The limits of evolutionary convergence in sympatry: reproductive interference and developmental constraints leading to local diversity in aposematic signals

Ludovic Maisonneuve[*,1], **Charline Smadi**[†] **and Violaine Llaurens**[*]

[*]Institut de Systématique, Evolution, Biodiversité (ISYEB), Muséum national d'Histoire naturelle, CNRS, Sorbonne Université, EPHE, Université des Antilles CP 50, 57 rue Cuvier, 75005 Paris, France, [†]Univ. Grenoble Alpes, INRAE, LESSEM, F-38402 St-Martin-d'Hères, France and Univ. Grenoble Alpes, CNRS, Institut Fourier, 38610 Gières, France.

ABSTRACT Mutualistic interaction between defended species is a striking case of evolutionary convergence in sympatry, driven by the increased protection against predators brought by mimicry. However, such convergence is often limited: sympatric defended species frequently display different or imperfectly similar warning traits. The phylogenetic distance between sympatric species may indeed prevent evolution towards the exact same signal. Moreover, warning traits are also implied in mate recognition, so that trait convergence might result in heterospecific courtship and mating. Here, we investigate the strength and direction of convergence in warning trait in defended species with different ancestral traits, using a mathematical model. We specifically determine the effect of phenotypic distances among ancestral traits of converging species, and costs of heterospecific sexual interactions on imperfect mimicry and trait divergence. Our analytical results confirm that reproductive interference limits the convergence of warning trait, leading to either imperfect mimicry or complete divergence. More surprisingly, our model pinpoints that reproductive interference can change the direction of convergence depending on the relative species densities. We also show that reproductive interference can generate imperfect mimicry only between species with different ancestral traits. Our model therefore highlights that convergence triggered by Müllerian mimicry not only depends on relative defence levels, but that relative species densities, heterospecific sexual interactions and ancestral traits interfere in the direction and strength of convergence between species.

KEYWORDS Mimicry; Reproductive interference; Convergence; Imperfect mimicry, Quantitative genetics

Introduction

Evolutionary convergence of traits between species is often promoted by natural selection, but trait similarity between species may be impaired by contrasted developmental and ecological constraints encountered by converging species. Large phylogenetic distances between species, as well as strikingly different ancestral traits may indeed affect the level of convergence. Furthermore, when trait convergence occurs in sympatry, traits similarity may generate costs linked to reproductive interference, counter-selecting trait resemblance between species. Here we investigate the limits of evolutionary convergence in sympatry by focusing on the striking case of mutualistic interactions.

Mutualistic interactions frequently provoke convergent evolution of different traits in sympatric species, but the ecological factors determining the strength and direction of convergence are still largely unknown. For example, avian vocal resemblance allows mixed-species flocks that has been suggested to benefit to all species (Goodale and Kotagama 2006); trait similarity between sympatric species may promote pollinator attraction in nectar rewarding flowers (Schemske 1981; Benitez-Vieyra et al. 2007). But the convergence in warning trait among defended species is certainly the most studied: the costs associated with predators' learning is indeed shared among all individuals dis-

Manuscript compiled: Friday 22nd January, 2021
[1]Corresponding author: ludovic.maisonneuve.2015@polytechnique.org.

playing the same trait, whatever their species (*e.g.* in plants (Lev-Yadun 2009), mollusks (Cortesi and Cheney 2010), vertebrates (Sanders et al. 2006; Springer and Smith-Vaniz 1972) and insects (Mallet and Gilbert Jr. 1995)). Field experiments report an intense selection exerted by predators favoring warning trait convergence in sympatry (Benson 1972; Mallet and Barton 1989; Kapan 2001; Arias et al. 2016; Chouteau et al. 2016). Surprisingly, despite this intense selection, sympatric defended species frequently exhibit only an imperfect resemblance (*e.g.* (Savage and Slowinski 1992)) or different warning traits (*e.g.* (Beccaloni 1997)).

The level of trait convergence between sympatric species can vary because of other developmental or selective constraints that may differ among species. For example diet (Grill and Moore 1998; Ojala et al. 2007) has been shown to influence the warning trait, and can strongly differ among sympatric species, leading to different species-specific trade-offs limiting convergence. As highlighted by the significant phylogenetic signal observed on the warning trait of mimetic butterflies of the tribe *ithomiines* (Elias et al. 2008; Chazot et al. 2014), developmental and historical constraints may indeed influence the evolution of warning traits. Theoretical studies suggest that ancestral trait may play a key role in the evolution of warning trait because the convergence of trait can be facilitated by an initial resemblance between species (Franks and Noble 2004; Balogh and Leimar 2005; Franks and Sherratt 2007). Other theoretical studies highlight that the level of standing genetic and phenotypic variance within species strongly influences convergence between species (Ruxton et al. 2008). The balance between (1) predation pressure shared by individuals from different sympatric species and (2) the developmental or selective constraints within each species may thus strongly shape the evolutionary convergence in warning traits, and especially its direction. The direction of convergence depends on the value of ancestral traits and the respective level of phenotypic evolution in the different sympatric species. While strict convergence (Sherratt 2008) involves trait resemblance jointly evolving in several sympatric species (*e.g.* Symula et al. (2001); Flanagan et al. (2004)), phenotypic evolution can also be more marked in some species (*i.e.* the 'mimic' species) leading to high resemblance to the ancestral trait displayed in other, usually better defended, species (*i.e.* the 'model' species) (see (Dalziell and Welbergen 2016) for the terminology). The direction of convergence might thus depend on the relative levels of defence between sympatric species shaping the intensity of selection favoring similarity in the mimetic species, acting together with the phenotypic distances between ancestral traits of each sympatric species, and the levels of constraints on those traits.

Moreover, the convergence of warning trait in different species may imply costs due to behavioral interference, therefore limiting this convergence. Mimetic traits are indeed frequently involved in mate choice (Jiggins et al. 2001; Naisbit et al. 2001; Kronforst et al. 2006; Merrill et al. 2014), leading to risk of confusion during sexual interactions. Species sharing similar aposematic trait may thus be exposed to substantial reproductive interference (named RI hereafter) implying fitness costs during mate acquisition due to interspecific interactions, including heterospecific courtship and mating as well as heterospecific male rivalry (Gröning and Hochkirch 2008). Such RI has been observed between species sharing similar aposematic traits (in *Heliconius* and *Mechanitis* species (Estrada and Jiggins 2008; Vasconcellos-Neto and Brown 1982)).

RI can generate reproductive character displacement (named RCD hereafter) (Gröning and Hochkirch 2008; Kyogoku 2015), whereby reproductive traits are more dissimilar between species in sympatric than in allopatric populations (Brown and Wilson 1956). Such RCD may thus impair convergence driven by mutualistic interactions. Theoretical studies have investigated how the evolution of females preferences may promote RCD in males (McPeek and Gavrilets 2006; Yamaguchi and Iwasa 2013): RI costs are predicted to favor the divergence of females preferences, therefore promoting the divergence of male reproductive traits through sexual selection. Nevertheless, the effect of RI on trait divergence is not straightforward when those traits are under strong selection promoting convergence. The outcomes of these antagonistic selective forces might result in trait divergence or limited convergence.

Here, we investigate the ecological factors limiting the convergence of traits involved in mutualistic interactions. We build a mathematical model, describing the evolution of quantitative traits in a focal species involved in mutualistic interactions with a fixed community of defended species. We specifically study the evolution of (1) the quantitative trait t involved in mutualistic interaction, displayed in both males and females and (2) the preference of females for the different values of this trait p. We assume that individuals gain protection against predators from the similarity of their warning trait t toward the trait displayed in the local community of defended species. However, trait similarity between species generates fitness costs of RI paid by females (McPeek and Gavrilets 2006; Yamaguchi and Iwasa 2013). We also investigate the effect of developmental and selective constraints by testing the effect of the ancestral trait value t_o on trait convergence. Using weak selection approximation (Barton and Turelli 1991; Kirkpatrick et al. 2002), we obtain equations describing the evolution of the mean trait and mean preference values. We then use analytical results and numerical analyses to investigate the effect of RI on the convergence of trait depending on different ecological factors. Finally, to specifically investigate the effect of RI and of ancestral trait values on the direction of convergence, we then extend our model to two interacting species, exhibiting different densities, individual levels of defence and ancestral trait values, and explore the joint evolution of displayed trait and females preference in both species.

Evolution of mimicry toward a fixed community of defended species

Methods

We consider a single species (refered to as *focal* species hereafter) living in sympatry with a community of defended species (referred to as the *local mimicry ring* hereafter). In the *local mimicry ring*, all individuals display a similar warning signal : we note $\bar{t}'$ the mean value of the trait displayed in the *local mimicry ring*, assumed to be fixed. We then investigate the evolution of a trait within the *focal* species, influenced by both predators' behavior (promoting mimicry toward the local mimicry ring) and mate choice. This trait t is expressed in both males and females and we assume that females only may express a mating preference for males with trait value close to p. Individuals of the *focal* species benefit from increased survival when they display a trait similar to the *local mimicry ring* (mean trait value in the community, $\bar{t}'$) because of predators warning signal learning. This resemblance toward the *local mimicry ring may* induce costs for individuals

from the *focal* species, because of RI. We thus assume increasing fitness costs due to RI, when the trait displayed by males from the *focal* species and trait displayed in the *local mimicry ring* becomes more similar (see Equation (8)).

We assume that the trait and preference in the *focal* species are quantitative traits, with an autosomal, polygenic basis, and additive effects. We consider discrete and non-overlapping generations. We assume weak selection, therefore we can approximate the change in the mean values of trait $\bar{t}$ and preference in the *focal* species $\bar{p}$ by (Iwasa et al. 1991):

$$\begin{pmatrix} \Delta\bar{t} \\ \Delta\bar{p} \end{pmatrix} = \frac{1}{2} \begin{pmatrix} G_t & C_{tp} \\ C_{tp} & G_p \end{pmatrix} \begin{pmatrix} \beta_t \\ \beta_p \end{pmatrix}, \tag{1}$$

where β_t and β_p describe the selective forces acting on the trait t and the preference p, respectively (defined below). G_t and G_p are the genetic variances of the trait t and the preference p, and C_{tp} is the genetic covariance between t and p in the *focal* species. Following (Iwasa et al. 1991), we assume that G_t and G_p are positive constants maintained by an equilibrium between selection and recurrent mutations. Under weak selection and preference, the genetic covariance between t and p, C_{tp}, is small in comparison with G_t and G_p (see Appendix 1.2), allowing us to approximate the change in the mean values of trait and preference by:

$$\begin{pmatrix} \Delta\bar{t} \\ \Delta\bar{p} \end{pmatrix} \approx \frac{1}{2} \begin{pmatrix} G_t & 0 \\ 0 & G_p \end{pmatrix} \begin{pmatrix} \beta_t \\ \beta_p \end{pmatrix} = \frac{1}{2} \begin{pmatrix} G_t\beta_t \\ G_p\beta_p \end{pmatrix}. \tag{2}$$

The selection vector is defined by:

$$\beta_t := \frac{\partial}{\partial t} \ln W_m(t|\bar{t},\bar{p},\bar{t}') + \frac{\partial}{\partial t} \ln W_f(t,p|\bar{t},\bar{t}')\Big|_{(t,p)=(\bar{t},\bar{p})}, \tag{3}$$

$$\beta_p := \frac{\partial}{\partial p} \ln W_f(t,p|\bar{t},\bar{t}')\Big|_{(t,p)=(\bar{t},\bar{p})}, \tag{4}$$

where W_m and W_f denote the fitness of males and females respectively, and will be defined in (5) and (6). Another important parameter is the trait value t_o in the *focal* species, representing the optimal trait value in this species, without interaction with other species. This optimal value is assumed to be shaped by developmental as well as selective constraints, specific to the *focal* species. The fitness of a male displaying a trait value t then depends on (1) sexual selection exerted by females preferences, (2) natural selection due to mimetic protection against predation, as well as (3) departure from the optimal trait value t_o, inducing a selective cost s. It is thus given by:

$$W_m(t|\bar{t},\bar{p},\bar{t}') = \exp\left[-a(t-\bar{p})^2\right] W_{pred}(t|\bar{t},\bar{t}') \exp\left[-s(t-t_o)^2\right], \tag{5}$$

where W_{pred} is the fitness component due to predation (see Equation (7) below). Females choosiness a, assumed constant among females, quantifies how much do females reject males with a non-preferred trait. We assume a being of order ε, with ε small, in line with the weak selection hypothesis and that $a > 0$. Under weak selection, the distributions of trait and preference values in the *focal* species are concentrated around the mean values $\bar{t}$ and $\bar{p}$ respectively (maintained by selection mutation equilibrium, see Appendix in (Iwasa et al. 1991)). The term $\exp\left[-a(t-\bar{p})^2\right]$ representing the effect of sexual selection therefore depends only on the mean value of females preference and not on the distribution of preferences among females. We assume a selective coefficient s of order ε^2, because we focus on the specific case where selective pressures due to selective constraints, predation and RI are of the same order of magnitude (see Appendix 1.1).

Females fitness is then affected by natural selection depending on the displayed trait value t, but also by the metabolic costs generated by the investment in reproduction. Females indeed suffer from a RI cost when mating with an heterospecific male. Females also pay a cost for seeking their preferred partners, increasing with the strength of their choosiness a.

The fitness of a female with trait and preference values t and p is given by:

$$W_f(t,p|\bar{t},\bar{t}') = \exp\left[-a(p-\bar{t})^2\right] W_{RI}(p|\bar{t},\bar{t}') W_{pred}(t|\bar{t},\bar{t}') \times \exp\left[-s(t-t_o)^2\right], \tag{6}$$

where W_{RI} describes the effect of RI on fitness (see Equation (8)). The term $\exp\left[-a(p-\bar{t})^2\right]$ describes the fitness cost due to females choosiness.

Predation depending on the level of mimicry of trait t. The evolution of the trait t, expressed by males and females from the *focal* species, is strongly influenced by the trait displayed by the individuals from the *local mimicry ring* (mean value $\bar{t}'$). Müllerian mimicry indeed generates positive density-dependent selection (Benson 1972; Mallet and Barton 1989; Chouteau et al. 2016), due to predators learning. This density-dependence is non linear and is often modelled as an hyperbolic decrease in mortality (see (Joron and Iwasa 2005; Llaurens et al. 2013) for example). The density-dependence is modulated by the individual defence level λ, shaping predator deterrence: the higher the defence, the quicker predators do learn. We note $\bar{\lambda}'$ the mean defence level of an individual in the *local mimicry ring*. The protection gained against predators then depends on the level of resemblance among defended prey, as perceived by predators and the number of individuals sharing the same signal. We note N and N' the densities of individuals in the *focal* species and in the *local mimicry ring*, respectively. We note $d \in (0,1)$ the basal predation rate and we assume this parameter to be of order ε, with ε small, in line with the weak selection hypothesis. The impact of predation on the fitness of an individual displaying the trait value t is given by:

$$W_{pred}(t|\bar{t},\bar{t}') = 1 - \frac{d}{1 + \lambda \exp\left[-b(t-\bar{t})^2\right]N + \bar{\lambda}' \exp\left[-b(t-\bar{t}')^2\right]N'}, \tag{7}$$

where $\exp\left[-b(t-\bar{t}')^2\right]$ (resp. $\exp\left[-b(t-\bar{t})^2\right]$) describes how much predators perceive the trait values t and $\bar{t}'$ (resp. $\bar{t}$) as similar. The predators discrimination coefficient b thus quantifies how much predators discriminate different trait values. We assume that this parameter is of the same order of magnitude as females choosiness a (*i.e* of order ε).

Cost induced by RI. Because males of the *local mimicry ring* may resemble males of the *focal* species, females will suffer from RI generated by erroneous mating attempts with heterospecific males (see (Gröning and Hochkirch 2008) for a review of RI costs). The risk of heterospecific mating depends on the relative densities of heterospecific and conspecific males. The cost of RI is thus assumed to be density-dependent in the model. We assume a balanced sex-ratio within both the *focal* species and the

local mimicry ring. Using Equation (1b) in (Yamaguchi and Iwasa 2013) the fitness of a female with preference p is modulated by:

$$W_{RI}(p|\bar{t},\bar{t}') = \frac{\exp\left[-a(p-\bar{t})^2\right]N/2}{\exp\left[-a(p-\bar{t})^2\right]N/2 + c_{RI}\exp\left[-a(p-\bar{t}')^2\right]N'/2},$$
$$= \frac{\exp\left[-a(p-\bar{t})^2\right]N}{\exp\left[-a(p-\bar{t})^2\right]N + c_{RI}\exp\left[-a(p-\bar{t}')^2\right]N'}, \quad (8)$$

where c_{RI} is the strength of RI, assumed of order ε, with ε small, in line with the weak selection hypothesis.

All variables and parameters used in the model are summed up in Table 1.

Model exploration. We assume that the *focal* species is ancestrally not in contact with the *local mimicry ring*, and therefore the initial mean trait value in the *focal* species is equal to the optimal trait t_o. We also assume that mean female preference value is initially equal to mean trait value. At the initial time, we assume that the *focal* species enters in contact with the *local mimicry ring*. The dynamics of trait and preference values then follow Equation (2).

Analytical resolution assuming weak females choosiness ($a = O(\varepsilon)$) and predators discrimination ($b = O(\varepsilon)$). Assuming that a and b are of order ε and s is of order ε^2 (hypothesis H_w), we analytically determine the mean trait and preference values in the *focal* species at equilibrium ($\bar{t}^*$ and $\bar{p}^*$). To identify the conditions enabling convergence toward the *local mimicry ring*, we estimate $|\bar{t}^* - \bar{t}'|$, *i.e.* the phenotypic distance between the mean trait within the *focal* species and the mean trait displayed in the *local mimicry ring*.

To estimate how much does sexual selection impair the phenotypic convergence toward the *local mimicry ring*, we then compute the discrepancy between trait and preference values in the focal species $|\bar{t}^* - \bar{p}^*|$.

Model exploration assuming strong females choosiness ($a = O(1)$) and predators discrimination ($b = O(1)$). Because the stringency of females choice, as well as the discrimination behavior of predators, are key drivers of the convergence toward the trait displayed in the *local mimicry ring*, we then explore large values of females choosiness and predators discrimination, *i.e.* $a = O(1)$ and $b = O(1)$. Under such assumptions, the effects of predation and RI on the changes in the mean trait and preference are of order ε. We also assume that s is of order ε, because if s is of order ε^2 the strength of developmental and selective constraints would be negligible compared with selection due to predation and RI. Assuming strong females choosiness violates the weak selection hypothesis, because strong females choosiness generates strong sexual selection for males and opportunity cost for females. However, because strong females choosiness leads to higher sexual selection, the discrepancy between preference and trait values ($|\bar{t}^* - \bar{p}^*|$) becomes limited. Therefore sexual selection and opportunity cost are actually weak and we can still use Equation (1) to study the evolution of the mean trait and preference values, assuming stronger females choosiness. However, under strong females choosiness ($a = O(1)$) and predators discrimination ($b = O(1)$), we cannot analytically determine the mean trait and preference values at equilibrium ($\bar{t}^*$ and $\bar{p}^*$). We then use numerical simulations to explore cases where a and b are of order 1 and s is of order ε (hypothesis H_s)

By default, we report results assuming (H_w). We also verified that the results stand assuming (H_s) (see Appendix 1.6). When the parameters have the same effect assuming (H_w) or (H_s), we do not provide the effect of parameters assuming (H_s).

Results

We study the evolution of the mean values of trait and preference in the *focal* species, described by Equation (2). When selection due to RI is stronger than the stabilizing constraints on the trait in the *focal* species (*i.e.* when inequality (11) is not satisfied), the distances between mean trait and preference values in the *focal* and $\bar{t}'$ become very large (not of order 1) (see Appendix 1.4). Otherwise, mean trait and preference values converge to the equilibrium values $\bar{t}^*$ and $\bar{p}^*$ with

$$\bar{t}^* = \frac{\frac{2bd\bar{\lambda}'N'}{(1+\lambda N+\bar{\lambda}'N')^2}\bar{t}' + 2st_o - ac_{RI}\frac{N'}{N}\bar{t}'}{\frac{2bd\bar{\lambda}'N'}{(1+\lambda N+\bar{\lambda}'N')^2} + 2s - ac_{RI}\frac{N'}{N}}, \quad (9)$$

and

$$\bar{p}^* = \bar{t}^* + c_{RI}\frac{N'}{N}(\bar{t}^* - \bar{t}'). \quad (10)$$

These analytical expressions allow us to study the impact of the different evolutionary forces on the level of convergence of the trait displayed in the *focal* species toward the signal exhibited in the *local mimicry ring* ($|\bar{t}^* - \bar{t}'|$).

Reproductive interference limits the convergence toward the local mimicry ring. Selection exerted by predators favors the convergence of trait in the *focal* species toward the signal exhibited by the *local mimicry ring*. When

$$c_{RI} < \frac{N}{aN'}\left(\frac{4bd\bar{\lambda}'N'}{(1+\lambda N+\bar{\lambda}'N')^2} + 4s\right), \quad (11)$$

the level of convergence toward the local mimicry ring is given by:

$$|\bar{t}^* - \bar{t}'| = \frac{2s}{\frac{2bd\bar{\lambda}'N'}{(1+\lambda N+\bar{\lambda}'N')^2} + 2s - ac_{RI}\frac{N'}{N}}|t_o - \bar{t}'|. \quad (12)$$

Hence, if we assume no RI ($c_{RI} = 0$), we have $|\bar{t}^* - \bar{t}'| < |t_o - \bar{t}'|$, implying that the trait displayed in the *focal* species gets closer to the trait displayed in the *local mimicry ring* (outcome (a) *mimicry* or (b) *imperfect mimicry*, see Figures 1(a) and 1(b)). The mimicry in the *focal* species becomes nearly perfect ($|\bar{t}^* - \bar{t}'|$ close to zero) when the strength of selection due to predation ($2bd\bar{\lambda}'N'/(1+\lambda N+\bar{\lambda}'N')^2$) is large enough, as compared to the intra-specific constraints (s) (outcome (a) *mimicry* see Figure 1(a)).

However assuming reproductive interference (RI) between females from the *focal* species and males from the *local mimicry ring* impairs this convergence. When RI is non null but has a limited strength, satisfying (11), Equation (12) implies that $\partial|\bar{t}^* - \bar{t}'|/\partial c_{RI} > 0$ (see Appendix 1.6.1). RI thus increases the gap between the traits displayed in the *focal* species and in *the local mimicry ring*, leading to imperfect mimicry in the *focal* species. Because females preferences generate sexual selection on males traits, RI promotes the phenotypic divergence between the *focal* species and the *local mimicry ring* (see Equation (10)). Thus RI limits the convergence generated by to Müllerian mimicry.

