THE INTERACTION BETWEEN LEARNING AND SPECIATION

D. Justin Yeh

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of PhD in the Biology Department in the School of Arts and Sciences.

Chapel Hill
2018

Approved by:

Maria R. Servedio

Marcus W. Feldman

Joel G. Kingsolver

David W. Pfennig

Karin S. Pfennig

© 2018
D. Justin Yeh
ALL RIGHTS RESERVED

ABSTRACT

D. Justin Yeh: The Interaction between Learning and Speciation
(Under the direction of Maria R. Servedio)

Assortative mating displays, preferences, or both can be affected by learning across a wide range of animal taxa, but the specifics of how this learning affects speciation with gene flow are not well understood. I use population genetic models with trait learning to investigate how the identity of the tutor affects the divergence of a self-referent phenotype-matching trait. I find that oblique learning (learning from unrelated individual of the previous generation) and maternal learning mask sexual selection and therefore do not allow the maintenance of divergence. In contrast, by enhancing positive frequency-dependent sexual selection, paternal learning can maintain more divergence than genetic inheritance, but leads to the loss of polymorphism more easily. Furthermore, paternal learning inhibits the invasion of a novel self-referent phenotype-matching trait, especially in a large population.

Reinforcement is the process through which assortative mating evolves by natural selection to reduce costly hybridization. Sexual imprinting could facilitate reinforcement by decreasing hybridization, or it could impede the process if heterotypic pairs imprint on each other. Either result could then subsequently affect speciation. Here, I use deterministic population genetic simulations to explore conditions under which sexual imprinting can evolve through reinforcement. I demonstrate that sexual imprinting can evolve as a one-allele assortative mating mechanism by reducing the risk of hybridization. The evolution of imprinting has the unexpected side effect of homogenizing an existing innate preference, because the imprinted preference

overrides the effect of innate preference, effectively making it neutral. I also find that the weight of the imprinting component in the female preference may evolve to a lower value when migration and divergent selection are strong and the cost of hybridization is low, conditions which make it beneficial for maladaptive immigrant females to acquire locally adaptive genes by hybridizing with the local males. Together, these results suggest that sexual imprinting has the capacity to promote or retard divergence through complex interactions, and can itself evolve as part of the speciation process.

The effect of learned culture (*e.g.*, birdsong dialects and human languages) on genetic divergence is unclear. Previous theoretical research suggests that because oblique learning allows phenotype transmission from individuals with no offspring to an unrelated individual in the next generation, the effect of sexual selection on the learned trait is masked. However, I propose that migration and spatially constrained learning can form a statistical association between cultural and genetic traits, which may allow selection on the cultural traits to indirectly affect the genetic traits. Here, I build a population genetic model that allows such a statistical association to form, and found that sexual selection and divergent selection on the cultural trait can indeed help maintain genetic divergence through such a statistical association. Furthermore I found that the genetic divergence maintained by this effect persists even when the cultural trait changes over time due to drift and mutation. These results suggest the role of obliquely transmitted traits in evolution may be underrated, and the lack of one-to-one associations between cultural and genetic traits may not be sufficient to disprove the role of culture in divergence.

ACKNOWLEDGEMENTS

I thank my advisor Dr. Maria Servedio for guiding me through each challenge in my doctoral research. I also thank everyone who were in the Servedio lab: Mr. Joel Adamson, Dr. Sumit Dhole, Dr. Caitlin Stern, Dr. Courtney Fitzpatrick, Dr. Jeremy Van Cleve, Dr. Olivier Cotto, and Dr. Lyu Nan. My committee and Dr. Haven Wiley gave wonderful suggestions. The second chapter of this dissertation was done in collaboration with Dr. Janette Boughman and Dr. Glenn-Peter Saetre, with enormous amount of input from Dr. Øistein Holen. The Biology Department's teaching assistantship supported me from 2012-2017 during the semesters, and NSF funding to Dr. Maria Servedio supported me through the summers. The Dissertation Completion Fellowship supported me from 2017-2018. I also thank the current and the former Graduate Student Services Managers, Ms. Kenlyn Merritt and Ms. Julie Cannefax, for helping me navigate through the administrative works. Finally I thank my family and friends for moral support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ACRONYMS

A and B: patches (in chapter 1) or genetic alleles (in Chapter 3)

a and b: genetic alleles

α: assortative mating preference (in chapters 1 and 3), or the imprinting weight that maximizes trait divergence (in chapter 2)

C: assortative mating locus (in chapter 1), or cultural trait "locus" (in chapter 3)

c: cost of imprinting

γ: skew of divergent selection

d: the difference in fitness of $T_{1,\mathrm{A}}$ between paternal learning and genetic inheritance (in chapter 1) or cultural trait divergence (in chapter 3)

EPP: Extra-pair paternity

$F_{k,l,j}$: the frequency of matings in population *j* between a female with allele T_k and a male with allele T_l

κ: the imprinting weight threshold below which polymorphism is lost

L: learning locus

l: learning allele frequencies

LD: linkage disequilibrium

λ: imprinting weight

m: migration rate

μ: mutation rate

N: population size

P: female preference locus

p: an Boolean index that indicates whether $k = l$ (in Chapter 1), or preference allele frequencies (in Chapter 2)

p^{imp}: imprinting component of female preference

p^{inn}: innate component of female preference

s: selection coefficient for divergent selection

s_D: divergent selection on the male trait (in chapter 2) or on the cultural trait (in chapter 3)

s_H: selection against hybrid

σ: the imprinting weight that is a stable strategy

T: male trait locus

T_a, T_b, and T_c: male trait loci

$t_{a,x}$, $t_{b,x}$, and $t_{c,x}$: allele frequencies of T_a, T_b, and T_c in patch x

t_{ij}: allele frequency of T_i in patch j

GENERAL INTRODUCTION

Learning is the process of modifying one's behavior based on previous experience, often adaptively. It is common across the animal kingdom (Thorndike 1989; Papini 2002). As a form of plasticity, learning can promote to evolution in multiple ways: It can generate new variation, including divergent phenotypes, move the population along the adaptive landscape, including crossing a valley, and allow the accumulation of genetic variation (Pfennig *et al.* 2010). Conversely, learning can hinder evolution by reducing heritability.

A form of learning of particular interest is social learning, which is learning from other individuals (Hoppitt and Laland 2013). Because it transfers information between individuals, and variations that can have different fitness arise when learning is erroneous or when innovation occurs (Cavalli-Sforza and Feldman 1981; Boyd and Richerson 1985), social learning allows cultural evolution, which can interact with genetic evolution (Simoons 1969; Feldman and Laland 1996). Although cultural evolution interacts with genetic evolution, the study of how learning interacts with speciation has only started relatively recently (Verzijden *et al.* 2012). One of the reasons cultural evolution is not widely studied is the traditional view that only genetically heritable traits are important to evolution (Laland *et al.* 2015).

Speciation is the ultimate origin of species diversity, and is the mechanism that links microevolution to macroevolution. It occurs by evolving reproductive isolation, which can be grouped into three categories based on when in the life cycle they occur: premating isolation, postmating prezygotic isolation, and postzygotic isolation. One type of speciation that is of

particular interest to evolutionary biologists is speciation with gene-flow, mainly due to its perceived complexity and difficulty. It is difficult because gene-flow allows recombination to disrupt the segregation of genes (Felsenstein 1981). Furthermore, speciation requires the maintenance of polymorphism across incipient species, which can be difficult under disruptive selection without negative frequency-dependence (Rueffler *et al.* 2006).

Here I focus on the learning of traits related to premating isolation. Premating isolation is known to be important in speciation with gene-flow as it protects locally adapted populations from homogenizing (van Doorn *et al.* 2009). It can be affected by learning because mating signals and preferences can be learned. Such learning not only affects premating isolation directly, but can also affects the intensity of divergent sexual selection (Verzijden *et al.* 2012).

One way for premating isolation to evolve is through reinforcement, the further evolution of reproductive isolation when partial isolation already exists. This occurs in part because it is more costly than not for parents to produce (and rear) offspring that have reduced fitness (Dobzhansky 1940). Reinforcement may be important for speciation with gene-flow as it allows populations that have evolved partial reproductive isolation in allopatry to complete the speciation process upon secondary contact (Howard 1993). Previous theoretical studies have shown that it is possible for reinforcement to occur through a learned trait, although it is uncertain whether learning itself may evolve through reinforcement (Servedio *et al.* 2009; Olofsson *et al.* 2011).

In this dissertation I use population genetic models that incorporate, among other factors, social learning, mate choice, natural selection, and migration to examine how social learning and speciation interact. In particular, I ask how different forms of learning affects divergence and the maintenance of polymorphism, and whether reinforcement can occur by the evolution of

learning. Because these systems are often complex, with emergent dynamics that may not be intuitive, empirical knowledge may be difficult to synthesize with only verbal logic. Mathematical models provide a good tool to untangle these dynamics and reveal the role played by each component is mathematical models (Servedio *et al.* 2014).

In Chapter 1, I compare how differences in the identity of a tutor affect trait divergence and the maintenance of polymorphism when traits are learned. Specifically, I model assortative mating by a learned trait used in self-referent phenotype-matching, and see how the resulting sexual selection affects trait divergence. I find that maternal and oblique learning mask sexual selection while paternal learning enhances sexual selection, because sexual selection only acts on males. Due to the fact that sexual selection is positive frequency-dependent, paternal learning can maintain more divergence than genetic inheritance can, although it sometimes leads to the loss of polymorphism more easily and inhibits the invasion of novel phenotypes. These results highlights the importance of knowing who the tutor is when trying to predict the effect of learning on speciation.

In Chapter 2, I examine the evolution of sexual imprinting, a form of learning, in the context of speciation. Sexual imprinting could facilitate reinforcement by decreasing hybridization, or it could impede the process because it potentially increases hybridization in genetically purebred offspring of heterospecific social pairs. I demonstrate that sexual imprinting can evolve because it is a one-allele assortative mating mechanism (Felsenstein 1981) that reduces hybridization. Furthermore, with increased imprinting, polymorphism in the mating signal is more easily maintained. However, divergence of the mating signal is usually not maximized by the evolution of imprinting, and the innate preference actually become homogenized as imprinting strength increases because imprinting overrides the effect of innate

preference, effectively making the innate preference locus neutral. I also find that imprinting sometimes evolve to a lower strength because maladaptive immigrant females can benefit by hybridizing with the local males to acquire locally adaptive genes. These results shed light on how sexual imprinting may have evolved, and how it can contribute to speciation with gene-flow.

In Chapter 3, I demonstrate that, although Chapter 1 suggests oblique learning masks sexual selection, migration and spatially constrained learning can lead to statistical associations between cultural and genetic traits, and this association can then allow natural and sexual selection on the cultural traits to indirectly promote genetic divergence. Furthermore I demonstrate that even when drift and mutation allows cultural traits to change over time the genetic divergence maintained by this effect can persists as long as the cultural trait in the two populations remain different. These results suggest that to better understand how obliquely learn traits (such as birdsong) affect speciation, empirical studies need to focus more on how learned traits affect survival and mating success.

CHAPTER 1 : REPRODUCTIVE ISOLATION WITH A LEARNED TRAIT IN A STRUCTURED POPULATION[1]

Summary

Assortative mating displays, preferences, or both can be affected by learning across a wide range of animal taxa, but the specifics of how this learning affects speciation with gene flow are not well understood. I use population genetic models with trait learning to investigate how the identity of the tutor affects the divergence of a self-referent phenotype-matching trait. I find that oblique learning (learning from unrelated individual of the previous generation) and maternal learning mask sexual selection and therefore do not allow the maintenance of divergence. In contrast, by enhancing positive frequency-dependent sexual selection, paternal learning can maintain more divergence than genetic inheritance, but leads to the loss of polymorphism more easily. Furthermore, paternal learning inhibits the invasion of a novel self-referent phenotype-matching trait, especially in a large population.

Introduction

Assortative mating is an important component of pre-mating isolation, and an essential element in speciation with gene flow (Coyne and Orr 2004; Smadja and Butlin 2011). Many mechanisms of speciation through the evolution of assortative mating have been proposed (*e.g.*, Martin and Hosken 2003; McKinnon *et al.* 2004; Rundle *et al.* 2005; van Doorn *et al.* 2009), among them, mechanisms that involve the learning of displays (sometimes called the "traits")

[1]Previously Published as Yeh, DJ and Servedio MR. 2015. Reproductive isolation with a learned trait in a structured population. *Evolution* 69(7):1938-1947

and preferences that lead to assortment (Marler 1957; Marler and Tamura 1962; Nottebohm 1969; Gill and Murray 1972; Baker 1975; Baptista 1975; Baker and Mewaldt 1978; Searcy *et al.* 1981; Payne 1986).

Learning influences displays and/or preferences involved in assortative mating across a wide range of animal taxa (Owens 1999; Slater 2003; Snell-Rood and Papaj 2009; Kozak *et al.* 2011). Six modes by which learning can be involved in assortative mating can be categorized by differences in what is changed by learning (the preference or the trait), what learning is based on, and whom is learned from (the "tutor"), (Table 1.1). However, the ways in which these differences affect speciation and how they compare to genetic inheritance have received only limited theoretical study (reviewed in Verzijden *et al.* 2012).

The learning of assortative mating displays and preferences has often been studied in birds, using song dialects and sexual imprinting, respectively. Because some birds mate assortatively within local dialect groups (Baker *et al.* 1981, 1987; Balaban 1988; Searcy *et al.* 1997), it has been speculated that learned dialects may function as learned "assortative mating displays", and hence contribute to speciation. However, empirical results remain conflicting (Grant and Grant 1997; Baker and Boylan 1999; MacDougall-Shackleton and MacDougall-Shackleton 2001). Support for the idea that sexual imprinting promotes assortative mating (as learned "assortative mating preferences") is also mixed (Irwin and Price 1999).

Spatial structure is likely to affect the potential for learning to impact speciation because learning can only occur between individuals that are close enough to be perceived. Many of the mechanisms that maintain local song dialects in birds, for example, are related to spatial structure, including local adaptation, a tendency to learn from local versus foreign individuals, limitations on dispersal, and intra-specific competition which inhibits immigration of foreign

individuals singing unusual songs (Lynch 1996; Slabbekoorn and Smith 2002; Salinas-Melgoza and Wright 2012). In particular, local adaptation is important because some signals may facilitate communication more efficiently, or, conversely, attract predators and brood parasites, in certain environments (Boncoraglio and Saino 2007).

Despite decades of interest in this topic by empiricists, theoretical studies integrating assortative mating, habitat-dependent selection and learning are sparse (Slabbekoorn and Smith 2002). Most models of mate choice and learning have assumed either direct learning from parents in a single, uniform environment (Aoki and Feldman 1987; Tramm and Servedio 2008; Chaffee *et al.* 2013), or that every individual in the population can learn from any other individual regardless of their location (Cavalli-Sforza and Feldman 1983; Wakano *et al.* 2004). Few theoretical studies (Ellers and Slabbekoorn 2003; Planqué *et al.* 2014) have investigated how spatial limitations on learning affect the potential for speciation.

In this study, we focus on a trait that is the target of self-referent phenotype-matching (Hauber and Sherman 2001), for which the maintenance of divergence with limited spatial structure has been studied (Servedio 2011), and ask how this divergence is affected by learning of the trait. Self-referent phenotype matching is a good proxy for sexual imprinting (Verzijden *et al.* 2005; Servedio *et al.* 2009), which is common in birds, the group upon which we base our model (*i.e.*, the trait can be thought of as bird song). In order to see how learning and differences in the identity of the tutor affect speciation and species maintenance, we first build a basic model of secondary contact with spatial structure, migration and assortative mating, and compare genetic inheritance of the trait to trait acquisition by paternal learning, maternal learning and oblique learning (learning from unrelated individuals of the previous generation). Horizontal learning (from unrelated peers) has no inheritance between generations and is therefore omitted.

After examining the basic model, we include viability selection on the trait in the form of local adaptation, to explore its effect on the maintenance of existing divergence in the secondary contact scenario. The inclusion of viability selection also allows us to address a final scenario: the initial build-up of divergence through the spread of a novel mutation that has higher fitness in one of the two populations. This last model, akin to ecological speciation through novel adaptation, can facilitate our understanding of how learning may affect speciation when allopatry is new, with no initial divergence.

We find that the identity of the tutor, trait frequencies in the populations, and the strength of mating preference can dramatically change the role of learning in speciation. Oblique and maternal learning mask sexual selection altogether and therefore do not allow the maintenance of divergence. In contrast, paternal learning can help maintain divergence when mating preferences are of moderate strength and there is high initial divergence. However, when initial divergence is low, paternal learning becomes an obstacle to speciation.

The Basic Model

We first construct a population genetic model based upon Servedio (2011), which describes the evolution of a phenotype-matching trait in two populations with gene flow and sexual selection (a *Mathematica* notebook with all analyses is archived on Dryad). We consider a trait that has two alleles, T_1 and T_2, in two populations, A and B, with migration between them. Individuals are assumed to be haploid for simplicity and to isolate the effect of the mating system on divergence. Migration affects both populations at the same rate m, and both sexes have the same migration rate (spot checks for deviations from this assumption, *e.g.,* m=0.01 for females and m=0.02 for males, and vice versa, indicate that it is not critical). The frequency of T_1 in population A right after migration, for example, is $t_{1A}^{m} = (1-m)t_{1A} + mt_{1B}$, where t_{ij} is the

frequency of T_i at population j before migration.

Mating occurs under polygyny, such that every female has an equal probability of mating. A female prefers to mate with a male that matches her own trait over one that does not by a factor of 1+α. As α increases from 0 to infinity, mating thus shifts from completely random to completely assortative. (We do not include search costs in our model, but consider their effects in the Discussion.) Therefore, the frequency of matings in population j between a female with allele T_k and a male with allele T_l is

$$F_{k,l,j} = \frac{(1+p\alpha)t_{kj}^m t_{lj}^m}{1+\alpha t_{kj}^m} \quad (1),$$

where p is 1 when $k = l$, and 0 when $k \neq l$. After mating, the trait is passed down to the next generation through genetic inheritance or by learning in one of three forms, namely, paternal, maternal, or oblique learning. In paternal (maternal) learning, individuals obtain their phenotypes from their fathers (mothers). In oblique learning, individuals learn randomly from an individual of the previous generation, with the probability of learning a trait equal to the frequency of the trait in the previous generation after reproduction. For genetic inheritance, an individual has an equal probability of inheriting a trait from its father or mother in this haploid system.

We next obtain the recursion equations necessary for solving for the equilibria. For simplicity we show only the equations for t_{1A}. The equations for t_{2A}, t_{1B}, and t_{2B} are exactly analogous to the equation for t_{1A} because in this model, every mechanism affects both alleles in both populations in the same way.

Maternal and oblique learning, while operating under biologically different mechanisms, both simplify to the same 2-island migration-only model (which cannot maintain divergence):

$$t_{1A}^{\text{mat}} = F_{1,1,A} + F_{1,2,A} = t_{1A}^{m} \quad (2),$$

$$t_{1A}^{\text{obl}} = (F_{1,1,A} + F_{1,2,A} + F_{2,1,A} + F_{2,2,A})t_{1A}^{m} = t_{1A}^{m} \quad (3),$$

For the maternal learning model, this is because while males have different numbers of offspring due to sexual selection, every female produces the same number of offspring due to the assumption of strict polygyny. If individuals learn their trait from their mother, every offspring will correspond to its mother, who was not under sexual selection, from the previous generation. It thus makes sense that maternal learning, even with sexual selection on males, will not lead to a change in trait frequencies over generations. Similar reasoning holds for oblique learning, under which every offspring represents an individual from the previous generation, regardless of its reproductive success. In short, under maternal or oblique learning, when the trait is only under sexual selection in males, the effect of sexual selection is effectively not heritable.

For paternal learning, the recursion equation is

$$t_{1A}^{\text{pat}} = F_{1,1,A} + F_{2,1,A} = t_{1A}^{m}\left(\frac{(1+\alpha)t_{1A}^{m}}{1+\alpha t_{1A}^{m}} + \frac{t_{2A}^{m}}{1+\alpha t_{2A}^{m}}\right) \quad (4),$$

This equation shows that the allele frequency of T_{1A} in the next generation will be the frequency of T_1 males in population A after migration, multiplied by the probability that they will mate with either T_1 or T_2 females in population A. The mother's trait does not matter because with paternal learning, the father's trait passes on to all of his offspring. We can compare this equation with the recursion equation for genetic inheritance in Servedio (2011):

$$t_{1A}^{\text{gen}} = F_{1,1,A} + \frac{1}{2}F_{1,2,A} + \frac{1}{2}F_{2,1,A} = t_{1A}^{m}\left(\frac{(1+\alpha)t_{1A}^{m}}{1+\alpha t_{1A}^{m}} + \frac{t_{2A}^{m}}{2(1+\alpha t_{1A}^{m})} + \frac{t_{2A}^{m}}{2(1+\alpha t_{2A}^{m})}\right) \quad (5),$$

In Equations (4) and (5) the first (identical) term can be thought of as an assortative mating term, and the second one(s) can be thought of as disassortative mating terms. This interpretation is important to keep in mind to understand the critical difference between the models, explained below.

Three sets of equilibria can be obtained for paternal learning (Table 1.2), which we compare to the results of genetic inheritance from Servedio (2011). The first set of equilibria we can immediately obtain from all recursion equations is the loss of polymorphism. This equilibrium is stable for both paternal learning and genetic inheritance. Simulations show that it is reached when asymmetry in trait frequencies (*i.e.*, the difference between t_{1A} and t_{2B}) and/or *m* are high, while α is moderately high or low (see Supplementary Figure S1, where the solid lines are above zero). We will discuss this result in more detail in the next section. The second and third equilibria in the models are only solvable if we assume the allele frequency is symmetric between populations, *i.e.*, $\hat{t}_{1A} = \hat{t}_{2B}$. These equilibria were also cross-checked with simulations with this assumption relaxed. The second set of equilibria in the model is an unstable homogenized coexistence of $\hat{t}_{1A} = \hat{t}_{1B} = 0.5$. The third equilibrium, which is the one we are most interested in, is the maintenance of divergence. As T_1 and T_2 are arbitrary up to this point, we can define divergence as having more T_1 in population A and more T_2 in population B. This equilibrium is stable if *m* is smaller than a function of *α* shown in Table 1.2. It can be demonstrated using the Reduce function of *Mathematica* that the maximum migration rate that allows stable divergence is always higher in the paternal learning model than in the genetic inheritance model (Figure 1.1a).

At the third equilibrium, paternal learning maintains higher divergence (Figure 1.1b; proof in Supplementary Material S1). To analyze the reason behind this, we subtract Equation (5) from Equation (4) and divide it by t_{1A}, which gives the difference, *d*, in reproductive success of T_1 in population A between paternal learning and genetic inheritance:

$$d = \frac{t_{1A}^m t_{2A}^m}{2t_{1A}} \left(\frac{1}{1+\alpha t_{2A}^m} - \frac{1}{1+\alpha t_{1A}^m}\right) \quad (6),$$

Expression (6) highlights the difference in disassortative terms, and is always positive because $t_{1A}^m > t_{2A}^m$ by definition. This difference emerges because of the following: A pairing of a T_1 female and T_2 male can produce a T_1 offspring half the time under genetic inheritance, but never under paternal learning, while a pairing of a T_2 female and T_1 male again produces a T_1 offspring half the time under genetic inheritance, but always produces it under paternal learning. Importantly, the latter pairing is more common in population A, because it is more difficult for rare T_2 females to find equally rare T_2 males than it is for common T_1 females to find T_1 males (and vice versa in population B). In short, the divergence is maintained by positive frequency-dependent sexual selection, and paternal learning exaggerates it because a male favored by sexual selection can pass on his attractive trait to all of his offspring under paternal learning, instead of just half under genetic inheritance (see Tramm and Servedio (2008) for an explanation of paternal imprinting exaggerating sexual selection; here we extend this to trait learning, and by adding spatial structure discover how this effect affects divergence).

Another interesting aspect of the equilibria in the genetic and paternal learning models is that the level of divergence between populations peaks at an intermediate value of α. As described in Servedio (2011), these peaks exist because the evolutionary force causing

divergence in this model, positive frequency-dependent sexual selection, starts to decrease as α gets too strong (above a peak level α_{opt}); at this point the mating success of rare males rises as rare females become less likely to compromise and mate with a mismatched male. This concept can also be demonstrated mathematically by taking the limit of recursion equations with α approaching infinity. In both models this results in the disassortative mating term approaching zero (and the assortative mating term approaching one), thus simplifying to a two-island migration-only model. The location of the peaks can be found by solving $\frac{\partial}{\partial \alpha}\hat{t}_{1A} = 0$ for the third equilibria. The peaks occur respectively at

$$\alpha_{\text{opt}}^{\text{gen}} = \frac{1}{\sqrt{2m}} - 1, \text{where } 0 < m < \frac{1}{18} \quad (7),$$

$$\alpha_{\text{opt}}^{\text{pat}} = \sqrt{\frac{1-m}{m}} - 1, \text{where } 0 < m < \frac{1}{10} \quad (8),$$

Comparing Expressions (7) and (8), we find that $\alpha_{\text{opt}}^{\text{pat}} > \alpha_{\text{opt}}^{\text{gen}}$. This is because the disassortative mating term in the recursions increases more slowly with paternal learning; in Equation (6) the first term (from paternal learning) has α in the denominator multiplied with the smaller t_{2A}^{m} instead of the larger t_{1A}^{m}, which means as α increases, the loss of positive frequency-dependent sexual selection in the paternal learning model occurs later than in genetic inheritance. In other words, the pairing of rare females with common males (which, as described above after Equation (6), contributes more to paternal learning) reduces at a lower rate than that of common females with rare males as females become choosier. Expressions (7) and (8) both decrease as migration increases because more immigrants carrying the less frequent trait makes the disassortative mating terms in Equations (4) and (5) larger, which accelerates the loss of positive frequency-

dependence as α increases.

The abovementioned properties of α_{opt} are noteworthy because α_{opt} is an ESS in both models (see Servedio 2011 and Supplementary Material S2 for proof), provided the migration rate is low (less than around 4% for genetic inheritance, 7% for paternal learning). The reason is that any mutant that leads to less divergence, *i.e.*, decreases positive frequency-dependent sexual selection, than the allele for α_{opt}will form linkage disequilibrium with the locally less frequent trait, and therefore will be indirectly selected against by sexual selection. The fact that α_{opt} is an ESS also means that paternal learning will result in higher divergence than genetic inheritance not only when comparing under the same α, but also when α is allowed to evolve.

We conclude that in a secondary contact scenario, paternal learning is better at maintaining divergence, for two reasons. First, for any given α, the maximum migration that allows stable divergence is always higher for paternal learning than genetic inheritance (Figure 1.1a). Second, under the same migration rate, paternal learning yields more divergence at a stable equilibrium (Figure 1.1b; Supplementary Material S1), even if preference strength can evolve. This is because paternal learning exaggerates the effect of sexual selection. Positive frequency-dependent sexual selection is thus inflated under paternal learning, yielding greater trait divergence.

Viability Selection on the Trait

We next analyze the interaction of natural selection and sexual selection by including local adaptation on the trait, in the form of viability selection before migration (the results remain qualitatively the same when we change the order of events in the life cycle). Specifically, we assume that selection favors the T_1 phenotype in population A and the T_2 phenotype in population B by a factor of 1+*s*. The minimum value of *s* that maintains divergence can be

understood as the "difficulty" of achieving divergence despite migration, because under our assumptions, selection always favors divergence. Since, as explained above, maternal and oblique learning do not contribute to the maintenance of divergence, from here on we focus on comparing paternal learning with genetic inheritance.

When an assortative mating trait is under disruptive natural selection, it is termed a "magic trait" (Gavrilets 2004). This dual function prohibits recombination from disrupting the association between genes under divergent selection and those responsible for assortative mating; they are one and the same. Magic traits may be more common than previously thought (Servedio *et al.* 2011), and can evolve from non-magic traits (Thibert-Plante and Gavrilets 2013). Birdsong has the potential to constitute a magic trait: it can be an assortative mating trait (Searcy *et al.* 1981, 1997) and can be under natural selection when certain frequencies, volume or patterns may, in different habitats, be more efficient in communication or attract more predators and brood parasites (Boncoraglio and Saino 2007).

After adding viability selection to the model we obtain new recursion equations which are too complex to analytically solve for the equilibria. We use deterministic simulations to analyze the models (the R code is available on Dryad). The simulations start with T_1 fixed in population A and T_2 at a frequency of $1 - 2^{-8}$ in population B. We included this deviation from complete symmetry to avoid potential artifacts. The simulations, written in R, iterate the life cycle with different sets of m, α and s until the equilibrium is reached, and finds the minimum s that maintains divergence above several different thresholds.

Figure 1.2a summarizes the results from one set of parameters; results with other selected sets of parameters are available in Supplementary Figure S1. The area above the solid lines in Figure 1.2a is where coexistence of the two phenotypes is possible. Two humps exist for the

solid line, and in between the humps is where speciation is most likely to occur (the minimum *s* required for substantial divergence is lowest). This space is larger for paternal learning than genetic inheritance. The dotted lines show the amount of selection required to reach high divergence (defined here as $\hat{t}_{1A} > 0.75$, other levels of divergence are shown in Supplementary Figure S2). Between the humps these values mostly overlap with the continuous line. In most of the parameter space we tested, genetic inheritance requires stronger selection than paternal learning to achieve high divergence. Genetic inheritance does however maintain polymorphism more easily (with lower *s*) than paternal learning when α is moderately low or high (shaded areas).

To interpret these results we must realize that positive frequency-dependent sexual selection has two effects in our models. First, it contributes to the maintenance of existing divergence by favoring a different trait within each population, therefore lowering the level of viability selection required to maintain divergence (Figure 1.2b). Second, it causes a loss of polymorphism by favoring the trait with the higher overall frequency across both populations, which must be counteracted by divergent viability selection for polymorphism to be maintained (Figure 1.2c). The lines in Figure 1.2b/1.2c have their bottom/peak in the middle because positive frequency-dependent sexual selection is highest at α_{opt}. The black lines (paternal learning) in Figure 1.2b/1.2c are lower/higher than the grey lines (genetic inheritance) because paternal learning exaggerates positive frequency-dependent sexual selection. The black curves in Figures 1.2b and 1.2c are wider than grey curves for two reasons. First, positive frequency-dependence in the paternal learning model is exaggerated by paternal learning and therefore drops less quickly as α moves away from α_{opt} (the explanation is similar to that for why $\alpha_{\mathrm{opt}}^{\mathrm{pat}} > \alpha_{\mathrm{opt}}^{\mathrm{gen}}$, see text under Expressions (7) and (8)). Secondly, under paternal learning viability

selection is less effective than under genetic inheritance: viability selection on males is always filtered by sexual selection before being passed on to offspring while viability selection on females is not inherited.

With these explanations in mind, we can now look into the exact mechanism underlying Figure 1.2a. Around α_{opt} positive frequency-dependent sexual selection seems more efficient at favoring divergence instead of removing polymorphism. At extremely high and low α, sexual selection against the rare trait is weak; therefore a low level of selection is sufficient to maintain polymorphism. However to reach high divergence, strong viability selection is required to counteract migration. At moderately high and low α, sexual selection against the trait with lower overall frequency seems stronger than sexual selection favoring it within populations through its higher local frequency. This indicates that positive-frequency dependent sexual selection on local trait frequency is more easily lost than that on overall frequency as we move away from α_{opt}.

