Dear Dr. Charles Mullon,

We are very pleased to re-submit our manuscript to PCI evolutionary biology. We just uploaded the revised version on the bioRXiv website, so it might take a few hours to be updated online.

The numerous questions and relevant comments raised by the reviewers and yourself have highlighted the lack of clarity of our material and methods. We entirely revised it to clarify our approach and our hypotheses. We also added a supplementary detailing the numerical scheme used for numerical simulations. We now have provided our scripts online at https://github.com/Ludovic-Maisonneuve/Evolution_and_genetic_architecture_of_disassortative_mating, allowing any reader to check and reuse it. In response to reviewer 3 comments we re-ran all of our simulations verifying that the numerical solution reaches a steady state. We carefully revised the result section as well, following all the comments. You will find below our detailed answers to all the comments. We hope that the profound revisions we carried out on our manuscript have improved it and that you will enjoy the reading.

Sincerely yours,

Ludovic Maisonneuve, on behalf of all authors.

Comments

1. The general structure of your model is still not clear. From eq. 1, one would assume that your model consists of just 8 differential equations (4 genotypes * 2 patches). Appendix S1 however suggests that you in fact track haplotypes rather than genotypes. So please state clearly at the beginning the structure of your model, clearly defining all variables involved. In short, readers need to know all the dynamic variables that you use in your numerical simulations in order to be able to replicate them.

The numerous questions and relevant comments raised by the reviewers and yourself have highlighted the lack of clarity of our material and methods. We entirely revised it to clarify our approach and our hypotheses. We also added a supplementary detailing the numerical scheme used for numerical simulations. We also provide our scripts online at https://github.com/Ludovic-Maisonneuve/Evolution_and_genetic_architecture_of_disassortative_mating, allowing readers to check and reuse it.

To specifically answer to your question: the model indeed aims at tracking the densities of each genotype within the populations living in patch 1 and 2 respectively. The number of possible genotypes depends on number of alleles at locus M, which is defined by the hypothesized genetic architecture of mate preference (either self-referencing (hyp.1) or recognition/trait (hyp.2). Assuming self-referencing (hyp.1), there are only two possible alleles at M locus within each
simulation (either random and assortative alleles \((r, \text{sim})\) or random and
disassortative alleles \((r, \text{dis})\). As there is three alleles at locus \(P\) \((a, b, c)\) the number
of genotype is \(3 \times 3 \times 3 \times 3 = 81\). Because we track down genotypic densities in both
patches, we obtain a dynamical system of \(81 \times 2 = 162\) equations. Assuming
recognition/\text{trait} (hyp. 2), we obtain \(3 \times 3 \times 4 \times 4 \times 2 = 288\) differential equations.

2. L.117, the ODE approach your model takes suggests that you track the
“density” rather than the “number of individuals”. This is for example what is
mentioned in the model by Joron and Iwasa, 2007, JTB, on which the current
analysis is based. In general, I have found this previous model much more
rigorously and clearly explained than your current one. You may therefore
find it useful to use the same terms and structure of explanation as Joron and
Iwasa, 2007, JTB.

This is correct, we track density rather than the number of individuals, this is now in
the manuscript (see lines 127-129, page 6). We profoundly changed the method
section to explained each term more rigorously (see lines 116-364, pages 6-20).

3. Like reviewer 3 I am baffled by the use of “d” on the right hand side of eq. 1.
In addition, this is defined later as a parameter for predation (eq. 3). In fact,
quite often you use the same or very similar notation to refer to different
things (e.g., \(P_i\) is the contribution of predation to the rate of change in
density in eq. 1 but also as “as the proportion of the phenotype” in eq. 7; the
use of a dash ‘ next to a variable is not consistent, with pop’ meaning
population other than pop, and f’ frequency in offspring rather than f in
parents). Simultaneously, you also use different notations to refer to similar
things (e.g., matrix notation is sometimes with subscript with or without
commas, and further with or without square brackets). This is not an
exhaustive list, so please go over your model carefully and make sure that
notations are consistent throughout. I also strongly recommend providing a
table with all variables and parameters of the model.

Our choice of variable names was highly misleading and we apologize for this. We
thus changed numerous notations in our model to avoid this problem, and checked
the manuscript for consistency in their use. We also added a table with the
description of all variables and updated the table with all parameters (Tables 1 and
2 pages 16-17).

4. Eq. 3 is overly complicated. First the denominator should simply be
\(\lambda N_i, \text{pop}\). Secondly, I don’t understand why in the numerator you
introduce the parameter \(\sigma\) when it is later fixed in all your analyses as
0.5. Why not simply have \(D_i, \text{pop}\) like in Joron and Iwasa, 2007, JTB?

Because we poorly explained our model, it appears as if there was a single non-
zero term in the sum. However, this sum counts the density of individuals displaying
the same color pattern as individual with genotype i in the patch pop. For example
individuals with genotypes (a,a,m1,m2), (a,b,m1,m2) and (a,c,m1,m2) have the
same color pattern irrespectively of the associated alleles m1, m2 at preference
locus.
Following your advice, we erased the parameter sigma and we used the D_i^pop
(named d_i,n in the new version of the manuscript) as in Joron and Iwasa 2017 JTB
(lines 198-206, pages 9-10).

5. Eq. 4, perhaps write out the intermediate step \( = - \text{mig} N_i,\text{pop} + \text{mig} N_i,\text{pop}' \).

In the revised version, we note write this term suggested and we added amore
thorough explanation of both terms - migN_i,\text{pop} and migN_i,\text{pop}' (lines 218-225, page 11).

6. The main difficulty in understanding the model concerns the sections “mate
preference” and “Reproduction” which require extensive revisions. Reviewer
3 points out many of the problems with these two sections (including
equations 5-6 that appear out of nowhere without defining variables, eq. 8
that comes before the sentence explain eq. 7 is finished). My suggestion to
aid readers is to start with eq. 10 (which is kind of clear but see point 7
below), then explain how the frequency of offspring of different genotypes is
calculated, to finally introduce the notions of preferences, mating probabilities
and recombination (also please choose carefully what you mean by “fertility”).

We followed your advice and we reorganize the whole method section and
specifically aim at clarifying the definition of fertility and mate preferences (lines
272-281, pages 13-14).

7. The way I read eq. 10, is that due to density dependent competition, each
offspring has a probability of survival given by r and K. In this case, I expect
the term \( r(1-N_{\text{tot,\text{pop}}}/K) \) to be multiplied to “the abundance of offspring with
genotype i”. However, in eq. 10 it is multiplied by \( N_{\text{tot,\text{pop}}} f'_i,\text{pop} \), where
\( N_{\text{tot,\text{pop}}} \) is the total abundance of adults, while \( f'_i,\text{pop} \) is the frequency of
genotype i in offspring. What is the meaning of the multiplication of these two
terms in the context of eq. 10?

Following Llaurens et al. (2013 JTB), we refer to as \( r(1-N_{\text{tot,\text{pop}}}/K)N_{\text{tot,\text{pop}}} \) the
total density of the offspring in the population at a given time unit. Out of this total
density of offspring, we compute the genotype frequencies depending on female
preferences, cost of preferring rare males and genotype densities in the parents.
We clarify this point in the manuscript (lines 232-240, page 12)
8. More generally, and as pointed out by reviewer three, make sure that the distinction between frequency and density (or abundance) is clear and that quantities defined as frequencies are in fact frequencies (e.g., does the sum of eq. 9 over $i$ really sum to one?)

Our reorganization of the methods should clarify this ambiguity and we also provide a demonstration validating that the sum of eq. 9 equals one in supplementary S2.

9. Eq. 9. What is “relative” probability?

The relative probability is computed among all the mated females: it is thus the probability for a mated female to be of genotype $j$ (See Otto et al. Genetics 2008). It is a conditionnal probability that a female has genotype $i$, knowing that this female has mated.

10. Eq. 9. Shouldn’t the denominator for the probability that male $k$ is the father be $F_{k,\text{pop}}$?

Once again, Eq. 9 was poorly explained, but the denominator should still be $F_{j,\text{pop}}$. The term $\text{Pref}_{j[k]f_{k,\text{pop}} / F_{j,\text{pop}}}$ is the probability that a female with genotype $j$ choose to mate with a male with genotype $k$ and can thus be written as the ratio between the frequency of male $k$ weighted by female preference, on the total frequency of males. (see lines 300-301, page 15)

11. Methods section. Please clarify your methods of numerical analysis as advised by reviewer 3. E.g. Where can one find the whole dynamical system you are solving (or at least find all the relevant building blocks)? How many time steps? What initial conditions?

