MECHANISMS AND GENOMIC PATTERNS OF REPRODUCTIVE ISOLATION IN *XIPHOPHORUS* FISHES

A Dissertation

by

RONGFENG CUI

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Gil G. Rosenthal
Committee Members,	Adam G. Jones
	Michael Smotherman
	Spencer T. Behmer
Head of Department,	Thomas McKnight

May 2014

Major Subject: Biology

ABSTRACT

Learned mate choice has a fundamental role in population dynamics and speciation. Social learning plays a ubiquitous role in shaping how individuals make decisions. Learning does not act on a blank slate, however, and responses to social experience depend on interactions with genetically-specified substrates – the so-called "instinct to learn". I develop a new software Admixsimul, which allows forward-time simulations of neutral SNP markers and functional loci, mapped to user-defined genomes with user-specified functions that allow for complex dominance and epistatic effects. Complex natural and sexual selection regimes (including indirect genetic effects) are available through user-defined, arbitrary fitness and mate-choice probability functions. Using simulation, I show that responses to learned stimuli can evolve to opposite extremes in the context of mating decisions, with choosers either preferring or avoiding familiar social stimuli, depending on the relative importance of inbreeding avoidance versus conspecific mate recognition. I also show that under certain scenarios, learned preference is sufficient to maintain reproductive isolation during secondary contact. Two sister species of swordtail fish have evolved such opposite responses to learned social stimuli. The interaction of learned and innate inputs in structuring mate-choice decisions can explain variation in genetic admixture in natural populations. Olfactory preference of *X. malinche* females is affected by previous experience with adult cues. I compare gene expression levels of transcriptome libraries prepared from pooled sensory and brain tissues between 3 treatment groups that differ by social

experience. I found genes involved in neural plasticity differentially expressed not only between control and exposure groups, but also between groups exposed to conspecific and heterospecific models. I also found evidence that certain vomeronasal receptor type 2 (V2R) paralogs may detect species-specific pheromone components and show differential expression between treatment groups. I then reconstruct evolutionary relationships among swordtails and platyfishes (*Xiphophorus*: Poeciliidae). Using genomic data, I resolve a high-confidence species tree of *Xiphophorus* that accounts for both incomplete lineage sorting and hybridization. The results allow me to re-examine a long-standing controversy about the evolution of the sexually selected sword in *Xiphophorus*, and demonstrate that hybridization has been strikingly widespread in the evolutionary history of this genus.

DEDICATION

To my grandparents, parents and family

To my childhood friends.

To my teachers.

ACKNOWLEDGEMENTS

My dissertation study would have not been possible without the unconditional help and mentorship from my advisor Prof. Gil Rosenthal. He not only taught me knowledge related to his expertise, but also instilled me with optimistic (and calorious) energy throughout my training. I thank my committee members Prof. Adam Jones, Prof. Mike Smotherman and Prof. Spencer Behmer for guiding the path of my study and providing constructive ideas. Prof. Mariana Mateos & Prof. Jim Woolley taught me phylogenetics, and generously helped me gain access to the TAMU Brazos cluster, where I conducted most of the computational work. I am indebted to my collaborators Machteld Verzijden, Molly Schumer and Prof. Peter Andolfatto for working with me on sexual imprinting and high-throughput genomics. I also thank my undergraduate assistant Michael Stanley for his hard work. Benjamin Philmus, Prof. Jim Smith, Prof. Prof. Larry Dangott and Prof. Yohannes Rezenom have provided generous assistance and equipment in an early project in pheromone characterization.

During the past five years I have received much help and moral support in study and life from many great friends and colleagues in the department, including Prof. Charles Criscione, Prof. Leslie Winemiller, Dr. Kay Goldman, Dr. Brad Johnson, Dr. Zach Culumber, Dr. Holly Kindsvater, Dr. Clay Small, Dr. Jill Detwiler, Dr. Michi Tobler, Dr. Kim Paczolt, Dr. Nick Ratterman, Courtney Passow, Olivia Ochoa, Melissa Giresi, Emily Rose, Emily Kasl, Mitchell Ramsey, Ying Fang, Nehemiah Cox, Sehresh

Saleem, Amrita Sherlekar, Victoria Smith, Huinan Li, Pablo Delclos, Mattie Squire, Sarah Flanagan, Dan Powell, Gastón Jofre and Laura Edelstein.

My work is supported by an NSF DDIG and a Rosemary Grant from S.S.E. I also received travel grants from TAMU Neuroscience, and SICB.

As a boy growing up in a large, economic-driven city Canton, I am fortunate enough to pursue my childhood dream of becoming an animal biologist, a rather outlandish career in the culture, thanking to the support and love of my parents and my grandparents. I thank my childhood friends, Mingkai Liu, Yue Zhuo, Yaqi Lin among many others, who have embraced my love for animals. I am also indebted to many outstanding educators, whose names I cannot fully list below, during my K-12 and college education. They have encouraged, equipped and helped me to develop my interest into a career path: Ms. An Zhang (maths), Ms. Yingpei He (Chinese), Mr. Hua Shen (Chinese), Ms. Xin Chen (biology), Ms. Shaoming Liu (biology), Mr. Jianrong Yi (biology), Ms. Jufen Zeng (chemistry), Wendy Lan (English), Aaron Yao (English), Mr. Yingyong Wang (undergraduate advisor) and Prof. Zhaorong Lun (undergrad. co-advisor).

Finally, with this work I would like to commemorate my beloved grandma, who seeded my interest in animal diversity, and passed away 55 days before my defense.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

CHAPTER I

INTRODUCTION

Complete reproductive isolation is the landmark of speciation, during which gene flow is largely ceased, allowing for continuous divergence between incipient species and eventually lead to complete species. Many studies have focused on the biochemical and genetic underpinnings of isolating mechanisms between divergent species (Presgraves et al. 2003), and often find drastic reduction of hybrid fitness such as sterility of the heterogametic sex (Haldane's rule) or lowered viability caused by incompatible interacting loci (Masly and Presgraves 2007). However, the seemingly strong isolating barriers observed in highly divergent species are unlikely to shed much light on the initial cause of divergence, because many of these strong barriers evolve much later after the divergence pattern has been canalized by genes that interact on a higher biological level, albeit the isolating effects of the initial barriers may be weak (Figure 1). During the early stage of divergence, pre-mating isolating may likely evolve by two mechanisms (Kaneshiro 1980): 1) lower relative fitness in hybrids (regardless of whether the selection against hybrid is "intrinsic" or "extrinsic") selects for pre-mating isolation, the so called "reinforcement" process (Servedio and Noor 2003) or 2) premating isolation evolves as a by-product of allopatric local adaptation of loci that affect female preference. Informed by the observation that closely related taxa can frequently hybridize, hybrid zones provide a window into the speciation process (Jiggins and Mallet

2000). In animals, pre-mating isolation often involves female mate choice (Andersson 1994).

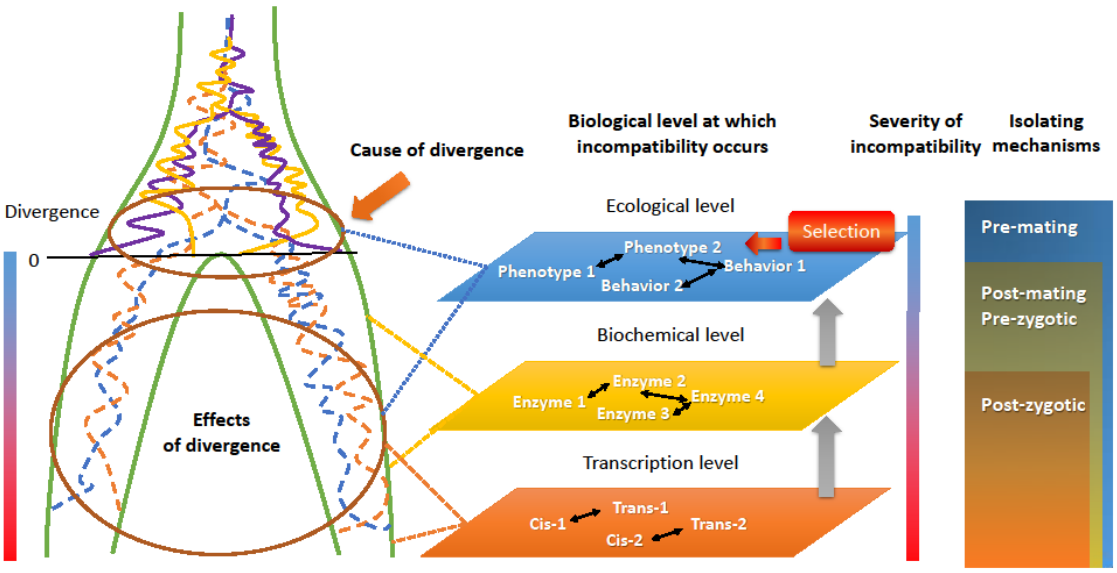


Figure 1. Schematic drawing of genetic loci as cause and effects of lineage divergence. Many loci incompatibilities that setup the divergence patterns before the final speciation event interact on an ecological level.

Many models on preference-trait co-evolution assume female mate choice to be genetic, in the sense that females have a genetic locus or several quantitative loci that target an "optimal" male trait value (Boake 1991). Although convenient for modeling, whether such innate preference is generally plausible mechanistically is unknown. One possible mechanism is the sensory bias hypothesis (Endler and Basolo 1998; Ryan 1998), where the female sensory periphery under natural selection constraints may become more sensitive to certain stimulus (e.g. wavelength of light, frequency of sound etc.). Under the strong assumption that stronger stimulus translates into stronger preference, male signals that "exploit" such pre-existing bias gain higher reproductive

success. Another mechanism is pleiotropic effects of a gene that controls both the signal and the preference (Marcillac et al. 2005). For example, both the call frequency and the auditory frequency sensitivity of frogs are correlated with body size, and thus the genes underlying body size have a pleiotropic effect on both the signal and the receiver (Ryan and Wilczynski 1988). Nonetheless, preference for signals of a more complex nature or a higher organization level, such as song structures in frogs and birds, complex color patterns, and behavioral courtship traits may not be easily explained by such a direct gene-to-gene targeting mechanism. Indeed, accumulating evidence suggests that mate choice is also affected by indirect genetic effects (Wolf et al. 1998) such as learning (Irwin and Price 1999). Learned mate preference alleviates the need for females to possess "innate preferences". For example, a wide range of animals are found to base their preference on the experience with parents, peers or previous mates (Verzijden et al. 2012b). The role of learned mate choice in speciation is debated, but at least in certain scenarios (e.g. paternal imprinting), learned mate choice generates similar results as if mate choice is genetically determined (Verzijden et al. 2005). Many previous models have assumed learning to have a positive effect on the outcome of mate choice. It has recently been shown in a model that negative effects of learning may prevent the Fisherian run-away process in signal-receiver coevolution (Bailey and Moore 2012). The evolution of learned mate avoidance *per se*, however, received little attention. For Chapter I, I develop a new individual-based simulation software package that allows convenient modeling of population dynamics incorporating learned/genetic mate choice, natural selection and complex genetic architectures. Then, I combine simulation and

empirical studies in two natural hybridizing *Xiphophorus* species to demonstrate that a negative effect of learning on mate preference can evolve under heterozygous advantage, and that opposing effects of learning may promote or prevent reproductive isolation.

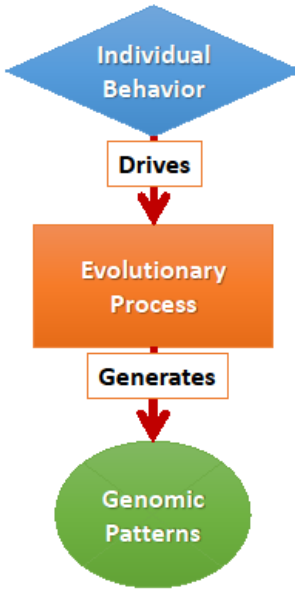


Figure 2. Studying mechanisms of mate choice may help to understand genomic patterns.

Behavioral mechanisms when operating on an evolutionary time scale can leave identifiable patterns in the genome. With the advent of the genomic era, many studies have strived to extract information about the underlying evolutionary processes from observable genomic patterns. A caveat is that very often, the genomic patterns are uninterpretable (i.e. contains no information) with regard to a parameter of interest due to the fact that different processes may result in similar patterns (Vitti et al. 2013). Hence, having *a priori* knowledge on the processes, and the mechanisms that promote

those processes, reduces the dimensionality of the inference problem, and helps ruling out alternative hypotheses (Figure 2).

Animals use multi-modal communication to make mate choice decisions (Partan and Marler 2005). Among these, visual and acoustic cues are probably the most well-studied ones. Many lab experiments have demonstrated that learning alters preference for these cues (Verzijden and ten Cate 2007; Verzijden and Rosenthal 2011), and mechanistically the change in the nervous system either occurs at the sensory periphery (e.g. differential expression of opsins) (Parry et al. 2005) or involves memory at a higher processing level (e.g. memory of bird songs) (Bolhuis et al. 2012). Less study has focused on the mechanism of plasticity in the olfactory modality in the context of mate choice (but see Niehuis et al. 2013), which is the most commonly used and the most ancient method of animal communication (Wyatt 2003). More and more recent studies suggest that olfactory cues are important in conveying social information and serve as effective identifiers for sex, individual, population and species (Lassance and Löfstedt 2013). In vertebrates, olfactory sensitivity (Nevitt et al. 1994) and preference (Crapon de Caprona and Ryan 1990) are both affected by previous experience. In Chapter II, I use high-throughput RNA-seq technique to study the molecular mechanisms underlying learned olfactory preference behavior by examining gene expression patterns in *X. malinche* females with different social exposure experiences.

Given the findings in Chapters I & II on learned female mate preference in *X. birchmanni* and *X. malinche*, I expect that the reproductive barriers between closely related *Xiphophorus* species should be porous, and particular parts of the genome may

be susceptible to introgression. In Chapter IV, I examined this prediction in the genus *Xiphophorus* by using high-throughput phylogenomics to test whether mate choice behavior observed in lab experiments leaves signatures of extensive hybridization among lineages.

CHAPTER II

EVOLUTION OF REVERSED HEDONIC VALUE IN LEARNED MATING PREFERENCES

Introduction

Social learning plays a ubiquitous role in shaping how individuals make decisions, including choosing a mate. Many animals base their mating preferences on the phenotype of other individuals they experience during development, a phenomenon known as sexual imprinting (Cate and Vos 1999; Owens et al. 1999; Griffiths 2003; Verzijden and ten Cate 2007; Kozak and Boughman 2009). Most models of sexual selection (Laland 1994a, b; Aoki et al. 2001; Ihara et al. 2003; Tramm and Servedio 2008) and speciation (Kirkpatrick and Dugatkin 1994; Agrawal 2001; Verzijden et al. 2005) assume that learning causes a positive association between an imprinted trait and later mate preference, such that choosers prefer to mate with individuals bearing familiar phenotypes. Yet the association between learned stimuli and subsequent behavior depends on a genetically-specified substrate – the so-called "instinct to learn" (Marler 1991). In some circumstances, notably in populations with a high degree of inbreeding depression, selection should favor avoidance rather than acceptance of familiar individuals, since unfamiliar males are less likely to be close relatives (Pusey 1980; Grob et al. 1998; Penn and Potts 1998; Weisfeld et al. 2003). Females often avoid the phenotypes of close kin (Fadao et al. 2000; Lehmann and Perrin 2003; Weisfeld et al. 2003; Gerlach and Lysiak 2006) or prefer novel, unfamiliar phenotypes (Burley et al.

1982; Qvarnström et al. 2004). By contrast, selection should favor imprinting on familiar individuals if it reduces the likelihood of mating with heterospecifics (Butlin 1987).

The effect of social stimuli on mating preferences can be parameterized as Ψ (Bailey and Moore 2012). When Ψ is positive, females have a higher probability of mating with males bearing phenotypes they have previously experienced. When Ψ is negative, females avoid mating with males bearing familiar phenotypes. A recent model (Bailey and Moore 2012) showed that positive Ψ can accelerate Fisherian runaway coevolution of female preference and male trait, while negative Ψ prevents runaway coevolution.

Here, I first use individual-based simulations to show that Ψ can evolve to be strongly positive or negative depending on the relative fitness costs of mating with close kin or heterospecifics. I further show that secondary contact of lineages with opposing Ψ's may either prevent or promote gene flow depending on genetic architecture and admixture proportions. I then report an empirical example of rapid evolution of Ψ in a young sister species pair by comparing the mating preferences of *Xiphophorus malinche* to those previously reported for its sister species, *X. birchmanni*, raised under identical social conditions (Verzijden and Rosenthal 2011). Given the divergent responses to social experience in these two species, my simulation model is consistent with dramatic differences in observed population-genetic patterns between two natural hybrid zones.

Materials and methods

First, I take an individual-based simulation approach (Appendix A) to address two problems: 1) how varying heterozygote viability affects the evolution of Ψ; 2) what types of admixture dynamics arise if two populations opposing Ψ's come into secondary contact. I first outline simulation settings that are common for both simulations 1 and 2, and then I detail the specifics for the two simulations. I follow my simulations with empirical studies to test model predictions.

Genetic architecture

I simulated diploid organisms with eight unlinked biallelic loci, with discrete allelic values of 1 or 0. These eight loci additively code a male-limited signaling trait (signal) in the range of [0,16] (Figure B-1). The value of Ψ is coded by a ninth, unlinked locus. The sex-determining locus is unlinked to the other nine loci, which results in a 1:1 sex ratio.

Sexual selection model

After natural selection (only applied in simulation 1), surviving individuals breed following these stages: 1) each female randomly samples 1000 males (expected number of males in the population) without incurring any cost; 2) during each sampling, she accepts the male with a probability defined below; 3) If she accepts a male, she goes on sampling other males; 4) when sampling ends she uses gametes randomly drawn from the pool of accepted males to produce offspring. 5) each female produces on average the same number of offspring such that the total number of offspring adds up to the carrying capacity.

