1994). Different mechanisms leading to mate preferences and associated genetic architecture can be hypothesized, that may involve the phenotype of the chooser. Based on the categories described by Kopp et al. (2018), we assume that disassortative mating can emerge from two main mechanisms. (1) *Self-referencing*, when an individual uses its own signal to choose its mate, which may generate a disassortative mating that depends on the phenotypes of both the choosing and the chosen partners. (2) Preferences for or rejection of a given phenotype in the available partners (*recognition/trait* hypothesis), independently from the phenotype of the choosing partner, may also enable the emergence of disassortative mate preferences. These two mechanisms could involve a two locus architecture where one locus controls the mating cue and the other one the preference towards the different cues (Kopp et al., 2018). 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 loci are linked (Kopp et al., 2018).

Here we explore the evolutionary forces leading to the emergence of disassortative mating. We use as a model system the specific case of the butterfly species *Heliconius numata*, where high polymorphism in wing pattern is maintained within populations (Joron et al., 1999) and strong disassortative mating operates between wing pattern forms (Chouteau et al., 2017). *H. numata* butterflies are chemically-defended (Arias et al., 2016; Chouteau et al., 2019), and their wing patterns act as warning signals against predators (Chouteau et al., 2016a). A a local scale, natural selection on local mimicry usually leads to the fixation of a single warning signal shared by multiple defended species (Müllerian mimicry) (Mallet and Barton, 1989). However, local polymorphism of mimetic color patterns is maintained in certain species for instance under a balance between migration and local selection on mimicry (Joron and Iwasa, 2005). Yet, the level of polymorphism observed within populations of *H. numata* (Joron et al., 1999) would require that the strong local selection is balanced by a very high migration rate. However, disassortative mating based on wing pattern operates in *H. numata*, with females rejecting males displaying the same color pattern (Chouteau et al., 2017). Such disassortative mating could enhance local polymorphism in color pattern within this species. Nevertheless, the mode of evolution of a disassortative mating is unclear, notably because preferences for dissimilar mates should not be favoured if natural selection by predators on adult wing pattern acts against rare morphs (Chouteau et al., 2016b). Building on this well-documented case study, we use a theoretical approach to provide general predictions on the evolution of disassortative mating in polymorphic traits, and on expected genetic architecture underlying this behavior.

Variation in wing color pattern in *H. numata* is controlled by a single genomic region, called the supergene P (Joron et al., 2006), displaying dictinct chromosomal inversions combinations, each associated with a distinct mimetic phenotype (Joron et al., 2011) . These inversions have

recently been shown to be associated with a significant genetic load, resulting in a strong heterozygote advantage (Jay et al., 2019). We thus investigate whether a genetic load associated with locally adaptive alleles may favor the evolution of mate preference and promote local polymorphism. We then explore two putative genetic architectures for mate preferences based on (1) *self referencing* and (2) based on a *recognition/trait* rule, and test for their respective impacts on the evolution of disassortative mating. Under both hypotheses, we assumed that the mating cue and the mating preference were controlled by two distinct loci, and investigate the effect of linkage between loci on the evolution of disassortative mating.

Methods

Model overview

Based on earlier models of Müllerian mimicry (Joron and Iwasa, 2005; Llaurens et al., 2013), we describe the evolution of mate preferences based on color pattern using ordinary differential equations (ODE). We track the density of individuals carrying different genotypes combining the alleles at the locus P controlling mimetic color pattern and at the locus M underlying sexual preference. We assume a diploid species, so that each genotype contains four alleles.

The set of all possible four-allele genotypes is defined as $\mathcal{G} = \mathcal{A}_P \times \mathcal{A}_P \times \mathcal{A}_M \times \mathcal{A}_M$ where $\mathcal{A}_P$, $\mathcal{A}_M$ are the set of alleles at locus P and M respectively. A given genotype is then an quadruplet of the form (p_m, p_f, m_m, m_f) with $p_m \in \mathcal{A}_P$ and $m_m \in \mathcal{A}_M$ (resp. p_f and m_f) being the alleles at loci P and M on the maternal (resp. paternal) chromosomes. A recombination rate ρ between the color pattern locus P and the preference locus M is assumed.

We consider two geographic patches numbered 1 and 2 where those genotypes can occur. For all $(i, n) \in \mathcal{G} \times \{1, 2\}$ we track down the density of individuals of each genotype i within each patch n, $N_{i,n}$ trough time. Following previous models, polymorphism in mimetic color pattern is maintained within each of the two patches, by a balance between (1) local selection on color pattern in opposite directions in the two patches and (2) migration between patches.

The evolution of genotype densities through time, for each patch, is influenced by predation, mortality, migration between patches and reproduction, following the general equations :

$$\forall (i,n) \in \mathcal{G} \times \{1,2\} \quad \frac{d}{dt} N_{i,n} = Pred_{i,n} + Mort_{i,n} + Mig_{i,n} + Rep_{i,n}, \tag{1}$$

where $Pred_{i,n}$, $Rep_{i,n}$, $Mig_{i,n}$, and $Mort_{i,n}$ described the respective contributions of these four processes to the change in density of genotype i within each patch n. The computation of each of these four contributions is detailed in specific sections below. All variables and parameters are summarized in Table 1 and 2 respectively.

Since our ODE model describes the change in genotype densities at a population level, this amounts to considering that predation, migration, reproduction and survival occur simultaneously (see Equation (1)). In a large population, we can assume that predation, migration, reproduction and survival indeed occur in different individuals at the same time. Such a model implies that generations are overlapping and that there is no explicit ontogenic development: each newborn individual instantaneously behaves as an adult individual and can immediately migrate and reproduce. Our deterministic model provides general predictions while ignoring the effects of stochastic processes such as genetic drift.

Mimetic color pattern alleles at locus P

At the color pattern locus P, three alleles are assumed to segregate, namely alleles a, b and c, encoding for phenotypes A, B and C respectively. The set of alleles at locus P is then $\mathcal{A}_P = \{a, b, c\}$. We assume 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* (Le Poul et al., 2014) and in other supergenes (Wang et al., 2013; Tuttle et al., 2016; Küpper et al., 2016). The three color pattern phenotypes are assumed to be perceived as categorically different by both mating partners and predators. We note CP the function translating each genotype i into the corresponding color pattern phenotype $\mathcal{G}$. For example, for all $(m_m, m_f) \in \mathcal{A}_M \times \mathcal{A}_M$, $CP((a, b, m_m, m_f)) = A$ because allele a in dominant over b and the color pattern phenotype

depends only on alleles at locus P. Each color pattern allele is also assumed to carry an individual genetic load expressed when homozygous.

Preference alleles at locus P

We investigate the evolution of mate preference associated with color patterns, exploring in particular the conditions enabling the evolution of disassortative mating. We assume a single choosy sex: only females can express preferences toward male phenotypes, while males have no preference and can mate with any accepting females. Female preferences toward males displaying different color patterns are controlled by the locus M. We assume two different models of genetic architecture underlying mate preferences: alleles at locus M determine either (1) a preference toward similar or dissimilar phenotypes, which therefore also depends on the phenotype of the choosing individual, following the *self-referencing* hypothesis or (2) a preference toward a given color pattern displayed by the mating partner, independent of the color pattern of the choosing individual, following the *recognition/trait* hypothesis.

Predation

The probability of predation on individuals depends on their mimetic color patterns controlled by the locus P. Predation is determined in our model by a basic (patch-specific) effect of the local community of prey favouring one of the wing patterns locally (local adaptation through mimicry), itself modulated by positive frequency dependence of the different wing patterns controlled by P, within the focal species population. This is detailed below.

Divergent local adaptation in color pattern

Local selection exerted by predators promotes convergent evolution of wing color patterns among defended species (*i.e.* Müllerian mimicry, (Müller, 1879)), forming so-called mimicry rings composed of individuals from different species displaying the same warning signal within a locality.

Mimicry toward the local community of defended prey therefore generates strong local selection on color pattern and the direction of this selection then varies across localities (Sherratt, 2006).

Here we assume two separate populations exchanging migrants of an unpalatable species involved in Müllerian mimicry with other chemically-defended species. Local communities of species involved in mimicry (*i.e.* mimicry rings) differ across localities. We consider two patches occupied by different mimetic communities: population 1 is located in a patch where the local community (*i.e.* other chemically-defended species, not including *H. numata*) mostly displays phenotype A, and population 2 in a patch where the mimetic community mostly displays phenotype B. This spatial variation in mimicry rings therefore generates a divergent selection favouring distinct locally adapted phenotypes. Note that the allele c, and corresponding phenotype C is non-mimetic in both patches and at a disadvantage in both patches. Every individual of the focal (polymorphic) species is exposed to a predation risk modulated by its resemblance to the local mimetic community of butterflies. Each genotype i in population n (with $(i,n) \in \mathcal{G} \times \{1,2\}$) suffers from a basic predation mortality factor $d_{i,n}$. This parameter is lower for individuals displaying the phenotype mimetic to the local community (*i.e.* the phenotype A in population 1 and B in population 2). Individuals displaying phenotype C being non-mimetic in both patches, suffer from a high predation risk in both patches.

Here, to simplify, we consider that this basic mortality factor takes the value d_m for the locally mimetic phenotype (A in patch 1, B in patch 2), and $d_{n\text{-}m}$ for the locally non-mimetic phenotypes (B and C in patch 1, A and C in patch 2). We therefore introduce parameters $d_{n\text{-}m}$ and d_m, with $d_{n\text{-}m} > d_m$, as follows: the basic predation mortality factors for individuals not displaying and displaying the same color pattern as the local community respectively. For $i \in \mathcal{G}$, the basic predation mortality factors of individuals with genotype i in patch 1 and 2 are

$$d_{i,1} = \mathbb{1}_{\{CP(i)=A\}} d_m + \mathbb{1}_{\{CP(i)\neq A\}} d_{n\text{-}m}, \tag{2}$$

$$d_{i,2} = \mathbb{1}_{\{CP(i)=B\}} d_m + \mathbb{1}_{\{CP(i)\neq B\}} d_{n\text{-}m}, \tag{3}$$

where $\mathbb{1}$ is the indicator function which return 1 if the condition under brace in true and 0 else.

Local positive frequency-dependent predation

Predation exerted on a given phenotype depends on its match to the local mimetic environment (described by the parameter $d_{i,n}$ for all $(i,n) \in \mathcal{G} \times \{1,2\}$, see previous paragraph), but also on its own abundance in the patch predators learn to associate warning patterns with chemical defense. This learning behavior generates positive frequency-dependent selection on color patterns (Chouteau et al., 2016b): displaying a widely shared color pattern decreases the risk of encountering a naive predator (Sherratt, 2006). Number-dependent predator avoidance in the focal species is assumed to depend on its unpalatability coefficient (λ) and on the density of each phenotype within the population: the protection gained by phenotypic resemblance is greater for higher values of the unpalatability coefficient λ. For $(i,n) \in \mathcal{G} \times \{1,2\}$, the change in the density of a genotype i in patch n due to predation thus takes into account both the spatial variation in mimetic communities (using $d_{i,n}$) modulated by the local frequency-dependent selection, and is thus described by the equation:

$$Pred_{i,n} = -\frac{d_{i,n} N_{i,n}}{1 + \lambda \sum_j \mathbb{1}_{\{CP(i)=CP(j)\}} N_{j,n}}, \tag{4}$$

where $\sum_j \mathbb{1}_{\{CP(i)=CP(j)\}} N_{j,n}$ is the total density, within patch n, of individuals sharing the same color pattern as individuals of genotype i.

Mortality

We assume a baseline mortality rate δ. The recessive genetic loads δ_a, δ_b, δ_c associated with the respective alleles a, b and c limit the survival probabilities of homozygous genotypes at locus P.

For $i = (p_m, p_f, m_m, m_f) \in \mathcal{G}, n \in \{1,2\}$ the change in density of individuals with genotype i in patch n is given by

$$Mort_{i,n} = -(\delta + (\mathbb{1}_{\{p_m=p_f=a\}}\delta_a + \mathbb{1}_{\{p_m=p_f=b\}}\delta_b + \mathbb{1}_{\{p_m=p_f=c\}}\delta_c))N_{i,n}. \quad (5)$$

Migration

We assume a constant symmetrical migration rate *mig* corresponding to a proportion of individuals migrating from one patch to the other, as classically assumed in population genetics models (see for instance Holt (1985); Kuang and Takeuchi (1994); Joron and Iwasa (2005)). The number of individuals of each of the genotypes migrating to the other patch is therefore directly proportional to their density in their source population. For $(i, n, n') \in \mathcal{G} \times \{1, 2\} \times \{1, 2\}, n \neq n'$, the change in the density of individuals with genotype i in patch n due to migration between patches n and n' is given by the difference between the density of individuals coming into the patch $migN_{i,n'}$ and those leaving the patch $migN_{i,n}$:

$$Mig_{i,n} = migN_{i,n'} - migN_{i,n}. \quad (6)$$

where *mig* is the migration coefficient $mig \in [0, 1]$.

Reproduction

In the model, the reproduction term takes into account the basic demographic parameter, the effect of mate preference controlled by locus M and the fecundity limitations associated with choosiness.