Invasion of a Novel Trait

Finally we consider the scenario where allopatry is newly established, *i.e.*, with no initial divergence, and examine if speciation can happen through the spread of a novel local adaptation. We conduct an evolutionary invasion analysis to find the level of divergent selection that allows T_2 to invade, in population B, a system initially dominated by T_1 in both populations (Supplementary Material S3).

Our analytical results show that, regardless of the value of α and m, it is always more difficult (requires stronger divergent selection) for a mutation to spread when there is paternal learning than when there is genetic inheritance (See Supplementary Figure S3). Furthermore, under paternal learning, as α increases, the difficulty of invasion increases roughly proportionally around a slope of 1, while with genetic inheritance, the minimum s instead approaches 1 as α

increases, at a much slower rate. This indicates the invasion of a new mutation is more difficult under paternal learning: in this case, the exaggerated positive frequency-dependent selection of paternal learning acts directly against rare phenotypes.

Invasion analysis assumes a very small deviation from $t_{2B} = 0$, but given that, under paternal inheritance, cultural mutation rates may be high, we are interested in non-trivial initial frequencies of t_{2B} as well. We therefore complement our analytical results with simulations (the R code is available on Dryad), which can examine the effects of these higher starting frequencies. Simulation results are consistent with our analytical predictions at low to moderate α. However when α gets higher, we find that the minimum selection required to establish the invading trait becomes weaker than the analytical prediction, and the difference between paternal learning and genetic inheritance decreases, sometimes drastically. This is because at high α, the rare T_2 is no longer selected against by positive frequency-dependent sexual selection and begins mating with other rare T_2, which is not considered in the invasion analysis. This explanation is confirmed when we find that simulation results become more similar to analytical predictions when we use a lower initial t_{2B} (Supplementary Figure S3).

These results suggest that positive frequency-dependent sexual selection against a novel trait would be strongest at the initial introduction of the new trait, and that invasion becomes easier if the invading trait were to increase in frequency, perhaps through drift. It also has implications for the evolution of a new song that emerges under paternal learning. In this case stronger positive frequency-dependent sexual selection may be easier to circumvent with smaller effective population sizes, and hence a higher frequency of introduction when a new song emerges by mutation or immigration.

Discussion

In this study, we ask how speciation and the maintenance of divergence are affected by the learning of a trait used in phenotype matching, and compare the effect of different tutors. Using population genetic models, we find that maternal and oblique learning mask sexual selection. In contrast, by exaggerating the effects of sexual selection, paternal learning facilitates the maintenance of divergence at secondary contact, but hinders the spread of a new mutation in a structured population. Our study also shows that, depending on the context, assortative mating by phenotype matching can have different effects on speciation. When divergence exists across populations, the positive frequency dependence caused by phenotype matching can maintain local alleles at high frequencies in different populations, and may even further their divergence. However, when starting without any divergence across populations, as in our final scenario, phenotype matching tends to maintain the original allele in both populations by hindering the invasion of a novel trait. The abovementioned mechanisms are unrelated to ploidy; therefore the assumption of haploidy in this model should not affect our conclusions.

In contrast to the case when there are separate preference and trait loci, assortative mating through self-referent phenotype matching does not require any linkage disequilibrium to establish divergence. The need for linkage disequilibrium is circumvented by joining the preference and the trait locus – having a certain trait automatically gives an individual a preference for that particular trait. This assumption may seem like an oversimplification, but there is empirical evidence for phenotype matching (for a review, see Hauber and Sherman 2001). Furthermore, sexual imprinting, for which there is substantial empirical evidence, has been shown to mimic the effects of phenotype matching (Verzijden *et al.* 2005, 2012). In fact, our models of a paternally learned phenotype-matching trait are mathematically equivalent to having male trait being learned or inherited from fathers and female preference also sexually

imprinted from fathers. This interpretation resembles preference and trait inheritance between brood-parasitic indigobirds (*Vidua chalybeata*) and their foster parents (Klein and Payne 1998; Payne *et al.* 1998, 2000; ten Cate 2000).

Among polygynous birds, oblique song learning is probably the most common learning mechanism (Lynch 1996). Our results show that in these cases, sexual selection alone has no effect on song evolution. Other mechanisms are required for sexual selection to drive the evolution of an obliquely learned trait. Possibilities include tendencies to learn from successful male rivalries (Searcy *et al.* 1981; Nelson 1992; Marler 1997; Vehrencamp 2001), or migration patterns that aggregate males with similar songs (Salinas-Melgoza and Wright 2012). There are some polygynous birds that learn songs paternally (Böhner 1983; Grant 1984; Zann 1985), and many other species for which the type of learning has not been established. For species that learn paternally, the exaggerated sexual selection resulting from paternal learning can promote divergence. Our results for paternal learning can also apply to other cases when there are constraints on song plasticity that cause birds to learn a song similar to their father's, or when offspring prefer to obliquely learn from a successful male (Clayton 1987). They may also apply to other forms of sexually-biased vertical transmission, *e.g.*, paternally-inherited epigenetic traits or cytoplasmically inherited traits in organisms with reversed sex roles.

Our assumption of no search costs to mating may be unrealistic when preferences are strong. Search costs can impede speciation in some cases (Schneider and Bürger 2006; Kopp and Hermisson 2008). However selection against strong preferences is not expected to affect our main conclusions, because even if a lower α evolves due to search costs, our results show that for a given preference strength, paternal learning will be able to maintain stronger divergence than genetic inheritance.

Geographically biased learning has been observed in numerous empirical studies (McGregor 1980; McGregor and Krebs 1982; Lachlan and Slater 2003). Theoretical studies by Planqué *et al.* (2014) show that the maintenance of dialects can be strongly affected by, among other things, spatial structure (modeled as dispersal). Our model allows learning to occur within each of two populations, and is therefore is more realistic than single-population models. A trait can, however, disperse to the other population after one generation through migration. Extensions assuming a more complex spatial structure that further restricts the geographic scope of learning would be interesting to pursue.

Tables and Figures

Table 1.1: Different ways learning affects assortative mating

Modes of learning	What is learned	What it is learned from	Tutor	Selected examples
Mate-choice copying	Preference	Preference	Individual of the same sex in the same or the previous generation	Birds (Galef and White 1998) Fruit flies (Mery *et al.* 2009) Fishes(Witte and Massmann 2003; Alonzo 2008) Humans (Little *et al.* 2008)
Changing preference based on social feedback	Preference	Preference	Potential and previous mates	Fruit flies (Siegel and Hall 1979; Dukas 2006)
Sexual imprinting / Learning of species recognition	Preference	Trait	Conspecifics or heterospecifics, consisting of: Parents (*vertical*; often of opposite sex), unrelated older individuals (*oblique*), or siblings and peers (*horizontal*)	Birds (Bateson 1966) Sheep and goats (Kendrick *et al.* 1998) Cichlids (Verzijden and ten Cate 2007) Sticklebacks (Kozak and Boughman 2009)
Discrimination based on prior exposure	Preference	Trait or the existence of trait variation	Potential mates	Guppies (Rosenqvist and Houde 1997) Crickets (Bailey and Zuk 2009) Fruit flies (Dukas 2006)
Learning from social feedback on own attractiveness	Trait	Preference	Potential mates	Zebra finches (Collins 1994; Royle and Pike 2010)
Trait learning	Trait	Trait	Vertical (paternal and/or maternal), oblique, or horizontal	Zebra finches (Böhner 1983; Eales 1985) Field sparrow (Nelson 1992) Humpback whales (Garland *et al.* 2011)

Table 1.2: Equilibria

Equilibria	Genetic inheritance (Servedio 2011)	Paternal Learning
Loss of polymorphism	Stable. Reached when starting with high asymmetry and/or migration rate.	Stable. Reached when starting with high asymmetry and/or migration rate.
Homogenized Coexistence	$\hat{t}_{1A} = \hat{t}_{1B} = 0.5$ Unstable	$\hat{t}_{1A} = \hat{t}_{1B} = 0.5$ Unstable
Divergence* $\hat{t}_{1A} = \hat{t}_{2B} > 0.5$	$\hat{t}_{1A} = \frac{1}{2}\left(1 + \frac{1}{1-2m}\sqrt{\frac{2m\alpha^2 + 10m\alpha + 8m - \alpha}{\alpha(2m\alpha + 2m - 1)}}\right)$	$\hat{t}_{1A} = \frac{1}{2}\left(1 + \frac{1}{1-2m}\sqrt{\frac{m\alpha^2 + 6m\alpha + 4m - \alpha}{\alpha(m\alpha + 2m - 1)}}\right)$
Conditions for stability**	$m < \frac{1}{4+2\alpha}\left(1 - \sqrt{\frac{2\alpha+3}{(\alpha+1)(\alpha+3)}}\right)$	$m < \frac{3+5\alpha+\alpha^2}{12+20\alpha+8\alpha^2+\alpha^3} - \frac{\sqrt{(\alpha+1)(\alpha+3)(2\alpha+3)}}{12+20\alpha+8\alpha^2+\alpha^3}$

* As stated it the main text, the divergence equilibrium for paternal learning is always higher than genetic inheritance, and the maximum *m* that allows stable divergence under paternal learning is always larger than genetic inheritance.

** We relaxed the symmetry assumption when finding the stability condition. Detailed analyses are available in the *Mathematica* notebook archived on Dryad.

Figure 1.1: The maintenance of divergence through paternal learning and genetic inheritance. (a) The highest migration rate that supports divergence. (b) The level of divergence maintained at m=0.02. Dotted lines indicate unstable equilibria. Black: paternal learning; Grey: genetic inheritance.

(a) (b)

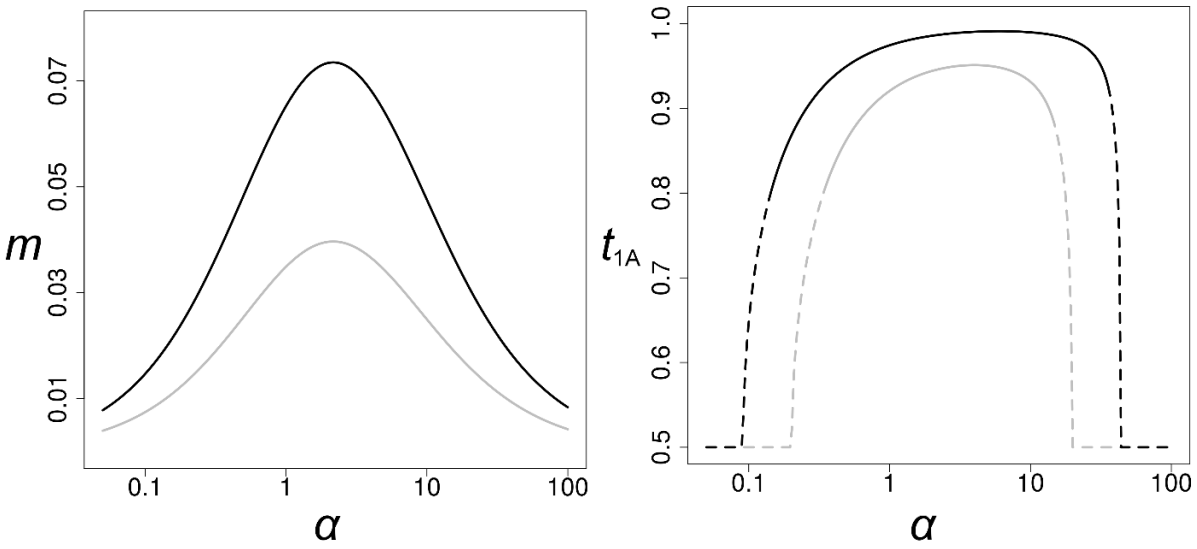
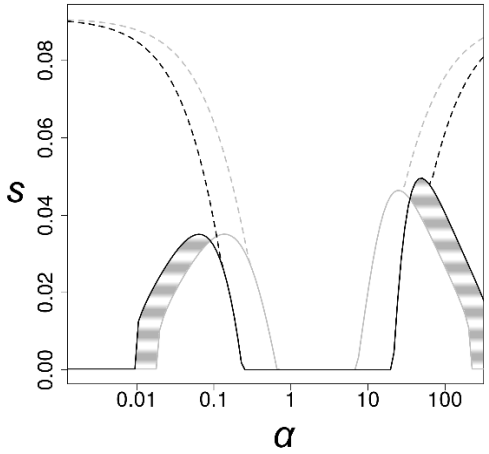
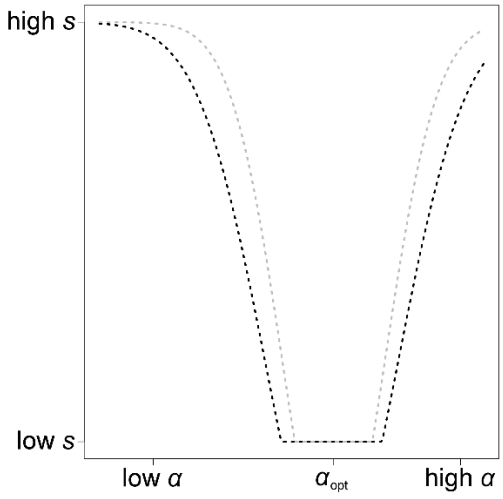


Figure 1.2: (a) Minimum selection required to maintain divergence at different levels of assortative mating (m=0.03, asymmetry $t_{1A} - t_{2B} = 2^{-8}$, divergence threshold $t_{1A} \geq 0.75$). The dotted lines are not visible when overlapping with the solid lines in the middle. (b)(c) Schematics illustrating that positive frequency-dependent sexual selection, which is strongest at α_{opt} and exaggerated when under paternal learning, has two effects: (b) It favors divergence because within each population a different trait is at higher frequency; this allows divergence to be maintained at lower levels of divergent selection. (c) It hampers the maintenance of polymorphism because at the global scale there is more T_1; this can be counteracted by divergent selection (polymorphism is maintained above the lines). Black: paternal learning; Grey: genetic inheritance; Solid lines: Minimum selection to maintain polymorphism; Dotted lines: Minimum selection to reach above divergence threshold; Shaded area: where genetic inheritance maintains polymorphism better than paternal learning.

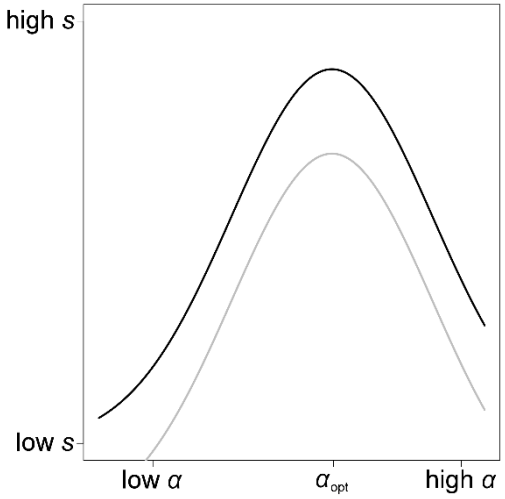
(a)



(b)



(c)



CHAPTER 2 : THE EVOLUTION OF SEXUAL IMPRINTING THROUGH REINFORCEMENT[2]

Summary

Reinforcement is the process whereby assortative mating evolves as a result of selection against costly hybridization. , Sexual imprinting could evolve as a mechanism of reinforcement, decreasing hybridization, or it could potentially increase hybridization in genetically purebred offspring of heterospecific social pairs. Here, we use deterministic population genetic simulations to explore conditions under which sexual imprinting can evolve through reinforcement. We demonstrate that a sexual imprinting component of female preference can evolve as a one-allele assortative mating mechanism by reducing the risk of hybridization. It will, however, often evolve to be a partial component rather than the sole determinant of female preference. The evolution of imprinting has the unexpected side effect of homogenizing existing innate preferences, because the imprinted preference neutralizes any innate preference by overriding it. We also find that the weight of the imprinting component of the female preference may evolve to a lower value when migration and divergent selection are strong and the cost of hybridization is low; these conditions render hybridization adaptive for immigrant females because they can acquire locally adaptive genes by mating with local males. Together, these results suggest that sexual imprinting can itself evolve as part of the speciation process, but in doing so has the capacity to promote or retard divergence through complex interactions.

[2] Coauthored with Jennette W. Boughman and Glenn-Peter Saetre.

Introduction

What is the evolutionary fate of incipient species that meet in secondary contact? Whether incipient species maintain trait divergence and genetic differentiation or instead become homogenized and fuse depends partly on whether they evolve stronger assortative mating. Enhanced assortative mating can be selected for when hybridization is costly, through the process of reinforcement (Dobzhansky 1940; Servedio and Noor 2003). Such increased assortment reduces hybridization, enabling further differentiation and independent evolution of the species. Assortative mating also strengthens genetic associations between mating preference and trait loci that form the basis of premating isolation.

Yet, the evolution of stronger assortative mating can be hampered by even low levels of gene flow (Kelly and Noor 1996; Servedio and Kirkpatrick 1997; Kirkpatrick 2000). This is in part because recombination is likely to break up associations between loci that have diverged in the two species that determine mate choice preferences and traits (Felsenstein 1981). Recombination will have this effect whenever assortative mating relies on unique alleles that determine the direction of mating preference becoming established in each population across each locus; this is an example of a "two-allele" mechanism in the terminology of Felsenstein (1981). Moreover, because such preference divergence during reinforcement is a relatively weak correlated response to selection at other loci (*e.g.*, those involved in hybrid identity), it is difficult for this weak force to counter the effects of gene flow. This results in preference alleles tending to homogenize between populations (*e.g.*, Servedio 2000. see van Doorn *et al.* 2009; Weissing *et al.* 2011 for a similar effect during sympatric speciation). Furthermore, such homogenized preferences will in turn tend to homogenize even locally adapted sexually selected traits (Servedio and Bürger 2014). Finally, preference variation across populations can be lost

altogether when gene flow between such populations is present, leading to one uniform preference being fixed in all areas (*e.g.*, Liou and Price 1994). Opportunities to observe these evolutionary changes in preference are rare, but several fishes show a loss of preference differentiation following human induced secondary contact (Rosenfield and Kodric-Brown 2003; Egger *et al.* 2010, 2012; Lackey and Boughman 2013).

At first glance, genetically based preferences may seem more likely than learned ones to contribute positively to divergence and reproductive isolation, because genetically based traits in general may be resistant to rapid change. However, learned preferences can also facilitate speciation and might not be as subject to the weaknesses of the two-allele mechanism described above (Kopp *et al.* 2018). Specifically, some of the challenges that gene flow poses can be circumvented when mating preferences result from sexual imprinting, a learning mechanism that occurs when early exposure to a parent shapes a mate preference that is expressed at maturation. Because the choosing individual is likely to share its trait phenotype with its parent, the evolution of sexually imprinted preferences would lead to assortative mating (via the evolution of choosiness). In such a case, a single allele (or set of alleles) for imprinting instead of random mating could spread across two incipient species resulting in assortative mating, causing sexual imprinting under what is known as a "one-allele" mechanism of Felsenstein (1981). Because the preference is controlled by the presence of what could hypothetically be a single allele across both populations (*e.g.*, an allele that causes imprinting), the issue of the homogenization of preferences (and hence, traits) is moot. Imprinting thus has significant potential to increase the amount of reproductive isolation and reduce hybridization when the phenotypes that are imprinted upon have diverged between the incipient species.

Even early work on sexual imprinting recognized its potential to prevent hybridization (*e.g.*, Bateson 1978) especially in geographic regions where two species meet, and suggested it might be important in speciation (Irwin and Price 1999). In birds, mate choice determined by sexual imprinting on parental phenotypes is common and widespread and has been reported in more than half of the bird orders (*e.g.*, ten Cate and Vos 1999). Immelmann (1972) argued that the most important function of imprinting is to ensure conspecific mating. Thus, imprinting has long been thought to play a key role in speciation. First recognized and most widely studied in birds (Bateson 1966), sexual imprinting has proven to be more widespread and some recent tests have found it in various fishes where it appears to strengthen reproductive isolation between diverging species (Verzijden and ten Cate 2007; Kozak *et al.* 2011), and in mammals (Kendrick *et al.* 1998; Penn and Potts 1998).

Although, as discussed above, imprinting may play a role in increasing reproductive isolation across populations because it is a one-allele mechanism of reproductive isolation that can lead to, or increase, choosiness, it may play a different role when the chance for hybridization is high. This may be reflected in the example of two exceptions to the rule of obligate imprinting in birds, both of which involve bird pairs that live in sympatry. Interspecific cross-fostering experiments on pied and collared flycatchers in a sympatric population in Öland, Sweden, found no effect of social environment but a strong, sex-linked genetic effect (Saether *et al.* 2007). Similar results were found in cross-fostering experiments in the Gouldian finch (Pryke 2010). One hypothesis, therefore is that the role of imprinting in mate choice is evolutionarily reduced in sympatry due to selection to reduce hybridization risk. One cost of imprinting would be mis-imprinting in cases of heterospecific pairing.

Given the potential dual effects of imprinting described above, we asked whether it is likely to evolve as a consequence of reinforcement. If so, does imprinting increase trait divergence, reduce hybridization, and maintain polymorphism in the traits and preferences that foster speciation? Alternatively, can imprinting be lost, as is sometimes found in natural populations? We specifically consider a situation in which both genetically based preferences and imprinting are present, and are allowed to interact evolutionarily. We find that genetic preferences often homogenize, while imprinting evolves over a wide range of parameters, fostering some (but not maximal) divergence and reduction of hybridization. Competing evolutionary forces on imprinting, including cases of adaptive hybridization, often, however, cause imprinting to evolve to contribute only partly to overall mating preferences.

Model

In this deterministic population genetics model, we consider 2 patches (labeled 0 and 1) both of infinite population size, and 5 diallelic haploid loci, in which the alleles are labeled 0 and 1. Three loci, T_a, T_b, and T_c, encode trait T additively, giving 4 discrete phenotypes, 0, 1, 2, and 3; alleles 0 are adapted to patch 0 and alleles 1 are adapted to patch 1. Locus P determines the innate preference and locus L determines the weight of the imprinting component ("imprinting weight"), λ. Recombination rates between all loci are assumed to be 0.5 (free recombination). Because the mating preference has a component that is paternally imprinted, the father's trait (T_{father}) is also tracked in the offspring as a phenotype with 4 states. We start by denoting the frequency of phenogenotype x in patch y as $f_{x,y}$, and track how it changes through the life cycle. The life cycle stages include migration, viability selection, sexual selection, reproduction, and paternal imprinting. Table 2.1 summarizes the life cycle, the function of each locus, and the evolutionary forces acting on them.

Migration between the two patches occurs at a rate of m. In other words, after migration the frequencies of phenogenotype x in patch y is $f_{x,y}^{\mathrm{mig}} = (1-m)f_{x,y} + mf_{x,1-y}$. Migration is thus generally assumed to be symmetric between patches, but all simulations were also run with slight, random asymmetries in m (up to 1% deviation) between patches to ensure that this did not qualitatively change the results.

The populations then proceed to viability selection, which acts on the T and L loci. Selection on the trait includes selection against hybrids (s_H) and divergent selection (s_D, where $0 \leq s_H, s_D, \leq \infty$) (Table 2.2). Additionally, the imprinting locus L is under an imprinting cost (c). The genotypic frequencies thus become

$$f_{x,y}^{\mathrm{vs}} = \frac{f_{x,y}^{\mathrm{mig}} v_{x,y}(1-d_x c)}{\overline{v_y}} \quad \text{..... (Eq. 1)}$$

after viability selection, where $d_x = 0$ if phenogenotype x corresponds to L_0 (weaker imprinting) or else $d_x = 1$ (L_1, stronger imprinting), $v_{x,y}$ is the relative viability of the trait in population y as shown in Table 2.2 and $\overline{v_y}$ is the mean viability in that population calculated as $\overline{v_y} = \sum_x v_{x,y}(1-d_x c) f_{x,y}^{\mathrm{mig}}$. The trait is expressed in both males and females; therefore the phenogenotypic frequencies after viability selection are identical between sexes. As with migration, the selection parameters s_H, s_D, and c were also given slight, random asymmetries (again up to 1% deviation) in replicas of all runs to assure that the qualitative results were robust to this change.

Mating occurs after viability selection. We assume random encounters and strict polygyny, where all females have equal mating success, resulting in direct sexual selection on males but not females. Thus the frequency of pairing between male with phenogenotype g and female with phenogenotype h is

$$F_{g,h,y}^{ss} = \frac{p_{g,h} f_{g,y}^{vs} f_{h,y}^{vs}}{z_{h,y}} \quad \text{(Eq. 2)}$$

where $z_{h,y} = \sum_j p_{j,h} f_{j,y}^{vs}$ and $p_{g,h}$ is the relative preference of female with genotype *h* for males with genotype *g*. For example, if $p_{1,1}/p_{1,2} = 2$ it means a female is twice as likely to pick a male with genotype 1 as one with genotype 2 in a two-choice test. Because there are only 4 distinct phenotypes for the male trait, although there are a total of 8 possible genotypes at the T loci, only 4 preference values needs to be defined for a given female. The female preference $p_{g,h}$ is a combination of innate preference ($p_{g,h}^{inn}$) determined by the P locus and paternal imprinting ($p_{g,h}^{imp}$), weighted by the imprinting weight (λ, encoded by the L locus), *i.e.*, $p_{g,h} = (1 - \lambda)p_{g,h}^{inn} + \lambda p_{g,h}^{imp}$. The innate and imprinting preference components are all obtained by using the mass functions of binomial distribution, and three innate preference functions are tested: which we term skewed, plateau, and peak (see Figure 2.1 for these functions and the parameters used to obtain them). The "skewed" innate preference is just as choosy (and divergent) as the most extreme case of imprinting (imprinting on a father with trait 0 or 3), thus if imprinting can evolve in this case, it would be because imprinting as a one-allele mechanism is more accurate than genetically inherited preferences at delineating conspecifics, and mating with conspecifics is beneficial. On the other hand, the preference functions "plateau" and "peak" are less choosy (and less divergent) than would be a female that imprinted on trait 0 or 3, thus learning could potentially also evolve in these two cases because it increases choosiness.

Because we assume free recombination, for each pairing offspring frequency is equally distributed across all possible phenogenotypes, with each phenogenotype constituting $1/2^i$ of that pairing frequency, where *i* is the number of genetic loci at which the two parents do not share the same allele. The frequency of phenogenotype *x* in the offspring is thus

$$f_{x,y}^{\text{rep}} = \sum_g \sum_h \frac{j_{g,h,x} F_{g,h,y}^{ss}}{2^i} \quad \text{(Eq. 3)}$$

where $j_{g,h,x} = 1$ if a pairing of g and h can produce offspring with x (all genetic alleles in x are present in g or h, and $\mathrm{T}_{\mathrm{father}}$ in x matches the trait value of g), or else $j_{g,h,x} = 0$.

The evolution of the imprinting weight λ, which is our main interest, occurs when the frequency of the L allele that encodes higher (or lower) λ increases. In the next section we explain how we use pairwise comparison of L alleles encoding different λ values to predict what value λ is expected to evolve. Change in the divergence and polymorphism at P and T loci, as well as hybridization, is also analyzed throughout.

Three variations of the models were also explored. The first one adds a skewness of divergent selection to allow viability to decrease non-linearly when moving away from the locally adapted trait value. The second one assesses the evolution of maternal imprinting instead of paternal imprinting. In the third variant the trait is only expressed in males (*i.e.*, there is sexual dimorphism), so that viability selection does not occur on females. The details of these variants are presented in full in Supplementary Material S1.

Simulation

To investigate the evolution of the imprinting weight, λ, using the abovementioned model, we use deterministic simulations written in C. Our first interest is to find out, given a secondary contact event, what imprinting value λ will evolve. Constraints in computation time prevent us from directly answering this question using a simulation with successive invasions and substitutions of different L alleles to explore the entire parameter space (although we did it for small sets of parameters). Instead, we show the range of conditions (including selection coefficients, migration rate, female preference distribution, and imprinting weight) under which a

small change in imprinting weight λ will be favored at secondary contact. Once we obtain this information, along with a few extra analyses and simulations (such as successive invasion for a few sets of parameters; see Supplementary Material S2 for details), we can infer what value of λ will ultimately evolve through successive invasion in most of the parameter space. This value is the stable strategy for imprinting to which the population is expected to evolve (and not evolve away from), which we call σ (from Greek σταθερός and στρατηγική, "stable" and "strategy"). Part of inferring the location of σ involves identifying the threshold that allows the maintenance of polymorphism in both the trait locus T and the preference locus P, which we call κ (from Greek κατώφλι, "threshold"). Because our second interest is in identifying the effects of the evolution of imprinting on speciation and divergence, we also identify a third point, which we call α (from Ancient Greek ἄκρον, "peak"), where the maximum level of trait divergence can be found.

To find the conditions during secondary contact that allow a small evolutionary change in imprinting weight, we let the invading L allele (introduced at a frequency of 0.01 in both patches) encode an imprinting weight λ that is higher or lower than the λ encoded by the resident allele, by an interval of 0.01. We examined all pairs of imprinting weights between 0 and 1 (the entire range of λ) with a difference of 0.01, in both directions of invasion (*i.e.*, starting with the higher value of λ fixed and starting with the lower value of λ fixed). The starting conditions for the T and P loci that represent secondary contact of divergent populations are as follows: The T_0 alleles at each of the three T loci are set to frequencies of 0.991, 0.992 and 0.993 in patch 0, and 0.01 in patch 1. The P_0 allele is set to a frequency of 0.991 in patch 0 and 0.01 in patch 1. The small asymmetries in allele frequencies across patches are used to avoid potential artifacts in the simulations that may arise if complete symmetry were assumed. Different magnitudes of these

asymmetries, *i.e.*, the frequency difference between T_0 or P_0 alleles in patch 0 and T_1 or P_1 alleles in patch 1 were also used to ensure that the numbers we pick are not unusual in any way. We iterate through the life cycle until equilibrium (defined as a change of less than 10^{-12} in the frequency of all phenogenotypes across a generation) is reached. We assess whether the L allele coding for the resident λ or for the mutant λ has become fixed at this equilibrium. Additionally we tried runs with larger mutational effects (*e.g.*, differences of 0.1, and λ_0=0.01 λ_1=0.99) to see whether this resulted in different dynamics. For the cost of learning, it is not necessary to model the cost as a function of λ (*e.g.*, cost = $\lambda c'$), as we assume there are only two L alleles competing at any time (mutation is rare). The relative viability cost for the higher imprinting strategy (λ_1-λ_0)c'/(1-$\lambda_0 c'$) can thus be approximated as (λ_1-λ_0)c' when c'<<1. Therefore, for all invasions, we assumed that the allele for the higher imprinting strength had a fixed viability cost while the other allele did not. When comparing between different sizes of mutational effect, we compare the results with the same c/(λ_1-λ_0) rather than the same *c*.

Finally, to confirm that our starting conditions for secondary contact are reasonable, we also ran the simulation with m=0 to see how the populations will evolve in allopatry. This analysis is presented at the end of the Results section, since a full understanding of this case follows from the findings of the main model.

We find that the effect size of mutation (λ_1-λ_0), small asymmetries in the parameter values, and the direction of asymmetry in initial allele frequencies have negligible effect on the results. For each invasion, the equilibrium generally takes a very long time to reach (millions of generations), but all results of interest (whether invasion is happening at the L locus, and the quasi-equilibrium frequencies reached at the P and T loci) can be observed within 50,000 generations, therefore after initial runs we set the simulations to stop after 80,000 generations.