Acceptance probability p is defined as:

$$p = \begin{cases} 1 & \Psi = 0 \\ \Phi(\Psi) & \Psi > 0 \\ \Phi(-\Psi) & \Psi < 0 \end{cases} \text{ where } \Phi(\Psi) = \begin{cases} 1 & \bar{s}' = s_{mid} \\ \max\left(\Omega(i=1), \exp\left[\frac{-(\bar{s}'-s)^2}{2(1/\Psi)^2}\right]\right) & \bar{s}' > s_{mid} \\ \max\left(\Omega(i=0), \exp\left[\frac{-(\bar{s}'-s)^2}{2(1/\Psi)^2}\right]\right) & \bar{s}' < s_{mid} \end{cases}$$

$$\textbf{where } \Omega(i) = \begin{cases} 1 & i=1, s \geq \bar{s}' \\ 0 & i=1, s < \bar{s}' \\ 1 & i=0, s \leq \bar{s}' \\ 0 & i=0, s > \bar{s}' \end{cases} \textbf{ and } \quad \bar{s}' = \begin{cases} \bar{s} & \Psi > 0 \\ s_{mid} + (s_{mid} - \bar{s}) & \Psi < 0 \end{cases}$$

Ψ – Ψ value encoded by the Ψ locus. | Ψ | is choosiness against suboptimal mates.

s – the signaling trait value of the courting male under assessment by the female.

s_{mid} – 16/2, the mid-point of the male trait range

$\bar{s}$ – the average trait value for model males in the previous generation (oblique imprinting; main text), the current generation (peer imprinting; Appendix B i) or the paternal trait (paternal imprinting; Appendix B i).

$\bar{s}'$ – the preferred male signal, defined as $\bar{s}$ when Ψ is positive and as s_{mid} + (s_{mid} – $\bar{s}$) when Ψ is negative (Figure B-2). Flipping the sign of Ψ therefore results in a mirrored preference function around the male trait mid-point s_{mid}.

Essentially, I assume that females divide male traits into two categories (the cutoff being the population-average male trait value) and learn to prefer (Ψ >0) / avoid (

$\Psi < 0$) one category depending on which category contains more males in the previous generation (Figure B-2A). I also explored a less realistic unimodal function (Appendix B iv). In my formulation, the absolute value of Ψ models the choosiness against suboptimal male traits, with higher values being choosier (Figure B-3A).

Simulation 1: Evolution of the sign of Ψ

Population setup

In simulation 1 I used 3 populations, where populations A and B merely provide polymorphism of male traits to the focal population F, in which I explore the evolution of Ψ. I first seeded two source populations A and B with fixed alleles in loci 1-9, such that population A has a fixed male signaling trait of 16 and population B is fixed for 0. Allelic values (and thus the phenotype value) at the Ψ locus were forced to 0 and not allowed to evolve in populations A and B. In population F, the allelic values of the Ψ locus are free to mutate in the range [-2,2] with mutation probability of 0.001 per meiosis. This high mutational rate was chosen to accelerate the simulation. I then populate my focal population F with 10 individuals from population B in generation 1. From generation 2 and on, 1 immigrant from A and 10 from B enter F every generation. The carrying capacity of populations A and B is 200, that of F is 2000. In recently bottlenecked populations, rare alleles (frequency <10%) are most prone to loss due to drift (Luikart et al. 1998); because the effect of negative Ψ is to maintain the rare allele (i.e., if the "rare allele" is already abundant, there's no need for another mechanism to maintain it), I set the rare allele frequency to 1/11 (9.1%) to mimic this realistic condition (see Appendix B ii for simulations with other rare allele frequencies). Here I

use continuous gene flow to maintain polymorphism of the male signaling loci in the focal population F. In reality polymorphisms can also be maintained by other mechanisms such as negative frequency-dependent selection.

Natural selection conditioned on heterozygosity

For convenience, I use "viability" to refer to the survival probability during natural selection. In each non-overlapping generation, adults in the population F were subjected to natural selection. I varied natural selection across runs to mimic the viability effects of varying degrees of heterozygote viability. I assume that heterozygosity increases viability in an inbred population, and decreases viability in an outbred population where there is a high risk of introgressing heterospecific alleles. I define H_{he} as the number of male signaling loci that confer higher viability when heterozygous, and H_{ho} to be those that confer higher viability when homozygous. I consider different proportions (H_{he}:H_{ho}) of the signaling loci. When H_{he}:H_{ho} = 8:0, it mimics an inbred population where all 8 loci are more fit in the heterozygous state while H_{he}:H_{ho} = 0:8 mimics an outbred population where complete homozygosity is more fit in terms of natural selection. Note that here I assume that the loci underlying the male trait are closely linked to or pleiotropic with loci underlying traits contributing to viability.

Survival probability is defined by the Gaussian function $p = \exp\left[\frac{-(f - f_{\max})^2}{2w^2}\right]$, where f is the number of signaling loci in a "more viable" state, f_{max}=8 (all loci in "more fit" state), with w (the strength of selection) set to 6 such that p=0.4 when f=0.

The runs

I repeated simulations of each H_{he}:H_{ho} proportion 10 times to calculate the mean and standard deviation of population means of Ψ. I ran the simulations for 5000 generations, sampling population statistics every 10 generations.

Simulation 2: Admixture of lineages with opposing Ψ's

I performed another set of simulations to explore if admixing populations with opposing Ψ's maintain reproductive isolation. Because I were now interested in short-term dynamics, I assumed that Ψ's have previously evolved to fixation parental species and mutation can be ignored. I fixed the allelic values at the Ψ locus to 0.25 and -0.25. In addition to the basic genomic setup, I placed 10 evenly-spaced (8.73cM between neighbors) neutral SNP markers on each of the 10 chromosomes from which I calculated genome-wide F_{IS} values (Figure B-1). Preference functions are as defined before. Unlike simulation 1, I did not impose natural selection in these simulations because I're now not interested in the evolution of Ψ. I model admixture of two parental populations ($Pop_{\Psi<0}$ and $Pop_{\Psi>0}$) fixed for their 8 signaling loci and their Ψ locus such that both traits take reciprocal extreme values (16 and -0.5, 0 and 0.5 respectively). I define N as the number of migrants per generation. A specified proportion of migrants from the two parental populations ($N_{\Psi<0}$: $N_{\Psi>0}$) moves into a focal hybrid population with a carrying capacity of 2000 (migration rate = 5%, or 100 migrants total) per generation. Admixture was conducted for 1000 generations, with F_{IS} and male trait distributions summarized every 10 generations. If a change in admixture proportions led to drastic change in admixture dynamics, I finely tuned the proportions to identify the parameter break point.

Preliminary analyses showed that results were highly repeatable across runs, so I only show data for a single run for simulation 2.

Simulations were performed using the software admixsimul (https://github.com/melop/admixsimul) (Cui & Rosenthal, in prep.). A copy of the current version of application code, configuration files and raw outputs are to be deposited at Dryad.

Fish collection and rearing

Two *X. malinche* populations and one *X. birchmanni* population were used in the study. Wild-caught female preference was tested in adult *X. malinche* collected using baited minnow traps from the type locality at the R ío Claro at Tlatzintla, Hidalgo, Mexico (20 º52' 51" N, 98 º47' 56" W). Since *X. malinche* appears to have been driven extinct at the type locality due to strip-mining for manganese prior to my exposure experiment, *X. malinche* for lab-rearing experiments came from the Arroyo Xontla near Chicayotla (20°55’ 30’ N, 98°34’36’ W), Mexico. Whenever *X. malinche* females were tested, adult *X. malinche* for olfactory cue production or imprinting models always came from the same populations as the subject females. All *X. birchmanni* in this study came from Río Garces (20°56’24’ N, 98°16’54’ W). The *X. malinche* for lab-rearing experiments and *X. birchmanni* both came from the same populations used in a previous sexual imprinting study (Verzijden and Rosenthal 2011). I stress that phenotypic (Rauchenberger et al. 1990) and genomic divergence (Schumer et al. In review) between species is much higher than intraspecific polymorphisms.

For the lab-rearing experiments, I collected 16 gravid *X. malinche* females. Females were housed in groups of 4 in 38L tanks installed with vertical perforated plastic fry dividers. Subject fry were collected within 24 hours after birth and were kept with visual and olfactory isolation from adults before exposure treatments. Due to the small brood size (on average 7-8 fry per brood) in *X. malinche* and because no family or tank effects were detected in a previous imprinting experiment with the sister species *X. birchmanni* (Verzijden and Rosenthal 2011), I decided to pool the families into two batches, based on date of birth. Fry were originally kept separated by brood until the pooling stage to ensure even distribution of families in each experimental group. Batch one of 55 fry was collected from 8 females in early May to early July in 2010 with an average date of birth in early June, 2010 (±30 days); batch two of 67 fry was collected from another set of 8 females in June to July in 2010 with an average date of birth in early July, 2010 (±20 days].

Batch one was used as a control group, which was housed in two 200L aquaria (28 per tank) in isolation from adults. Batch two was divided into two exposure groups, each (33 per group) with an approximately equal number of fry from each family. Exposure treatments started on 31 August 2010, when the average age of batch two was 60±20 days. Each of the exposure tanks (200L) was divided by a perforated Plexiglas board at 1/4 along the length to allow transmission of both visual and olfactory signals from model adults housed in the smaller compartment. Exposure group one (B-Exp) was exposed to two male and two female adult *X. birchmanni*. Exposure group two (M-Exp) was exposed to two male and two female adult *X. malinche* from the same population.

None of the mothers of the subject fry were used as models. A 30% water change was performed weekly. A previous study in *X. birchmanni* showed that a similar setup allowed sufficient visual and olfactory stimuli for sexual imprinting (Verzijden and Rosenthal 2011). Although I did not randomize batch one and batch two between the control and exposure groups due to the large age difference between the two batches, both the control group and the exposure groups were tested in visual and olfactory trials at an average age of 11 months thus the difference in behavior is unlikely an age effect.

To minimize the effects of peer learning and to prevent focal females from mating, maturing males in all three groups were removed as soon as signs of sexual differentiation (thickening of the anal fin into the gonopodium) were seen. Peer imprinting on fellow females in the cohort could not be eliminated, but such effect, if any, is expected to be equal in the control, M-EXP and B-EXP groups. There is also evidence that male-specific courtship behavior is correlated with olfactory cue production (Rosenthal et al. 2011), thus olfactory imprinting of courtship-related pheromones from fellow females is unlikely. Exposure lasted for 272 days before the first behavioral trial. After each trial, subjects were temporarily housed in a 5L aquarium containing fresh conditioned water for about 5 hours, then back to their home tanks before subsequent trials to avoid carrying over olfactory cues. The fish room was maintained a 12:12 h light:dark cycle at 22-23 ℃. Fry were fed twice a day with *Spirulina* flake food (Angelsplus, USA) and decapsulated brine shrimp. Older fry and adults were fed frozen bloodworms in place of brine shrimp. The artificially controlled

rearing and testing conditions preclude the possibility of behavioral difference in control and exposure groups caused by seasonal effects.

Visual preference trials

I used previously described computer animation playback techniques (Wong and Rosenthal 2006; Fisher et al. 2009; Verzijden and Rosenthal 2011) to test focal females' preferences for *X. birchmanni* versus *X. malinche* visual cues. I used previously described *X. birchmanni* and *X. malinche* computer animations representing mean morphological traits of each species. (Verzijden and Rosenthal 2011) were used. As previously described (Verzijden and Rosenthal 2011), CRT monitors displayed male courtship animations with mean *X. birchmanni* and *X. malinche* phenotypes controlled by a video server synchronized with the Viewer (Biobserve GmbH, Bonn, Germany) recording software. For wild-caught females, videos were started manually and association times (see below) were scored by direct visual observation.

The trial tanks (51 x 28 cm filled to a depth of 20 cm) were opaque on all sides except at the short ends, which are placed against the CRT screens. Female position during the trial was recorded at 8 frames per second using an overhead camera. The tank was equally divided along its length into three virtual zones with the middle one defined as the neutral zone.

CRT monitors displayed a blank screen with a uniform color identical to the background in the animation stimulus during the 10-min acclimatization period. At the end of acclimatization, standard-length matched *X. birchmanni* and *X. malinche* visual stimuli started playing on each side for 5 min followed by 5 min blank screen and then

another 5 min of stimuli with the sides switched. Association time for each stimulus was summed from the two 5-min stimulation periods. Association time is a repeatable measure of preference and predicts mate choice in poeciliids (Cummings and Mollaghan 2006; Walling et al. 2010).

If the female stayed in one zone for >290s (out of 300s) in both trial periods, she was operationally defined as non-responsive and excluded from data analysis. Side of first presentation was systematically altered across trials. I tested 26 (19 responsive) control, 18 (18 responsive) B-Exp and 17 (13 responsive) M-Exp females.

Olfactory preference trials

Two days after the visual trials, I tested female preference for conspecific versus heterospecific male odors following the protocol described in previous studies (McLennan and Ryan 1999; Fisher et al. 2006; Verzijden and Rosenthal 2011). To produce the olfactory cues, 20L aquaria were thoroughly cleaned with 1:1 mixture of hydrogen peroxide and soap and rinsed with carbon-filtered water at least 6 times. Groups of four male *X. birchmanni* and four male *X. malinche* were separately placed in 16 liters of carbon-filtered water and visually exposed to 6 females from their own population in adjacent identical tanks for 4 hours to interact. In total, 6 male *X. birchmanni* and 6 male *X. malinche* were used to make the stimulus water. Model males in the exposure treatments were never used to produce stimulus water.

During the preference tests, trial tank configurations were as reported above except that the two transparent ends were blocked with opaque white paper. Ten minutes before each trial, the focal female was introduced to the testing tank for acclimation.

After the 10-minute period, stimulus water started dripping on both far ends of the tank driven by a custom circuit board-controlled peristaltic pump (VWR) until the end of trial, at a flow rate of approximately 5 ml/min. When the cues started dripping, I allowed 5 min for the focal female to visit both preference zones. If the subject failed to do so she was operationally defined as unresponsive and excluded from data analysis. Starting at the moment the subject entered the third zone, the time in each zone was recorded for a total of 5 min. Each female was tested twice back-to-back with the first presentation of cues randomized by sides. Then the cue sides were switched in the second trial. I averaged the association time in the two trials for data analysis. If the female was unresponsive in one trial, I only included association time from the other trial in analysis. The definition of operational unresponsiveness differs between visual and olfactory trials, because I assume females can only sense the olfactory cue by physically approaching the zone, while visual cues become immediately available as the monitors turned on regardless of the physical position of the female. I tested 26 (18 responsive) control, 18 (17 responsive) B-Exp and 17 (17 responsive) M-Exp females.

Results and discussion

Increased heterozygote viability promotes negative Ψ

I simulated a population of sexual, diploid organisms with 8 male signaling loci (SLs) and a single locus coding for the value of Ψ. Baseline polymorphism of the 8 SLs was maintained by gene flow from two populations that are reciprocally fixed at all

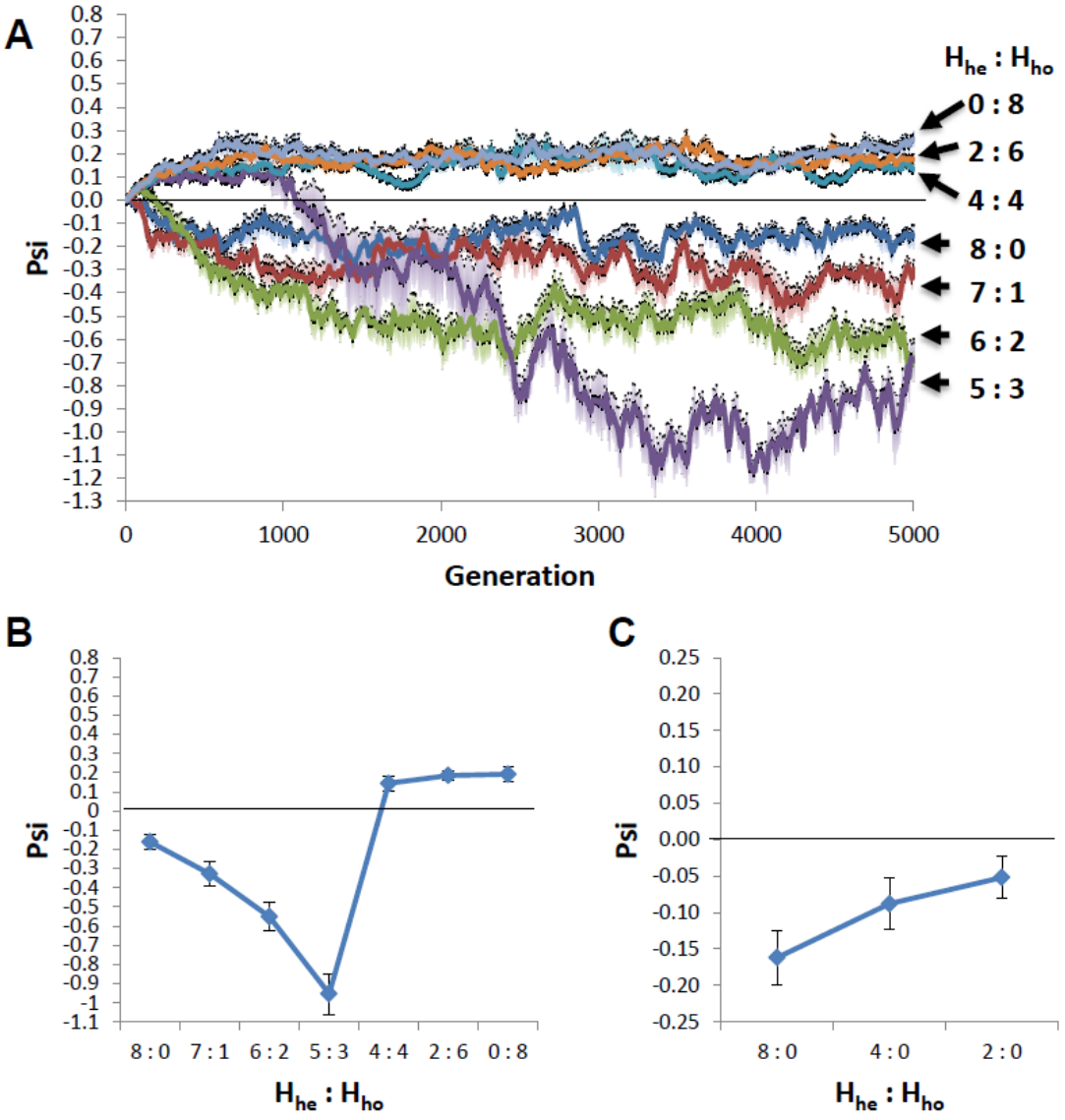


Figure 3. Evolution of Ψ under different proportions of loci that are fit in heterozygous state (H_{he}) vs. homozygous state (H_{ho}), assuming oblique imprinting (see Fig S4 for peer imprinting). A) Mean and S.E.M. of population average of Ψ across 10 independent runs plotted against generations for each proportion setting, except for H_{he} : H_{ho} = 5:3, where 4 runs were excluded because they did not reach equilibrium by the end of 5000 generations. B) Mean and S.E.M of Ψ summarized from generations 3000-5000 of different H_{he} : H_{ho} proportions. C) Mean and S.E.M of Ψ summarized from generations 3000-5000 simulated with 8, 4, 2 H_{he} and 0 H_{ho} loci.