Local demography

We assume that the populations from both patches have identical carrying capacity K and growth rate r. We name $N_{tot,n}$ the total density in patch n. The change in the total density due to reproduction is given by the logistic regulation function $r(1 - \frac{N_{tot,n}}{K})N_{tot,n}$. Thus for $(i, n) \in \mathcal{G} \times$

$\{1,2\}$, the change in the density of genotype i in patch n generated by sexual reproduction is given by:

$$Rep_{i,n} = r(1 - \frac{N_{tot,n}}{K})N_{tot,n}F_{i,n}, \tag{7}$$

where $(F_{i,n})_{i\in\mathcal{G}}$ are the frequencies of each genotype in the progeny $(F_{i,n})_{i\in\mathcal{G}}$. This frequencies depend on the behavior of the female, controlled by the preference locus M and on the availability of the preferred partners in the population, as detailed in the following section.

Mate preferences

During sexual reproduction, we assume that only one out of the two sexes expresses a mate preference, as often observed in sexual reproduction where females are usually choosier. Thus we assume females to be the choosy sex. The mate preference of female is then considered strict, implying that choosy individuals never mate with individuals displaying their non-preferred phenotype. Two hypothetical mate preference mechanisms are investigated.

Under the *self-referencing* hypothesis (hyp 1), three alleles are assumed at loci M, coding for (i) random mating (r), (ii) assortative mating *sim* and (iii) disassortative *dis*) respectively (see fig. S5 for more details, ($\mathcal{A}_M = \{r, sim, dis\}$). We assume that the *self-referencing* preference alleles *sim* and *dis* are dominant to the random mating allele r (see fig. S5 for more details). The dominance relationship between the *sim* and *dis* alleles is not specified however, because we never introduce these two alleles together. Note that under the *self-referencing* hypothesis (hyp. 1), mate choice depends not only on the color pattern of the male, but also on the phenotype of the female expressing the preference.

The alternative mechanism of mate preference investigated, assumes a specific recognition of color patterns acting as mating cue (*recognition/trait*, hyp. 2). Under hyp. 2, four alleles segregate at locus M: allele m_r, coding for an absence of color pattern recognition (leading to random mating behavior), and m_a, m_b and m_c coding for specific recognition of color pattern phenotypes A, B and C ($\mathcal{A}_M = \{m_r, m_a, m_b, m_c\}$). The *no preference* allele m_r is recessive to all the

preference alleles m_a, m_b and m_c, and preference alleles are co-dominant, so that females with heterozygous genotype at locus M may recognize two different color pattern phenotypes. Then, the recognition enabled by preference alleles m_a, m_b and m_c triggers either *attraction* (hyp. 2.a) or *rejection* (hyp. 2.b) toward the recognized color pattern, leading to assortative or disassortative mating depending on the genotype i of the female and the color pattern phenotype of the male (see figure S6 and S7 for more details).

Genotype frequencies in the progeny

We assume separate sexes and obligate sexual reproduction, and therefore compute explicitly the Mendelian segregation of alleles during reproduction, assuming a recombination rate ρ between the color pattern locus P and the preference locus M. We assume that the frequency of male and female of a given phenotype is the same. For $(i,n) \in \mathcal{G} \times \{1,2\}$, the frequency of genotype i in the progeny in patch n ($F_{i,n}$) then also depends on the frequencies of each genotype in the patch and on the mate preferences of females computed in equation (13). We introduce the preference coefficients $(Pref_{i,J})_{(i,J)\in\mathcal{G}\times\{A,B,C\}}$. These coefficients depend on the alleles at locus M as detailed in the next section. For $(i,J) \in \mathcal{G} \times \{A,B,C\}$ the preference coefficient $Pref_{i,J}$ is defined as $Pref_{i,J} = 1$ when females with genotype i accept males with phenotype J as mating partners and $Pref_{i,J} = 0$ otherwise.

For $i \in \mathcal{G}, n \in \{1,2\}$, we define $T_{i,n}$ as the probability that a female of genotype i in patch n accepts a male during a mating encounter (see (Otto et al., 2008)):

$$T_{i,n} = Pref_{i,A}P_{A,n} + Pref_{i,B}P_{B,n} + Pref_{i,C}P_{C,n}, \tag{8}$$

where for $J \in \{A,B,C\}$, $P_{J,n} = \frac{\sum_{i\in\mathcal{G}} N_{i,n}\mathbb{1}_{\{CP(i)=J\}}}{\sum_{i\mathcal{G}} N_{i,n}}$ denotes the frequency of phenotype J in patch n.

Because choosy individuals might have a reduced reproductive success due to limited mate availability (Kirkpatrick and Nuismer, 2004; Otto et al., 2008), we also assume a relative fitness cost associated with choosiness. This cost is modulated by the parameter c_r. When this cost is

absent ($c_r = 0$), females have access to a large quantity of potential mates, so that their mating rate is not limited when they become choosy ("Animal" model). When this cost is high ($c_r = 1$), females have access to a limited density of potential mates, so that their mating rate tends to decrease when they become choosy ("Plant" model). Intermediate values of c_r implies that females can partially recover the fitness loss due to the encountering of non-preferred males towards reproduction with other males. This cost of choosiness is known to limit the evolution of assortative mating (Otto et al., 2008) and may thus also limit the emergence of disassortative mating.

Following (Otto et al., 2008) we compute the mating rate $M_{i,n}$ of a female with genotype i in patch n :

$$M_{i,n} = 1 - c_r + c_r T_{i,n}. \tag{9}$$

We note $\overline{M}_n$ the average mating rate in patch n defined as

$$\overline{M}_n = \sum_i f_{i,n} M_{i,n}, \tag{10}$$

where for $(i,n) \in \mathcal{G} \times \{1,2\}$ f_i is the frequency of genotype i in patch n.

For $(j,k) \in \mathcal{G}^2$, the quantity

$$\frac{f_{j,n} M_{j,n}}{\overline{M}_n}, \tag{11}$$

is the probability that, given that a female has mated in the patch n, this female is of genotype j, and

$$\frac{Pref_{j,CP(k)} f_{k,n}}{T_{j,n}} = \frac{Pref_{j,CP(k)} f_{k,n}}{Pref_{j,A} P_{A,n} + Pref_{j,B} P_{B,n} + Pref_{j,C} P_{C,n}}, \tag{12}$$

is the probability that, given that a female of genotype j has mated in patch n, its mate is a male genotype k, depending on female preference and availability of males carrying genotype k.

For $(i,n) \in \mathcal{G} \times \{1,2\}$, the frequency of genotype i in the progeny of the population living in patch n is

$$F_{i,n} = \sum_{(j,k)\in\mathcal{G}^2} coef(i,j,k,\rho) \times \underbrace{\frac{f_{j,n}M_{j,n}}{\overline{M}_n}}_{\substack{\textit{probability that, given that}\\ \textit{a female has mated, this}\\ \textit{female is of genotype } j}} \times \underbrace{\frac{Pref_{j,CP(k)}f_{k,n}}{T_{j,n}}}_{\substack{\textit{probability that, given that a female}\\ \textit{of genotype } j \textit{ has mated,}\\ \textit{its mate is a male of genotype } k}}, \quad (13)$$

where $coef(i,j,k,\rho)$ controls the mendelian segregation of alleles during reproduction between an individual of genotype j and an individual of genotype k, depending on the recombination rate ρ between the color pattern locus P and the preference locus M (see Supp. S1 for detailed expression of $coef(i,j,k,\rho)$). We checked that for all n in $\{1,2\}$ the sum of $F_{i,n}$ over all i is always equal to one, as expected (see Supp. S2).

Model exploration

The complexity of this two-locus diploid model prevents comprehensive exploration with analytical methods, we therefore used numerical simulations to identify the conditions promoting the evolution of disassortative mating. All parameters and parameter intervals used in the different simulations are summarized in Table 2. The values of the basic predation mortality factor d_m and d_{n-m}, the unpalatability λ and migration rate *mig* are chosen as conditions maintaining balanced polymorphism at the color pattern locus P in the absence of disassortative mating, taken from (Joron and Iwasa, 2005).

Simulations are performed using Python v.3. and by using discrete time steps as an approximation (Euler method) (see Supp. S3 for more details about the numeric resolution). We checked that reducing the magnitude of the time step provided similar dynamics (see fig. S8), ensuring that our discrete-time simulations provide relevant outcomes. Note that all scripts used in this study are available on GitHub: https://github.com/Ludovic-Maisonneuve/Evolution_and_genetic_architecture_of_disassortative_mating.

Abbreviation	Description
$N_{i,n}$	Density of individuals with genotype i in patch n
$Pred_{i,n}$	Change in the density of individuals with genotype i caused by to predation
$Rep_{i,n}$	Change in the density of individuals with genotype i caused by to reproduction
$Mig_{i,n}$	Change in the density of individuals with genotype i caused by to migration
$Mort_{i,n}$	Change in the density of individuals with genotype i caused by to mortality
$CP(i)$	Color pattern phenotype of individuals with genotype i
$Pref_{i,J}$	Preference of individuals with genotype i towards individuals with phenotype J
$f_{i,n}$	Frequency of genotype i in patch n
$P_{I,n}$	Frequency of phenotype I in patch n
$T_{i,n}$	Probability that a female of genotype i in patch n accepts a male as mating partner during one mating encounter
$M_{i,n}$	Mating rate of females with genotype i in patch n
$\overline{M}_n$	Average female mating rate in patch n
$F_{i,n}$	Frequency of genotype i in the progeny of the population living in patch n
$P_{s\text{-}acc}$	Proportion of individuals expressing a self-accepting behavior
$P_{s\text{-}av}$	Proportion of individuals expressing a self-avoidance behavior

Table 1: **Description of variables used in the model.**

Abbreviation	Description	Parameter interval
$\mathcal{A}_P$	Set of all possible alleles at locus P	$\{a, b, c\}$
$\mathcal{A}_M$	Set of all possible alleles at locus M	$\{r, sim, dis\}$ (hyp. 1) $\{m_r, m_a, m_b, m_c\}$ (hyp. 2)
$\mathcal{G}$	Set of all possible genotypes	$\mathcal{A}_P \times \mathcal{A}_P \times \mathcal{A}_M \times \mathcal{A}_M$
$N^0_{tot,n}$	Initial population density in patch n	100
d_m	Basic predation mortality factor for individuals displaying the color pattern matching the local community	0.05
d_{n-m}	Basic predation mortality factor for individuals displaying a color pattern different from the local community	0.15
λ	Unpalatability coefficient	0.0002
mig	Migration rate	[0,1]
ρ	Recombination rate	[0, 0.5]
r	Growth rate	1
K	Carrying capacity within each patch	2000
δ	Baseline mortality rate	0.1
δ_i	Genetic load linked to allele i	[0, 1]
c_r	Relative cost of choosiness	[0, 1]

Table 2: **Description of parameters used in the model and range explored in simulations.**

Introduction of preference alleles

We assume that random mating is the ancestral preference behavior. Before introducing preference alleles, we therefore introduce color pattern alleles in equal proportions, and let the population evolves under random mating until the dynamical system reaches an equilibrium. We assume that a steady point is reached when the variation of genotype frequencies in the numerical solution during one time unit is below 10^{-5} (see Supp. S4 for more details). At this steady state, we then introduce the preference allele *dis* in proportion 0.01 (when exploring hyp. 1) or the preference alleles m_a, m_b, m_c in proportion $\frac{0.01}{3}$ (when exploring hyp. 2).

After the introduction of preference alleles, we follow the evolution of disassortative mating and its consequences in the two populations:

- Early dynamic : First, we identify the range of parameters enabling the emergence of disassortative mating, by tracking genotype numbers during the first 100 time steps after the introduction of preference alleles.

- Steady state : Then, we study the long-term evolutionary outcome associated with the changes in mating behavior, by computing genotype numbers at equilibrium, *i.e.* by running simulations until the variation of genotype frequency during one time unit is below 10^{-5} (see Supp. 4 for more details).

Summary statistics

To facilitate the interpretation of our results, we compute a number of summary statistics from the outcomes of our simulations. We define haplotypes as the pairs of alleles in $\mathcal{A}_P \times \mathcal{A}_M$ containing two alleles located on the same chromosome or inherited from the same parent. We then calculate haplotype frequencies in patch n $(f^{haplo}_{p,m,n})_{(p,m)\in\mathcal{A}_P\times\mathcal{A}_M}$ for $n \in \{1,2\}$. Then for $(p,m,n) \in \mathcal{A}_P \times \mathcal{A}_M \times \{1,2\}$, the frequency of haplotype (p,m) in patch n is given by:

$$f^{haplo}_{p,m,n} = \frac{\sum_{i=(p_m,p_f,m_m,m_f)\in\mathcal{G}} N_{i,n}\left(\frac{1}{2}\mathbb{1}_{\{p_m=p\}}\mathbb{1}_{\{m_m=m\}} + \frac{1}{2}\mathbb{1}_{\{p_f=p\}}\mathbb{1}_{\{m_f=m\}}\right)}{\sum_{i=(p_m,p_f,m_m,m_f)\mathcal{G}} N_{i,n}}. \tag{14}$$

The estimation of haplotype frequencies allows to characterize the association between color pattern alleles and preference alleles, leading to different mating behaviors among partners with different color patterns, specifically under the *recognition/trait* hypothesis (Hyp.2). To characterize female mating preferences generated by the different genotypes at locus M and the link with their own color pattern phenotype, we then distinguish two main behaviors emerging under hyp. 2 (fig. S6 and S7) 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 pattern phenotype.