Results

We find that imprinting weight can evolve because of the selective pressures in this system, but the direction and endpoint of this evolution can vary based on the parameters of the model. Furthermore, this evolution can stop at an equilibrium that constitutes "partial" imprinting, with an intermediate imprinting weight that allows both imprinted and innate contributions to preference. We additionally find that the evolution of imprinting can in many cases not maximize either preference or trait divergence.

Figure 2.2 presents a schematic (Figure 2.2a) and example (Figure 2.2b) of the relationship of imprinting weight (σ, in a case when it is partial) to two other points of interest. One of these is a threshold (κ), which represents a conservative value below which polymorphism in either T (or in rare cases, P), cannot be maintained, rendering the question of speciation moot. The existence of $\kappa > 0$ implies that variation in the trait cannot be maintained by innate preferences alone in in some ranges of parameter space; sufficient reliance on imprinting must also be present. The stable strategy for imprinting weight (σ) can be found when it is located above the polymorphism threshold (κ). We therefore begin our presentation of the results of the model with a detailed explanation of how the parameters of the model affect the position of κ. The other, α, presents the peak in trait divergence between the populations. It is interesting to note that is not usually coincident with the stable strategy σ (Figure 2.2b). The position of α is discussed further in the section "Trait Divergence and the Effect of Hybrid and Immigrant Populations".

Maintenance of Polymorphism

We define loss of polymorphism as any allele of T or P being lost across both patches, and stop the simulation whenever that happens. In some extreme parameter space which we are

not interested in (high m, high s_D, s_H=0) the population may evolve towards an intermediate trait by losing one or two T_0 alleles and fixing the other(s), as this can be a response to strong disruptive selection s_D. However, most cases of polymorphism loss occur on all three T loci in the same direction, but not on the P locus. We find such loss of polymorphism at the trait loci when the level of imprinting in the population starts below κ (Figures 2.2 & 2.3). The value of κ positively correlates with migration (m), the level of initial asymmetry in allele frequencies of P and/or T (*e.g.*, the frequency difference between Ta_0 in patch 0 and Ta_1 in patch 1), and usually negatively correlates with viability selection (s_D, and s_H). It is also affected by the extremity of the innate preference (*i.e.*, polymorphism seems most likely to be lost under skewed innate preference; Figure S3-1 in Supplementary Material S3), but this last pattern is not entirely clear. We cannot detect any effect on κ of the imprinting cost (c, not shown). Population divergence and the evolution of imprinting are facilitated when κ=0. If κ>0, innate preference alone cannot maintain polymorphism, then it is difficult for imprinting to evolve because a population with low or no imprinting will quickly lose polymorphism at the trait and preference loci. During the transient period before the loss of polymorphism, using our starting values representing secondary contact, we found that the change in the allele frequency of L is generally negligible (<0.1%). After the polymorphism is lost, the allele for stronger imprinting either becomes lost (if c>0) or stops changing (if c=0).

The loss of polymorphism in T occurs at least in part because the alleles present at the preference locus, which is only under indirect selection, tend towards homogenization across populations due to migration; such homogenization makes divergence at the P locus less than that at the T loci, thus causing sexual selection to in-turn reduce trait divergence (Servedio and Bürger 2014). As the system evolves towards homogenization at the P and T loci, the entire

system can increasingly be seen as one population with one of the P alleles having a higher frequency, due to initial asymmetry making one P allele more common globally than the other. The more common P allele causes sexual selection for the corresponding T alleles across the entire system, and the eventual loss of polymorphism in the T loci will occur as both the preference and trait alleles continue to increase in frequency through the Fisher Process (Fisher 1915). The P locus evolves slower and does not lose polymorphism in most of the parameter space because it is only under indirect selection, and because the preference locus only contributes (1- λ) of the female preference phenotype.

When $\lambda > \kappa$, however, the homogenization at the P locus becomes less relevant for evolution at the T loci because the imprinted preference overwrites the innate preference, effectively neutralizing the effects of P and thus inhibiting it from further homogenizing the T loci. Importantly, imprinted preferences are re-established every generation based on paternal trait frequencies, and thus cannot homogenize like a genetic allele. Therefore, even though migration occurs after imprinting (moving some imprinted females to the "wrong" population) the imprinted preference does not homogenize the trait. The neutralization of P as a result of imprinting can be seen in Figure 2.2b (dashed line), where P loses divergence as λ increases (although initially near κ divergence at P may increase due to indirect selection through T). Because of this key difference between imprinting and a preference locus, even in the scenario of "skewed" innate preferences, where the imprinted preference is identical to or even less extreme than the innate preference, increasing imprinting weight still increases the overall divergence of female preference and thus facilitates the maintenance of polymorphism in the trait.

Another force that affects the maintenance of polymorphism is selection against hybrids (s_H). It leads to positive frequency-dependent viability selection at the T loci within and across

populations, because individuals with rare traits have a higher chance of encountering individuals of the opposite trait and producing less fit offspring. This helps maintain polymorphism in this secondary contact model because different traits are more common in each patch. Because of this effect, in Figure 2.3, κ generally decreases with stronger selection against hybrids (s_H). Some exceptions exist (*e.g.*, with low to moderate s_D in the several panels of Figure 2.3), however; one reason could be because globally the asymmetry in our initial frequencies makes one set of alleles more common than the other, which can lead to the loss of polymorphism under positive frequency dependent selection especially when there is no high s_D to maintain divergence.

The location of κ is dependent on the initial asymmetry of the allele frequencies at the T loci across the two patches, although easily countered by divergent selection (*e.g.*, polymorphism is always maintained when s_D>0.1 and s_H>0 even if we increase the initial asymmetry by two orders of magnitude). If the frequencies are more symmetric, polymorphism at the T loci is more easily maintained. This is important especially when inferring the imprinting weight that will evolve (our second question), because when polymorphism is maintained, the alleles at the T and P loci evolve towards increased symmetry in their frequencies given roughly symmetrical selection and preference strengths. Therefore, in the sets of simulations where new L alleles are allowed to invade sequentially after P and T reach equilibrium at the current λ, it is possible to evolve to a lower imprinting weight without losing polymorphism, because the more symmetric allele frequencies at the T and P loci have shifted κ to a lower value. In the most extreme cases where we have complete symmetry (T_0 and P_0 in patch 0 have the exact same frequencies as T_1 and P_1 in patch 1), κ=0 and the polymorphism will never be lost, whereas when we have

complete asymmetry (T_0 and P_0 in patch 0 are 1 and T_1 and P_1 in patch 1 are 0) there is by definition no polymorphism.

Evolution of Imprinting

When $\lambda > \kappa$, polymorphism is maintained and the system evolves towards an imprinting weight of σ. Populations starting at σ cannot be invaded by either higher or lower λ, and populations starting above or below this point can always be invaded by individuals closer to this point. Thus this is the most likely outcome of evolution upon secondary contact.

When there is no cost of imprinting, complete imprinting usually evolves, *i.e.*, σ=1 (Figure 2.4cf, there is an exception explained in the next paragraph). The evolution of imprinting reduces hybridization (Figure 2.2b, dotted line). This is consistent with the expectations for the process of reinforcement occurring by a one-allele mechanism, which imprinting would fall under. In these models, the allele that encodes a stronger assortative mating preference, in this case resulting from higher λ, increases its frequency by forming a statistical association (linkage disequilibrium) with the purebreds (trait value 0 or 3), which have higher fitness.

There are exceptions to the evolution of complete imprinting even when there is no cost of imprinting (c = 0; Figure 2.4cf, lighter orange areas). These are mostly observed when m and s_D are moderately high, s_H is low, and the innate preference is not very extreme (plateau or peak; Supplementary Material S3). In these cases we find σ to be somewhere between 0.9 and 1. In this region, there are a significant number of immigrants (due to m being somewhat high and s_D not too high) and hybrids (due to low s_H, less extreme innate preference, and large number of immigrants). Under these conditions we find that adaptive hybridization occurs. Paternal imprinting decreases the fitness of female immigrants and hybrids in these cases (moderately

high disruptive selection s_D) because it reduces the chance of them mating with locally adapted individuals to produce locally adapted offspring. This "good-genes" benefit of adaptive hybridization is able to counter the indirect selection for stronger imprinting, which is weak at high λ and low s_H. This effect thus creates an upper limit for the evolution of imprinting even when there is no direct cost against it. When migration is at 0.5 the aforementioned effect disappears (not shown; figures are all black), because half of the offspring will end up migrating again (m = 0.5 can also be thought of as corresponding to full sympatry); no trait is locally adaptive anymore. The above explanation is further supported by the fact that, in a variant of the model where immigrant females are not selected against by divergent selection (*i.e.* the population is sexually dimorphic), we observe, 1) this "good-genes" effect become stronger (there are more immigrant females at the time of mating; Supplementary Material S1), and 2) increasing s_D always leads to lower σ (because immigrant females always survive to reproduction) (Supplementary Material S1).

When there is a cost to imprinting (c>0), σ generally becomes lower, sometimes reaching 0, which means that no evolution of imprinting can occur at all (Figure 2.4ad, white areas). Most trends observed when c>0 can be explained by the risk (chance multiplied by cost) of hybridization. Any conditions that lead to more surviving immigrants increase the chance of hybridization which will in turn lead to stronger selection for premating isolation, raising the value of σ. Forces that increase the chance of hybridization include high migration (m; Figure 2.4d-f), low viability selection (s_H and s_D; Figure 2.4), and low sexual selection (low λ and the innate preference $p_{g,h}^{\text{inn}}$ is "peak" or "plateau"; Supplementary Material S3). In fact, when the migration rate is at maximum (m=0.5) the chance of hybridization is so high that complete imprinting always evolves under the imprinting costs we tested (not shown). Because sexual

selection itself reduces the chance of hybridization (because sexual selection favors the local trait), there are diminishing returns for reinforcement, thus creating an upper limit on how high imprinting can evolve when there are opposing forces such as direct costs (see arguments of Moore 1957).

We arbitrarily picked results from 10 sets of parameter values from the parameter space in which stronger viability selection (s_H or s_D) leads to weaker imprinting (e.g., in the middle of Figure 2.4a) for further analysis. In these parameter spaces, the frequency of hybrids (T_1 and T_2 individuals) generally occurs at an order of magnitude of 10^{-4}-10^{-3}. With such a low chance of hybridization, the indirect selection for imprinting becomes very weak, and is thus easily counteracted by the imprinting cost (c). However, when s_D is low, most immigrants will survive viability selection, therefore the chance of hybridization is high (we again analyzed results from 10 sets of parameter values and found that the frequency of hybrid offspring is at an order of magnitude of 10^{-2}-10^{-1}). As a result, stronger selection against hybrids (s_H) leads to higher σ, which fits what is generally predicted in reinforcement models, *i.e.*, stronger selection against hybrids leads to more reinforcement (reviewed in Servedio and Noor 2003). This is most noticeable near the bottom right part of Figure 2.4d.

One unusual pattern we find is that in Figure 2.4d near the bottom right region σ is lower than the top left region, and increasing migration lowers σ (compare to Figure 2.4a). Here, the benefit of evolving a higher λ is not as high as in other parameter spaces because the high m, low s_D, intermediate s_H, and low λ due to high c allow a high frequency of hybrids in the population (around 20%). The high probability of imprinting on a hybrid parent means imprinting is not as effective in increasing assortative mating preference, especially if the innate preference is

already very divergent (*i.e.*, the "skewed" innate preference). Indeed, this pattern is not observed in the other two innate preference functions (Supplementary Material S3).

Trait Divergence and the Effect of Hybrid and Immigrant Populations

We find that the evolution of imprinting weight does not map directly to an increase in trait divergence. As is shown in Figure 2.2, there is a peak in the degree to which the trait can diverge as λ increases (Figure 2.2b, solid line). The point where maximum trait divergence is achieved (α) is often below 1, at a value different from σ. This peak exists because as imprinted mating preferences become strong, the immigrant females start to mate exclusively with immigrant males, thus reducing positive frequency-dependent sexual selection against the locally rare trait (Servedio 2011). In other words, the allele encoding for stronger imprinting will lead to preferences not only for the locally adapted purebreds, but for any males who are identical to the choosing females' fathers. Depending on the exact value of various parameters, this can include a small but non-trivial number of maladaptive immigrant purebreds or even hybrids, since females with these fathers are also present in the population. The immigrant population thus starts to increase with the increase of λ past the value α, reducing trait divergence. As shown in Supplementary Material S3, the peak in trait divergence occurs at the lowest levels of λ when m is high, s_H is high, s_D is low, and the innate preference is less extreme ("peaked"), because this parameter space best supports the maintenance of a maladaptive immigrant population.

The Initial Conditions for Secondary Contact

In our simulations above we started with imprinting weights across a wide range of values, but whether imprinting would evolve to be stronger or weaker in nature will of course depend in part on whether it is already present at secondary contact. To better understand what

value of imprinting the population is expected to start from at the onset of secondary contact, we ran simulations with m=0 to see what value of imprinting a population is expected to evolve to in allopatry. We find that imprinting weight evolves towards an intermediate-to-high level. In our model, trait polymorphism is quickly lost in a single allopatric population as the locally favored trait becomes fixed; this limits the evolution of imprinting, which requires linkage disequilibrium between the trait and the imprinting locus (Supplementary Material S4). However imprinting is expected to slowly evolve to high levels if polymorphism were maintained by mutation. This is consistent with the high level of imprinting observed in birds. There may also be selective benefits to imprinting in nature that are not included in the model. The evolution of imprinting becomes very slow relative to other loci when the population is (or is becoming) monomorphic, suggesting that other mechanisms not captured by our model (*e.g.* drift, phylogenetic constraints, direct selection through pleiotropy) will probably determine the initial imprinting weight when secondary contact occurs.

If a population starts from high imprinting upon secondary contact, as these results suggest, it will often evolve a lower imprinting weight to arrive at the stable strategy σ. Note that, as described in the section on Maintenance of Polymorphism (and see Supplementary Material S2), if there is successive invasion, the trait symmetry between the patches could be higher than the one used in our simulation making it is possible to evolve towards a low σ that is located below the κ shown in Figures 2.3 and 2.4.

Discussion

In this study we set out to find whether imprinting would evolve to be stronger or weaker after secondary contact, and what effect such evolution would have on hybridization and divergence. One possible expectation was that imprinting weight would evolve to be stronger

due to its effect in decreasing hybridization via a “one-allele” assortative mating mechanism. The alternative hypothesis is that potential increases in hybridization due to mis-imprinting in heterospecific pairs would lead instead to the genetic determination of preferences and the loss of sexual imprinting. We find that stronger imprinting always leads to lower hybridization in our model, imprinting help maintains polymorphism, and imprinting tends to evolve through reinforcement. However, imprinting weight often evolves so that there is only a partial contribution of imprinting to overall female preference, and trait divergence is often not maximized.

Increasing imprinting weight is found to help maintain trait polymorphism across the system because it avoids the problem of preference homogenizing across patches. This is consistent with our expectation that imprinting as a one-allele mechanism can avoid some of the processes that hinders speciation in two-allele mechanisms. It is interesting to note that previous theoretical studies using sympatric models (which can be understood as having preference and traits already homogenized) found the opposite result, that genetically inherited preference is better than imprinting at maintaining polymorphism because genetic preference creates a line of polymorphic equilibria (Verzijden et al 2005).

The evolution of imprinting can be countered by two forces. First, direct costs to imprinting can prevent or cap its evolution. Second, adaptive hybridization can occur, such that immigrant females that benefit by hybridizing with locally adapted males can, if common enough, cause imprinting to evolve to an intermediate weight. In this case, immigrant females benefit by gaining locally adapted "good genes" for their offspring. Because there are opposing forces bother favoring and selecting again imprinting, if a population starts with weak imprinting, stronger imprinting will evolve providing that trait polymorphism has been

maintained, while if the population instead starts with strong imprinting (and there are direct costs or good genes effects for immigrant females) the level of imprinting may evolve to be weaker.

In general selection against hybrids can be attributed to be an important cause of imprinting evolution. When costs to imprinting are present, the population evolves stronger imprinting with weaker divergent selection and with higher migration, both of which increase the chance of hybridization. Costs to imprinting also, however, lead selection against hybrids to generate inconsistent trends because s_H both increase the fitness cost of hybridizing but also ultimately decreases the chance of disassortative pairing.

Because imprinting often evolves to an intermediate weight (σ), hybridization is not always minimized. Trait divergence between populations is also not maximized ($\sigma \neq \alpha$) because the evolved weight of imprinting does not maximize positive frequency-dependent sexual selection (which increases trait divergence; see Servedio 2011). We find that the differences in trait divergence and hybridization across imprinting weights are generally small, thus the above two observations may be concealed by noise in empirical systems. However, it is interesting to note that if a system contains imprinting that is above the stable strategy at the time of secondary contact, as our simulations of allopatry hint may often be the case, imprinting can evolve to be weaker while increasing hybridization. If the trait allele frequency is very asymmetric (which can occur when the two populations are very different in size) and there is little to no divergent selection, it is also possible that the evolutionary weakening of imprinting will cause the system to evolve towards the loss of trait polymorphism.

We also see diminishing returns of reinforcement in reducing hybridization. Because the chance of hybridization is reduced with increased imprinting, the benefit of evolving yet-stronger

imprinting becomes progressively smaller as imprinting weight increases (Moore 1957). This ultimately results in reinforcement being balanced out by imprinting cost or the good genes effect described above, stopping at an intermediate weight of imprinting. Similar results have been studied in previous reinforcement studies, such as the evolution of plant selfing (Holsinger *et al.* 1984).

Sexual selection, when generated by separate loci determining preferences, can oppose trait divergence, because the preference alleles tend to homogenize, leading in turn to homogenization of the trait (Servedio and Bürger 2014). This study finds support for the argument that when there is sexual imprinting, because imprinting is a one-allele mechanism under which sexual selection is based directly on trait frequencies, we can avoid this phenomenon and thus maintain or promote trait divergence. This suggests that taxa in which imprinting is present may have a higher ability to withstand population fusion or extinction during secondary contact than would taxa in which imprinting is not present. This prediction could potentially be tested empirically. One caveat here is the model assumes no bias in imprinting (the imprinted preference always peaks at the father's trait). If imprinting had, for example, a natural skew towards the hybrid phenotypes, it might lead to a reduction in trait divergence. Ten Cate and Vos (1999) suggest that bias in imprinting is usually for traits carried by the parent of the opposite sex that are more extreme (displaced in a direction away from the same-sex parent). This has been demonstrated experimentally *e.g.* in zebra finches. A bias towards phenotypes even more extreme than the opposite parent is very likely a general feature of cognitive discrimination processes (Ghirlanda and Enquist 1999; ten Cate and Rowe 2007). A model by Gilman and Kozak (2015) has also shown that if the imprinting occurs to favor for more extreme traits, in their case by shifting the learner's preference away from an "avoided

phenotype" obtained from oblique imprinting, imprinting would lead to more speciation than if preference is genetically inherited.

One unexpected result from our model is that the evolution of imprinting can lead to the loss of divergence of genetically differentiated preferences. This occurs because the learned preference masks the effect of innate preference as imprinting evolves, rendering the latter effectively neutral, and subject to homogenization across populations or even to the loss of genetic variation altogether. Premating isolation therefore can increase despite genetic preference becoming more similar across populations. This highlights the importance of allowing individuals to imprint naturally when assessing reproductive isolation in a population, as tracking only genetic preference can lead to misleading conclusions regarding reproductive isolation. A good example occurs in threespine sticklebacks, where the imprinted preference seems to add to the innate preference; only offspring that imprinted on a conspecific father show substantial assortative mating. Kozak *et al.* (2011) showed that the imprinted preference of offspring that imprinted on heterospecific fathers was found to cancel their innate preference, leading to a preference for heterospecifics and reversing assortative mating. Offspring raised without a father had neutral preference and no assortative mating. Thus, studies that rear offspring without paternal care may be likely to underestimate the strength of assortative mating that would be found in natural populations where fathers rear offspring.

Also unexpected is the discovery that under some conditions the population does not evolve to complete imprinting even when there is no direct cost, because of adaptive hybridization. Although this outcome is restricted to conditions of moderately strong divergent selection and low selection against hybrids in our model, it is possible that in many taxa this phenomenon will be more prevalent. With a full genome, there may be many loci at which

certain foreign alleles may perform well in a novel genetic background, even when hybrids suffer a more substantial fitness cost at the trait loci, leading to adaptive introgression (Abbott *et al.* 2013; Seehausen *et al.* 2014). Selection against hybridization may be relaxed or counterbalanced in such cases, which may even cause a switch from preferring conspecifics to heterospecifics in the local population. Such environmentally dependent switches to choosing heterospecific mates have been demonstrated in the hybridizing spadefoot toads *Spea bombifrons* and *S. multiplicata* (Pfennig 2007), and may also be present in the hybridization of collard and pied flycatchers (Veen *et al.* 2001).

Effects of model assumptions

Because the purpose of this model is to qualitatively, as opposed to quantitatively, answer our research question, we do not (and sometimes cannot, due to lack of data) attempt to match the parameter values in the model to any specific empirical system. This allows us to make general statements about reinforcement and the evolution of imprinting, as opposed to taxonomically specific statements. For this reason, the long generation times that the model requires for the evolution of imprinting by small intervals should not be taken to mean that the mechanism uncovered in this model is trivial. There are however a few choices we had to make which may affect the qualitative results, which we discuss below. Some of these choices make predictable effects to the model and thus have only limited effect on the general applicability of our results, though, as we can see from this model, unintuitive effects sometimes may arise.

Our most thoroughly analyzed model assumes imprinting is paternal and viability selection affects both sexes equally. These two assumptions may not hold in all empirical systems. Thus we also ran a set of simulations with maternal imprinting, and another where viability selection only acts on males (*i.e.*, there is sexual dimorphism). The results are similar to

those in the original model (Supplementary Material S1), as these two assumptions are not required for any of the most fundamental mechanisms of the causes and effects of imprinting evolution that we describe (imprinting decreasing hybridization and maintaining polymorphism, adaptive hybridization, imprinting masking innate preference, *etc*.) One difference between these cases is that in the sexual dimorphic model, immigrant females always survive to reproduction. Increasing divergent selection, which makes it more beneficial for immigrant females to mate with local males via adaptive hybridization will thus lead to lower levels of imprinting than in the sexually monomorphic model (Supplementary Material, Figure S1). Note that deviation from these two assumptions may be the result of differences in other life history traits (Lande 1980; Tramm and Servedio 2008; Chaffee *et al.* 2013; Invernizzi and Gilman 2015), which may also contribution to the evolution of imprinting weight.

Another assumption in our model that may not always hold in nature is that learning occurs only in the form of sexual imprinting that occurs before migration. There are species in which learning occurs after migration, for example in the form of dialect matching or mate-choice copying from peers (horizontal) or unrelated individual of the previous generation (oblique) (*e.g.*, Verner 1975; Galef jr and White 1998). In these cases, individuals are learning from a set of individuals that is less diverged (since there are migrants present), thus learning would tend to homogenize the female preference and increase hybridization. As such, we would expect an increased chance of losing polymorphism due to homogenized preference, and reinforcement could potentially lead to evolution toward less learning.

Our model has broad implications because imprinting is found across a wide range of taxa. Still, the underlying mechanism of imprinting may vary across taxa (Immelmann and Suomi 1981; Oetting *et al.* 1995). The operation of imprinting is best-studied in birds, but our

current knowledge is too sparse to suggest which modelling choice is more appropriate or allow a taxonomic comparison. Still, we can speculate that in some cases preferences may be the result of an additive combination of an imprinted preference with an innate preference, as modeled here, while in other cases imprinting may affect preference perhaps by narrowing and/or shifting the innate preference distribution (increasing discrimination), or through a weighted geometric or harmonic mean instead of a weighted arithmetic mean. If an innate preference serves as a basis for imprinting to modify, it may not homogenize as drastically when imprinting weight increases. while if innate and imprinted preferences are combined though a weighted geometric mean, for example, it would increase the preference for hybrids in females whose innate component and the imprinted components are drastically different. This latter case will potentially make imprinting less likely to evolve to a high value through reinforcement. In short, we recommend more empirical research on how innate and the imprinted components of preference may be combined because it can provide information crucial for understanding how imprinting evolves.

In species with extra-pair paternity (EPP), imprinting would not be as effective in decreasing hybridization as without EPP, since the social father of an EPP offspring may not carry the same trait as the genetic father. However, because the female preference for a social partner may not be the same as that for an extra-pair mate, and the degree to which imprinting affects the two preferences may be different, it remains to be investigated how EPP would affect the evolution of imprinting through reinforcement. In sympatric flycatcher populations the rate of heterospecific pairing is much higher than the rate of hybridization. Females in mixed species pairs seek extra-pair copulations with conspecific males (Veen *et al.* 2001; Cramer *et al.* 2016). Accordingly, a large proportion of the chicks in such clutches are genetically of the same species

as their mother. Female chicks in such nests would therefore be prone to hybridize due to mis-imprinting on their social, heterospecific father. Such mis-imprinting is not possible in our model which does not include EPP, which may explain why imprinting has evolved to be weaker in sympatry in pied and collared flycatchers (Saether *et al.* 2007) and Gouldian finch (Pryke 2010), both of which have EPP. In both systems mate preferences appear to be innate and cross-fostering experiments have failed to demonstrate any effect of sexual imprinting on mate preferences (Saether *et al.* 2007; Pryke 2010). A previous theoretical study shows EPP also affects whom the individuals evolve to imprint on (Invernizzi and Gilman 2015).

Learning and Speciation

Servedio *et al.* (2009) show that imprinted preference can be used as a mechanism for reinforcement, here we show that imprinting itself can evolve as a result of reinforcement. Imprinting can of course also evolve through other evolutionary mechanisms, for example it may be beneficial (rather than just not costly) if it evolves initially to recognize gender (Vos 1995), or as a "second chance" to obtain traits that are favored when other transmission modes are not efficient (Creanza *et al.* 2016). Another form of imprinting, habitat imprinting, can evolve under disruptive selection if the cost is low, which leads to immediate speciation (Beltman and Metz 2005).

Other forms of learning are also known to have positive effects on speciation (Verzijden *et al.* 2012). Preference learning based on interactions with prospective mates can increase assortative mating (Servedio and Dukas 2013). Social learning that leads to exploitation of novel food source can also lead to speciation if selection favors specialization (*e.g.*, Payne *et al.* 2000). More broadly, learning as a form of plasticity reduces heritability but can also accelerate evolution by increasing phenotypic variation, shifting the phenotype distribution to somewhere

with a steeper selection gradient, and helping the crossing of fitness valleys (Ancel 2000; Borenstein *et al.* 2006). We believe that further studies will continue to reveal important processes by which learning interacts with speciation.