The important revisions carried our on the material and method sections and one the appendix should now allow any reader to replicate the dynamical system. The numeric scheme used is also specifically detailed now in supplementary 3. We also provide our scripts online at https://github.com/Ludovic-Maisonneuve/Evolution_and_genetic_architecture_of_disassortative_mating for readers to check and reuse it.

12. I did not review your results in detail as I don’t understand the model. I suggest that you aim for the same level of clarity in your results and the one I am advising for your model. So please, revise your results as extensively as required by reviewers’ comments.

We also revised the result section, following reviewers comments and hope you will be able to enjoy reading them now.

Reviews
Reviewed by Tom Van Dooren, 2020-03-19 21:44
This is my review of the revision "" by Maisonneuve et al. My previous remarks have been dealt with adequately. I am pleased by the changes made by the authors in this revision. However, I have many remaining small issues with the writing. Please also use English or American spelling consistently. I didn't make that choice in my suggested corrections.

Specific comments
L17: behaviors -> behavior
L20: Please remove "that are starting to emerge"
L23: traits -> trait
L25: trait -> traits
L36: fungy -> fungi
L36: change sentences to: "Disassortative mating can be based on different traits. In Amphidromus inversus snails,..." Write species names such as Amphidromus inversus italic.
L41: "controlling for" refers to correcting for nuisance variables in statistics. "controlling for variations in" -> "affecting the"
Also check the legend with figure one, three, five. L411.
L44: promotes -> promote
L46: heterozygotes -> heterozygote
L49: differ -> differs
L56: for which -> where; controls for -> determines
L58: contains -> contain
L70: when individual used -> when an individual uses
L76: toward -> towards
L80 locus -> loci
L84: population -> populations
L97: promotes -> promote
L98: such -> a
L110: based on recognition -> based on a recognition
L111: impact -> impacts
One line below L114: on earlier -> on an earlier
L116: controlling variations in wing colour pattern -> controlling wing colour patterns
L117: the number -> the number of individuals
L118: populations -> population
L120: described -> describe
L121: populations -> population
L133: pattern -> patterns
L139: "The environment ... " -> "Local communities of species involved in mimicry (i.e. mimicry rings) vary."
L140: "we": start a new sentence here.
L145: remove ")"
L146: Strength -> strength
L151: higher -> larger
L158: was -> is
L172: each -> a
I enjoyed this manuscript. It describes an interesting model with results that have interesting implications for understanding disassortative mating. The authors have greatly improved the manuscript and have addressed my previous comments. I have just a few additional comments:

Thank you Tom for pointing out these numerous errors. We corrected each of them and all authors checked the manuscript again after corrections to avoid these mistakes.

Reviewed by anonymous reviewer, 2020-03-25 08:06
1) As far as I understand, the only difference between the two "patches" in the model comes from differences in phenotype specific predation rates. $d$ is the base line predation rate and sigma adjusts that rate. But in some of the figures (e.g. fig 1) the caption says that $d = 0$ for the simulation. Doesn't this make the patches identical? What is going on here? Perhaps I still don't understand how the patches differ.

We agree that simulations assuming $d = 0$ and $\delta = 0$ this are irrelevant to the question of the evolution of disassortative mating. We now have run relevant simulations using $d = 0.1$ and $\delta = 0.1$, and have corrected figure 1 accordingly.

2) I believe that Figure S7 presents results from the same simulations as Figure 2, but run for a longer time. Just looking at Fig 2, it looks like there is a threshold (shown be the purple line) that divides parameter space into a region where invasion occurs and a region where invasion does not occur. But I think this is a bit misleading because Fig S7 indicates that invasion happens everywhere if you wait long enough. So my question is, does the purple line on Fig 2 tell us anything useful in the end?

Most results shown in the manuscript indeed described equilibrium state but we also studied the mutant invasion during the first 100 times unit to estimate the mutants advantage during early dynamic. This allows to estimate how likely the invasion of such mutant would be in natural populations where genetic drift may wipe out mutants with a slow increase in frequency. This is now clarified in the Material and Method section (lines 324-339, page 18).

3) There are still a few typos scattered throughout the manuscript. I suggest another thorough read through.

All authors checked the manuscript again after corrections to avoid these mistakes.

Reviewed by anonymous reviewer, 2020-03-08 17:30

I reviewed this manuscript previously. After the revision, I can follow the set-up of the model much better, although this is still not an easy task. Overall, the manuscript still reads very rough in many places and somewhat unfinished, as I will detail below.

The numerous questions and relevant comments raised by the reviewers and yourself have highlighted the lack of clarity of our material and methods. We entirely revised it to clarify our approach and our hypotheses. We also added a supplementary detailing the numerical scheme used for numerical simulations. We also provide our scripts online at https://github.com/Ludovic-Maisonneuve/Evolution_and_genetic_architecture_of_disassortative_mating, allowing readers to check and reuse it.
By commenting our scripts to make it accessible, we spotted an error in the frequencies of introduction of preference alleles under recognition/trait rule (hyp.2), we this run again some simulations, explaining why some results reported in the revised version are slightly different.

MODEL DESCRIPTION

Eq. (2) is unclear. What is the letter d in the four terms on the right-hand side supposed to mean. At a later point d is introduced as predation strength. But I do not think that this is what the authors mean here. Instead, I suspect that the authors want to refer here to a differential. But why? In this revised version, the model is formulated as an ODE and each term on the right-hand side describes the rate of change in abundance of a genotype in one of the populations due to a different mechanism. I cannot see why a differential d should appear here.

Our choice of variable names was highly misleading and we apologize for this. We thus changed numerous notations in our model to avoid this problem, and checked the manuscript for consistency in their use. We also added a table with the description of all variables and updated the table with all parameters (Tables 1 and 2, pages 16-17).

Formulating the model as an ODE requires some implicit assumptions, namely, that generations are overlapping and that each newborn individual instantaneously behaves as an adult individual that can migrate and reproduce. There is no time delay due to an ontogenetic development. I understand that the model is meant as an approximation of real dynamics but I would like to see these conceptual problems being pointed out.

We agree that it is crucial to point out the implicit assumptions. We thus clarified them during the revision of the method part (see lines 139-146, page 7).

In eq. (3) I do not understand the denominator. Given that Res[i,j]=1 for i=j and zero otherwise I understand that the sum only contains a single non-zero term, namely, Res[i,i]N_i, pop. So, why the sum? Is this not more complicated than necessary?

Because we poorly explained our model, it appears as if there was a single non-zero term in the sum. However, this sum counts the density of individuals displaying the same color pattern as individual with genotype i in the patch pop. For example individuals with genotypes (a,a,m1,m2), (a,b,m1,m2) and (a,c,m1,m2) have the same color pattern irrespectively of the associated alleles m1, m2 at preference locus.

In my opinion eq. (4) is not well motivated. It says that individuals of genotype i migrate proportionally to the difference in abundance of this genotype in the two patches. This seems to suggest that individuals could assess the abundance of their genotype in both patches, an idea that I would consider unrealistic. A much simpler
and, in my opinion, more realistic idea would be that a fixed proportion of individuals present in each patch migrate. More realistic might, however, be that migration is not random but takes into account some form of habitat matching. I realise that I made the same comment in my previous review. The response by the authors (“We assume a constant migration rate mig i.e. a proportion mig of the population migrates in the other patch. Equation (3) is a result of individuals entering and exiting the patch pop.”) does not address my point.

In our model, we do not assume that individuals could assess the abundance of their genotype in both patches and adjust their migration behavior. On the contrary, we assume a constant migration rate mig, so that the density of migrant in one direction only depends on the total density in the patch they leave. Eq. (4) thus describes a density of individuals mig x N_i,pop leaving the patch and a proportion mig x N_i,pop' migrate entering the patch at each time unit. We modified this section of the material and method to clarify the meaning of that eq. 4 (Lines 218-225, page 11).