alleles, at a proportion of 1:10 (baseline rare allele frequency 9.1%, see Appendix B ii for other proportions). The starting value of Ψ was 0 and allowed to mutate with a probability of 0.001 per meiosis. I modeled the viability of individuals as a function of the heterozygote states of SLs. I define H_{he} as the number of SLs that confer higher viability when heterozygous, and H_{ho} to be those that do so when homozygous. I consider different proportions (H_{he}:H_{ho}) of the SLs. Note that here I assume that the SLs are closely linked to or pleiotropic with loci underlying traits contributing to viability. When H_{he}:H_{ho} = 8:0, analogous to a highly inbred population where heterozygotes are more fit, the estimates of population means of Ψ rapidly evolved to and generally stayed at negative values across 10 independent simulations, although occasionally Ψ can become positive (Figure 3A). The absolute value of Ψ , which corresponds to female choosiness, fluctuated over time. There were periodic fluctuations of the absolute value of Ψ, which in my model represented the choosiness of the females.. When I modeled heterozygotes at all 8 loci as unfit, analogous to an outbred population with a high fitness cost to hybridization, Ψ evolved stably to positive values (Figure **3**). Similar results were also found assuming peer (Figure B-4A) and paternal imprinting (Figure B-4B). Ψ evolved to positive values when $H_{he} <= H_{ho}$ and negative values when $H_{he} > H_{ho}$, with more negative Ψ (-0.5511 ±0.0752) observed when H_{he} :H_{ho} is closer to 1:1. There is an abrupt switch from negative to positive Ψ as the H_{he} :H_{ho} ratio approaches 1:1 (Figure B-5, Figure B-6). Reducing the number of QTLs of the male signaling trait to 4 and 2, or assuming oblique (Figure 3), peer (Figure B-4A) or parental imprinting (Figure B-4B) do not change the conclusions qualitatively.

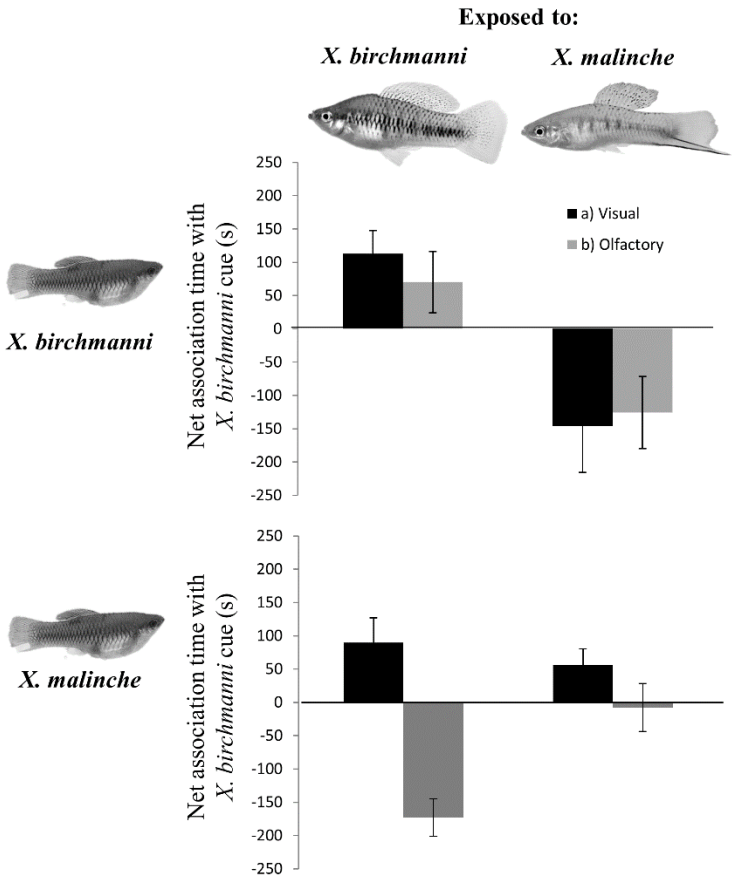


Figure 4. Contrasting effects of early social experience in the sister species *X. birchmanni* and *X. malinche*. Results for *X. birchmanni* were published in a previous study (Verzijden and Rosenthal 2011).

Thus, negative Ψ's are expected to evolve when heterozygosity confers higher viability than homozygous genotypes. The evolutionary dynamics of Ψ differed depending on its sign. When Ψ is positive, stable equilibrium can be reached by driving allele frequencies to extreme values (assuming, though, polymorphism is maintained by other mechanisms such as gene flow, negative frequency-dependent selection or

mutation). However when Ψ is negative, the absolute values of Ψ fluctuate periodically as a function of the population heterozygosity (Figure B-7).

As an empirical example, I compare learned preference in a recently diverged sister species of swordtail fishes *Xiphophorus malinche* and *X. birchmanni*. I raised female *X. malinche* with exposure to adult conspecifics, heterospecifics and no adult models and then tested their preference for conspecific versus heterospecific cues. Female *X. malinche* showed experience-independent preferences for visual cues for heterospecific male cues and negatively experience-dependent preferences for olfactory cues (Figure 4; Table B-1), in striking contrast to females of the sister species *X. birchmanni*, which showed positively experience-dependent preferences in both modalities (Verzijden and Rosenthal 2011). These patterns mirror the effects of short-term social experience on olfactory preferences in adult *X. malinche* and *X. birchmanni* females, where *X. malinche* females avoided heterospecific males after a one-week exposure to heterospecifics, while *X. birchmanni* females did not (Verzijden et al. 2012a). *X. malinche* from a different population also avoided familiar barring symmetry they first experienced (Tudor and Morris 2009). This learned disdain likely has a genetic component in this species and affects learned mate preference across different ontogenetic periods. *X. malinche* differs from its congener *X. birchmanni* in that they have lower standing genome-wide polymorphism (Cui et al. 2013) due to its smaller effective population size resulted from the limited distribution in high-land habitats. Morphologically *X. malinche* are exceptionally homogeneous compared to other

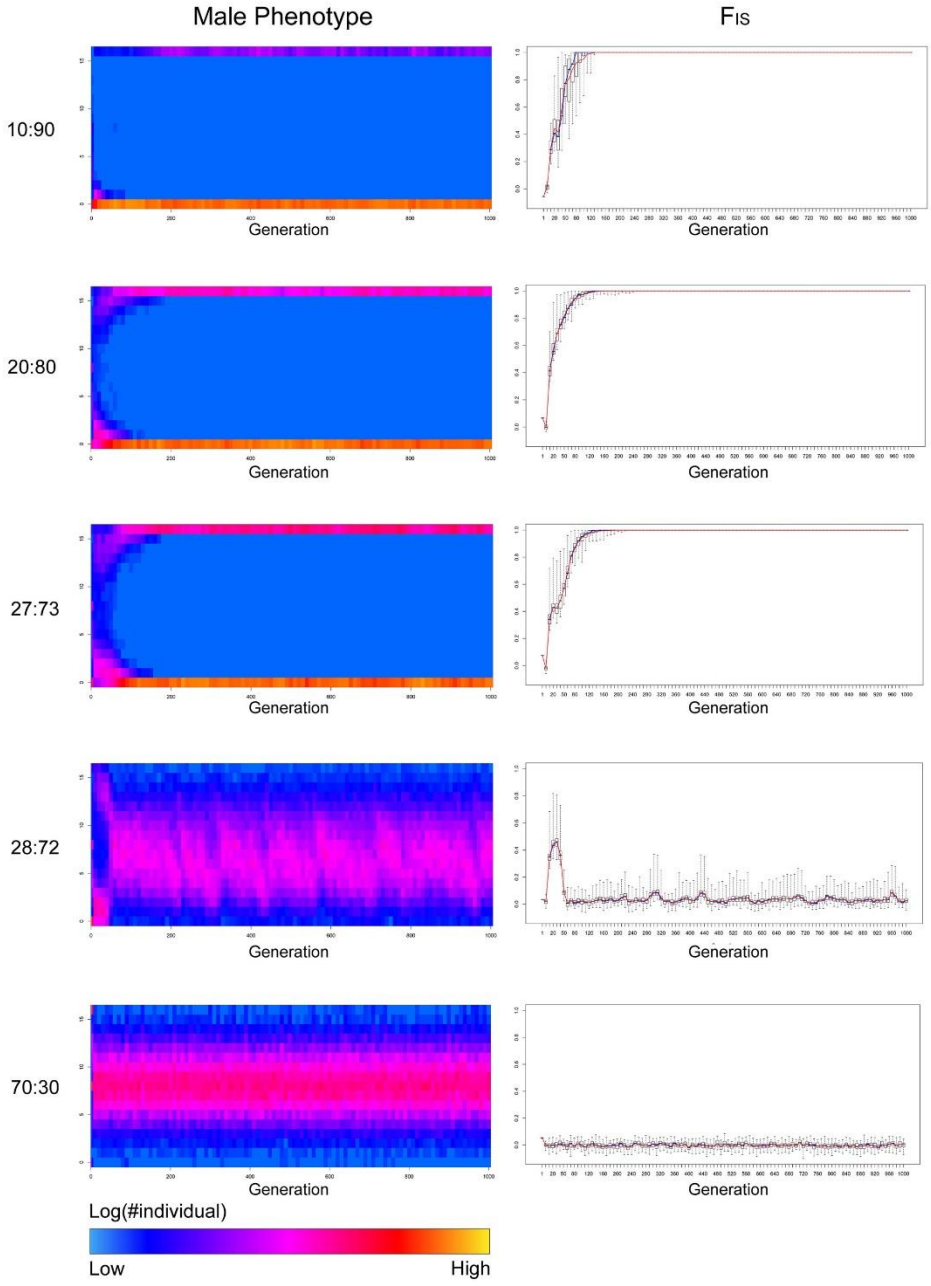


Figure 5. Reproductive isolation by mixing lineages with divergent Ψ's can be achieved when gene flow from the lineage with $\Psi < 0$: $\Psi > 0$ is less than 28:72 individuals per generation, in a hybrid population with carrying capacity of 2000 and assuming oblique imprinting. This limit is alleviated by physically linking loci underlying Ψ with the male signaling loci (Figure 7). Also see Figure B-8 for peer imprinting.

Xiphophorus species, for example, they lack many macromelanophore (Culumber 2013) and male-size category (Ryan et al. 1990) polymorphisms commonly found in lowland species with large population size. These conditions are predicted by my modeling to result in negative Ψ's.

Reproductive isolation between lineages with opposing Ψ's

I next asked if opposite directions of *per se* can maintain reproductive isolation between divergent taxa. I assumed that Ψ's have previously evolved to fixation in

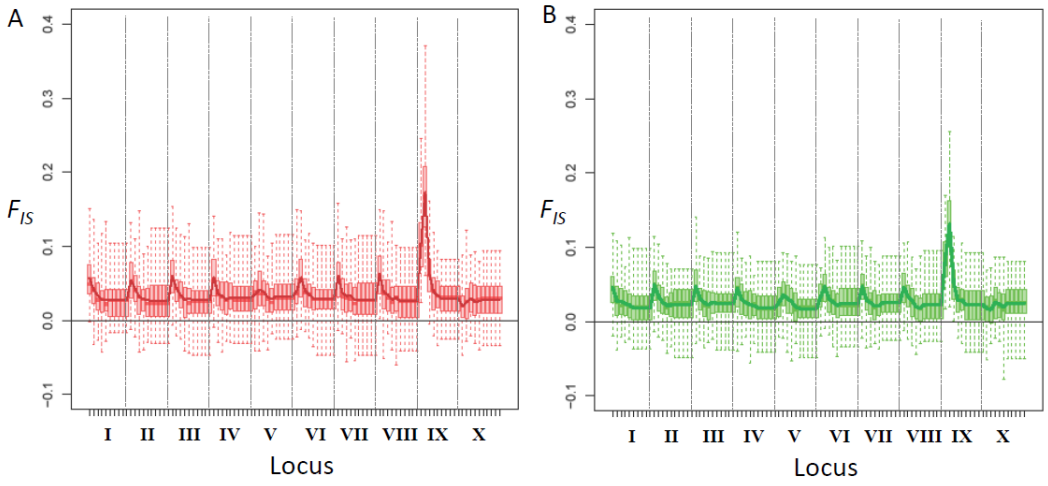


Figure 6. F_{IS} plotted for 100 neutral SNP markers on 10 chromosomes, summarized across generations 200-1000 for simulation 2. Admixture proportions A) Ψ < 0 : Ψ > 0 = 28:72; B) Ψ < 0 : Ψ > 0 = 40:60 (also See Figure 5). Both of these admixture proportions lack complete genome-wide isolation, but still result in elevated reproductive isolation compared to the null (F_{IS}=0). Notice that SNP markers closely linked to the 8 signaling loci on chromosomes 1-8 and the locus of Ψ on chromosome 9 show elevated F_{IS} values compared to the genomic background. Chromosome boundaries are marked with dotted vertical lines. When gene flow proportions exceeded this threshold, there was an abrupt shift towards extensive admixture, although genome-average F_{IS} was still slightly, but significantly different from 0 when the proportion < 50:50 (Figure 6), especially around the 8 signaling loci and the Ψ locus (Figure 6).

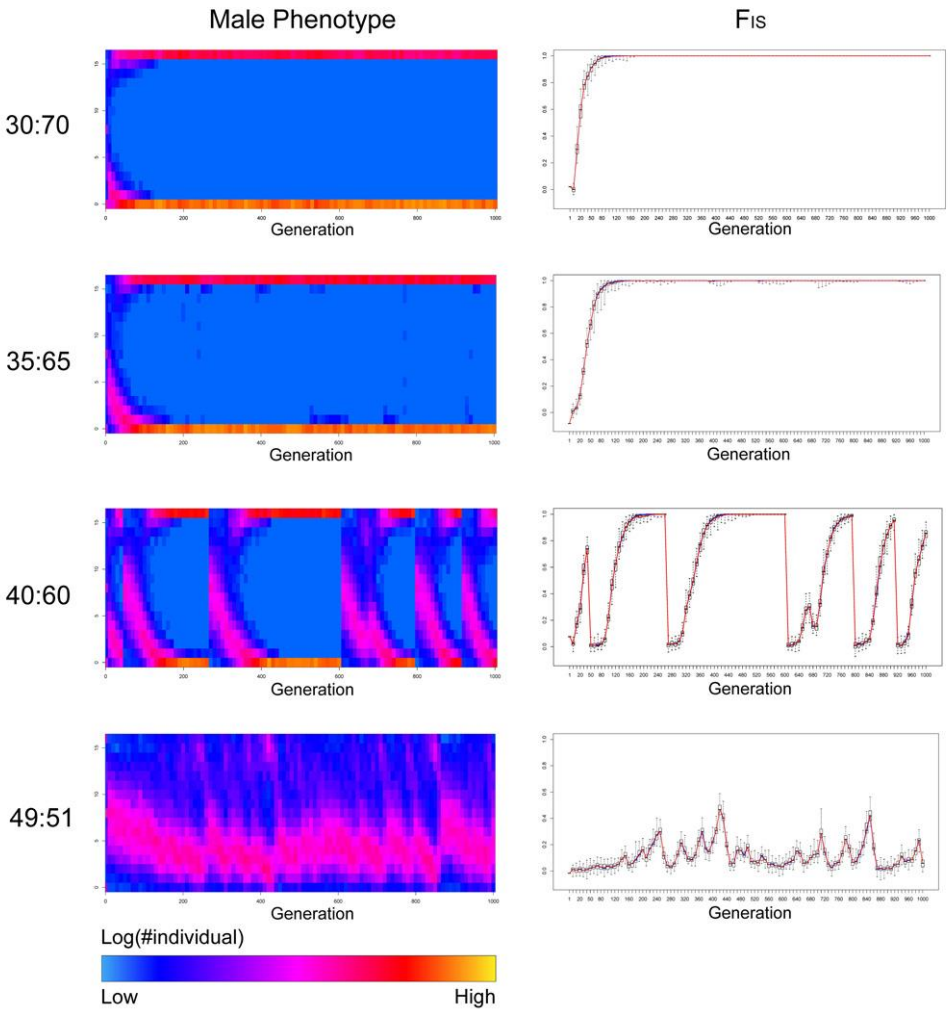


Figure 7. Physical linkage between loci underlying Ψ and the male signaling loci (Figure B-1) extends admixture proportions of lineages with $\Psi < 0$: $\Psi > 0$ until they start admixing. Note that partial reproductive isolation is still visible when proportions are close to 50:50.

parental species and mutation can be ignored. For every generation, I allowed 5% gene flow into the hybrid zone with varying ratios of two parental populations with Ψ_1=0.25 and Ψ_2=-0.25. The two parental forms were re-established genome-wide (F_{IS} fixed at 1) with sustained gene flow into the hybrid zone when $N_{\Psi>0}$: $N_{\Psi<0}$ <= 27:73 (or 1.35% : 3.65%, population size = 2000, Figure 5). Past the 50:50 threshold, admixture was

enhanced compared to the null, evidenced by the genome-wide excess of heterozygotes, resulting in negative F_{IS} values (Figure B-13A). The critical threshold of $N_{\Psi>0} : N_{\Psi<0}$ is caused by breakdown of linkage disequilibrium between the Ψ locus and the signaling loci. Indeed, when Ψ is encoded by additive QTLs tightly linked to signaling loci, complete reproductive isolation can be at least periodically achieved ($F_{IS} = 1$) up to 40:60 admixture proportions (Figure 7). Similar patterns also hold for peer imprinting (Appendix B iii; Figure B-8) and unimodal preference functions (Appendix B iv; Figure B-9). The critical threshold is slightly higher (32:68) assuming peer imprinting.