In order to compare the mating behaviors observed under *self-referencing* (hyp. 1) *attraction* (hyp. 2.a) and *rejection* (hyp. 2.b) hypotheses, we compute population statistics, $P_{s\text{-}acc}$ (see equation (15)) and $P_{s\text{-}av}$ (see equation (16)) as the proportion of individuals exhibiting respectively a self-acceptance or a self-avoidance behavior throughout both patches. These two inferred behaviors can be directly compared with mate preferences empirically estimated. For example, in experiments where females can choose partners among males displaying different color patterns (Chouteau et al., 2017), the proportion of females mating with males displaying their own phenotype color pattern can be easily scored and compared to the proportion of self-accepting individuals computed in our model.

$$P_{s\text{-}acc} = \sum_{i \in \mathcal{G}} f_i Pref_{i,CP(i)}, \tag{15}$$

$$P_{s\text{-}av} = \sum_{i \in \mathcal{G}} f_i (1 - Pref_{i,CP(i)}). \tag{16}$$

Results

Effect of mate choice on polymorphism

The emergence of disassortative mating requires initial polymorphism at the trait used as mating cue. Because the costs associated with mate searching and courting penalize females preferring rare phenotypes, the distribution of color pattern variation in the population may be an important condition for the emergence of disassortative mating. In turn, the evolution of disassortative mating is likely to generate a positive selection on rare phenotypes, therefore enhancing polymorphism at the color pattern locus P. To disentangle the feedbacks between polymorphism of the cue and evolution of disassortative mating, we first investigate the impact of different mating behaviors on the distribution of color pattern phenotypes within populations.

Under random mating, the frequencies of color pattern alleles at equilibrium computed for different migration rates *mig* show that polymorphism can be maintained through an equilibrium between spatially heterogeneous selection and migration (fig.1 (a)), consistent with previous results from the literature (Joron and Iwasa, 2005). In the absence of migration however, phenotypes A and B are fixed in the populations living in patch 1 and 2 respectively, owing to their mimetic advantage within their respective communities. Polymorphism with persistence of phenotypes A and B within each population can only be maintained with migration, but in all cases the non-mimetic phenotype C is not maintained in any of the two populations (fig.1 (a)).

To test for an effect of mate choice on this selection/migration equilibrium, we then compare those simulations assuming random mating (*i.e.* with preference alleles r) with simulations where *self-referencing* preference alleles generating either assortative (*sim* allele) or disassortative (*dis* allele) behavior were introduced at the mate choice locus M (hyp. 1), assumed to be fully linked to the color pattern locus P ($\rho = 0$). Assuming assortative mating via *self-referencing* (hyp. 1) the results are similar to whose observed under random mating (fig.1 (a),(b)). Nevertheless, the proportion of locally adapted alleles is higher than under random mating because assortative

mating reinforces positive frequency dependent selection on those alleles. In contrast, disassortative mating maintains a higher degree of polymorphism, with the two mimetic phenotypes A and B and the non-mimetic phenotype C persisting within both populations, for all migration rates (fig.1 (c)). The non-mimetic phenotype C is rarely expressed because allele c is recessive. Nevertheless, individuals displaying phenotype C benefit from a high reproductive success caused by disassortative mating. Indeed, the strict disassortative preference assumed here strongly increases the reproductive success of individuals displaying a rare phenotype such as C. Negative frequency-dependent selection (FDS hereafter) on color pattern thus generated by disassortative mating counteracts the positive FDS due to predator behavior acting on the same trait. Therefore, disassortative mate preferences can strongly promote polymorphism within the two populations living in patch 1 and 2 respectively. When polymorphism is high, the cost of finding a dissimilar mate should be reduced, therefore selection acting against disassortative preferences be reduced. Our results thus highlight the decrease in the cost of finding a dissimilar mate when disassortative mating becomes established.

Linked genetic load favors the persistence of maladaptive alleles

In the following simulations, the migration parameter *mig* is set to 0.1, to allow for the persistence of polymorphism of color pattern phenotype A and B when assuming random mating. We then investigate the influence of a genetic load associated with the different color pattern alleles on polymorphism at the color pattern locus P, under random mating. This allows quantifying the effect of heterozygote advantage, independently of the evolution of mating preferences. We observe that the non-mimetic phenotype C is maintained together with phenotypes A and B within both populations, when (i) all three alleles carry a genetic load of similar strength, *i.e.* $\delta_a = \delta_b = \delta_c > 0$ or (ii) when allele c is the only one without any associated genetic load ($\delta_a = \delta_b > 0$ and $\delta_c = 0$) (fig. S9). In contrast, phenotype C is not maintained when a genetic load is associated with the non mimetic allele c only ($\delta_a = \delta_b = 0$ and $\delta_c > 0$), or when this load is stronger than the one associated with alleles a and b (fig. S9). The heterozygote advantage

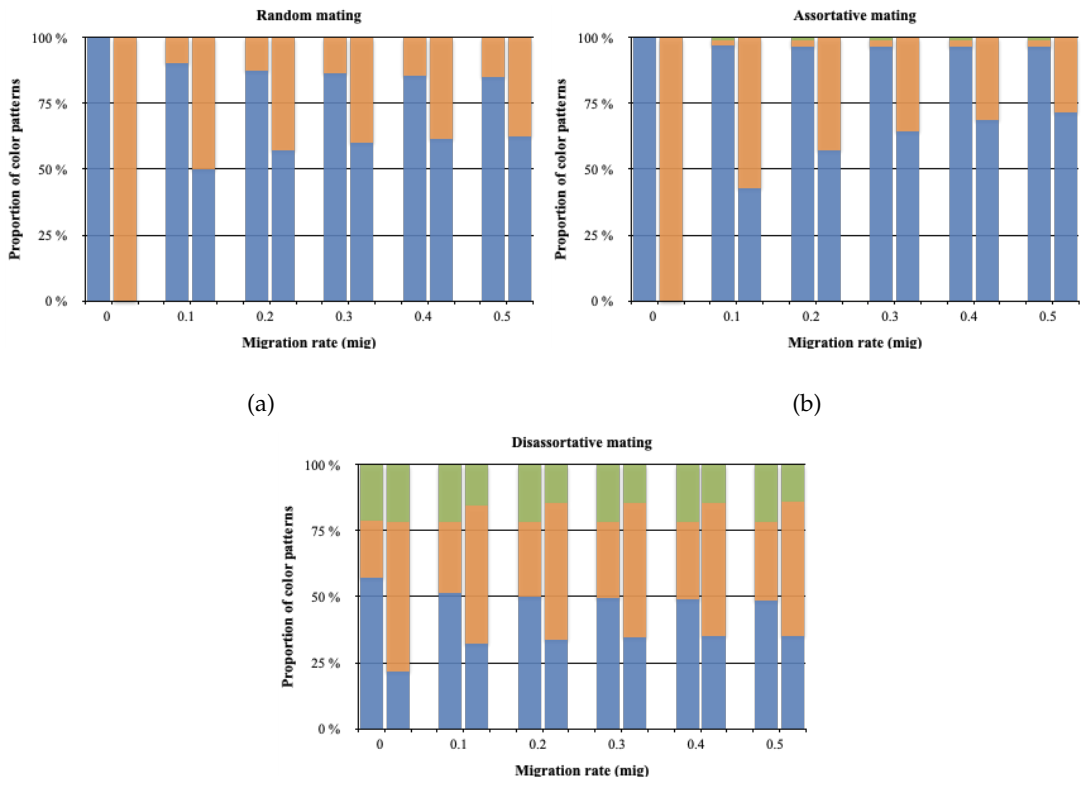


Figure 1: **Influence of mate preferences on color pattern diversity within both patches**. The equilibrium frequencies of color pattern phenotypes in patches 1 and 2 for different migration rates *mig* are computed assuming different mating behaviors, *i.e.* , random (a), assortative (b) or disassortative (c). The heights of the colored stacked bars indicate the frequencies of color pattern phenotypes A, B and C (blue, orange and green areas respectively) in patches 1 and 2 (on the left and right side respectively, for each migration level). The three alleles at the locus P controlling color pattern variations are introduced in proportion $\frac{1}{3}$ in each patch. The locus M controls for the *self-referencing* based mate preferences (hyp. 1): preferences alleles r, *sim* and *dis* were introduced in simulations shown in panel (a), (b) and (c) respectively. Simulations are run assuming $r = 1$, $K = 2000$, $N^0_{tot,1} = N^0_{tot,2} = 100$, $\lambda = 0.0002$, $d_m = 0.05$, $d_{n-m} = 0.15$, $\rho = 0$, $c_r = 0.1$, $\delta_a = \delta_b = \delta_c = 0$ and $\delta = 0.1$.

generated by genetic load associated with the dominant mimetic alleles at locus P therefore favors the persistence of a balanced polymorphism and more specifically promotes the maintenance of allele c in both patches, even though this allele does not bring any benefit through local (mimicry) adaptation.

Evolution of disassortative mating

Because we expect heterozygote advantage at the color pattern locus P to enhance the evolution of disassortative mating preferences at locus M, we first investigate the influence of a genetic load on the evolution of disassortative behavior by testing the invasion of *self-referencing* mutation triggering self-avoidance *dis* (hyp. 1) in a population initially performing random mating with genotype frequencies at equilibrium. We compute the frequency of mutants 100 time units after their introduction, assuming full linkage between loci P and M. Figure 2 shows that the genetic load associated with alleles a and b ($\delta_a = \delta_b$), has a strong positive impact on the emergence of disassortative mating. The genetic load associated with the recessive allele c (δ_c) has a weaker positive effect on the evolution of disassortative mating. Simulations assuming different relative cost of choosiness (c_r) show a similar effect of associated genetic loads (see fig. 2). However the cost of choosiness reduces the range of genetic load values allowing the emergence of disassortative preference. When this cost is high, the invasion of mutant allele *dis* is prevented, regardless of the strength of genetic load (see fig. 2(d)). Although an increased cost of choosiness slows down the invasion of the disassortative mating mutant *dis* (see fig. 2), a genetic load linked to the color pattern locus P generally favors the emergence of disassortative mating in both patches.

To investigate the long-term evolution of disassortative mating promoted by the genetic loads associated with color pattern alleles, we then compute the frequency of mutant allele *dis* at equilibrium in conditions previously shown to promote its emergence (*i.e.* assuming limited cost of choosiness). Figure 3 shows that the mutant preference allele *dis* is never fixed within populations. This suggests that the heterozygote advantage at locus P allowing the emergence of disassortative mating decreases when this behavior is common in the population. The *dis* mutant nevertheless reaches high frequencies when the genetic load associated with the recessive allele c is intermediate ($\delta_c \approx 0.35$) and that the genetic load associated with dominant alleles a and b is strong (see fig. 3). This result seems surprising because the highest level of disassortative mating is not reached when the genetic load is at the highest in all the three alleles at locus P. On the

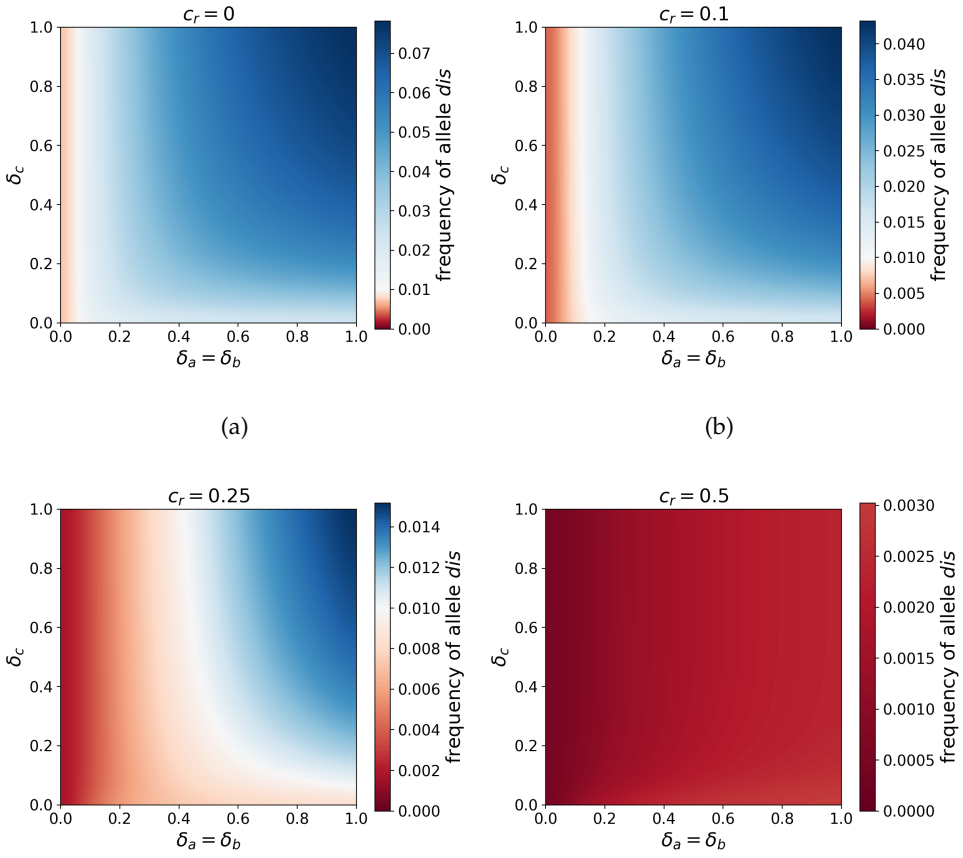


Figure 2: **Influence of a linked genetic load on the emergence of disassortative mating for different costs of choosiness, assuming *self-referencing* (hyp. 1).** The frequency of the mutant allele *dis* is shown 100 time units after its introduction depending on the strength of genetic load associated with the dominant alleles a and b ($\delta_a = \delta_b$) and to the recessive allele c, δ_c. The initial frequency of allele *dis* was 0.01, the area where mutant allele increase (resp. decrease) is shown in blue (resp. red). Simulations are run assuming either (a) no cost of choosiness $c_r = 0$, (b) a low cost of choosiness $c_r = 0.1$, (c) an intermediate cost of choosiness $c_r = 0.25$ or (d) an elevated cost of choosiness $c_r = 0.5$. Simulations are run assuming $r = 1$, $K = 2000$, $N^0_{tot,1} = N^0_{tot,2} = 100$, $\lambda = 0.0002$, $d_m = 0.05$, $d_{n-m} = 0.15$, $mig = 0.1$ and $\rho = 0$.