In eq.(5) and (6) $f_i$ and $\text{Pref}_i[i]$ appear. Neither of these have been introduced in the text so far. Thus, at this point I cannot understand these equations. Thank you for pointing out this mistake, indeed we need to introduce Pref earlier so that these equations can be properly understood by the readers. This is now corrected in the revised version of the method section (Lines 272-276, page 13 and lines 294, page 14).

In ll. 245 $\text{Pref}_{i,[j]}$ is defined as a preference of one genotype for another genotype. In eq. (7) $\text{Pref}_{i,A}$ is used, where A is a phenotype due to alleles at the P-locus. This seems to be an abuse of notation and does not facilitate reading the manuscript. The Authors also might want in some place in their manuscript explain clearly their notation. What is the rational for either putting subscripts into hard brackets or not? Also, sometimes two subscripts are separated by a comma and sometimes not. Our choice of variable names was highly misleading and we apologize for this. We thus changed numerous notations in our model to avoid this problem, and checked the manuscript for consistency in their use. We also added a table with the description of all variables and updated the table with all parameters (Tables 1 and 2, pages 16-17).

ll. 249 this sentence belongs directly after eq. (7) because that is where $P_{i,\text{pop}}$ appears for the first time. Modified

ll. 249 refer to as -> denotes Modified

What is meant by proportion here? Frequency or abundance? For abundance $N_{i,\text{pop}}$ has been used in eq. (2), for frequency $f_{i,\text{pop}}$ is used in eq. (9).
P_I^pop is the frequency of the phenotype I in the population pop. We now provide an explicit formula for this term in the methods to clarify (line 280, page 14).

Furthermore, in eq.(2) P_i,pop denotes reproduction. Quite confusing. Our choice of variable names was highly misleading and we apologize for this. We thus changed numerous notations in our model to avoid this problem, and checked the manuscript for consistency in their use. We also added a table with the description of all variables and updated the table with all parameters (Tables 1 and 2, pages 16-17). For the sake of clarity Pred_i^pop denotes the predation in the revised version of the methods (lines 135-138, page 7).

More generally, I do not understand the rationale behind eq. (7). What exactly is meant by fertility? What is the logic behind the right-hand side of eq. (7)? I do not see a proper explanation. Is it the proportion of the male population a given female can mate with? What is assumed about female mating behaviour? Can they search for mating partners until they are mated or can they choose only one mating partner, which then maybe suitable or not, per time unit?

After reading your comment, we realized that this section was unclear. We thus revised it completely. Our modelling of preferences follow the framework described in Otto et al. (2008). We thus assume that choosy females can pay a relative cost (c_r) when they prefer males with rare phenotype. This fitness cost depends on the parameter c_r. In our revised notations, the fertility of female with genotype i is now refer to as M_i.

When c_r = 0, M_i is always equal to 1 and every female have the same fertility (i.e. there is no cost for being choosy). When c_r > 0 the females who have preference for rare males pay a fitness cost, associated to lost mating opportunities.

In line 239 it is said that f'I,pop is a frequency. However, it is not obvious that eq. (9) truly defines a frequency in the sense that the f'I,pop’s sum up to one. I wonder whether for this to be true a normalisation by the total f’_i,pop would be necessary? Our reorganization of the methods should clarify this ambiguity (see eq 13 page 15) and we also provide a demonstration validating that the sum of eq. 9 equals one in supplementary S2.

I. 261 Here, the authors refer to the appendix for an explanation of coef(i,j,k,rho). Without further explanation I find it very difficult to understand the formula for coef_haplotype that is given in the appendix. First, no explanation is given for the set-theoretic notation that is used here. Second, no explanation is given for the different terms on the right-hand side. As a reader of a manuscript I do not want to be a detective.

The appendix S1 describing the functions coef(i,j,k,rho) and coef_haplotype has been entirely revised to improve clarity (pages 43). We also clarified the definition of haplotype (see lines 341-345, page 19).
NUMERICAL METHODS

I find the numerical methods explained in insufficient detail. We made a lot of effort to explain our numerical method in more detail, both in the method section and also in the supplementary S3.

First, how is it assured that the dynamics reach a fixed point equilibrium. Is it always so that simulations run for 10000 time steps do show no further change? In several of the results I wonder whether the numbers presented in the figures truly represent frequencies at a fixed point equilibrium or whether these frequencies are still transient.
We agree with your comment, it is unknown whether the solution reaches an equilibrium point after 10000 time units in any simulation. To insure the dynamical system reached an equilibrium, we run simulations until the change in genotype frequencies during one unit of time stays below a threshold of 10^-5 (see supplementary S4 for more details).

Second, how is it assured that, if a fixed point equilibrium exists, it is unique? In other words, how is it assured that the dynamics does not depend on the initial conditions?

Third, in the legend of fig. 3 it is mentioned that the dynamics are first run for 10000 time units with random mating before the other mating alleles are introduced. Why this procedure? If the dynamics would have a unique equilibrium, then it should not matter whether one starts the simulation with all alleles present simultaneously at both loci.
Ancestral state => random, how does preference evolves ?

We agree that it is important to study the number of steady states of an ODE system. Here, the complexity of our ODE system prevents the identification of whole set of the steady states and we cannot determine the number of equilibrium points. We are aware that the initial conditions may have an importance on the trajectory and the final equilibrium point reached by the dynamical system (if the dynamical system reaches one).
The manuscript aims at investigating the emergence and the evolution of disassortative mating. We thus focused on a specific scenario enabling to tackle this question and thus specifically looked at the emergence of disassortative mutant.
We also assumed that the ancestral mate preference behavior is random mating and therefore introduce the preference allele in an initial population where individuals mated randomly. For this reason, we first let our population evolved under random mating until equilibrium during initial simulations. We then introduced preference allele in small proportion (1%) and we follow the fate of this preference alleles by looking at the steady state reached after the introduction of this allele. We clarify this in the manuscript (see lines 324-330, page 18).
Fourth, I wonder whether 10000 time units is a “long time” or a “short time”. To how many generations does this correspond? It is clear that this is not easy to answer this but methods to estimate generation time for overlapping generations do exist. To the extend that the results are transient an interpretation based purely based on the number of time units is difficult.

We can not directly translate our time units into generation time, but we now run our simulations until the numerical resolution per time unit does not change above $10^{-5}$ to make sure we reached equilibrium. We also study the mutant invasion during the first 100 times unit to estimate the advantage of the mutant during the early. This allows to estimate how likely the invasion of such mutant would be in natural populations where genetic drift may wipe out mutants with a slow increase in frequency. This is now clarified in the Material and Method section (see lines 331-339, page 18).

Fig. S5 shows the temporal dynamics of a the mutant allele dis. I am not sure about the meaning of the scale on the y-axis but what this graph suggests to me that 100 time units is a very short time span since the $P_{\text{mut}}$ seems to grow exponentially without any signs of slowing down.

We run new simulations for 2500 units of time and corrected the figure accordingly (see Fig.8, page 50). Using this longer simulation time, we noticed the expected slow down.

Fifth, this is very parameter rich model but only a relatively small range of the parameter space seems to have been explored. Specifically, the parameters $d$, $\sigma$, $\lambda r$, and $K$ have not been varied. I understand that this are chosen such that these result in disruptive selection at the $p$-locus. Is there another justification for not varying them?

We agree that our model contains many parameters. It is based on previous model on polymorphic mimicry (Joron & Iwasa 2005 and Llaurers et al. 2013 for the diploid version) that have extensively explored the effects of these parameters on color pattern polymorphism. We thus focuses on the few parameters directly linked to the evolution of disassortative mating and the underlying genetic architecture of preference, namely the cost of choosiness, the level of genetic burden associated with color pattern alleles and the recombination rate.

RESULTS

Fig. 1
For this figure $d=0$ and $\delta=0$ for all deltas. In other words, in these simulations mortality is absent, individuals can reproduce and migrate but never die. This means that once the carrying capacity its reached in both patches no more population wide changes in genotype frequencies can occur since reproduction comes to a halt. Investigating such a case seems a very odd choice. I would expect
that the final gene frequencies are highly sensitive to initial conditions and do not reflect a unique equilibrium fixed point. This analysis seems close to irrelevant to me.