These results show that in certain scenarios, learned preference by itself can maintain reproductive isolation of interfertile species during secondary contact. One prediction from my admixture model is that there is more assortative mating when the hybrid index distribution is more biased towards the parental species with $\Psi>0$, while excess of heterozygosity should occur when it is biased towards the species with $\Psi<0$. My simulation predicts that few *X. malinche* in sympatry with more *X. birchmanni* would be unlikely to hybridize. On the other hand, if *X. birchmanni* were rare and *X. malinche* were common, hybridization should occur in both directions (Figure 5) and there should be a genome-wide excess of heterozygotes due to disassortative mating (Figure B-13).

Consistent with my model predictions, genotyping results using MSG (Andolfatto et al. 2011) show that a *X. birchmanni*-biased hybrid population (Rio Calnali-Mid, ~48% *X. malinche*) is highly structured with a bimodal distribution of hybrid indices, and a *X. malinche*-biased (Tlatemaco, ~80% *X. malinche*) population is

evenly admixed evidenced by a unimodal distribution of hybrid indices, inferred from high-throughput genotyping of ~150 individuals from each population (Schumer et al. In review; Figure S12). In Tlatemaco, about 4% of sites show significant excess of heterozygosity after accounting for potential bias in genotyping methods. This excess is absent from Calnali-Mid (Appendix B v; Figure B-13B). Furthermore, effectively unlinked loci showing significant linkage disequilibrium in Tlatemaco have much higher instances of heterospecific associations compared to Calnali samples, also suggestive of heightened outbreeding (Schumer et al. in review). The overall pair-wise REAP (Thornton et al. 2012) kinship coefficient of Calnali-Mid samples (0.09860) is higher ($t=4.385$, $d.f.=6411$, $p=0.00001$) than Tlatemaco (0.09305), consistent with higher level of disassortative mating in the latter (Appendix B v). Both of these populations were inferred to have admixed for > 150 generations based on ancestry block size and thus unlikely to reflect a non-equilibrium state (Appendix B v). Thus, I conclude that the puzzling population structures previously observed in different *X. birchmanni-X. malinche* hybrid zones can be partly explained by the learned preference behavior both observed in the lab and predicted by modeling.

Conclusions

In this study, I used simulations and empirical data to demonstrate that heterozygote viability has a dispositive impact on the evolution of Ψ, the hedonic value of learned social stimuli. I also demonstrated that divergent Ψ's can maintain reproductive isolation between sympatric species even in the absence of other isolating

barriers. My model may explain the evolution of sexual imprinting in a broad range of taxa, because it does not require females to possess any "preference loci" directly targeting specific male signals, and does not require parental care. The empirical example in a young sister species pair shows that opposing Ψ's have rapidly evolved in the directions predicted by my model. My study also highlights the potentially important effects of learned mate choice in explaining population-genetic patterns in the wild.

CHAPTER III

EARLY LEARNING TRIGGERS EXPRESSION CHANGES IN NEURAL PLASTICITY GENES AND ODORANT RECEPTORS IN *XIPHOPHORUS MALINCHE*

Introduction

Learned mate preference is the process through which early social experience affects mate choice decisions later in life (Grant and Grant 1997). Experienced with phenotypes can either positively influence mate preference, promoting assortative mating (Verzijden et al. 2005; Servedio et al. 2009); or negatively influence preference so as to increase outbreeding. Females can be genetically predisposed to prioritize imprinting on conspecific signals (Marler 1991), or may imprint on novel signals (Witte and Sawka 2003; Westerman et al. 2012). Sexual imprinting can strengthen reproductive isolation between species (Verzijden and ten Cate 2007; Verzijden et al. 2008; Kozak and Boughman 2009; Kozak et al. 2011) and can be an important mechanism in preventing hybridization between closely related species. In sticklebacks, where benthic and limnetic forms live in sympatry but are ecologically specialized to different microenvironments, sexual imprinting on odor and color is a major mechanism preventing hybridization between the two species (Kozak et al. 2011). In theory, learned preference can also retard signal-receiver divergence (Bailey and Moore 2012). In Darwin's finches, imprinting on male songs determines later song preference, and mis-imprinting on heterospecific songs promotes hybridization (Grant and Grant 2008);

similar patterns have been observed in a large number of species (reviewed in Irwin and Price 1999).

Mate choice often involves multiple sensory modalities (Rowe 1999), many of which are affected by learning (Vos 1995; Verzijden and Rosenthal 2011; Westerman et al. 2012). Previous studies have mainly focused on the signal-receiver co-evolution in the visual modality, which is likely under strong selective constraint due to the low number of photoreceptors involved in a wide range of tasks (Rosenthal and Ryan 2000). Not surprisingly, sensory drive, or the exploitation by male signals of female preferences that evolved in other contexts (Ryan 1990; Endler and Basolo 1998), has been most often associated with visual cues (Seehausen et al. 2008). On the other hand, the olfactory modality is thought to be more flexible during evolution as a result of the rich repertoire of odorant receptor genes found in vertebrate genomes (Amadou et al. 2003). Olfactory signals are used by a wide range vertebrate species in mate selection (Wyatt 2003), and can convey information including reproductive state, species, population and individual identities (Sorensen et al. 1990; Sorensen et al. 2005; Sorensen et al. 2011). Very often, olfactory preferences in vertebrates are also affected by learning (Plenderleith et al. 2005; Verzijden et al. 2008; Verzijden and Rosenthal 2011), providing further plasticity.

The mechanisms of olfactory learning likely involve both the sensory periphery (e.g. the olfactory epithelium) and higher processes in the central nervous system (CNS) (Wilson and Stevenson 2003; Lazarini and Lledo 2011). Vertebrate odorant receptor neurons (ORNs) express at least four families of olfactory receptors: ORs, V2Rs, V1Rs

and TAARs (Bazáes et al. 2013). Each ORN expresses a single olfactory receptor allele (Lewcock and Reed 2004). Axon projections of ORNs expressing the same olfactory receptor allele coalesce to the same glomerulus in the olfactory bulb (OB), forming a chemotopic map (topographical arrangements) of odorant features (Kermen et al. 2013).

Some evidence suggests that exposure to olfactory cues can result in "tuning" of the sensory periphery, or increased sensitivity of olfactory epithelum to different odors. Electrophysiological studies in salmon (Nevitt et al. 1994) and zebrafish (Whitlock et al. 2006) found that imprinting on artificially supplied chemicals during development altered sensitivity to these odorants later in life. In zebrafish, learned sensitivity to new odors was accompanied by increased expression of a gene involved in odorant receptor cell neurogenesis (*otx2*) (Whitlock et al. 2006).

Recent studies have shed some light on the mechanisms underlying sensitivity tuning of the olfactory epithelium. A study found that cells expressing an artificially disrupted OR are supplemented by another stochastically selected OR (Lewcock and Reed 2004). This suggests that maintenance of the allele-specific expression of odorant receptors involves a feedback loop that depends on functional activation by odor stimuli (reviewed in Shykind 2005). This model provides a natural mechanism for olfactory learning through tuning peripheral sensitivity to encountered odor cues. It is therefore likely that exposure to different odorants will elicit changes in expression of odorant receptors. In addition to sensitivity tuning at the olfactory epithelium, olfactory learning is also associated with adult neurogenesis in the OB (Lazarini and Lledo 2011). Plasticity on higher-order processes is also likely important. In fact, olfactory

discrimination is critically dependent on long-term memory in the piriform cortex (Wilson and Stevenson 2003), a homolog of the dorsal posterior part of fish telencephalon (Kermen et al. 2013). Behavioral outcome in response to social olfactory cues is partly mediated by neural plasticity in the *nucleus accumbens* (Kelz et al. 1999). Given that neural plasticity is involved in all levels of olfactory processing and learning (Wilson and Stevenson 2003), I expect that many genes related to these processes will be differentially expressed to facilitate olfactory learning.

In the swordtail fish species *Xiphophorus malinche* and *X. birchmanni*, learned mate choice is particularly important to understand because these two species naturally hybridize, and learned mate preferences may have significant impacts on the dynamics of hybridization. While *X. birchmanni* females learn to prefer both visual and olfactory cues of familiar males, *X. malinche* show the opposite pattern--*X. malinche* females select against familiar olfactory cues. This aversion has been attributed to the small population sizes of *X. malinche* which could favor outbreeding. Interestingly, female *X. malinche* show this pattern in response to olfactory cues whether they are exposed to conspecifics or heterospecifics. Regardless of the cause, this asymmetric learning is likely to have a major impact on reproductive barriers between *X. malinche* and *X. birchmanni*; preference for novel stimuli could drive *X. malinche* females to hybridize.

In the present study, I use an RNAseq based approach to investigate the mechanisms of sexual imprinting in the swordtail fish *X. malinche*. I chronically expose groups of juvenile *X. malinche* females to olfactory cues generated by conspecific and heterospecific adults, and compare gene expression between these groups and a control

group with no adult olfactory experience, and analyze gene expression in tissues that include the olfactory epithelium, the olfactory bulb and the telencephalon. My results indicate that major gene expression changes occur during olfactory learning in *X. malinche* when I compared between control and exposure, as well as females exposed to different species. Many of these genes are related to neural plasticity and provide clues into the mechanisms of learning to discriminate between species-specific stimuli.

Methods

Fish collection and exposure treatments

16 wild-caught female *X. malinche* from Chicayotla (Culumber et al. 2011) were allowed to drop fry in the lab. When fry reached approximately 2 months of age, families were evenly pooled and separated into 3 treatment groups: 1) M-EXP exposed to 2 pairs of adult *X. malinche* from the same population; 2) B-EXP exposed to 2 pairs of adult *X. birchmanni* from Garces (Culumber et al. 2011) and 3) Control did not receive adult stimulus. Exposure treatments were performed in adjacent 208L aquaria where adult stimuli and subject fry were divided by a perforated Plexiglas board. Maturing males from the cohort were isolated from the subject females as soon as I identified signs of formation of the gonopodium. At an average age of 11 months I tested their preference behavior to visual and olfactory cues of *X. birchmanni* and *X. malinche* (Cui et al. in prep). After each behavioral trial, females were rinsed and placed back to their original exposure tanks. After all behavioral trials were concluded, females were

restored to their respective exposure tanks for an additional 2 months before sample collection to minimize possible short-term effects from behavioral trials.

Sample collection

I randomly selected 4 females from each of the treatment groups (3 groups, 12 samples). Females were euthanized with an overdose of MS-222 and whole heads were preserved in Trizol solution and stored in -80 until use. Female bodies were dissected to ensure that females were virgins and showed no sign of parasites. Sensory tissue was dissected from the head by making a single 45-degree cut with dissection scissors in front of the anterior edge of the orbits. This tissue sample includes lower lip, tongue, upper lip, olfactory epithelium, the olfactory bulb and the telencephalon (Figure C-1).

RNA extraction and library preparation

RNA was extracted from the above tissue using a standard Trizol reagent protocol following manufacturer's instructions and quantified and assessed for quality on a BioAnalyzer. One micrograms of total RNA were used to prepare libraries following Illumina's TruSeq mRNA Sample Prep Kit with minor modifications. All libraries were prepared simultaneously. Briefly, mRNA was purified from total RNA using manufacturer provided beads. Following cDNA synthesis, mRNA was chemically fragmented and following end repair and A-tailing, samples were uniquely indexed by ligation. Libraries were PCR-amplified for 18 cycles and library size distribution and quality was verified on a Bioanalyzer 2100. Libraries were quantified on a Qubit fluorimeter, pooled in equal quantities, and sequenced on two Illumina HiSeq 2500 lanes (with 101 bp reads).

Adaptor and PCR primer sequences, and low quality bases in the raw reads were removed and trimmed by Trimnomatic with the following parameters: ILLUMINACLIP:princeton_adaptors2.fa:3:7:7 LEADING:20 TRAILING:20 SLIDINGWINDOW:4:20 MINLEN:80. Only reads >80 bp after filtering were kept for the downstream analyses.

Read mapping and de novo assembly

I use the previously described pseudogenome assembly for *X. malinche* at ~35X (Cui et al. 2013) coverage as the reference sequence for read mapping. First, I mapped pooled reads from all individuals using tophat to obtain a comprehensive alternative junction list. I then mapped reads for each individual sample separately guided by this junction list. I allowed three mismatches to the reference per read (3/101 bp) and used default settings for the other parameters. I also repeated the differential expression analyses with mapped reads using two mismatches with tophat, as well as Stampy with an expected divergence of 2% (Appendix C).

Because divergent paralogs of odorant receptors may not be mappable to the reference genome (scaffolds based on *X. maculatus*; approximately 2 million years diverged), unmapped reads were assembled with Velvet-Oases using a kmer size of 31 to discover possible odorant receptors. I filtered the final transcript isotigs such that only the longest isotig for each locus was kept. I used this filtered list to look for extra odorant receptors.

Odorant receptor identification and molecular evolution

I followed a workflow described in a previous study (Hashiguchi et al. 2008). A total of 25 V2R, 55 OR, 17 TAAR and 84 V1R sequences from teleost species were downloaded from genbank as the probe sequences. They were translated into amino acid sequences, aligned by Muscle invoked from MEGA 5, and translated back to in-frame nucleotide sequences. An HMM model was built with this initial amino acid alignment with hmmer 2.3.2. Genewise 2.2.0 is able to identify full coding sequences in correct open-reading frames (ORFs) given the HMM model and a genomic or transcriptome sequence. Because the algorithm is intractable on full genomes or transcriptomes, I first performed discontiguous megablast (megablast-dc) using the initial probe sequences, collapsed overlapping hits, and extracted the hit region plus 5kb 5'- and 3'- flanking regions. These extracted sequence fragments were used as inputs for Genewise. Inferred coding sequences were then added back to the alignment, realigned to the correct reading frames, and used as probes and HMM inputs for the next iteration, repeating the above procedure. I iterated until no more new V2R, OR, TAAR or V1R sequences were identified by Genewise. In the *de novo* assemblies, only the longest isotig was retained in the final dataset.

I extracted orthologous odorant receptor sequences from pseudogenomes of 5 *Xiphophorus* species using the exon structure identified by Genewise. The pseudogenomes of these species have the same coordinate system as the reference *X. maculatus* genome, so realignment is unnecessary. Each gene alignment was then checked for premature stop codons, in which case they were discarded from further

analysis. I then inferred gene trees for each odorant receptor gene for the 5 species using GTR+Gamma model with RAxML. The ML tree was used in Codeml to test for positive selection, using a likelihood ratio test between the M8 and the M8a models (d.f. = 1) or M8-M7 (d.f.=2). I also reconstructed phylogenies for each of the odorant receptor families using RAxML. I were not able to perform these analyses in receptors resulted from *de novo* assembly because transcriptome data are not available for other species.

Differential expression analyses

The gene models for *X. maculatus* were downloaded from Ensembl and the coordinates were translated into the current genome assembly version using a custom script. Because the current version of the gene models are incomplete for odorant receptors in *X. malinche*, I added entries into the GTF file for the V2R, OR, TAAR and V1R receptors identified above. With this updated GTF file, I counted the number of reads to each gene using a python package htseq-count, requiring a mapping quality (MAPQ) of 30 and allowed any reads partially overlapped with the feature to be counted. These raw counts were imported into the edgeR package for differential expression (DE) analyses. After raw read counts were imported into edgeR, genes were only kept for further analyses if the read counts per million reads (cpm) exceed 0.5 in ≥4 samples. I treat the genes that do not pass this initial filter effectively unexpressed. First, edgeR calculates normalization factors for each sample using the TMM method (Robinson and Oshlack 2010) to account for library size variations. Second, maximum conditional likelihood is used to estimate the common dispersion parameter for all genes (Robinson and Smyth 2008). Third, the gene-wise dispersion parameters are estimated

by an empirical Bayes method (Robinson and Smyth 2007). Finally, an exact test on the means between two groups of normalized count values (negative binomially distributed) was used to assess differential expression (Robinson and Smyth 2008).

I performed two sets of analyses for differential expression by treatment 1) restricting my analysis to known candidate genes and 2) a transcriptome-wide analysis. For my candidate gene analysis, I extracted raw *p* values of transcripts annotated as members of the V2R, OR, TAAR or V1R gene families from edgeR's global exact test. I then calculated q-values using the R package qvalues for all the raw *p* values for these genes. Three comparisons were made: B-EXP to M-EXP, B-EXP to Control and M-EXP to Control.

GO enrichment analysis

To determine whether particular functional categories were enriched in genes differentially expressed between the treated and control groups and the *birchmanni*-treated and *malinche* treated females, I performed Gene Ontology (GO) analysis. I used the annotated *X. maculatus* genome to assign HUGO gene symbols. HUGO genes symbols were matched to GO categories using the Bioconductor human gene ontology database (http://www.bioconductor.org/packages/2.13/data/annotation/html/org.Hs.eg.db.html). A custom GO database was built using the GOstats package in R. All genes that passed coverage filtering in edgeR (see above) were included as part of the gene universe and I tested for significant enrichment of different functional categories (biological processes, cellular component, molecular function) by comparing the gene universe to the significant genes lists using the GOstats and GSEAbase

packages in R with a hypergeometric test (Falcon and Gentleman 2007). I use Revigo (http://revigo.irb.hr/) to visualize GO categories, clustered by semantic similarities.