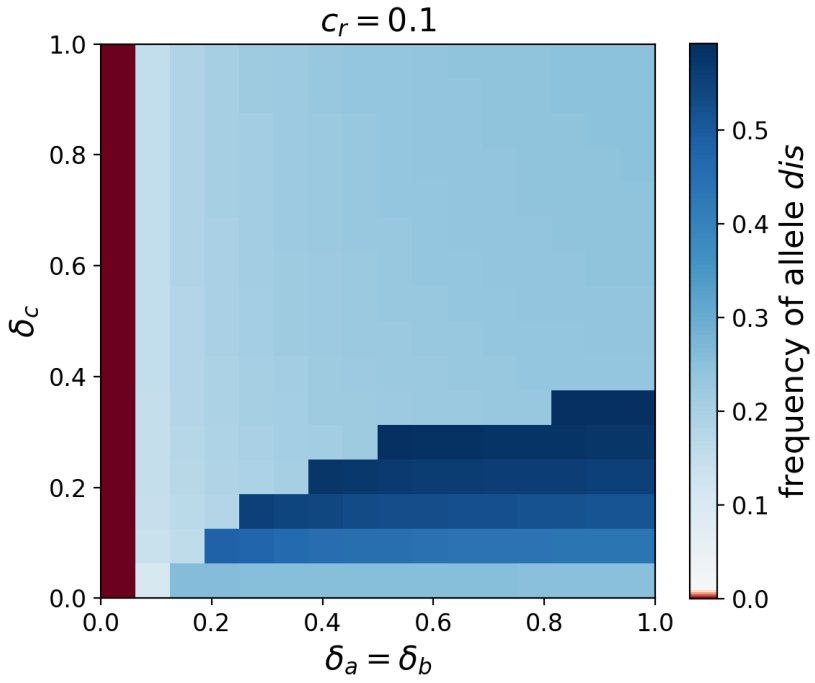


Figure 3: **Influence of a linked genetic load on the level of disassortative mating at equilibrium for low cost of choosiness ($c_r = 0.1$), assuming *self-referencing* (hyp. 1).** The frequency of the mutant allele *dis* is shown at equilibrium after its introduction depending on the strength of genetic load associated with the dominant alleles a and b ($\delta_a = \delta_b$) and with the recessive allele c, δ_c. The initial frequency of allele *dis* is 0.01. The area where the frequency of the mutant allele increases (resp. decrease) is shown in blue (resp. red). Simulations are run assuming $r = 1$, $K = 2000$, $N^0_{tot,1} = N^0_{tot,2} = 100$, $\lambda = 0.0002$, $d_m = 0.05$, $d_{n-m} = 0.15$, $mig = 0.1$, $\rho = 0$ and $c_r = 0.1$.

contrary, disassortative mating is favoured when a genetic load is associated with the dominant alleles only: disassortative mating limits more the cost of producing unfit offspring when a genetic load is associated with dominant alleles, because these alleles are always expressed as color pattern phenotypes, and therefore avoided by females with disassortative preferences.

How does the genetic architecture of mating preference influence the evolution of disassortative mating ?

To study the impact of the genetic architecture of mate preferences on the evolution of disassortative mating, we then compare the invasion of *self-referencing* alleles *dis* with the invasion of *recognition/trait* alleles (*i.e.* alleles m_r, m_a, m_b and m_c controlling random mating and specific recognition of phenotype A, B and C respectively, hyp. 2). We assume loci P and M to be fully linked ($\rho = 0$), and compare simulations where mate preference alleles trigger either disassortative preference (hyp. 1), *attraction* (hyp. 2.a) or *rejection* (hyp. 2.b) of the recognized color pattern phenotype. We report the frequencies of haplotypes, in order to follow the association of color pattern and preference alleles (fig.4(a), fig.4(b) and fig.4(c) respectively).

Under a *self-referencing* rule, alleles a and b are associated with preference allele *dis* as soon as the genetic load associated with the dominant alleles (alleles a and b) is greater than 0. Indeed disassortative mating favors the production of heterozygotes and reduces the expression of the genetic load in the offspring. In contrast, the non-mimetic allele c, not associated with any genetic load, is preferentially linked with the random mating allele r. This result is surprising because heterozygotes carrying a c allele have a lower predation risk than homozygotes with two c alleles, which are maladaptive in both patches. However, the benefit for allele c to be associated with disassortative preference may be weaker than expected. Because of the genetic load associated with the dominant color alleles a and b, the c allele is common in the population, so that this allele c often occurs at homozygous state. The alleles a and b are associated with disassortative preference, thus these alleles are likely to be at heterozygote state. Since the c allele is recessive, disassortative crosses between individuals with phenotype C and A or B may frequently produce progeny with half of the offspring carrying two c alleles, suffering from increased predation. The benefit of disassortative preference associated with allele c may be lower than the cost of choosiness, therefore favoring the association between the allele c and the random mating allele r.

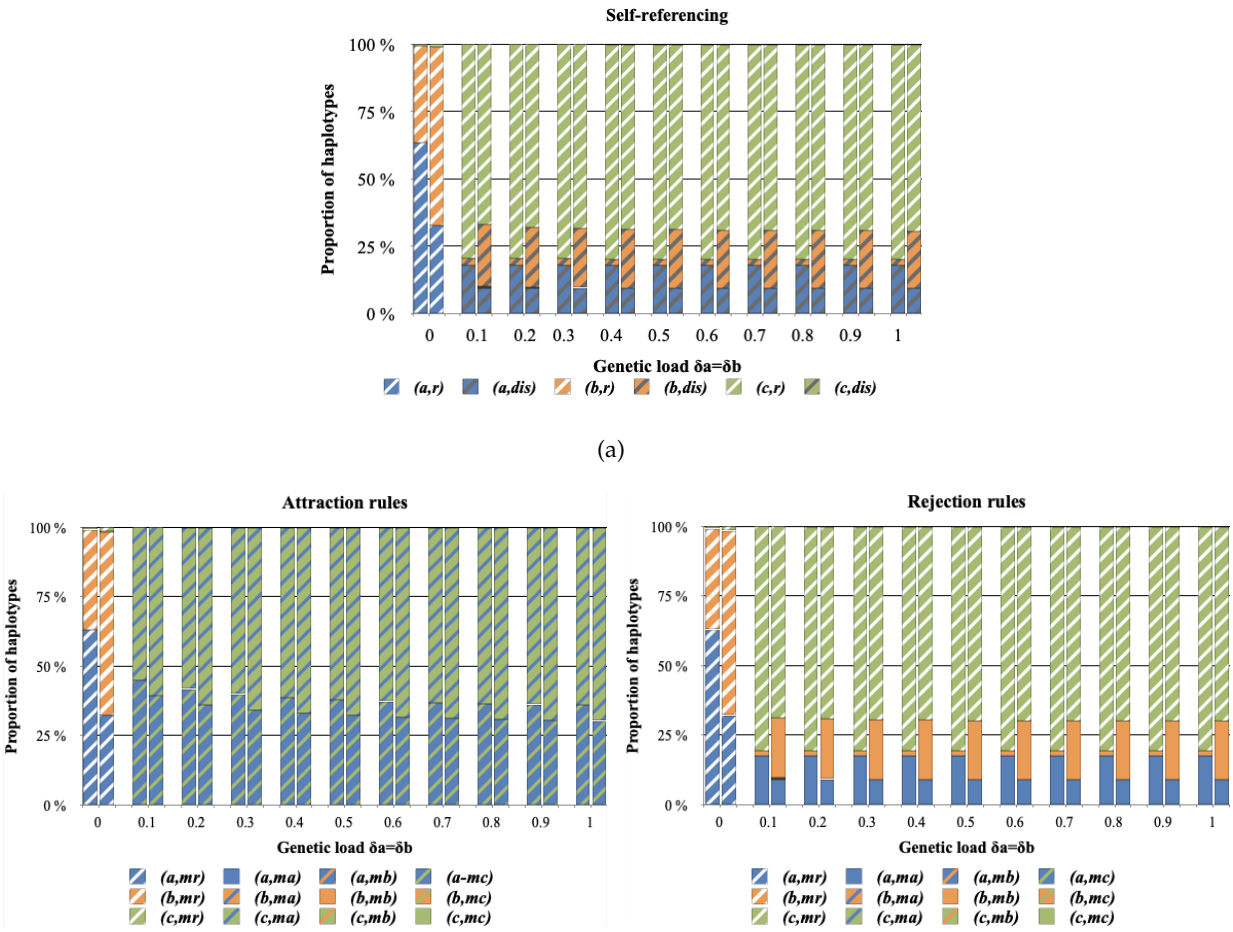


Figure 4: **Influence of a genetic load on haplotype diversity, assuming (a) *self-referencing* (hyp. 1), (b) *attraction* rule (hyp. 2.a) or (c) *rejection* rule (hyp. 2.b) at the preference locus (*recognition/trait*)**. The proportion of haplotypes at equilibrium after the introduction of preference alleles in both patches are shown for different values of genetic load associated with alleles a and b ($\delta_a = \delta_b$). For each value of genetic load ($\delta_a = \delta_b$) the first and second bars show the frequencies of haplotypes in the patches 1 and 2 respectively. Simulations are run assuming $r = 1$, $K = 2000$, $N^0_{tot,1} = N^0_{tot,2} = 100$, $\lambda = 0.0002$, $d_m = 0.05$, $d_{n-m} = 0.15$, $\rho = 0$, $mig = 0.1$, $\delta_c = 0$, $\delta = 0.1$ and $c_r = 0.1$.

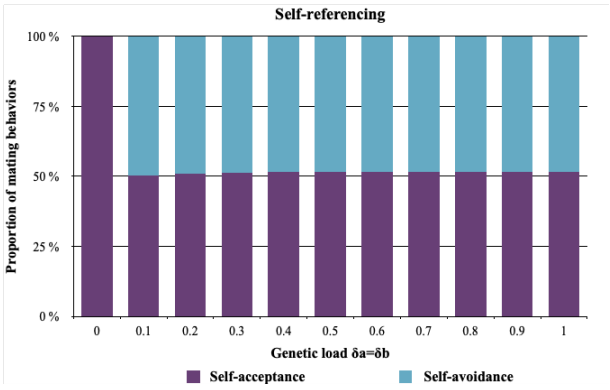
When preference alleles cause female attraction to males exhibiting a given phenotype (hyp. 2.a), only haplotypes (a, m_c) and (c, m_a) are maintained in both patches at equilibrium (fig.4(b)). The haplotype (a, m_c) benefits from both positive selection associated with mimicry and limited expression of the genetic load due to the preferential formation of heterozygotes. Haplotype

(c, m_a) is maintained because of the benefit associated with the choice of the most frequent mimetic phenotype A, and the limited expression of the non-mimetic phenotype C due to c being recessive. The proportion of haplotype (a, m_c) decreases as the genetic load associated with allele a increases. Indeed the mating between two individuals of genotype (a, c, m_c, m_a) becomes more likely and leads to the formation of individuals (a, a, m_c, m_c) suffering from the expression of the genetic load. The b allele is then lost mainly because of the dominance relationships between alleles a and b, making the phenotype A more commonly expressed and therefore favouring the haplotype (c, m_a) rather than (c, m_b). The sexual selection generated by the disassortative mating of individuals thus exerts a strong disadvantage against the b allele, provoking its elimination.