We agree that simulations assuming $d = 0$ and $\delta = 0$ are irrelevant to the question of the evolution of disassortative mating. We now have run relevant simulations using $d = 0.1$ and $\delta = 0.1$, and have corrected figure 1 accordingly.

p. 18-19 and Figs. S7 and S8
The legend for figure S7 seems to be wrong. Also, Fig. S7 does not seem to show what is mentioned in the main text, at least, I cannot read from Fig. S7 that widespread fixation of the allele $\text{dis}$ occurs.

We agree with the reviewer’s comment. Figure S7 (figure 3 in the new version) indeed does not indicate that the fixation of allele $\text{dis}$ occurs. We corrected the manuscript and the figures to highlight that indeed the $\text{dis}$ allele is never fixed (lines 438-439, page 23).

Figure S8 shows that for higher costs (cost=0.25) in most of the parameter space the frequency of the allele $\text{dis}$ drops below the initial frequency 0.01. Thus, I would disagree with the authors who state “Simulations assuming different costs associated with choosiness (cost) show a similar effect of associated genetic loads, although increasing this cost slows down the invasion of the choosy disassortative mating mutant $\text{dis}$ (see Sup. fig. S8).” (ll. 345-347)

This was indeed misleading: increasing the cost of choosiness indeed reduces the range of parameters allowing the emergence of allele $\text{dis}$. We corrected the manuscript accordingly (lines 429-432, page 23).

In the legend of figures S7 and S8 the value of the parameter $\delta$ is not given. Is figure S8b supposed to be identical to figure 2? It is not clear to me whether the two alleles $r$ and $\text{dis}$ at the mating locus can coexist long term for some parameter combinations.

In the new version of the manuscript, figure 3 and 4.a clearly indicate the coexistence of preference alleles $r$ and $\text{dis}$ when color pattern alleles $a$ and $b$ suffer from a genetic load. Indeed in this case allele $a$, $b$ and $c$ are maintained in the population and no recombination is assumed. Color pattern alleles $a$ and $b$ are advantaged while linked to dis preference allele and color pattern $c$ is advantage while linked with $r$ preference allele. This allows the coexistence of allele $r$ and $\text{dis}$ at the preference locus $M$. We clarify this point in the manuscript (lines 439-441, pages 23-24)

Why are in fig. 2 the frequencies computed for only 100 time units? As mentioned above, according to fig. S5, this seems rather short. For fig. 2 I miss information about the frequencies of the different color pattern alleles. As I explained above, the results from fig. 1 are uninformative for this case.

p. 20-22
In figure 2, we indeed show the dynamics of the mutant, by displaying its frequency 100 time units after introduction. This allows us to quickly estimate the conditions enabling the invasion of the mutant. If the mutant frequency increase slowly immediately after its introduction in the population, we assume that the mutant would be lost by drift, in an actual population where stochastic events happen.

In figure 3b the results seem to be identical for all \( \delta_i > 0 \). This seems rather surprising to me. If it is really true, it needs an explanation. Similarly, in fig. 3a the results seem to be identical for \( \delta_i > 0.3 \). Given that in figure 3a the results for \( \delta_i = 0.1 - 0.3 \) are clearly different from those for \( \delta_i = 0.4 - 1 \) I find it very surprising that the results in figure 4a, which in my understanding are derived from those in figure 3a, hardly differ for different \( \delta_i > 0 \). This does not seem to be discussed but in my opinion, if true, requires an explanation.

The previously shown figure 3a reported values obtained after 2000 time steps, so that equilibrium was not always reached. To improve clarity, we thus performed new simulations and modified the figures from the manuscript: we now run our simulations until equilibrium was reached (see S4 for more detail). On the corrected Figures, we now observe that \( \delta_1 \) (now \( \delta_a \)) hardly affects haplotype frequencies as soon as \( \delta_1 > 0 \). (fig.4).

p. 23, ll. 408-414 and Fig. S10

These results are quite incomprehensible. Fig. S10 shows results that in my understanding are based on simulations shown in figure 2 (and S7, although I am unsure whether S7 is the correct figure). In ll. 342-343 it is mentioned that over longer evolutionary time the dis allele is fixed (something that is in fact not visible from figure S7). If this were to be the case, I wonder whether it is really possible that half of the population acts according to self-acceptance and half according to self-avoidance? Furthermore, fig. 2 (and fig. S7) show a clear gradient for increasing \( \delta_i \). In contrast, fig. S10 shows that results do not change with increasing \( \delta_i \) for \( \delta_i > 0 \). Are these results compatible? I recommend to add a figure in the style of fig. 3 accompanying fig. S10.

Thank you for pointing this discrepancy: you are absolutely right, we do not observe the fixation of allele dis at equilibrium. Figure 3 now clearly shows the coexistence of alleles \( r \) and dis when color pattern alleles \( a \) and \( b \) suffer from a genetic load, but also highlights that the dis allele is not fixed in the population. In these simulations, alleles \( a, b \) and \( c \) are maintained in the population and there is no recombination. Color pattern alleles \( a \) and \( b \) are advantaged when linked to the dis preference allele and color pattern \( c \) is advantaged when linked with randing mating allele \( r \) (fig. 4a). This allows the coexistence of allele \( r \) and dis. This coexistence between these two alleles explain why different mating behavior are observed within populations (fig.6a).

On figure S7 (now fig 2) when \( \delta_c = 0 \), the frequency of allele dis does not vary trough \( \delta_a, \delta_b \) since \( \delta_a > 0.1 \). This was not observable on figure S7.
because of the way the fonction matplotlib.pyplot.colormesh was doing the graphic by ignoring the values at the border of the graphic.

p. 24-26 and figs. S10, S11 and S12
Several results in this section are difficult to understand.
We are deeply sorry for this lack of clarity and we have now used your comments to improve our description of the results.

First, in l. 434 the authors write that recombination further promotes the invasion of the allele dis. I wonder what that means given that the authors write in l. 343 that over long time span the allele dis reaches fixation (without recombination). Do they mean that the speed of invasion is increased? If this is the case, then I come back to my earlier question to what extend the presented results show transient dynamics and to what extend they show equilibria?
A comparison between the column corresponding to delta_i=0.5 in fig. S10 and fig. S11 shows that any positive recombination rate completely changes the result compared to the case with no recombination. Again, it is not clear to me whether this concerns the outcome at a stable point equilibrium or transient dynamics as observed after 10000 time units? Naively, I would expect that also in the simulations shown in fig. S10 the allele dis becomes fixed and that therefore in the long run the results with and without recombination should be equal to each other. If this is indeed true, it should be clearly pointed out, and if it not true, then the reason should be made clear.

Early-dynamics vs. long term equilibrium
To improve clarity, we performed new simulations and modified the figures from the manuscript: we now run our simulations until equilibrium was reached (see S4 for more detail) for all figures except fig 2 and S10 where we specifically investigate the early dynamics of preference mutants.

Effect of the recombination rate when assuming a self-referencing architecture of preference (Hyp. 1).
Under self referencing (hyp. 1), the frequency of the disassortative allele dis at steady state does increase with the recombination rate. The major difference is observed when comparing simulations assuming null and non-null recombination rates between the preference locus M and the color pattern locus P.
Assuming self-referencing (hyp. 1), increasing recombination rate strongly promotes self-avoiding behavior. Selection generated by the genetic load associated to the color pattern alleles a and b promotes their linkage with the disassortative self-referencing allele dis, while the genetic-load free allele c is preferentially linked to the random mating allele r (fig. S11.a). Because the allele dis reaches a high frequency in the population, as soon as recombination occurs, it 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 a genetic load. In contrast, the \textit{c-dis} haplotype leads to the production a viable offspring that may benefit from mimicry. This is likely to explain why, under the self-referencing hypothesis (hyp. 1), recombination thus significantly increases the proportion of disassortative mating. We clarified this point in the manuscript (see lines 520-529, page 30)

\textit{Comparing the effect of the recombination rate when assuming either self-referencing (hyp. 1) or recognition/trait (hyp. 2) architecture of preference.}

Under self-referencing rule (hyp. 1), mate preference depends on the phenotype displayed by the female, so that the allele \textit{dis} always translates into a disassortative behavior. The self-referencing architecture (hyp. 1) is thus very similar to a single locus architecture, where a single pleiotropic gene controls both the mating cue and the rejection of this cue. 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 \textit{M}, independently from the color pattern of the female. Thus the physical linkage is important between mate preference and phenotype locus. By breaking these associations, recombination breaks may thus prevent the evolution of disassortative preferences.