Abbreviation	Description
$\bar{t}/\bar{p}$	Mean trait/preference value in the *focal* species
G_t/G_p	Genetic variance of trait t/preference p
C_{tp}	Genetic covariance between trait t and preference p
β_t/β_p	Selection coefficient on trait t/preference p
$\bar{t}'$	Mean trait value in the mimetic community
a	Females choosiness
s	Strength of selective and developmental constraints promoting the optimal trait value t_o
t_o	Optimal trait value in the focal species without ecological interaction with other species
d	Basal predation rate
b	Predators discrimination coefficient
λ	Defence level carried out by individuals of the focal species
$\bar{\lambda}'$	Mean defence level of an individual of the local mimicry ring
N/N'	Density of the focal species/local mimicry ring
c_{RI}	Strength of reproductive interference

Table 1 Description of variables and parameters used in the model assuming a fixed *local mimicry ring*

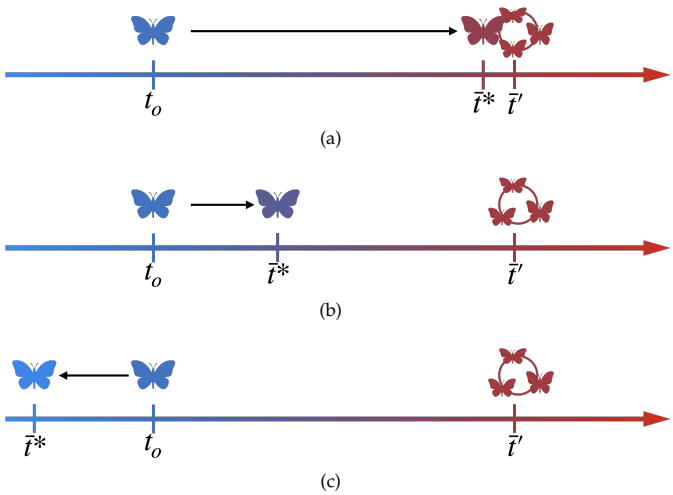


Figure 1 Illustration of the three main outcomes: (a) *mimicry*: the value of the trait in the *focal* species $\bar{t}$ becomes very close to the mean value displayed in the *local mimicry ring* $\bar{t}'$, (b) *imperfect mimicry*: the value of the trait in the *focal* species $\bar{t}$ gets closer but stay distant to the mean value displayed in the *local mimicry ring* $\bar{t}'$, (c) *no mimicry*: the value of the trait in the *focal* species $\bar{t}$ gets closer but stay distant to the mean value displayed in the *local mimicry ring* $\bar{t}'$.

However, when the cost associated with RI crosses a threshold, *i.e.* when

$$c_{RI} \geq \frac{N}{aN'} \frac{4bd\bar{\lambda}' N'}{(1+\lambda N+\bar{\lambda}' N')^2}, \tag{13}$$

then

$$|\bar{t}^* - \bar{t}'| > |t_o - \bar{t}'|. \tag{14}$$

When assuming such elevated cost of RI, imperfect mimicry is thus no longer observed, and RI rather promotes a strong divergence of the trait in the *focal* species (see Figure 1(c) for an illustration).

When inequality (11) is not satisfied, the phenotypic distance between the *focal* species and the *local mimicry ring* becomes very large. This very large divergence is biologically unrealistic but suggests that RI can promote phenotypic divergence between defended species living in sympatry. This unrealistic divergence stems from the weak female choosiness ($a = O(\varepsilon)$) assumed: indeed only a large distance between females preference and the trait displayed in the *local mimicry ring* allows female choice to reduce the fitness cost due to RI. Elevated cost of RI thus promotes large inter-specific phenotypic distances by sexual selection. Relaxing the weak preference and predator discrimination hypothesis, *i.e.* assuming that $a = O(1)$ and $b = O(1)$, confirms that RI does limit the mimicry in the *focal* species. However, in this case, when a strong divergence is favored, this divergence becomes high but stays at order $O(1)$ (see Appendix 1.5). Indeed, as females choosiness is strong, this divergence strongly reduces fitness cost due to RI. Therefore stabilizing constraints on the trait in the *focal* species become more important than RI, avoiding very large divergence.

Strict females choosiness and moderate predators discrimination favor convergence of mimetic signals despite reproductive interference. The evolution of the trait in the *focal* species is shaped by two antagonistic evolutionary forces, generated by RI and Müllerian mimicry, respectively. RI indirectly limits the convergence toward the *local mimicry ring* by impacting females preference. Therefore, females choosiness a may be a key feature to understand the evolution of the trait in the *focal* species. The

selection exerted by predation also depends on animal behavior through predators learning. We thus investigate the impact of the strength of females choosiness a and of the predators discrimination coefficient b on the evolution of the mimetic trait, allowing strong values of a and b (*i.e.* assuming (H_s)).

When the female choosiness is low, the selection promoting divergence generated by RI is limited. Indeed, females are not choosy and tend to accept all encountered males, including heterospecific males, whatever the direction of their preference. The difference in fitness cost due to RI between females with different preferences is then low, leading to poor selection due to RI. With higher level of females choosiness, fitness cost due to RI depends more on the direction of preference, leading to higher selection caused by RI. Therefore, higher level of females choosiness generates stronger limit to the convergence toward the *local mimicry ring* (see Figure 2).

By contrast, above a threshold, higher level of females choosiness allows females to accurately distinguish males from their own species from individuals from the local mimicry ring. Figure 2 highlights that increased choosiness limits divergence of the trait in the focal species. Accurate females choice allows a quasi similarity of the traits of the focal species and of the local mimicry ring, without implying heterospecific mating, relaxing the selection promoting divergence generated by RI. The evolutionary outcome is also impacted by selection exerted by predation. The predation rate depends on predators discrimination coefficient b, modulating their capacity to differentiate traits displayed in different mimetic species. The higher discrimination of predators reduces the distances between the average traits displayed in the *focal* species and the *local mimicry ring* (Figure 2), leading to more accurate mimicry (see Appendix 1.6.2 for more detail). Surprisingly, when predators discrimination increases above a certain threshold, increased discrimination no longer promotes accurate mimicry (Figure 2). When b is high, individuals from the *focal* species benefit from protection gained by mimicry only when they look very similar to the individuals of the local mimicry ring. RI makes such high similarity too costly, and therefore divergence of the trait in the *focal* species becomes promoted when predators discrimination is too high (Figure 2).

Assuming reproductive interference, weakly defended with low density species do not mimic the local mimicry ring. The defence level of the *focal* species, *i.e.* λN, is likely to impact the level of trait divergence. Without RI, $\partial|\bar{t}^* - \bar{t}'|/\partial\lambda N > 0$ indicates that the defence level in the *focal* species increases divergence from the signal displayed in the *local mimicry ring*. As expected, the convergence toward the mimetic signal is thus more likely to happen when the *focal* species is weakly defended (λN small). Indeed, individuals from weakly defended species gain a limited protection from conspecific individuals and have a great advantage from mimicry to the local community of well-defended species. By contrast, individuals of a well-defended species get protection from their conspecifics, therefore the selective pressure favoring similarity to the *local mimicry ring* is weaker. In well-defended species, developmental and selective constraints exceed selection favoring similarity and promote the persistence of traits in the *focal* species distinct from the *local mimicry ring*. To understand the interaction between the two components of the defence level, we then compared the effect of the density N and of the individual defence level λ of the focal species, for different costs of choosiness (Figure 3). Without RI ($c_{RI} = 0$), Figure 3(a) confirms that low individual defence level

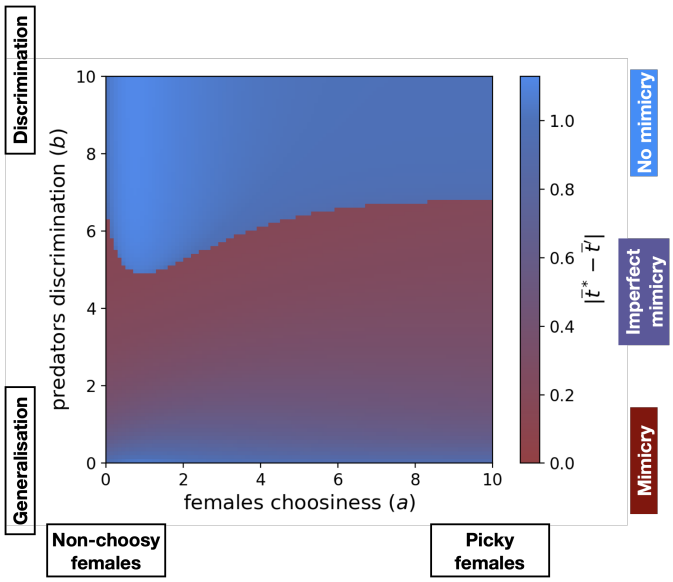


Figure 2 Influence of predators discrimination b and of the strength of female choosiness a on the phenotypic distance between the *focal* species and the *local mimicry ring*. The plot reports values generated by numerical simulations performed using parameters values of the (H_s) assumption. The parameter range generating *mimicry*, *imperfect mimicry* and *no mimicry* are shown in red, purple and blue respectively. We assume: $G_t = 0.01$, $G_p = 0.01$, $c_{RI} = 0.002$, $d = 0.02$, $\lambda = 0.1$, $N = 20$, $\bar{\lambda}' = 0.1$, $N' = 20$, $s = 0.005$, $t_o = 0$, $\bar{t}' = 1$.

promotes mimicry in the *focal* species.

By contrast, assuming RI ($c_{RI} > 0$), Figures 3(b)-(c) show that the evolutionary outcome depends more on the individual defence level, than on the density of the *focal* species. The convergence toward the signal displayed in the *local mimicry ring* is observed when the individual defence level λ is low and when the density of the *focal* species is sufficiently large. When the defence level of the *focal* species (λN) is low, the protection gained by positive frequency-dependent selection within species is low, and the convergence toward the *local mimicry ring* is thus strongly promoted. Surprisingly, this convergence is impaired for similar values of defence level, when the density of the focal species is low. When the density of the *focal* species us low, females pay higher fitness costs due to RI, because they encounter more often heterospecific than conspecific males. Altogether, our results suggest that convergence of the mimetic traits is likely to happen for low level of individual defence in the *focal* species (*i.e.* Batesian ($\lambda = 0$) or quasi-Batesian ($\lambda > 0$) mimicry) and when the density of the *focal* species is high.

Evolution of mimicry between two interacting species

Method

Because RI may impact the direction of convergence, and because ancestral traits that might differ among species may influence the evolution of traits, we build an extension of the previous model by assuming two sympatric species, whose traits can evolve. This two-species model allows investigating the eco-evolutionary dynamics of trait convergence. We thus follow the evolution of traits (t_1 and t_2 in species 1 and 2 respectively) and females choosiness (p_1 and p_2 in species 1 and 2 respectively).

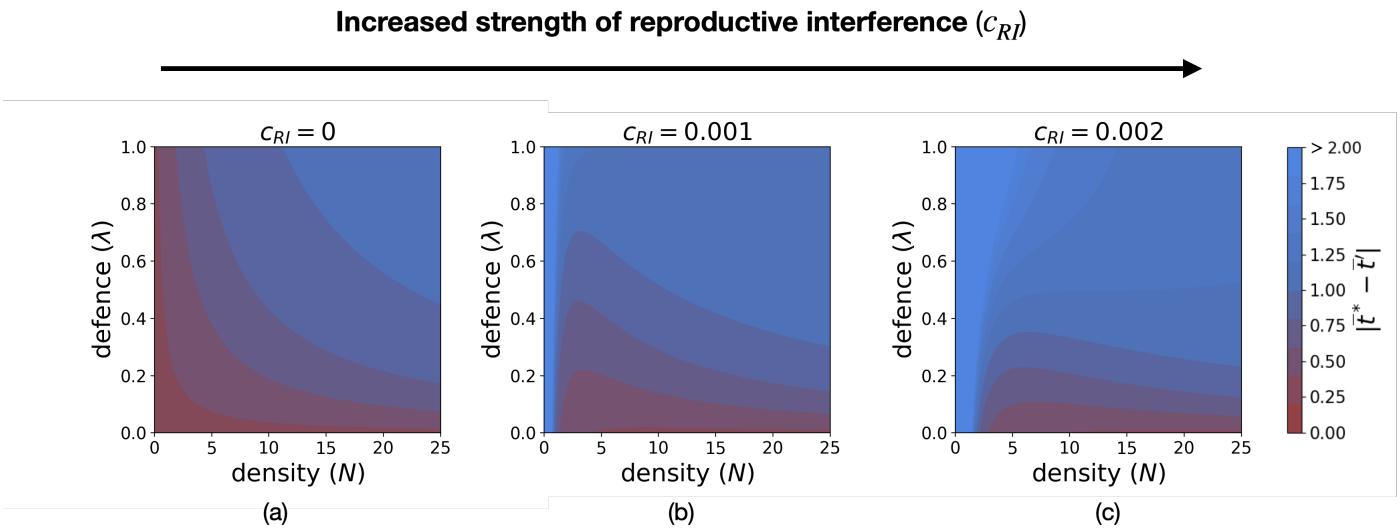


Figure 3 Influence of the density of the *focal* species N and of individual defence level in the *focal* species λ on the phenotypic distance between the *focal* species and the *local mimicry ring*, for different strengths of reproductive interference c_{RI}. The plot reports values generated by the analytic approximation assuming (H_w). The parameter range generating *mimicry*, *imperfect mimicry* and *no mimicry* are shown in red, purple and blue respectively. Phenotypic distances greater than 2 are shown are blue. Different values of strengths of reproductive interference are assumed: (a) $c_{RI} = 0$, (b) $c_{RI} = 0.001$, (c) $c_{RI} = 0.002$. We assume: $a = 0.01$, $d = 0.05$, $b = 0.01$, $\bar{\lambda}' = 0.1$, $N' = 20$, $s = 0.00005$, $t_o = 0$, $\bar{t}' = 1$.

Because the two species are assumed to be fully sympatric, individuals from both species face the same community of predators, and can encounter each others. The evolution of trait and preference in both species are thus shaped by both (1) Müllerian mimicry and (2) RI caused by heterospecific mating.

Assuming that traits and preferences have an autosomal polygenic basis with additive effects in both species, the change in the mean values of traits $\bar{t}_1$, $\bar{t}_2$ and preferences $\bar{p}_1$ and $\bar{p}_2$ can be approximated under a weak selection hypothesis, using:

$$\begin{pmatrix} \Delta\bar{t}_i \\ \Delta\bar{p}_i \end{pmatrix} = \frac{1}{2}\begin{pmatrix} G_{t_i} & C_{t_ip_i} \\ C_{t_ip_i} & G_{p_i} \end{pmatrix} \approx \frac{1}{2}\begin{pmatrix} G_{t_i}\beta_{t_i} \\ G_{p_i}\beta_{p_i} \end{pmatrix}, \qquad (15)$$

where for $i \in \{1,2\}$ the variables β_{t_i} and β_{p_i} describe the selective forces acting on the traits and preferences t_i and p_i (defined in Appendix 2.2). G_{t_i} and G_{p_i} are the additive genetic variances of t_i and p_i and $C_{t_ip_i}$ is the additive genetic covariance between t_i and p_i. As in the previously described model, following (Iwasa et al. 1991), we assume that G_{t_i} and G_{p_i} are positive constants maintained by an equilibrium between selection and recurrent mutations. As we already showed (see Appendix 1.2), the additive genetic covariance $C_{t_ip_i}$ is small in comparison with G_{t_i} and G_{p_i}, allowing us to approximate the change in the mean values of trait and preference as done in (15).

Females choosiness (a_1 and a_2), strength of selective and developmental constraints (s_1 and s_2), optimal trait without ecological interaction with the other species (t_{o1} and t_{o2}), defence level (λ_1 and λ_2) carried out by individuals, as well as density (N_1 and N_2) can differ between species 1 and 2. All variables and parameters used in the two-population model are summed up in the supplementary Table A1.

Model exploration. To specifically test the effect of species interactions on trait convergence, we assume the two species to be initially isolated and explore trait evolution when they become sympatric. We thus assume the initial mean trait values to be equal to the optimal traits t_{o1} and t_{o2} respectively, favored by developmental constraints, that may differ between species. Because of opportunity costs, we also assume the initial mean females preference value to equal mean initial trait values within each species. The dynamics of the mean traits and preferences values when both species become in contact are then given by Equation (15).

Exploration assuming weak females choosiness ($a_1 = O(\varepsilon)$ and $a_2 = O(\varepsilon)$) and predators discrimination ($b = O(\varepsilon)$) (Hypothesis (H_w)). When we assume that a_1, a_2 and b are of order ε and s is of order ε^2, we can analytically determine the mean traits and preferences values at equilibrium in most cases. In the few remaining cases, we use numerical analyses to estimate these values (see Appendix 2.4). We note $\bar{t}_1^*$, $\bar{t}_2^*$, $\bar{p}_1^*$ and $\bar{p}_2^*$ the mean traits and preferences at equilibrium. We follow the quantity $|\bar{t}_1^* - \bar{t}_2^*|$, *i.e.* the mean phenotypic distance between both species at equilibrium. To determine the direction of convergence, we check whether the trait values observed at equilibrium (*i.e.* $\bar{t}_1^*$ and $\bar{t}_2^*$) are closer to the ancestral trait value of species 1 or of species 2 (t_{o1} or t_{o2} respectively).

Exploration assuming strong females choosiness ($a_1 = O(1)$ and $a_2 = O(\varepsilon)$) and predators discrimination ($b = O(1)$) (Hypothesis $((H_s))$). Similarly to the previous model exploration, we use numerical simulations to study cases where females choosiness and predators discrimination are strong ($a_1 = O(1)$, $a_2 = O(1)$, $b = O(1)$ and $s = O(\varepsilon)$).

As in the exploration of the previous model, by default we report results assuming (H_w) and check the validity of these results when assuming (H_s) (see Appendix 2.5).

Results

Reproductive interference may reverse the direction of convergence. As previously unraveled, RI may limit the convergence of aposematic traits, depending on the level of defence and on

the density of the two mimetic species (Figure 3(c)). By studying the two-species model, we can also investigate how RI may influence the direction of convergence, depending on the relative levels of defence of the two species, generated by their densities and individual strength of defence ($\lambda_1 N_1$ and $\lambda_2 N_2$, respectively). The fitness costs caused by RI then also differ in the two species, depending on their relative abundance, leading to unbalanced selective pressures on trait evolution in the two species: females from the rarest species suffer more from RI, because they encounter heterospecific more often than conspecific males.

To investigate how the relative defence levels of the two species affect the direction of convergence, we study the phenotype at equilibrium in both species, for different values of the two components of the defence level of species 1: the individual defence level (λ_1) and the density (N_1). We assumed than the species 2 is already well protected.

When assuming no reproductive interference ($c_{RI} = 0$), species traits converge toward the optimal trait t_o initially displayed in the best defended species. In Figures 4(a) and 4(b), individuals from the poorly defended species 1 get weak protection from conspecific individuals and thus have a greater advantage to look similar to individuals of species 2. The more species 1 is defended, the more its mean trait value is close to the optimal trait value t_{o1} (see Figure 4(a)). Such increase in the defence level of species 1 also impacts the evolution of trait in the sympatric species 2 (see Figure 4(b)): when the individual defence level in species 1 (λ_1) is below a threshold, the more individuals from species 1 are protected, the more the mean trait value in species 2 moves away from its ancestral trait (t_{o2}). Surprisingly, above this threshold, the more species 1 is protected, the more the mean trait value in species 2 gets closer to its ancestral trait value (t_{o2}). The weak discrimination capacities of predators assumed favors such evolution of divergent traits. Individuals from species 2 may thus gain some protection from species 1 even with a poor resemblance. When species 1 is hugely protected (high $\lambda_1 N_1$), this protection is still high, even when individuals of species 2 display trait close to the optimal trait value t_{o2}.

When assuming positive strength of reproductive interference ($c_{RI} > 0$), the trait value of species 1 almost never converges toward the trait value initially displayed in species 2 (t_{o2}). On the contrary, individuals of species 2 can mimic individuals of species 1 (see Figure 4(d)), when the defence level of individual of species 1 is high and when species 1 is rare. Because individuals from both species are well defended (high λ_1 and λ_2), individuals of both species have an advantage to look similar. However, when species 1 is rarer, this leads to an increased cost of RI, inhibiting convergence towards the optimal trait value displayed in the alternative species (t_{o2}) (see Figure 4(c)). Since predation pressure promotes convergence of traits of both species, the mean trait in species 2 becomes closer to the species 1 ancestral trait t_{o1}. Surprisingly, this advergence also happens when individuals of species 2 contributes more to predator learning those of species 1 (see Figure 4(c) and (d)). By contrast, when the density in species 1 exceeds the density in species 2, individuals from both species exhibit traits close to their ancestral traits. As previously explained, both species are well-protected and then gain little from mimicry. However, the phenotypic distance between the two species is higher than in the previous case, because RI further promotes trait dissimilarity (see Figure 4).

Our results highlight that RI impacts the evolution of mimetic traits, and may even reverse the direction of convergence, depending on the relative abundances and defence levels of sympatric species.

Optimal traits in interacting species strongly modulate the phenotypic divergence driven by reproductive interference. Our previous results highlight that RI limits the convergence of mimetic signals. This effect is also observed in the model investigating the evolution of traits in two interacting species. However, the effect of RI on trait divergence strongly depends on the optimal traits initially displayed in the two species in absence of ecological interactions (t_{o1} and t_{o2}). When these optimal traits differ, an increase in strength of RI leads to a progressive increase in the phenotypic distance between both species (see Figure 5).

Surprisingly, when the optimal traits without ecological interaction are the same in both species ($t_{o1} = t_{o2}$), an increase in the strength of RI (c_{RI}) has no effect on the phenotypic distance between both species, as long as this cost of RI remains below a threshold (see Figures 6(a) and 6(b)). This effect is observed assuming weak (H_w) or strong (H_s) females choosiness and predators discrimination.

However, when the strength of RI (c_{RI}) is greater than this threshold, assuming weak females choosiness and predators discrimination (H_w), the phenotypic distance between both species becomes very large, while their initially displayed optimal trait values t_{o1} and t_{o2} are equal. A similar trend is observed when females choosiness and predators discrimination are strong (H_s): the phenotypic distance is null when the strength of RI remains below a threshold, while it quickly increases to a great value, when the strength of RI is above the threshold (see Figure 6(b)).