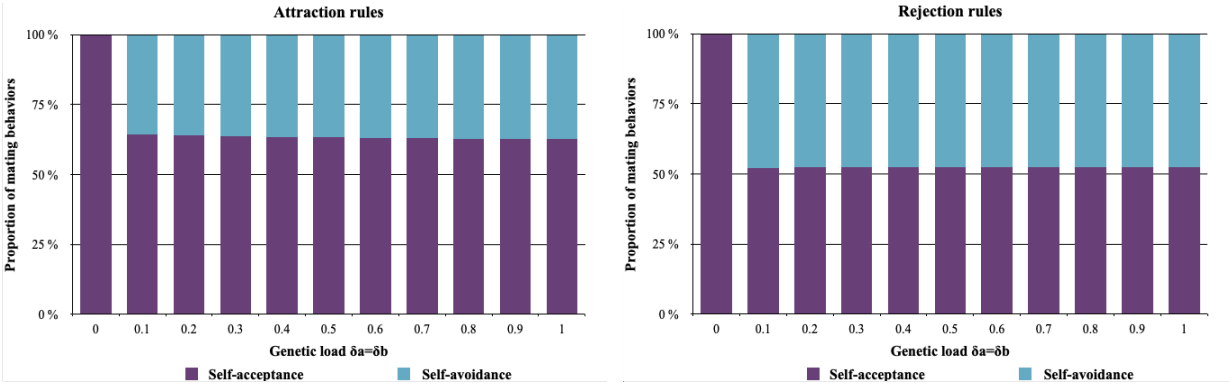
By contrast, when mate preference is based on alleles causing *rejection* behavior (hyp. 2.b) and when a genetic load is associated with the mimetic alleles a and b at locus P, these alleles become associated with the corresponding rejection alleles at locus M (*i.e.* (a, m_a) and (b, m_b) have an intermediate frequencies in both patches) (fig.4(c)). Non mimetic allele c becomes associated with random mating preference allele r. The three alleles (a, b and c) persist within patches for all positive values of genetic load. This contrasts with the evolutionary outcome observed under attraction rule (hyp. 2.a) where mimetic allele b is lost if the genetic load is greater than 0 (fig. 4(b)).

We then investigate how these haplotype frequencies translate into individual behaviors in the populations at equilibrium. As highlighted in fig.5, the proportion of each behavior depends more on the existence of a genetic load linked to dominant alleles, than on its strength. The proportion of disassortative mating is similar when assuming *self-referencing* (hyp. 1) and *recognition/trait* leading to rejection (hyp. 2.b) ($P_{s\text{-}av} \approx 48\%$) (fig.5(a) and5(c)).

By contrast, when we consider preference alleles leading to *attraction* (hyp. 2.a), the disassortative behavior is scarcer at equilibrium ($P_{s\text{-}av} \approx 36\%$) (fig. 5(b)). This may seem surprising given that most haplotypes are formed by a color pattern allele linked with an *attraction* allele for a different color pattern (fig. 4(b)). Nevertheless, the color pattern allele c is linked to m_a coding for attraction to A. As a consequence, most individuals formed are heterozygous at both the



(a)



(b) (c)

Figure 5: **Influence of a genetic load on the distribution of mating behavior observed at the population level, assuming (a) *self-referencing* (hyp. 1), (b) *attraction* rule (hyp. 2.a) or (c) *rejection* rule (hyp. 2.b) at the preference locus (*recognition/trait*)**. The proportion of individuals displaying self-acceptance $P_{s\text{-}acc}$ (in purple) and self-avoidance $P_{s\text{-}av}$ (in blue) obtained at equilibrium after the introduction of preference alleles are shown for different values of the level of genetic load of δ_a and δ_b. Simulations are run assuming $r = 1$, $K = 2000$, $N^0_{tot,1} = N^0_{tot,2} = 100$, $\lambda = 0.0002$, $d_m = 0.05$, $d_{n-m} = 0.15$, $\rho = 0$, $mig = 0.1$, $\delta_c = 0$, $\delta = 0.1$ and $c_r = 0.1$.

color pattern locus P (with one allele a and one allele c) and at the preference locus M (with one preference allele coding for attraction toward phenotype A and another preference allele triggering attraction toward phenotype C). These double heterozygotes thus benefit from mimicry and avoid the expression of deleterious mutations, and are self-accepting. However, under the *self-referencing* (hyp. 1) or *rejection* (hyp. 2.b) rules disassortative mating is more likely to emerge.

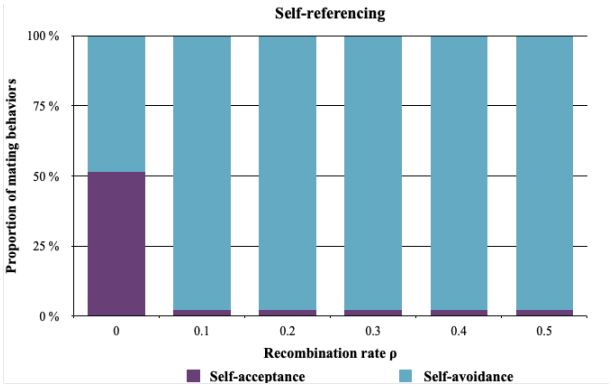
Indeed under hyp. 2.b, haplotypes composed by a phenotype allele and its corresponding preference allele ((a, m_a) for example) generally immediately translates into a self-avoiding behavior, whatever the genotypic combinations within individuals. Moreover under hyp. 1 disassortative haplotype, *i.e.* an haplotype where the preference allele is *dis*, always generates a disassortative behavior.

This highlights that the genetic architecture of mate preference plays a key role in the evolution of the mating behavior of diploid individuals: the evolution of disassortative haplotypes inducing disassortative preferences do not necessarily cause disassortative mating at the population level. At equilibrium, the proportion of self-avoidance behavior in the population hardly depends of the strength of the genetic load (figure 5). However, the strength of the genetic load does increase the speed of evolution of disassortative mating (see fig. S10 exploring the dynamics of invasion of self-avoiding behavior for different levels of genetic load), therefore suggesting stronger positive selection on disassortative mating when the genetic load associated with dominant wing color pattern alleles is higher.

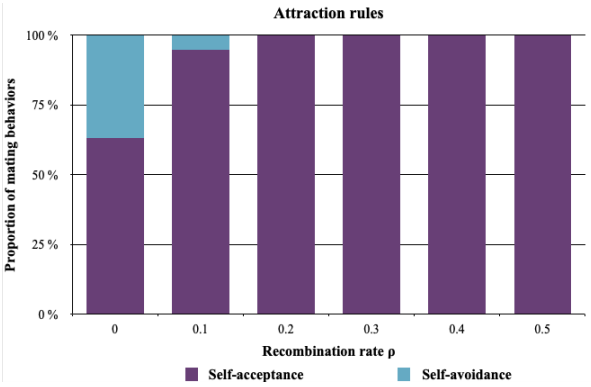
Impact of linkage between loci P and M on the evolution of disassortative mating

In previous sections, we observed that the genetic load associated with the two most dominant alleles at the color pattern locus P impacts the evolution of mate choice, and that disassortative mating is more prevalent under *self-referencing* (hyp. 1) or under *rejection* (hyp. 2.b) than under *attraction* (hyp. 2.a) rules, when the color pattern locus P and the preference locus M are fully linked. We then test for an 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.

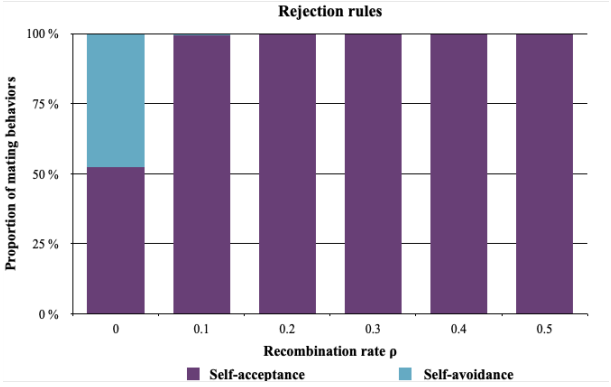
Assuming *self-referencing* (hyp. 1), increasing recombination rate strongly promotes the self-avoidance behavior ($P_{s\text{-}av} \approx 98\%$) (see fig. 6(a)). Selection generated by the genetic load associated to color pattern alleles a and b promotes their linkage with the disassortative *self-referencing* allele *dis*, while the genetic-load free allele c tends to be linked to the random mating allele r (as observed in simulations assuming no recombination, fig. S11(a)). Because the allele *dis* reaches



(a)



(b)



(c)

Figure 6: **Influence of the recombination rate between color pattern and preference alleles on the distribution of mating behavior observed at the population level, assuming different genetic architectures of mate preferences: either (a) *self-referencing* (hyp. 1), or *recognition/trait* leading to (b) *attraction* rule (hyp. 2.a) or (c) *rejection* rule (hyp. 2.b)**. The proportion of individuals displaying self-acceptance $P_{s\text{-}acc}$ (in purple) and self-avoidance $P_{s\text{-}av}$ (in blue) obtained at equilibrium are shown for different values of recombination rate ρ between the preference locus M and the color pattern locus P. Simulations are run assuming $r = 1$, $K = 2000$, $N^0_{tot,1} = N^0_{tot,2} = 100$, $\lambda = 0.0002$, $d_m = 0.05$, $d_{n-m} = 0.15$, $mig = 0.1$, $\delta_a = \delta_b = 0.5$, $\delta_c = 0$, $\delta = 0.1$ and $c_r = 0.1$.

a high frequency in the population, recombination generates a large density of recombinant haplotypes (a, r), (b, r), (c, dis). Haplotypes (a, r) and (b, r) are disfavored because they lead to a the production of offspring suffering from the expression of a genetic load, whereas (c, dis) leads to the production of viable offspring. Therefore, under the *self-referencing* hypothesis (hyp. 1), recombination thus significantly increases the proportion of disassortative mating.

Under *self-referencing* rule (hyp. 1), mate preference depends on the phenotype displayed by the individual, so that the allele *dis* always translates into a disassortative behavior. By contrast, when assuming *recognition/trait* for a given color pattern allele (hyp. 2), mating behavior depends only on the genotype at the preference locus M, independently from the color pattern of the female. We therefore expect a stronger effect of recombination rate on mate choice evolution. Figure 6 indeed confirms this prediction. Under *attraction* (hyp. 2.a) and *rejection* (hyp. 2.a) rules, the most striking effect is observed when comparing simulations assuming $\rho = 0$ *vs* $\rho > 0$: self-avoidance behavior is scarcely observed in the population ($P_{s\text{-}av} \approx 1\%$) when there is recombination ($\rho > 0$).

Our results suggest that disassortative mating can emerge under *self-referencing* rule or with a single-locus architecture (or tight linkage between the cue and the preferences) under *recognition/trait* rule. Nevertheless, in natural populations, strict *self-referencing* behaviour whereby manipulating of the phenotype of the chooser change its preference, is scarcely observed (see Kopp et al. (2018) for a review). Thus, we expect that disassortative mating might emerge when the mating cue and preferences have tightly linked genetic basis or are controlled by a single gene.

Discussion

Genetic architecture of disassortative mating: theoretical predictions

Our model shows that without recombination between color pattern (locus P) and preference alleles (locus M), disassortative mating is more likely to emerge when the genetic architecture is with *self-referencing* (hyp. 1) or with color pattern recognition triggering *rejection* (hyp. 2.b). When preference alleles cause *attraction* to males exhibiting a given phenotype (hyp. 2.a), heterozygote advantage favors haplotypes formed by a color pattern allele linked with an attraction allele for a different color pattern. However, these haplotypes do not necessarily imply a complete self-avoidance behavior in females carrying them. The co-dominance relationships assumed at

preference locus indeed generate preferences for two different phenotypes in heterozygotes at the locus M, favoring self-acceptance. This effect is reinforced by the mate choice, favoring the association between a color allele and the corresponding attraction allele in the offspring, increasing the emergence of self-accepting genotypes. This might explain the low proportion of self-avoidance behavior observed within populations, when assuming the *attraction* rule (hyp. 2.a). By contrast, when recombination between the two loci does occur, a *self-referencing* architecture (hyp. 1) may facilitate the evolution of disassortative mating. The genetic basis of disassortative mating is largely unknown in natural populations. Assortative mating is better documented, for instance in *Heliconius* butterflies where it is generally associated with attraction towards a specific cue. The locus controlling preference for yellow *vs.* white in *H. cydno* maps close to the gene *aristaless*, whose expression differences determine the white/yellow switch in this species (Kronforst et al., 2006; Westerman et al., 2018). In *H. melpomene*, a major QTL associated with preference towards red was identified in crosses between individuals displaying a red pattern and individuals with a white pattern (Merrill et al., 2019). This QTL is also located close to the gene *optix* involved in the variation of red patterning in *H. melpomene*. Assortative mating in *Heliconius* thus seems to rely on alleles encoding preference for specific cues, linked to with loci involved in the variation of these cues. Similarly, our model suggests that the genetic architecture of disassortative mating might involved tight linkage between the cue and the preference. However, our results show that gene coding for rejection toward certain cue are more likely to promote disassortative mating, in contrast with the attraction alleles documented in species where assortative mating behavior is observed.

Similar mate preference is obtained with some *recognition/trait* (hyp.2) genotypes than with some *self-referencing* (hyp. 1) genotypes: for example, under the *rejection* rule (hyp. 2.b), the genotype (a, a, m_a, m_a) leads to the same mate preference as the genotype (a, a, dis, dis) under the *self-referencing* genetic architecture. Introducing recombination in the *recognition/trait* architecture then enables the decoupling of the mating cue and of its corresponding preference alleles, thereby disrupting the self rejection behavior. Furthermore, under the *recognition/trait* architecture, our

model distinguishes whether the specific recognition of the cue leads to *rejection* or *attraction*, and highlights that these two hypotheses lead to the evolution of different preference regimes: disassortative mating is more likely to emerge assuming a *rejection rule*. This rule indeed generates a greater density of self-rejecting haplotypes than the *attraction* rule, although recombination limits this effect.

The dominance relationships assumed at both the cue and preference loci are likely to impact our predictions on the evolution of disassortative mating. Disassortative mating is advantaged when it favors the production of offspring free from the expression of genetic load, and dominance relationship at the color pattern locus P may hide the signalling of genetic load associated with recessive cue alleles. This explains why disassortative mating is favored when genetic load is low in recessive cue alleles and large in dominant cue alleles. The co-dominance assumed at the preference locus generates preferences toward two different phenotypes in heterozygotes at the preference locus and we suspect that alternative hypothesis may modulate our predictions on the evolution of disassortative mating.