In fig. 5, as in many place before, it is not clear to me whether we are here looking at transient dynamics or frequencies as they are obtained when the system has reached a point equilibrium. Is it really such that in fig. 5b a different point equilibrium is reached for rho=0.1 and 0.2 than for rho>0.2. Naively, I would expect that rho affects the speed of evolution but not so much the equilibrium frequencies. This needs to be clarified.

In fig. 5, we reported the frequencies obtained at equilibrium, assuming that the equilibrium was always reached after 2000 units of time. To make sure that equilibrium was reached in every simulations, we now use a threshold of variations in the numerical resolution, above which we consider that the genotype densities do not chance, and therefore that the equilibrium is reached (see supplementary 4 for more details). As you expected, the recombination does affect the speed of evolution but its effect on equilibrium frequencies is very limited.

Fig. S12 shows figures analogous to fig. 3. While in fig. 3 each column consists of two sub-columns, one for each habitat, this is not the case in figure S12. Why? I think something went wrong with the labelling along the x-axis in this figure. We modified fig S12 (now figure S11a) to show haplotype frequency within each of the two patches.

MINOR COMMENTS
Generally, the writing still contain many small mistakes. The below is such a small selection.
l. 3 acting on -> with respect to
II. 7-8 on mimetic color pattern -> either: on a mimetic color pattern, or: on mimetic color patterns
I. 97 promotes -> promote
I. 116 track down -> trackv
I. 151 decreased by 1-sigma or decreased by sigma?
I. 186 delete ) at the end
I. 190 Why two alleles. The following text specifies three alleles: r, sim, dis.
I. 195 I do not understand this explanation
I. 197 of choosen partner -> of the chosen partner
I. 267 I wonder whether Mortality would not be a more appropriate header for this section. After all, eq. (12) seems to describe a reduction in population size due to mortality.
No explanation is given for the functional form on the right-hand side of eq. (12). It seems here that baseline survival and survival due to genetic load act multiplicatively. Can this be motivated on biological grounds? I would have found it at least as intuitive if survival would have been given by 1-delta-delta_i, thus that the two mortality factors act additively.
We agree with the reviewer on this point: it makes more sense to assume that the two mortality factors act additively. We thus modified this part of the material and method accordingly.
I. 277 genetic densities -> genotype densities
I. 278 (see first Equation) -> (see equation (2))
II.277-279 Why does tracking genotype densities at a population level makes it an obvious choice that the four demographic processes occur at the same time? What do not see the logic behind the construction of this sentence.
It happens to different individuals
I. 285 range values -> maybe better: parameter intervals
I. 288 drawn from -> taken from
I. 289 and use discrete time step -> and by using discrete time steps
I. provide -> provides
I. 341 slighter -> weaker or smaller
I. 444 I do not follow why what is said in the previous sentence implies (therefore) what is said in this sentence.
I. 447 under attraction rule -> under the attraction rule
I. 480 for trait -> for the color trait (?)
Thank you for pointing these errors, we carefully corrected each of them.
Evolution and genetic architecture of disassortative mating at a locus under heterozygote advantage

Ludovic Maisonneuve¹,* , Mathieu Chouteau², Mathieu Joron³ and Violaine Llaurens¹

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 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 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 framwork 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 also 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). At 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 distinct 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 het-
108 erozygote 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 polymor-
110 phism. We then explore two putative genetic architectures for mate preferences based on (1) self
112 referencing and (2) based on a recognition/trait rule, and test for their respective impacts on the
114 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
116 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
118 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
120 preference. We assume a diploid species, so that each genotype contains four alleles.

The set of all possible four-allele genotypes is defined as $G = A_P \times A_P \times A_M \times A_M$ where $A_P,$
$A_M$ are the set of alleles at locus $P$ and $M$ respectively. A given genotype is then an quadruplet
122 of the form $(p_m, p_f, m_m, m_f)$ with $p_m \in A_P$ and $m_m \in A_M$ (resp. $p_f$ and $m_f$) being the alleles at
loci $P$ and $M$ on the maternal (resp. paternal) chromosomes. A recombination rate $\rho$ between
124 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
128 all $(i, n) \in G \times \{1, 2\}$ 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:

$$\forall (i, n) \in G \times \{1, 2\} \quad \frac{d}{dt} N_{i,n} = \text{Pred}_{i,n} + \text{Mort}_{i,n} + \text{Mig}_{i,n} + \text{Rep}_{i,n},$$

where $\text{Pred}_{i,n}$, $\text{Rep}_{i,n}$, $\text{Mig}_{i,n}$, and $\text{Mort}_{i,n}$ 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 $A_P = \{a, b, c\}$. 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 $\text{CP}$ the function translating each genotype $i$ into the corresponding color pattern phenotype $G$. For example, for all $(m_m, m_f) \in A_M \times A_M$, 


\[ CP((a,b,m_m,m_f)) = A \] because allele \( a \) is 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) ∈ \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 $d_{n-m}$ for the locally non-mimetic phenotypes (B and C in patch 1, A and C in patch 2). We therefore introduce parameters $d_{n-m}$ and $d_m$, with $d_{n-m} > d_m$, 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
where 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 $(i, n) \in G \times \{1, 2\}$, 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 ($\lambda$) 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 $\lambda$. For $(i, n) \in G \times \{1, 2\}$, 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:

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

where $\sum_j 1_{\{CP(i) = CP(j)\}} N_{j,n}$ 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 $\delta$. The recessive genetic loads $\delta_a$, $\delta_b$, $\delta_c$ associated with the respective alleles $a$, $b$ and $c$ limit the survival probabilities of homozygous genotypes at locus $P$.

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

$$\text{Mort}_{i,n} = -\left(\delta + (\mathbb{1}_{p_m=p_f=a}\delta_a + \mathbb{1}_{p_m=p_f=b}\delta_b + \mathbb{1}_{p_m=p_f=c}\delta_c)\right)N_{i,n}.$$

(5)

Migration

We assume a constant symmetrical migration rate $\text{mig}$ 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 $\text{mig}N_{i,n'}$ and those leaving the patch $\text{mig}N_{i,n}$:

$$\text{Mig}_{i,n} = \text{mig}N_{i,n'} - \text{mig}N_{i,n}.$$

(6)

where $\text{mig}$ is the migration coefficient $\text{mig} \in [0, 1]$.

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 $N_{\text{tot},n}$ the total density in patch $n$. The change in the total density due to reproduction is given by the logistic regulation function $r\left(1 - \frac{N_{\text{tot},n}}{K}\right)N_{\text{tot},n}$. Thus for $(i, n) \in G \times \{1,2\}$, the change in the density of genotype $i$ in patch $n$ generated by sexual reproduction is given by:

$$Rep_{i,n} = r\left(1 - \frac{N_{\text{tot},n}}{K}\right)N_{\text{tot},n}F_{i,n}. \quad (7)$$

The frequencies of each genotype in the progeny $(F_{i,n})_{i \in G}$ depends 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 $\text{sim}$ and (iii) disassortative $\text{dis}$ respectively (see fig. S5 for more details, $A_M = \{r, \text{sim}, \text{dis}\}$). We assume that the self-referencing preference alleles $\text{sim}$ and $\text{dis}$ are dominant to the random mating allele $r$ (see fig. S1 for more details). The dominance relationship between the $\text{sim}$ and $\text{dis}$ 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 $C$ ($A_M = \{m_r, m_a, m_b, m_c\}$). 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 $\rho$ 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 $(i, n) \in G \times \{1, 2\}$, 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 $(Pref_{i,J})_{(i,J) \in G \times \{A,B,C\}}$. These coefficients depend on the alleles at locus $M$ as detailed in the next section. For $(i, J) \in G \times \{A,B,C\}$ 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 $i \in G, n \in \{1, 2\}$, 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)): 
\[ T_{i,n} = \text{Pref}_{i, A} P_{A,n} + \text{Pref}_{i, B} P_{B,n} + \text{Pref}_{i, C} P_{C,n}, \]  