Our results highlight that the optimal traits without ecological interaction strongly modulate the effect of RI on the convergence of mimetic signals. Surprisingly, drastic divergence might be promoted by strong cost of RI, even when the optimal phenotypes are the same in the two interacting species.

Discussion

Reproductive interference leads to unexpected evolution of warning trait

Our model shows that RI limits the convergence of aposematic traits in defended species living in sympatry. The effect of RI strongly depends on the relative abundances of interacting species, leading to surprising evolutionary outcomes. For example, in rare defended species, selection favoring mimicry towards the local community of defended species is expected to be strong. Nevertheless, our model shows that the elevated costs of RI prevents the evolution of mimicry in the rarest species, because females then encounter way more heterospecific than conspecific males.

Our results thus shed some light on the persistence of distinct warning signals within local communities of defended species in the wild: RI is indeed likely to promote the persistence of a distinct warning pattern in low-density populations of aposematic species becoming in contact with a local mimicry ring exhibiting a different warning trait. Nevertheless, our model does not take into account the dynamics of population density, and therefore ignores the extinction risk of low-density populations. Such non-mimetic populations with low density can nevertheless persist in the wild, when the level of individual defence is sufficiently high.

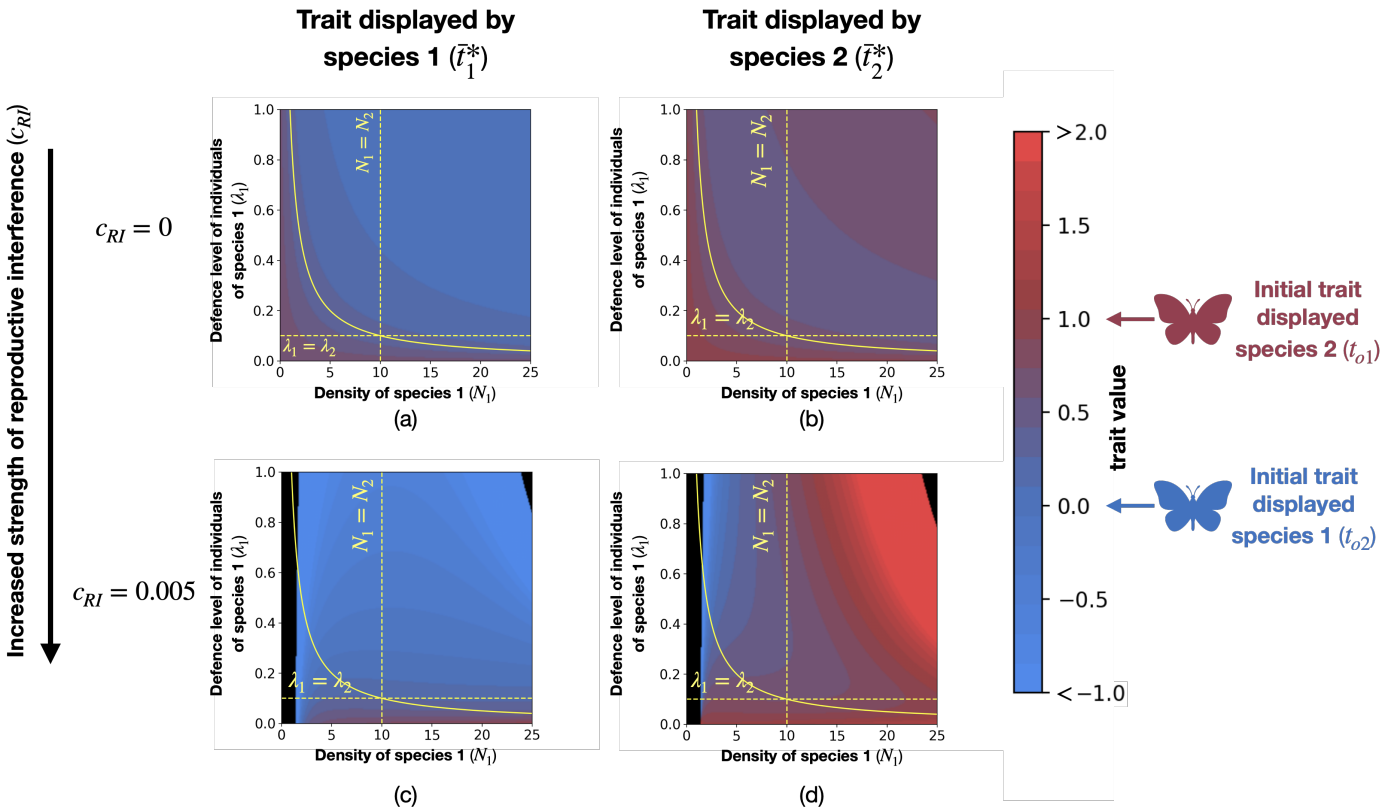


Figure 4 Influence of the density and of the individual defence level in species 1 (N_1 and λ_1) on the traits displayed in both species, for different strengths of RI (c_{RI}). Plots report values obtained with the analytic approximation, assuming (H_w). Mean traits values can become very large (black zone). Trait values greater than 2 (resp. lower than -1) are shown in blue (resp. red). The yellow solid line shows the case where both species have the same level of defence ($\lambda_1 N_1 = \lambda_2 N_2$). Different values of strengths of reproductive interference are assumed: (a) and (b) $c_{RI} = 0$, (c) and (d) $c_{RI} = 0.005$. We assume: $a_1 = a_2 = 0.01$, $d = 0.05$, $b = 0.01$, $\lambda_2 = 0.1$, $N_2 = 10$, $s_1 = s_2 = 0.000005$, $t_{o1} = 0$, $t_{o2} = 1$.

Our model also brings new insights on the ecological processes driving the direction of convergence in warning traits. In absence of RI, warning traits are expected to evolve in the less defended species (e.g. low density populations and/or low level of individual defence), becoming mimetic toward better-defended species living in sympatry (Balogh and Leimar 2005; Franks and Sherratt 2007). When considering the costs of RI however, this general trend no longer holds, because the evolution of warning trait can be observed in the most abundant species toward the signal of the less abundant one.

Our model therefore highlights the interplay between mutualism and RI in sympatric species, driving the strength and the direction of convergence of traits involved in these ecological interactions.

The evolution of imperfect mimicry

Our results suggest that RI alone is unlikely to promote imperfect mimicry, in contradiction with previous predictions (Pfennig and Kikuchi 2012). When sympatric species share the same ancestral trait, we indeed observe either perfect mimicry or strong trait divergence, depending on the strength of RI. In our model, imperfect mimicry is observed only when interacting species have different ancestral traits. The contrasted developmental and selective constraints underwent by the different species may thus play an important role in imperfect mimicry. These different constraints are likely strongly correlated with the phylogenetic distances between species: closely-related species are likely to share same genetic basis and developmental pathway leading to the warning trait and to also share similar environments, due to niche conservatism (Chazot et al. 2014; Joshi et al. 2017). Our results suggest that imperfect mimicry could not be promoted among closely-related species experiencing high levels of RI but limited differences in ancestral traits. Imperfect mimicry may rather be observed between phylogenetically distant species, subject to more strikingly different developmental and selective constraints, where RI might be more limited. Distantly related species indeed might have diverged in other traits, facilitating mate recognition through different cues.

How important are developmental and selective constraints in the evolution of warning trait?

Estimating the level of developmental and selective constraints potentially shaping the evolution of warning traits is not straightforward. Genetic studies, reviewed in (Joron et al. 2006), show that *Heliconius* species share the same 'toolkit' of genes, repeatedly recruited during both convergent and divergent evolution of warning patterns within and between species. This important lability in color patterns observed in this genus suggests a limited level of developmental constraints, facilitating the evolution of mimicry, even between species from different sub-clades within the genus (Hines et al. 2011). By contrast, in butterflies from the tribe *ithomiines*, a strong phylogenetic signal on the

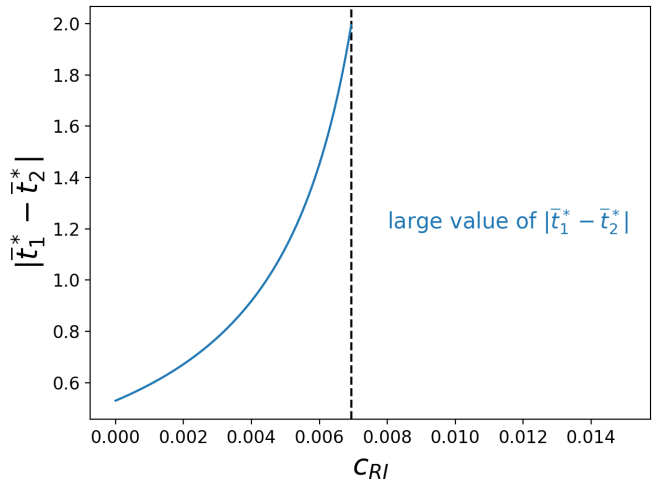


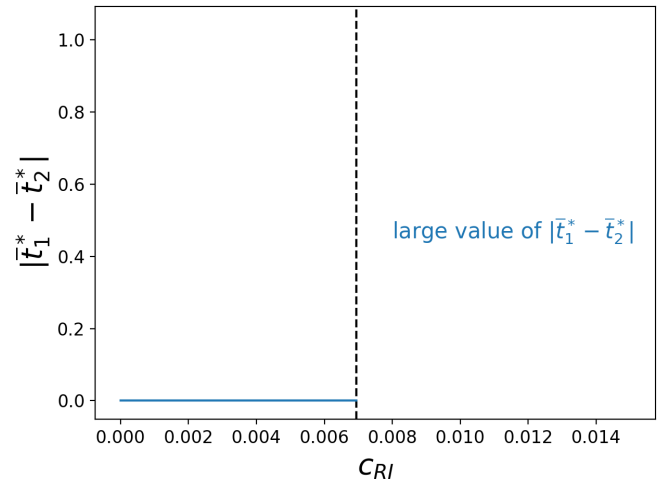
Figure 5 Influence of the strength of reproductive interference c_{RI} on the phenotypic distances between the two species, when they have different optimal traits without ecological interaction ($t_{o1} \neq t_{o2}$). The plot reports values obtained by the analytic approximation assuming (H_w). We assume: $a_1 = a_2 = 0.01$, $d = 0.02$, $b = 0.01$, $\lambda_1 = \lambda_2 = 0.1$, $N_1 = N_2 = 10$, $s_1 = s_2 = 0.000005$, $t_{o1} = 0$, $t_{o2} = 1$.

warning trait is observed (Elias et al. 2008; Chazot et al. 2014), suggesting that developmental and selective constraints may limit the evolution of mimicry among distantly related species.

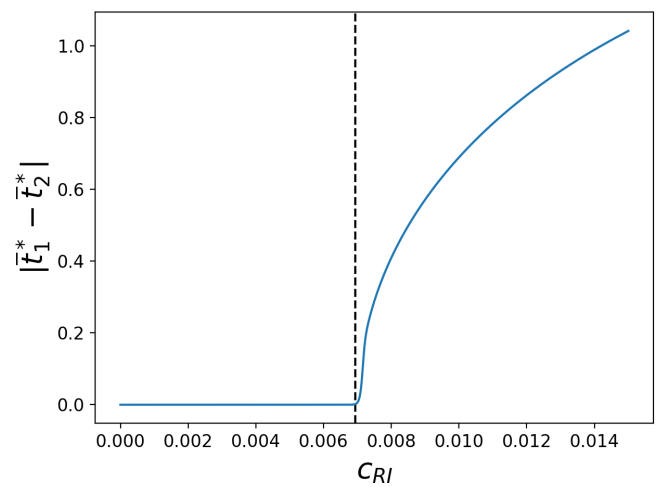
The level of developmental constrains and RI that are both linked to the phylogenetic distances among species are thus interacting factors shaping the mutualistic interactions between defended species living in sympatry. The direction and strength of convergent evolution is thus not entirely defined by relative densities and levels of defence, but also by the evolutionary history of interacting species.

Evolution of females preference limiting the RI costs generated by mimicry

Our results show the evolution of close similarity between species, despite the costs of RI. When females choosiness is high, successful species recognition might occur without decreasing the protection brought by mimicry. Such situation arises when predators largely generalize, and therefore do not discriminate imperfect mimics. Such differences in the discrimination of color patterns between prey and predators are frequently observed in the wild. For instance, Llaurens et al. (2014) showed that the variations in color pattern between co-mimetic species from the distantly related genus *Heliconius* and *Melinaeae* might be better perceived by *Heliconius* butterflies than by avian predators. The evolution of visual perception in females could also enhance species discrimination without impairing mimicry. The evolution of vision in females from the butterfly species *Heliconius erato* indeed coincides with the evolution of the yellow pigments 3-OH-kinurenin displayed on the wings of individuals from their own species (Bybee et al. 2012). The evolution of high discrimination capacities in mimetic prey, as well as the evolution of mating cue undetected by predators could thus limit the cost of RI in mimetic prey. In butterflies, mate choice indeed often relies on pheromones that may strongly differ among closely-related mimetic species (Darragh et al. 2017; González-Rojas et al. 2020). Micro-habitat differences among mimetic species



(a)



(b)

Figure 6 Influence of the strength of reproductive interference c_{RI} on the phenotypic distances between the two species when they share the same optimal traits without ecological interaction ($t_{o1} = t_{o2}$), assuming (a) (H_w) weak or (b) (H_s) strong females choosiness and predators discrimination. Plots reports values obtained by (a) the analytic approximation assuming (H_w) and (b) numerical simulations to use parameters values corresponding to (H_s) assumption. Different values of females choosiness, predators discrimination coefficient and strength of selective and developmental constraints are assumed : (a) $a = 0.01, b = 0.01$ and $s = 0.00005$, (b) $a = 1, b = 1$ and $s = 0.005$. We assume: $G_t = 0.01$, $G_p = 0.01$, $d = 0.02$, $\lambda_1 = \lambda_2 = 0.1$, $N_1 = N_2 = 10$, $t_{o1} = t_{o2} = 1$.

may also allow reducing heterospecific encounters while still enjoying the benefit of mimicry by sharing the same predator community (Estrada and Jiggins 2002). In *Morpho* butterflies displaying local convergences in wing patterns, temporal segregation in patrolling activity has been observed between species sharing similar color patterns (Le Roy et al. 2020), that may strongly limit heterospecific rivalry .

The levels of RI among mimetic species might thus be modulated by the evolution of the converging traits themselves, as

well as the evolution of other traits involved in species interactions.

Conclusion

Our analytical and numerical results highlight the role of RI in the evolution of warning trait. They confirm that RI limits the convergence of warning trait promoted by mutualistic interactions. Surprisingly, RI can change the direction of the convergence of warning trait when the species densities are unbalanced. Our results also suggest that RI alone cannot explain imperfect mimicry, highlighting the role of ancestral traits as well as developmental and selective constraints in the evolution of imperfect mimicry.

Data Availability

Codes are available online: github.com/Ludovic-Maisonneuve/limits-of-ev-conv

Acknowledgments

LM would like to thank Dorian Ni for feedbacks on the mathematical part of the study. LM and VL would like to thank Marianne Elias, Charline Pinna and the whole 'Evolution and Development of Phenotypic Variations' team for stimulating discussions on the evolution of the warning traits. The authors would like to thank the ANR SUPERGENE (ANR-18-CE02-0019) for funding the PhD of LM, and the Emergence program from Paris city council for supporting the team of VL. This work was also supported by the Chair "Modélisation Mathématique et Biodiversité" of VEOLIA- Ecole Polytechnique-MNHN-F.X.

Conflict of interest disclosure

The authors of this preprint declare that they have no conflict of interest with the content of this article.

Literature Cited

Arias, M., Y. le Poul, M. Chouteau, R. Boisseau, N. Rosser, et al., 2016 Crossing fitness valleys: empirical estimation of a fitness landscape associated with polymorphic mimicry. Proceedings of the Royal Society B: Biological Sciences **283**: 20160391.

Balogh, A. C. and O. Leimar, 2005 Müllerian mimicry: an examination of fisher's theory of gradual evolutionary change. Proceedings of the Royal Society B: Biological Sciences **272**: 2269–2275.

Barton, N. H. and M. Turelli, 1991 Natural and sexual selection on many loci. Genetics **127**: 229–255.

Beccaloni, G., 1997 Ecology, natural history and behaviour of ithomiine butterflies and their mimics in ecuador (lepidoptera: Nymphalidae: Ithomiinae). Tropical Lepidoptera **8**: 103–124.

Benitez-Vieyra, S., N. H. de Ibarra, A. M. Wertlen, and A. A. Cocucci, 2007 How to look like a mallow: evidence of floral mimicry between turneraceae and malvaceae. Proceedings of the Royal Society B: Biological Sciences **274**: 2239–2248.

Benson, W. W., 1972 Natural selection for mullerian mimicry in heliconius erato in costa rica. Science **176**: 936–939.

Brown, W. L. and E. O. Wilson, 1956 Character displacement. Systematic Zoology **5**: 49–64.

Bybee, S. M., F. Yuan, M. D. Ramstetter, J. Llorente-Bousquets, R. D. Reed, et al., 2012 Uv photoreceptors and uv-yellow wing pigments in heliconius butterflies allow a color signal to serve both mimicry and intraspecific communication. The American Naturalist **179**: 38–51, PMID: 22173459.

Chazot, N., K. R. Willmott, P. G. Santacruz Endara, A. Toporov, R. I. Hill, et al., 2014 Mutualistic mimicry and filtering by altitude shape the structure of andean butterfly communities. The American Naturalist **183**: 26–39, PMID: 24334733.

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

Cortesi, F. and K. L. Cheney, 2010 Conspicuousness is correlated with toxicity in marine opisthobranchs. Journal of Evolutionary Biology **23**: 1509–1518.

Dalziell, A. H. and J. A. Welbergen, 2016 Mimicry for all modalities. Ecology Letters **19**: 609–619.

Darragh, K., S. Vanjari, F. Mann, M. F. Gonzalez-Rojas, C. R. Morrison, et al., 2017 Male sex pheromone components in *Heliconius* butterflies released by the androconia affect female choice. PeerJ **5**: e3953.

Elias, M., Z. Gompert, C. Jiggins, and K. Willmott, 2008 Mutualistic interactions drive ecological niche convergence in a diverse butterfly community. PLOS Biology **6**: 1–1.

Estrada, C. and C. Jiggins, 2002 Patterns of pollen feeding and habitat preference among heliconius species. Ecological Entomology **27**: 448 – 456.

Estrada, C. and C. D. Jiggins, 2008 Interspecific sexual attraction because of convergence in warning colouration: is there a conflict between natural and sexual selection in mimetic species? Journal of Evolutionary Biology **21**: 749–760.

Flanagan, N. S., A. Tobler, A. Davison, O. G. Pybus, D. D. Kapan, et al., 2004 Historical demography of müllerian mimicry in the neotropical heliconius butterflies. Proceedings of the National Academy of Sciences **101**: 9704–9709.

Franks, D. W. and J. Noble, 2004 Batesian mimics influence mimicry ring evolution. Proceedings of the Royal Society of London. Series B: Biological Sciences **271**: 191–196.

Franks, D. W. and T. N. Sherratt, 2007 The evolution of multicomponent mimicry. Journal of Theoretical Biology **244**: 631 – 639.

González-Rojas, M. F., K. Darragh, J. Robles, M. Linares, S. Schulz, et al., 2020 Chemical signals act as the main reproductive barrier between sister and mimetic heliconius butterflies. Proceedings of the Royal Society B: Biological Sciences **287**: 20200587.

Goodale, E. and S. W. Kotagama, 2006 Vocal mimicry by a passerine bird attracts other species involved in mixed-species flocks. Animal Behaviour **72**: 471 – 477.

Grill, C. P. and A. Moore, 1998 Effects of a larval antipredator response and larval diet on adult phenotype in an aposematic ladybird beetle. Oecologia **114**: 274–282.

Gröning, J. and A. Hochkirch, 2008 Reproductive interference between animal species. The Quarterly Review of Biology **83**: 257–282, PMID: 18792662.

Hines, H. M., B. A. Counterman, R. Papa, P. Albuquerque de Moura, M. Z. Cardoso, et al., 2011 Wing patterning gene redefines the mimetic history of heliconius butterflies. Proceedings of the National Academy of Sciences **108**: 19666–19671.

Iwasa, Y., A. Pomiankowski, and S. Nee, 1991 The evolution of costly mate preferences ii. the "handicap" principle. Evolution **45**: 1431–1442.

Jiggins, C. D., R. E. Naisbit, R. L. Coe, and J. Mallet, 2001 Reproductive isolation caused by colour pattern mimicry. Nature

411: 302–305.

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

Joron, M., C. D. Jiggins, A. Papanicolaou, and W. O. McMillan, 2006 Heliconius wing patterns: an evo-devo model for understanding phenotypic diversity. Heredity **97**: 157–167.

Joshi, J., A. Prakash, and K. Kunte, 2017 Evolutionary assembly of communities in butterfly mimicry rings. The American Naturalist **189**: E58–E76.

Kapan, D. D., 2001 Three-butterfly system provides a field test of müllerian mimicry. Nature **409**: 338–340.

Kirkpatrick, M., T. Johnson, and N. Barton, 2002 General models of multilocus evolution. Genetics **161**: 1727–1750.

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

Kyogoku, D., 2015 Reproductive interference: ecological and evolutionary consequences of interspecific promiscuity. Population Ecology **57**: 253–260.

Le Roy, C., C. Roux, E. Authier, H. Bastide, V. Debat, et al., 2020 Convergent morphology and divergent phenology: unravelling the coexistence of mimetic morpho butterfly species. bioRxiv .

Lev-Yadun, S., 2009 Müllerian mimicry in aposematic spiny plants. Plant Signaling & Behavior **4**: 482–483, PMID: 19816137.

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

Llaurens, V., M. Joron, and M. Théry, 2014 Cryptic differences in colour among müllerian mimics: how can the visual capacities of predators and prey shape the evolution of wing colours? Journal of Evolutionary Biology **27**: 531–540.

Mallet, J. and N. H. Barton, 1989 Strong natural selection in a warning-color hybrid zone. Evolution **43**: 421–431.

Mallet, J. and L. E. Gilbert Jr., 1995 Why are there so many mimicry rings? correlations between habitat, behaviour and mimicry in heliconius butterflies. Biological Journal of the Linnean Society **55**: 159–180.

McPeek, M. A. and S. Gavrilets, 2006 The evolution of female mating preferences: Differentiation from species with promiscious males can promotes speciation. Evolution **60**: 1967 – 1980.

Merrill, R. M., A. Chia, and N. J. Nadeau, 2014 Divergent warning patterns contribute to assortative mating between incipient heliconius species. Ecology and Evolution **4**: 911–917.