Altogether, our theoretical model shows that the genetic basis of mate preferences has a strong impact on the evolution of disassortative mating at loci under heterozygote advantage. This emphasizes the need to characterize the genetic basis of mate preference empirically and the linkage disequilibrium with the locus controlling variation in the mating cues.

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

Here, we confirm that the evolution of disassortative mating is promoted by the heterozygote advantage associated with alleles determining the mating cue. As mentioned above, the phenotype of the chosen individuals depends on the dominance relationships at the color pattern locus. Our model highlights that a genetic load associated with the dominant alleles contributes more to disassortative mating than a genetic load associated with the most recessive haplotype. This

theoretical prediction is in accordance with the few documented cases of polymorphism promoted by disassortative mating. In the polymorphic butterfly *Heliconius numata* for instance, the top dominant haplotype *bicoloratus* is associated with a strong genetic load (Jay et al., 2019). Similarly, in the white throated sparrow, the dominant *white* allele is also associated with a significant genetic load (Tuttle et al., 2016). Again, in the self-incompatibility locus of the *Brassicaceae*, dominant haplotypes carry a higher genetic load than recessive haplotypes (Llaurens et al., 2009). Disassortative mating is beneficial because it increases the number of heterozygous offspring with higher fitness. Once disassortative mating is established within a population, recessive deleterious mutations associated with the dominant haplotype become sheltered because the formation of homozygotes carrying two dominant alleles is strongly reduced, thereby limiting the opportunities for purging via recombination (Llaurens et al., 2009). Falk and Li (1969) proved that disassortative mate choice promotes polymorphism, and therefore limits the loss of alleles under negative selection. Disassortative mating might thus shelter deleterious mutations linked to dominant alleles, and reinforce heterozygote advantage. The sheltering of deleterious mutations is favored by the interaction between two aspects of the genetic architecture: dominance at the mating cue locus and limited recombination. This is likely to happen in polymorphic traits involving chromosomal rearrangements, where recombination is limited. Many rearranged haplotypes are indeed associated with serious fitness reduction as homozygotes (Faria et al., 2019), such as in the derived haplotypes of the supergene controlling plumage and mate preferences in the white-throated sparrow (Thomas et al., 2008). The deleterious elements in the inverted segment can be due to an initial capture by the inversions (Kirkpatrick, 2010), but they could also accumulate through time, resulting in different series of deleterious mutations associated to inverted and non-inverted haplotypes (Berdan et al., 2019).

Here, we assume that mate choice relied purely on a single cue. Nevertheless, mate choice could be based on other cues, controlled by linked loci and enabling discrimination between homozygotes and heterozygotes, thereby further increasing the proportion of heterozygous offsprings with high fitness. We also modelled strict preferences regarding color patterns, but

choosiness might be less stringent in the wild, and may limit the evolution of disassortative mating. Depending on the cues and dominance relationships among haplotypes, different mate choice behaviors may also evolve, which might modulate the evolution of polymorphism within populations. Our model thus stresses the need to document dominance relationships among haplotypes segregating at polymorphic loci, as well as mate choice behavior and cues, to understand the evolutionary forces involved in the emergence of disassortative mating.

Conclusions

Inspired by a well-documented case of disassortative mating based on cues subject to natural selection, our model shows that ~~balancing selection promoting local polymorphism and~~ heterozygote advantage is likely to favor the evolution of disassortative mating preferences. We highlight that disassortative mating is more likely to emerge when loci code for self-referencing disassortative preference or rejection of specific cues. However rejection locus only promotes disassortative mating when they are in tight linkage with the locus controlling mating cue variation.

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. We are also grateful to Roger Butlin, Charles Mullon, Tom Van Dooren and two anonymous reviewers for their thoughtful review of a previous versions of this manuscript. We also thank Sylvain Gerber for his help in improving the English of this article. This work was supported by the Emergence program from Paris City Council to VL and the ANR grant SUPERGENE to VL and MJ.

References

M. Arias, A. Meichanetzoglou, M. Elias, N. Rosser, D. L. De-Silva, B. Nay, and V. Llaurens. Variation in cyanogenic compounds concentration within a heliconius butterfly community: does mimicry explain everything? BMC evolutionary biology, 16(1):272, 2016.

E. L. Berdan, A. Blanckaert, R. K. Butlin, and C. Bank. Muller's ratchet and the long-term fate of chromosomal inversions. bioRxiv, 2019. 10.1101/606012. URL https://www.biorxiv.org/content/early/2019/04/13/606012.

R. K. Butlin, P. M. Collins, and T. H. Day. The effect of larval density on an inversion polymorphism in the seaweed fly coelopa frigida. Heredity, 1984.

L. Casselton. Mate recognition in fungi. Heredity, 88(2):142, 2002.

M. Chouteau, M. Arias, and M. Joron. Warning signals are under positive frequency-dependent selection in nature. Proceedings of the National Academy of Sciences, 113(8):2164–2169, 2016a. ISSN 0027-8424. 10.1073/pnas.1519216113. URL https://www.pnas.org/content/113/8/2164.

M. Chouteau, M. Arias, and M. Joron. Warning signals are under positive frequency-dependent selection in nature. Proceedings of the national Academy of Sciences, 113(8):2164–2169, 2016b.

M. Chouteau, V. Llaurens, F. Piron-Prunier, and M. Joron. Polymorphism at a mimicry supergene maintained by opposing frequency-dependent selection pressures. Proceedings of the National Academy of Sciences, 114(31):8325–8329, 2017.

M. Chouteau, J. Dezeure, T. N. Sherratt, V. Llaurens, and M. Joron. Similar predator aversion for natural prey with diverse toxicity levels. Animal Behaviour, 153: 49 – 59, 2019. ISSN 0003-3472. https://doi.org/10.1016/j.anbehav.2019.04.017. URL http://www.sciencedirect.com/science/article/pii/S000334721930137X.

T. H. Day and Butlin. Non-random mating in natural populations of the seaweed fly, coelopa frigida. Heredity, 1987.

M. de Cara, N. Barton, and M. Kirkpatrick. A model for the evolution of assortative mating. The American Naturalist, 171(5):580–596, 2008. 10.1086/587062. URL https://doi.org/10.1086/587062. PMID: 18419568.

C. T. Falk and C. C. Li. Negative assortative mating: Exact solution to a simple model. Genetics, 1969.

R. Faria, K. Johannesson, R. K. Butlin, and A. M. Westram. Evolving inversions. Trends in ecology & evolution, 2019.

S. J. Hiscock and S. M. McInnis. Pollen recognition and rejection during the sporophytic self-incompatibility response: Brassica and beyond. Trends in plant science, 8(12):606–613, 2003.

R. D. Holt. Population dynamics in two-patch environments: Some anomalous consequences of an optimal habitat distribution. Theoretical Population Biology, 28(2): 181 – 208, 1985. ISSN 0040-5809. https://doi.org/10.1016/0040-5809(85)90027-9. URL http://www.sciencedirect.com/science/article/pii/0040580985900279.

B. M. Horton, Y. Hu, C. L. Martin, B. P. Bunke, B. S. Matthews, I. T. Moore, J. W. Thomas, and D. L. Maney. Behavioral characterization of a white-throated sparrow homozygous for the zal2 m chromosomal rearrangement. Behavior genetics, 43(1):60–70, 2013.

A. M. Houtman and J. B. Falls. Negative assortative mating in the white-throated sparrow, zonotrichia albicollis: the role of mate choice and intra-sexual competition. Animal Behaviour, 48(2):377–383, 1994.

P. Jay, M. Chouteau, A. Whibley, H. Bastide, V. Llaurens, H. Parrinello, and M. Joron. Mutation accumulation in chromosomal inversions maintains wing pat-

tern polymorphism in a butterfly. bioRxiv, 2019. 10.1101/736504. URL https://www.biorxiv.org/content/early/2019/08/15/736504.

Y. Jiang, D. I. Bolnick, and M. Kirkpatrick. Assortative mating in animals. The American Naturalist, 181(6):E125–E138, 2013. 10.1086/670160. URL https://doi.org/10.1086/670160. PMID: 23669548.

M. Joron and Y. Iwasa. The evolution of a müllerian mimic in a spatially distributed community. Journal of Theoretical Biology, 237(1):87–103, 2005.

M. Joron, I. R. Wynne, G. Lamas, and J. Mallet. Variable selection and the coexistence of multiple mimetic forms of the butterfly heliconius numata. Evolutionary Ecology, 13(7-8):721–754, 1999.

M. Joron, R. Papa, M. Beltrán, N. Chamberlain, J. Mavárez, S. Baxter, M. Abanto, E. Bermingham, S. J. Humphray, J. Rogers, et al. A conserved supergene locus controls colour pattern diversity in heliconius butterflies. PLoS biology, 4(10):e303, 2006.

M. Joron, L. Frezal, R. T. Jones, N. L. Chamberlain, S. F. Lee, C. R. Haag, A. Whibley, M. Becuwe, S. W. Baxter, L. Ferguson, et al. Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. Nature, 477(7363):203, 2011.

M. Kirkpatrick. How and why chromosome inversions evolve. PLoS biology, 8(9):e1000501, 2010.

M. Kirkpatrick and S. L. Nuismer. Sexual selection can constrain sympatric speciation. Proceedings of the Royal Society of London. Series B: Biological Sciences, 271(1540):687–693, 2004.

M. Kopp, M. R. Servedio, T. C. Mendelson, R. J. Safran, R. L. Rodríguez, M. E. Hauber, E. C. Scordato, L. B. Symes, C. N. Balakrishnan, D. M. Zonana, et al. Mechanisms of assortative mating in speciation with gene flow: connecting theory and empirical research. The American Naturalist, 191(1):1–20, 2018.

M. R. Kronforst, L. G. Young, D. D. Kapan, C. McNeely, R. J. O'Neill, and L. E. Gilbert. Linkage of butterfly mate preference and wing color preference cue at the genomic location of wingless. Proceedings of the National Academy of Sciences, 103(17):6575–6580, 2006.

Y. Kuang and Y. Takeuchi. Predator-prey dynamics in models of prey dispersal in two-patch environments. Mathematical Biosciences, 120(1):77 – 98, 1994. ISSN 0025-5564. https://doi.org/10.1016/0025-5564(94)90038-8. URL http://www.sciencedirect.com/science/article/pii/0025556494900388.

C. Küpper, M. Stocks, J. E. Risse, N. dos Remedios, L. L. Farrell, S. B. McRae, T. C. Morgan, N. Karlionova, P. Pinchuk, Y. I. Verkuil, A. S. Kitaysky, J. C. Wingfield, T. Piersma, K. Zeng, J. Slate, M. Blaxter, D. B. Lank, and T. Burke. A supergene determines highly divergent male reproductive morphs in the ruff. Nature Genetics, 48(1):79–83, 2016. 10.1038/ng.3443. URL https://doi.org/10.1038/ng.3443.

Y. Le Poul, A. Whibley, M. Chouteau, F. Prunier, V. Llaurens, and M. Joron. Evolution of dominance mechanisms at a butterfly mimicry supergene. Nature Communications, 5:5644, 2014.

V. Llaurens, L. Gonthier, and S. Billiard. The sheltered genetic load linked to the s locus in plants: new insights from theoretical and empirical approaches in sporophytic self-incompatibility. Genetics, 183(3):1105–1118, 2009.

V. Llaurens, S. Billiard, and M. Joron. The effect of dominance on polymorphism in müllerian mimicry. Journal of theoretical biology, 337:101–110, 2013.

J. Mallet and N. H. Barton. Strong natural selection in a warning-color hybrid zone. Evolution, 43(2):421–431, 1989. ISSN 00143820, 15585646. URL http://www.jstor.org/stable/2409217.

C. Mérot, V. Llaurens, E. Normandeau, L. Bernatchez, and M. Wellenreuther. Balancing selection via life-history trade-offs maintains an inversion polymorphism in a seaweed fly. bioRxiv, 2019. 10.1101/648584. URL https://www.biorxiv.org/content/early/2019/05/24/648584.

R. M. Merrill, P. Rastas, S. H. Martin, M. C. Melo, S. Barker, J. Davey, W. O. McMillan, and C. D. Jiggins. Genetic dissection of assortative mating behavior. PLoS biology, 17(2):e2005902, 2019.

F. Müller. 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, 1879.

S. P. Otto, M. R. Servedio, and S. L. Nuismer. Frequency-dependent selection and the evolution of assortative mating. Genetics, 179(4):2091–2112, 2008.

D. J. Penn and W. K. Potts. The evolution of mating preferences and major histocompatibility complex genes. The American Naturalist, 153(2):145–164, 1999.

S. Piertney and M. Oliver. The evolutionary ecology of the major histocompatibility complex. Heredity, 96(1):7, 2006.

T. N. Sherratt. Spatial mosaic formation through frequency-dependent selection in müllerian mimicry complexes. Journal of Theoretical Biology, 2006.

X. Thibert-Plante and S. Gavrilets. Evolution of mate choice and the so-called magic traits in ecological speciation. Ecology letters, 16, 06 2013. 10.1111/ele.12131.