Results

Read mapping and de novo assembly

Following quality trimming, an average of 24 million reads were retained per individual, and 81.3% of these were mapped by TopHat (Table C-1). In addition to the 115 OR, 71 V2R, 58 TAAR and 2 V1R sequences present in the *X. maculatus* genome assembly, I identified 8 OR, 8 V2R and 2 TAAR sequences by *de novo* assembly of unmapped reads. Among the receptors identified from the genome, 9 OR (7.8%), 48 V2R (67.6%), 8 TAAR (13.8%) and 0 V1R (0%) pass initial coverage filtering. All *de novo* assembled receptors passed initial filtering, leaving a total of 17 ORs, 56 V2Rs and 10 TAARs.

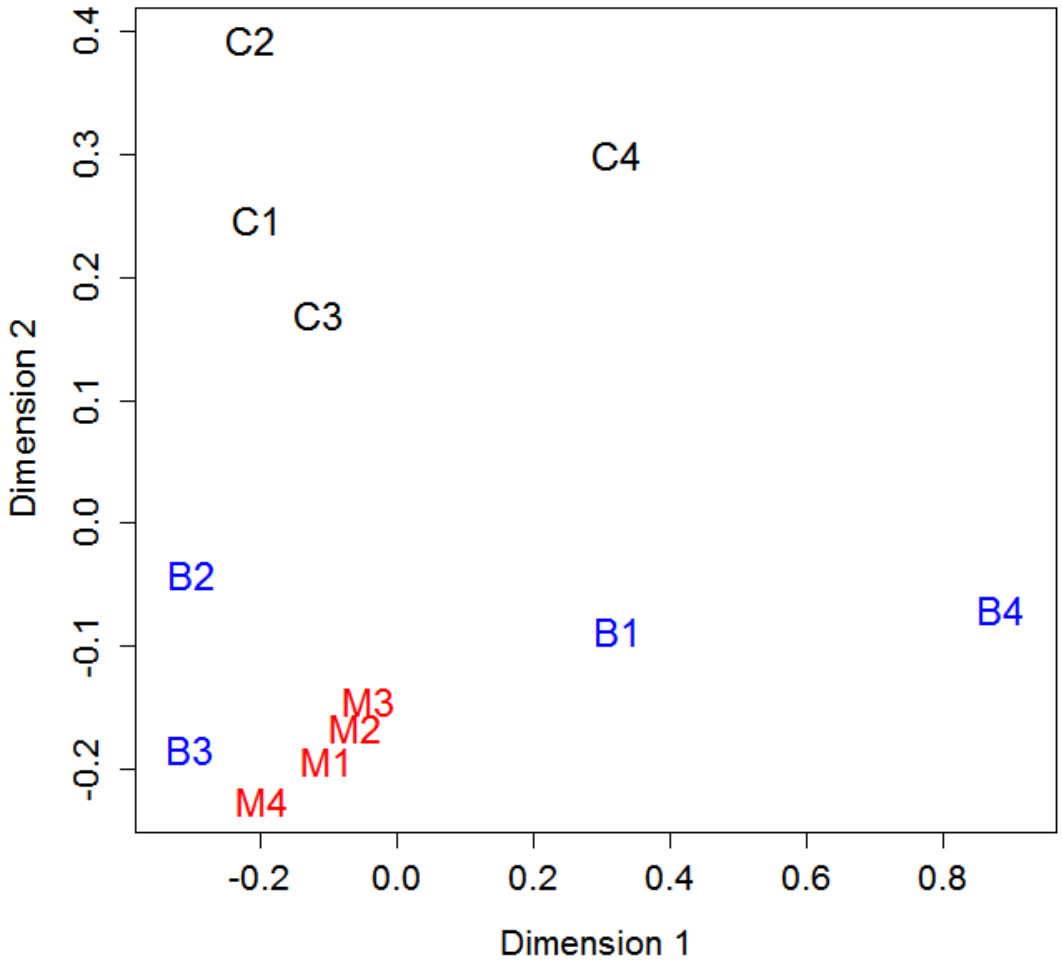


Figure 8. MDS plot of gene expression. Dimensions 1 & 2 for the 1000 most informative genes generated by edgeR. Dimension 2 shows clear separation in gene expression patterns between treated and control individuals.

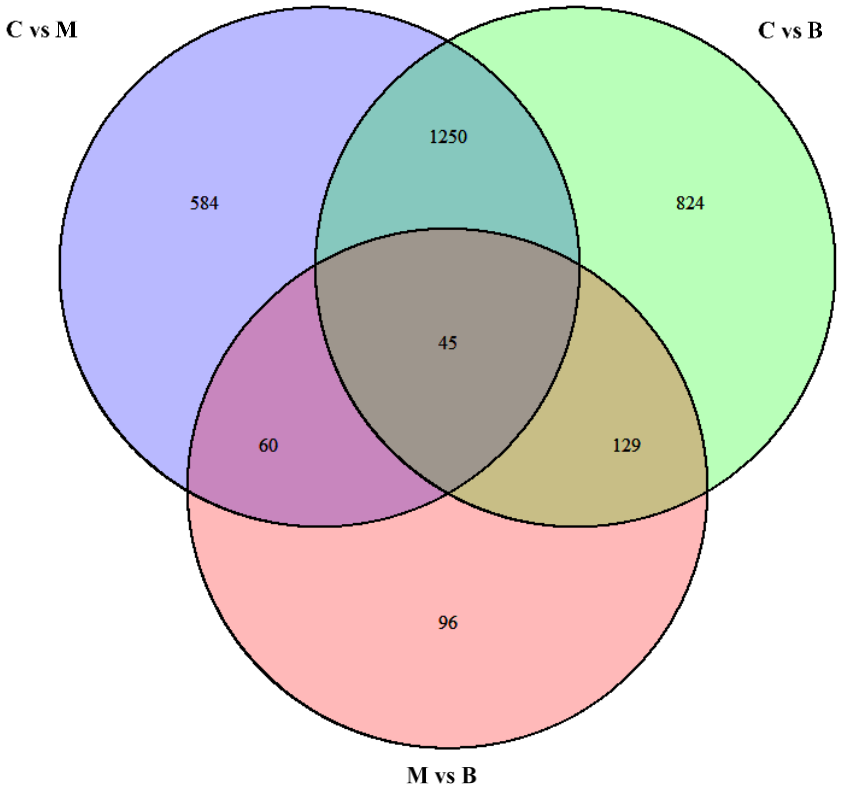


Figure 9. Venn diagram showing overlaps of differentially expressed genes between three comparisons. 1) C vs M: Control vs. *malinche*-exposed 2) C vs B: Control vs. *birchmanni* exposed and 3) M vs B: *birchmanni* vs. *malinche*-exposed.

Differential expression

Social environment had a dramatic effect on gene regulation (Figure 8). After coverage filtered with edgeR 17,765 genes were retained. Control group clusters separately compared to the two exposure groups (Figure 1). After false discovery correction at FDR=0.05, 2248 and 1939 of these were significantly differentially expressed between the control group and *birchmanni* and *malinche* stimulus treated groups respectively (Table C-2). Most genes differentially expressed between the treated and control groups were consistent; 58% of genes differentially expressed in the

birchmanni treated group were shared with the *malinche* treated group and 67% of genes differentially expressed in the *malinche* treated group were shared with the *birchmanni* treated group (Figure 9; Figure 10). Of the 1279 genes with shared significance between both treated groups, 99% had the same direction of expression relative to the contol (e.g. up or downregulated). Three hundred and thirty genes were significantly differentially expressed between the *malinche* and *birchmanni* treated stimulus groups (Figure 2; Table C-2).

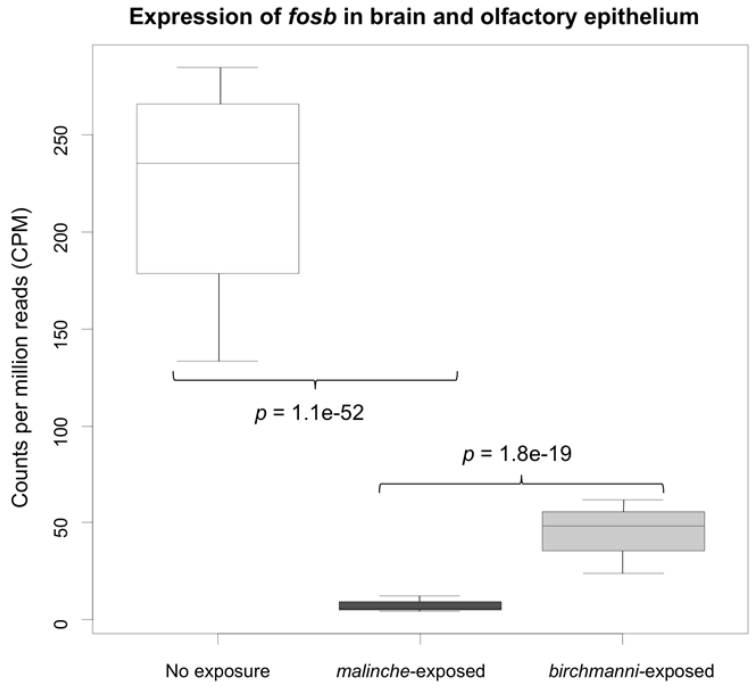


Figure 10. Boxplot showing expression levels of *fosb* (in counts per million reads) by group. *fosb* is strongly differentially expressed by treatment and by group. P-values indicated are corrected for multiple testing.

Differential expression of candidate genes

I specifically assessed expression of olfactory learning candidate genes. Fifteen V2R receptors were significantly differentially expressed between *malinche* and

birchmanni treated females, though none of these were still significant following multiple testing corrections. Six V2R receptors were significantly differentially expressed between the control and *malinche* treated individuals, and 5 ORs were differentially expressed, two of which survived FDR correction. In the control vs *birchmanni* treated comparison 2 ORs, 1 TAAR, and 2 V2Rs were differentially expressed; 1 V2R survived FDR correction.

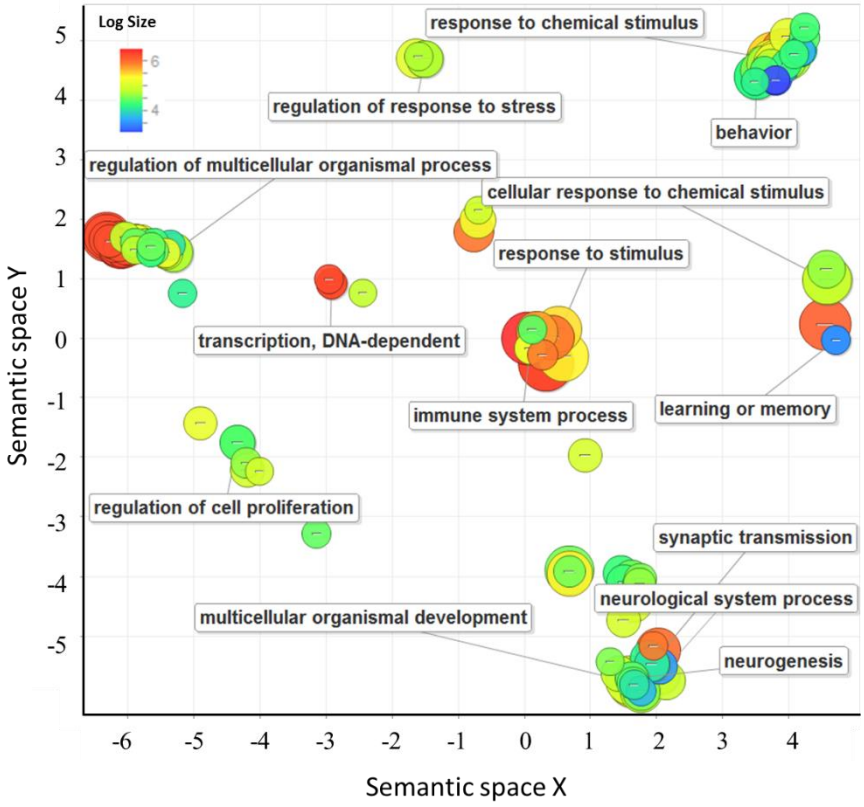


Figure 11. Gene ontology terms of differentially expressed genes for the *birchmanni*-, *malinche*-exposed comparison. Visualized with Revigo (http://revigo.irb.hr/) using the SimRel (Schlicker et al. 2006) term similarity index. Size of circle represents –log(p value), log(size) refers to sub-categories contained within the category.

Gene ontology analysis

To investigate whether particular functional categories were enriched in the differentially expressed genes, I performed Gene Ontology analysis. In genes differentially regulated between *X. birchmanni*- treated and *X. malinche*-treated females, response to chemical stimulus (GO:0042221) was the most significantly enriched biological process category (p=6.4e-06; Figure 11). In total, 60 GO categories were significantly enriched at p<0.01 between the two olfactory treatments (Table C-3). Between treated and control individuals, 4 and 26 GO categories were significantly enriched for *malinche* and *birchmanni* treated groups respectively (Table C-4 & Table C-5).

Functional divergence in differentially expressed odorant receptors

One hypothesis for the difference in olfactory responses between *X. birchmanni* and *X. malinche* females is that the genes involved in the olfactory periphery have functional divergence between species. The V2R family is best represented in the transcriptome within the four OR families (Figure 12; Figure 13). Using orthologous sequences from 5 *Xiphophorus* species, representing all 3 major clades, I found signature of positive selection or relaxed selection on 4 V2R receptors (Table C-7; M8-M7 $p < 0.05$, M8-M8a $p < 0.02$), 1 OR receptor (M8-M7 $p < 0.05$, M8-M8a $p < 0.01$) and 5 TAAR receptors (M8-M7 $p < 0.05$, M8-M8a $p < 0.02$). Both V1R receptors are conserved in the phylogeny (M8-M7 $p > 0.4$, M8-M8a $p > 0.1$).

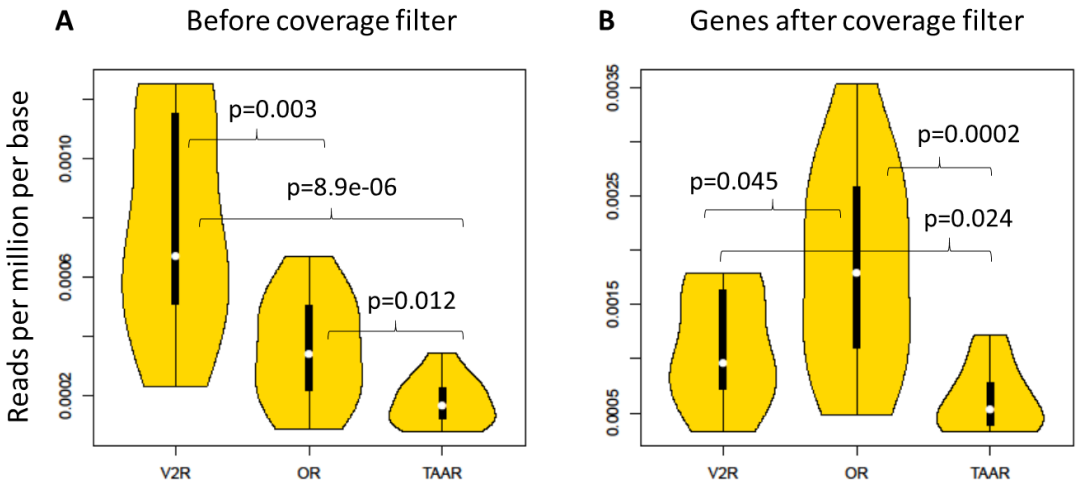


Figure 12. Odorant receptor families have different expression levels on the olfactory epithelium in *Xiphophorus malinche*. Variations in 12 female samples are represented as violin plots normalized to reads per million reads divided by the summed gene length of all included genes in that gene family. A) all identified odorant receptor genes are included regardless of their coverage level. (82 V2Rs, 126 ORs and 61 TAARs) B) only odorant receptor genes that passed the initial coverage cutoff were included. (56 V2Rs, 17 ORs and 10 TAARs). Note that two V1R receptors have no reads in any samples.