Naisbit, R. E., C. D. Jiggins, and J. Mallet, 2001 Disruptive sexual selection against hybrids contributes to speciation between *Heliconius cydno* and *Heliconius melpomene*. Proceedings of the Royal Society of London. Series B: Biological Sciences **268**: 1849–1854.

Ojala, K., L. Lindsröm, and J. Mappes, 2007 Life-history constraints and warning signal expression in an arctiid moth. Functional Ecology **21**: 1162–1167.

Pfennig, D. W. and D. W. Kikuchi, 2012 Competition and the evolution of imperfect mimicry. Current Zoology **58**: 608–619.

Ruxton, G. D., D. W. Franks, A. C. V. Balogh, and O. Leimar, 2008 Evolutionary implications of the form of predator generalization for aposematic signals and mimicry in prey. Evolution **62**: 2913–2921.

Sanders, K., A. Malhotra, and R. Thorpe, 2006 Evidence for a müllerian mimetic radiation in asian pitvipers. Proceedings of the Royal Society B: Biological Sciences **273**: 1135–1141.

Savage, J. M. and J. B. Slowinski, 1992 The colouration of the venomous coral snakes (family elapidae) and their mimics (families aniliidae and colubridae). Biological Journal of the Linnean Society **45**: 235–254.

Schemske, D. W., 1981 Floral convergence and pollinator sharing in two bee-pollinated tropical herbs. Ecology **62**: 946–954.

Sherratt, T. N., 2008 The evolution of müllerian mimicry. Die Naturwissenschaften **95**: 681–695.

Springer, V. and W. Smith-Vaniz, 1972 Mimetic relationships involving fishes of the family blenniidae. Smithsonian Contributions Zoology **112**.

Symula, R., R. Schulte, and K. Summers, 2001 Molecular phylogenetic evidence for a mimetic radiation in peruvian poison frogs supports a müllerian mimicry hypothesis. Proceedings of the Royal Society of London. Series B: Biological Sciences **268**: 2415–2421.

Vasconcellos-Neto, J. and K. S. Brown, 1982 Interspecific hybridization in mechanitis butterflies (ithomiinae): a novel pathway for the breakdown of isolating mechanisms. Biotropica **14**: 288–294.

Yamaguchi, R. and Y. Iwasa, 2013 Reproductive character displacement by the evolution of female mate choice. Evolutionary Ecology Research **15**: 25–41.

Appendix: The limits of evolutionary convergence in sympatry: reproductive interference and developmental constraints leading to local diversity in aposematic signals

1 Model 1: Evolution of mimicry toward a fixed community of defended species

1.1 Selection coefficients

1.1.1 Computation of β_t

For the sake of readability, we first recall the definitions of the quantities under study.

The selective forces acting on the trait value write:

$$\begin{aligned}
\beta_t &= \frac{\partial}{\partial t} \ln \left(W_m(t|\bar{t}, \bar{p}, \bar{t}') \right) + \frac{\partial}{\partial t} \ln \left(W_f(t, p|\bar{t}, \bar{t}') \right) \Big|_{(t,p)=(\bar{t},\bar{p})} \\
&= \frac{\partial}{\partial t} \left\{ -a(t-\bar{p})^2 + \ln W_{pred}(t|\bar{t}, \bar{t}') - s(t-t_o)^2 \right\} \Big|_{(t,p)=(\bar{t},\bar{p})} \\
&\quad + \frac{\partial}{\partial t} \left\{ -a(p-\bar{t})^2 + \ln W_{pred}(t|\bar{t}, \bar{t}') + \ln W_{RI}(p|\bar{t}, \bar{t}') - s(t-t_o)^2 \right\} \Big|_{(t,p)=(\bar{t},\bar{p})} \\
&= -2a(\bar{t}-\bar{p}) + 2 \frac{\partial}{\partial t} \ln \left(W_{pred}(\bar{t}|\bar{t}, \bar{t}') \right) \Big|_{(t,p)=(\bar{t},\bar{p})} - 4s(\bar{t}-t_o).
\end{aligned}$$

Recall the definition of W_{pred} in (7) The second term of the right hand side equals

$$2\frac{\partial}{\partial t}\ln\left(W_{pred}(t|\bar{t},\bar{t}')\right) = -4b\left((t-\bar{t}')\overline{\lambda}'\exp\left[-b(t-\bar{t}')^2\right]N' + (t-\bar{t})\lambda\exp\left[-b(t-\bar{t})^2\right]N\right)$$
$$\left(\frac{1}{1+\lambda N\exp\left[-b(t-\bar{t})^2\right]+\overline{\lambda}'\exp\left[-b(t-\bar{t}')^2\right]N'-d}\right.$$
$$\left.-\frac{1}{1+\lambda N\exp\left[-b(t-\bar{t})^2\right]+\overline{\lambda}'\exp\left[-b(t-\bar{t}')^2\right]N'}\right).$$

When the parameters b and d are of order ε (under (H_w)), we obtain

$$\beta_t = -2a(\bar{t}-\overline{p}) - \frac{4bd(\bar{t}-\bar{t}')\overline{\lambda}'N'}{\left(1+\lambda N+\overline{\lambda}'N'\right)^2} - 4s(\bar{t}-t_o) + O(\varepsilon^3). \tag{A1}$$

1.1.2 Computation of β_p

Here again we recall the expression of selective forces acting on the preference value:

$$\begin{aligned}\beta_p &= \frac{\partial}{\partial p}\ln\left(W_f(t,p|\bar{t},\bar{t}')\right)\Big|_{(t,p)=(\bar{t},\overline{p})}\\ &= \frac{\partial}{\partial p}\left\{-a(p-\bar{t})^2+\ln W_{pred}(t|\bar{t},\bar{t}')+\ln W_{RI}(p|\bar{t},\bar{t}')-s(t-t_o)^2\right\}\Big|_{(t,p)=(\bar{t},\overline{p})}\\ &= \frac{\partial}{\partial p}\left\{-a(p-\bar{t})^2-\ln\left(1+c_{RI}\exp\left[a(\bar{t}-\bar{t}')(\bar{t}+\bar{t}'-2p)\right]\frac{N'}{N}\right)\right\}\Big|_{(t,p)=(\bar{t},\overline{p})}\end{aligned}$$

The last term of the right hand side equals

$$\begin{aligned}\frac{\partial}{\partial p}\ln\left(W_{RI}(p|\bar{t},\bar{t}')\right) &= -\frac{-2c_{RI}a(\bar{t}-\bar{t}')\exp\left[a(\bar{t}-\bar{t}')(\bar{t}+\bar{t}'-2p)\right]\frac{N'}{N}}{1+c_{RI}\exp\left[a(\bar{t}-\bar{t}')(\bar{t}+\bar{t}'-2p)\right]\frac{N'}{N}}\\ &= \frac{2c_{RI}a(\bar{t}-\bar{t}')}{c_{RI}+\exp\left[a(\bar{t}-\bar{t}')(2p-\bar{t}-\bar{t}')\right]\frac{N}{N'}}.\end{aligned}$$

When the parameters a and c_{RI} are of order ε (under (H_w)), we finally obtain

$$\beta_p = -2a(\overline{p}-\bar{t}) + 2c_{RI}a(\bar{t}-\bar{t}')\frac{N'}{N} + O(\varepsilon^3). \tag{A2}$$

1.2 Computation of the matrix of correlation

In this part we approximate the genetic covariance between trait and preference C_{tp}, using the results from [Kirkpatrick et al., 2002]. Trait and preference are controled by different sets of unlinked loci with additive effects, denoted T and P, respectively. For each i in T (resp. P), we note ξ_i^t (resp. ξ_i^p) the contribution of the locus i on trait (resp. preference) value. The trait and preference values of an individual are then given by

$$t = \sum_{i \in T} \xi_i^t \quad \text{and} \quad p = \sum_{i \in P} \xi_i^p. \tag{A3}$$

As in [Lande, 1981] we assume that the distributions of ξ_i^t and ξ_i^p are multivariate Gaussian. Let C_{ij} be the genetic covariance between loci i and j. Then the elements of the matrix of correlation are given by:

$$G_t = \sum_{i,j \in T} C_{ij}, \quad G_p = \sum_{i,j \in P} C_{ij} \quad \text{and} \quad C_{tp} = \sum_{i \in T, j \in P} C_{ij}. \tag{A4}$$

To compute the change on genetic correlation we need to identify various selection coefficients (see [Barton and Turelli, 1991, Kirkpatrick et al., 2002]). These coefficients are obtained using the contribution to the next generation of a mating between a male with trait t_m and a female with trait t_f and preference p_f due to natural selection and mating preference:

$$\begin{aligned} W(t_m, t_f, p_f) = &\exp\left[-a(p_f - t_m)^2\right] W_{pred}(t_m|\bar{t}, \bar{t}') \exp\left[-s(t_m - t_o)^2\right] \\ &\times W_{pred}(t_f|\bar{t}, \bar{t}') \exp\left[-s(t_f - t_o)^2\right] W_{RI}(p_f|\bar{t}, \bar{t}'). \end{aligned} \tag{A5}$$

For simplicity we consider only leading terms in the change in genetic correlation. For

$(i,j) \in T \times P$, combining Equations (9), (12), (15) from Kirkpatrick et al. [2002] gives the change in the genetic covariance between loci i and j:

$$\Delta C_{ij} = -\frac{C_{ij}}{2} + \frac{1}{4}\tilde{a}_{t_m t_m} \sum_{k,l \in T} (C_{ik}C_{jl} + C_{il}C_{jk}) + \frac{1}{4}\tilde{a}_{p_f p_f} \sum_{k,l \in P} (C_{ik}C_{jl} + C_{il}C_{jk}) + \frac{1}{4}\tilde{a}_{t_m p_f} \sum_{k \in T, l \in P} C_{ik}C_{jl} + O(\varepsilon^2) \tag{A6}$$

with $\tilde{a}_{\mu\rho}$ for $(\mu,\rho) \in \{t_m, t_f, p_f\}^2$ being the leading term of

$$a_{\mu\rho} := \frac{1}{2}\frac{\partial^2}{\partial\mu\partial\rho} \ln(W(t_m, t_f, p_f))\Big|_{(t_m,t_f,p_f)=(\bar{t},\bar{t},\bar{p})},$$

the selection coefficients calculated from the contribution to the next generation. The expressions of these coefficients are:

$$a_{p_f p_f} = -a - \frac{2c_{RI}a^2(\bar{t}-\bar{t}')^2 \frac{N}{N'} \exp\left[a(\bar{t}-\bar{t}')(2\bar{p}-\bar{t}-\bar{t}')\right]}{\left(c_{RI} + \frac{N}{N'}\exp\left[a(\bar{t}-\bar{t}')(2\bar{p}-\bar{t}-\bar{t}')\right]\right)^2},$$

$$\begin{aligned} a_{t_m t_m} = &-a - s + 2b^2\left(\overline{\lambda}' N'\left(1 + (\bar{t}-\bar{t}')^2\right)\exp\left[b(\bar{t}-\bar{t}')^2\right] + \lambda N\right) \\ &\times \left(\frac{1}{1+\lambda N + \overline{\lambda}' N' \exp\left[b(\bar{t}-\bar{t}')^2\right] - d} - \frac{1}{1+\lambda N + \overline{\lambda}' N' \exp\left[b(\bar{t}-\bar{t}')^2\right]}\right) \\ &+ \left(2b\overline{\lambda}' N'(\bar{t}-\bar{t}')\exp\left[b(\bar{t}-\bar{t}')^2\right]\right)^2 \\ &\times \left(\frac{1}{\left(1+\lambda N + \overline{\lambda}' N' \exp\left[b(\bar{t}-\bar{t}')^2\right] - d\right)^2} - \frac{1}{\left(1+\lambda N + \overline{\lambda}' N' \exp\left[b(\bar{t}-\bar{t}')^2\right]\right)^2}\right), \end{aligned}$$

and

$$a_{t_m p_f} = 2a,$$

therefore $\tilde{a}_{p_f p_f} = -a$, $\tilde{a}_{t_m t_m} = -a$ and $\tilde{a}_{t_m p_f} = 2a$.

By summing Equations (A6) over each i, j in T and P we obtain:

$$\Delta C_{tp} = -\frac{C_{tp}}{2} - \frac{1}{2} a G_t C_{tp} - \frac{1}{2} a G_p C_{tp} + \frac{1}{4} a G_t G_p + O(\varepsilon^2) \tag{A7}$$

For the sake of simplicity we assumed that the genetic correlations between traits and preferences are at equilibrium (as in [Barton and Turelli, 1991, Pomiankowski and Iwasa, 1993]). Neglecting the terms of order ε^2 we obtain from (A7) that the value at equilibrium is given by

$$C^*_{tp} = \frac{a G_t G_p}{1 + a(G_t + G_p)} = a G_t G_p + O(\varepsilon^2).$$

When we relax the hypothesis $a = O(\varepsilon)$ and $b = O(\varepsilon)$ the previous calculation still stand, except the last approximation.

1.3 Quasi equilibria

We search mean values of trait and preference at equilibrium $\overline{t}^*$ and $\overline{p}^*$ such as the leading terms of $\Delta\overline{t}^*$ and $\Delta\overline{p}^*$ are null. We remind that we assume assume weak females choosiness and predators discrimination (default hypothesis (H_w)). The leading terms of $\Delta\overline{t}^*$ and $\Delta\overline{p}^*$ are thus $\tilde{\beta}_t = \tilde{\beta}_p = 0$ with

$$\tilde{\beta}_t = -2a(\overline{t}^* - \overline{p}^*) - \frac{4bd(\overline{t}^* - \overline{t}')\overline{\lambda}' N'}{\left(1 + \lambda N + \overline{\lambda}' N'\right)^2} - 4s(\overline{t}^* - t_o), \tag{A8}$$

$$\tilde{\beta}_p = -2a(\overline{p}^* - \overline{t}^*) + 2c_{RI} a(\overline{t}^* - \overline{t}')\frac{N'}{N}. \tag{A9}$$

We call such points quasi equilibria because at these points the term of order ε^3 in the selection coefficient may be non-zero.

Lemma 1. *If*

$$\frac{2bd\overline{\lambda}'N'}{(\lambda N+\overline{\lambda}'N'+1)^2}+2s-\frac{ac_{RI}N'}{N}\neq 0, \tag{A10}$$

there is one quasi steady point $(\overline{t}^*,\overline{p}^*)$*, where*

$$\overline{t}^*=\frac{\frac{2bd\overline{\lambda}'N'}{(1+\lambda N+\overline{\lambda}'N')^2}\overline{t}'+2st_o-\frac{ac_{RI}N'}{N}\overline{t}'}{\frac{2bd\overline{\lambda}'N'}{(1+\lambda N+\overline{\lambda}'N')^2}+2s-\frac{ac_{RI}N'}{N}}\quad \textit{and}\quad \overline{p}^*=\overline{t}^*+c_{RI}(\overline{t}^*-\overline{t}')\frac{N'}{N}. \tag{A11}$$

Proof. Assume that $a\neq 0$. From the definitions of $\tilde{\beta}_t$ and $\tilde{\beta}_p$ in (A8) and (A9), we deduce that the values of trait and preference $(\overline{t}^*,\overline{p}^*)$ for which the selections on trait and preference are null satisfy:

$$\tilde{\beta}_t=0\iff a(\overline{p}^*-\overline{t}^*)=\frac{2bd(\overline{t}^*-\overline{t}')\overline{\lambda}'N'}{\left(1+\lambda N+\overline{\lambda}'N'\right)^2}+2s(\overline{t}^*-t_o). \tag{A12}$$

and

$$\tilde{\beta}_p=0\iff a(\overline{p}^*-\overline{t}^*)=ac_{RI}(\overline{t}^*-\overline{t}')\frac{N'}{N}. \tag{A13}$$

From (A12) and (A13) we derivate:

$$\frac{2bd(\overline{t}^*-\overline{t}')\overline{\lambda}'N'}{\left(1+\lambda N+\overline{\lambda}'N'\right)^2}+2s(\overline{t}^*-t_o)=ac_{RI}(\overline{t}^*-\overline{t}')\frac{N'}{N}.$$

By factorizing by $\overline{t}$ we get:

$$\overline{t}^*\left(\frac{2bd\overline{\lambda}'N'}{\left(1+\lambda N+\overline{\lambda}'N'\right)^2}+2s-ac_{RI}\frac{N'}{N}\right)=\overline{t}'\left(\frac{2bd\overline{\lambda}'N'}{\left(1+\lambda N+\overline{\lambda}'N'\right)^2}-ac_{RI}\frac{N'}{N}\right)+2st_o. \tag{A14}$$

Using (A10), we obtain the value of $\overline{t}^*$ given in (A11). From (A13) we deduce the associated mean preference and conclude the proof. □

1.4 Fast and slow dynamics.

We suppose that the parameters are such as $\frac{2bd\overline{\lambda}'N'}{(\lambda N+\overline{\lambda}'N'+1)^2}+2s-\frac{ac_{RI}N'}{N}$ is of order 1. We may have two long term behaviours. In case of convergence (case 1. of Lemma 2), the trajectories quickly approach the line $\overline{t}=\overline{p}$, before evolving according to a slower dynamic along this line (see the proof of Lemma 2 and Figures 1 and 2).

Lemma 2. *1. If*

$$\frac{2bd\overline{\lambda}'N'}{(\lambda N+\overline{\lambda}'N'+1)^2}+2s-\frac{ac_{RI}N'}{N}>0,$$

the quantities $(\overline{t}-\overline{t}^*)$ *and* $(\overline{p}-\overline{p}^*)$ *become of order* $\sqrt{\varepsilon}$ *when the number of generation goes to infinity, where the values of* $\overline{t}^*$ *and* $\overline{p}^*$ *have been given in* (A11).

2. If

$$\frac{2bd\overline{\lambda}'N'}{(\lambda N+\overline{\lambda}'N'+1)^2}+2s-\frac{ac_{RI}N'}{N}<0,$$

trait and preference becomes very large.

Proof. As we have said, we can then decompose the dynamics into two steps. In the first one the leading order terms of the selection coefficient are the terms describing sexual selection and cost of choosiness. In this case we can approximate $\Delta(\overline{t}-\overline{p})^2$.

$$\begin{aligned}
\Delta(\overline{t}-\overline{p})^2 &= (\overline{t}+\Delta\overline{t}-\overline{p}-\Delta\overline{p})^2-(\overline{t}-\overline{p})^2\\
&= 2(\overline{t}-\overline{p})(\Delta\overline{t}-\Delta\overline{p})+(\Delta\overline{t}-\Delta\overline{p})^2\\
&= 2(\overline{t}-\overline{p})(\frac{1}{2}G_t\beta_t-\frac{1}{2}G_p\beta_p)+(\frac{1}{2}G_t\beta_t-\frac{1}{2}G_p\beta_p)^2\\
&= -(\overline{t}-\overline{p})\Bigg(2a(G_t+G_p)(\overline{t}-\overline{p})-a^2(G_p+G_t)^2(\overline{t}-\overline{p})\\
&\quad -2\left(\frac{ac_{RI}G_pN'(\overline{t}'-\overline{t})}{N}+\frac{2bdG_t\overline{\lambda}'N'(\overline{t}'-\overline{t})}{(lN+\overline{\lambda}'N'+1)^2}+G_ts(t_o-\overline{t})\right)\Bigg)+O(\varepsilon^3),
\end{aligned}$$

where we used (2 for the third inequality and (A1) and (A2) for the last one. As by assumption, a, c_{RI}, b and d are of order ε, and s of order ε^2, we obtain

$$\Delta(\bar{t}-\overline{p})^2 = -(\bar{t}-\overline{p})\Big(2a(G_t+G_p)(\bar{t}-\overline{p}) + O(\varepsilon^2)\Big) + O(\varepsilon^3). \tag{A15}$$

Denote by r_n the value of $(\bar{t}-\overline{p})^2$ at generation n. We will prove the following statement:

$$\text{There exists } \alpha > 0 \text{ such that for } n \text{ large enough, } r_n \leq 2\alpha\varepsilon^2. \tag{A16}$$

If we introduce the parameter

$$\mathcal{A} := 1 - 2a(G_t + G_p),$$

Equation (A15) may be rewritten:

$$r_{n+1} = \mathcal{A} r_n + \sqrt{r_n} f_n(1,\varepsilon) + f_n(2,\varepsilon),$$

for every $n \in \mathbb{N}$ $|f_n(1,\varepsilon)| \leq c_1\varepsilon^2$, $|f_n(2,\varepsilon)| \leq c_2\varepsilon^3$ where c_1 and c_2 are finite constants. Now introduce the variable

$$\rho_n := r_n - r_0\mathcal{A}^n.$$

As a is of order ε, and G_t and G_p of order one, there exists a positive and finite constant $C_{\mathcal{A}}$ such that for ε small enough,

$$0 < \mathcal{A} < 1 - C_{\mathcal{A}}\varepsilon.$$

We will choose in the sequel a positive real number α satisfying

$$\sqrt{\frac{2}{\alpha}}c_1 + \frac{c_2}{\alpha} < C_{\mathcal{A}} \tag{A17}$$

and

$$\frac{c_1}{\sqrt{\alpha}} + \frac{\sqrt{\alpha}c_1 + c_2}{\alpha} < C_{\mathcal{A}}. \tag{A18}$$

As $0 < \mathcal{A} < 1$, we see that for n large enough, $r_0\mathcal{A}^n \leq \alpha\varepsilon^2$. Thus to prove (A16), it is enough to prove that for n large enough, $|\rho_n| \leq \alpha\varepsilon^2$. We obtain for ρ the recurrence equation

$$\rho_{n+1} = \mathcal{A}\rho_n + \sqrt{\rho_n + r_0\mathcal{A}^n} f_n(1,\varepsilon) + f_n(2,\varepsilon). \tag{A19}$$

Assume first that $|\rho_n| \leq \alpha\varepsilon^2$. In this case,

$$\begin{aligned} |\rho_{n+1}| &\leq \mathcal{A}\alpha\varepsilon^2 + \sqrt{2\alpha\varepsilon^2}c_1\varepsilon^2 + c_2\varepsilon^3 \\ &= \alpha\varepsilon^2\left(\mathcal{A} + \sqrt{\frac{2}{\alpha}}c_1\varepsilon + \frac{c_2}{\alpha}\varepsilon\right) \leq \alpha\varepsilon^2 \end{aligned}$$

thanks to (A17) provided that ε is small enough. Hence it is enough to prove that there exists one n such that $\rho_n \leq \alpha\varepsilon^2$, and the inequality will hold for later generations.