Tables and Figures

Table 2.1: Model summary

Life History Stage	Relevant Loci and Parameters	Biological Meaning
Birth	$L \xrightarrow{\text{defines}} \lambda$	L locus encodes learning strength
	$P \xrightarrow{\text{defines}} p_{g,h}^{\text{inn}}$	P locus encodes innate preference
Paternal Imprinting	$T_{\text{father}} \xrightarrow{\text{defines}} p_{g,h}^{\text{imp}}$	Father's trait is imprinted on
	$\lambda, p_{g,h}^{\text{imp}}$ and $p_{g,h}^{\text{inn}} \xrightarrow{\text{defines}} p_{g,h}$	Female preference is formed
Migration	m	Two island migration at constant rate
Viability Selection	s_D and $s_H \xRightarrow{\text{select on}} T_a, T_b$ and T_c	Divergent and disruptive selection on T
	$c \xRightarrow{\text{selects on}} L$	Cost of learning
Sexual Selection	$p_{g,h} \xRightarrow{\text{selects on}} T_a, T_b$ and T_c	Female preference selects on T

Table 2.2: Relative viability ($v_{x,y}$) of the trait phenotypes due to viability selection

Genotype (Ta,Tb,Tc)	Trait Phenotype (T)	Relative Viability in Patch 0	Relative Viability in Patch 1
(0,0,0)	0	$(1 + s_H)(1 + 3s_D)$	$1 + s_H$
(0,0,1), (0,1,0), or (1,0,0)	1	$1 + 2s_D$	$1 + s_D$
(0,1,1), (1,0,1), or (1,1,0)	2	$1 + s_D$	$1 + 2s_D$
(1,1,1)	3	$1 + s_H$	$(1 + s_H)(1 + 3s_D)$

Figure 2.1: Histograms for female preference components, $p_{g,h}^{inn}$ and $p_{g,h}^{imp}$. P_1 females' innate preferences are exact opposite of those of P_0 females. Imprinted preference on T_2 and T_3 father are exact opposite those of T_1 and T_0 father. (The parameters for these binomial distributions are n=3 and p= (a) 1/6, (b) 1/4, (c) 1/3, (d) 1/6, (e) 7/18)

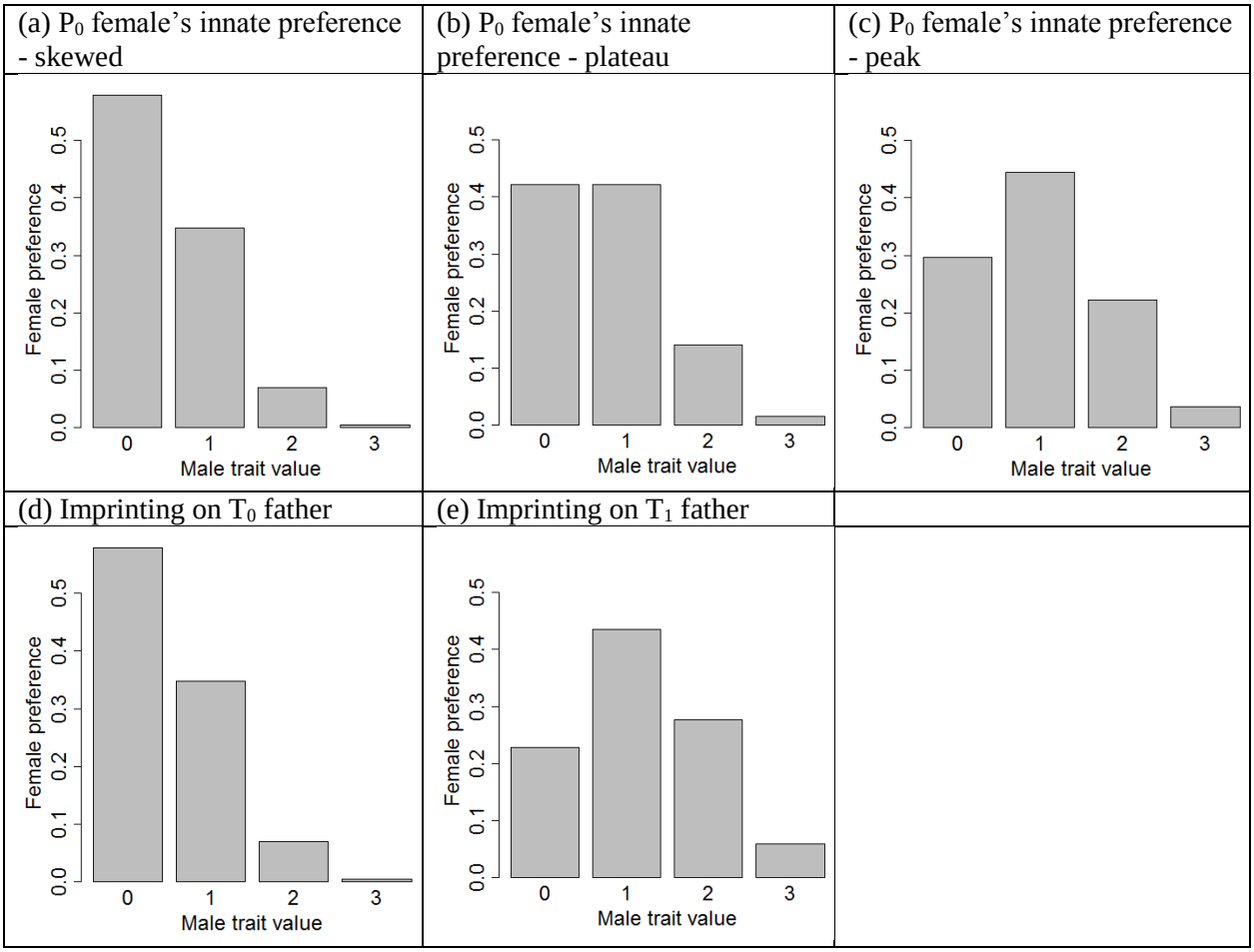


Figure 2.2: (a) An illustration of κ and σ, and how the increase of various parameters would affect their locations. κ is also positively affected by the degree of initial asymmetry in allele frequencies of P and/or T across patches, and both κ and σ are higher if the innate preference is more extreme (*e.g.*, "skewed" rather than "peak"). (b) As λ increases, once past κ, polymorphism is maintained, and the trait divergence (the continuous line, representing the average difference in the frequency of T_0 alleles across patches, *i.e.*,$(t_{a,1,0}- t_{a,1,1}+ t_{b,1,0}- t_{b,1,1}+t_{c,1,0}- t_{c,1,1})/3$) starts to increase until it peaks at α. The divergence at the preference locus P (the dashed line, representing the difference in the P_0 allele frequency across patches, *i.e.*, $p_{a,1,0} - p_{a,1,1}$) initially increase with the T loci, but then decreases due to indirect selection being masked by learning. The frequency of hybrids (the dotted line, representing the sum of the frequencies of individuals with trait values 1 and 2) jumps up from 0 at κ, and then starts to decrease. In the case shown here, $\sigma<\alpha$. (s_H=0.5, s_D=0.071, m=0.01, c=0, innate preference=peak)

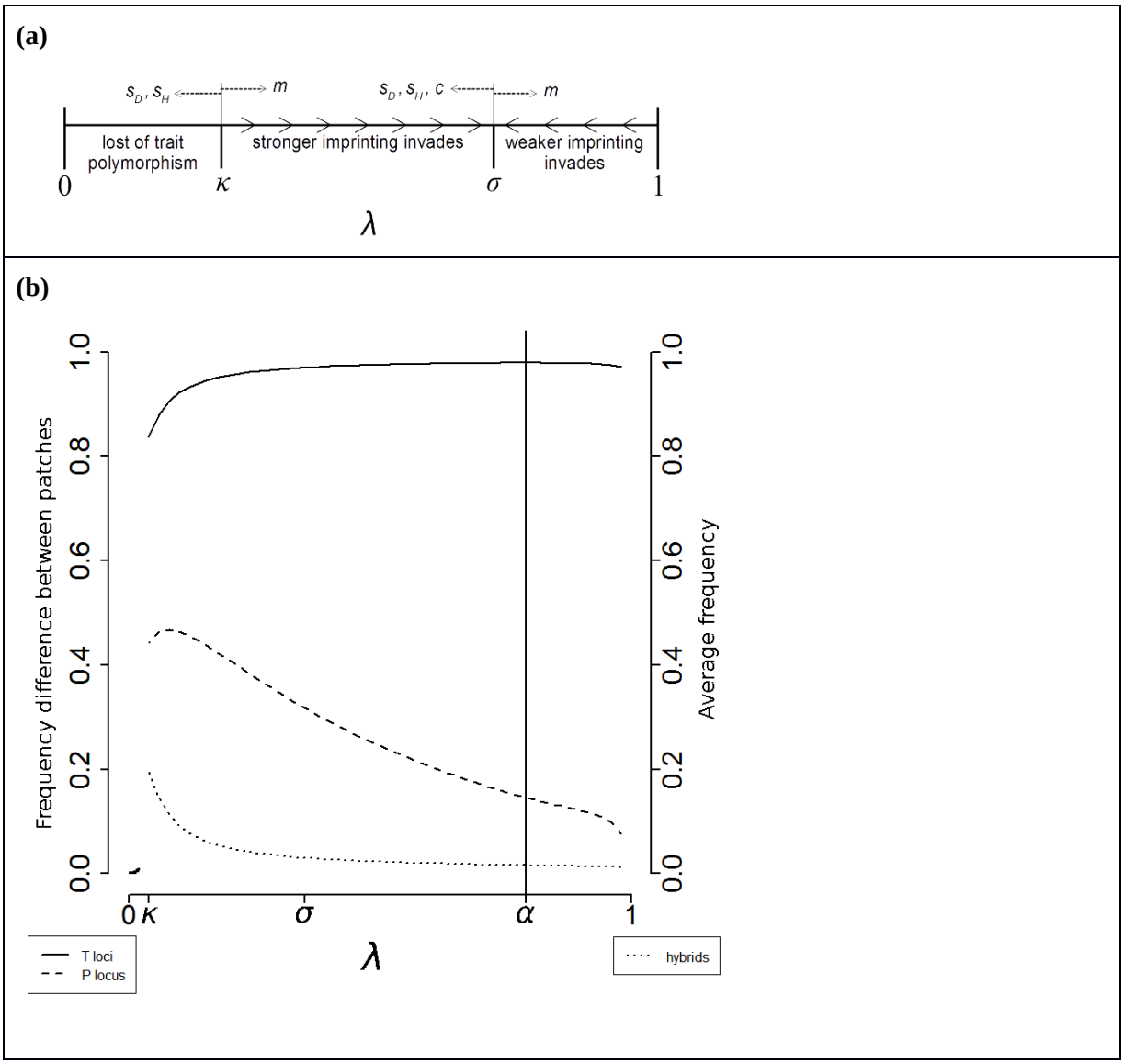


Figure 2.3: κ, the threshold below which polymorphism at the trait loci is lost, and as a result imprinting cannot evolve. Polymorphism at the preference and trait loci is maintained above this threshold. In non-white regions (κ>0) innate preference alone cannot maintain polymorphism, and in black regions (κ=1) polymorphism is always lost, and as a result imprinting cannot evolve. (Linear interpolation used for plotting, grid density: 8x8; c=0, innate preference=peak)

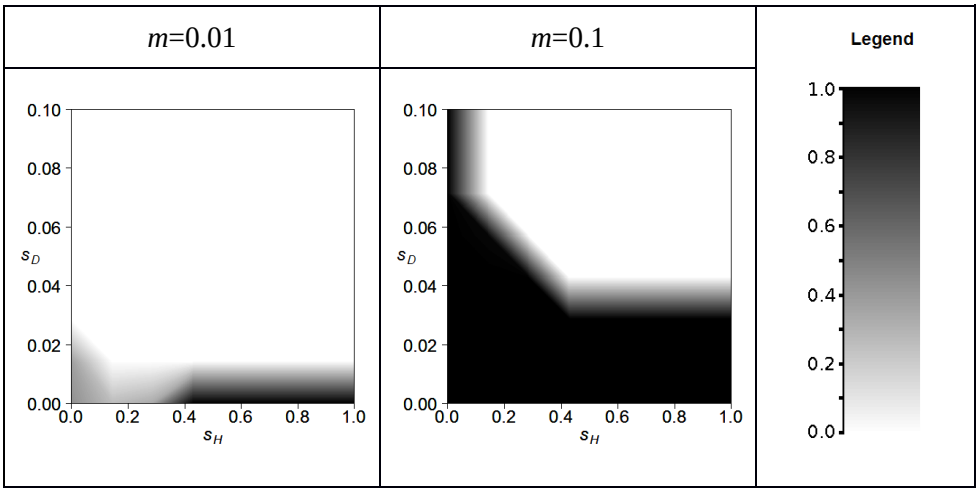
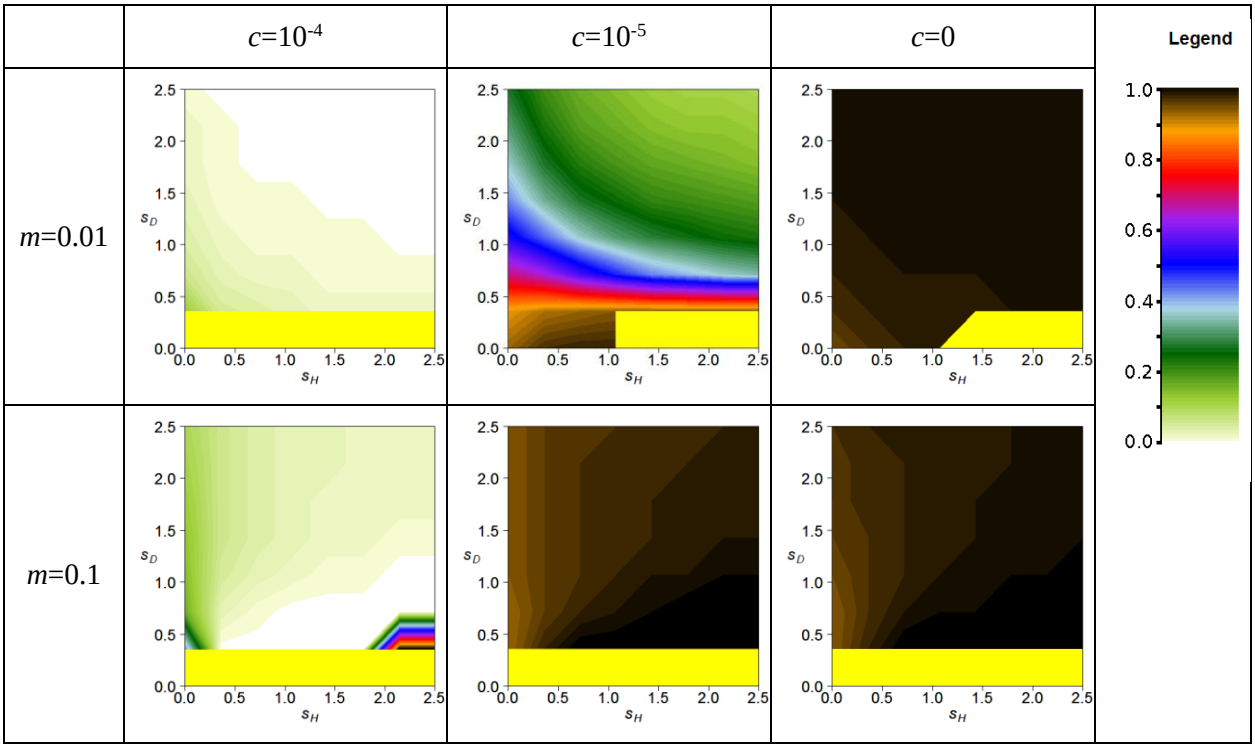


Figure 2.4: σ, the level of imprinting that evolves when polymorphism is maintained. In the yellow area, the threshold for maintaining polymorphism is above the stable strategy ($\sigma < \kappa$) given the starting conditioned described in the Simulation section. (Innate preference=peak; linear interpolation used for plotting, grid density: 8x8). Note that the color scheme has a different meaning in this figure versus Figure 2.3. In Figure 2.3 black (κ=1) corresponds to loss of polymorphism even with complete imprinting, so no imprinting can evolve, whereas here black (σ =1) refers to complete imprinting.



CHAPTER 3 : ASSORTATIVE MATING BY AN OBLIQUELY TRANSMITTED LOCAL CULTURAL TRAIT PROMOTES GENETIC DIVERGENCE

Summary

The effect of learned culture (*e.g.*, birdsong dialects and human languages) on genetic divergence is unclear. Previous theoretical research suggests that because oblique learning allows phenotype transmission from individuals with no offspring to an unrelated individual in the next generation, the effect of sexual selection on the learned trait is masked. However, I propose that migration and spatially constrained learning can form a statistical association between cultural and genetic traits, which may allow selection on the cultural traits to indirectly affect the genetic traits. Here, I build a population genetic model that allows such a statistical association to form, and found that sexual selection and divergent selection on the cultural trait can indeed help maintain genetic divergence through such a statistical association. Furthermore I found that the genetic divergence maintained by this effect persists even when the cultural trait changes over time due to drift and mutation. These results suggest the role of obliquely transmitted traits in evolution may be underrated, and the lack of one-to-one associations between cultural and genetic traits may not be sufficient to disprove the role of culture in divergence.

Introduction

Cultural transmission is the propagation of behavior between individuals through learning. It can be found across a wide range of animal taxa, including insects, fish, amphibians, birds, and mammals, in numerous forms, such as sexual imprinting, mate-choice copying, the

learning of foraging strategies, birdsong learning, and language acquisition (Hoppitt and Laland 2013). The interaction between cultural transmission and evolution has been gaining attention in recent years (Laland *et al.* 2015). One source of interest is in understanding human evolution. Many have suggested that human languages may serve as a barrier to gene flow in human populations, with some empirical support (Barbujani 1997). Another field in which the role of learning is being heavily examined is speciation (Verzijden *et al.* 2012). In particular, learning can change assortative mating signals and preferences, thus affecting the evolution of pre-mating isolation (Verzijden *et al.* 2012; Pfennig and Servedio 2013). Because speciation bridges micro- and macro-evolution, the contribution of learning to speciation suggests that the short-term mechanism can have long-term evolutionary consequences.

Speciation with gene flow is considered particularly complex because recombination can break up associations between loci, hampering the divergence of genes that are not directly under divergent selection even at very low migration rates (Felsenstein 1981). To counter the effect of recombination there needs to be mechanisms to build linkage disequilibrium (LD), the non-random statistical association between loci due to mechanisms such as non-random mating, migration, selection, epistatic effects, and physical linkage. Considered one of the most important factors in speciation with gene flow, LD allows loci that are not directly under divergent selection to diverge by hitchhiking along with other loci that are under direct selection (Servedio 2009; Smadja and Butlin 2011).

Similar to LD which occurs between genetic loci, cultural traits can form statistical association with other cultural traits ("cultural linkage disequilibrium"), or genes ("phenogenotypic association") (Feldman and Zhivotovsky 1992). The latter is particularly interesting as it allows forces affecting a cultural trait to indirectly affect genetic evolution

(Feldman and Zhivotovsky 1992). Analogous to LD, phenogenotypic association can form through many mechanisms, including genetic traits affecting the ability to learn (Feldman and Zhivotovsky 1992), the learning of assortative mating preferences or signals (Tramm and Servedio 2008), and spatially localized learning which can coincide with local genetic markers. The formation of phenogenotypic association due to localized learning and its effect on speciation has yet to be studied, although I suspect it has already been observed in multiple empirical systems but with its significance overlooked (*e.g.*, in MacDougall-Shackleton and MacDougall-Shackleton 2001; Kenyon *et al.* 2016; Camacho 2017; Lipshutz *et al.* 2017). The learning of assortative mating traits is also particularly interesting because assortative mating alone already creates pre-mating isolation that reduces the effective gene flow between subpopulations, favouring the formation of LD (Gavrilets 2004).

One example of cultural transmission interacting with assortative mating is birdsong dialects (Verzijden *et al.* 2012). It has long been suggested that learned birdsong dialects may act as barriers to dispersal at contact zones (Baker and Mewaldt 1978; Petrinovich *et al.* 1981), either because there are female preferences for local dialects (Searcy *et al.* 1981), or because foreign songs elicit less response from competing males and thus are less effective for territorial defence (Kenyon *et al.* 2016). Attempts to find empirical support for this hypothesis have yielded conflicting results (Gill and Murray 1972; Baker 1975; Baker and Thompson 1982; Zink and Barrowclough 1984; Lachlan and Servedio 2004; Kenyon *et al.* 2011, 2016; Lipshutz *et al.* 2017; Mason *et al.* 2017), which correlation between genetic and cultural distance found in some research but not in others. A recurring problem with empirical attempts to answer this question is that the researchers look for correlation, which does not imply causation. In this study I demonstrate how such correlation can be generated without causation, and how causation can

exist when correlation is hard to observe. On the theoretical front, it is known that dialects are easier to maintain if learning is prevalent (Olofsson and Servedio 2008), and various conditions during secondary contact can maintain spatial correlations between genes that affects the production and recognition of songs and the learned song (Rowell and Servedio 2012). What is not known is whether genes that do not directly interact with birdsong, which likely include most genes sequenced in aforementioned empirical research, can diverge because of birdsong dialects.

The most common form of song learning in wild birds is oblique learning, *i.e.*, learning from unrelated individuals of the previous generation (Lynch 1996). Previous theoretical work has shown that obliquely learned traits may hamper speciation through divergent sexual selection because sexual selection on these traits is masked when unmated individuals obliquely transmit their traits to unrelated juveniles (Yeh and Servedio 2015). However, because spatial limitations to cultural transmission can form local cultures, which may lead to phenogenotypic association between the learned trait and genes, sexual selection on an obliquely learned trait may indirectly promote the divergence of genetic loci associated with the learned trait. Indeed, at least in some systems, spatial constraints on learning do occur on a scale similar to that of migration rate (Searcy *et al.* 2002; Laiolo and Tella 2005; Garland *et al.* 2011), *i.e.*, individuals do not learn from individuals far outside their migrating distance and migration does not swamp out local dialects. Even in birds that migrate for some distance, migration does not remove the social relationship that determines the tutor-pupil relationship (Templeton *et al.* 2012).

Other mechanisms that may lead to the formation of local culture include cultural mutation and cultural drift (Marler and Tamura 1962; Potvin and Clegg 2015). Reasons cultural traits mutate include innovation and erroneous learning occurring at varying rates (Lachlan and Slater 2003; Catchpole and Slater 2008). While some birdsong is known to change over time,

with males more responsive to current than historic songs (Luther and Derryberry 2012), it is not known whether such change over time affects genetic divergence across space. It is possible that the replacement of local culture over time ("cultural trait turnover") allows for the maintenance of genetic divergence without a long-term association between a specific cultural trait and genetic markers, *viz.*, that as long as the cultural traits remain different across space, genetic divergence can be maintained, even if local culture is changing over time through drift and mutation. If that is the case, local variation in birdsong may be large enough (because the local culture is in the process of changing) to obscure the difference between subpopulations, while the genetic divergence across space is maintained by the obscured cultural trait difference. This adds a layer of complexity for empirical research since cluster analysis using birdsong may not identify actual breeding boundaries even if they exist.

Here I build a population genetic model analysed with both deterministic and stochastic simulations. The model consists of two patches, each having locally adapted cultural traits that are transmitted obliquely within the patches and are the basis of assortative mating. Besides the cultural trait, individuals also carry two loci that lower hybrid viability through a Dobzhansky-Muller incompatibility. Lower hybrid viability is included to reflect a secondary contact scenario. The variable of interest is the LD between the two loci, which indicates both the genetic divergence across the patches, as well as the reduction in the frequency of individuals with hybrid genotypes. I investigate 1) Whether phenogenotypic association can form due to spatial limitations to cultural transmission, 2) whether divergent selection and assortative mating based on the cultural trait increase LD between the two genetic loci through phenogenotypic association, 3) whether cultural drift and cultural mutation, both of which can contribute to

diverged local culture, can also increase LD, and 4) whether a novel cultural trait can replace the local resident trait through drift while maintaining the divergence of genetic loci.

The Basic Model

I start with 2 patches populated by a haploid population with 2 diallelic genetic loci (A and B) that are involved in a Dobzhansky-Muller incompatibity and a cultural trait "locus" (C) that has two "alleles" (*e.g.*, two types of birdsong). $p_{i,x}$ is the frequency of phenogenotype *i* in patch *x*. For now I assume no drift and no mutation in all three loci. The life cycle starts with migration (m) between the two patches, followed by divergent selection (s_D) on the cultural trait and selection against hybrids (s_H) on the genetic loci, after which sexual selection (α) occurs based on phenotype matching (Hauber and Sherman 2001) at the C locus. After reproduction and recombination, the offspring then learn the cultural trait obliquely within each patch. The measurement for progress towards speciation is the LD between A and B loci, regarding which I assess whether it would positively correlate with cultural trait divergence (the frequency difference of C alleles across patches).

For migration I use a 2-island migration model to allow spatial structure. The frequency of phenogenotype *i* in patch *x* after migration is $p_{i,x}^{\mathrm{mig}} = (1-m)p_{i,x} + mp_{i,1-x}$, where m is the migration rate. Migration, the only step in the life cycle that leads to interaction between patches, is what forms a phenogenotypic association, an essential component of this model. I assume the migration rates are identical across sexes, which is not expected to be critical to the qualitative results presented below.

After migration the population is subject to viability selection. The cultural trait is assumed to be under divergent selection (s_D). Examples of environmental selection on a cultural trait include high-frequency vocal signals being more effective in environments with low-

frequency ambient noise (Luther and Derryberry 2012), tonal languages developing more easily in humid environment, and, looking beyond vocal signals (Everett *et al.* 2015), long striking tools being less efficient in aquatic environments (Mann and Patterson 2013). For loci A and B, genotypes Ab and aB are considered hybrids and there is selection against these hybrids (s_H). Therefore, the phenogenotype frequency after viability selection is

$$p_{i,x}^{\mathrm{vs}} = \frac{(1-d_{i,x}s_D)(1-h_i s_H)}{\overline{w_x}} p_{i,x}^{\mathrm{mig}} \quad \text{..... (Eq 1)}$$

where $d_{i,x} = 1$ for those carrying the locally maladaptive cultural trait (C_2 if x=1 and C_1 if x=2), otherwise $d_{i,x} = 0$; $h_i = 1$ for hybrids, otherwise $h_i = 0$; and $\overline{w_x} = \sum_i \left(1 - d_{i,x}s_D\right)(1 - h_i s_H) p_{i,x}^{\mathrm{mig}}$ is the local mean viability.

For sexual selection the model assumes polygyny, *i.e.*, all females have the same reproductive fitness while males are under direct sexual selection. Mate choice is based on phenotype matching, such that females are 1+α time more likely to prefer males with whom they share the same song than to prefer other males (this follows (Servedio 2000b), and is referred to as "similarity-based" in (Gavrilets 2004)). Thus pairings between males with phenogenotype *u* and females with phenogenotype *v* occurs in patch *x* at frequency

$$F_{v,u,x} = \frac{(1+y_{v,u}\alpha)p_{v,x}^{\mathrm{vs}} p_{u,x}^{\mathrm{vs}}}{\sum_t (1+y_{t,u}\alpha)p_{t,x}^{\mathrm{vs}}} \quad \text{..... (Eq 2)}$$

where $y_{v,u} = 1$ if *u* and *v* have the same cultural trait, otherwise $y_{v,u} = 0$.

Reproduction then allows the alleles of A and B to be passed on to the offspring, with free recombination, *e.g.*, $p^{\mathrm{rep}}_{AB,x} = F_{AB,AB,x} + \left(F_{AB,Ab,x} + F_{Ab,AB,x} + F_{AB,aB,x} + F_{aB,AB,x}\right)/2 + \left(F_{AB,ab,x} + F_{ab,AB,x} + F_{Ab,aB,x} + F_{aB,Ab,x}\right)/4$ for offspring genotype AB in patch *x*. The offspring then obtain their cultural traits through oblique learning. I assume equal transmission rates for all cultural alleles, meaning there is no "fecundity" selection on C, and the rates are also independent of tutors and offspring's genotype. The probability that an offspring will learn a particular trait is simply the frequency of that trait in the previous generation in the same patch. For example, the frequency of phenogenotype ABC_1 in patch *x* after learning is $p^{\mathrm{lrn}}_{ABC_1,x} = p^{\mathrm{rep}}_{AB,x}\, p^{\mathrm{vs}}_{C_1,x}$. This equation suggests that preferences for a song do not affect song frequency in the next generation, which correspond to the fact that offspring can learn from unmated males.

Note that due to learning being oblique, right after this step the phenogenotypic associations in the offspring become 0 within individual patches. (Phenogenotypic associations are calculated exactly analogous to linkage disequilibria. For example, the phenogenotypic association between A and C is $p_{AC_1}p_{aC_2} - p_{aC_1}p_{AC_2}$.) Phenogenotypic associations do however exist if calculated across patches. Such associations also become present in individual patches when migration admixes the two patches; these associations are the foundation of all main discoveries in this model.

The recursion equations for the frequencies of alleles at the A, B, and C "loci", and for LD can be obtained, but are too complex to solve for equilibria analytically. I therefore use deterministic simulation to find the equilibria. The program, written in Wolfram Mathematica 10.0 (Inc. 2014), iterates through the lifecycle repeatedly for 4000 generations, by which time the changes in phenogenotypic frequencies per generation become negligible. To simulate a secondary contact scenario, which is the scenario in which contacts between subpopulations with

previously established different cultures are expected to be found, I start with the two patches diverged, but with small deviation from exact symmetry to avoid potential artefacts ($p_{A,1} = 0.99$, $p_{B,1} = 0.99$, $p_{C_1,1} = 0.99$, $p_{A,2} = 0.011$, $p_{B,2} = 0.012$, $p_{C_1,2} = 0.013$).

Results

The first thing to note is that cultural trait divergence (measured as frequency of C_1 in patch 1) is maintained mainly by migration-selection balance. The main parameters affecting it are s_D and m. Assortative mating preference α has limited, if any, effect on the cultural trait because regardless of how many offspring a male has, its chance of passing on the song to an individual in the next generation through oblique learning is the same. The selection against hybrids s_H has very limited effect on the cultural trait as well. This is because hybrid frequencies are low across both patches, so migration does not lead to strong association between the hybrids and any particular cultural trait (*i.e.*, the 3-way phenogenotypic association is close to 0). A few spot checks with drastically different initial conditions lead to results indistinguishable from the ones above, as long as the genetic divergence starts high (if not the genetic polymorphism may be lost). As the cultural trait evolves very quickly under migration-selection balance, it reaches equilibrium before the genetic alleles equilibrate.

If the cultural trait does not exist, or is not diverged, loci A and B will simply homogenize due to migration, and then lose polymorphism because s_H selects against rare alleles. Here, with the cultural trait divergence maintained by migration-selection balance, the phenogenotypic associations then enables mate choice based on cultural trait C to reduce hybridization rate for the genetic loci, and allow divergent selection and sexual selection acting on C to promote genetic divergence, leading to positive LD between the loci A and B. This LD is maximized at intermediate α (Figure 3.1*a*), increases with selection coefficients s_H and s_D

(Figure 3.1*b*), and decreases with migration rate *m* (not shown). The reason an intermediate α is the most effective at maintaining the divergence of the cultural trait (and thus LD) is that under phenotype matching, when mating preference is too strong, disassortative mating decreases and males with a rare trait becomes just as successful as individuals with a common trait, leading to the loss of positive frequency-dependent sexual selection (Servedio 2011; Yeh and Servedio 2015). The peaks of LD occur at higher α with increased s_D because when s_D is high, the immigrant population is small, therefore the loss of sexual selection due to immigrants mating with their own kind requires stronger preferences. In all parameter space tested, s_H must be above some threshold for LD to be maintained, suggesting pre-mating isolation alone is not sufficient to satisfy Udovic's condition to stop introgression (Udovic 1980).

Drift, Cultural Mutation and Cultural Trait Turnover

I then add drift and cultural mutation to the life cycle. Drift occurs along with viability selection, calculated as $p_{i,x}^{\mathrm{dft}} = X^*/N$, where *N* is the subpopulation size within a patch and X^* is a random number sampled from the binomial distribution $\Pr(X = k) = \binom{N}{k}\left(p_{i,x}^{\mathrm{vs}}\right)^k\left(1 - p_{i,x}^{\mathrm{vs}}\right)^{N-k}$. Pervious occurrence of $p_{i,x}^{\mathrm{vs}}$ in Equation 2 and the equation for p^{lrn} are then replaced with $p_{i,x}^{\mathrm{dft}}$. Drift can also potentially be modelled as occurring with reproduction, which I also investigated but did not observe any difference from the current model.

I assume cultural mutation occurs at rate μ, while genetic mutation is assumed to be negligible compared to cultural mutation and thus not included. Using ABC_1 as an example, the frequency after oblique learning is thus

$$p^{\text{lrn}}_{ABC_1,x} = p^{\text{rep}}_{AB,x}\left((1-\mu)p^{\text{vs}}_{C_1,x} + \mu p^{\text{vs}}_{C_2,x}\right) \text{..... (Eq 3)}$$

The simulation conditions are the same as those in the previous section, except because the simulation is now stochastic, I run 20 replications for each parameter set and record the frequencies every 500 generations.

Finally, the cultural trait C is then expanded to include multiple possible song types or "alleles" to see whether a novel cultural trait that is equally locally adapted can replace the resident trait through drift, while maintaining the LD. Increasing the total possible number of C alleles can also be thought of as decreasing physiological constraints on the production of birdsong, or increasing perception to notice previously unperceivable differences in the signal. $C_1, C_3, C_5 \cdots$ are adapted to patch 1, and $C_2, C_4, C_6 \cdots$ patch 2. I assume C_j and C_{j+2} have identical viability so that they can potentially replace each other through drift. I also assume all cultural traits have the same mutation rate, and all traits are equally likely to arise through mutation. Equation 3 is thus modified to:

$$p^{\text{lrn}}_{ABC_1,x} = p^{\text{rep}}_{AB,x}\left((1-\mu)p^{\text{vs}}_{C_1,x} + \frac{\mu}{n-1}\left(p^{\text{vs}}_{C_2,x} + p^{\text{vs}}_{C_3,x} + \cdots + p^{\text{vs}}_{C_n,x}\right)\right) \text{..... (Eq 4)}$$

where *n* is the number of possible C alleles in total. The simulation starts with the same conditions as before, with patch 1 and patch 2 populated by C_1 and C_2 respectively. All other cultural traits arise through mutation. Cultural trait divergence with multiple alleles is measured by adapting Nei's genetic distance: $D = 1 - \frac{\sum_u p_{u,1}p_{u,2}}{\sqrt{(\sum_u p^2_{u,1})(\sum_u p^2_{u,2})}}$, where *u* indicates the C alleles.

This measurement captures the probability that two individuals randomly chosen from the two patches have different traits.

Results

Figures 3.1*c* and 3.1*d*, obtained from a representative set of parameters, illustrates three noteworthy findings of the model. First, there is a chance of losing polymorphism at the genetic loci across both patches when drift is strong, which appear as LD=0 on the figure. Second, when polymorphism is maintained, there exists a positive correlation between cultural trait divergence and LD. Third, cultural trait turnover can occur, which does not have noticeable effect on LD.

Polymorphism is more easily maintained when the population is large, α is at an intermediate value (Figure 3.2*a*), s_D is high (Figure 3.2*a*) or s_H is very high or very low (not shown). Large population size means lower drift, thus better maintenance of polymorphism. Intermediate α and high s_D both increase cultural trait divergence, which helps maintain polymorphism through phenogenotypic associations. s_H selects against rare alleles, leading to loss of polymorphism; however when starting with a diverged population, a very high s_H removes the immigrant population very effectively, thus help maintain polymorphism. The fact that s_H selects against rare alleles also means that re-establishing a lost allele would be difficult even if there is genetic mutation to bring it back.