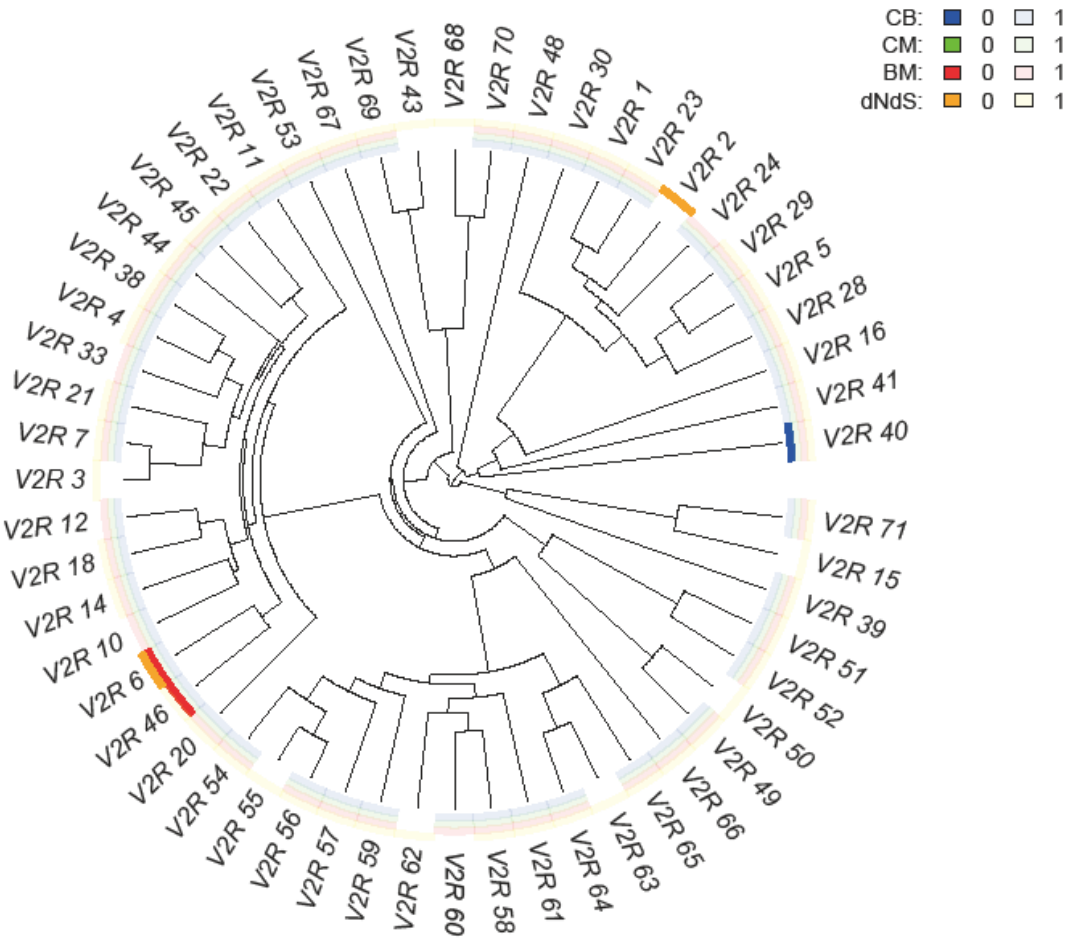


Figure 13. Differential expression and signature of positive selection plotted on a V2R phylogeny of 5 *Xiphophorus* spp. Because all orthologs are monophyletic for the 5 species, the tips are pruned to leave a single representative gene. Dark colors indicate significant tests at α=0.01, light colors indicate that the gene passed initial filtering criteria but the test statistic >0.01. Blank slots mean that the gene did not pass initial filtering (for dN/dS it's due to premature stop codon in at least 1 of the 5 species, for Differential expression (DE) it's due to low coverage in the transcriptome). Blue (1st ring) – DE between control and birchmanni-exposed; Green (2nd ring) – DE between control and malinche-exposed; Red (3rd ring) – DE between birchmanni and malinche exposed; Orange (outer-most) – signature for elevated dN/dS ratio tested by likelihood ratio test with 5 species for that gene using M8-M8a models in Codeml.

Discussion

Social exposure alters gene expression related to synaptic plasticity

Gene Ontology analysis clearly demonstrated an enrichment of genes involved in response to chemical stimulus in all three comparisons, confirming that olfactory learning from adult cues leaves a signature on the cellular level (Figure 11). The gene expression response to adult olfactory stimulus is dramatic (Figure 1), while many fewer genes were differentially expressed between the two exposure treatments. Though male pheromone components are uncharacterized in *Xiphophorus,* in other fish species pheromones contain metabolites which indicate species identity, reproductive state, relatedness, and male condition, among other information (Dulka et al. 1987; Sorensen et al. 1990; Sorensen et al. 2005; Sorensen et al. 2011). Thus, a strong gene expression response to male pheromone exposure likely reflects olfactory response to these stimuli. Interestingly, much of the gene expression response I document is likely to be a general response to olfactory learning since groups treated with conspecific and heterospecific stimuli shared a striking proportion of differentially regulated genes, and >99% of these showed the same direction of expression relative to the control.

Interestingly, I found enrichment of GO terms related to neurogenesis and synaptic transmission between the conspecific and heterospecific olfactory treatments. Studies of auditory learning in white crown sparrows have demonstrated that learning is sometimes biased towards conspecifics, or the so-called “instinct to learn” (Marler 1991). Mechanistically, olfactory learning requires reorganization of neuron circuitry in both the olfactory bulb and the piriform cortex (Fletcher and Wilson 2003; Wilson and

Stevenson 2003). Therefore, enrichment of genes suggests that early olfactory learning in *Xiphophorus* triggers a different level of neural plasticity depending on whether the cue is conspecific or heterospecific. Though behavioral research demonstrates that females can distinguish between con- and heterospecific cues (Crapon de Caprona and Ryan 1990; McLennan and Ryan 1999), it is unknown how this information is encoded in the pheromone cue (e.g. distinct ligands or distinct ratios). While my results support that the different cues produced by the two species of *Xiphophorus* elicit different gene expression responses in *X. malinche* females, particularly in relation to genes involved in neural repatterning, I do not know yet if the difference is due to ratio/concentration or structurally different ligands. Though pooled tissues were used in this library including the olfactory epithelium, the OB and the telencephalon, future experiments will focus on localizing these enriched genes in the brain to verify whether candidate brain regions involved in olfactory learning show distinct responses to conspecific versus heterospecific olfactory cues.

Although less GO enrichment was found between the control group and the two exposure groups, this is likely due to a much longer list of differentially expressed genes (~2000 out of 17,765 genes vs. 330). When limiting to fewer genes using a more stringent FDR cutoff at 0.01, more GO terms were found to be enriched (Table C-4 & Table C-5). Interestingly, GO analysis did suggest that there is significant enrichment of immune response genes in the two experimental groups compared to the control groups. It is possible that the cues of adult males trigger female defense responses as has been documented in other species (Lawniczak and Begun 2004). This may be a result of high

levels of sexual conflict in poeciliids due to internal fertilization and sperm competition (Pilastro et al. 2003).

Highly significant DE in fosb

Among the three pair-wise comparisons, *fosb* was strongly differentially expressed. The control group had the highest level of *fosb*, followed by the heterospecific-exposed group and then the conspecific-exposed group (Figure 10). A previous study demonstrated that fosb had decreased expression in the accessory olfactory bulb in *Rattus* mothers exposed to pups (Canavan et al. 2011). In mammals the accessory olfactory bulb receives signals from the pheromone-detecting cells that express V1R and V2Rs, suggesting that *fosb* is involved in plastic responses to social cues. In the CNS, increased fosb expression mediates long-term plasticity. ΔFosB, a stable isoform of *fosb* (Ulery et al. 2006), is thought to reinforce addictive behavior (Nestler et al. 2001) byincreasing sensitivity to the rewarding effects of drugs after chronic exposure in the *nucleus accumbens* (Kelz et al. 1999) and promotes drug seeking behavior (Nestler et al. 2001). In other rodents, ΔFosB overexpression in the same brain region is also associated with sexual reward (Hedges et al. 2009; Pitchers et al. 2010). ΔFosB expression is also associated with plasticity and learning in other brain regions, for example, it is activated by olfactory cues associated with social defeat in mice in the prefrontal cortex (Bourne et al. 2013).

Long term memory storage is regulated by the CREB pathway (Silva et al. 1998). The CREB protein regulates genes that contain CREB-binding sites (cAMP response-element), which includes *fosb* among other genes such as *dusp1*, *egr1* and *crem* (Wood

et al. 2006). Intriguingly, *dusp1* and *egr1* are also differentially expressed in all three comparisons, while *crem* is differentially expressed in B-M and C-M comparisons. All three genes show the same expression patterns as *fosb*, where C>B>M (Figure C-3). This suggest that downstream genes of the CREB pathway (Silva et al. 1998) are co-regulated in different social exposure treatments.

Two mutually non-exclusive hypotheses could explain the observed expression pattern in *fosb*. First, *fosb* down-regulation may be associated with exposure to social cues in the olfactory bulb (Canavan et al. 2011), and the conspecific cue may have resulted in the strongest down regulation. Second, if the down-regulation happens in the CNS, it may be explained by the fact that *X. malinche* differs behaviorally from *X. birchmanni* in that females learn to disdain familiar olfactory cues (Verzijden et al. 2012a). An intriguing possibility is that this species-specific learned avoidance is associated with the depression of *fosb* level after social exposure. Future experiments in the sister species *X. birchmanni*, in which females prefer familiar phenotypes, will allow me to test these hypotheses.

V2Rs as candidates for species-specific cue detectors

Traditionally vomeronasal receptor families – V1R and V2R—have been associated with the pheromone detection (Rodriguez et al. 2000). Recently, studies have suggested that the main olfactory pathways (expressing canonical ORs) may also detect certain components of sex pheromones. Despite V1R being suggested in pheromone detection in fish (Pfister and Rodriguez 2005), I did not detect expression of this gene family. To the contrary, I detected expression of a large number of V2R receptors mined

from the *Xiphophorus* genome and *de novo* transcriptome. Per-base expression level of V2R and OR are only marginally different (OR>V2R, Wilcox sign rank p = 0.045; Figure 12B). This suggest that the olfactory epithelium of *X. malinche* is contains a larger variety of V2Rs and expresses fewer paralogs of ORs, each at approximately the same level. V2R receptors are of particular interest because they have been implicated in pheromone detection in both mammals and fish. V2Rs are co-expressed with MHC (Dulac et al. 2006) and detect MHC-1 ligands (Leinders-Zufall et al. 2004), and in general they respond to proteinaceous pheromone components in mice (Chamero et al. 2007) or amino acids in fish (Speca et al. 1999). Interestingly, all 7 differentially expressed odorant receptors between conspecific vs. heterospecific treatments are V2Rs (raw $p<0.02$, $q=0.09$, expected number of false positives: 0.63). The most differentially expressed V2R, V2R_6 (raw $p=0.0056$) also has the highest rate of molecular evolution (M8-M8a, $p=0.0013$) among 5 *Xiphophorus* species. This paralog has the highest expression level in conspecific exposure treatment, lowest in heterospecific treatment and intermediate in controls. There are 13 (at least 2 unique between *X. birchmanni* and *X. malinche*) amino acid substitutions between *X. birchmanni* and *X. malinche* sequences at this gene, 9 (2 unique) of which are located on the N-terminus (Figure C-4), where the putative ligand-binding domain is located in V2Rs (Yang et al. 2005) Together, these results suggest that the V2R paralog may detect a species-specific component, and conspecific cues, presumably co-evolving with the receptor, may most effectively increase receptor expression level due to high affinity. Other differentially expressed V2Rs in this comparison do not show signature of positive selection, which could

suggest that ratio differences in certain chemical components also convey species-specific information (Sorensen et al. 2011).

Differentially expressed odorant receptors in control vs. exposure comparisons are all structurally conserved in *Xiphophorus*. This observation is consistent with the idea that certain components in the sex pheromone blend are chemically conserved, and information is conveyed through ratio or concentration variations. For example, chemical components signaling for sex, reproductive and nutritional state are likely derivatives of intrinsic hormones or metabolites (Sorensen et al. 1990), and are thus likely to be structurally conserved.

Conclusions

My study showed that olfactory learning in *X. malinche* evokes strong gene expression responses in the sensory periphery and possibly even the brain regions involved in higher-order processes. Specifically, in addition to presence or absence of adult cues, species-specific cues can also alter neural plasticity. Genes related to long-term memory, such as fosb, egr1 and dusp1, show depression after conspecific exposure. Finally I suggest that olfactory learning can alter sensitivity at the periphery and I identified a candidate V2R paralog that potentially detects species-specific odorant.

CHAPTER IV

PHYLOGENOMICS REVEALS EXTENSIVE RETICULATE EVOLUTION IN *XIPHOPHORUS* FISHES[*]

Introduction

A growing body of work has begun to recognize the importance of hybridization in the evolutionary process (Mallet 2007, 2008; Abbott et al. 2010; Arnold and Martin 2010). There is increasing awareness that many species diversify in the presence of ongoing gene flow, or experience post-speciation gene flow without the collapse of reproductive isolation. Some of the most rapidly diversifying groups, such as African cichlids (e.g. Seehausen 2004; Schwarzer et al. 2012) and *Heliconius* butterflies (Mallet 2005), have weak post-zygotic isolation and frequently hybridize. In certain species groups, the spread of adaptive alleles through hybridization is thought to underlie phenotypic diversification (Rieseberg et al. 2003; Heliconius Genome Consortium 2012). It has also been suggested that hybridization can lead to speciation in some cases (Mallet 2007, 2008; Abbott et al. 2010; Arnold and Martin 2010). Understanding the extent of hybridization, and the role of hybridization in evolution, is an important focus of current evolutionary research.

---

[*] Reprinted from Cui, R., M. Schumer, K. Kruesi, R. Walter, P. Andolfatto, and G. G. Rosenthal. 2013. Phylogenomics reveals extensive reticulate evolution in *Xiphophorus* fishes. Evolution 67:2166-2179 with permission from the publisher.

Large, genome-wide datasets can be used both to address phylogenetic relationships between species and examine patterns of incomplete lineage sorting and hybridization. Historically, it has been difficult to determine phylogenetic relationships in groups with hybridization or high levels of incomplete lineage sorting (reviewed in Degnan and Rosenberg 2009), but the availability of genome-wide data and new computational techniques (An é et al. 2007; Kubatko 2009; Pickrell and Pritchard 2012) allows researchers to explicitly account for incomplete lineage sorting and hybridization when constructing species trees (Pollard et al. 2006; Cranston et al. 2009). Previous studies have shown that next-generation sequencing data has the potential to generate well resolved gene trees for testing hypotheses of hybridization (Hittinger et al. 2010). Despite this, very few phylogenetic studies to date have examined hybridization, particularly in a large group of species. Many phylogenetic studies that have examined hybridization have not explicitly accounted for incomplete lineage sorting (Decker et al. 2009; Schwarzer et al. 2012), used candidate genes (Heliconius Genome Consortium 2012), small numbers of genes (Hailer et al. 2012), cyto-nuclear discordance (Meyer et al. 2006; Aboim et al. 2010; Kang et al. 2013), or been unable to investigate hybridization due to the use of concatenated datasets (Nabholz et al. 2011; dos Reis et al. 2012).

Swordtails and platyfishes (Poeciliidae, genus *Xiphophorus*) are longstanding models of sexual selection (Darwin 1859; Ryan 1990), evolutionary genetics (Schartl 1995; Basolo 2006), and oncology (reviewed in Meierjohann and Schartl 2006). Numerous phylogenetic studies have failed to reach consensus on many aspects of their

evolutionary relationships (Rosen 1960; Rosen 1979; Rauchenberger et al. 1990; Haas 1992; Marcus and McCune 1999; Morris et al. 2001; Meyer et al. 2006; Kang et al. 2013). Early studies placed the swordtails as a monophyletic group derived from the more basal platyfishes (Figure 14 A). More recent studies have proposed conflicting topologies for three clades: northern swordtails, southern swordtails and platyfishes (northern and southern;(Figure 1B & 1C, Meyer et al. 1994).

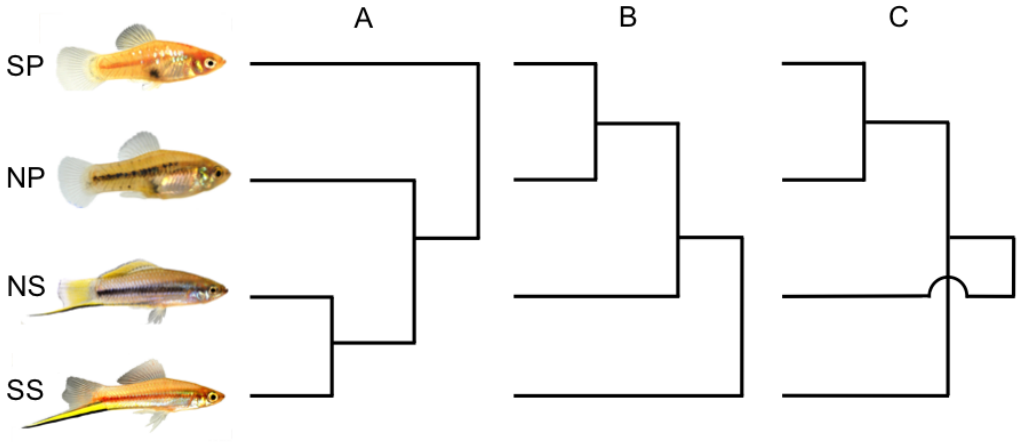


Figure 14. Previous phylogenetic hypotheses for interclade relationships in *Xiphophorus*. Abbreviations: NP—northern platyfishes, SP—southern platyfishes, NS—northern swordtails, SS—southern swordtails. A) Platyfishes are paraphyletic and basal to genus (Rosen, 1979; Rauchenberger, 1990; Basolo, 1990b). B) Swordtails are paraphyletic and northern swordtails are grouped with platyfishes (Meyer et al., 2006). C) Swordtails are paraphyletic and southern swordtails are grouped with platyfishes (Meyer et al., 1994).

While the inconsistency among phylogenetic studies may be the result of insufficient data or errors in inference, another likely cause is interspecific hybridization. *Xiphophorus* species have weak postzygotic isolation (Morizot et al. 1991; Walter et al. 2004b) and there are contemporary natural hybrid zones between multiple pairs of

species (Rosenthal et al. 2003; Schartl 2008). Premating barriers play a primary role in reproductive isolation between species (Schartl 2008). As a rule, females prefer to mate with conspecifics over heterospecifics (McLennan and Ryan 1999; Hankison and Morris 2003), but these preferences are susceptible to ecological disturbances and plasticity in preferences (Fisher et al. 2006; Verzijden and Rosenthal 2011; Willis et al. 2011; Willis et al. 2012). It is therefore likely that episodes of hybridization have played an important role in *Xiphophorus* evolution and may account for conflicting phylogenies.

Uncertainties about the *Xiphophorus* phylogeny have also contributed to controversies about trait evolution in the genus. *Xiphophorus* has been intensively studied in the context of the preexisting bias hypothesis of sexual selection (Basolo 1990; Rosenthal and Evans 1998). The sexually-dimorphic "sword" extension of the caudal fin in males of some *Xiphophorus* species was proposed to have evolved in response to a latent bias in females, but this hypothesis hinged on the basal placement of platyfishes and monophyly of the swordtails (Basolo 1990). This view was challenged by subsequent phylogenies that did not support the monophyly of northern and southern swordtails (Meyer et al. 1994); but see (Basolo 1995b). Without a robust phylogenetic framework, it is difficult to draw inferences about whether or not preexisting preferences drove the evolution of the sword ornament in *Xiphophorus*.