Now assume that $\rho_n > \alpha\varepsilon^2$. Using that

$$\sqrt{\rho_n + r_0\mathcal{A}^n} \leq \sqrt{|\rho_n|} + \sqrt{r_0\mathcal{A}^n} \leq \sqrt{|\rho_n|} + \sqrt{\alpha}\varepsilon$$

for n large enough, we obtain from (A19)

$$\begin{aligned} |\rho_{n+1}| &\leq \mathcal{A}|\rho_n| + \sqrt{|\rho_n|}c_1\varepsilon^2 + (\sqrt{\alpha}c_1 + c_2)\varepsilon^3 \\ &\leq \mathcal{A}|\rho_n| + \sqrt{|\rho_n|}c_1\sqrt{\frac{|\rho_n|}{\alpha}}\varepsilon + (\sqrt{\alpha}c_1 + c_2)\frac{|\rho_n|}{\alpha}\varepsilon \\ &= |\rho_n|\left(\mathcal{A} + \frac{c_1}{\sqrt{\alpha}}\varepsilon + \frac{\sqrt{\alpha}c_1 + c_2}{\alpha}\varepsilon\right). \end{aligned}$$

Due to (A18), the term in bracket belongs to $(0,1)$ if ε is small enough. We conclude that there exists n_0 such that $|\rho_n| \leq \alpha\varepsilon^2$ for any $n \geq n_0$. This concludes the proof of (A16) and of the fast convergent phase.

We now want to prove that $\bar{t}$ becomes close to $\bar{t}^*$, and $\bar{p}$ to $\bar{p}^*$. To this aim, we look at the variation of

$$\frac{(\bar{t} - \bar{t}^*)^2}{G_t} + \frac{(\bar{p} - \bar{p}^*)^2}{G_p},$$

where the values of $\bar{t}^*$ and $\bar{p}^*$ are the ones defined in (A11). By definition,

$$\Delta\left(\frac{(\bar{t}-\bar{t}^*)^2}{G_t}+\frac{(\bar{p}-\bar{p}^*)^2}{G_p}\right)=\frac{2\Delta\bar{t}(\bar{t}-\bar{t}^*)+\left(\Delta\bar{t}\right)^2}{G_t}+\frac{2\Delta\bar{p}(\bar{p}-\bar{p}^*)+(\Delta\bar{p})^2}{G_p}.$$

When $\bar{t}-\bar{p}$ is of order ε, $\Delta\bar{t}$ and $\Delta\bar{p}$ are of order ε^2. We then neglect the terms proportional to $\left(\Delta\bar{t}\right)^2$ and $(\Delta\bar{p})^2$. This yields

$$\begin{aligned}\Delta\left(\frac{(\bar{t}-\bar{t}^*)^2}{G_t}+\frac{(\bar{p}-\bar{p}^*)^2}{G_p}\right)=&\frac{2\Delta\bar{t}(\bar{t}-\bar{t}^*)}{G_t}+\frac{2\Delta\bar{p}(\bar{p}-\bar{p}^*)}{G_p}+O(\varepsilon^4)\\ =&\beta_t(\bar{t}-\bar{t}^*)+\beta_p(\bar{p}-\bar{p}^*)+O(\varepsilon^4),\end{aligned}$$

where we used (2). Now notice that

$$\bar{t}-\bar{t}^*=\bar{t}-\bar{p}+\bar{p}-\bar{p}^*+\bar{p}^*-\bar{t}^*=\bar{p}-\bar{p}^*+O(\varepsilon)=\frac{\bar{t}-\bar{t}^*}{2}+\frac{\bar{p}-\bar{p}^*}{2}+O(\varepsilon), \tag{A20}$$

as $\bar{p}-\bar{t}$ is of order ε and $\bar{p}^*-\bar{t}^*$ of order ε. We can make the same computation for $\bar{p}-\bar{p}^*$. Adding that β_t and β_p are of order ε^2, we obtain:

$$\Delta\left(\frac{(\bar{t}-\bar{t}^*)^2}{G_t}+\frac{(\bar{p}-\bar{p}^*)^2}{G_p}\right)=(\beta_t+\beta_p)\frac{\bar{t}-\bar{t}^*}{2}+(\beta_t+\beta_p)\frac{\bar{p}-\bar{p}^*}{2}+O(\varepsilon^3). \tag{A21}$$

Using (A1) and (A2) and simplifying, we get:

$$\beta_t+\beta_p=-\frac{4bd(\bar{t}-\bar{t}')\bar{\lambda}'N'}{\left(1+\lambda N+\bar{\lambda}'N'\right)^2}-4s(\bar{t}-t_o)+2c_{RI}a(\bar{t}-\bar{t}')\frac{N'}{N}+O(\varepsilon^3).$$

By substracting $-\frac{4bd(\bar{t}^*-\bar{t}')\bar{\lambda}'N'}{\left(1+\lambda N+\bar{\lambda}'N'\right)^2}-2s(\bar{t}^*-t_o)+2c_{RI}a(\bar{t}^*-\bar{t}')\frac{N'}{N}$ which is equal to zero (because $\beta_t|_{(t,p)=(\bar{t}^*,\bar{p}^*)}=0$ and $\beta_p|_{(t,p)=(\bar{t}^*,\bar{p}^*)}=0$, we get:

$$\beta_t+\beta_p=-\left(\frac{2bd\bar{\lambda}'N'}{\left(1+\lambda N+\bar{\lambda}'N'\right)^2}+2s-ac_{RI}\frac{N'}{N}\right)2(\bar{t}-\bar{t}^*)+O(\varepsilon^3) \tag{A22}$$

$$=-\left(\frac{2bd\bar{\lambda}'N'}{\left(1+\lambda N+\bar{\lambda}'N'\right)^2}+2s-ac_{RI}\frac{N'}{N}\right)2(\bar{p}-\bar{p}^*)+O(\varepsilon^3), \tag{A23}$$

where we used (A20). Replacing $\beta_t + \beta_p$ using (A22) and (A23) into (A21) yield:

$$\Delta \left(\frac{(\bar{t} - \bar{t}^*)^2}{G_t} + \frac{(\bar{p} - \bar{p}^*)^2}{G_p} \right) = -A \left((\bar{t} - \bar{t}^*)^2 + (\bar{p} - \bar{p}^*)^2 \right) + O(\varepsilon^3), \tag{A24}$$

with

$$A := \left(\frac{2bd\overline{\lambda}' N'}{\left(1 + \lambda N + \overline{\lambda}' N'\right)^2} + 2s - ac_{RI} \frac{N'}{N} \right). \tag{A25}$$

First case: $A > 0$

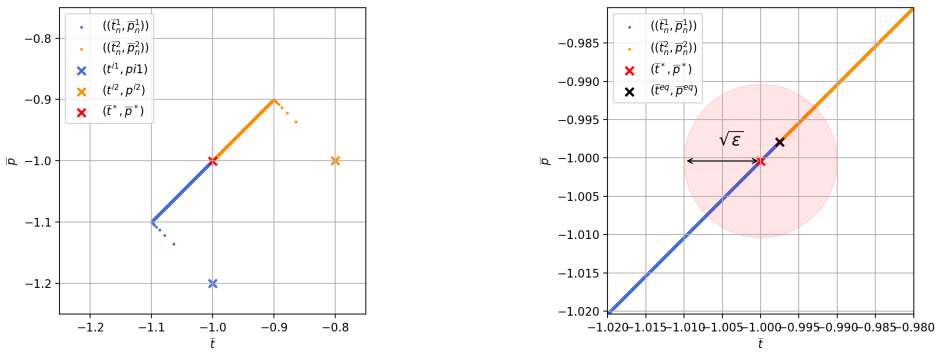
We suppose that the parameters are such that A is not of an order inferior to ε^2. Using computations similar to the ones derived to prove (A16), we may prove that after a number of generations large enough, $(\bar{t} - \bar{t}^*)^2$ and $(\bar{p} - \bar{p}^*)^2$ become of order ε, and thus $\bar{t} - \bar{t}^*$ and $\bar{p} - \bar{p}^*$ become of order $\varepsilon^{1/2}$.

To illustrate this result we simulated two trajectories of the system (2) $((\bar{t}_n^1, \bar{p}_n^1))_{n \in \mathbb{N}}$ and $((\bar{t}_n^2, \bar{p}_n^2))_{n \in \mathbb{N}}$ for two different initial conditions: (t^{i1}, p^{i1}) and (t^{i2}, p^{i2}). The trajectories are shown in Figure 1 and the parameter values are given in the caption. The parameter values correspond to the weak selection hypothesis and are such that $A > 0$.

The fast and slow dynamics are visible in Figure 1(a) where the trajectories quickly become close to the line defined by $\bar{t} = \bar{p}$ and then slowly converge toward the quasi equilibrium points. We noted $(\bar{t}^{eq}, \bar{p}^{eq})$ the equilibrium point reached by the two simulated trajectories.

Second case: $A < 0$**.**

Recall (A24) and that we assumed that A is not of order inferior to ε^2. This entails that $(\bar{t} - \bar{t}^*)^2/G_t + (\bar{p} - \bar{p}^*)^2/G_p$ goes to infinity. However when $t - \bar{t}'$ or $p - \bar{p}'$ becomes high



(a) (b)

Figure 1: (a) Evolution of mean trait value $\bar{t}$ and mean female preference value $\bar{p}$ for two different initial conditions. The blue (resp. orange) points represent the numerical trajectories for the initial condition (t^{i1}, p^{i1}) (resp. (t^{i2}, p^{i2})) (b) Zoom on the previous figure. The red circle has for center the quasi equilibrium point $(\bar{t}^*, \bar{p}^*)$ and for radius $\sqrt{\varepsilon}$. This illustrates that the numerical steady point $(\bar{t}^{eq}, \bar{p}^{eq})$ is at a distance of order $\sqrt{\varepsilon}$. The trajectories were obtained with the parameter values: $(t^{i1}, p^{i1}) = (-1, -1.2)$, $(t^{i2}, p^{i2}) = (-0.8, -1)$, a $= 10^{-4}$, b $= 2 \times 10^{-4}$, $d = 2 \times 10^{-4}$, $\lambda = 0.1$, $\overline{\lambda}' = 0.1$, $N = 10$, $N_p = 20$, $c_{RI} = 10^{-4}$, $s = 10^{-8}$, $\bar{t}' = 1$ and $t_o = 0$. The trajectories are computed over 1,000,000 generations.

(for example of order $1/\varepsilon$) the approximation

$$\exp\left[a(t - \bar{t}')(2p - t - \bar{t}')\right] = a(t - \bar{t}')(2p - t - \bar{t}') + O(\varepsilon^2)$$

does not stand. Therefore from a certain number of generations the approximation does not correctly describe the evolution of trait and preference. We can only conclude that trait and preference become very large. This is illustrated by Figure 2 which shows a simulation of two trajectories when $A < 0$. □

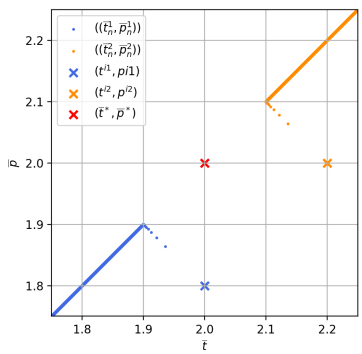


Figure 2: Evolution of mean trait value $\overline{t}$ and mean female preference value $\overline{p}$ for two different initial conditions. The blue (resp. orange) points represent the numerical trajectories for the initial condition (t^{i1}, p^{i1}) (resp. (t^{i2}, p^{i2})). The trajectories were obtained with the parameter values: $(t^{i1}, p^{i1}) = (-1, -1.2)$, $(t^{i2}, p^{i2}) = (-0.8, -1)$, a $= 10^{-4}$, b $= 2 \times 10^{-4}$, $d = 2 \times 10^{-4}$, $\lambda = 0.1$, $\overline{\lambda}' = 0.1$, $N = 10$, $N_p = 20$, $c_{RI} = 10^{-4}$, $s = 0.25 \times 10^{-8}$, $\overline{t}' = 1$ and $t_o = 0$. The trajectories are computed over 1,000,000 generations.

1.5 The phenotypic distances between the focal species and the local mimicry ring remain of order 1 assuming strong females choosiness

We obtain an analytic approximation of the phenotypic distance between the focal species and the local mimicry ring when the strengths of females choosiness (a) and predators discrimination (b) are weak (*i.e.* $a = O(\varepsilon)$ and $b = O(\varepsilon)$ (H_w)). To obtain this approximation we also assume that the strength of developmental and selective constraints is of the same order of magnitude as selection coefficients due to predation and reproductive interference (*i.e.* $s = O(\varepsilon^2)$). Under this assumption, the phenotypic distance between the focal species

and the local mimicry ring becomes very large (see Figure 3) when

$$c_{RI} \geq \frac{2N}{aN'} \left(\frac{bd\overline{\lambda}' N'}{(1 + \lambda N + \overline{\lambda}' N')^2} + s \right). \tag{A26}$$

As explained in the main manuscript, this unrealistic divergence (under the weak choosiness hypothesis ($a = O(\varepsilon)$)) suggests that RI promotes strong phenotype divergence between the focal species and the local mimicry ring. Under weak choosiness, only a large distance between females preference and the trait displayed by the local mimicry ring leads to a reduction of fitness cost generated by RI. Because females preference generates sexual selection on males trait, the phenotype divergence between the focal species and the local mimicry ring becomes high. When we relax the weak choosiness hypothesis (H_s) and use numerical simulations, then a great cost of RI provokes large but still of order 1 phenotypic distance between the focal species and the local mimicry ring (see Figure 3). However, we observe that the lower female choosiness, the higher is phenotypic distance when the cost of RI increases. As explained above, when females are not very choosy, RI favors the evolution of larger phenotypic distances limiting heterospecific mating.

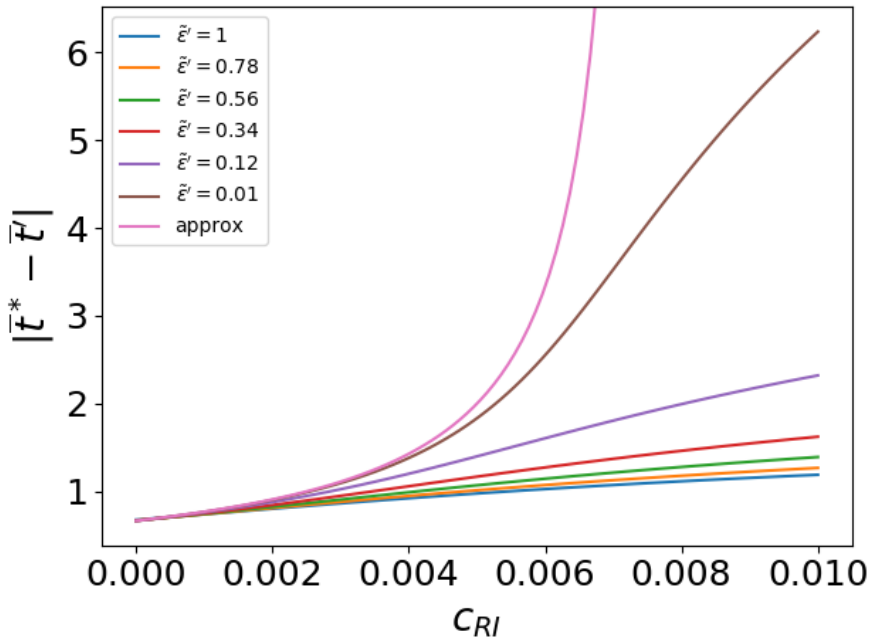


Figure 3: **Influence of the cost generated by reproductive interference c_{RI} on the phenotypic distances between the focal species and the local mimicry rings, using the analytical approximation ('approx' curve) or numerical simulations.** Different values of females choosiness (a), predators discrimination coefficient (b) and strength of selective and developmental constraints (s) are assumed for numerical simulations. This illustrate that curves obtained by numerical simulations tend to look similar to those using the analytical approximation where parameters values tends to satisfies (H_w). We assume: $a = \tilde{\varepsilon}'$, $b = \tilde{\varepsilon}'$, $s = 0.5\tilde{\varepsilon}\tilde{\varepsilon}'$,$d = 2\tilde{\varepsilon}$, $\lambda = 0.1$, $\overline{\lambda}' = 0.1$, $N = 10$, $N' = 20$, $t_o = 0$, $\bar{t}' = 1$.

1.6 Effect of different parameters on the level of trait divergence $|\bar{t}^* - \bar{t}'|$

We then aim at disentangling the effects of the different parameters of the model on trait divergence. We thus focus on the case where $A > 0$ (recall definition (A25)). In this case,

assuming weak females choosiness and predators discrimination (H_w), we know that the phenotypic distance between the focal species and the local mimicry ring, $|\bar{t}^* - \bar{t}'|$, equals

$$|\bar{t}^* - \bar{t}'| = \frac{2s|t_o - \bar{t}'|}{\frac{2bd\bar{\lambda}'N'}{(1+\lambda N+\bar{\lambda}'N')^2} + 2s - ac_{RI}\frac{N'}{N}},$$

allowing to investigate the relative impact of the different parameters on trait divergence. We also run numerical simulations to investigate the effect of the different parameters assuming strong females preference and predators discrimination (*i.e.* $a = O(1)$ and $b = O(1)$ (H_s)).

1.6.1 Effect of reproductive interference

Assuming weak females choosiness and predators discrimination (H_w), the impact of the strength of RI on the phenotypic distance between the focal species and the local mimicry ring is given by the sign of $\partial|\bar{t}^* - \bar{t}'|/\partial c_{RI}$. In our case,

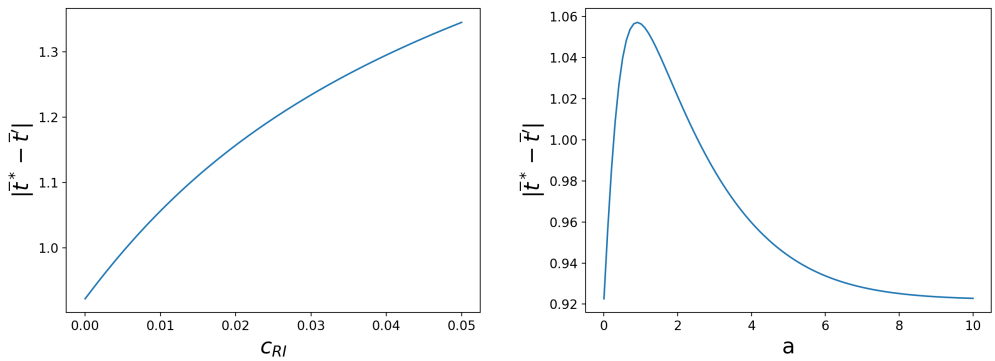
$$\frac{\partial|\bar{t}^* - \bar{t}'|}{\partial c_{RI}} = \frac{2sa\frac{N'}{N}|t_o - \bar{t}'|}{\left(\frac{2bd\bar{\lambda}'N'}{(1+\lambda N+\bar{\lambda}'N')^2} + 2s - ac_{RI}\frac{N'}{N}\right)^2} > 0.$$

Hence the level of divergence increases with the strength of RI. A similar effect is observed assuming strong females choosiness and predators discrimination (H_s) (see Figure 4(a)). Similarly, assuming weak females choosiness and predators discrimination (H_w)

$$\frac{\partial|\bar{t}^* - \bar{t}'|}{\partial a} = \frac{2sc_{RI}\frac{N'}{N}|t_o - \bar{t}'|}{\left(\frac{2bd\bar{\lambda}'N'}{(1+\lambda N+\bar{\lambda}'N')^2} + 2s - ac_{RI}\frac{N'}{N}\right)^2} > 0,$$

thus the phenotypic distance (along with the strength of selection due to RI) increases with the strength of females choosiness. However, when females choosiness is strong and exceeds a threshold, females are able to distinguish males from their own species from individuals from the local mimicry rings, which limits divergence of the trait in the focal species (see

Figure 4(b)). Accurate female choice allows a quasi-similarity of the traits of the focal species and of the local mimicry ring, without implying heterospecific mating, relaxing the divergent selection exerted by RI.



(a) (b)

Figure 4: **Influence of the strength of (a) reproductive interference c_{RI} and of (b) females choosiness a on the phenotypic distances between the focal species and the local mimicry rings, assuming strong females choosiness and predators discrimination.** The default parameters values are as follows: $a = 1$, $c_{RI} = 0.01$, $b = 1$, $d = 0.02$, $\lambda = 0.1$, $\overline{\lambda}' = 0.1$, $N = 10$, $N' = 20$, $s = 0.0025$, $t_o = 0$, $\bar{t}' = 1$.

1.6.2 Effect of predation

When a and b are of order ε (H_w),

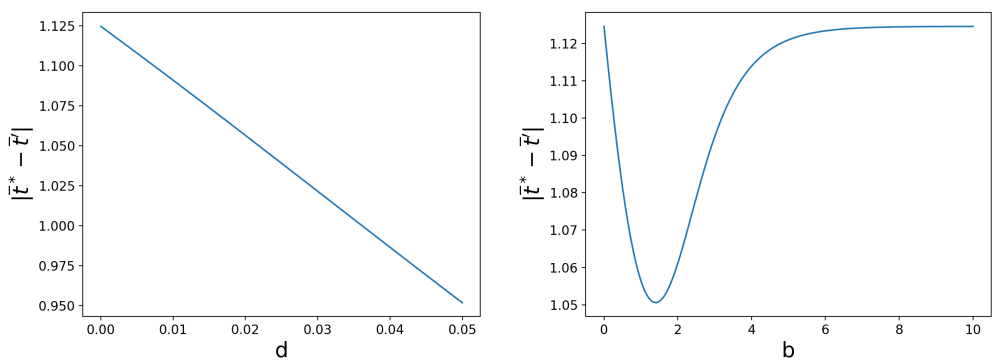
$$\frac{\partial|\bar{t}^* - \bar{t}'|}{\partial d} = \frac{-2b\overline{\lambda}'N'}{(1+\lambda N+\overline{\lambda}'N')^2}\frac{2s|t_o - \bar{t}'|}{\left(\frac{2bd\overline{\lambda}'N'}{(1+\lambda N+\overline{\lambda}'N')^2} + 2s - ac_{RI}\frac{N'}{N}\right)^2} < 0,$$

hence the phenotypic distance decreases with the basal predation rate d. Assuming strong females choosiness and predators discrimination (H_s) leads to similar effect (Figure 5(a)).