J. W. Thomas, M. Cáceres, J. J. Lowman, C. B. Morehouse, M. E. Short, E. L. Baldwin, D. L. Maney, and C. L. Martin. The chromosomal polymorphism linked to variation in social behavior in the white-throated sparrow (zonotrichia albicollis) is a complex rearrangement and suppressor of recombination. Genetics, 179(3):1455–1468, 2008. ISSN 0016-6731. 10.1534/genetics.108.088229. URL https://www.genetics.org/content/179/3/1455.

H. B. Throneycroft. A cytogenetic study of the white-throated sparrow, zonotrichia albicollis (gmelin). Evolution, 29(4):611–621, 1975. ISSN 00143820, 15585646. URL http://www.jstor.org/stable/2407072.

T. Tregenza and N. Wedell. Genetic compatibility, mate choice and patterns of parentage: invited review. Molecular Ecology, 9(8):1013–1027, 2000.

E. M. Tuttle, A. O. Bergland, M. L. Korody, M. S. Brewer, D. J. Newhouse, P. Minx, M. Stager, A. Betuel, Z. A. Cheviron, W. C. Warren, et al. Divergence and functional degradation of a sex chromosome-like supergene. Current Biology, 26(3):344–350, 2016.

J. Wang, Y. Wurm, M. Nipitwattanaphon, O. Riba-Grognuz, Y.-C. Huang, D. Shoemaker, and L. Keller. A y-like social chromosome causes alternative colony organization in fire ants. Nature, 493(7434):664–668, 2013. 10.1038/nature11832. URL https://doi.org/10.1038/nature11832.

C. Wedekind and S. Füri. Body odour preferences in men and women: do they aim for specific mhc combinations or simply heterozygosity? Proceedings of the Royal Society of London. Series B: Biological Sciences, 264(1387):1471–1479, 1997.

C. Wedekind, T. Seebeck, F. Bettens, and A. J. Paepke. Mhc-dependent mate preferences in humans. Proceedings of the Royal Society of London. Series B: Biological Sciences, 260(1359): 245–249, 1995.

E. L. Westerman, N. W. VanKuren, D. Massardo, A. Tenger-Trolander, W. Zhang, R. I. Hill, M. Perry, E. Bayala, K. Barr, N. Chamberlain, et al. Aristaless controls butterfly wing color variation used in mimicry and mate choice. Current Biology, 28(21):3469–3474, 2018.

S. Wright. The distribution of self-sterility alleles in populations. Genetics, 24(4):538, 1939.

S1: Mendelian segregation

To compute the proportion of a given genotype in the progeny of the different crosses occurring in the population, we define a function $coef(g^O, g^M, g^F, \rho)$ summarizing the Mendelian segregation of alleles assuming two diploid loci and a rate of recombination ρ between these loci. Let $g^O = (p^O_m, p^O_f, m^O_m, m^O_f), g^M = (p^M_m, p^M_f, m^M_m, m^M_f)$ and $g^F = (p^F_m, p^F_f, m^F_m, m^F_f)$ be the offspring, maternal and paternal genotypes respectively, all in $\mathcal{G}$. For $I \in \{O, M, F\}$, p^I_m and m^I_m (resp. p^I_f and m^I_f) are the alleles on the maternal (resp. paternal) chromosomes. $coef(g^O, g^M, g^F, \rho)$ is the average proportion of genotype g^O in the progeny of a mother of genotype g^M mating with a father of genotype g^F given a recombination rate ρ.

Each diploid mother can produce four types of haploid gametes containing alleles (p^M_m, m^M_m), (p^M_f, m^M_f), (p^M_f, m^M_m) or (p^M_m, m^M_f), in proportion $\frac{1-\rho}{2}$,$\frac{1-\rho}{2}$,$\frac{\rho}{2}$ and $\frac{\rho}{2}$ respectively. Then the proportion of gametes with alleles $(p, m) \in \mathcal{A}_P \times \mathcal{A}_M$ produced by the mother is given by the function $coef_{haplotype}(p, m, g^M, \rho)$, where

$$\begin{aligned} coef_{haplotype}(p, m, g^M, \rho) =& \frac{1-\rho}{2}\mathbb{1}_{\{p=p^M_m\}}\mathbb{1}_{\{m=m^M_m\}} + \frac{1-\rho}{2}\mathbb{1}_{\{p=p^M_f\}}\mathbb{1}_{\{m=m^M_f\}} \\ &+ \frac{\rho}{2}\frac{1-\rho}{2}\mathbb{1}_{\{p=p^M_f\}}\mathbb{1}_{\{m=m^M_m\}} + \frac{\rho}{2}\frac{1-\rho}{2}\mathbb{1}_{\{p=p^M_m\}}\mathbb{1}_{\{m=m^M_f\}}. \end{aligned}$$

Similarly, each diploid father can produce four types of haploid gametes. The proportion of genotype $(p, m) \in \mathcal{A}_P \times \mathcal{A}_M$ in the gametes of a given father is given by the function $coef_{haplotype}(p, m, g^F, \rho)$.

The average proportion of genotype g^O in the progeny of a cross between a mother of genotype g^M and a father of genotype g^F given a recombination rate ρ is given by:

$$coef(g^O, g^M, g^F, \rho) = coef_{haplotype}(p^O_m, m^O_m, g^M, \rho) coef_{haplotype}(p^O_f, m^O_f, g^F, \rho).$$

S2: Checking of the computed genotype frequencies in the progeny of all crosses

We then check that the sum of the computed frequencies of the different genotypes i in the progeny of all crosses occuring in patch n ($(F_{i,n})_{i\in\mathcal{G}}$ for $n \in \{1,2\}$) actually equals to one. Let n be in $\{1,2\}$, we have:

$$
\begin{aligned}
\sum_{i\in\mathcal{G}} F_{i,n} &= \sum_{i\in\mathcal{G}} \sum_{(j,k)\in\mathcal{G}^2} coef(i,j,k,\rho) f_{j,n} \frac{M_{j,n}}{\overline{M}_n} Pref_{j,CP(k)} \frac{f_{k,n}}{T_{j,n}}, \\
&= \sum_{(j,k)\in\mathcal{G}^2} f_{j,n} \frac{M_{j,n}}{\overline{M}_n} Pref_{j,CP(k)} \frac{f_{k,n}}{T_{j,n}} \underbrace{\sum_{i\in\mathcal{G}} coef(i,j,k,\rho)}_{=1}, \\
&= \sum_{j\in\mathcal{G}} f_{j,n} \frac{M_{j,n}}{\overline{M}_n} \frac{\sum_{k\in\mathcal{G}} Pref_{j,CP(k)} f_{k,n}}{T_{j,n}}, \\
&= \sum_{j\in\mathcal{G}} f_{j,n} \frac{M_{j,n}}{\overline{M}_n} \frac{\sum_{k\in\mathcal{G}} (\mathbb{1}_{\{CP(k)=A\}} + \mathbb{1}_{\{CP(k)=B\}} + \mathbb{1}_{\{CP(k)=C\}}) Pref_{j,CP(k)} f_{k,n}}{T_{j,n}} \\
&\text{because } \forall k \in \mathcal{G}, \quad \mathbb{1}_{\{CP(k)=A\}} + \mathbb{1}_{\{CP(k)=B\}} + \mathbb{1}_{\{CP(k)=C\}} = 1, \\
&= \sum_{j\in\mathcal{G}} f_{j,n} \frac{M_{j,n}}{\overline{M}_n} \frac{\sum_{k\in\mathcal{G}} \mathbb{1}_{\{CP(k)=A\}} Pref_{j,A} f_{k,n} + \sum_{k\in\mathcal{G}} \mathbb{1}_{\{CP(k)=B\}} Pref_{j,B} f_{k,n} + \sum_{k\in\mathcal{G}} \mathbb{1}_{\{CP(k)=C\}} Pref_{j,C} f_{k,n}}{T_{j,n}}, \\
&= \sum_{j\in\mathcal{G}} f_{j,n} \frac{M_{j,n}}{\overline{M}_n} \frac{Pref_{j,A} \sum_{k\in\mathcal{G}} \mathbb{1}_{\{CP(k)=A\}} f_{k,n} + Pref_{j,B} \sum_{k\in\mathcal{G}} \mathbb{1}_{\{CP(k)=B\}} f_{k,n} + Pref_{j,C} \sum_{k\in\mathcal{G}} \mathbb{1}_{\{CP(k)=C\}} f_{k,n}}{T_{j,n}}, \\
&= \sum_{j\in\mathcal{G}} f_{j,n} \frac{M_{j,n}}{\overline{M}_n} \frac{Pref_{j,A} P_{A,n} + Pref_{j,B} P_{B,n} + Pref_{j,C} P_{C,n}}{T_{j,n}}, \\
&= \sum_{j\in\mathcal{G}} f_{j,n} \frac{M_{j,n}}{\overline{M}_n} \frac{T_{j,n}}{T_{j,n}}, \\
&= \sum_{j\in\mathcal{G}} f_{j,n} \frac{M_{j,n}}{\overline{M}_n}, \\
&= \frac{\overline{M}_n}{\overline{M}_n}, \\
&= 1.
\end{aligned}
$$

S3: Numerical resolution

In this study, we used a numerical scheme to simulate our dynamical system. For $(i,n) \in \mathcal{G} \times \{1,2\}$, let $N_{i,n}^t$ be the numerical approximation of $N_{i,n}(t)$. We use a explicit Euler scheme, therefore we approximate the quantity $\frac{d}{dt}N_{i,n}(t)$ by

$$\frac{N_{i,n}^{t+\Delta t} - N_{i,n}^t}{\Delta t},$$

with Δt being the step time in our simulations.

For $(i,n) \in \mathcal{G} \times \{1,2\}$, an approximation of equation 1 becomes:

$$\frac{N_{i,n}^{t+\Delta t} - N_{i,n}^t}{\Delta t} = Pred_{i,n}^t + Mort_{i,n}^t + Mig_{i,n}^t + Rep_{i,n}^t.$$

This equation is equivalent to:

$$N_{i,n}^{t+\Delta t} = N_{i,n}^t + \Delta t \left(Pred_{i,n}^t + Mort_{i,n}^t + Mig_{i,n}^t + Rep_{i,n}^t \right).$$

Given $(N_{i,n}^0)_{(i,n)\in\mathcal{G}\times\{1,2\}}$, we can simulate an approximation of the dynamical system.

S4: Numerical approximation of equilibrium states

To estimate the equilibrium reached by our dynamical system using simulations assuming different initial conditions, we define the variable Var^t quantifying the change in the numerical solution :

$$Var^t = \sqrt{\sum_{(i,n)\in\mathcal{G}\in\{1,2\}} \left(\frac{N_{i,n}^{t+\Delta t} - N_{i,n}^{t}}{\Delta t}\right)^2}.$$

When $\frac{Var^t}{N_{tot}} < 10^{-5}$, we assume that the dynamical system has reached equilibrium, with N_{tot} being the total density in both patches.

Supplementary Figures

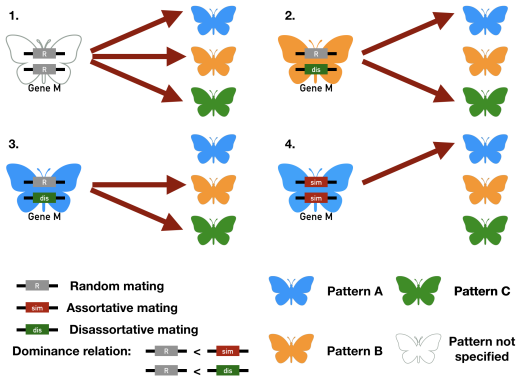


Figure S5: **Mate preferences expressed by individuals carrying different genotypes at the preference locus *M*, assuming *self-referencing* (hyp. 1).** 1. Butterflies carrying two *r* alleles mate at random, independently of either their own color pattern or the color pattern displayed by mating partners. 2-3. Butterflies carrying a *dis* allele display disassortative mating, and mate preferentially with individuals with a color pattern different from their own. 4. Butterflies carrying a *sim* allele display an assortative mating behavior and therefore 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*.

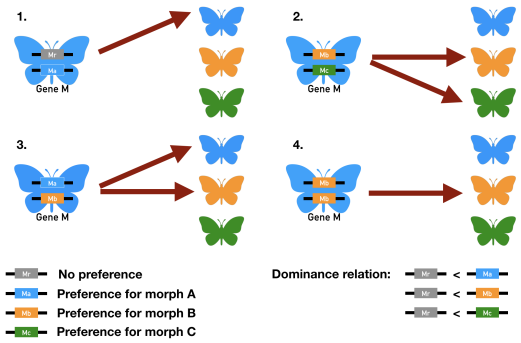


Figure S6: **Mate preferences expressed by individuals carrying different genotypes at the preference locus *M*, assuming preference alleles encoding for *attraction* of specific color patterns (*recognition/trait*) (hyp. 2.a)**. 1. A butterfly displaying phenotype *A* (in blue) carries one allele coding for specific attraction toward partners 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 partners 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 partners 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*.