Hybridization could also account for the discontinuous distribution of traits in the genus. For example, the geographically separated *X. andersi* and *X. xiphidium* are the only two platyfishes naturally expressing swords (albeit short and unpigmented). Also, the sympatric swordtail *X. birchmanni* and platyfish *X. variatus* (Fisher et al. 2009)

share several male sexually-dimorphic traits, despite no evidence of current hybridization (Rauchenberger et al. 1990; GGR, personal observation). Though these shared male sexually dimorphic traits could be the result of convergent evolution, interfertility between species raises the possibility that gene flow may be responsible for some shared traits.

To evaluate the role of hybridization in shaping the interrelationships among *Xiphophorus* species, I use an RNA-seq-based approach to collect transcriptome sequence data for 24 of the 26 described *Xiphophorus* species and two outgroups. Despite their increasing use in population genetics and gene mapping studies, next-generation sequencing techniques have had limited applications in phylogenetics thus far. Though a variety of next-generation sequencing methods can be used generate phylogenetic datasets, for recently diverged species such as *Xiphophorus* (~4-6 mya; Mateos et al. 2002; Kallman and Kazianis 2006), techniques such as RNAseq that generate long alignments are promising (e.g. Kocot et al. 2011; Smith et al. 2011). Though RNA-seq datasets have been concatenated to produce highly supported phylogenies (Hittinger et al. 2010; Nabholz et al. 2011), concatenated datasets cannot be used to investigate gene flow. I use RNA-seq data to resolve the *Xiphophorus* species tree with high confidence and apply phylogenetic approaches to identify cases of interspecific hybridization. I find evidence for widespread hybridization in *Xiphophorus* and discuss my findings in the context of sexual selection and trait evolution.

Methods

RNA extraction and library preparation

I obtained a single live male of each of 24 *Xiphophorus* species and two outgroups from a variety of sources (Cui et al. 2013, Table S1). Fresh specimens were photographed and killed with an overdose of tricaine methanesulfate (MS-222) in compliance with Texas A&M University IACUC protocol # 2012-164. Whole brains were dissected immediately and preserved in RNAlater (Ambion, Austin, TX) at -20˚C until use.

RNA was extracted using a standard TRI Reagent protocol (Ambion, Austin, TX) following manufacturer's instructions and quantified with a Nanodrop 1000 (Thermo Scientific, Wilmington, DE). 1-4 μg of total RNA was used to prepare libraries following Illumina's TruSeq mRNA Sample Prep Kit with minor modifications. Briefly, mRNA was purified using a bead-based protocol and chemically fragmented; first and second strand cDNA was synthesized from these fragments and end repaired. Following end repair, 3' ends were adenylated and a custom adapter was ligated. This allowed for one of 23 unique indices (Cui et al. 2013, Table S2) to be added to each library during PCR amplification (16-18 cycles). Following agarose gel purification of the desired size distribution (350-500 bp) and quality verification on the Bioanalyzer 2100 (Agilent, Santa Clara, CA), this 23-plex library was sequenced on one paired-end Illumina HiSeq 2000 lane at the Lewis-Sigler Institute Microarray Facility (Princeton, New Jersey). Samples from three additional species that could not be uniquely indexed were sequenced on partial lanes (*X. birchmanni*, *X. malinche*, and *X. nigrensis*). Raw 101 base

pair (bp) reads were trimmed to remove low quality bases (Phred quality score<20) and reads with fewer than 30 bp of contiguous high quality bases were removed using the script TQSfastq.py (http://code.google.com/p/ngopt/source/browse/trunk/SSPACE/tools/TQSfastq.py).

All raw data have been deposited in NCBI's Sequence Read Archive (Accession #: SRA061485).

Transcriptome assembly

For phylogenetic analysis (see below), I analyzed alignments to the *X. maculatus* genome and two transcriptomes, to demonstrate that my results are robust to the effects of reference species and reference sequence type (transcriptome or genome). The two transcriptomes were generated using paired-end data for the two species with the highest coverage, *X. birchmanni* and *X. mayae*. 29,535,466 (*X. birchmanni*) and 31,993,860 (*X. mayae*) paired-end reads were assembled using velvet (Zerbino and Birney 2008) with a range of kmers (21, 31, and 41). I initially used the merge assemblies option with Oases (Schulz et al. 2012), but found that this increased the incidence of highly divergent alignments in my later analysis, likely as a result of the formation of chimeric transcripts. I therefore used Oases to construct transcript isoforms without merging and used a custom perl script to select the longest isoform for my transcriptome reference. I compared the quality of different assemblies based on the assembly N50 and total number of bases assembled. In both cases, using a kmer of 31 produced the highest quality transcriptome: *X. birchmanni* –N50 of 2,441 bp, total size 108 Mbp; *X. mayae* – N50 of 3,545 bp, total size 147 Mbp.

Using this preliminary transcriptome assembly for each species, I then identified regions that were not unique in the *X. maculatus* genome by BLASTing all transcripts to the *X. maculatus* genome using the blastn algorithm (Camacho et al. 2009). To be conservative, I masked regions of transcripts that had multiple hits to the genome at an e-value < 1e-5, by converting the bases in that region to Ns. Many transcripts had short regions that BLASTed with high confidence to multiple sites in the genome. In order to avoid excluding regions that would not result in incorrect mapping I performed simulations to determine whether reads originating from another transcript would incorrectly map to small homologous regions (<100 bp) in the transcriptome. I found that regions with multiple blast hits shorter than 70 bp could be included in the analysis (Cui et al. 2013, Supporting information i). I completed all subsequent steps in the analysis using the *X. birchmanni* transcriptome, but repeated the analyses using both the *X. mayae* transcriptome assembly and *X. maculatus* genome to examine the effect of choice of reference sequence, and effects of using the transcriptome vs the genome assembly (Cui et al. 2013, Supporting Information ii, Figure S1). I chose the *X. birchmanni* transcriptome assembly for my main analysis because it resulted in more alignments that passed my criteria than the *X. mayae* assembly (see below), and is less likely to combine heterogeneous sequences in gene tree analysis than alignments to the *X. maculatus* genome (see Supporting Information ii).

Alignment generation and total evidence phylogeny

Trimmed reads were aligned to the *X. birchmanni* transcriptome sequence and the *X. maculatus* mitochondrial genome (GenBank Accession: AP005982.1) using

STAMPY v1.0.17 (Lunter and Goodson 2011). The number of reads per species and alignment statistics are summarized in Table S2 (Cui et al. 2013). Mapped reads were analyzed for variant sites and sequence depth in each species using the samtools/bcftools pipeline (Li et al. 2009) with a mapping quality cutoff of 20. Because the mapping process generates aligned sequences to the reference, traditional gene by gene global alignment is unnecessary.

A custom PHP script was used to generate sequence alignments for all 26 species based on the bcf file that is the output of the samtools/bcftools pipelines. Before analyzing phylogenetic relationships based on these sequence alignments, I performed a number of quality control steps to exclude low-quality and low-information sites from my dataset. For each species, transcripts with average per-site coverage <5X were excluded. Individual sites within a transcript were masked as N if coverage at that site was <5X or the variant quality score was <20. Sites containing polymorphism or indels were also masked. After this initial masking, I compared the remaining sites between species. If a site had been masked in 90% or more of the *Xiphophorus* species, or the transcript had been excluded in both outgroup species, I excluded that site from out analysis. I also excluded regions of high divergence (more than 7 character differences from the *X. birchmanni* sequence in 21 bp) using a sliding window. All remaining sites were included in the analysis. This resulted in 10,999 alignments ≥500 bp with a total alignment length of 16.85 Mbp (22.91% missing data).

To produce a total evidence phylogeny as a first approximation of the likely species tree, all alignments were concatenated for analysis. Due to the large *a priori*

partition numbers I did not allow free model parameters for each partition. Instead, I analyzed the dataset as a single partition with 100 rapid bootstraps (Stamatakis et al. 2008) followed by maximum likelihood tree estimation using the General Reversible Time substitution model (GTR+T) in RAxML 7.2.8 (Stamatakis 2006). Similar methods were used to produce total evidence phylogenies based on alignments to the *X. mayae* transcriptome and *X. maculatus* genome (Supporting Information ii).

Gene tree analysis with BUCKy

Gene trees can sometimes produce topologies that are different from the species tree because of incomplete lineage sorting or hybridization (Degnan and Rosenberg 2009). To explicitly account for these factors, I performed gene tree analysis using BUCKy (An é et al. 2007). I initially attempted gene tree analysis using both the programs BEST (Liu 2008) and BUCKy (An é et al. 2007) but were unable to use BEST due to computational limitations. The BUCKy program uses Bayesian concordance analysis to estimate the likely species tree topology, and what proportion of loci support the dominant topology. It has the advantage of making no assumptions about the source of the discordance. Therefore, in addition to incomplete lineage sorting, BUCKy accommodates hybridization (Larget et al. 2010). I treat instances of major discordance identified by BUCKy as potential cases of hybridization for further investigation (see below).

Due to limitations in computational speed I used a smaller dataset for BUCKy analysis (transcripts > 1.5 kb in length, 7.6 Mb) and excluded both outgroups from analysis due to extensive missing data in *Priapella* (Cui et al. 2013, Table S3). As a

result of computational constraints, I divided this 2,366 transcript dataset into two 1,183 transcript datasets for analysis with MrBayes 3.2.1 to obtain a posterior distribution of gene trees for each partition. Stationary stage was determined by inspecting parameter traces using Tracer. The following chain length (and burn-ins) were used: run 1, 31.2 million (10 million); run 2, 28.6 million (10 million). BUCKy was run with a range of α priors, which describes the expected number of unique gene trees given the number of taxa and the total number of genes in the genome. I did not find differences in the topology or concordance factor values between α values of 1, 2 and 5, and report results obtained with α=1. The concordance factor (CF) describes the estimated proportion of the genome significantly supporting a topology and BUCKy systematically underestimates the CF for the major topology, especially with short branch lengths and in the presence of ILS or hybridization (Chung and An é2011). Alternative bipartitions with concordance factors higher than 10% were further investigated for evidence of hybridization (see below). This large concordance factor cutoff was used to identify species with strong evidence for gene tree-species tree discordance. I also repeated the BUCKy analysis using alignments to the *X. maculatus* genome (Supporting Information iii, iv).

Mitochondrial phylogeny

Using the same methods outlined above, but with a 20x coverage cutoff, I obtained concatenated mitochondrial alignments of all coding regions for 26 species (15,787 bp, 42.63% missing). I repeated the analysis excluding both outgroups to

prevent long-branch attraction. Both datasets were analyzed with RAxML 7.8.2 using GTR+Γ. Nodal support was determined with 100 rapid bootstraps in RAxML.

Investigating potential hybridization

Strong discordance between gene trees and the species tree can be caused by incomplete lineage sorting or hybridization. To investigate whether incomplete lineage sorting can be rejected as the cause of discordance, I used the approximately unbiased (AU) test (Shimodaira 2002) and Patterson's D-statistic (Green et al. 2010; Schumer et al. 2013) to further investigate all groups in which I found strong evidence of discordance using BUCKy (CF ≥ 10%). Both tests were used to measure whether there was significant asymmetry in support for the two minor topologies in a four species tree; significant asymmetry in support for one of the two minor topologies can be a sign of hybridization (Meng and Kubatko 2009; Durand et al. 2011). I only investigated potential hybridization that occurred between extant species and I did not investigate potential hybridization between *X. andersi* and *X. milleri* because it was not detected in BUCKy analysis of alignments to the *X. maculatus* genome (Supporting Information iv, Cui et al. 2013). Two potential hybridization events could not be examined using my approach because of species interrelationships; these were the potential hybridization events between *X. hellerii* and *X. mayae*, and *X. couchianus* and *X. meyeri*, since higher discordance occurred between *X. hellerii* and *X. alvarezi*, and *X. couchianus* and *X. meyeri*, respectively (see Results).

I exported 4 (D-statistic) or 5 (AU test) taxon alignments for the 8 pairs of species for which I found strong evidence of gene tree discordance from BUCKy

(excluding *X. hellerii –X. mayae* and *X. couchianus—X.gordoni*, see above), a closely related species, and two outgroups (detailed in Cui et al. 2013, Table S2). To investigate a case of mito-nuclear discordance, I also tested for gene flow between *X. birchmanni* and *X. pygmaeus* despite the fact that I failed to detect nuclear discordance between these two species (see Results). Quality criteria for alignments were as described above, except missing data at a site in any species resulted in that site being excluded from the analysis.

For the AU test, I enforced monophyly of the two outgroups and tested support for the three possible topologies of an unrooted 4-taxa tree. Site likelihoods were calculated using RAxML 7.8.2 with the General Time Reversible model and a gamma distribution of substitution rates (GTR+GAMMA). These likelihoods were used as input for AU tests implemented in Consel 0.2 (Shimodaira and Hasegawa 2001). The AU *p*-value is the probability that a tree is as likely as the Maximum Likelihood tree. If a particular topology had an AU *p*-value greater than 0.95 for an alignment, I concluded that the alignment supported that topology. I excluded partitions that had an observed likelihood difference of 0 because these are likely caused by low numbers of informative characters in the alignment (Schmidt 2009). I determined whether there was significant asymmetry in support for the two minor topologies by calculating 95% confidence intervals using 1000 replicates of non-parametric bootstrapping.

As a secondary method, I calculated Patterson's D-statistic (Green et al. 2010) for the same alignments using only one outgroup (Cui et al. 2013, Table S2). Sites with ABBA (shared sites between species 2 and 3) and BABA (shared sites between species 1

and 3) patterns were counted in the four species alignments. With the null hypothesis of incomplete lineage sorting, the number of ABBA and BABA sites is expected to be equal. Significant deviation of D from 0 suggests that incomplete lineage sorting can be rejected as a null hypothesis. Significance of D was determined using a two sample z-test. Standard error was determined by jack-knife bootstrapping of D for each transcript using the bootstrap package in R (R Development Core Team 2010), I then compared the D-statistic results with the AU results (Table 1); in cases in which all three methods (BUCKy, AU test, and Patterson's D-statistic) support hybridization, I conclude that hybridization likely occurred between the taxa in question.

Character mapping and independent contrasts

To examine patterns of sword evolution and sword preference evolution in the context of my species tree topology, I compiled a dataset of sword index (sword length/standard length, (Basolo 1995a), machinery for sword production (MSP, the ability to generate sword after androgen treatment, (Gordon et al. 1943; Dzwillo 1963, 1964; Zander and Dzwillo 1969), and sword preference in *Xiphophorus* females based on previously published studies (Cui et al. 2013, Supporting Information v, Table S4, Table S5). Though previous researchers have distinguished between the long and pigmented sword found in southern and some northern swordtails, and the short unpigmented sword found in two platyfishes, I do not make that distinction here. However, I recognize that the sword is a complex trait and has many morphological differences between clades.

Sword index and female preference index were coded as continuous variables while MSP was coded as a binary character. I traced the characters on the total evidence tree produced by alignment to the *X. birchmanni* transcriptome by maximum parsimony in Mesquite 2.75 (Maddison and Maddison 2011), because maximum likelihood methods cannot be implemented for continuous variables in this program. I used the PDAP-PDTREE package (Midford et al. 2011) in Mesquite to perform independent contrast analysis (Felsenstein 1985) between sword index and preference for swords and linear regressions between sword and preference characteristics and node height.

Table 1. Proportion of the gene trees supporting alternative topologies estimated by an approximately unbiased test (AU test) and D-statistic calculated for all species with discordance >10% identified by BUCKy. Confidence intervals are calculated by 1000 replications of non-parametric bootstraps for AU tests and jack-knife bootstrapping for the D-statistic. Positive values support gene flow between the two species in question. Divergence refers to the average sequence divergence between the two sister species in the genomic regions significantly (AU $p > 0.95$) supporting the indicated sister relationships.

Species pair	CF	D-statistic (jack-knife SE, p-value)	Percent of AU test support for the two minor topologies (95% confident intervals)	Divergence
X. nezahualcoyotl —	0.25	0.38	((*X. nezahualcoyotl, X. montezumae*), *X. cortezi*): 32% (30-33%)	0.36%
X. montezumae		(0.07, p <3.1e-14)	((*X. montezumae, X. cortezi*), *X. nezahualcoyotl*): 3% (2-3%)	0.36%
X. signum —	0.21	0.56	((*X. signum, X. mayae*), *X. hellerii*): 51% (49-53%)	0.40%
X. mayae		(0.03, p < 4.4e-65)	((*X. signum, X. hellerii*), *X. mayae*): 7.6% (7-8.5%)	0.40%
X. hellerii —	0.21	-0.41	((*X. hellerii, X. alvarezi*), *X. mayae*): 16% (15-17%)	0.32%
X. alvarezi		(0.01, p < 6.4e-100)	((*X. hellerii, X. mayae*), *X. alvarezi*): 13% (12-14%)	0.36%
X. variatus—	0.11	0.31	((*X. xiphidium, X. variatus*), *X. evelynae*): 30% (29-32%)	0.37%
X. xiphidium		(0.014, p < 1.5e-86)	((*X. xiphidium, X. variatus*), *X. evelynae*): 10% (9-11%)	0.44%
X. milleri—	0.11	0.09	((*X. evelynae, X. milleri*), *X. couchianus*): 16% (15-18%)	0.43%
X.evelynae		(0.013, p<0.00043)	((*X. milleri, X. couchianus*), *X. evelynae*): 14% (13-15%)	0.43%
X. couchianus—	0.29	0.11	((*X. meyeri, X. couchianus*), *X. gordoni*): 26% (24-29%)	0.06%
X. meyeri		(0.04, p <0.0030)	((*X. couchianus, X. gordoni*), *X. meyeri*): 17% (15-19%)	0.08%

Table 1 Continued.