When a and b are of order ε (H_w),

$$\frac{\partial|\bar{t}^* - \bar{t}'|}{\partial b} = \frac{-2d\overline{\lambda}' N'}{(1+\lambda N + \overline{\lambda}' N')^2} \frac{2s|t_o - \bar{t}'|}{\left(\frac{2bd\overline{\lambda}' N'}{(1+\lambda N+\overline{\lambda}' N')^2} + 2s - ac_{RI}\frac{N'}{N}\right)^2} < 0.$$

An increase on predator discriminating ability b thus also promotes phenotypic convergence as it increases the positive selection on mimicry generated by predation. Surprisingly, when a and b are of order 1 and predators discrimination exceeds a certain threshold, increased discrimination no longer promotes accurate mimicry (Figure 5(b)). When b is high, individuals from the focal species benefit from protection gained by mimicry only when they look very similar to the individuals of the local mimicry ring. RI makes such high similarity too costly, and therefore divergence of the trait in the focal species becomes promoted (Figure 5(b)).



(a) (b)

Figure 5: **Influence of the strength of predation d (a) and of predators discrimination b (b) on the phenotypic distances between the focal species and the local mimicry rings, assuming strong females choosiness and predators discrimination.** The default parameters values are as follows: $a = 1$, $c_{RI} = 0.01$, $b = 1$, $d = 0.02$, $\lambda = 0.1$, $\overline{\lambda}' = 0.1$, $N = 10$, $N' = 20$, $s = 0.0025$, $t_o = 0$, $\bar{t}' = 1$.

1.6.3 Effect of the defence level of the focal species

Without RI ($c_{RI} = 0$) and when a and b are of order ε (H_w), we have

$$\frac{\partial|\bar{t}^* - \bar{t}'|}{\partial \lambda N} = \frac{2s|t_o - \bar{t}'|}{\left(\frac{2bd\bar{\lambda}'N'}{(1+\lambda N+\bar{\lambda}'N')^2} + 2s\right)} \frac{4bd\bar{\lambda}'N'}{(1+\lambda N+\bar{\lambda}'N')^3} > 0.$$

Therefore, the more defended the focal species is, the greater is the phenotypic distance between the focal species and the local mimicry ring.

Assuming RI generates cost ($c_{RI} > 0$), the density of the focal species then modulates the fitness cost incurred by RI. Therefore, we study the effect of the two components of the level of defence of the focal species: the individual defence level λ and the density N.

$$\frac{\partial|\bar{t}^* - \bar{t}'|}{\partial \lambda} = \frac{2bd\bar{\lambda}'N'N}{(1+\lambda N+\bar{\lambda}'N')^3} \frac{2s|t_o - \bar{t}'|}{\left(\frac{2bd\bar{\lambda}'N'}{(1+\lambda N+\bar{\lambda}'N')^2} + 2s - ac_{RI}\frac{N'}{N}\right)^2} > 0,$$

indicates that the individual defence level increases the phenotypic distance. Assuming strong females choosiness and predators discrimination leads to similar effect (Figure 6).

We also have

$$\frac{\partial|\bar{t}^* - \bar{t}'|}{\partial N} = -\frac{1}{N}\left(ac_{RI}\frac{N'}{N} - \frac{4bd(\lambda N)(\bar{\lambda}'N')}{(1+\lambda N+\bar{\lambda}'N')^3}\right) \frac{2s|t_o - \bar{t}'|}{\left(\frac{2bd\bar{\lambda}'N'}{(1+\lambda N+\bar{\lambda}'N')^2} + 2s - ac_{RI}\frac{N'}{N}\right)^2}.$$

While strong divergence is promoted when the focal species has a low abundance, for larger abundances, the effect of the density on the phenotypic distance between the focal species and the mimicry ring depends on the strength of RI and predation (Figure 7). When fitness cost due to RI is great compared to fitness cost due to predation, an increase in the density of the focal species decreases the phenotypic distance, because its reduces the fitness cost due to RI (similar effect observed assuming strong females choosiness and predators discrimination (H_s) (Figure 7(a))). By contrast, when fitness cost due to RI is low compared to fitness cost due to predation, an increase in the density of the focal species

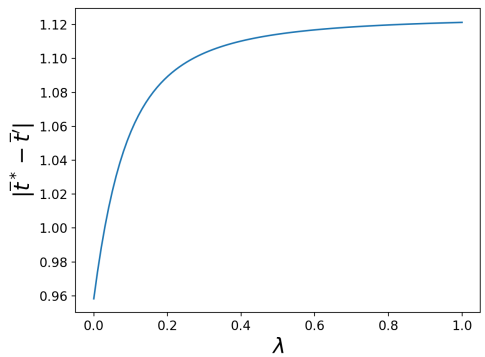


Figure 6: **Influence of the individual defence level λ on the phenotypic distances between the focal species and the local mimicry rings, assuming strong females choosiness and predators discrimination.** The default parameters values are as follows: $a = 1$, $c_{RI} = 0.01$, $b = 1$, $d = 0.02$, $\lambda = 0.1$, $\overline{\lambda}' = 0.1$, $N = 10$, $N' = 20$, $s = 0.0025$, $t_o = 0$, $\bar{t}' = 1$.

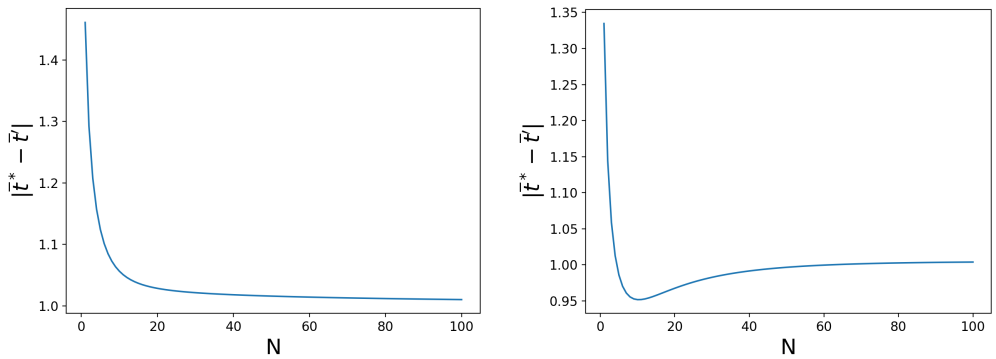
increases the phenotypic distance because the focal species is better defended (similar effect observed assuming strong females choosiness and predators discrimination (H_s) (see right part of Figure 7(b))).

1.6.4 Effect of selective and developmental constraints promoting the optimal trait value t_o.

When a and b are of order ε,

$$\frac{\partial|\bar{t}^* - \bar{t}'|}{\partial s} = \frac{2|t_o - \bar{t}'|\left(\frac{2bd\overline{\lambda}'N'}{(1+\lambda N+\overline{\lambda}'N')^2} - ac_{RI}\frac{N'}{N}\right)}{\left(\frac{2bd\overline{\lambda}'N'}{(1+\lambda N+\overline{\lambda}'N')^2} + 2s - ac_{RI}\frac{N'}{N}\right)^2}.$$

Thus when selection due to predation is stronger than selection due to RI (i.e. $\frac{2bd\overline{\lambda}'N'}{(1+\lambda N+\overline{\lambda}'N')^2} > ac_{RI}\frac{N'}{N}$), selective and developmental constraints promotes the divergence of trait between the focal species and the local mimicry ring (similar effect observed assuming strong females



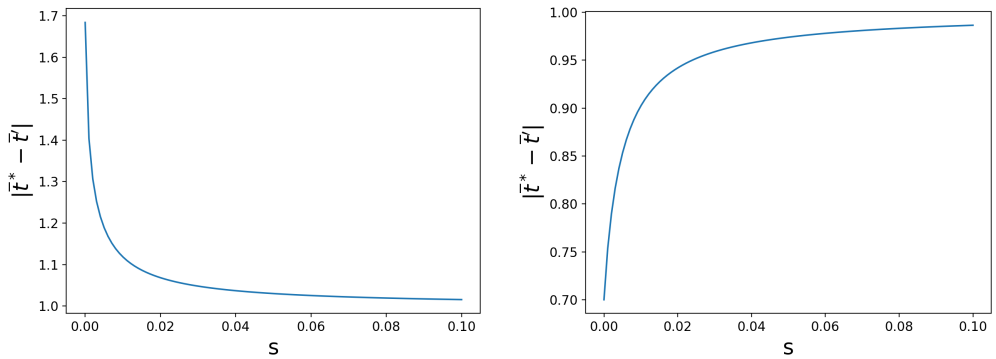
(a) (b)

Figure 7: **Influence of the density of the focal species N on the phenotypic distances between the focal species and the local mimicry rings, when (b) predation is low compared to RI and (b) when predation is high compared to RI, assuming strong females choosiness and predators discrimination.** Simulations were run assuming: (a) $d = 0.02$ or (b) $d = 0.05$. The default parameters values are as follows: $a = 1$, $c_{RI} = 0.01$, $b = 1$, $\lambda = 0.1$, $\overline{\lambda}' = 0.1$, $N = 10$, $N' = 20$, $s = 0.0025$, $t_o = 0$, $\bar{t}' = 1$.

choosiness and predators discrimination (Figure 8(b))). By contrast, when the selection due to RI is stronger than selection due to predation (i.e. $ac_{RI}\frac{N'}{N} > \frac{2bd\overline{\lambda}'N'}{(1+\lambda N+\overline{\lambda}'N')^2}$), selective and developmental constraints limits the trait divergence due to RI (similar effect observed assuming strong females choosiness and predators discrimination (Figure 8(a))).

1.6.5 Effect of the mimetic community.

Finally, we investigate the impact of the composition of the local mimicry ring (density N' and defense level λ') on trait divergence. When a and b are of order ε (H_w), we have:



(a) (b)

Figure 8: **Influence of the strength of selective constraints s on the phenotypic distances between the focal species and the local mimicry rings, when (b) predation is low compared to RI and (b) when predation is great compared to RI, assuming strong females choosiness and predators discrimination.** Simulations were run assuming: (a) $d = 0.02$ or (b) $d = 0.05$. The default parameters values are as follows: $a = 1$, $c_{RI} = 0.01$, $b = 1$, $\lambda = 0.1$, $\overline{\lambda}' = 0.1$, $N = 10$, $N' = 20$, $s = 0.0025$, $t_o = 0$, $\bar{t}' = 1$.

$$\frac{\partial|\bar{t}^* - \bar{t}'|}{\partial\overline{\lambda}'} = \frac{2bdN'(\overline{\lambda}'N' - 1 - \lambda N)}{(1 + \lambda N + \overline{\lambda}'N')^3} \frac{2s|t_o - \bar{t}'|}{\left(\frac{2bd\overline{\lambda}'N'}{(1+\lambda N+\overline{\lambda}'N')^2} + 2s - ac_{RI}\frac{N'}{N}\right)^2},$$

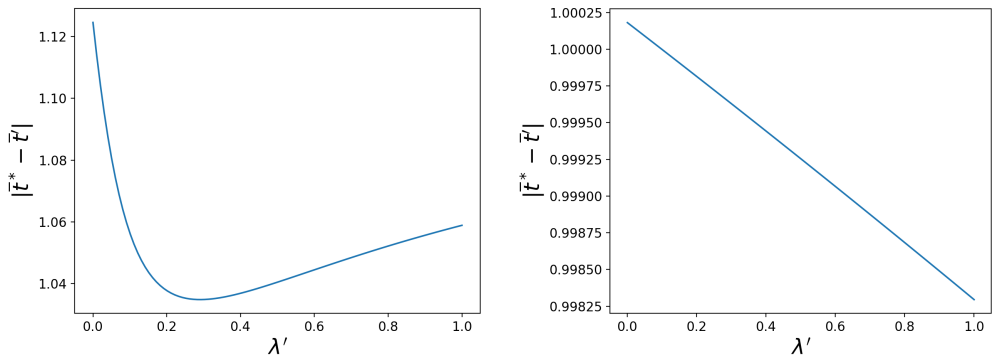
and

$$\frac{\partial|\bar{t}^* - \bar{t}'|}{\partial N'} = \left(\frac{2bd\overline{\lambda}'(\overline{\lambda}'N' - 1 - \lambda N)}{(1 + \lambda N + \overline{\lambda}'N')^3} + \frac{ac_{RI}}{N}\right) \frac{2s|t_o - \bar{t}'|}{\left(\frac{2bd\overline{\lambda}'N'}{(1+\lambda N+\overline{\lambda}'N')^2} + 2s - ac_{RI}\frac{N'}{N}\right)^2}.$$

Thus, when the defence level of the local mimicry ring ($\overline{\lambda}'N'$) is lower than the defence level of the focal species plus one ($\overline{\lambda}'N' < 1 + \lambda N$), an increase in the defence level of individuals of the local mimicry ring promotes phenotypic similarity between the focal

species and the local mimicry ring (similar effect observed assuming strong but moderate females choosiness and predators discrimination (H_s) (Figure 9(a))). Indeed, an increase of $\overline{\lambda}'N'$ increases the advantage of looking similar to individuals of the local mimicry ring. More surprisingly, when the defence level of the local mimicry ring is greater ($\overline{\lambda}'N' > 1 + \lambda N$), the phenotypic distance increases with $\overline{\lambda}'N'$ (similar effect observed assuming strong but moderate females choosiness and predators discrimination (Figure 9(a))). Under weak ($b = O(\varepsilon)$) or strong but moderate (*e.g.* $b = 1$) predators discrimination, predators tend to generalize mimetic trait. Then individuals of the focal species gain at least a little protection from the local mimicry ring, even when displaying a imperfectly mimetic trait. When the level of protection of the mimicry ring is high, the fitness costs linked to predation is low compared to the RI and decreases with the defence level of individuals of the local mimicry ring. This decrease in selection pressure generated by predation decreases the phenotype similarity between the focal species and the local mimicry ring. When predators discrimination is strong and high (*e.g.* $b = 10$), there is no generalization of imperfect mimics by predators, this effect is no longer observed (see Figure 9(b)).

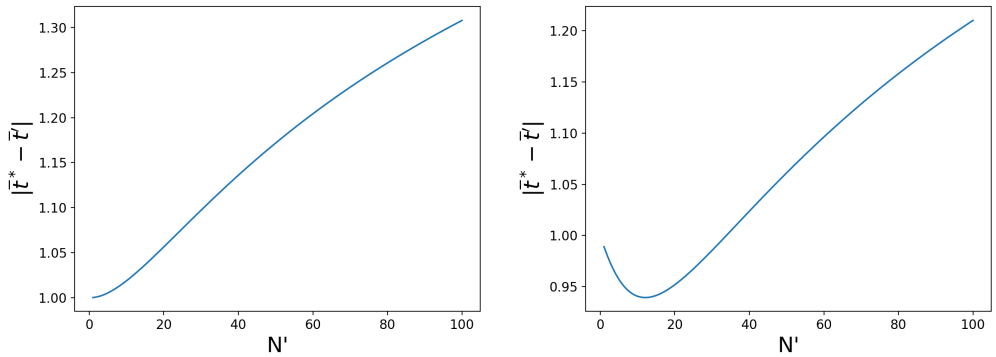
The density N' and the individual defence level of the local mimicry ring λ' have the same effect on the overall defence level of the local mimicry ring (function of the parameter $\overline{\lambda}'N'$), and therefore have a similar effect on the phenotypic distance. However, the density of the local mimicry ring N' also plays a role on the fitness costs generated by RI. When the strength of RI is high as compared to predation, the phenotypic distance increases with the density of the local mimicry ring (N'), because it increases the cost of RI (similar effect observed assuming strong females choosiness and predators discrimination (Figure 10(a))). When the strength of RI is low compared to predation, the density of the local mimicry ring (N') has a similar effect than the defence level of individuals of the local mimicry



(a) (b)

Figure 9: **Influence of the mean defence level of individuals of the local mimicry ring $\overline{\lambda}'$ on the phenotypic distances between the focal species and the local mimicry rings for (a) moderate ($b = 1$) and (b) high ($b = 10$) predators discrimination, assuming strong females choosiness and predators discrimination (*i.e.* $a = O(1)$ and $b = O(1)$).** Simulations were run assuming: (a) $a = 1$ and (b) $a = 10$. The default parameters values are as follows: $a = 1$, $c_{RI} = 0.01$, $d = 0.02$, $\lambda = 0.1$, $\overline{\lambda}' = 0.1$, $N = 10$, $N' = 20$, $s = 0.0025$, $t_o = 0$, $\bar{t}' = 1$.

ring λ' on the phenotypic distance (similar effect observed assuming strong but moderate females choosiness and predators discrimination (Figure 10(b))).



(a) (b)

Figure 10: **Influence of the density of the local mimicry ring N' on the phenotypic distances between the focal species and the local mimicry rings, when (b) predation is low compared to RI and (b) when predation is high compared to RI, assuming strong females choosiness and predators discrimination.** Simulations were run assuming: (a) $d = 0.02$ or (b) $d = 0.05$. The default parameters values are as follows: $a = 1$, $c_{RI} = 0.01$, $b = 1$, $\lambda = 0.1$, $\overline{\lambda}' = 0.1$, $N = 10$, $N' = 20$, $s = 0.0025$, $t_o = 0$, $\bar{t}' = 1$.

2 Model 2: Evolution of mimicry between two interacting species

2.1 Parameters of the model

The parameters used in the model are summed up in Table 1 .

Abbreviation	Description
$\bar{t}_i/\bar{p}_i$	Mean trait/preference value in the species i
G_{t_i}/G_{p_i}	Genetic variance of trait t_i/preference p_i
$C_{t_i p_i}$	Genetic covariance between trait t_i and preference p_i
β_{t_i}/β_{p_i}	Selection coefficient on trait t_i/preference p_i
a_i	Females choosiness in species i
s_i	Strength of developmental constraints in species i
t_{oi}	Optimal trait favored by developmental constraints in species i
d	Basal predation rate
b	Predators discrimination
λ_i	Unpalatability coefficient of individuals of the species i
N_i	Density of the species i
c_{RI}	Strength of reproductive interference

Table 1: Description of variables and parameters used in the model with two species. The subscript $i \in \{1, 2\}$ denotes the species.

2.2 Selection coefficients

For $i \in \{1,2\}$, the selection coefficients are defined by:

$$\beta_{t_i} = \frac{\partial}{\partial t_i} \ln\left(W_m^i(t_i|\overline{t_1},\bar{t}_2,\overline{p}_i)\right) + \frac{\partial}{\partial t_i} \ln\left(W_f^i(t_i,p_i|\bar{t}_1,\bar{t}_2)\right)\Big|_{(t_i,p_i)=(\bar{t}_i,\overline{p}_i)},$$

$$\beta_{p_i} = \frac{\partial}{\partial p_i} \ln\left(W_f^i(t_i,p_i|\bar{t}_1,\bar{t}_2)\right)\Big|_{(t_i,p_i)=(\bar{t}_i,\overline{p}_i)},$$

where W_m^i and W_f^i are the fitness of males and females in species i, and write

$$W_m^i(t_i|\overline{t_1},\bar{t}_2,\overline{p}_i) = \exp\left[-a_i(t_i-\overline{p}_i)\right]W_{pred}(t_i|\bar{t}_1,\bar{t}_2)\exp\left[-s_i(t_i-t_{oi})\right],$$

$$W_f^i(t_i,p_i|\bar{t}_1,\bar{t}_2) = \exp\left[-a_i(p_i-\bar{t}_i)\right]W_{RI}^i(p_i|\bar{t}_1,\bar{t}_2)W_{pred}(t_i|\bar{t}_1,\bar{t}_2)\exp\left[-s_i(t_i-t_{oi})\right].$$

Here W_{pred} is the impact of predation on fitness and W_{RI}^i is the impact RI on fitness of individuals of species i.