(8)

where for \( J \in \{A, B, C\} \), \( P_{J,n} = \frac{\sum_{i \in G} N_{i,n} A_{j(i)}}{\sum_{i \in G} N_{i,n}} \) 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 low (\( c_r = 0 \)), females have access to a large quantity of potential mates, so that their fertility 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 fertility 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 “fertility” \( M_{i,n} \) of a female with genotype \( i \) in patch \( n \):

\[ M_{i,n} = 1 - c_r + c_r T_{i,n}. \]  

(9)

We note \( \overline{M}_n \) the average fertility in patch \( n \) defined as

\[ \overline{M}_n = \sum_{i} f_{i,n} M_{i,n}, \]  

(10)

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

For \((j, k) \in G^2\), the quantity

\[ \frac{M_{j,n}}{\overline{M}_n}, \]  

(11)
is the relative probability that a female of genotype $j$ in patch $n$ has mated, and

$$\frac{\text{Pref}_{j,CP(k)} f_{k,n}}{T_{j,n}} = \frac{\text{Pref}_{j,CP(k)} f_{k,n}}{\text{Pref}_{j,A,n} P_{A,n} + \text{Pref}_{j,B,n} P_{B,n} + \text{Pref}_{j,C,n}}$$ \hspace{1cm} (12)$$

is the probability that a female of genotype $j$ in patch $n$ mates with a male of genotype $k$, depending on female preference and availability of males carrying genotype $k$. For $(i,n) \in G \times \{1,2\}$, the frequency of genotype $i$ in the progeny of the population living in patch $n$ is

$$F_{i,n} = \sum_{(j,k) \in G^2} \text{coef}(i,j,k,\rho) \times f_{j,n} \times \frac{M_{j,n}}{M_n} \times \frac{\text{Pref}_{j,CP(k)} f_{k,n}}{T_{j,n}}$$ \hspace{1cm} (13)$$

where $\text{coef}(i,j,k,\rho)$ 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 $\rho$ between the color pattern locus $P$ and the preference locus $M$ (see Supp. S1 for detailed expression of $\text{coef}(i,j,k,\rho)$). We checked that for all $n$ in $\{1,2\}$ the sum of all $F_{i,n}$ 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 $\lambda$ and migration rate $\text{mig}$ are chosen as conditions maintaining balanced polymorphism at the color pattern locus $P$, taken from (Joron and Iwasa, 2005).
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N_{i,n}$</td>
<td>Density of individuals with genotype $i$ in patch $n$</td>
</tr>
<tr>
<td>$\text{Pred}_{i,n}$</td>
<td>Change in the density of individuals with genotype $i$ caused by predation</td>
</tr>
<tr>
<td>$\text{Rep}_{i,n}$</td>
<td>Change in the density of individuals with genotype $i$ caused by reproduction</td>
</tr>
<tr>
<td>$\text{Mig}_{i,n}$</td>
<td>Change in the density of individuals with genotype $i$ caused by migration</td>
</tr>
<tr>
<td>$\text{Mort}_{i,n}$</td>
<td>Change in the density of individuals with genotype $i$ caused by mortality</td>
</tr>
<tr>
<td>$\text{CP}(i)$</td>
<td>Color pattern phenotype of individuals with genotype $i$</td>
</tr>
<tr>
<td>$\text{Pref}_{i,j}$</td>
<td>Preference of individuals with genotype $i$ towards individuals with phenotype $j$</td>
</tr>
<tr>
<td>$f_{i,n}$</td>
<td>Frequency of genotype $i$ in patch $n$</td>
</tr>
<tr>
<td>$P_{I,n}$</td>
<td>Frequency of phenotype $I$ in patch $n$</td>
</tr>
<tr>
<td>$T_{i,n}$</td>
<td>Probability that a female of genotype $i$ in patch $n$ accepts a male as mating partner during one mating encounter</td>
</tr>
<tr>
<td>$M_{i,n}$</td>
<td>Fertility of females with genotype $i$ in patch $n$</td>
</tr>
<tr>
<td>$\overline{M}_n$</td>
<td>Average female fertility in patch $n$</td>
</tr>
<tr>
<td>$F_{i,n}$</td>
<td>Frequency of genotype $i$ in the progeny of the population living in patch $n$</td>
</tr>
<tr>
<td>$P_{s-acc}$</td>
<td>Proportion of individuals expressing a self-accepting behavior</td>
</tr>
<tr>
<td>$P_{s-av}$</td>
<td>Proportion of individuals expressing a self-avoidance behavior</td>
</tr>
</tbody>
</table>

Table 1: Description of variables used in the model.
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
<th>Parameter interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_P$</td>
<td>Set of all possible alleles at locus $P$</td>
<td>{$a, b, c$}</td>
</tr>
<tr>
<td>$A_M$</td>
<td>Set of all possible alleles at locus $M$</td>
<td>{$r, sim, dis$} (hyp. 1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>{$m_r, m_a, m_b, m_c$} (hyp. 2)</td>
</tr>
<tr>
<td>$G$</td>
<td>Set of all possible genotypes</td>
<td>$A_P \times A_P \times A_M \times A_M$</td>
</tr>
<tr>
<td>$N_{tot,n}^0$</td>
<td>Initial population density in patch $n$</td>
<td>100</td>
</tr>
<tr>
<td>$d_m$</td>
<td>Basic predation mortality factor for individuals displaying the color pattern matching the local community</td>
<td>0.05</td>
</tr>
<tr>
<td>$d_{n\sim m}$</td>
<td>Basic predation mortality factor for individuals displaying a color pattern different from the local community</td>
<td>0.15</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>Unpalatability coefficient</td>
<td>0.0002</td>
</tr>
<tr>
<td>$mig$</td>
<td>Migration rate</td>
<td>[0,1]</td>
</tr>
<tr>
<td>$\rho$</td>
<td>Recombination rate</td>
<td>[0, 0.5]</td>
</tr>
<tr>
<td>$r$</td>
<td>Growth rate</td>
<td>1</td>
</tr>
<tr>
<td>$K$</td>
<td>Carrying capacity within each patch</td>
<td>2000</td>
</tr>
<tr>
<td>$\delta$</td>
<td>Baseline mortality rate</td>
<td>0.1</td>
</tr>
<tr>
<td>$\delta_i$</td>
<td>Genetic load linked to allele $i$</td>
<td>[0, 1]</td>
</tr>
<tr>
<td>$c_r$</td>
<td>Relative cost of choosiness</td>
<td>[0, 1]</td>
</tr>
</tbody>
</table>

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

**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 $A_P \times A_M$ containing two alleles located on the same chromosome or inherited from the same parent. We then calculate haplotype frequencies in patch $n$ ($f_{haplo}^{p,m,n} \in A_P \times A_M$ for $n \in \{1, 2\}$). Then for $(p, m, n) \in A_P \times A_M \times \{1, 2\}$, the frequency of haplotype $(p, m)$ in patch $n$ is given by:

$$f_{haplo}^{p,m,n} = \frac{\sum_{i=(p_{nu},p_{mf},m_{nu},m_{mf})} N_{i,n} \left( \frac{1}{2} \mathbb{1}_{\{p_{nu}=p\}} \mathbb{1}_{\{m_{nu}=m\}} + \frac{1}{2} \mathbb{1}_{\{p_{mf}=p\}} \mathbb{1}_{\{m_{mf}=m\}} \right)}{\sum_{i=(p_{nu},p_{mf},m_{nu},m_{mf})} N_{i,n}}.$$  

(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, $P_{s-acc}$ (see equation (15)) and $P_{s-avr}$ (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.

\[ P_{s-acc} = \sum_{i \in G} f_i \text{Pref}_{i, CP(i)}, \quad (15) \]

\[ P_{s-av} = \sum_{i \in G} f_i (1 - \text{Pref}_{i, CP(i)}). \quad (16) \]

**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 \( m_i \) 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 those 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 is reduced, and selection acting against disassortative preferences is reduced.

*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 $B$
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 $r = 1$, $K = 2000$, $N_{tot,1}^0 = N_{tot,2}^0 = 100$, $\lambda = 0.0002$, $d_m = 0.05$, $d_n = 0.15$, $\rho = 0$, $c_r = 0.1$, $\delta_a = \delta_b = \delta_c = 0$ and $\delta = 0.1$.