When polymorphism is maintained, increasing cultural trait divergence has a positive effect on LD, as shown by the exponential regression lines in Figure 3.1*c*. As discussed in the previous section, divergent and sexual selection on the cultural trait can indirectly increase LD through phenogenotypic associations. Cultural trait divergence caused by drift, on the other hand, does not noticeably increase LD (*i.e.*, whenever polymorphism is maintained frequently enough to run a regression, no positive trend is found within the same coloured dots in Figures

3.1*c* and 3.1*d*). While the effect of drift should be passed to the genetic loci and thus increase LD, the results show that such a mechanism is too weak to overcome the stochastic noise also caused by drift.

The data points with a black x on top in Figure 3.1*c*, *d* are cases in which C_3 is replacing C_1 in patch 1 (defined as having ten times more of C_3 than C_1, while $p_{C_3,1} > 0.5$ to avoid cases in which both frequencies are low). Such cultural trait turnover is most common when drift is strong and the cultural mutation rate μ is moderate (Figure 3.2*b*). Strong drift allows novel traits to drift to high frequency, while a moderate μ provides enough novel traits for drift to act on, but not so high that it prevents the novel trait from reaching high frequency. Cultural trait turnover is not affected by s_H, s_D, m, α, or whether polymorphism in the genetic loci is maintained or not. This is because, as stated before, the C locus is maintained by migration-selection balance, and s_D and m make no distinction between C_{v} and $C_{\mathrm{v}+2}$. All data points that show turnover fall right within the distribution of the other dots, and the rate of losing polymorphism in the genetic loci are similar as well. This suggests that once the novel trait has replaced the original trait through cultural turnover, LD can be maintained as before.

However, increasing the number of possible cultural alleles decreases the positive correlation between cultural trait divergence and LD (Figure 3.2*c*). This is most likely because having multiple coexisting cultural alleles dilutes the effect of positive-frequency dependent sexual selection. In other words, LD may reduce slightly *during* turnovers. Nonetheless, even in a simulation with 16 alleles, the positive correlation between LD and cultural trait divergence is still present. Similarly, higher μ reduces cultural trait divergence, and as a result, LD (Supplementary Material Figure S1, compare between rows). The effect is again relatively small. In fact, cultural trait divergence and LD can be maintained even at μ=0.1. If drift or other sources

of noise are strong, the reduction in LD due to cultural mutation or multiple coexisting cultural alleles may be hard to detect empirically.

Selection before Migration

So far viability selection (s_H and s_D) has been assumed to occur after migration but before mate choice. Some empirical systems may be better described with viability selection occurring *before* migration. Furthermore, because in the above model divergent selection on the cultural trait happens right after the phenogenotypic association is formed through migration, it is not clear whether the correlation between LD and cultural trait divergence is mainly maintained by divergent selection or sexual selection. In the case where viability selection occurs before migration, any correlation between LD and cultural trait divergence will have to be the result of sexual selection. I examine this case by simply replacing p, p^{vs}, and p^{mig} respectively with p^{vs}, p^{mig}, and p on the right hand side of all previous equations, and then performing the same analyses as above.

The results show that all previously described trends are still present after changing the order of the life stages (Supplementary Material, compare Figures S1 and S2). The only difference between the two models is that the positive correlation between the cultural trait divergence and LD is lower when selection occurs before migration than when selection occurs second (Figure 3.2*c*), which mainly happens because strong divergent selection no longer contributes immediately to LD through the phenogenotypic association newly formed by migration. Instead, divergent selection functions to help maintain cultural trait divergence. This maintained trait divergence in turn contribute to phenogenotypic association formed during migration, which allows the effect of positive frequency-dependent sexual selection to maintain on LD. Trait divergence is also required for sexual selection in this model since sexual selection

is positively frequency-dependent. The different results between the two models are unrelated to s_H occurring at different times, because the hybrid frequencies are equally low across the two patches (except due to drift), so the whether migration occurs before or after viability selection is not expected to have any directional effect on the hybrids. Figure 3.3*c* also shows that for the set of parameter values presented here, in the "selection before migration" model, without the immediate effect of s_D on LD through phenogenotypic association newly formed during migration, sexual selection can still maintain about half of the positive correlation between cultural trait divergence and LD found previously.

Discussion

Although previous study find that the effect of sexual selection on a cultural trait is masked by oblique learning (Yeh and Servedio 2015), the model presented here shows that genetic loci associated with the cultural trait can still be affected by sexual selection (as well as viability selection) through phenogenotypic associations. Even though the cultural trait is not vertically inherited, it is still an effective barrier against hybridization. More broadly this suggests that evolutionary biologists should not neglect traits that lack heritability, because these traits may still form association with other traits and pass the effect of selection on them to the associated traits. Similarly, although geography appears to be the main driver for both genetic and linguistic divergence in human populations, with little to no direct interaction between the two (Rosser *et al.* 2000; Creanza *et al.* 2015), the correlation between gene and culture alone may allow selection on one to affect the other. A potential example of selection on culture affecting the genetic traits is the recent gerrymandering in North Carolina, USA, which disproportionally affect black voters (a genetic trait), even though the lawmakers drew the voting

districts by voting behaviour (a cultural trait, although see (Charney 2008)) (Supreme Court of the United States 2017).

In this chapter I demonstrate that correlation between gene and culture is not a sufficient proof that birdsong dialect prevents hybridization, because such correlation can form easily when both diverged as a result of spatial structure. Such explanation has been used to explain the genetic and song divergence found in Timberline Wren (Camacho 2017). Conversely, I also demonstrate that not finding any correlation between gene and culture does not disprove the possibility that birdsong dialects serve as a barrier to gene flow, because oblique learning removes the correlation at the local scale. This phenomenon potentially explains why birdsong does not predict genotype in the sympatric area of several warbler hybrid zones despite sharp clines in both gene and culture (Kenyon *et al.* 2011, 2016). Furthermore, during cultural trait turnover, multiple cultural traits can coexist in a given patch, all associated with the same genetic alleles, while still limiting hybridization. Together, these results suggest that correlational studies are insufficient for understand the role of obliquely learned trait in preventing gene flow. Empiricists working on this question should instead measure reproductive barriers more directly by observing how the learned traits affect survival and mating success across populations, which can be done with playback experiments (Lipshutz *et al.* 2017). It is also important to measure migration rate, as it affects the level of phenogenotypic association.

Drift is not found to have any significant contribution to speciation with gene flow in this model. Cultural drift can make empirical research on birdsong more difficult by allowing cultural trait turnover and increasing stochastic noise, whereas genetic drift leads to the loss of polymorphism. Drift in allopatry however can be an important for establishing the initial genetic divergence, which is important in this model since the genes are not locally adapted.

The model assumes conservatively that all locally adaptive cultural alleles have the same fitness; therefore cultural trait turnover only occurs through drift. If the novel songs are instead more adaptive than the old song, perhaps because the population is adapting to a new or changing environment, such as increasing traffic noise (Luther and Derryberry 2012), it would lead to a higher rate of cultural trait turnover, which would still maintain LD when the turn-over is complete, and reduce the time in which LD is reduced because of multiple coexisting alleles.

Oblique learning is assumed to be unbiased in this model; the probability for an individual to learn a song is equal to the frequency of that song in the population. It is unclear whether this assumption is reasonable due to a lack of data. What is known is that many birds preferentially learn from neighbouring rivals (Beecher *et al.* 1994) and counter-singing adults (Templeton *et al.* 2010). If individuals are more inclined to learn a song that is locally more common (*i.e.*, there is positive frequency dependent "fecundity" selection on the song), this should increase cultural trait divergence. That increased cultural trait divergence, as described before, would then lead to stronger phenogenotypic associations, passing even more effect of sexual and divergent selection onto the genetic loci, although the magnitude of such effect might be small. Such a learning inclination would also select against novel songs, making it more difficult for cultural trait turnover to occur through drift.

The results of this study should be robust against mechanisms that may lead to an association between a specific cultural trait and the hybrids (3-way phenogenotypic association), such as asymmetric mating preferences between subpopulations, which has been found in warblers (Secondi *et al.* 2003) and fruit flies (Kaneshiro 1976). One may argue that such a mechanism would challenge the finding that s_H does not affect C because the cultural trait of the population that is more likely to hybridize would be indirectly selected against. However, if

selection against hybrids is strong enough to impact evolution through a 3-way association, pre-mating isolation should evolve through reinforcement more quickly because it occurs more easily through 2-way associations. Indeed, in shifting hybrid zones, the species with more introgression is usually the expanding species, suggesting that the cost of hybridization is not very high (Dasmahapatra *et al.* 2002; Secondi *et al.* 2003).

While I examined the effect of swapping migration and viability selection in the life cycle, all versions of the model assume that learning occurs right after reproduction, in the same patch as the parents. This assumption should be valid for species that only learn during a short sensitive period in early life, such as marsh wren (Kroodsma 1978), zebra finches (Zann 1990) and song sparrows (Marler and Peters 1987), but not for species with open-ended learning, such as northern mockingbirds (Howard 1974) and canaries (Nottebohm and Nottebohm 1978). If learning mainly occurs after migration instead of before, such as for maintaining territory by matching the local song (Verner 1975), phenogenotypic association would not form.

The model presented here focuses on how phenogenotypic association can form through migration and spatial constraint to learning. There are of course other mechanisms that may lead to phenogenotypic associations. It may be fruitful to systematically examine all mechanisms known to maintain LD to assess whether a parallel mechanism for phenogenotypic association exists and is of importance to the study of speciation and gene-culture coevolution.

Figures

Figure 3.1: Factors affecting LD. (a) LD is maximized at intermediate levels of α, and positively correlates with s_D.(m=0.01, $s_H = 0.09$, no mutation, no drift) **(b)** LD positively correlates with s_H, and a minimal level of s_H is needed to maintain LD. (m=0.01, α=2.07, no mutation, no drift). **(c)** The polymorphism in the genetic loci is sometimes lost, leading to LD=0. (m=0.01; s_H=0.1; μ=10^{-3}; α=1; N=100; 3 cultural alleles) **(c) (d)** Where polymorphism is maintained, there is a positive correlation between LD and cultural distance. The black x's indicate simulations in which C_3 has replaced C_1 in patch 1 (defined as $\frac{p_{C_3,1}}{p_{C_1,1}} > 10 \cap p_{C_3,1} > 0.5$). (Panel d parameters same as panel c, except N=1000).

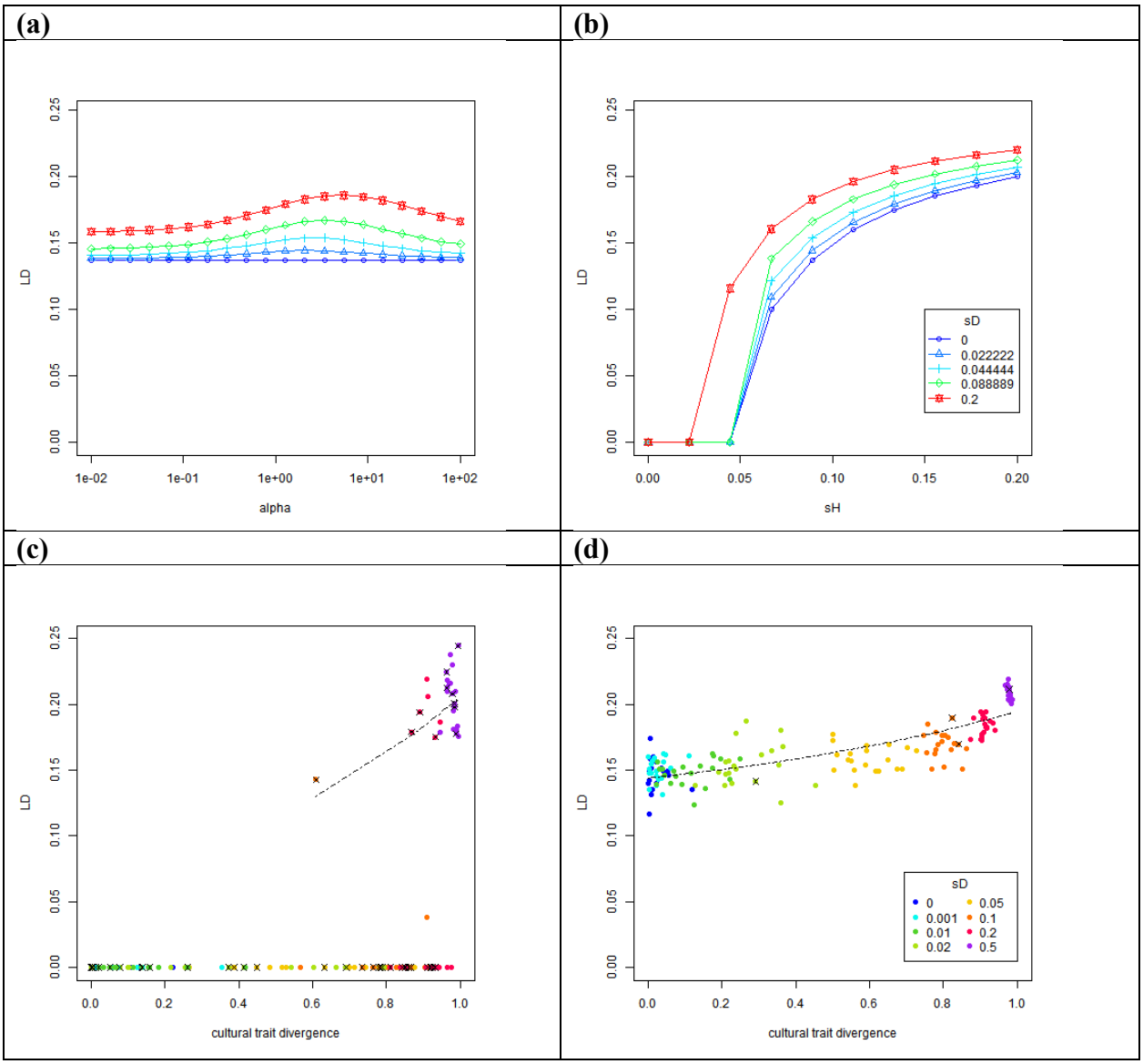
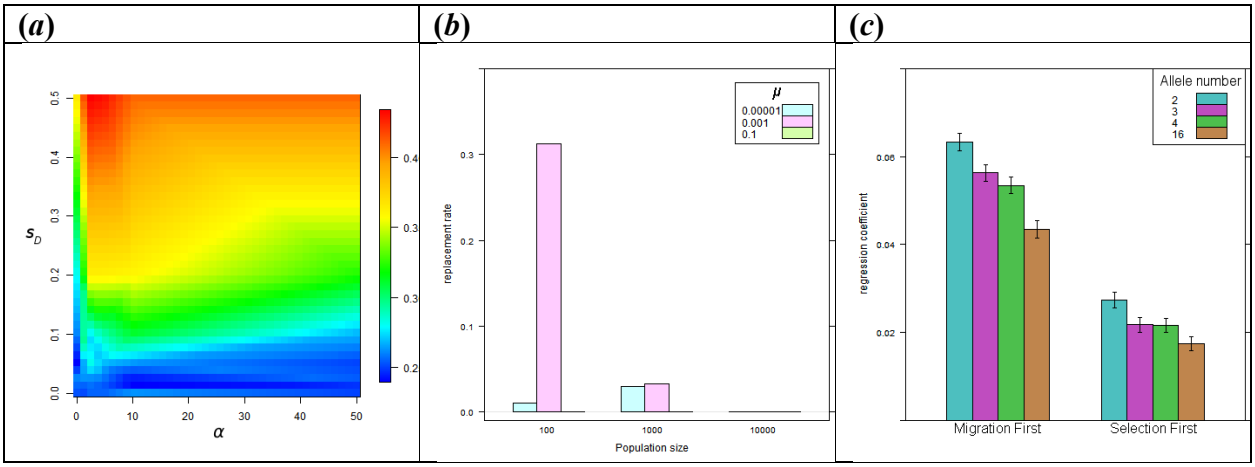


Figure 3.2: Maintenance of polymorphism at the genetic loci, and cultural trait turnover. **(*a*)** How α and s_D affect genetic polymorphism across both patches. Colour indicates the proportion of simulation output (20 runs for each parameter, using data from 2000 and 4000 generations) in which the polymorphism is maintained (m=0.01; s_H=0.1; N=100; μ=0.001; 2 cultural alleles; migration first) **(*b*)** The rate of cultural trait turnover (m=0.01; s_H=0.1; s_D=0-0.5; α=1; 3 cultural alleles; migration first). **(*c*)** The values of the exponential regression coefficient of LD on cultural trait divergence (all coefficients p<0.001; error bars show 95% CI; m=0.01; s_H=0.1; s_D=0-0.5; α=1).



OVERALL CONCLUSIONS

Social learning is common across the animal kingdom, and it even exists in non-animal species such as slime molds (Vogel and Dussutour 2016). It is possible that the course of biological evolution has been heavily influenced by cultural evolution and perhaps even more mechanisms of information transmission (Laland *et al.* 2015). The results of my dissertation show that not only microevolution, but also macroevolution may be affected by learning. Understanding learning and cultural evolution in various species is therefore important for evolutionary biology.

It many systems, the underlying genetic architecture for mating signals, preferences, and learning are not known. My models assumed the most simplistic case, but it is very likely that more complex architectures exist, at least in some of the taxa. It is well known that genetic architecture affects evolution. In additive traits each locus is under weaker selection and there are more chance for recombination to disrupt divergence, which can make speciation more difficult (Barton 1983; Gourbiere 2004). On the other hand, in additive traits generalist-specialist dynamics may lead to evolutionary branching (Dieckmann and Doebeli 1999; Polechová and Barton 2005). Empirical data on these genetic architectures and models that compare different architectures are therefore important. Similarly, it is possible that cultural traits across different domains may have different underlying architecture (*e.g.*, discrete religious institutions versus continuous political spectra), and this something that needs to be further studied in order to understand how to properly compare biological evolution to cultural evolution.

The extent to which fecundity selection and directional mutation play a role in cultural evolution is currently a widely discussed topic (Acerbi and Mesoudi 2015). It has been documented that there are biases in birdsong learning and human language acquisition due to both physical constraints and cognitive predispositions (Doupe and Kuhl 1999). Furthermore there can be direction mutation in cultural transmission (Cardoso and Atwell 2011), and the fidelity of birdsong learning (directly related to mutation rate) varies across species (Catchpole and Slater 2008). Directional mutation is not expected to alter any of the qualitative results in my models. However, if selection is primarily due to variation in transmission rates as opposed to the survival or the mating success, the outcome of my third model may be drastically different.

The models I studied in this dissertation deal with how one genepool becomes two and how populations remain distinct upon secondary contact. Similar processes may occur in cultural group formation, cultural integration, and radicalization, which are very relevant to the current affairs. While there is no direct analog of reproductive isolation in cultural evolution, it may be possible to borrow other concepts from speciation to further our understanding of human cultures. For example, reduction in recombination rate, built up of LD, and reduced fitness of intermediates may also occur in cultural evolution. Identifying these processes and expanding speciation models may contribute to the understanding and management of abovementioned issues.

APPENDIX A: CHAPTER 1 SUPPLEMENTARY MATERIALS

Supplementary Material S1: Comparison of Divergence Equilibria

Here we examine whether, at the divergence equilibria, $\hat{t}_{1A}$ for paternal learning is larger than $\hat{t}_{1A}$ for genetic inheritance, that is, whether

$$\frac{1}{2}\left(1+\frac{1}{1-2m}\sqrt{\frac{2m\alpha^2+10m\alpha+8m-\alpha}{\alpha(2m\alpha+2m-1)}}\right)<\frac{1}{2}\left(1+\frac{1}{1-2m}\sqrt{\frac{m\alpha^2+6m\alpha+4m-\alpha}{\alpha(m\alpha+2m-1)}}\right) \quad \text{(S1.1)}$$

holds for all *m* and *α*. Everything outside of the square root can be canceled out, and we are only interested in real solutions. Therefore, Expression (*S1.1)* is equivalent to

$$\frac{2m\alpha^2+10m\alpha+8m-\alpha}{\alpha(2m\alpha+2m-1)}<\frac{m\alpha^2+6m\alpha+4m-\alpha}{\alpha(m\alpha+2m-1)} \quad \text{(S1.2)}$$

Subtracting 1 from both sides, we get

$$\frac{8m(\alpha+1)}{2m\alpha+2m-1}<\frac{4m(\alpha+1)}{m\alpha+2m-1} \quad \text{(S1.3)}$$

Dividing by $8m(\alpha+1)$ and further simplifying we find that

$$2m\alpha+2m-1>2m\alpha+4m-2 \quad \text{(S1.4)}$$

and

$$m < \frac{1}{2} \quad \text{(S1.3)}$$

This is true by the definition of the migration rate, therefore paternal learning always gives higher divergence than genetic inheritance.

Supplementary Material S2: Evolution of α

We add in a locus C with two alleles, C_1 and C_2, which code for two different levels of assortative mating preference strength, α_1 and α_2 respectively. The life cycle is identical to that of the basic model described in the main text. The mating table for population A is therefore

Female \ Male	C_1T_1	C_1T_2	C_2T_1	C_2T_2
C_1T_1	$\frac{{x_{1A}}^2(1+\alpha_1)}{1+\alpha_1 t_1}$	$\frac{x_{1A}x_{2A}}{1+\alpha_1 t1}$	$\frac{x_{1A}x_{3A}(1+\alpha_1)}{1+\alpha_1 t1}$	$\frac{x_{1A}x_{4A}}{1+\alpha_1 t1}$
C_1T_2	$\frac{x_{1A}x_{2A}}{1+\alpha_1 t_2}$	$\frac{{x_{2A}}^2(1+\alpha_1)}{1+\alpha_1 t_2}$	$\frac{x_{2A}x_{3A}}{1+\alpha_1 t_2}$	$\frac{x_{2A}x_{4A}(1+\alpha_1)}{1+\alpha_1 t_2}$
C_2T_1	$\frac{x_{3A}x_{1A}(1+\alpha_2)}{1+\alpha_2 t_1}$	$\frac{x_{3A}x_{2A}}{1+\alpha_2 t_1}$	$\frac{{x_{3A}}^2(1+\alpha_2)}{1+\alpha_2 t_1}$	$\frac{x_{3A}x_{4A}}{1+\alpha_2 t_1}$
C_2T_2	$\frac{x_{4A}x_{1A}}{1+\alpha_2 t_2}$	$\frac{x_{4A}x_{2A}(1+\alpha_2)}{1+\alpha_2 t_2}$	$\frac{x_{4A}x_{3A}}{1+\alpha_2 t_2}$	$\frac{{x_{4A}}^2(1+\alpha_2)}{1+\alpha_2 t_2}$

where $x_{1A}, x_{2A}, x_{3A}, x_{4A}$ are genotypic frequencies of $C_1T_1, C_1T_2, C_2T_1, C_2T_2$, respectively, in population A after migration.

After paternal learning of trait T and genetic inheritance of locus C, we can obtain 8 recursion equations for the 4 genotypes in 2 populations. Because allele frequencies add up to 1

we can substitute in

$$\begin{cases} x_{1A} = 1 - x_{2A} - x_{3A} - x_{4A} \\ x_{1B} = 1 - x_{2B} - x_{3B} - x_{4B} \end{cases} \quad \textit{(S2.1)},$$

This allows us to remove the redundant equations for C_1T_1 in both populations. Taking partial derivatives on the remaining 6 recursion equations allows us to construct a 6×6 Jacobian matrix:

$$\mathbf{J} = \begin{bmatrix} \frac{\partial}{\partial x_{2A}} x'_{2A} & \frac{\partial}{\partial x_{2B}} x'_{2A} & \frac{\partial}{\partial x_{3A}} x'_{2A} & \frac{\partial}{\partial x_{3B}} x'_{2A} & \frac{\partial}{\partial x_{4A}} x'_{2A} & \frac{\partial}{\partial x_{4B}} x'_{2A} \\ \frac{\partial}{\partial x_{2A}} x'_{2B} & \frac{\partial}{\partial x_{2B}} x'_{2B} & \frac{\partial}{\partial x_{3A}} x'_{2B} & \frac{\partial}{\partial x_{3B}} x'_{2B} & \frac{\partial}{\partial x_{4A}} x'_{2B} & \frac{\partial}{\partial x_{4B}} x'_{2B} \\ \frac{\partial}{\partial x_{2A}} x'_{3A} & \frac{\partial}{\partial x_{2B}} x'_{3A} & \frac{\partial}{\partial x_{3A}} x'_{3A} & \frac{\partial}{\partial x_{3B}} x'_{3A} & \frac{\partial}{\partial x_{4A}} x'_{3A} & \frac{\partial}{\partial x_{4B}} x'_{3A} \\ \frac{\partial}{\partial x_{2A}} x'_{3B} & \frac{\partial}{\partial x_{2B}} x'_{3B} & \frac{\partial}{\partial x_{3A}} x'_{3B} & \frac{\partial}{\partial x_{3B}} x'_{3B} & \frac{\partial}{\partial x_{4A}} x'_{3B} & \frac{\partial}{\partial x_{4B}} x'_{3B} \\ \frac{\partial}{\partial x_{2A}} x'_{4A} & \frac{\partial}{\partial x_{2B}} x'_{4A} & \frac{\partial}{\partial x_{3A}} x'_{4A} & \frac{\partial}{\partial x_{3B}} x'_{4A} & \frac{\partial}{\partial x_{4A}} x'_{4A} & \frac{\partial}{\partial x_{4B}} x'_{4A} \\ \frac{\partial}{\partial x_{2A}} x'_{4B} & \frac{\partial}{\partial x_{2B}} x'_{4B} & \frac{\partial}{\partial x_{3A}} x'_{4B} & \frac{\partial}{\partial x_{3B}} x'_{4B} & \frac{\partial}{\partial x_{4A}} x'_{4B} & \frac{\partial}{\partial x_{4B}} x'_{4B} \end{bmatrix}$$

To see if α_{opt} is an ESS, we conduct an invasion analysis where C_1 is set so that $\alpha_1 = \alpha_{\text{opt}}$ and is fixed in both populations, and let C_2, which encodes an arbitrary $\alpha_2 = \alpha_{\text{mut}}$ that differs from α_{opt}, invade. We set the frequencies of T according to the divergent equilibrium obtained from the basic model (see Table 1.2, main text). The resulting Jacobian matrix takes the form

$$\mathbf{J} = \begin{bmatrix} \mathbf{J_{res}} & \mathbf{V} \\ \mathbf{0} & \mathbf{J_{mut}} \end{bmatrix} \quad \textit{(S2.3)}$$

Where $\mathbf{J_{res}}$ is a 2x2 matrix and $\mathbf{J_{mut}}$ is a 4x4 matrix (see Otto and Day 2007). The submatrix $\mathbf{0}$ exists in the bottom left-hand corner because the corresponding recursion equations reduce to zero when C_2 is absent. To show that α_{opt} is an ESS and cannot be invaded by any other level of assortative mating, the absolute value of the leading eigenvalues of $\mathbf{J_{res}}$ and $\mathbf{J_{mut}}$, which are real (see below), must be smaller than 1.