2.2.1 Computation of β_{t_1} and β_{t_2}

Similarly as the case of one population in a mimicry ring studied previously (recall the definition of W_{pred} in (7)), the impact of predation on an individual of trait t is:

$$\begin{aligned}\frac{\partial}{\partial t_1}\ln\left(W_{pred}^2(t_1|\bar{t}_1,\bar{t}_2)\right) = &-2b\Big((t_1-\bar{t}_1)\lambda_1\exp\left[-b(t_1-\bar{t}_1)^2\right]N_1 + (t_2-\bar{t}_2)\overline{\lambda}_2\exp\left[-b(t_2-\bar{t}_2)^2\right]N_2\Big)\\ &\times\Bigg(\frac{1}{1+\lambda_1N_1\exp\left[-b(t_2-\bar{t}_1)^2\right]+\lambda_2\exp\left[-b(t_2-\bar{t}_2)^2\right]N_2-d}\\ &-\frac{1}{1+\lambda_1N_1\exp\left[-b(t_2-\bar{t}_1)^2\right]+\lambda_2\exp\left[-b(t_2-\bar{t}_2)^2\right]N_2}\Bigg).\end{aligned}$$

By symmetry we obtain the expression of $\frac{\partial}{\partial t_2}\ln\left(W_{pred}^2(t_2|\bar{t}_1,\bar{t}_2)\right)$ by exchanging 1 and 2. When b is of order ε (H_w), similar computations as in the case of one population in a

mimicry ring yield:

$$\beta_{t_1} = -2a_1(\bar{t}_1 - \bar{p}_1) - \frac{4bd(\bar{t}_1 - \bar{t}_2)\lambda_2 N_2}{(1 + \lambda_1 N_1 + \lambda_2 N_2)^2} - 4s_1(\bar{t}_1 - t_{o1}) + O(\varepsilon^3), \tag{A27}$$

$$\beta_{t_2} = -2a_2(\bar{t}_2 - \bar{p}_2) - \frac{4bd(\bar{t}_2 - \bar{t}_1)\lambda_1 N_1}{(1 + \lambda_1 N_1 + \lambda_2 N_2)^2} - 4s_2(\bar{t}_2 - t_{o2}) + O(\varepsilon^3). \tag{A28}$$

2.2.2 Computation of β_{p_1} and β_{p_2}

Let $W^i_{RI}(p|\bar{t}_1, \bar{t}_2)$ indicate the impact of RI on a female of species i with preference p. If we denote by j the complement of i in $\{1, 2\}$, we get:

$$W^i_{RI}(p_i|\bar{t}_1, \bar{t}_2) = \frac{\exp\left[-a_i(p_i - \bar{t}_i)^2\right] N_i}{\exp\left[-a_i(p_i - \bar{t}_i)^2\right] N_i + c_{RI} \exp\left[-a_j(p_j - \bar{t}_j)^2)\right] N_j}.$$

In the same way as in the previous model we thus get, when we assume that a_1, a_2 and c_{RI} are of order ε (H_w),

$$\beta_{p_1} = -2a_1(\bar{p}_1 - \bar{t}_1) + 2c_{RI}a_1(\bar{t}_1 - \bar{t}_2)\frac{N_2}{N_1} + O(\varepsilon^3), \tag{A29}$$

$$\beta_{p_2} = -2a_2(\bar{p}_2 - \bar{t}_2) + 2c_{RI}a_2(\bar{t}_2 - \bar{t}_1)\frac{N_1}{N_2} + O(\varepsilon^3). \tag{A30}$$

2.3 Quasi equilibria

We search mean values of traits and preferences at equilibrium $\bar{t}^*_1$, $\bar{p}^*_1$, $\bar{t}^*_2$ and $\bar{p}^*_2$ such as the leading terms of $\Delta\bar{t}^*_1$, $\Delta\bar{p}^*_1$, $\Delta\bar{t}^*_2$ and $\Delta\bar{p}^*_2$ are null, or equivalently, $\tilde{\beta}_{t_1} = \tilde{\beta}_{p_1} = \tilde{\beta}_{t_2} = \tilde{\beta}_{p_2} = 0$ with

$$\tilde{\beta}_{t_1} = -2a_1(\bar{t}^*_1 - \bar{p}^*_1) - \frac{4bd(\bar{t}^*_1 - \bar{t}^*_2)\lambda_2 N_2}{(1 + \lambda_1 N_1 + \lambda_2 N_2)^2} - 4s_1(\bar{t}^*_1 - t_{o1}), \tag{A31}$$

$$\tilde{\beta}_{p_1} = -2a_1(\bar{p}^*_1 - \bar{t}^*_1) + 2c_{RI}a_1(\bar{t}^*_1 - \bar{t}^*_2)\frac{N_2}{N_1}, \tag{A32}$$

$$\tilde{\beta}_{t_2} = -2a_2(\bar{t}^*_2 - \bar{p}^*_2) - \frac{4bd(\bar{t}^*_2 - \bar{t}^*_1)\lambda_1 N_1}{(1 + \lambda_1 N_1 + \lambda_2 N_2)^2} - 4s_2(\bar{t}^*_2 - t_{o2}), \tag{A33}$$

$$\tilde{\beta}_{p_2} = -2a_2(\bar{p}^*_2 - \bar{t}^*_2) + 2c_{RI}a_2(\bar{t}^*_2 - \bar{t}^*_1)\frac{N_1}{N_2}. \tag{A34}$$

Lemma 3. *Let us introduce for $i \neq j \in \{1,2\}^2$,*

$$A_i = \frac{2bd\lambda_j N_j - a_i c_{RI} \frac{N_j}{N_i} \left(1 + \lambda_1 N_1 + \lambda_2 N_2\right)^2}{2bd\lambda_j N_j + \left(2s_i - a_i c_{RI} \frac{N_j}{N_i}\right) \left(1 + \lambda_1 N_1 + \lambda_2 N_2\right)^2}, \tag{A35}$$

$$B_i = \frac{2s_i t_{oi} \left(1 + \lambda_1 N_1 + \lambda_2 N_2\right)^2}{2bd\lambda_j N_j + \left(2s_i - a_i c_{RI} \frac{N_j}{N_i}\right) \left(1 + \lambda_1 N_1 + \lambda_2 N_2\right)^2}. \tag{A36}$$

If $2bd\lambda_j N_j + \left(2s_i - a_i c_{RI}(N_j/N_i)\right) \left(1 + \lambda_1 N_1 + \lambda_2 N_2\right)^2 \neq 0$ for $i \in \{1,2\}$ and if $A_1 A_2 \neq 1$ there is a unique quasi steady point given for $i \in \{1,2\}$ by

$$\bar{t}_i^* = \frac{B_i + A_i B_j}{1 - A_1 A_2}, \quad \bar{p}_i^* = \bar{t}_i^* + c_{RI}(\bar{t}_i^* - \bar{t}_j^*) \frac{N_j}{N_i}, \tag{A37}$$

Remark. We do not give the quasi equilibrium traits and preferences when one of the following conditions is verified:

- $2bd\lambda_2 N_2 + \left(2s_1 - a_1 c_{RI}(N_2/N_1)\right) \left(1 + \lambda_1 N_1 + \lambda_2 N_2\right)^2 = 0,$
- $2bd\lambda_1 N_1 + \left(2s_2 - a_2 c_{RI}(N_1/N_2)\right) \left(1 + \lambda_1 N_1 + \lambda_2 N_2\right)^2 = 0,$
- $A_1 A_2 = 1.$

Indeed investigating those cases leads to a long enumeration of subcases with different quasi steady points, and those cases are negligible compared with all the possible combinations of parameters. We thus choose to not give this enumeration.

Proof. Let $2bd\lambda_j N_j + \left(2s_i - a_i c_{RI}(N_j/N_i)\right) \left(1 + \lambda_1 N_1 + \lambda_2 N_2\right)^2 \neq 0$ for $i \neq j \in \{1,2\}^2$. Using the results of the first model (see Appendix 1.3) we have

$$\bar{t}_1^* = \frac{\frac{2bd\lambda_2 N_2}{(1+\lambda_1 N_1 + \lambda_2 N_2)^2} \bar{t}_2^* + 2s_1 t_{o1} - a_1 c_{RI} \frac{N2}{N1} \bar{t}_2^*}{\frac{2bd\lambda_2 N_2}{(1+\lambda_1 N_1 + \lambda_2 N_2)^2} + 2s_1 - a_1 c_{RI} \frac{N_2}{N_1}},$$

and

$$\bar{t}_2^* = \frac{\frac{2bd\lambda_1 N_1}{(1+\lambda_1 N_1 + \lambda_2 N_2)^2} \bar{t}_1^* + 2s_2 t_{o2} - a_2 c_{RI} \frac{N1}{N2} \bar{t}_1^*}{\frac{2bd\lambda_1 N_1}{(1+\lambda_1 N_1 + \lambda_2 N_2)^2} + 2s_2 - a_2 c_{RI} \frac{N_1}{N_2}},$$

which can be rewritten

$$\bar{t}_1^* = A_1\bar{t}_2^* + B_1, \quad \bar{t}_2^* = A_2\bar{t}_1^* + B_2,$$

where $(A_i, B_i), i \in \{1, 2\}$ have been defined in (A35)-(A36).

If $A_1A_2 \neq 1$ there are unique values for traits at quasi equilibrium given by $\bar{t}_1^*$ and $\bar{t}_2^*$ defined in (A37). Using previous results we have the associated mean preferences at quasi equilibrium $\bar{p}_1^*$ and $\bar{p}_2^*$ defined in (A37).

□

2.4 Fast and slow dynamics.

Lemma 4. *Let us introduce*

$$B_{pred} = \frac{2bd}{(1 + \lambda_1 N_1 + \lambda_2 N_2)^2},$$
$$C_1 = 2\lambda_2 N_2 B_{pred} + 4s_1 - 2c_{RI}a_1\frac{N_2}{N_1},$$
$$C_2 = 2\lambda_1 N_1 B_{pred} + 4s_2 - 2c_{RI}a_2\frac{N_1}{N_2},$$

and recall the definitions of A_1 and A_2 in (A35).

1. *If $C_1 > 0$, $C_2 > 0$ and $C_1C_2(1 - A_1A_2) - \frac{1}{4}(A_1C_1 - A_2C_2)^2 > 0$ the quantities $\bar{t}_1 - \bar{t}_1^*$, $\bar{t}_2 - \bar{t}_2^*$, $\bar{p}_1 - \bar{p}_1^*$ and $\bar{p}_2 - \bar{p}_2^*$ become of order $\sqrt{\varepsilon}$ when the number of generation goes to infinity, with $\bar{t}_1^*$, $\bar{t}_2^*$, $\bar{p}_1^*$ and $\bar{p}_2^*$ defined by (A37).*

2. *If $C_1 < 0$, $C_2 < 0$ and $C_1C_2(1 - A_1A_2) - \frac{1}{4}(A_1C_1 - A_2C_2)^2 > 0$, traits and preferences become very large.*

Remark. When $C_1C_2(1 - A_1A_2) - \frac{1}{4}(A_1C_1 - A_2C_2)^2 > 0$ or $C_1C_2 < 0$ we are not able to analytically determine the convergence or the divergence of traits and preferences. To

infer such information we simulate the dynamic of traits and preferences considering the leading term of the selection coefficients.

Proof. As previously, we can decompose the dynamics into two steps. In the first one the leading order terms of the selection coefficient are the terms describing sexual selection and cost of choosiness. Using computations similar to the ones in the previous model we can approximate $\Delta(\bar{t}_i - \overline{p}_i)^2$ for $i \in \{1, 2\}$ by

$$\Delta(\bar{t}_i - \overline{p}_i)^2 = -(\bar{t}_i - \overline{p}_i)\Big(2a_i(G_{t_i} + G_{p_i})(\bar{t}_i - \overline{p}_i) + O(\varepsilon^2)\Big) + O(\varepsilon^3).$$

Using the same reasoning that in the previous model we may prove that above a certain number of generations $t_1 - p_1$ and $t_2 - p_2$ are of order ε.

We now look at the variation of

$$\mathcal{A} := \frac{(\bar{t}_1 - \bar{t}_1^*)^2}{G_{t_1}} + \frac{(\bar{t}_2 - \bar{t}_2^*)^2}{G_{t_2}} + \frac{(\overline{p}_1 - \overline{p}_1^*)^2}{G_{p_1}} + \frac{(\overline{p}_2 - \overline{p}_2^*)^2}{G_{p_2}},$$

with $\bar{t}_1^*$, $\bar{t}_2^*$, $\overline{p}_1^*$ and $\overline{p}_2^*$ defined by (A37).

By definition,

$$\begin{aligned}\Delta\mathcal{A} =& \frac{2\Delta\bar{t}_1(\bar{t}_1 - \bar{t}_1^*) + \left(\Delta\bar{t}_1\right)^2}{G_{t_1}} + \frac{2\Delta\bar{t}_2(\bar{t}_2 - \bar{t}_2^*) + \left(\Delta\bar{t}_2\right)^2}{G_{t_2}} \\ &+ \frac{2\Delta\overline{p}_1(\overline{p}_1 - \overline{p}_1^*) + (\Delta\overline{p}_1)^2}{G_{p_1}} + \frac{2\Delta\overline{p}_2(\overline{p}_2 - \overline{p}_2^*) + (\Delta\overline{p}_2)^2}{G_{p_2}}.\end{aligned}$$

When $\bar{t}_1 - \overline{p}_1$ and $\bar{t}_2 - \overline{p}_2$ are of order ε, $\Delta\bar{t}_1$, $\Delta\bar{t}_2$, $\Delta\overline{p}_1$ and $\Delta\overline{p}_2$ are of order ε^2. We then neglect the terms proportional to $\left(\Delta\bar{t}_1\right)^2$, $\left(\Delta\bar{t}_2\right)^2$, $(\Delta\overline{p}_1)^2$ or $(\Delta\overline{p}_2)^2$:

$$\Delta\mathcal{A} = \frac{2\Delta\bar{t}_1(\bar{t}_1 - \bar{t}_1^*)}{G_{t_1}} + \frac{2\Delta\bar{t}_2(\bar{t}_2 - \bar{t}_2^*)}{G_{t_2}} + \frac{2\Delta\overline{p}_1(\overline{p}_1 - \overline{p}_1^*)}{G_{p_1}} + \frac{2\Delta\overline{p}_2(\overline{p}_2 - \overline{p}_2^*)}{G_{p_2}} + O(\varepsilon^4).$$

By replacing the terms $\Delta\bar{t}_1$, $\Delta\bar{t}_2$, $\Delta\overline{p}_1$ and $\Delta\overline{p}_2$ using (12) and simplifying, we obtain:

$$\Delta\mathcal{A} = \beta_{t_1}(\overline{t_1} - \overline{t_1}^*) + \beta_{p_1}(\overline{p_1} - \overline{p_1}^*) + \beta_{t_2}(\overline{t_2} - \overline{t_2}^*) + \beta_{p_2}(\overline{p_2} - \overline{p_2}^*) + O(\varepsilon^4).$$

Using that for $i \in \{1,2\}$, $\bar{t}_i - \bar{p}_i$ and $\bar{t}_i^* - \bar{p}_i^*$ are of order ε, we obtain similarly as in (A21)

$$\Delta\mathcal{A} = (\beta_{t_1} + \beta_{p_1})\left(\frac{\bar{t}_1 - \bar{t}_1^*}{2} + \frac{\bar{p}_1 - \bar{p}_1^*}{2}\right) + (\beta_{t_2} + \beta_{p_2})\left(\frac{\bar{t}_2 - \bar{t}_2^*}{2} + \frac{\bar{p}_2 - \bar{p}_2^*}{2}\right) + O(\varepsilon^3). \quad \text{(A38)}$$

Using (A27) and (A29) and simplifying, we obtain:

$$\beta_{t_1} + \beta_{p_1} = -\frac{4bd(\bar{t}_1 - \bar{t}_2)\lambda_2 N_2}{(1 + \lambda_1 N_1 + \lambda_2 N_2)^2} - 4s_1(\bar{t}_1 - t_{o1}) + 2c_{RI}a_1(\bar{t}_1 - \bar{t}_2)\frac{N_2}{N_1} + O(\varepsilon^3).$$

By substracting $-\frac{4bd(\bar{t}_1^* - \bar{t}_2^*)\lambda_2 N_2}{(1+\lambda_1 N_1 + \lambda_2 N_2)^2} - 4s_1(\bar{t}_1^* - t_{o1}) + 2c_{RI}a_1(\bar{t}_1^* - \bar{t}_2^*)\frac{N_2}{N_1}$ which is equal to zero (because $\beta_{t_1}|_{(t_1,t_2,p_1,p_2)=(\bar{t}_1^*,\bar{t}_2^*,\bar{p}_1^*,\bar{p}_2^*)} = 0$ and $\beta_{p_1}|_{(t_1,t_2,p_1,p_2)=(\bar{t}_1^*,\bar{t}_2^*,\bar{p}_1^*,\bar{p}_2^*)} = 0$, we get:

$$\begin{aligned}(\beta_{t_1} + \beta_{p_1}) &= \left(-\frac{4bd\lambda_2 N_2}{(1 + \lambda_1 N_1 + \lambda_2 N_2)^2} - 4s_1 + 2c_{RI}a_1\frac{N_2}{N_1}\right)(\bar{t}_1 - \bar{t}_1^*) \\ &\quad - \left(-\frac{4bd\lambda_2 N_2}{(1 + \lambda_1 N_1 + \lambda_2 N_2)^2} + 2c_{RI}a_1\frac{N_2}{N_1}\right)(\bar{t}_2 - \bar{t}_2^*) + O(\varepsilon^3) \\ &= -C_1(\bar{t}_1 - \bar{t}_1^*) + A_1 C_1(\bar{t}_2 - \bar{t}_2^*) + O(\varepsilon^3).\end{aligned}$$

Because for $i \in \{1,2\}, \bar{t}_i - \bar{t}_i^* = \bar{p}_i - \bar{p}_i^* + \bar{t}_i - \bar{p}_i + \bar{p}_i^* - \bar{t}_i^* = \bar{p}_i - \bar{p}_i^* + O(\varepsilon)$ we also have

$$(\beta_{t_1} + \beta_{p_1}) = -C_1(\bar{p}_1 - \bar{p}_1^*) + A_1 C_1(\bar{p}_2 - \bar{p}_2^*) + O(\varepsilon^3).$$

We thus get

$$(\beta_{t_1} + \beta_{p_1})(\bar{t}_1 - \bar{t}_1^*) = -C_1(\bar{t}_1 - \bar{t}_1^*)^2 + A_1 C_1(\bar{t}_1 - \bar{t}_1^*)(\bar{t}_2 - \bar{t}_2^*) + O(\varepsilon^3),$$

and

$$(\beta_{t_1} + \beta_{p_1})(\bar{p}_1 - \bar{p}_1^*) = -C_1(\bar{p}_1 - \bar{p}_1^*)^2 + A_1 C_1(\bar{p}_1 - \bar{p}_1^*)(\bar{p}_2 - \bar{p}_2^*) + O(\varepsilon^3).$$

Hence using (A38), the symmetry of the model between populations 1 and 2, and that for $i \in \{1,2\}$, $\bar{t}_i - \bar{p}_i$ and $\bar{t}_i^* - \bar{p}_i^*$ are of order ε, we obtain:

$$\begin{aligned}2\Delta\mathcal{A} = &-C_1(\bar{t}_1 - \bar{t}_1^*)^2 + (A_1 C_1 + A_2 C_2)(\bar{t}_1 - \bar{t}_1^*)(\bar{t}_2 - \bar{t}_2^*) - C_2(\bar{t}_2 - \bar{t}_2^*)^2 \\ &- C_1(\bar{p}_1 - \bar{p}_1^*)^2 + (A_1 C_1 + A_2 C_2)(\bar{p}_1 - \bar{p}_1^*)(\bar{p}_2 - \bar{p}_2^*) - C_2(\bar{p}_2 - \bar{p}_2^*)^2,\end{aligned}$$

with can be rewritten in matrix form

$$2\Delta\mathcal{A} = -T^t M T - P^t M P + O(\varepsilon^3),$$

with

$$M = \begin{pmatrix} C_1 & -\frac{A_1C_1+A_2C_2}{2} \\ -\frac{A_1C_1+A_2C_2}{2} & C_2 \end{pmatrix}, \quad T = \begin{pmatrix} \bar{t}_1 - \bar{t}_1^* \\ \bar{t}_2 - \bar{t}_2^* \end{pmatrix}, \quad P = \begin{pmatrix} \bar{p}_1 - \bar{p}_1^* \\ \bar{p}_2 - \bar{p}_2^* \end{pmatrix},$$

and T^t and P^t being the transposed of T and P respectively.

First case: $C_1 > 0$, $C_2 > 0$ and $C_1C_2(1 - A_1A_2) - \frac{1}{4}(A_1C_1 - A_2C_2)^2 > 0$.

According to the Sylvester's criterion the matrix M is positive-definite. Using computations similar to the ones derived to prove (A16), we may prove that after a number of generations large enough, $\bar{t}_1 - \bar{t}_1^*$, $\bar{t}_2 - \bar{t}_2^*$, $\bar{p}_1 - \bar{p}_1^*$ and $\bar{p}_2 - \bar{p}_2^*$ become of order $\varepsilon^{1/2}$.

Second case: $C_1 < 0$, $C_2 < 0$ and $C_1C_2(1 - A_1A_2) - \frac{1}{4}(A_1C_1 - A_2C_2)^2 > 0$.

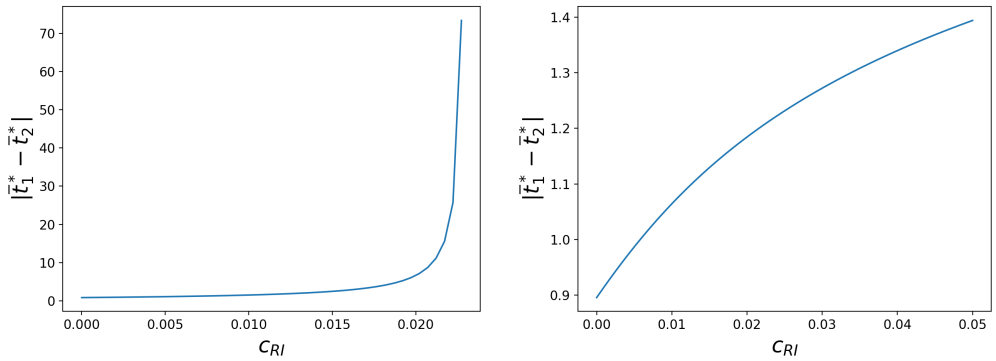
According to the Sylvester's criterion the matrix $-M$ is positive-definite. Therefore $\bar{t}_1$, $\bar{t}_2$, $\bar{p}_1$ and $\bar{p}_2$ become very large.

□

2.5 Effect of different parameters on the level of trait divergence $|\bar{t}_1^* - \bar{t}_2^*|$ assuming weak or strong females choosiness and predators discrimination

We want to verify that the different parameters have similar effect on phenotypic distance between species in this model than in the first. To this aim we plot the effect of different parameters on the level of trait divergence assuming whether (H_w) (using the analytical

approximation when possible or numerical simulations) or (H_s) (using numerical simulations). The following plots confirm that the parameters have a similar effect on the phenotypic distance.



(a) (b)

Figure 11: **Influence of the strength of reproductive interference c_{RI} on the phenotypic distances between the focal species and the local mimicry rings, assuming (a) weak or (b) strong females choosiness and predators discrimination.** Simulations were run assuming (a) $a_1 = a_2 = b = 0.01$ or (b) $a_1 = a_2 = b = 1$. The default parameters values are as follows: $d = 0.02$, $\lambda_1 = 0.1$, $\lambda_2 = 0.1$, $N_1 = 10$, $N_2 = 20$, $s_1 = 0.0000025$, $s_2 = 0.0000025$, $t_{o1} = 0$, $t_{o2} = 1$.