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 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 $\text{dis}$ (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 $c$ ($\delta_c$) 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 $\text{dis}$ 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 $\text{dis}$ (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 $\text{dis}$ 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 $\text{dis}$ is never fixed within populations. This suggests that the heterozygote advantage at locus $P$ allowing the emergence
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 \( c \), \( \delta_c \). 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 \(\text{dis}\) 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 \(c, \delta_c\). The initial frequency of allele \(\text{dis}\) 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 \(r = 1, K = 2000, N_{\text{tot},1}^0 = N_{\text{tot},2}^0 = 100, \lambda = 0.0002, d_m = 0.05, d_{n-m} = 0.15, \text{mig} = 0.1, \rho = 0\) and \(c_r = 0.1\).

Of disassortative mating decreases when this behavior is common in the population. The \(\text{dis}\) mutant nevertheless reaches high frequencies when the genetic load associated with the recessive allele \(c\) is lower than the genetic load associated with dominant alleles \(a\) and \(b\) (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 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 $r$, probably because the cost of choosiness limits the association with the preference allele in absence of heterozygote advantage.

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, allele $b$ being lost. (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
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 \(r = 1\), \(K = 2000\), \(N_0^1 = N_0^2 = 100\), \(\lambda = 0.0002\), \(d_m = 0.05\), \(d_{n-m} = 0.15\), \(\rho = 0\), \(mig = 0.1\), \(\delta_c = 0\), \(\delta = 0.1\) and \(c_r = 0.1\).

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)).

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 \( P_{\text{self-acc}} \) (in purple) and self-avoidance \( P_{\text{self-av}} \) (in blue) obtained at equilibrium after the introduction of preference alleles are shown for different values of the level of genetic load of \( \delta_a \) and \( \delta_c \). Simulations are run assuming \( r = 1, K = 2000, N_{\text{tot,1}}^0 = N_{\text{tot,2}}^0 = 100, \lambda = 0.0002, d_m = 0.05, d_{n-m} = 0.15, \rho = 0, mig = 0.1, \delta_c = 0, \delta = 0.1 \) and \( c_r = 0.1. \)
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) \( (P_{s-av} \approx 48\%) \) (fig. 5(a) and 5(c)).

By contrast, when we consider preference alleles leading to attraction (hyp. 2.a), the disassortative behavior is scarcer at equilibrium \( (P_{s-av} \approx 36\%) \) (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 color pattern locus \( P \) (with one allele \( a \) and another allele) and at the preference locus \( M \) (with one preference allele coding for attraction toward phenotype \( A \) and another preference allele triggering attraction toward another color pattern phenotype). 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.a 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 impact the speed of evolution of disassortative mating (fig. S10 for the proportions of behaviors during the emergence of disassortative mutants), 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 favored by *self-referencing* (hyp. 1) or by *rejection* (hyp. 2.b) 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-av} ≈ 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 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 ρ = 0 vs ρ >
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 $P_{s-acc}$ (in purple) and self-avoidance $P_{s-av}$ (in blue) obtained at equilibrium are shown for different values of recombination rate $r$ between the preference locus $M$ and the color pattern locus $P$. Simulations are run assuming $r = 1$, $K = 2000$, $N_{tot,1}^0 = N_{tot,2}^0 = 100$, $\lambda = 0.0002$, $d_m = 0.05$, $d_{n-m} = 0.15$, $mig = 0.1$, $\delta_a = \delta_b = 0.5$, $\delta_c = 0$, $\delta = 0.1$ and $c_r = 0.1$.

0: self-avoidance behavior is scarcely observed in the population ($P_{s-av} \approx 1\%$) when there is recombination ($r > 0$).

Our results suggests that single-locus architecture or tight linkage between the cue and the preferences are likely to promote the evolution of disassortative mating. The self-referencing architecture (hyp. 1) generates an immediate rejection of matching phenotype: it is thus very similar
to a single locus architecture, where a single pleiotropic gene controls both the mating cue and the rejection of this cue. The negative effect of recombination on the evolution of disassortative behavior observed when assuming recognition/trait also highlight that the linkage between color pattern and preference locus also promotes the emergence of disassortative mating.

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. 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. Contrastingly, our model suggests that the genetic architecture of disassortative mating might differ from those documented in species showing assortative mating behavior.

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.a), 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 effect of dominance at the color pattern locus has a variable impact on the evolution of disassortative mating depending on the genetic architecture of preference. Under both recognition/trait rules, mate choice is based on the phenotype of the chosen individual, so dominance relationships at the color pattern locus influence the evolution of disassortative mating. Nevertheless, under self-referencing mate choice also depends on the phenotype of the choosing individual, so that dominance at the color pattern locus of both the choosing and chosen individuals determines the choice. Under recognition/trait, however, mate choice does not depend on the phenotype of the choosing individual, but on the dominance relationships at the mate preference locus, allowing for different types of preference to emerge, including individuals reproducing with different phenotypes only, or individuals mating with either similar or different phenotypes.

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 below, 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 density 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). Similarly, the model of Karlin and Feldman (1968) suggests that disassortative mating slows down the purge of deleterious alleles. 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 archi-
tecture: 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 feedback 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


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.


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 $\text{coef}(g^O, g^M, g^F, \rho)$ summarizing the Mendelian segregation of alleles assuming two diploid loci and a rate of recombination $\rho$ between these loci. Let $g^O = (p_m^O, p_f^O, m_m^O, m_f^O), g^M = (p_m^M, p_f^M, m_m^M, m_f^M)$ and $g^F = (p_m^F, p_f^F, m_m^F, m_f^F)$ be the offspring, maternal and paternal genotypes respectively, all in $G$. 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. $\text{coef}(g^O, g^M, g^F, \rho)$ 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 $\rho$.

Each diploid mother can produce four types of haploid gametes containing alleles $(p_m^M, m_m^M)$, $(p_f^M, m_f^M)$, $(p_m^M, m_f^M)$ or $(p_f^M, m_m^M)$, in proportion $\frac{1 - \rho}{2}, \frac{1 - \rho}{2}, \rho$ and $\frac{\rho}{2}$ respectively. Then the proportion of gametes with alleles $(p, m) \in A_p \times A_m$ produced by the mother is given by the function $\text{coef}_{\text{haplotype}}(p, m, g^M, \rho)$, where

$$\text{coef}_{\text{haplotype}}(p, m, g^M, \rho) = \frac{1 - \rho}{2} 1\{p = p_m^M\} 1\{m = m_m^M\} + \frac{1 - \rho}{2} 1\{p = p_f^M\} 1\{m = m_f^M\}$$

$$+ \frac{\rho}{2} \frac{1 - \rho}{2} 1\{p = p_f^M\} 1\{m = m_m^M\} + \frac{\rho}{2} \frac{1 - \rho}{2} 1\{p = p_m^M\} 1\{m = m_f^M\}.$$

Similarly, each diploid father can produce four types of haploid gametes. The proportion of genotype $(p, m) \in A_p \times A_m$ in the gametes of a given father is given by the function $\text{coef}_{\text{haplotype}}(p, m, g^F, \rho)$.