The two eigenvalues of $\mathbf{J_{res}}$ can be found analytically and can easily be seen to be real, but are too complicated to show here (see *Mathematica* notebook on Dryad). Using the Reduce function of *Mathematica*, we find the two eigenvalues are between -1 and 1 if

$$0 < m < \tfrac{1}{12}(5 - \sqrt{17}) \approx 0.073 \qquad \textit{(S2.4)}$$

The four eigenvalues of matrix $\mathbf{J_{mut}}$ are too complicated to solve analytically. We will use the Perron-Frobenius theorem, partly evaluated by a graphical analysis, to ensure the leading eigenvalue is real and positive, and then use a modification of Routh-Hurwitz condition to show that the leading eigenvalue is below 1. Together, these would show the absolute value of the leading eigenvalues is below 1, a sufficient condition for the matrix to be stable. To apply the Perron-Frobenius theorem, we start by checking if the matrix is positive. Because some elements in $\mathbf{J_{mut}}$ are similar or identical, the matrix can be rewritten as

$$\mathbf{J_{mut}} = \begin{bmatrix} (1-m)a & ma & (1-m)c & mc \\ mb & (1-m)b & md & (1-m)d \\ (1-m)d & md & (1-m)b & mb \\ mc & (1-m)c & ma & (1-m)a \end{bmatrix} \qquad \textit{(S2.5),}$$

where

$$a=\frac{k^6(\alpha_{\mathrm{mut}}+1)+k^4\left((n+2)\alpha_{\mathrm{mut}}+n+3\right)+k^2(2n\alpha_{\mathrm{mut}}+n+2)+n\alpha_{\mathrm{mut}}}{k^2(k^2+2)(k^2(\alpha_{\mathrm{mut}}+2)+n\alpha_{\mathrm{mut}})} \quad \textit{(S2.6a)}$$

$$b=\frac{-k^6(\alpha_{\mathrm{mut}}+1)+k^4\left((n-2)\alpha_{\mathrm{mut}}+n-3\right)+k^2(2n\alpha_{\mathrm{mut}}+n-2)+n\alpha_{\mathrm{mut}}}{k^2(k^2+2)(k^2(\alpha_{\mathrm{mut}}+2)+n\alpha_{\mathrm{mut}})} \quad \textit{(S2.6b)}$$

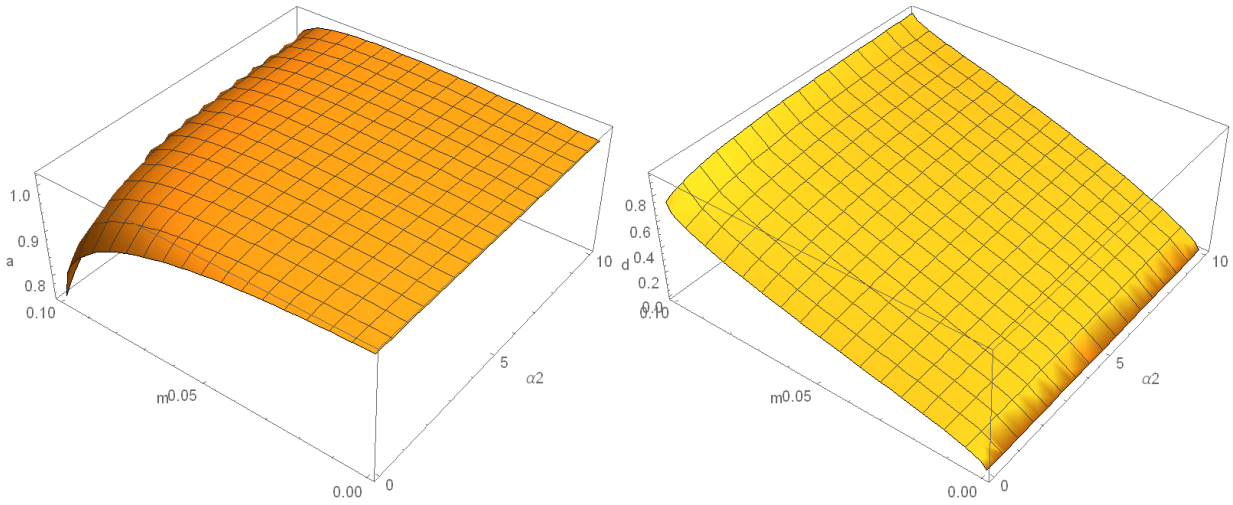
$$c=-\frac{n+k^2}{2(k^2(\alpha_{\mathrm{mut}}+2)+n\alpha_{\mathrm{mut}})} \quad \textit{(S2.6c)}$$

$$d=-\frac{n-k^2}{2(k^2(\alpha_{\mathrm{mut}}+2)+n\alpha_{\mathrm{mut}})} \quad \textit{(S2.6d)}$$

$$k=\sqrt{\sqrt{\frac{1}{m}-1}-1} \quad \textit{(S2.6e)}$$

$$n=\sqrt{k^4-4} \quad \textit{(S2.5f)}$$

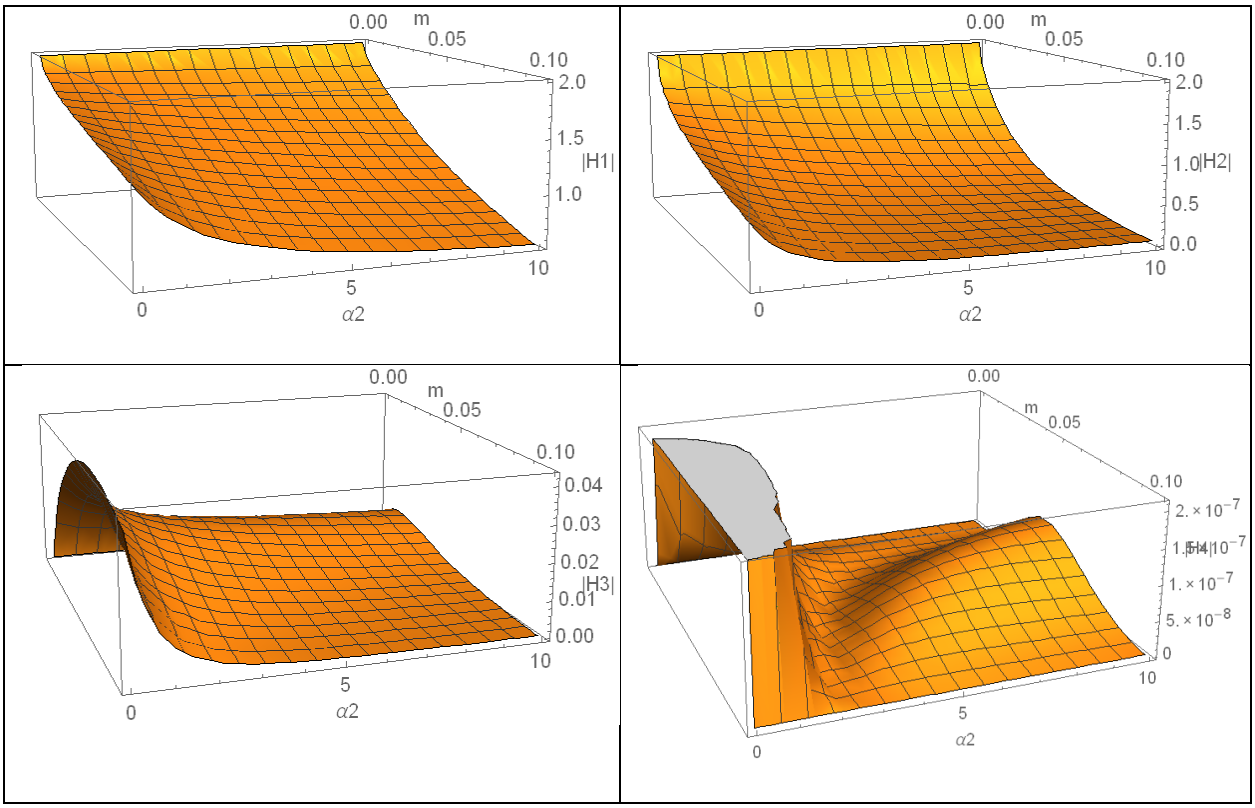
Plotting *a* and *b* shows they are positive when $m<1/10$, which is consistent with Equation 8 of the main text:



and using the Reduce function of *Mathematica*, c and *d* are also found to be positive when $m<1/10$. When all elements are positive, the Perron-Frobenius theorem states that the leading eigenvalue must be real and positive.

To show that the eigenvalues are smaller than 1 using the Routh-Hurwitz condition (again with a graphical analysis), we substitute in $\lambda=r+1$ and use the characteristic polynomial of *r* to

build Hurwitz matrices (see Otto and Day 2007). If the Routh-Hurwitz condition are met, this guarantees $r<0$ and thus $\lambda<1$. Plotting the determinants of the four Hurwitz matrices, we find that all 4 determinants are positive, thus satisfies the Routh-Hurwitz condition.



Supplementary Material S3: Evolutionary Invasion Analysis for the Spread of a Novel Trait in Population B

The Jacobian Matrices for paternal learning, $\mathrm{J_p}(t_{1A}, t_{2B})$, and genetic inheritance, $\mathrm{J_g}(t_{1A}, t_{2B})$, are obtained by taking partial derivatives of the recursion equations. After substituting $t_{1A} = 1$ and $t_{2B} = 0$, we obtain

$$\mathrm{J_p}(1,0) = \frac{1}{(1+\alpha)(1+s)}\begin{bmatrix} 1-m & -m(1+s)^2 \\ -m & (1-m)(1+s)^2 \end{bmatrix} \qquad (S3.1)$$

$$\mathrm{J_g}(1,0) = \frac{\alpha+2}{2}\mathrm{J_p}(1,0) \qquad (S3.2)$$

Expression (*S3.2)* includes the scalar because for genetic inheritance the male half of the population is under sexual selection, which gives $\frac{\alpha}{2}$, and female half of the population is not, which gives $\frac{2}{2}$. We can then calculate the respective leading Eigenvalues:

$$\lambda_{\mathrm{p}}(1,0) = \frac{(1-m)\big(2+s(2+s)\big)+\sqrt{s^2(2+s)^2(1-m)^2+4m^2(1+s(2+s))}}{2(1+s)(1+\alpha)} \qquad (S3.3)$$

$$\lambda_{\mathrm{g}}(1,0) = \frac{\alpha+2}{2}\lambda_{\mathrm{p}}(1,0) \qquad (S3.4)$$

The conditions for λ>1 describe when T_{2B} can increase its frequency and thus invade the population (to establish in population B). For paternal learning this condition is met when divergent selection is

$$s_p > \frac{\alpha^2+2m\alpha+\sqrt{\alpha(\alpha+2)\big(4m(1-m)+\alpha(\alpha+2)\big)}}{2(1-m)(1+\alpha)} \qquad (S3.5)$$

and for genetic inheritance, when divergent selection is

$$s_g > \frac{\alpha^2+2m\alpha(\alpha+2)+\sqrt{\alpha(3\alpha+4)\left(4m(1-m)(\alpha+2)^2+\alpha(3\alpha+4)\right)}}{4(1-m)(\alpha^2+3\alpha+2)} \qquad (S3.6)$$

The difference between the two minimum viability selections is

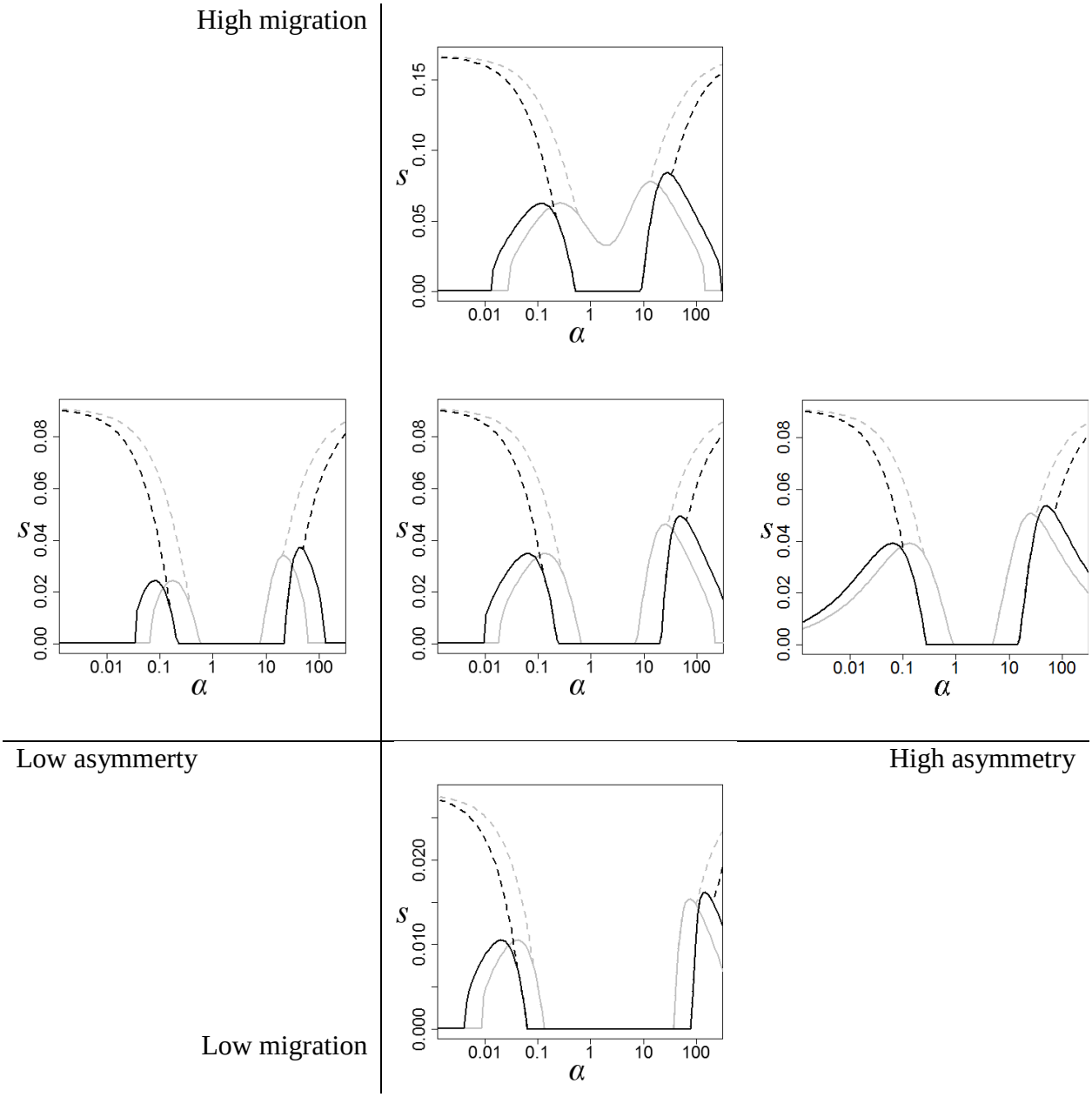
$$d = \frac{\sqrt{\alpha(4+3\alpha)\big(4(1-m)+\alpha(3-2m)\big)\big(\alpha+2m(2+\alpha)\big)}+2(2+\alpha)\sqrt{\alpha(2+\alpha)\big(4(1-m)m+\alpha(2+\alpha)\big)}+\alpha(2m(2+\alpha)+\alpha(3+2\alpha))}{4(1-m)(1+\alpha)(2+\alpha)}$$

(S3.7)

This is always positive because all the terms are positive. Furthermore, it is also positive after taking partial derivative with respect to α or m, meaning stronger assortative mating or migration intensifies the difference between paternal learning and genetic inheritance.

Supplementary Figure S1

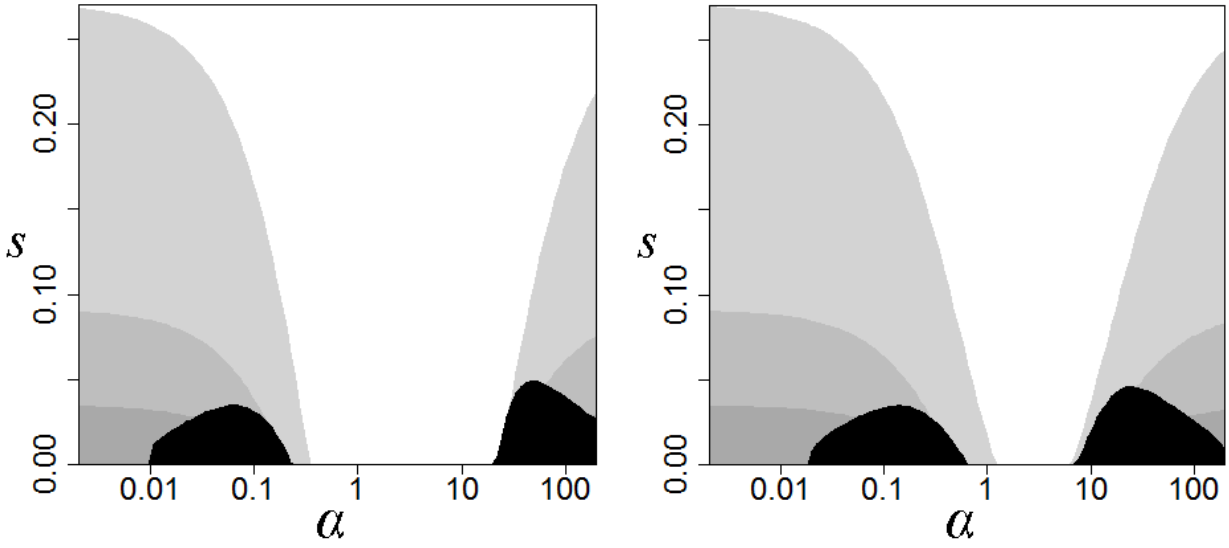
Minimum selection required to maintain divergence (dotted lines) and polymorphism (solid lines), under different migration rates and initial asymmetries. Black: paternal learning; Grey: genetic inheritance; Divergence threshold = 0.75; m=0.01, 0.03, 0.05; Asymmetries $(t_{1A} - t_{2B})$ = 2^{-2}, 2^{-8}, 2^{-14}; The dotted lines are not visible when overlapping with the solid lines in the middle. Note that vertical axes differ in scale.



Supplementary Figure S2

Level of divergence under (*a*) paternal learning and (*b*) genetic inheritance. Black: loss of polymorphism; Dark grey: low divergence (<0.625); Grey: low divergence (0.625-0.75); Light grey: medium divergence (0.75-0.875); White: high divergence (>0.875). (m=0.03, asymmetry = 2^{-8})

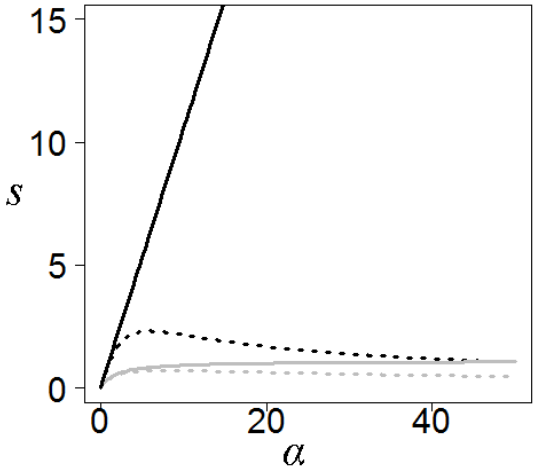
(*a*) (*b*)



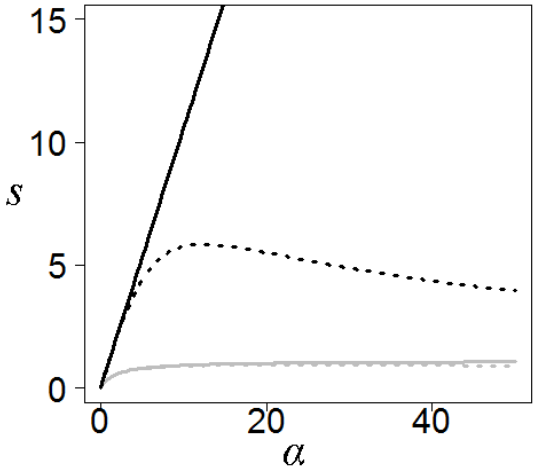
Supplementary Figure S3

Figure S3. Minimum selection required for invasion of T_2 in population B. Dotted lines shows the results from simulation. The exact sizes of invasion, defined by the initial t_{2B}, are *(a)* 10^{-2} *(b)* 10^{-3} and *(c)* 10^{-4}. The solid lines are prediction from invasion analysis, which assumes initial $t_{2B} = 0$. For all graphs, *m*=0.05. Black: paternal learning; Grey: genetic inheritance.

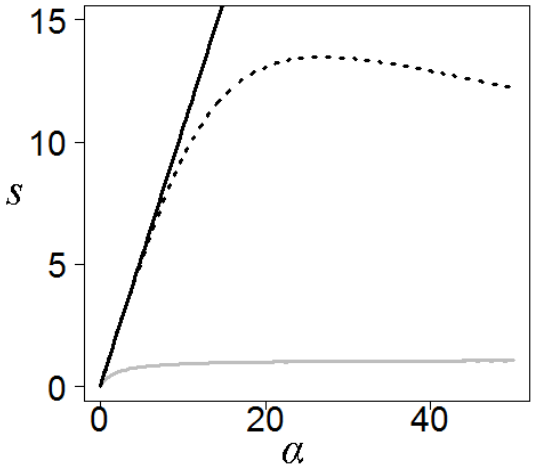
(a) 10^{-2}



(b) 10^{-3}



(c) 10^{-4}



APPENDIX B: CHAPTER 2 SUPPLEMENTARY MATERIALS

Supplementary Material S1: Variations of the model

Skewness of divergent selection

We add a parameter, γ, to allow viability to decrease faster when moving away from the locally adaptive trait value ($0 \leq \gamma \leq 1$, where 0 is most skewed and 1 is not skewed). Table S1 shows the viability table after including the γ parameter.

We did not find any trends that cannot be explain by the mechanisms described in the main text. High skewness (low γ) increases divergent selection, which help maintain polymorphism, leading to lower κ. High skewness also reduces the opportunity of disassortative pairing by reducing the amount of surviving hybrids, thus lowering σ. Finally, because low γ select against hybrids but not immigrant purebreds, high skewness actually help maintain the presence of a small maladaptive immigrant population, leading to a low α.

Maternal imprinting

In general there are only minor differences between the results of our model assuming paternal versus maternal imprinting (Figures S1-1 and S1-2, compare 1st and 2nd column). All mechanisms described in the main text that affect the maintenance of polymorphism and the evolution of paternal imprinting still hold for maternal imprinting. Regarding maintaining polymorphism (Figure S1-1), because an imprinted preference is formed based on paternal/maternal trait frequencies, it is at least as diverged as the trait (more so with paternal imprinting, see Tramm and Servedio 2008), thus imprinting can maintain polymorphism by favoring the local trait (avoiding the loss of polymorphism through the homogenization that can occur at the preference locus, Servedio and Bürger 2014). For the stable strategy of imprinting weight (Figure S1-2), both maternal and paternal imprinting increase assortative mating and

reduce hybridization, therefore in both models imprinting evolves. The only difference between paternal and maternal imprinting arises when hybridization occurs, *i.e.*, when the father and the mother carry different traits. This causes the level of assortative mating created through imprinting to be slightly different between maternal and paternal imprinting. The difference is a consequence of assortative and disassortative mating that arises in this haploid system. First consider matings between maladapted immigrant females and locally adapted males. Given that hybrid offspring that survive will always be more likely to be the ones that carry the locally adapted trait, paternal imprinting in this pairing, which results in a preference for the father's locally adapted trait, will lead to assortative mating. In the opposite case, the hybridization between a locally adapted female and maladaptive immigrant males, surviving offspring that imprint on their father are in contrast likely to mate disassortatively (they are likely to be locally adapted but paternal imprinting will cause them to develop a preference for maladapted males). The former case occurs more frequently, because it is easier for immigrant female to encounter a local male than for a local female to encounter an immigrant male. Therefore, paternal imprinting creates stronger assortative mating than maternal imprinting (since with maternal imprinting the outcomes of assortative versus disassortative mating in the two scenarios are swapped, making disassortative mating slightly more common). Regardless, as seen in Figures S1-2 the difference between the two imprinting modes is small.

Sexual dimorphism

In many species sexually selected traits are sexually dimorphic. Therefore we briefly consider the case where females do not express the trait (*i.e.* females do not undergo viability selection on the trait). Removing selection on females increases the number of immigrant females still present in the population after the viability selection step of the life cycle. As

explained in the main text, these females can get a "good-genes" benefit by hybridizing with the locally adapted males. This effect counters the effect of reinforcement and causes the ESS for imprinting to be lower than in the sexually monomorphic model (Figure S1-2, compare 1st and 3rd column). Polymorphism is also more easily lost with sexual dimorphism (Figure S1-1, compare 1st and 3rd columns), because the preference alleles homogenize across the populations more easily when immigrant females are not selected against (see the comparison of these cases in Servedio and Bürger 2014); homogenization of the preference can lead to the loss of trait polymorphism, as explained in the main text.

Table S1: The viability table after including a parameter to account for the skewness of divergent selection.

Genotype (Ta,Tb,Tc)	**Trait Phenotype (T)**	**Relative Viability in Patch 0**	**Relative Viability in Patch 1**
(0,0,0)	0	$(1 + s_H)(1 + 3s_D)$	$1 + s_H$
(0,0,1), (0,1,0), or (1,0,0)	1	$1 + 2\gamma\, s_D$	$1 + \gamma\, s_D$
(0,1,1), (1,0,1), or (1,1,0)	2	$1 + \gamma\, s_D$	$1 + 2\gamma\, s_D$
(1,1,1)	3	$1 + s_H$	$(1 + s_H)(1 + 3s_D)$

Figure S1-1: Polymorphism threshold (κ) under maternal imprinting, or with sexual dimorphism (females not under any viability selection). Note the x and y axes are not on the same scale. (c=0, $p_{g,h}^{\mathrm{inn}}$ is skewed).

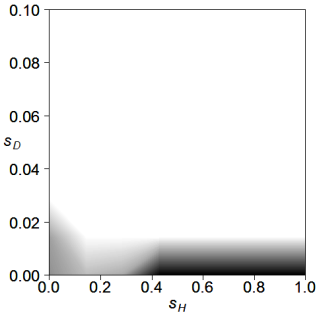
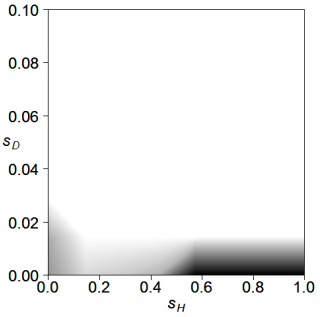
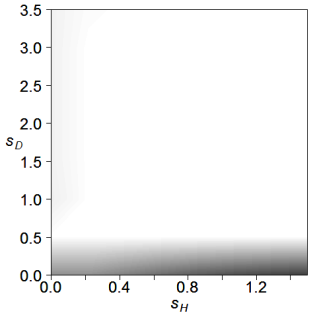
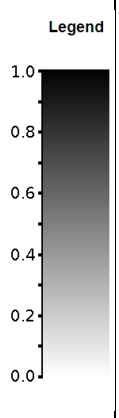
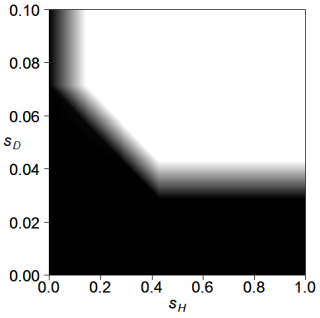
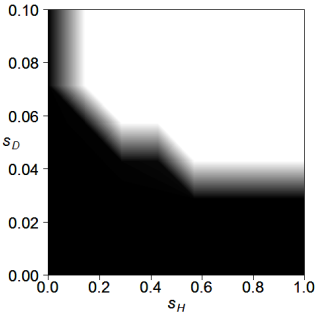
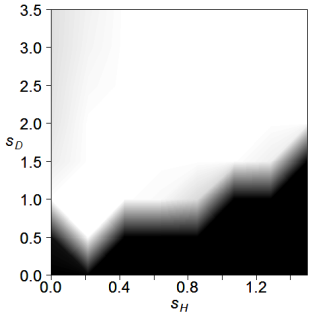
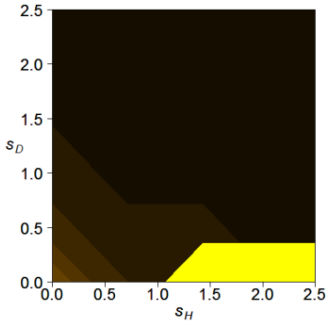
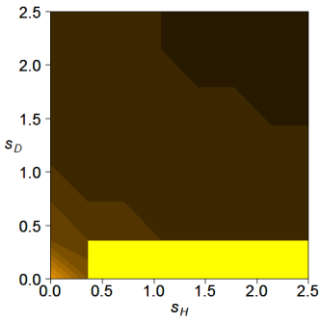
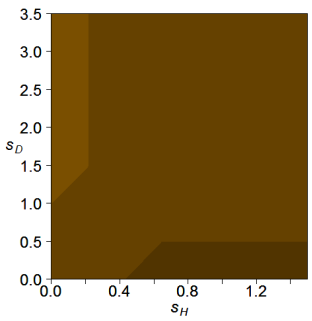
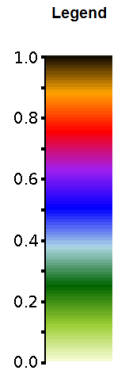
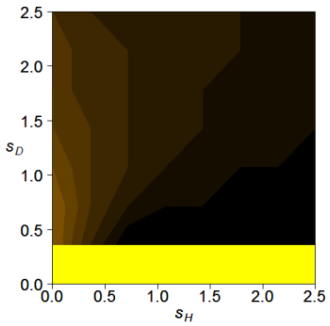
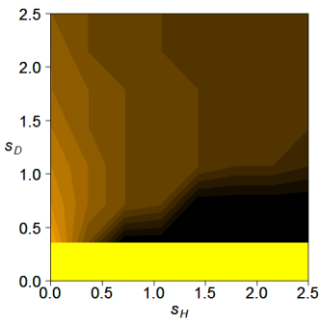
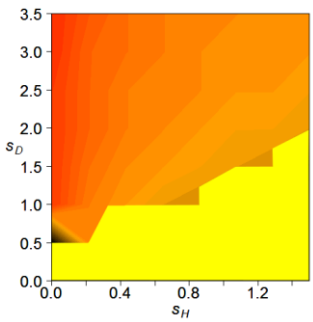
m	Paternal Imprinting & Sexual Monomorphism (the model in main text)	Maternal Imprinting	Sexual Dimorphism	
0.01				
0.1				

Figure S1-2: Stable strategy of imprinting weight (σ) under maternal imprinting, or with sexual dimorphism (females not under any viability selection). Yellow region indicate stable strategy masked by loss of polymorphism ($\kappa>\sigma$). Note the x and y axes are not on the same scale. (c=0, $p_{g,h}^{\mathrm{inn}}$ is skewed).

m	Paternal Imprinting & Sexual Monomorphism (the model in main text)	Maternal Imprinting	Sexual Dimorphism	
0.01				
0.1				

Supplementary Material S2: Inferring the evolution of imprinting weight without simulating successive invasions

The results from our main simulation in which we reset the allele frequency for each invasion can be used to infer the results for successive invasion. The inference is made by assuming that the imprinting weights that the population will evolve to are the ones that can invade others, but cannot be invaded by others, both from a higher and a lower value of λ, in the

simulation described in the main text. One potential objection to this assumption is that, because in the simulation all λ values the invasions restart from the initial secondary-contact frequencies at the preference and trait loci, the starting conditions for each new invasion of λ the P and T loci is not an equilibrium, thus the population might not really be able to evolve as we infer through successive invasion and substitution. For this reason, the inference above is only valid when the following two conditions are met.

The first condition that must be met for this inference to be valid is that the population starts above the polymorphism threshold κ. If we start from below κ, the initial variation will be lost in critical components of the system in the very early stages of imprinting evolution, and there is no way to rescue polymorphism because there is positive frequency-dependent sexual selection. As such, imprinting will not evolve.

The second condition that must be met is that $\sigma>\kappa$. This is because the symmetry in the frequencies of P and T across the two patches increase over time when the polymorphism is maintained in successive invasion, making polymorphism easier to maintain under successive invasion (*i.e.*, κ itself will shift lower if it is not already at 0). When $\sigma<\kappa$, imprinting weight will evolve to a value lower than the κ value found in our main simulation, reaching a σ value that is unobservable in the main simulation due to loss of polymorphism.

The abovementioned dynamics is confirmed for a subset of parameter values by running simulations where new invasions arise sequentially after the trait and preference alleles are allowed to reach equilibrium starting from the ending point of the previous simulation after the invading L allele is artificially fixed. More specifically, in this simulation, once it was determined that invasion is occurring, we reset the invading allele at the L locus to a frequency of 1 as the new resident allele, but left other alleles as they were, and ran 1000 generations to let

T and P reach the new equilibrium, before we introduced the next L allele. The equilibrium obtained in these simulations are identical to the stable strategy of the main simulation, except when the two conditions described above (starting above κ and $\sigma>\kappa$) are not met.

Although computational constraints prohibits us from running simulations with actual successive invasion, we believe the abovementioned inference is accurate for the following reasons:

1) Separation of time scale. T and P are found to be evolving at least 100 times faster than λ, so resetting the frequencies at the T and P loci for each invasion should not have any drastic effect.

2) For the purpose of arriving at the stable strategy for imprinting, restarting the frequency is more conservative than successive invasion, as it increases the chance of losing polymorphism at the trait loci. The frequency of T at the quasi-equilibrium is more symmetric across patches than at the starting frequencies which we use, so it is expected that if we use successive invasion it would be even less likely to lose polymorphism. This does not affect the evolution of imprinting except in cases in which κ were previously above σ and now have a lower κ that is below σ, allowing evolution to a low stable strategy without losing polymorphism at the trait.

3) Successive invasion would basically keep the population near the basin of the stable strategy that we found. The line of trait divergence in Figure 2.2b is very smooth and flat, even though the figure is obtained by restarting the frequencies for every λ to high divergence. In the simulation where we did not restart the allele frequencies for the T and P loci (*i.e.*, where a new equilibrium was found for a fixed L before introduction of the new L allele, so that the invasion

occurs while the population is near the line), the population remained on that line. The trait distributions did not change and the three trait loci always have the same allele frequencies when polymorphism is maintained (save for a small difference that is less than 0.01% and only occur in simulations that resets the allele frequencies for each invasion, likely caused by the tiny difference in initial allele frequencies).

4) Multiple different variations of the simulations were performed in the course of our analyses, and we never found any equilibrium for λ that is not already presented in this chapter. These additional variants included applying a small perturbation to the genotype frequencies every few generations, adding small arbitrary numbers to parameter values to avoid exact symmetry, using arbitrary starting frequencies (*e.g.* $p_0 = 0.9$ in both populations, T highly asymmetric), letting the original strategy re-invade after reaching equilibrium, etc.

Supplementary Material S3: The effect of different innate preference functions

Here we show how polymorphism maintenance and imprinting evolution is affected by different innate preference functions. The "skewed" innate preference in the top rows is the choosiest, such that imprinting cannot increase choosiness, whereas the "peak" innate preference in the bottom rows is the least choosy.