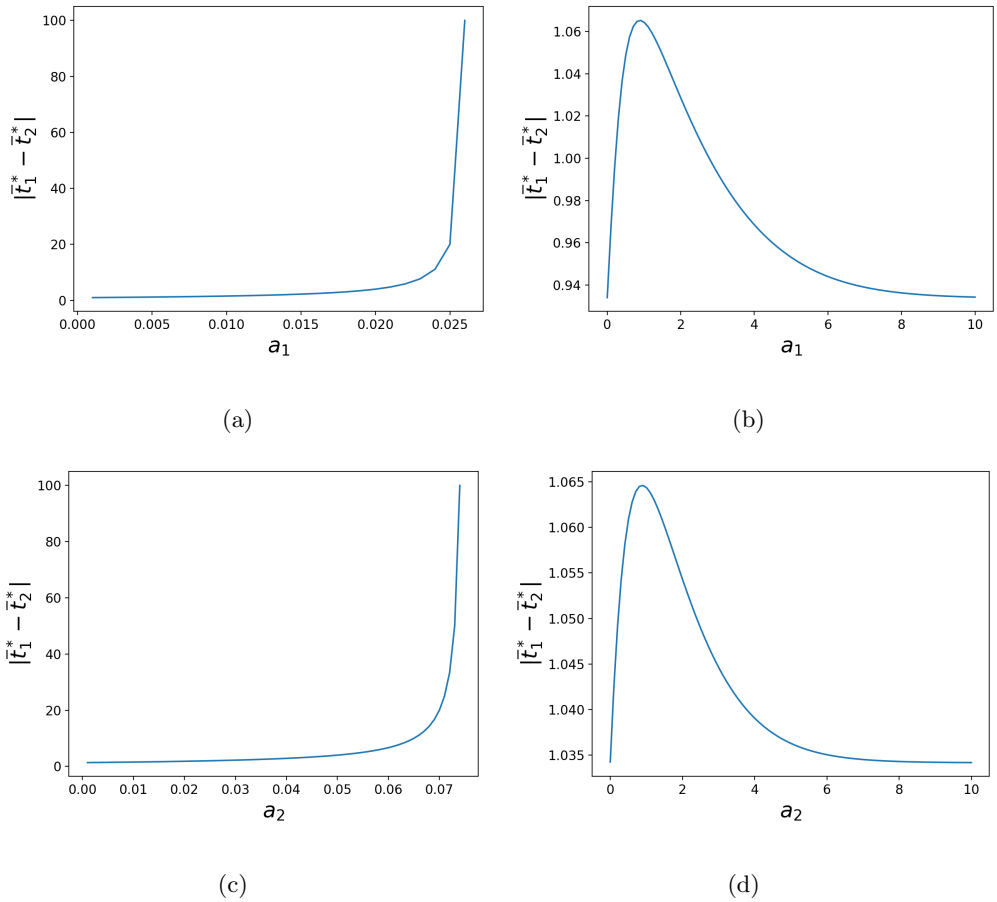
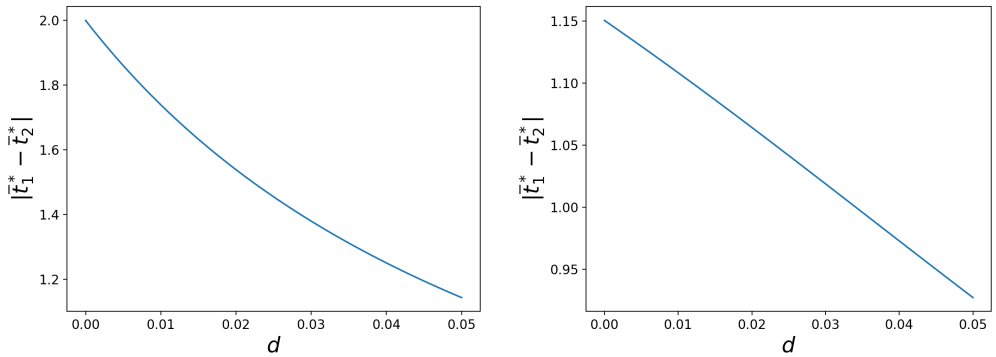
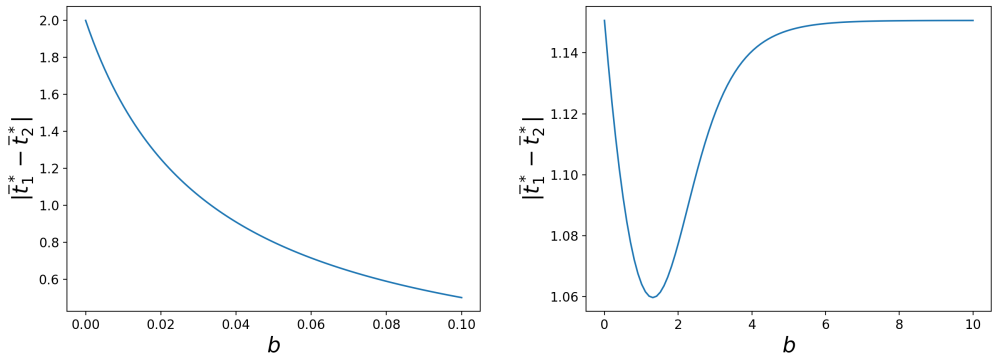


Figure 12: **Influence of females choosiness in species (a)(b) 1 a_1 and (c)(d) 2 a_1 on the phenotypic distances between the focal species and the local mimicry rings, assuming (a)(c) weak or (b)(d) strong females choosiness and predators discrimination.** Simulations were run assuming (a) $a_2 = b = 0.01$, (b) $a_2 = b = 1$, (c) $a_1 = b = 0.01$ or (d) $a_1 = b = 1$. The default parameters values are as follows: $c_{RI} = 0.01$, $d = 0.02$, $\lambda_1 = 0.1$, $\lambda_2 = 0.1$, $N_1 = 10$, $N_2 = 20$, $s_1 = 0.0000025$, $s_2 = 0.0000025$, $t_{o1} = 0$, $t_{o2} = 1$.



(a) (b)

Figure 13: **Influence of the strength of predation d on the phenotypic distances between the focal species and the local mimicry rings, assuming (a) weak or (b) strong females choosiness and predators discrimination.** Simulations were run assuming (a) $a_1 = a_2 = b = 0.01$, (b) $a_1 = a_2 = b = 1$. The default parameters values are as follows: $c_{RI} = 0.01$, $\lambda_1 = 0.1$, $\lambda_2 = 0.1$, $N_1 = 10$, $N_2 = 20$, $s_1 = 0.0000025$, $s_2 = 0.0000025$, $t_{o1} = 0$, $t_{o2} = 1$.



(a) (b)

Figure 14: **Influence of predators discrimination b on the phenotypic distances between the focal species and the local mimicry rings, assuming (a) weak or (b) strong females choosiness and predators discrimination.** Simulations were run assuming (a) $a_1 = a_2 = 0.01$, (b) $a_1 = a_2 = 1$. The default parameters values are as follows: $c_{RI} = 0.01$, $d = 0.02$, $\lambda_1 = 0.1$, $\lambda_2 = 0.1$, $N_1 = 10$, $N_2 = 20$, $s_1 = 0.0000025$, $s_2 = 0.0000025$, $t_{o1} = 0$, $t_{o2} = 1$.

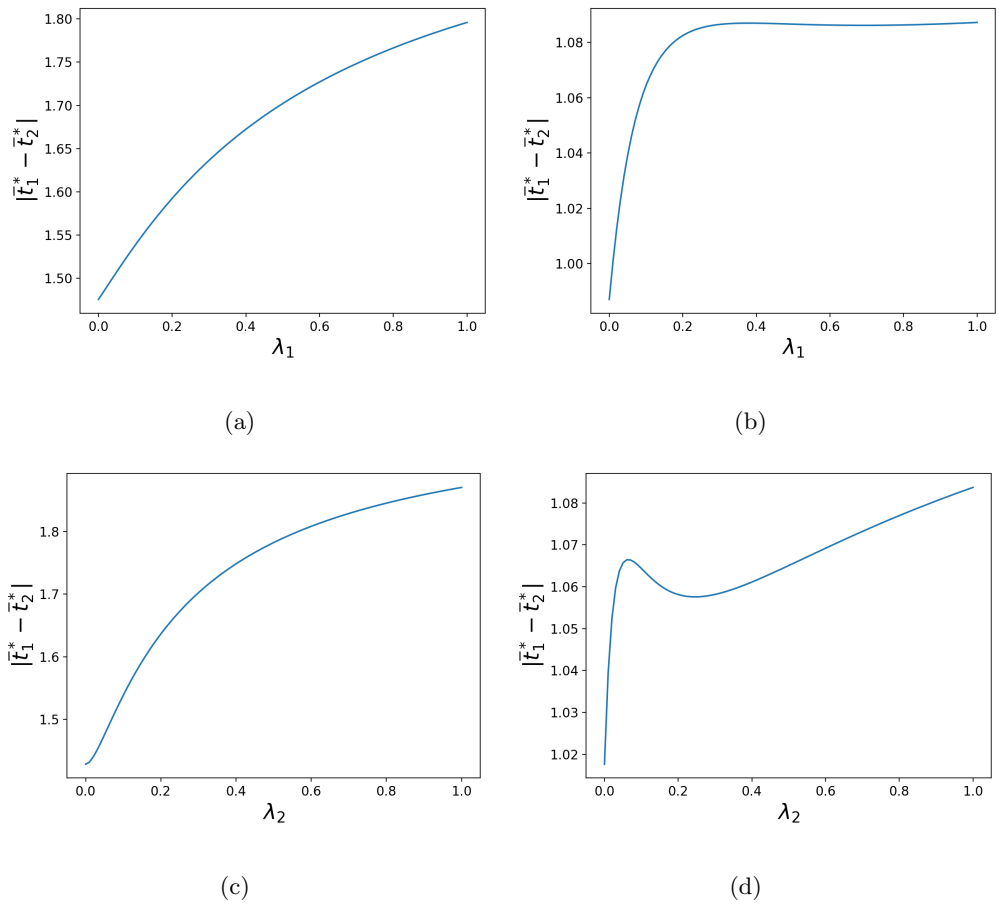


Figure 15: **Influence of the defence level of (a)(b) species 1 λ_1 and of (c)(d) species 2 λ_2 on the phenotypic distances between both species, assuming (a)(c) weak or (b)(d) strong females choosiness and predators discrimination.** Simulations were run assuming (a)(c) $a_1 = a_2 = b = 0.01$, (b)(d) $a_1 = a_2 = b = 1$. The default parameters values are as follows: $c_{RI} = 0.01$, $d = 0.02$, $\lambda_1 = 0.1$, $\lambda_2 = 0.1$, $N_1 = 10$, $N_2 = 20$, $s_1 = 0.0000025$, $s_2 = 0.0000025$, $t_{o1} = 0$, $t_{o2} = 1$.

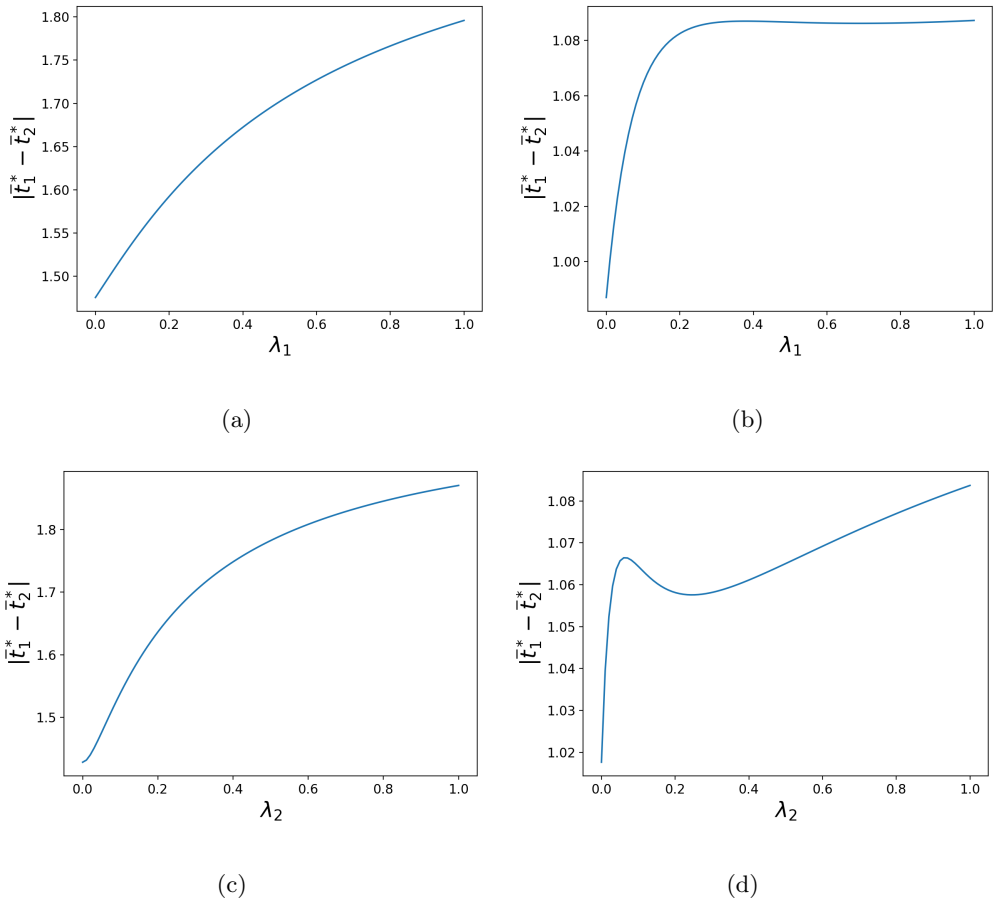


Figure 16: **Influence of the defence level of (a)(b) species 1 λ_1 and of (c)(d) species 2 λ_2 on the phenotypic distances between both species, assuming (a)(c) weak or (b)(d) strong females choosiness and predators discrimination.** Simulations were run assuming (a)(c) $a_1 = a_2 = b = 0.01$, (b)(d) $a_1 = a_2 = b = 1$. The default parameters values are as follows: $c_{RI} = 0.01$, $d = 0.02$, $\lambda_1 = 0.1$, $\lambda_2 = 0.1$, $N_1 = 10$, $N_2 = 20$, $s_1 = 0.0000025$, $s_2 = 0.0000025$, $t_{o1} = 0$, $t_{o2} = 1$.

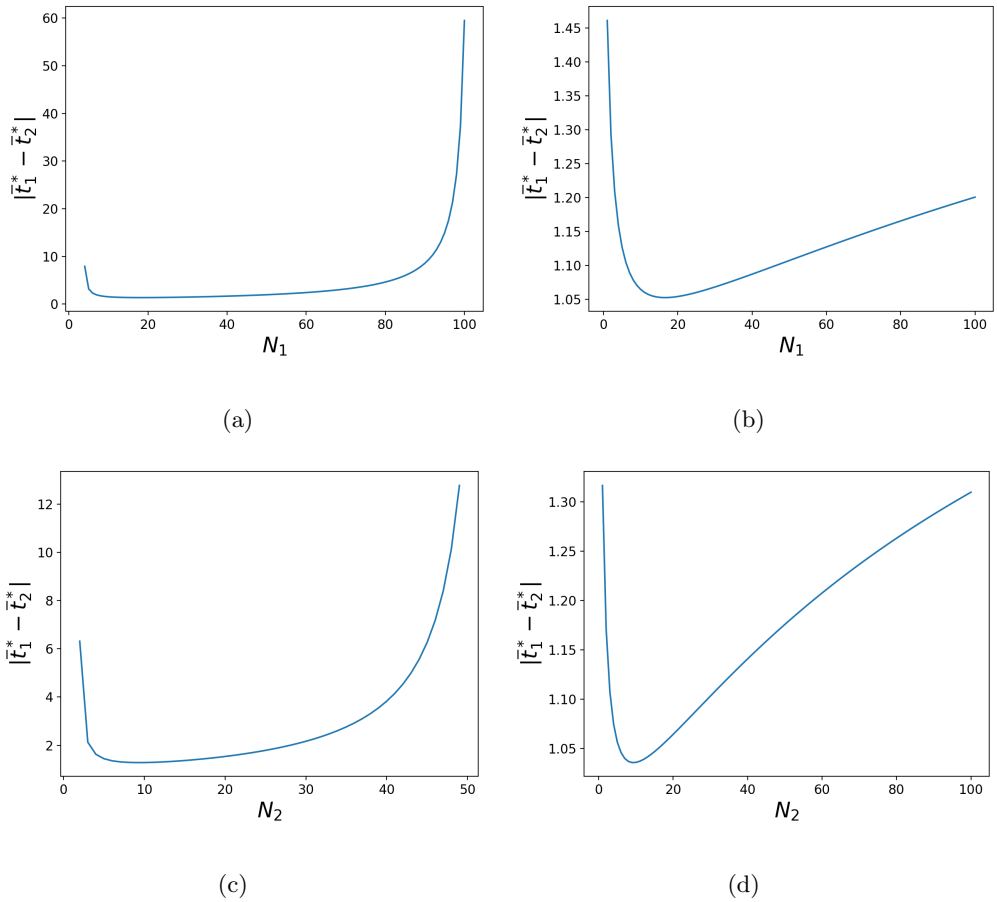


Figure 17: **Influence of the density of (a)(b) species 1 N_1 and of (c)(d) species 2 N_2 on the phenotypic distances between the focal species and the local mimicry rings, assuming (a)(c) weak or (b)(d) strong females choosiness and predators discrimination.** Simulations were run assuming (a) $a_1 = a_2 = b = 0.01$, (b) $a_1 = a_2 = b = 1$. The default parameters values are as follows: $c_{RI} = 0.01$, $d = 0.02$, $\lambda_1 = 0.1$, $\lambda_2 = 0.1$, $N_1 = 10$, $N_2 = 20$, $s_1 = 0.0000025$, $s_2 = 0.0000025$, $t_{o1} = 0$, $t_{o2} = 1$.

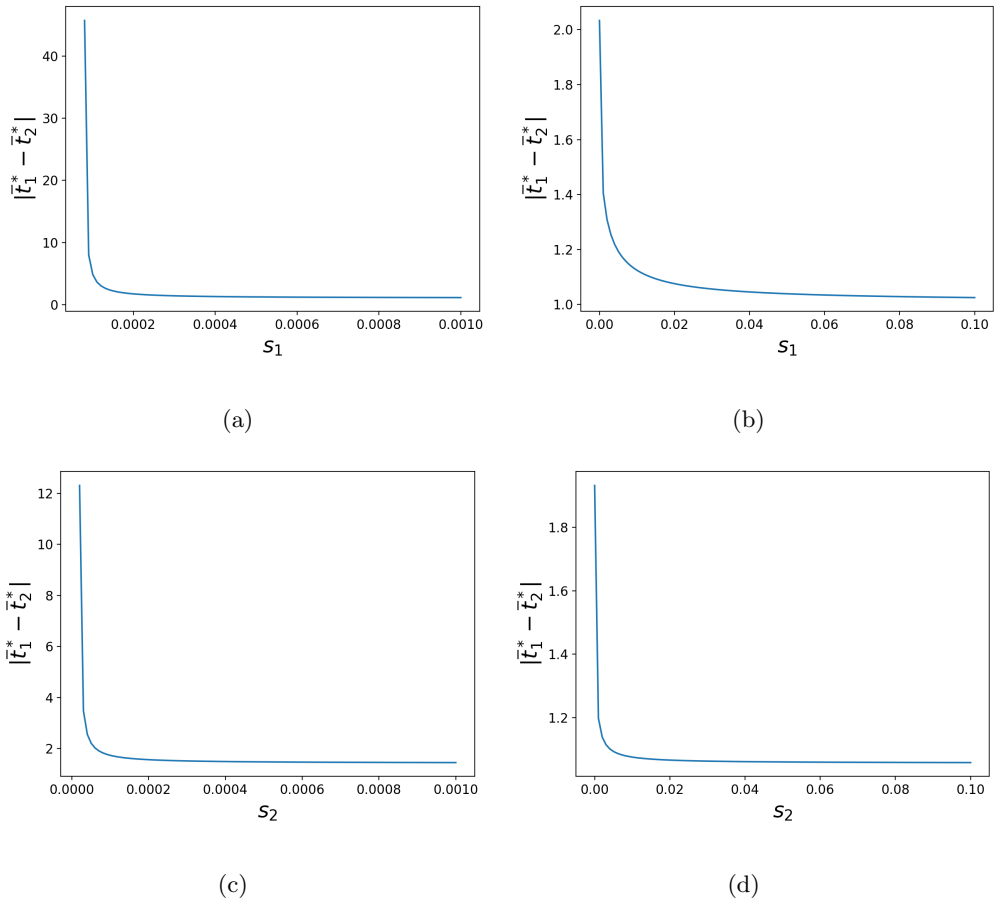
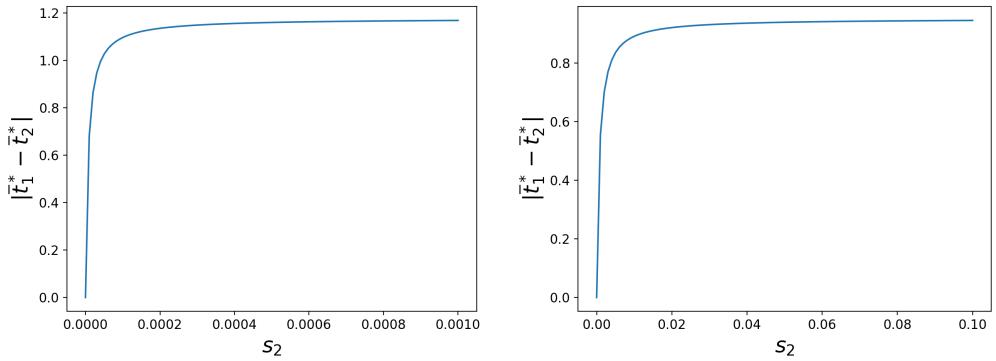


Figure 18: **Influence of the strength of selective constraints in (a)(b) species 1 s_1 and in (c)(d) species 2 s_2 on the phenotypic distances between the focal species and the local mimicry rings, assuming (a)(c) weak or (b)(d) strong females choosiness and predators discrimination.** Simulations were run assuming (a)(c) $a_1 = a_2 = b = 0.01$, (b)(d) $a_1 = a_2 = b = 1$. The default parameters values are as follows: $c_{RI} = 0.01$, $d = 0.02$, $\lambda_1 = 0.1$, $\lambda_2 = 0.1$, $N_1 = 10$, $N_2 = 20$, $s_1 = 0.0000025$, $s_2 = 0.0000025$, $t_{o1} = 0$, $t_{o2} = 1$.



(a) (b)

Figure 19: **Influence of the strength of selective constraints in species 2 s_2 on the phenotypic distances between the focal species and the local mimicry rings, when predation is great compare to RI, assuming (a) weak or (b) strong females choosiness and predators discrimination.** Simulations were run assuming (a) $a_1 = a_2 = b = 0.01$, (b) $a_1 = a_2 = b = 1$. The default parameters values are as follows: $c_{RI} = 0.01$, $d = 0.05$, $\lambda_1 = 0.1$, $\lambda_2 = 0.1$, $N_1 = 10$, $N_2 = 20$, $s_1 = 0.0000025$, $s_2 = 0.0000025$, $t_{o1} = 0$, $t_{o2} = 1$.

References

N. H. Barton and M. Turelli. Natural and sexual selection on many loci. Genetics, 127 (1):229–255, 1991. ISSN 0016-6731. URL https://www.genetics.org/content/127/1/229.

M. Kirkpatrick, T. Johnson, and N. Barton. General models of multilocus evolution. Genetics, 161(4):1727–1750, 2002. ISSN 0016-6731. URL https://www.genetics.org/content/161/4/1727.

R. Lande. Models of speciation by sexual selection on polygenic traits. Proceedings of the National Academy of Sciences, 78(6):3721–3725, 1981. ISSN 0027-8424. 10.1073/pnas.78.6.3721. URL https://www.pnas.org/content/78/6/3721.

A. Pomiankowski and Y. Iwasa. Evolution of multiple sexual preferences by fisher's runaway process of sexual selection. Proceedings: Biological Sciences, 253(1337):173–181, 1993. ISSN 09628452. URL http://www.jstor.org/stable/49806.