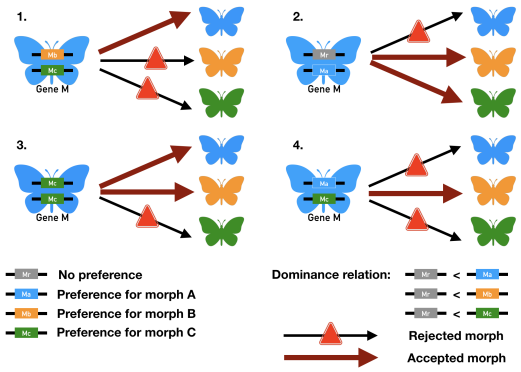
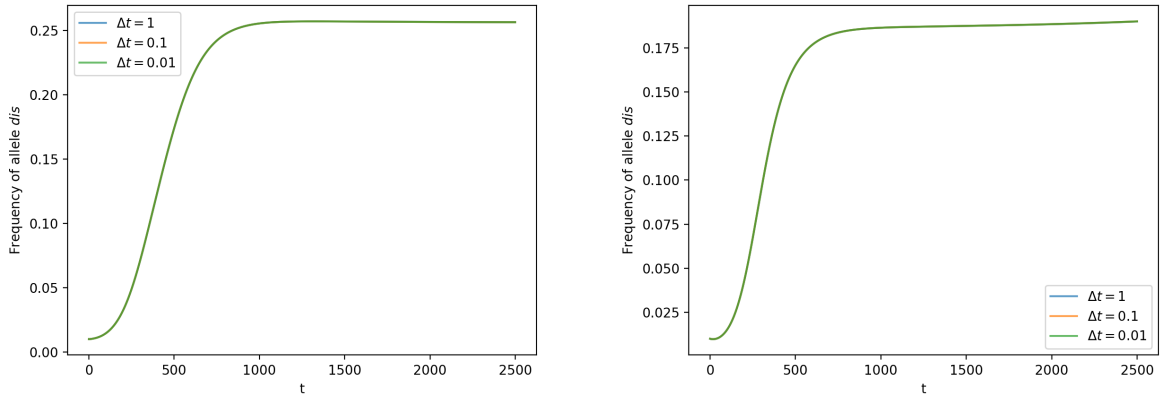


Figure S7: **Mate preferences expressed by the different individuals carrying different genotypes at the preference locus *M*, assuming preference alleles encoding for *rejection* of specific color patterns (*recognition/trait*) (hyp. 2.a)**. 1. A butterfly displaying phenotype *A* (in blue) carries one allele coding for specific rejection toward partners displaying phenotype *B* (in orange) and one allele coding for specific rejection toward partners 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) carries one allele coding for specific rejection toward partners 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) carries 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) carries one allele coding for specific rejection toward partners displaying phenotype *A* (in blue) and one allele coding for specific rejection toward partners 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*.

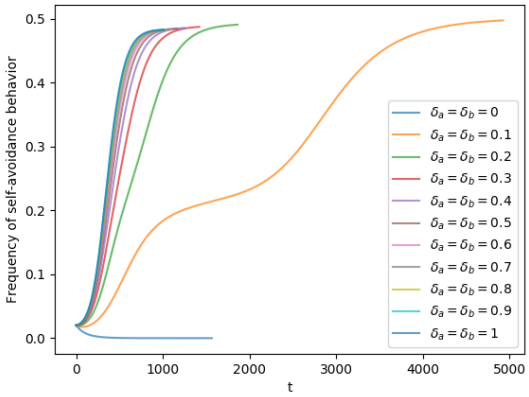


(a) (b)

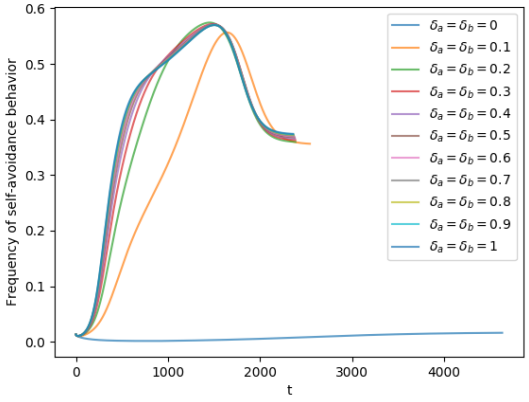
Figure S8: **Evolution of the proportion of a mutant *dis* in the population immediately after its introduction, using simulations with three different time units ($\Delta t = 1$ in blue, $\Delta t = 0.1$ in orange or $\Delta t = 0.01$ in green), under the *self-referencing* hypothesis (hyp. 1).** All simulations give similar dynamics, assuming (a) $\delta_a = \delta_b = 0.5$, $\delta_c = 0$ or (b) $\delta_a = \delta_b = \delta_c = 0.2$, confirming that using discrete time simulations provides relevant estimations of the evolution of disassortative mating. Simulations are run during 2500 time steps and assuming , $r = 1$, $K = 2000$, $N^0_{tot,1} = N^0_{tot,2} = 100$, $\lambda = 0.0002$, $d_m = 0.05$, $d_{n-m} = 0.15$, $\rho = 0$, $mig = 0.1$, $\delta = 0.1$ and $c_r = 0.1$.

$\delta_1 = \delta_2$	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	90,5 %	9,5 %	0,0 %	49,8 %	50,2 %	0,0 %
0,00	0,25	90,5 %	9,5 %	0,0 %	49,8 %	50,2 %	0,0 %
0,00	0,50	90,5 %	9,5 %	0,0 %	49,8 %	50,2 %	0,0 %
0,00	1,00	90,5 %	9,5 %	0,0 %	49,8 %	50,2 %	0,0 %
0,25	0,00	61,8 %	7,7 %	30,6 %	22,3 %	52,1 %	25,6 %
0,25	0,25	78,8 %	17,9 %	3,3 %	36,2 %	57,6 %	6,2 %
0,25	0,50	80,5 %	17,8 %	1,7 %	39,3 %	57,2 %	3,5 %
0,25	1,00	81,6 %	17,6 %	0,8 %	41,5 %	56,6 %	1,8 %
0,50	0,00	54,5 %	5,7 %	39,8 %	18,7 %	49,6 %	31,7 %
0,50	0,25	76,3 %	18,6 %	5,1 %	33,9 %	57,8 %	8,3 %
0,50	0,50	78,7 %	18,7 %	2,6 %	37,5 %	57,7 %	4,8 %
0,50	1,00	80,2 %	18,5 %	1,3 %	40,2 %	57,3 %	2,5 %
1,00	0,00	49,9 %	4,6 %	45,5 %	16,9 %	47,7 %	35,4 %
1,00	0,25	74,6 %	18,9 %	6,5 %	32,7 %	57,4 %	9,8 %
1,00	0,50	77,5 %	19,1 %	3,3 %	36,6 %	57,7 %	5,7 %
1,00	1,00	79,3 %	19,0 %	1,7 %	39,6 %	57,4 %	3,0 %

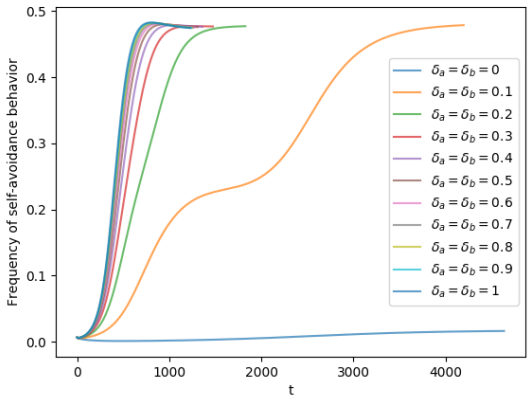
Figure S9: **Influence of genetic load on color pattern polymorphism, assuming random mating.** The proportions of phenotypes A, B and C in the populations living in patch 1 and 2 respectively at equilibrium depend on the different values of genetic load associated with the dominant allele a (δ_a), intermediate-dominant allele b (δ_b) and recessive allele c (δ_c). Simulations are run assuming $r = 1$, $K = 2000$, $N^0_{tot,1} = N^0_{tot,2} = 100$, $\lambda = 0.0002$, $d_m = 0.05$, $d_{n-m} = 0.15$, $\rho = 0$, $mig = 0.1$, $\delta = 0.1$ and $c_r = 0.1$.



(a)



(b)



(c)

Figure S10: **Frequency of *self-avoidance* behavior at the population level through time for different levels of genetic load, assuming (a) *self-referencing* (hyp. 1), (b) *attraction* rule (hyp. 2.a) or (c) *rejection* rule (hyp. 2.b) at the preference locus (*recognition/trait*)**. The evolution of the proportion of individuals displaying *self-avoidance* $P_{s\text{-}av}$ after the introduction of preference alleles until equilibrium are shown for different values of genetic load δ_a and δ_b. Simulations are run assuming $r = 1$, $K = 2000$, $N^0_{tot,1} = N^0_{tot,2} = 100$, $\lambda = 0.0002$, $d_m = 0.05$, $d_{n-m} = 0.15$, $\rho = 0$, $mig = 0.1$, $\delta_c = 0$, $\delta = 0.1$ and $c_r = 0.1$.

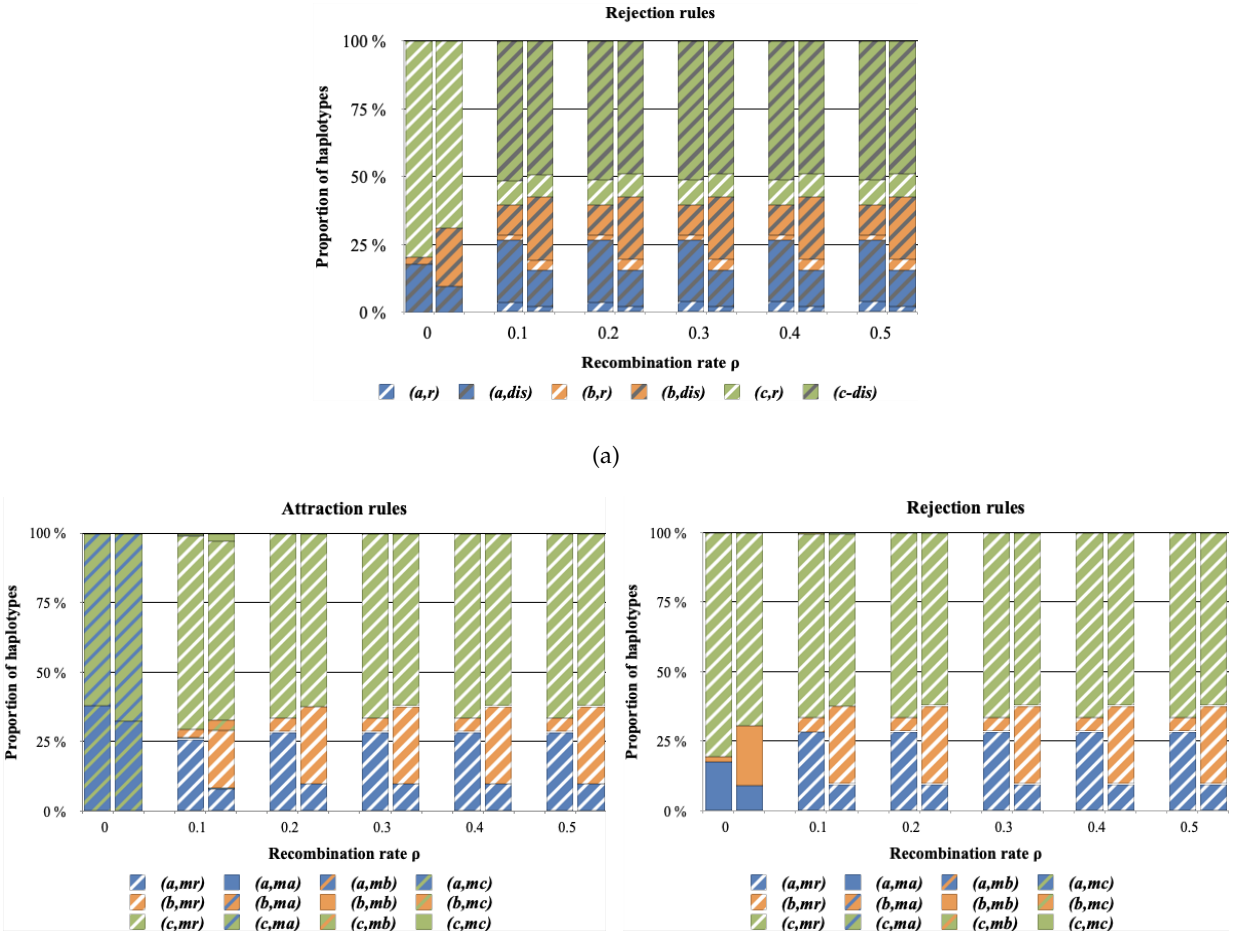


Figure S11: **Influence of the recombination between color pattern and preference alleles on haplotype diversity, assuming (a) *self-referencing* (hyp. 1), (b) *attraction* rule (hyp. 2.a) or (c) *rejection* rule (hyp. 2.b) at the preference locus (*recognition/trait*)**. The proportion of haplotypes at equilibrium after the introduction of preference alleles in both patches are shown for different values of recombination rate ρ between the preference locus M and the color pattern locus P. For each value of recombination rate (ρ) the first and second bars represented haplotype proportions in the populations living in the patch 1 and 2 respectively. Simulations are run assuming $r = 1$, $K = 2000$, $N^0_{tot,1} = N^0_{tot,2} = 100$, $\lambda = 0.0002$, $d_m = 0.05$, $d_{n-m} = 0.15$, $mig = 0.1$, $\delta_a = \delta_b = 0.5$, $\delta_c = 0$, $\delta = 0.1$ and $c_r = 0.1$.