In Figure S3-1, we see that as the innate preference become choosier (towards the top rows), polymorphism become easier to maintain (lower κ, shown as lighter color), and there exists some complex interactions between the innate preference function and s_H.

In Figure S3-2, we see that as innate preference become less choosy (towards the bottom rows), reducing imprinting weight becomes more effective in helping immigrant females mate

with local males, thus lowering σ (as seen in the orange regions found in the bottom right panels).

Figure S3-1: The threshold κ, below which polymorphism at the trait loci is lost, and as a result imprinting cannot evolve. Polymorphism at the preference and trait loci is maintained above this threshold. In non-white regions (κ>0), innate preference alone cannot maintain polymorphism. In black regions (κ=1) polymorphism is always lost, and as a result imprinting cannot evolve. (Linear interpolation used for plotting, grid density: 8x8; c=0)

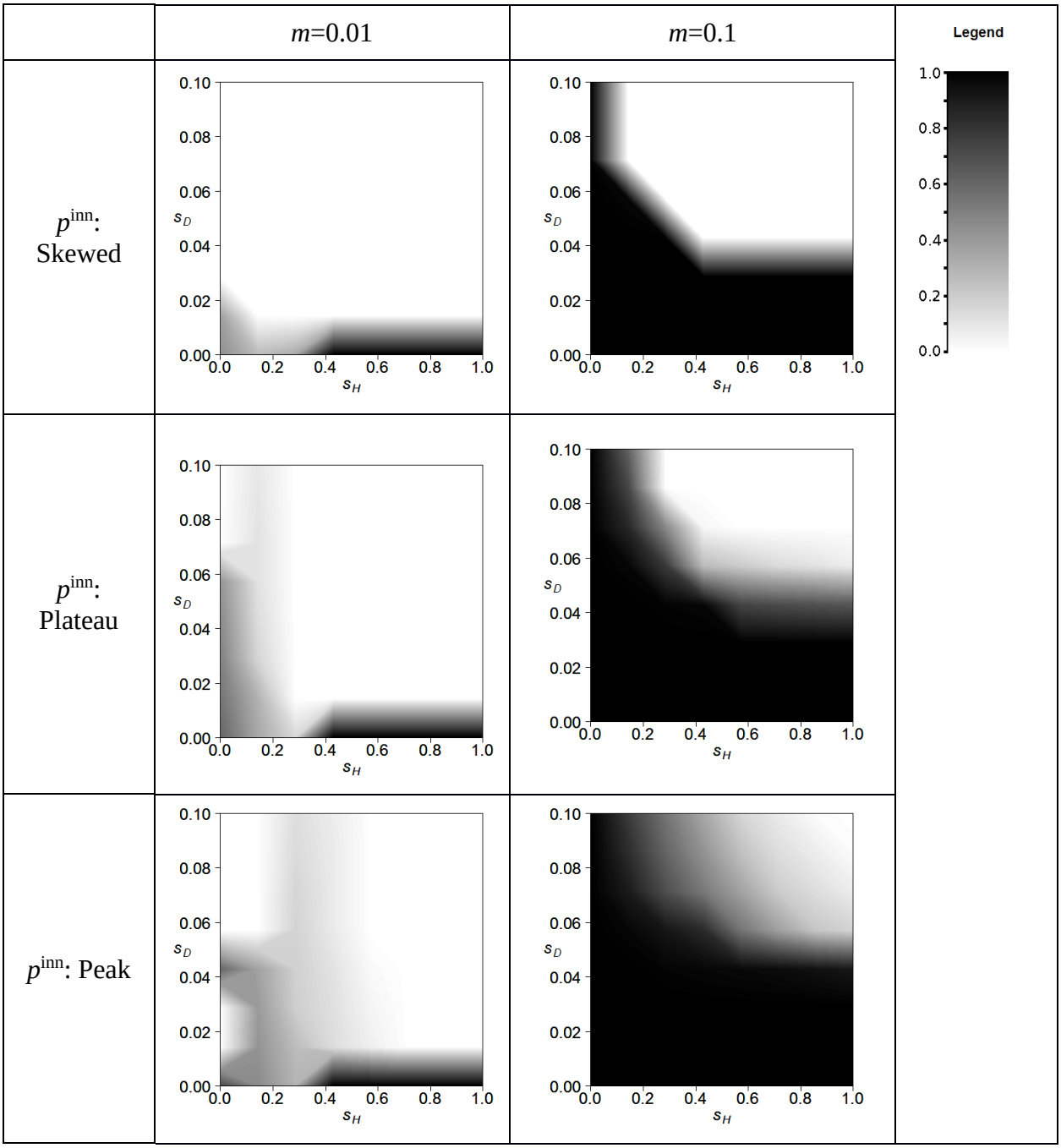
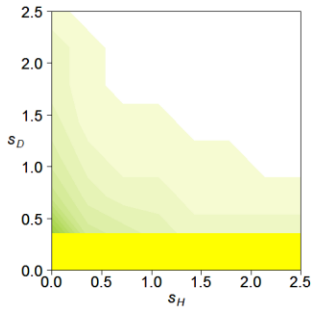
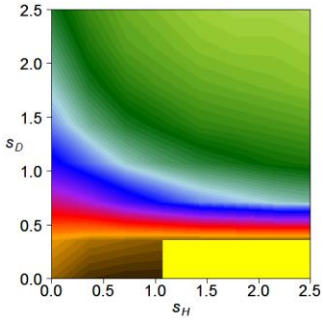
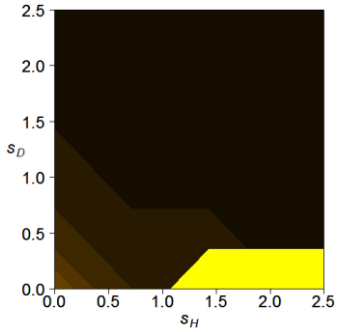
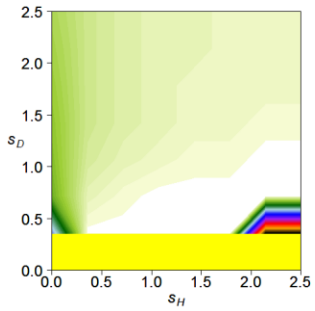
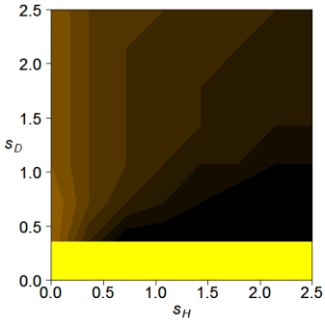
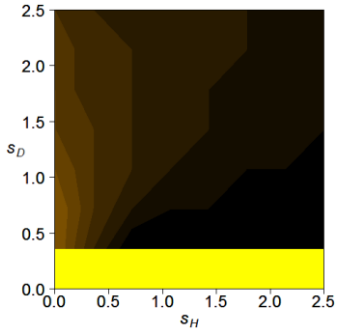
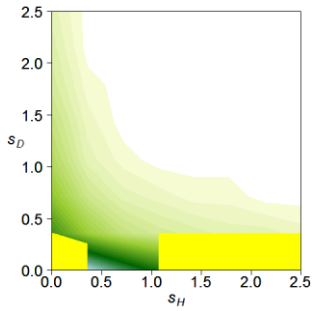
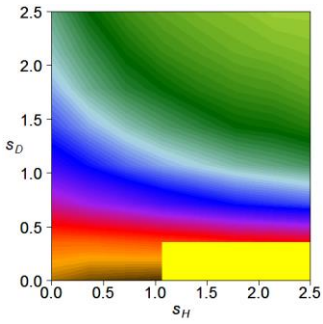
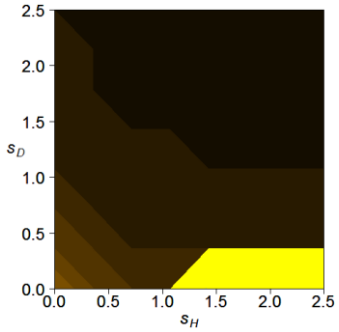
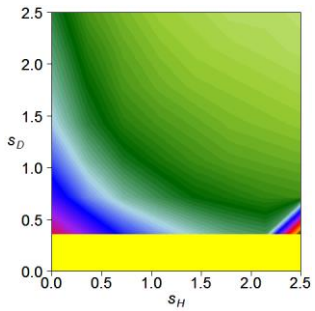
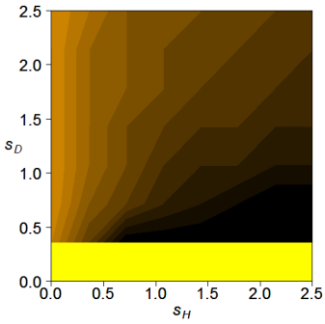
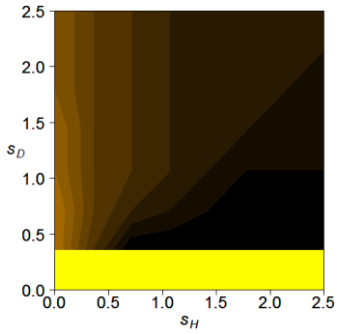
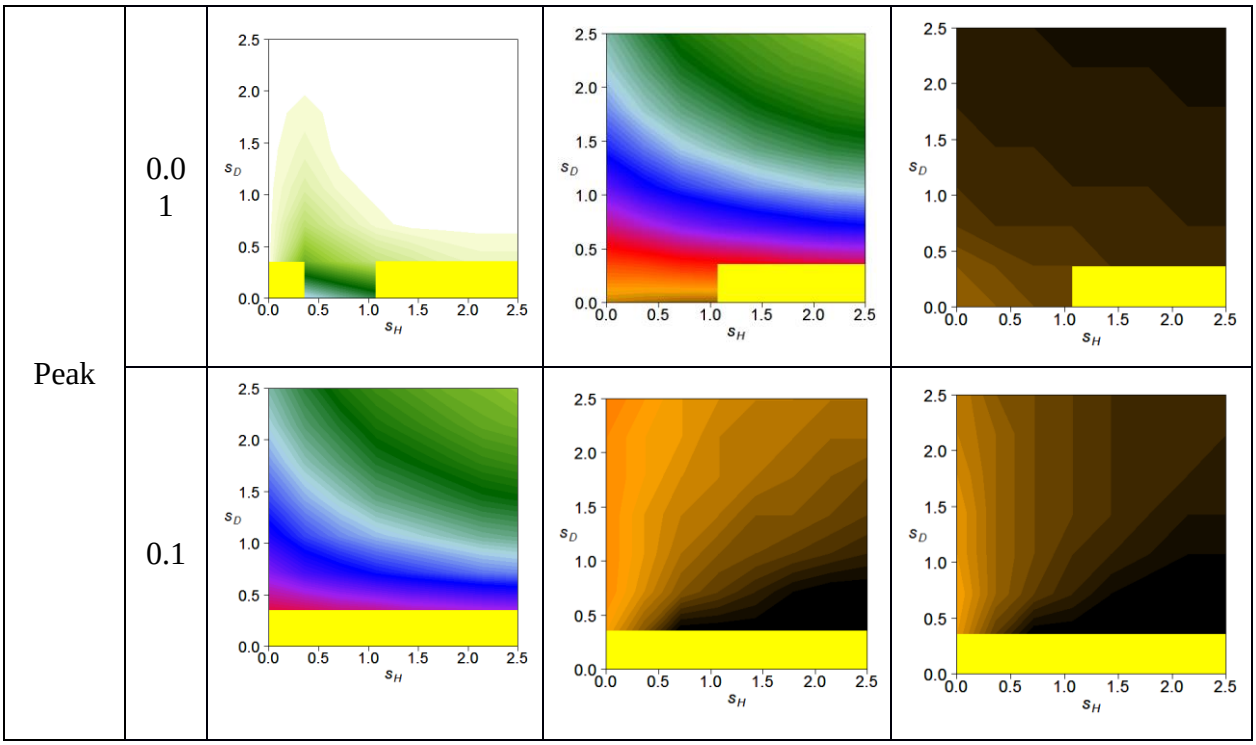
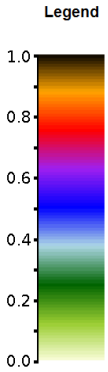


Figure S3-2: σ, the level of imprinting that evolves when polymorphism is maintained. In the yellow area, the threshold for maintaining polymorphism is above the stable strategy ($\sigma < \kappa$) given the starting conditioned described in the Simulation section. (Linear interpolation used for plotting, grid density: 8x8).

p^{inn}	m	c=10^{-4}	c=10^{-5}	c=0
Skewed	0.01			
	0.1			
Plateau	0.01			
	0.1			



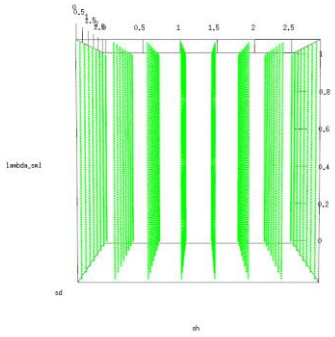
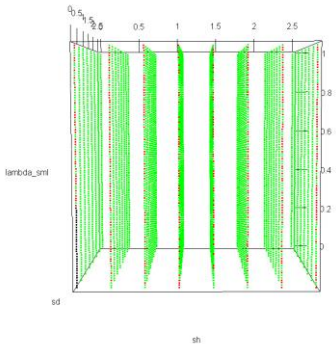
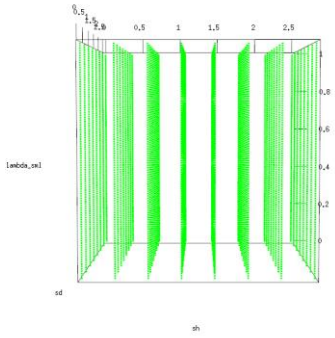
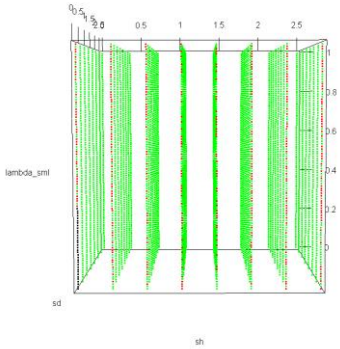
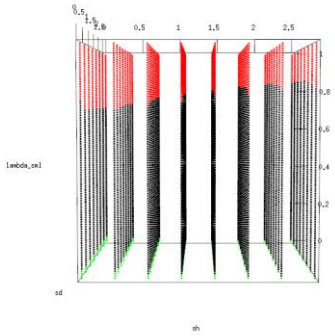
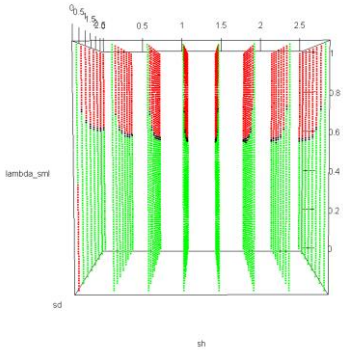


Supplementary Material S4. Evolution in allopatry before secondary contact

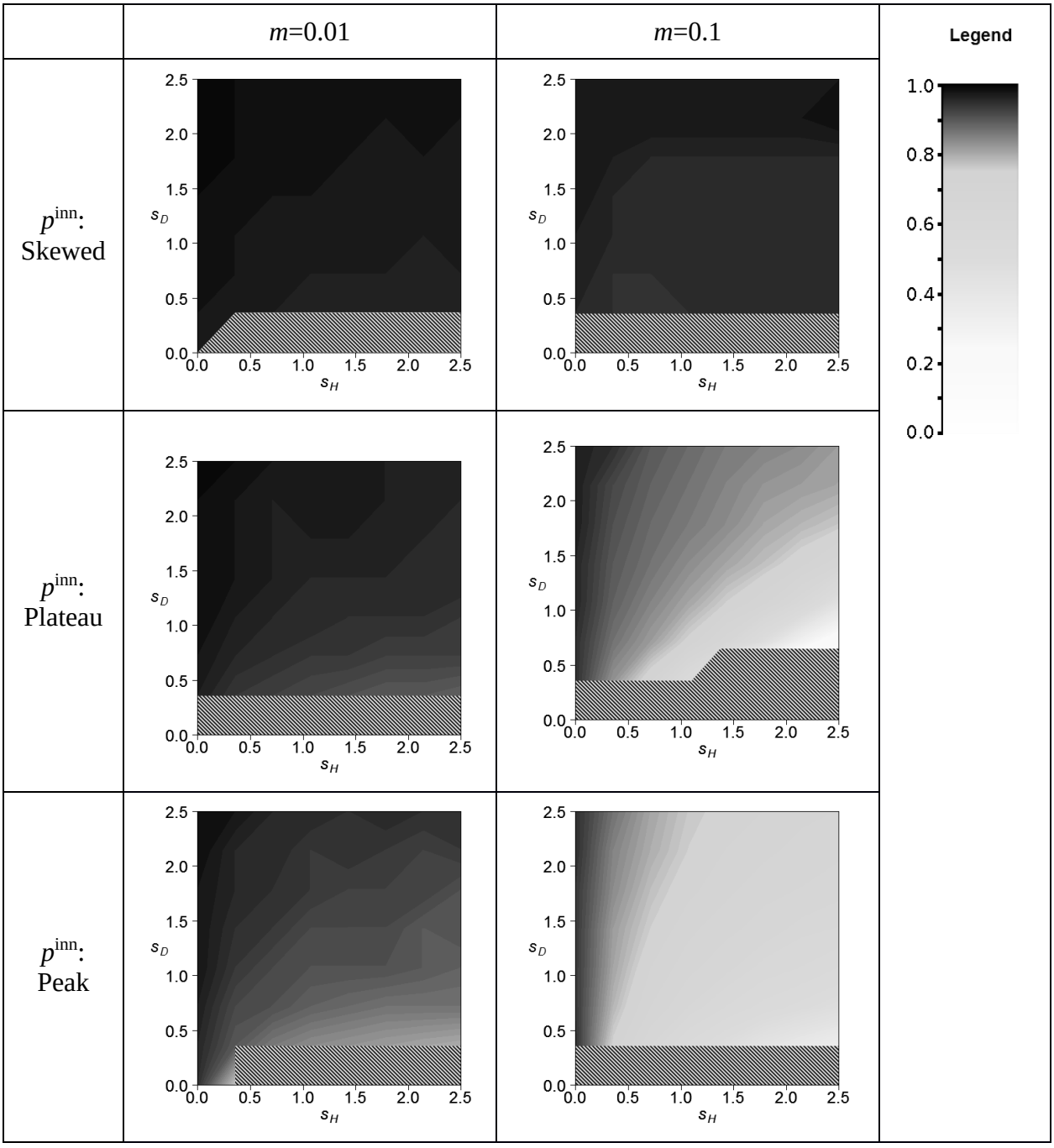
To give us better understanding of where the initial frequencies should be at the initiation of secondary contact, we run simulations to check how the population will evolve in allopatry, *i.e.*, with m=0. Results are summarized in Figure S3, which shows the change in allele frequency for each locus (indicated in the first column) for either the adaptive allele (T_0 and P_0 in patch 0) or allele for higher λ (L_1). With the exception of s_D=0 when starting with equal frequencies, T_0 always increase and reach fixation very quickly. The evolution of T is what drives the evolution

of the other loci. P increases with T, but do not reach fixation since it becomes neutral after T reaches fixation. Regardless of where it starts, the L locus is evolving towards an intermediate value of λ, but because L is evolving very slowly while T is fixing very quickly, the changes are not very noticeable (hence the black region in the figures, which represents a frequency change of less than 10^{-6}). If there is mutation to maintain polymorphism at T loci, we can infer that P and L would reach fixation eventually, although it would take at least several thousand generations.

Figure S4: X axis (left-right): s_H; Y axis (in-out): s_D; Z axis (up-down): λ. m=0, c=0. Innate preference=peak. Initial conditions: l_0=0.99, l_1=0.01, or the reverse; $\Delta\lambda$=0.01; initial frequency of T and P alleles shown in the top row. Green=increase; red=decrease; black=no observable change (precision: 10^{-6}).

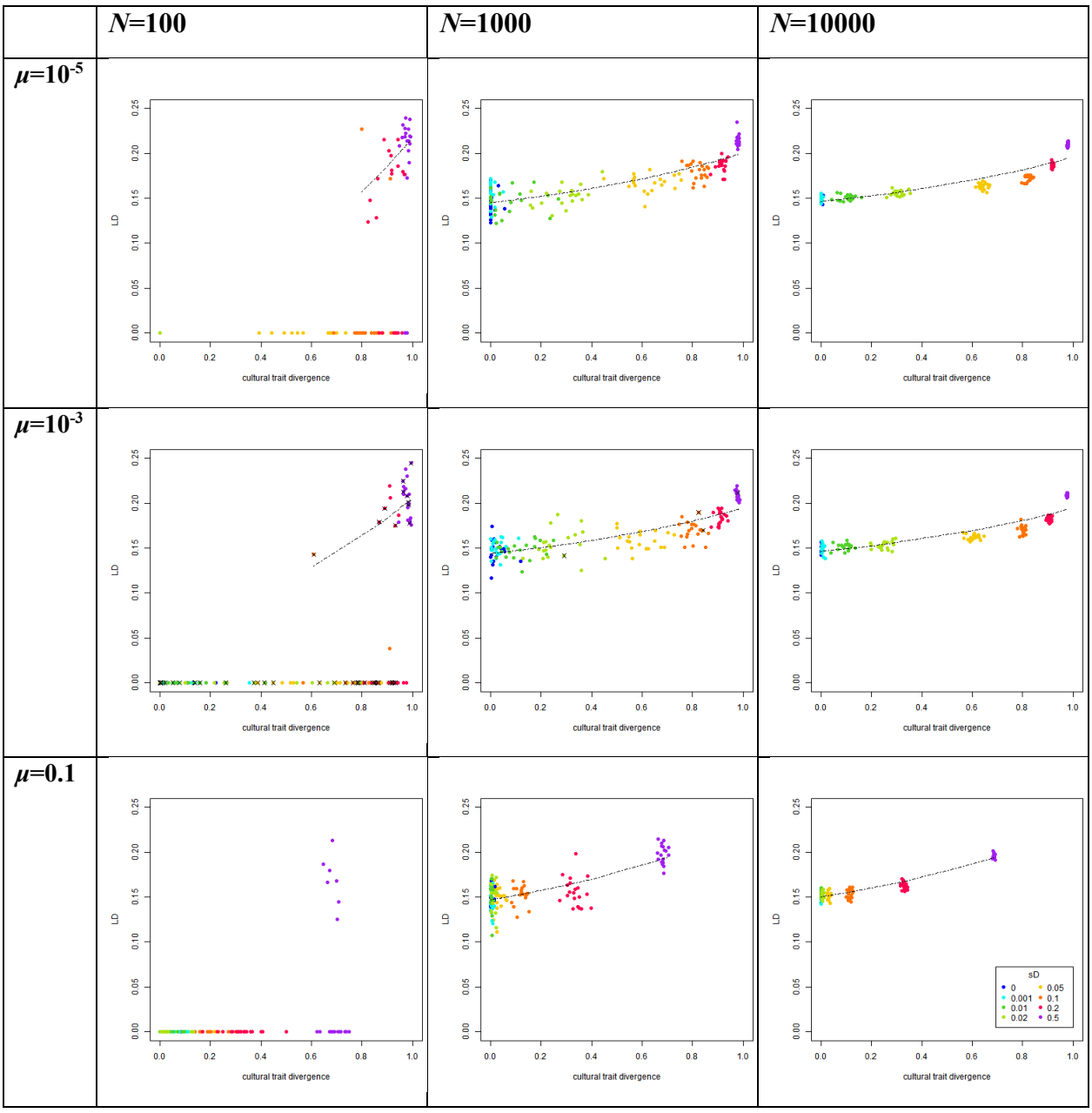
	Starting adapted (t_0=0.99, p_0=0.99)	**Starting in the middle (t_0=0.5, p_0=0.5)**
T (reach fixation in green area)		
P (frequency change: 0.0001 - 0.2)		
L (frequency change: ~0.00001 in green & red)		

Supplementary Figure S5: *α*, the level of imprinting that leads to maximum divergence of the trait. In the hatched area, the level of λ that would have maximized trait divergence is not sufficiently high to maintain polymorphism, therefore whenever polymorphism is maintained, increasing λ only reduces divergence. (c=0; linear interpolation used for plotting, grid density: 8x8)

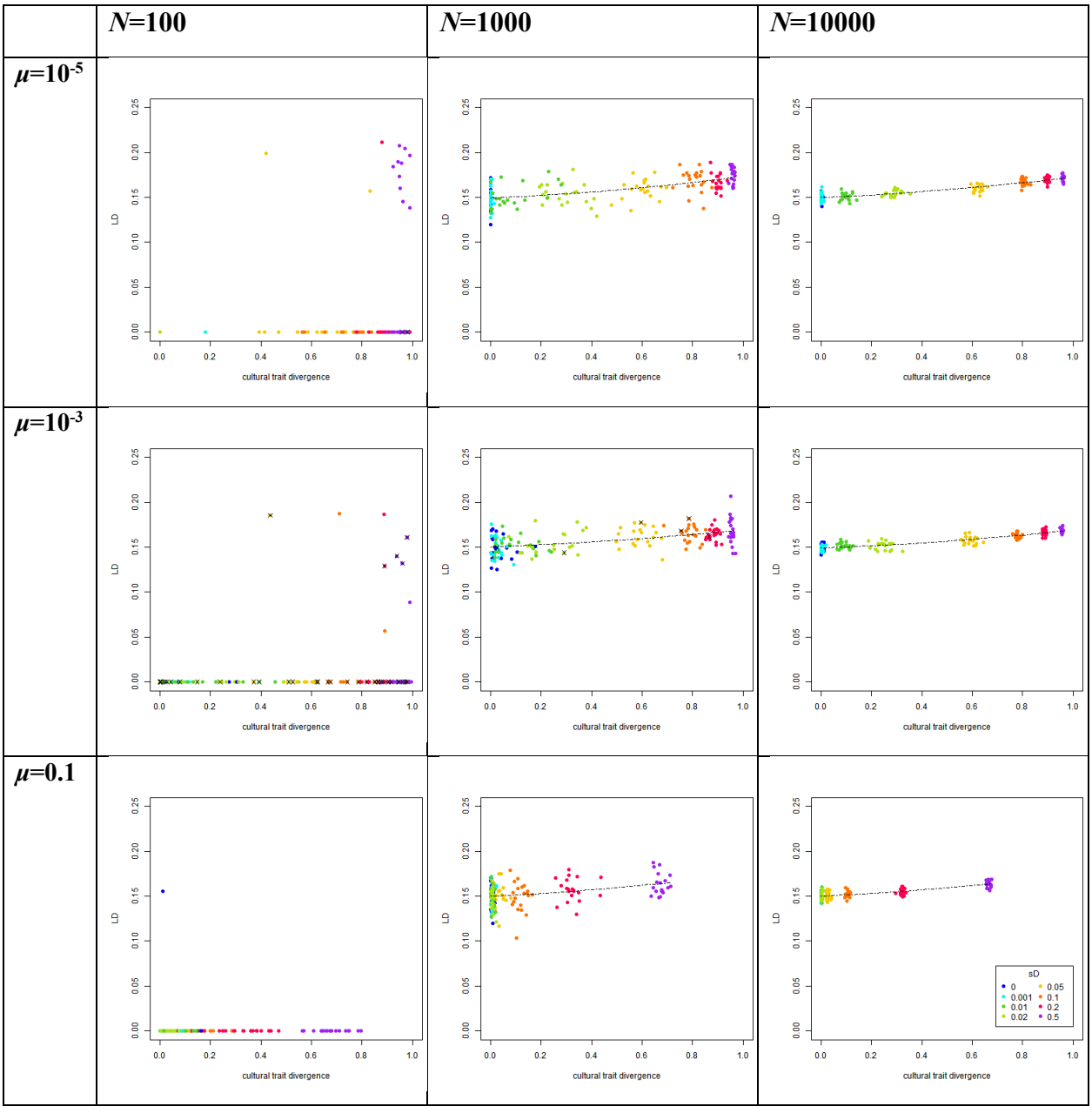


APPENDIX C: CHAPTER 3 SUPPLEMENTARY MATERIALS

Supplementary Figure S1. Cultural mutation (μ) lowers cultural trait divergence and LD. Drift only makes everything more diffused. Cultural trait divergence caused by drift does not increase LD. Strong drift leads to polymorphism loss, which is most common with low population size. The black x's indicate simulations in which C_3 is replacing C_1 in patch 1. This replacement does not affect LD, and is more common when drift is strong and cultural mutation rate is at intermediate values. (All panels m=0.01; α=1; s_H=0.1; s_D=0-0.5 indicated by colour; 3 cultural alleles)



Supplementary Figure S2: Selection before Migration. Compared with the previous graph, the positive correlation between cultural trait divergence and LD is weaker. This is because divergent selection on culture can no longer affects the genes immediately through the phenogenotypic association. Instead, the role of divergent selection is only to help maintain the across-patch phenogenotypic association by increasing cultural trait divergence; the positive correlation between cultural trait divergence and LD is now maintained mainly by sexual selection. (m=0.01; α=1; s_H=0.1; s_D=0-0.5 indicated by colour; 3 cultural alleles)



REFERENCES

Abbott, R., D. Albach, S. Ansell, J. W. Arntzen, S. J. E. Baird, N. Bierne, J. Boughman, A. Brelsford, C. A. Buerkle, R. Buggs, R. K. Butlin, U. Dieckmann, F. Eroukhmanoff, A. Grill, S. H. Cahan, J. S. Hermansen, G. Hewitt, A. G. Hudson, C. Jiggins, J. Jones, B. Keller, T. Marczewski, J. Mallet, P. Martinez-Rodriguez, M. Möst, S. Mullen, R. Nichols, A. W. Nolte, C. Parisod, K. Pfennig, A. M. Rice, M. G. Ritchie, B. Seifert, C. M. Smadja, R. Stelkens, J. M. Szymura, R. Väinölä, J. B. W. Wolf, and D. Zinner. 2013. Hybridization and speciation. J. Evol. Biol. 26:229–246.

Acerbi, A., and A. Mesoudi. 2015. If we are all cultural Darwinians what's the fuss about? Clarifying recent disagreements in the field of cultural evolution. Biol. Philos. 30:481–503. Springer Netherlands.

Alonzo, S. H. 2008. Female mate choice copying affects sexual selection in wild populations of the ocellated wrasse. Anim. Behav. 75:1715–1723.

Ancel, L. W. 2000. Undermining the Baldwin expediting effect: does phenotypic plasticity accelerate evolution? Theor. Popul. Biol. 58:307–19.

Aoki, K., and M. W. Feldman. 1987. Toward a theory for the evolution of cultural communication: coevolution of signal transmission and reception. Proc. Natl. Acad. Sci. U. S. A. 84:7164–8.

Bailey, N. W., and M. Zuk. 2009. Field crickets change mating preferences using remembered social information. Biol. Lett. 5:449–51.

Baker, M. 1975. Song dialects and genetic differences in White-crowned Sparrows (Zonotrichia leucophrys). Evolution 29:226–241.

Baker, M., and J. Boylan. 1999. Singing behavior, mating associations and reproductive success in a population of hybridizing Lazuli and Indigo Buntings. Condor 101:493–504.