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 $\rho$ is given by:

$$\text{coef}(g^O, g^M, g^F, \rho) = \text{coef}_{\text{haplotype}}(p_m^O, m_m^O, g^M, \rho) \text{coef}_{\text{haplotype}}(p_f^O, m_f^O, g^F, \rho).$$
**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 occurring 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:

\[
\sum_{i \in \mathcal{G}} F_{i,n} = \sum_{i \in \mathcal{G}} \sum_{(j,k) \in \mathcal{G}^2} \text{coef}(i, j, k, \rho) f_{j,n} \frac{M_{n,i} \text{Pref}_{i,\mathcal{CP}(k)}}{M_n} \frac{f_{k,n}}{T_{j,n}},
\]

\[
= \sum_{(j,k) \in \mathcal{G}^2} f_{j,n} \frac{M_{n,i} \text{Pref}_{i,\mathcal{CP}(k)} f_{k,n}}{M_n} \sum_{i \in \mathcal{G}} \text{coef}(i, j, k, \rho),
\]

\[
= \sum_{j \in \mathcal{G}} f_{j,n} \frac{M_{n,i} \sum_{k \in \mathcal{G}} \text{Pref}_{i,\mathcal{CP}(k)} f_{k,n}}{M_n} T_{j,n},
\]

\[
= \sum_{j \in \mathcal{G}} f_{j,n} \frac{M_{n,i} \sum_{k \in \mathcal{G}} \{1_{\{\mathcal{CP}(k)=A\}} + 1_{\{\mathcal{CP}(k)=B\}} + 1_{\{\mathcal{CP}(k)=C\}}\} \text{Pref}_{i,\mathcal{CP}(k)} f_{k,n}}{T_{j,n}}
\]

because \( \forall k \in \mathcal{G}, \quad 1_{\{\mathcal{CP}(k)=A\}} + 1_{\{\mathcal{CP}(k)=B\}} + 1_{\{\mathcal{CP}(k)=C\}} = 1, \)

\[
= \sum_{j \in \mathcal{G}} f_{j,n} \frac{M_{n,i} \text{Pref}_{i,A} \sum_{k \in \mathcal{G}} 1_{\{\mathcal{CP}(k)=A\}} f_{k,n} + \text{Pref}_{i,B} \sum_{k \in \mathcal{G}} 1_{\{\mathcal{CP}(k)=B\}} f_{k,n} + \text{Pref}_{i,C} \sum_{k \in \mathcal{G}} 1_{\{\mathcal{CP}(k)=C\}} f_{k,n}}{T_{j,n}},
\]

\[
= \sum_{j \in \mathcal{G}} f_{j,n} \frac{M_{n,i} \text{Pref}_{i,A} P_{A,n} + \text{Pref}_{i,B} P_{B,n} + \text{Pref}_{i,C} P_{C,n}}{T_{j,n}},
\]

\[
= \sum_{j \in \mathcal{G}} f_{j,n} \frac{M_{n,i} T_{j,n}}{T_{j,n}},
\]

\[
= \sum_{j \in \mathcal{G}} f_{j,n} \frac{M_{n,i}}{M_n},
\]

\[
= \frac{M_n}{M_n} = 1.
\]
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 an 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 \(\Delta t\) being the step time in our simulations.

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

\[
\frac{N_{i,n}^{t+\Delta t} - N_{i,n}^t}{\Delta t} = \text{Pred}_{i,n}^t + \text{Mort}_{i,n}^t + \text{Mig}_{i,n}^t + \text{Rep}_{i,n}^t.
\]

This equation is equivalent to:

\[
N_{i,n}^{t+\Delta t} = N_{i,n}^t + \Delta t \left( \text{Pred}_{i,n}^t + \text{Mort}_{i,n}^t + \text{Mig}_{i,n}^t + \text{Rep}_{i,n}^t \right).
\]

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 $Var_t^d$ quantifying the change in the numerical solution:

$$Var_t^d = \left( \sum_{(i,n) \in \Omega \in \{1,2\}} \left( \frac{N_{i,n}^{t+\Delta t} - N_{i,n}^t}{\Delta t} \right)^2 \right)^{1/2}.$$

When $\frac{Var_t^d}{N_{\text{tot}} < 10^{-5}}$, we assume that the dynamical system has reached equilibrium, with $N_{\text{tot}}$ being the total density in both patches.
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 $\text{dis}$ allele display disassortative mating, and mate preferentially with individuals with a color pattern different from their own. 4. Butterflies carrying a $\text{sim}$ 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.
Figure S8: Evolution of the proportion of a mutant dis in the population immediately after its introduction, using simulations with three different time units (Δt = 1 in blue, Δt = 0.1 in orange or Δt = 0.01 in green), under the self-referencing hypothesis (hyp. 1). All simulations give similar dynamics, assuming (a) δ_a = δ_b = 0.5, δ_c = 0 or (b) δ_a = δ_b = δ_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_{tot,1}^0 = N_{tot,2}^0 = 100, λ = 0.0002, d_m = 0.05, d_{m-n} = 0.15, ρ = 0, mig = 0.1, δ = 0.1 and c_r = 0.1.
<table>
<thead>
<tr>
<th>δ₁ = δ₂</th>
<th>δ₃</th>
<th>Population 1</th>
<th>Population 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Proportion of morph A</td>
<td>Proportion of morph B</td>
</tr>
<tr>
<td>0.00</td>
<td>0.00</td>
<td>90.5 %</td>
<td>9.5 %</td>
</tr>
<tr>
<td>0.00</td>
<td>0.25</td>
<td>90.5 %</td>
<td>9.5 %</td>
</tr>
<tr>
<td>0.00</td>
<td>0.50</td>
<td>90.5 %</td>
<td>9.5 %</td>
</tr>
<tr>
<td>0.00</td>
<td>1.00</td>
<td>90.5 %</td>
<td>9.5 %</td>
</tr>
<tr>
<td>0.25</td>
<td>0.00</td>
<td>61.8 %</td>
<td>7.7 %</td>
</tr>
<tr>
<td>0.25</td>
<td>0.25</td>
<td>78.8 %</td>
<td>17.9 %</td>
</tr>
<tr>
<td>0.25</td>
<td>0.50</td>
<td>80.5 %</td>
<td>17.8 %</td>
</tr>
<tr>
<td>0.25</td>
<td>1.00</td>
<td>81.6 %</td>
<td>17.8 %</td>
</tr>
<tr>
<td>0.50</td>
<td>0.00</td>
<td>54.5 %</td>
<td>5.7 %</td>
</tr>
<tr>
<td>0.50</td>
<td>0.25</td>
<td>76.3 %</td>
<td>18.6 %</td>
</tr>
<tr>
<td>0.50</td>
<td>0.50</td>
<td>78.7 %</td>
<td>18.7 %</td>
</tr>
<tr>
<td>0.50</td>
<td>1.00</td>
<td>80.2 %</td>
<td>18.5 %</td>
</tr>
<tr>
<td>1.00</td>
<td>0.00</td>
<td>49.9 %</td>
<td>4.8 %</td>
</tr>
<tr>
<td>1.00</td>
<td>0.25</td>
<td>74.8 %</td>
<td>18.9 %</td>
</tr>
<tr>
<td>1.00</td>
<td>0.50</td>
<td>77.5 %</td>
<td>19.1 %</td>
</tr>
<tr>
<td>1.00</td>
<td>1.00</td>
<td>79.3 %</td>
<td>19.0 %</td>
</tr>
</tbody>
</table>

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 \( a \) (\( \delta_a \)), intermediate-dominant allele \( b \) (\( \delta_b \)) and recessive allele \( c \) (\( \delta_c \)). Simulations are run assuming \( r = 1, K = 2000, N_{tot,1}^0 = N_{tot,2}^0 = 100, \lambda = 0.0002, d_m = 0.05, d_{m-m} = 0.15, \rho = 0, mig = 0.1, \delta = 0.1 \) and \( c_r = 0.1 \).
Figure S10: Influence of a genetic load on the distribution of mating behavior at the population level, shortly after the introduction of the mutants, 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 $P_{s-acc}$ (in purple) and self-avoidance $P_{s-av}$ (in blue) obtained 100 time units after the introduction of preference alleles are shown for different values of genetic load $\delta_a$ and $\delta_b$. Simulations are run assuming $r = 1$, $K = 2000$, $N_{tot,1}^0 = N_{tot,2}^0 = 100$, $\lambda = 0.0002$, $d_m = 0.05$, $d_{n-m} = 0.15$, $p = 0$, $mig = 0.1$, $\delta_c = 0$, $\delta = 0.1$ and $c_r = 0.1$. 

![Figure S10 Parts a, b, and c](image-url)
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 $\rho$ between the preference locus $M$ and the color pattern locus $P$. For each value of recombination rate ($\rho$) the first and second bars represented haplotype proportions in the populations living in the patch 1 and 2 respectively. Simulations are run assuming $r = 1$, $K = 2000$, $N_{tot,1}^0 = N_{tot,2}^0 = 100$, $\lambda = 0.0002$, $d_m = 0.05$, $d_{n-m} = 0.15$, $mig = 0.1$, $\delta_d = \delta_b = 0.5$, $\delta_c = 0$, $\delta = 0.1$ and $c_r = 0.1$. 