Baker, M. C., T. K. Bjerke, H. U. Lampe, and Y. O. Espmark. 1987. Sexual response of female yellowhammers to differences in regional song dialects and repertoire sizes. Anim. Behav. 35:395–401.

Baker, M., and L. Mewaldt. 1978. Song dialects as barriers to dispersal in white-crowned Sparrows, Zonotrichia leucophrys nuttalli. Evolution 32:712–722.

Baker, M., K. Spitler-Nabors, and D. Bradley. 1981. Early experience determines song dialect responsiveness of female sparrows. Science (80-.). 214:819–821.

Baker, M., and D. Thompson. 1982. Allozyme frequencies in a linear series of song dialect populations. Evolution 36:1020–1029.

Balaban, E. 1988. Cultural and genetic variation in Swamp Sparrows (Melospiza georgiana): II.

Behavioral salience of geographic song variants. Behaviour 105:292–322.

Baptista, L. 1975. Song dialects and demes in sedentary populations of the white-crowned sparrow (Zonotrichia leucophrys nuttalli). Publ. Zool. Inst. Univ. Calif. 105:1–52. University of California Press, Berkeley.

Barbujani, G. 1997. DNA variation and language affinities. Am. J. Hum. Genet. 61:1011–1014.

Barton, N. H. 1983. Multilocus Clines. Evolution (N. Y). 37:454.

Bateson, P. 1978. Sexual imprinting and optimal outbreeding. Nature 273:659–660.

Bateson, P. P. 1966. The characteristics and context of imprinting. Biol. Rev. Camb. Philos. Soc. 41:177–211.

Beecher, M. D., S. E. Campbell, and P. K. Stoddard. 1994. Correlation of song learning and territory establishment strategies in the song sparrow. Proc. Natl. Acad. Sci. U. S. A. 91:1450–1454.

Beltman, J. B., and J. a J. Metz. 2005. Speciation: more likely through a genetic or through a learned habitat preference? Proc. Biol. Sci. 272:1455–63.

Böhner, J. 1983. Song learning in the Zebra Finch (Taeniopygia guttata): Selectivity in the choice of a tutor and accuracy of song copies. Anim. Behav. 31:231–237.

Boncoraglio, G., and N. Saino. 2007. Habitat structure and the evolution of bird song: a meta-analysis of the evidence for the acoustic adaptation hypothesis. Funct. Ecol. 21:134–142.

Borenstein, E., I. Meilijson, and E. Ruppin. 2006. The effect of phenotypic plasticity on evolution in multipeaked fitness landscapes. J. Evol. Biol. 19:1555–1570.

Boyd, R., and P. J. Richerson. 1985. An informal sketch of the dual inheritance model. Pp. 3–10 *in* Culture and the Evolutionary Process1. The University of Chicago Press, Chicago.

Camacho, A. 2017. Geographical barriers as drivers for song and genetic divergence. P. *in* Evolution 2017. Portland, OR.

Cardoso, G. C., and J. W. Atwell. 2011. Directional cultural change by modification and replacement of memes. Evolution (N. Y). 65:295–300.

Catchpole, C. K., and P. J. B. Slater. 2008. How accurately do they learn? Pp. 68–70 *in* Bird Song: Biological Themes and Variations. Cambridge University Press, Cambridge.

Cavalli-Sforza, L., and M. Feldman. 1981. Culture as the object of evolution. P. 10 *in* Cultural transmission and evolution: a quantitative approach. Princeton University Press, Princeton, NJ.

Cavalli-Sforza, L. L., and M. W. Feldman. 1983. Cultural versus genetic adaptation. Proc. Natl. Acad. Sci. U. S. A. 80:4993–6.

Chaffee, D. W., H. Griffin, and R. T. Gilman. 2013. Sexual imprinting: what strategies should we expect to see in nature? Evolution 67:3588–99.

Charney, E. 2008. Genes and Ideologies. Perspect. Polit. 6:299–319.

Clayton, N. 1987. Song tutor choice in zebra finches. Anim. Behav. 35:714–721.

Collins, S. 1994. Male displays: cause or effect of female preference? Anim. Behav. 48:371–375.

Coyne, J. A., and A. H. Orr. 2004. Behavioral and nonecological isolation. 1st ed. Sinauer Associates, Inc., Sunderland, MA.

Cramer, E. R. A., M. Ålund, S. E. McFarlane, A. Johnsen, and A. Qvarnström. 2016. Females discriminate against heterospecific sperm in a natural hybrid zone. Evolution (N. Y). 70:1844–1855.

Creanza, N., L. Fogarty, and M. W. Feldman. 2016. Cultural niche construction of repertoire size and learning strategies in songbirds. Evol. Ecol. 30:285–305. Springer International Publishing.

Creanza, N., M. Ruhlen, T. J. Pemberton, N. A. Rosenberg, M. W. Feldman, and S. Ramachandran. 2015. A comparison of worldwide phonemic and genetic variation in human populations. Proc. Natl. Acad. Sci. 112:1265–1272.

Dasmahapatra, K. K., M. J. Blum, A. Aiello, S. Hackwel, N. Davies, E. P. Bermingham, and J. Mallet. 2002. Inferences from a rapidly moving hybrid zone. Evolution (N. Y). 56:741–753.

Dieckmann, U., and M. Doebeli. 1999. On the origin of species by sympatric speciation. Nature 400:354–357.

Dobzhansky, T. 1940. Speciation as a stage in evolutionnary divergence. Am. Nat. 74:312–321.

Doupe, A. J., and P. K. Kuhl. 1999. BIRDSONG AND HUMAN SPEECH: Common Themes and Mechanisms. Annu. Rev. Neurosci. 22:567–631.

Dukas, R. 2006. Learning in the context of sexual behaviour in insects. Anim. Biol. 56:125–141.

Eales, L. 1985. Song learning in zebra finches: some effects of song model availability on what is learnt and when. Anim. Behav. 1293–1300.

Egger, B., K. Mattersdorfer, and K. M. Sefc. 2010. Variable discrimination and asymmetric preferences in laboratory tests of reproductive isolation between cichlid colour morphs. J. Evol. Biol. 23:433–439.

Egger, B., K. M. Sefc, L. Makasa, C. Sturmbauer, and W. Salzburger. 2012. Introgressive hybridization between color morphs in a population of cichlid fishes twelve years after human-induced secondary admixis. J. Hered. 103:515–522.

Ellers, J., and H. Slabbekoorn. 2003. Song divergence and male dispersal among bird

populations: a spatially explicit model testing the role of vocal learning. Anim. Behav. 65:671–681.

Everett, C., D. E. Blasi, and S. G. Roberts. 2015. Climate, vocal folds, and tonal languages: Connecting the physiological and geographic dots. Proc. Natl. Acad. Sci. 112:1322–1327.

Feldman, M., and K. Laland. 1996. Gene-culture coevolutionary theory. Trends Ecol. Evol.

Feldman, M. W., and L. a Zhivotovsky. 1992. Gene-culture coevolution: toward a general theory of vertical transmission. Proc. Natl. Acad. Sci. U. S. A. 89:11935–8.

Felsenstein, J. 1981. Skepticism towards Santa Rosalia, or why are there so few kinds of animals? Evolution 35:124–138.

Fisher, R. A. 1915. The evolution of sexual preference. Eugen. Rev. 7:184–192.

Galef jr, B., and D. White. 1998. Mate-choice copying in Japanese quail, Coturnix coturnix japonica. Anim. Behav. 55:545–52.

Garland, E. C., A. W. Goldizen, M. L. Rekdahl, R. Constantine, C. Garrigue, N. D. Hauser, M. M. Poole, J. Robbins, and M. J. Noad. 2011. Dynamic horizontal cultural transmission of humpback whale song at the ocean basin scale. Curr. Biol. 21:687–91. Elsevier Ltd.

Gavrilets, S. 2004. Fitness landscapes and the origin of species. Princeton University Press, Princeton, NJ.

Ghirlanda, S., and M. Enquist. 1999. The geometry of stimulus control. Anim. Behav. 58:695–706.

Gill, F., and B. Murray. 1972. Discrimination behavior and hybridization of the blue-winged and golden-winged warblers. Evolution 26:282–293.

Gilman, R. T., and G. M. Kozak. 2015. Learning to speciate: The biased learning of mate preferences promotes adaptive radiation. Evolution (N. Y). 69:3004–3012.

Gourbiere, S. 2004. How do natural and sexual selection contribute to sympatric speciation? J. Evol. Biol. 17:1297–1309.

Grant, B. R. 1984. The Significance of Song Variation in a Population of Darwin's Finches. Behaviour 89:90–116.

Grant, P., and B. Grant. 1997. Hybridization, sexual imprinting, and mate choice. Am. Nat. 149:1–28.

Hauber, M. E., and P. W. Sherman. 2001. Self-referent phenotype matching: theoretical considerations and empirical evidence. Trends Neurosci. 24:609–16.

Holsinger, K. E., M. W. Feldman, and F. B. Christiansen. 1984. The Evolution of Self-Fertilization in Plants : A Population Genetic Model. Am. Nat. 124:446–453.

Hoppitt, W., and K. Laland. 2013. Social learning: an introduction to mechanisms, methods, and models. Pp. 1–9 *in*. Princeton University Press, Princeton, NJ.

Howard, D. J. 1993. Reinforcement: origins, dynamics, and fate of an evolutionary hypothesis. Pp. 46–69 *in* R. G. Harrison, ed. Hybrid Zones and the Evolutionary Process. Oxford University Press, New York, NY.

Howard, R. D. 1974. the Influence of Sexual Selection and Interspecific. Evolution (N. Y). 28:428–438.

Immelmann, K. 1972. Sexual and Other Long-Term Aspects of Imprinting in Birds and Other Species. Adv. Study Behav. 4:147–174.

Immelmann, K., and S. J. Suomi. 1981. Sensitive phases in development. Pp. 395–431 *in* Behavioral development.

Inc., W. R. 2014. Mathematica. Version 10. Wolfram Research Inc., Champaign, Illinois, USA.

Irwin, D., and T. Price. 1999. Sexual imprinting, learning and speciation. Heredity (Edinb). 82:347–354.

Kaneshiro, K. 1976. Ethological Isolation and Phylogeny in the Planitibia Subgroup of Hawaiian Drosophila. Evolution (N. Y). 30:740–745.

Kelly, J. K., and M. A. F. Noor. 1996. Speciation by reinforcement: A model derived from studies of Drosophila. Genetics 143:1485–1497.

Kendrick, K. M., M. R. Hinton, K. Atkins, M. A. Haupt, and J. D. Skinner. 1998. Mothers determine sexual preferences. Nature 395:229–30.

Kenyon, H. L., M. Alcaide, D. P. L. Toews, and D. E. Irwin. 2016. Cultural isolation is greater than genetic isolation across an avian hybrid zone.

Kenyon, H. L., D. P. L. Toews, and D. E. Irwin. 2011. Can Song Discriminate between MacGillivray's and Mourning Warblers in a Narrow Hybrid Zone? Condor 113:655–663.

Kirkpatrick, M. 2000. Reinforcement and divergence under assortative mating. Proc. R. Soc. B Biol. Sci. 267:1649–1655.

Klein, N., and R. Payne. 1998. Evolutionary associations of brood parasitic finches (Vidua) and their host species: analyses of mitochondrial DNA restriction sites. Evolution 52:566–582.

Kopp, M., and J. Hermisson. 2008. Competitive speciation and costs of choosiness. J. Evol. Biol. 21:1005–23.

Kopp, M., R. J. Safran, M. R. Servedio, R. L. Rodr, T. C. Mendelson, M. E. Hauber, E. C. Scordato, C. N. Balakrishnan, L. B. Symes, D. M. Zonana, and G. S. Van Doorn. 2018. Mechanisms of assortative mating in speciation with gene flow: connecting theory and

empirical research. Am. Nat. 191:1–20.

Kozak, G. M., and J. W. Boughman. 2009. Learned conspecific mate preference in a species pair of sticklebacks. Behav. Ecol. 20:1282–1288.

Kozak, G. M., M. L. Head, and J. W. Boughman. 2011. Sexual imprinting on ecologically divergent traits leads to sexual isolation in sticklebacks. Proc. Biol. Sci. 278:2604–10.

Kroodsma, D. 1978. Aspects of learning in the ontogeny of bird song: where, from whom, when, how many, which, and how accurately? Pp. 215–230 *in* M. Burghardt, G. and Bekoff, ed. The development of behavior. Garland, New York, NY.

Lachlan, R. F., and M. R. Servedio. 2004. Song learning accelerates allopatric speciation. Evolution 58:2049–63.

Lachlan, R. F., and P. J. B. Slater. 2003. Song learning by chaffinches: how accurate, and from where? Anim. Behav. 65:957–969.

Lackey, A. C. R., and J. W. Boughman. 2013. Loss of sexual isolation in a hybridizing stickleback species pair. Curr. Zool. 59:591–603.

Laiolo, P., and J. L. Tella. 2005. Habitat fragmentation affects culture transmission: Patterns of song matching in Dupont's lark. J. Appl. Ecol. 42:1183–1193.

Laland, K. N., T. Uller, M. W. Feldman, K. Sterelny, G. B. Müller, A. Moczek, E. Jablonka, and J. Odling-Smee. 2015. The extended evolutionary synthesis : its structure , assumptions and predictions. Proc. R. Soc. B Biol. Sci. 282.

Liou, L. W., and T. D. Price. 1994. Speciation by Reinforcement of Premating Isolation. Evolution (N. Y). 48:1451–1459.

Lipshutz, S. E., I. A. Overcast, M. J. Hickerson, R. T. Brumfield, and E. P. Derryberry. 2017. Behavioural response to song and genetic divergence in two subspecies of white-crowned sparrows (Zonotrichia leucophrys). Mol. Ecol. 26:3011–3027.

Little, A. C., R. P. Burriss, B. C. Jones, L. M. DeBruine, and C. a. Caldwell. 2008. Social influence in human face preference: men and women are influenced more for long-term than short-term attractiveness decisions. Evol. Hum. Behav. 29:140–146.

Luther, D. a., and E. P. Derryberry. 2012. Birdsongs keep pace with city life: Changes in song over time in an urban songbird affects communication. Anim. Behav. 83:1059–1066. Elsevier.

Lynch, A. 1996. The population memetics of birdsong. Pp. 181–197 *in* D. E. M. Kroodsma and E. H. Miller, eds. Ecology and evolution of acoustic communication in birds. Cornell Univ. Press, Ithaca, NY.

MacDougall-Shackleton, E. A., and S. A. MacDougall-Shackleton. 2001. Cultural and genetic

evolution in mountain White-crowned Sparrows: song dialects are associated with population structure. Evolution 55:2568–75.

Mann, J., and E. M. Patterson. 2013. Tool use by aquatic animals. Philos. Trans. R. Soc. B Biol. Sci. 368.

Marler, P. 1957. Specific distinctiveness in the communication signals of birds. Behaviour 11:13–39.

Marler, P. 1997. Three models of song learning: evidence from behavior. J. Neurobiol. 33:501–16.

Marler, P., and S. Peters. 1987. A Sensitive Period for Song Acquisition in the Song Sparrow, Melospiza melodia: A Case of Age- limited Learning. Ethology 76:89–100.

Marler, P., and M. Tamura. 1962. Song "dialects" in three populations of White-crowned Sparrows. Condor 64:368–377.

Martin, O., and D. Hosken. 2003. The evolution of reproductive isolation through sexual conflict. Nature 423:979–982.

Mason, N. A., K. J. Burns, J. A. Tobias, S. Claramunt, N. Seddon, and E. P. Derryberry. 2017. Song evolution, speciation, and vocal learning in passerine birds. Evolution (N. Y). 71.

McGregor, P. 1980. Song dialects in the corn bunting (Emberiza calandra). Z. Tierpsychol. 297:285–297.

McGregor, P., and J. Krebs. 1982. Song types in a population of great tits (Parus major): their distribution, abundance and acquisition by individuals. Behaviour 79:126–152.

McKinnon, J., S. Mori, and B. Blackman. 2004. Evidence for ecology's role in speciation. Nature 429:1–5.

Mery, F., S. A. M. Varela, E. Danchin, S. Blanchet, D. Parejo, I. Coolen, and R. H. Wagner. 2009. Public versus personal information for mate copying in an invertebrate. Curr. Biol. 19:730–4.

Nelson, D. 1992. Song overproduction and selective attrition lead to song sharing in the field sparrow (Spizella pusilla). Behav. Ecol. Sociobiol. 30:415–424.

Nottebohm, F. 1969. The song of the chingolo, Zonotrichia capensis, in Argentina: description and evaluation of a system of dialects. Condor 71:299–315.

Nottebohm, F., and M. E. Nottebohm. 1978. Relationship between Song Repertoire and Age in the Canary, Serinus canarius. Zeitschrift f??r Tierpsychologie 46:298–305.

Oetting, S., E. Prove, and H. J. Bischof. 1995. Sexual Imprinting as a 2-Stage Process - Mechanisms of Information-Storage and Stabilization. Anim. Behav. 50:393–403.

Olofsson, H., A. M. Frame, and M. R. Servedio. 2011. Can reinforcement occur with a learned trait? Evolution 65:1992–2003.

Olofsson, H., and M. R. Servedio. 2008. Sympatry affects the evolution of genetic versus cultural determination of song. Behav. Ecol. 19:596–604.

Otto, S. P., and T. Day. 2007. Box 8.2: Routh-Hurwitz Conditions for Local Stibility. Pp. 309–312 *in* A Biologist's Guide to Mathematical Modeling in Ecology and Evolution. Princeton University Press, Princeton, NJ, USA.

Owens, I. 1999. Sexual selection, speciation and imprinting: separating the sheep from the goats. Trends Ecol. Evol. 5347:131–132.

Papini, M. R. 2002. Pattern and process in the evolution of learning. Psychol. Rev. 109:186–201.

Payne, R. 1986. Bird songs and avian systematics. Curr. Ornithol. 3:87–126. Springer US.

Payne, R., L. Payne, and J. Woods. 1998. Song learning in brood-parasitic indigobirds Vidua chalybeata: song mimicry of the host species. Anim. Behav. 55:1537–53.

Payne, R., L. Payne, J. Woods, and M. Sorenson. 2000. Imprinting and the origin of parasite-host species associations in brood-parasitic indigobirds, Vidua chalybeata. Anim. Behav. 59:69–81.

Penn, D., and W. Potts. 1998. MHC-disassortative mating preferences reversed by cross-fostering. Proc. R. Soc. B Biol. Sci. 265:1299–1306.

Petrinovich, L., T. Patterson, and L. Baptista. 1981. Song dialects as barriers to dispersal: a re-evaluation. Evolution 35:180–188.

Pfennig, D. W., and M. R. Servedio. 2013. The role of transgenerational epigenetic inheritance in diversification and speciation. Non-Genetic Inherit. 1:17–26.

Pfennig, D. W., M. a Wund, E. C. Snell-Rood, T. Cruickshank, C. D. Schlichting, and A. P. Moczek. 2010. Phenotypic plasticity's impacts on diversification and speciation. Trends Ecol. Evol. 25:459–67.

Pfennig, K. S. 2007. Facultative Mate Choice Drives Adaptive Hybridization. Science (80-.). 318:965–967.

Planqué, R., N. F. Britton, and H. Slabbekoorn. 2014. On the maintenance of bird song dialects. J. Math. Biol. 68:505–31.

Polechová, J., and N. H. Barton. 2005. Speciation through competition: a critical review. Evolution 59:1194–1210.

Potvin, D. A., and S. M. Clegg. 2015. The relative roles of cultural drift and acoustic adaptation in shaping syllable repertoires of island bird populations change with time since

colonization. Evolution (N. Y)., doi: 10.1111/evo.12573.

Pryke, S. R. 2010. Sex chromosome linkage of mate preference and color signal maintains assortative mating between interbreeding finch morphs. Evolution (N. Y). 64:1301–1310.

Rosenfield, J. A., and A. Kodric-Brown. 2003. Sexual selection promotes hybridization between Pecos pupfish, Cyprinodon pecosensis and sheepshead minnow, C. variegatus. J. Evol. Biol. 16:595–606.

Rosenqvist, G., and A. Houde. 1997. Prior exposure to male phenotypes influences mate choice in the guppy, Poecilia reticulata. Behav. Ecol. 8:194–198.

Rosser, Z. H., T. Zerjal, M. E. Hurles, M. Adojaan, D. Alavantic, A. Amorim, W. Amos, M. Armenteros, E. Arroyo, G. Barbujani, G. Beckman, L. Beckman, J. Bertranpetit, E. Bosch, D. G. Bradley, G. Brede, G. Cooper, H. B. Côrte-Real, P. de Knijff, R. Decorte, Y. E. Dubrova, O. Evgrafov, A. Gilissen, S. Glisic, M. Gölge, E. W. Hill, A. Jeziorowska, L. Kalaydjieva, M. Kayser, T. Kivisild, S. a Kravchenko, A. Krumina, V. Kucinskas, J. Lavinha, L. a Livshits, P. Malaspina, S. Maria, K. McElreavey, T. a Meitinger, a V Mikelsaar, R. J. Mitchell, K. Nafa, J. Nicholson, S. Nørby, A. Pandya, J. Parik, P. C. Patsalis, L. Pereira, B. Peterlin, G. Pielberg, M. J. Prata, C. Previderé, L. Roewer, S. Rootsi, D. C. Rubinsztein, J. Saillard, F. R. Santos, G. Stefanescu, B. C. Sykes, A. Tolun, R. Villems, C. Tyler-Smith, and M. A. Jobling. 2000. Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. Am. J. Hum. Genet. 67:1526–43.

Rowell, J. T., and M. R. Servedio. 2012. Vocal communications and the maintenance of population specific songs in a contact zone. PLoS One 7:e35257.

Royle, N. J., and T. W. Pike. 2010. Social feedback and attractiveness in zebra finches. Behav. Ecol. Sociobiol. 64:2015–2020.

Rueffler, C., T. J. M. Van Dooren, O. Leimar, and P. a Abrams. 2006. Disruptive selection and then what? Trends Ecol. Evol. 21:238–45.

Rundle, H. D., S. F. Chenoweth, P. Doughty, and M. W. Blows. 2005. Divergent selection and the evolution of signal traits and mating preferences. PLoS Biol. 3:e368.

Saether, S. a, G.-P. Saetre, T. Borge, C. Wiley, N. Svedin, G. Andersson, T. Veen, J. Haavie, M. R. Servedio, S. Bures, M. Král, M. B. Hjernquist, L. Gustafsson, J. Träff, and A. Qvarnström. 2007. Sex chromosome-linked species recognition and evolution of reproductive isolation in flycatchers. Science 318:95–7.

Salinas-Melgoza, A., and T. F. Wright. 2012. Evidence for vocal learning and limited dispersal as dual mechanisms for dialect maintenance in a parrot. PLoS One 7:e48667.

Schneider, K. a, and R. Bürger. 2006. Does competitive divergence occur if assortative mating is costly? J. Evol. Biol. 19:570–88.

Searcy, W. A., P. Marler, and S. S. Peters. 1981. Species song discrimination in adult female song and swamp sparrows. Anim. Behav. 29:997–1003.

Searcy, W. A., S. Nowicki, and M. Hughes. 1997. The response of male and female song sparrows to geographic variation in song. Condor 99:651–657.

Searcy, W. a, S. Nowicki, M. Hughes, and S. Peters. 2002. Geographic song discrimination in relation to dispersal distances in song sparrows. Am. Nat. 159:221–30.

Secondi, J., V. Bretagnolle, C. Compagnon, and B. Faivre. 2003. Species-specific song convergence in a moving hybrid zone between two passerines. Biol. J. Linn. Soc. 80:507–517.

Seehausen, O., R. K. Butlin, I. Keller, C. E. Wagner, J. W. Boughman, P. A. Hohenlohe, C. L. Peichel, G. Saetre, C. Bank, Å. Brännström, A. Brelsford, C. S. Clarkson, F. Eroukhmanoff, J. L. Feder, M. C. Fischer, A. D. Foote, P. Franchini, C. D. Jiggins, F. C. Jones, A. K. Lindholm, K. Lucek, M. E. Maan, D. A. Marques, S. H. Martin, B. Matthews, J. I. Meier, M. Möst, M. W. Nachman, E. Nonaka, D. J. Rennison, J. Schwarzer, E. T. Watson, A. M. Westram, and A. Widmer. 2014. Genomics and the origin of species. Nat. Rev. Genet. 15:176–192.

Servedio, M., Y. Brandvain, S. Dhole, C. Fitzpatrick, E. Goldberg, C. Stern, J. Van Cleve, and D. Yeh. 2014. Not just a theory-the utility of mathematical models in evolutionary biology. PLoS Biol. 12:1–5.

Servedio, M., and M. Kirkpatrick. 1997. The effects of gene flow on reinforcement. Evolution 51:1764–1772.

Servedio, M., and M. Noor. 2003. The role of reinforcement in speciation: theory and data. Annu. Rev. Ecol. Evol. Syst. 34:339–364.

Servedio, M. R. 2011. Limits to the evolution of assortative mating by female choice under restricted gene flow. Proc. Biol. Sci. 278:179–87.

Servedio, M. R. 2000a. Reinforcement and the genetics of nonrandom mating. Evolution 54:21–9.

Servedio, M. R. 2000b. The effects of predator learning, forgetting, and recognition errors on the evolution of warning coloration. Evolution (N. Y). 54:751–763.

Servedio, M. R. 2009. The role of linkage disequilibrium in the evolution of premating isolation. Heredity (Edinb). 102:51–6.

Servedio, M. R., and R. Bürger. 2014. The counterintuitive role of sexual selection in species maintenance and speciation. Proc. Natl. Acad. Sci. 111:8113–8118.

Servedio, M. R., and R. Dukas. 2013. Effects on population divergence of within-generational learning about prospective mates. Evolution 67:2363–75.

Servedio, M. R., G. S. Van Doorn, M. Kopp, A. M. Frame, P. Nosil, and G. S. Van Doorn. 2011. Magic traits in speciation: “magic” but not rare? Trends Ecol. Evol. 26:389–97.

Servedio, M., S. Sæther, and G. Sætre. 2009. Reinforcement and learning. Evol. Ecol. 23:109–123.

Siegel, R. W., and J. C. Hall. 1979. Conditioned responses in courtship behavior of normal and mutant Drosophila. Proc. Natl. Acad. Sci. U. S. A. 76:3430–4.

Simoons, F. J. 1969. Primary Adult Lactose Intolerance and the Milking Habit: A Problem in Biological and Cultural Interrelations. Am. J. Dig. Dis. 14(12)819-836 15 14:816–836.

Slabbekoorn, H., and T. B. Smith. 2002. Bird song, ecology and speciation. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 357:493–503.

Slater, P. J. B. 2003. Fifty years of bird song research: a case study in animal behaviour. Anim. Behav. 65:633–639.

Smadja, C. M., and R. K. Butlin. 2011. A framework for comparing processes of speciation in the presence of gene flow. Mol. Ecol. 20:5123–40.

Snell-Rood, E. C., and D. R. Papaj. 2009. Patterns of phenotypic plasticity in common and rare environments: a study of host use and color learning in the cabbage white butterfly Pieris rapae. Am. Nat. 173:615–31.

Supreme Court of the United States. 2017. Cooper v. Harris. 581.

Templeton, C. N., C. Akçay, S. E. Campbell, and M. D. Beecher. 2010. Juvenile sparrows preferentially eavesdrop on adult song interactions. Proc. Biol. Sci. 277:447–453.

Templeton, C. N., V. A. Reed, S. E. Campbell, and M. D. Beecher. 2012. Spatial movements and social networks in juvenile male song sparrows. Behav. Ecol. 23:141–152.

ten Cate, C. 2000. How learning mechanisms might affect evolutionary processes. Trends Ecol. Evol. 5347:179–181.

ten Cate, C., and C. Rowe. 2007. Biases in signal evolution: learning makes a difference. Trends Ecol. Evol. 22:380–387.

ten Cate, C., and D. R. Vos. 1999. Sexual Imprinting and Evolutionary Processes in Birds: A Reassessment. Adv. Study Behav. 28:1–31.

Thibert-Plante, X., and S. Gavrilets. 2013. Evolution of mate choice and the so-called magic traits in ecological speciation. Ecol. Lett. 16:1004–13.

Thorndike, E. L. 1989. Animal Intelligence: An Experimental Study of the Associative Processes in Animals. Psychol. Rev. Monogr. Suppl. 2:i-109.

Tramm, N. A., and M. R. Servedio. 2008. Evolution of mate-choice imprinting: competing

strategies. Evolution 62:1991–2003.

Udovic, D. 1980. Frequency-dependent selection, disruptive selection, and the evolution of reproductive isolation. Am. Nat. 116:621–641.

van Doorn, G. S., P. Edelaar, and F. J. Weissing. 2009. On the origin of species by natural and sexual selection. Science 326:1704–7.

Veen, T., T. Borge, S. C. Griffith, G. P. Saetre, S. Bures, L. Gustafsson, and B. C. Sheldon. 2001. Hybridization and adaptive mate choice in flycatchers. Nature 411:45–50.

Vehrencamp, S. L. 2001. Is song-type matching a conventional signal of aggressive intentions? Proc. Biol. Sci. 268:1637–42.

Verner, J. 1975. Complex song repertoire of male long billed marsh wrens in eastern Washington. Living Bird 14:263–300.

Verzijden, M. N., R. F. Lachlan, and M. R. Servedio. 2005. Female mate-choice behavior and sympatric speciation. Evolution 59:2097–108.

Verzijden, M. N., and C. ten Cate. 2007. Early learning influences species assortative mating preferences in Lake Victoria cichlid fish. Biol. Lett. 3:134–6.

Verzijden, M. N., C. ten Cate, M. R. Servedio, G. M. Kozak, J. W. Boughman, and E. I. Svensson. 2012. The impact of learning on sexual selection and speciation. Trends Ecol. Evol. 27:1–9. Elsevier Ltd.

Vogel, D., and A. Dussutour. 2016. Direct transfer of learned behaviour via cell fusion in non-neural organisms. Proc. R. Soc. B Biol. Sci. 283:20162382.

Vos, D. R. 1995. The role of sexual imprinting for sex recognition in zebra finches: A difference between males and females. Anim. Behav. 50:645–653.

Wakano, J. Y., K. Aoki, and M. W. Feldman. 2004. Evolution of social learning: a mathematical analysis. Theor. Popul. Biol. 66:249–58.

Weissing, F. J., P. Edelaar, and G. S. van Doorn. 2011. Adaptive speciation theory: A conceptual review. Behav. Ecol. Sociobiol. 65:461–480.

Witte, K., and R. Massmann. 2003. Female sailfin mollies, Poecilia latipinna, remember males and copy the choice of others after 1 day. Anim. Behav. 65:1151–1159.

Yeh, D. J., and M. R. Servedio. 2015. Reproductive isolation with a learned trait in a structured population. Evolution 69:1938–1947.

Zann, R. 1985. Ontogeny of the Zebra Finch Distance Call .1. Effects of Cross-Fostering To Bengalese Finches. Zeitschrift Fur Tierpsychologie-Journal Comp. Ethol. 68:1–23.

Zann, R. 1990. Song and call learning in wild zebra finches in south-east Australia. Anim.

Behav. 40:811–828.

Zink, R. M., and G. F. Barrowclough. 1984. Allozymes and song dialect: A reassessment. Evolution (N. Y). 38:444–448.