Species pair	CF	D-statistic (jack-knife SE, p-value)	Percent of AU test support for the two minor topologies (95% confident intervals)	Divergence
X. xiphidium—	0.13	0.03	((*X. andersi*, *X. xiphidium*), *X. meyeri*): 19% (18-21%)	0.54%
X. andersi		(0.014, p <0.041)	((*X. meyeri*, *X. andersi*), *X. xiphidium*): 11% (9-12%)	0.55%
X. evelynae—	0.14	0.18	((*X. variatus*, *X. evelynae*), *X. couchianus*): 19% (18-20%)	0.34%
X. variatus		(0.013, p<5.3e-28)	((*X. evelynae*, *X. couchianus*), *X. variatus*): 10% (10-11%)	0.36%
X. birchmanni—	<0.10	-0.03	((*X. pygmaeus*, *X. malinche*)), *X. birchmanni*): 3% (3-4%)	0.57%
X. pygmaeus		(0.017 , p=0.19)	((*X. birchmanni*, *X. pygmaeus*)), *X. malinche*): 3% (3-3%)	0.61%

Results

A high confidence species tree based on nuclear data

I used multiple methods to construct a high confidence species tree for *Xiphophorus*. As an initial approach to determine the likely species tree, I constructed a total evidence phylogeny using RAxML based on alignments to the *X. birchmanni* transcriptome; this concatenated nuclear dataset produced a fully resolved phylogeny with 100% bootstrap support for all nodes (Figure 15 A).

To confirm that my total evidence topology was not dependent on the reference sequence used for assembly, I repeated the same analysis using the *X. maculatus* genome and the *X. mayae* transcriptome as reference sequences (Supporting information ii). The total evidence topology produced from alignment to the *X. maculatus* genome was nearly identical to the topology produced from alignment to the *X. birchmanni* transcriptome; this topology placed *X. nezahualcoyotl* as sister to *X. cortezi* rather than *X. montezumae* (Cui et al. 2013, Supporting Information ii, Figure S1A). The total evidence topology produced from alignment to the *X. mayae* transcriptome also resulted in a highly similar topology, albeit with lower bootstrap support; for details see Supporting Information ii, Figure S1B in Cui et al. (2013).

I also investigated the likely species tree with gene tree analysis using BUCKy. With this method I infer a nearly identical species tree topology to the topology produced by the analysis of the concatenated *X. birchmanni* aligned dataset with RAxML (Figure 16; see Table S6 in Cui et al. 2013 for CFs). As with the topology

produced by aligning to the *X. maculatus* genome, the only difference between the species tree inferred by BUCKy and the concatenated topology was the placement of *X. nezahualcoyotl* sister to *X. cortezi* rather than *X. montezumae*. Given that the species trees produced by the different analyses are nearly identical, I focus subsequent discussion on the species tree determined by BUCKy.

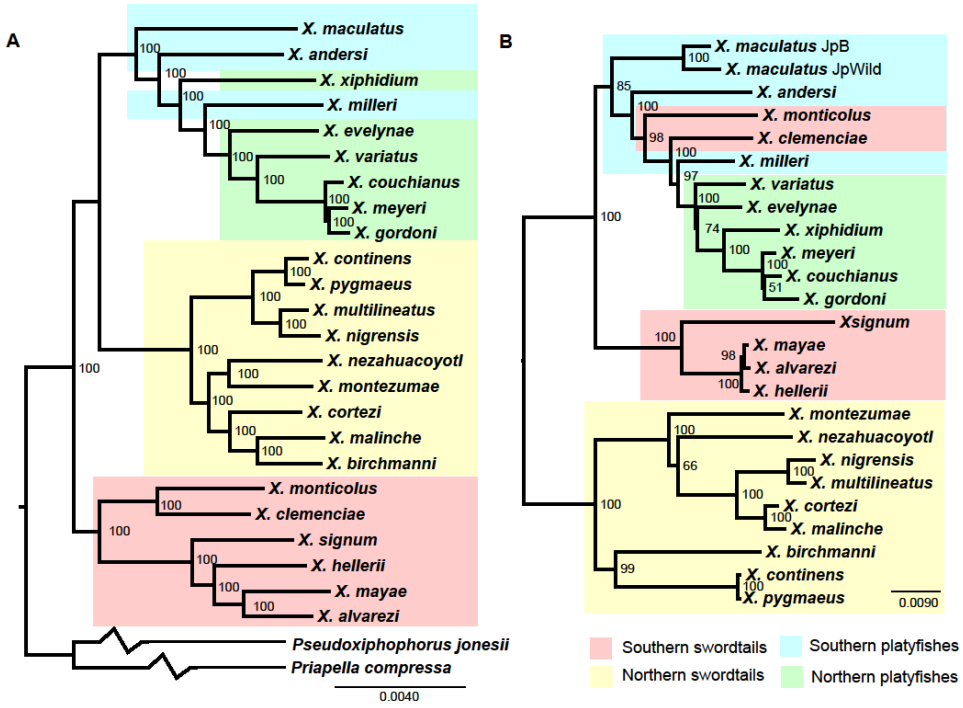


Figure 15. Total evidence trees. A) Total evidence nuclear phylogeny produced by concatenating 10,999 transcripts totaling 16,855,549 (22.91% missing) sites using RAxML 7.2.8 with GTR+ Γ model. Nodal support generated by 100 rapid bootstraps with GTR+CAT. Log likelihood = -33380190.53. B) Unrooted mitochondrial tree (coding regions only, 15,787 bp, 42.63% missing) of 24 *Xiphophorus* species excluding two outgroup species. Log likelihood= -40371.19.

My results resolve many of the uncertainties of the *Xiphophorus* phylogeny (Figure 15 & Figure 16). The monophyly of the three major *Xiphophorus* clades (northern

swordtails, platyfish, and southern swordtails) is strongly supported. My results also strongly support the placement of platyfish and northern swordtails as sister clades, with southern swordtails at the base of the genus (corresponding to Figure 14 B). In within clade relationships, I also find a number of major differences from previous studies. *X. xiphidium,* which is a northern sworded platyfish, is resolved at a more basal position that is more closely related to the southern platyfish *X. andersi* than in previous studies (Meyer et al. 2006). Excluding *X. xiphidium,* southern platyfishes are paraphyletic while northern platyfishes are monophyletic. Out of three hypothesized clades within the northern swordtails (Rauchenberger et al. 1990), two (the *cortezi* and *pygmaeus* clades) were found to be paraphyletic, while the *montezumae* clade is polyphyletic. *X. montezumae* was resolved as the sister to the *cortezi* clade, which here included *X. nezahualcoyotl,* while *X. continens* was supported as sister of *X. pygmaeus* (but see Cui et al. 2013, Table S1).

Extensive cytonuclear conflict

Previous phylogenetic studies in many taxa have historically relied heavily on mitochondrial DNA (mtDNA) as a marker. I constructed mtDNA relationships based on concatenated expressed mitochondrial sequences using RAxML. I find low concordance between mitochondrial and nuclear relationships in *Xiphophorus* (Figure 15). The placement of *X. maculatus* was not well resolved in the mtDNA tree. When outgroups are included, *X. maculatus* resides at the base of a clade including all southern swordtails and other platyfishes, with low bootstrap support (Cui et al. 2013, Figure S2). Excluding both outgroups places *X. maculatus* sister to other platyfishes, but again with low

bootstrap support (Figure 15 B). In the rooted trees, mtDNA sequences support monophyly of southern swordtails and platyfishes, in direct conflict with results from nuclear data. As reported previously (Meyer et al. 2006; Kang et al. 2013), *X. clemenciae* and *X. monticolus* are nested within the platyfishes, also in contradiction to the species tree inferred from nuclear genome sequences.

Hybridization inferred by BUCKy, AU tests, and the D-statistic

Gene tree discordance can result from incomplete lineage sorting (Pollard et al. 2006), reconstruction errors, or from hybridization (Degnan and Rosenberg 2009). I used BUCKy to identify instances of major discordance between individual nuclear gene trees and the consensus species tree (Figure 16 & Figure 17). A large number of *Xiphophorus* species had high levels of nuclear gene tree discordance based on the results of my BUCKy analysis. I found greater than 10% discordance in all three major groups: northern swordtails (*X. nezahualcoyotl, X. montuzumae*), platyfishes (*X. xiphidium, X. andersi, X. evelynae, X. variatus, X. couchianus, X. milleri, X. meyeri, X. gordoni*), and southern swordtails (*X. hellerii, X. signum, X. alvarezi*, and *X. mayae*). I find nearly identical patterns of discordance in both the alignments to the *X. birchmanni* transcriptome assembly and the alignments to the *X. maculatus* genome. (Cui et al. 2013, Supporting Information iii, iv; Figure S3).

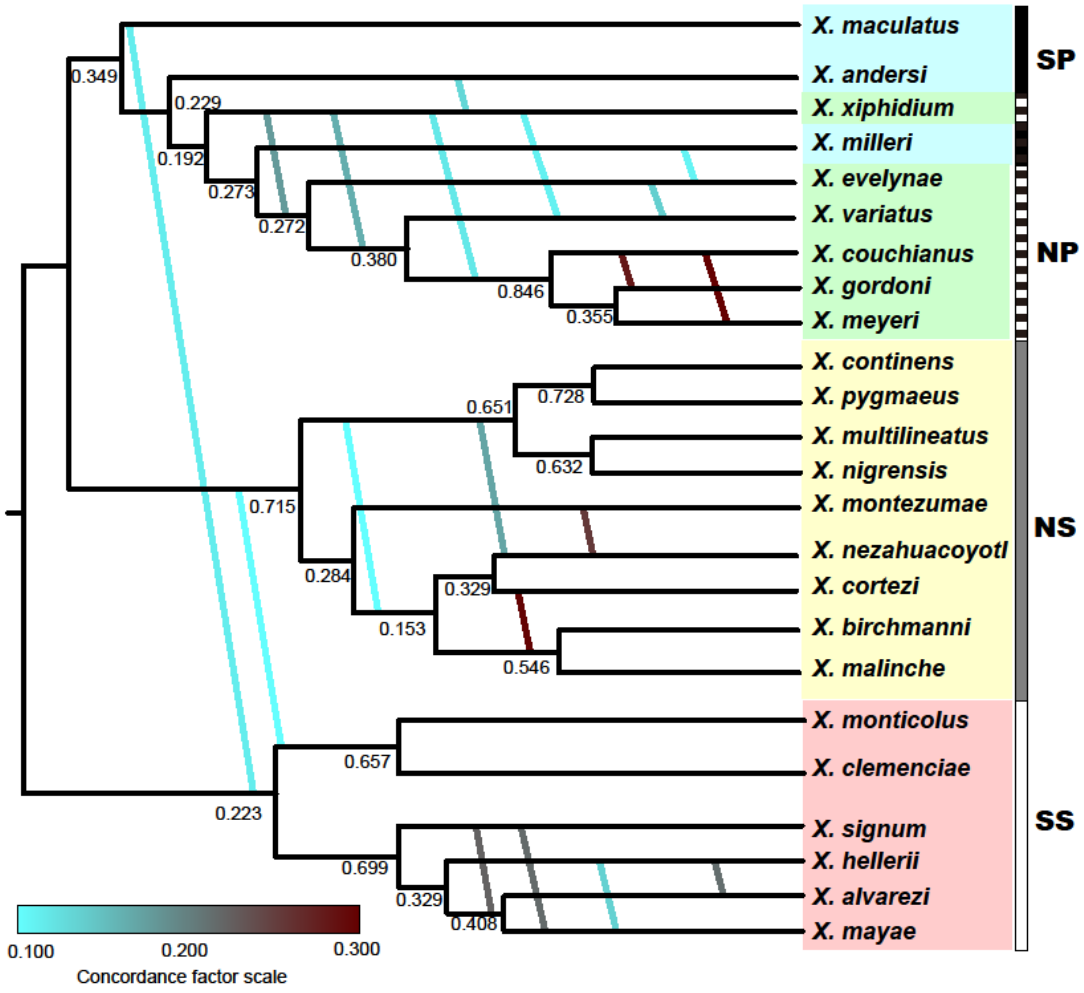


Figure 16. A primary concordance tree produced by BUCKy (α=1) from 2,366 gene trees inferred by MrBayes 3.2.1. Nodal values are Bayesian concordance factors. Discordance which is detected using both the *X. birchmanni* transcriptome and the *X. maculatus* genomic reference (see Cui et al. 2013, Figure S3) with concordance factors > 10% is indicated with solid lines using a color gradient. See Table S6 (Cui et al. 2013) for concordance factors and confidence intervals.

For all species that showed high levels of discordance (i.e. ≥ 10%; Figure 16), I performed two additional analyses to differentiate between incomplete lineage sorting and hybridization. To evaluate the evidence for hybridization, I applied the

approximately unbiased (AU) test (Shimodaira 2002) and Patterson's D-statistic (Green et al. 2010) to these species (Cui et al. 2013, Table S2). Both tests allow me to determine whether incomplete lineage sorting can be rejected as the cause of discordance by testing for asymmetry of the two less frequent topologies. Given agreement between all three analyses (BUCKy, the AU test, and Patterson's D-statistic), I conclude that hybridization is a likely contributor to gene tree discordance among the species in question. Specifically, I find that 7 out of 8 tested cases of discordance are likely the result of hybridization (Table 2). Some species had levels of hybridization consistent with extensive admixture or hybrid speciation. For example, the majority of the *X. xiphidium* alignments (70%) cluster with *X. meyeri* in AU tests, but a significant proportion (19%) are grouped with the southern platyfish *X. andersi*. In *X. nezahualcoyotl*, 66% of alignments were most closely related to *X. cortezi* but 32% are more closely related to *X. montezumae*. This hybridization is unlikely to be recent however, because divergence between species was still high in regions supporting minor topologies (Table 1, compared to Table S7 in Cui et al. 2013).

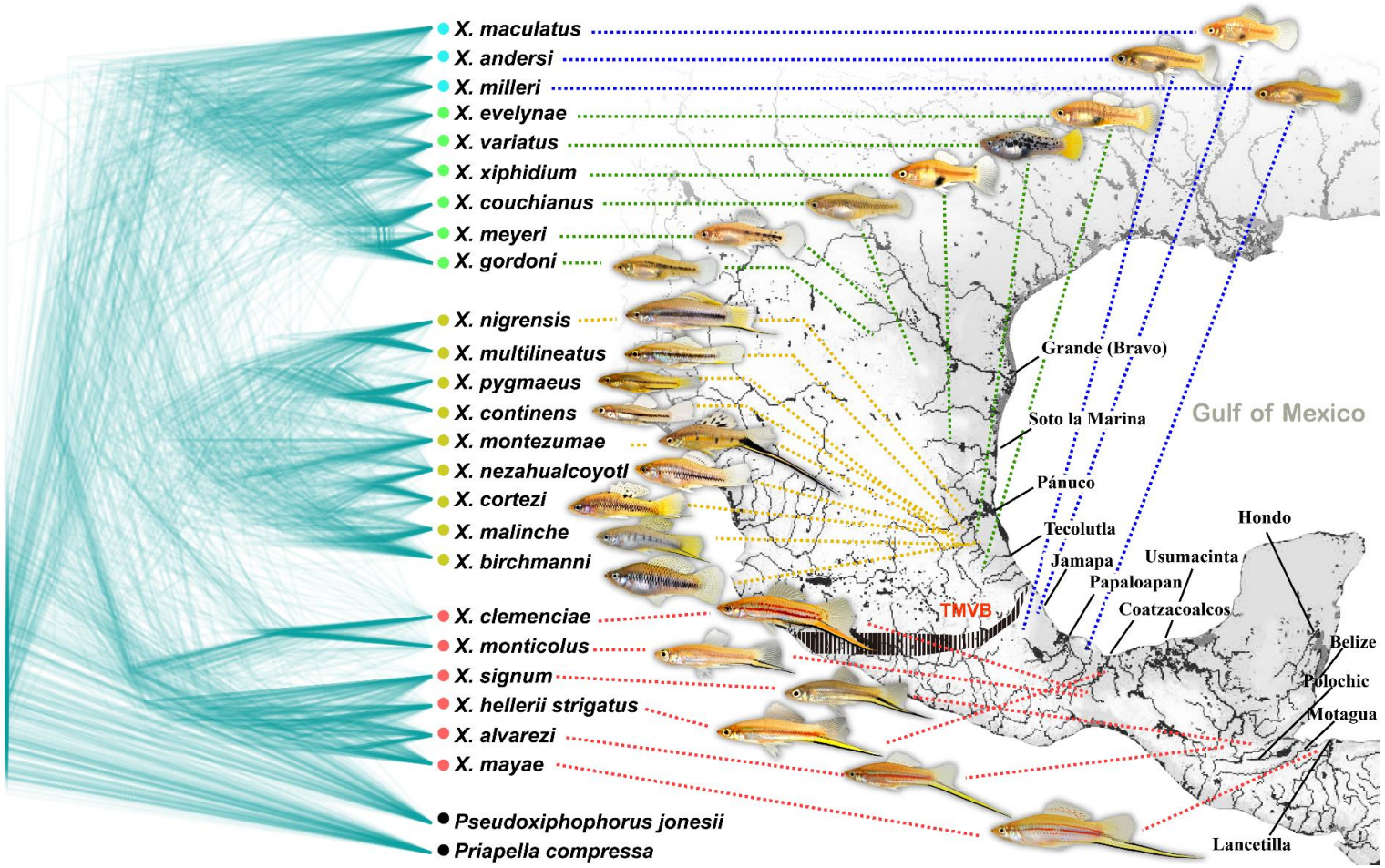


Figure 17. DensiTree of *Xiphophorus*. Localities and photographs of the specimens used in this study mapped on a DensiTree produced by overlaying 160 gene trees inferred by MrBayes 3.2.1. Black vertical lines indicate the location of a major geographical barrier, the Trans-Mexican Volcanic Belt (TMVB). Blue lines: southern platyfishes; green lines: northern platyfishes; yellow lines: northern swordtails; pink lines: southern swordtails.

Character mapping and independent contrasts

I used the total evidence tree to investigate evolutionary trends for sword length, sword preference, and whether female preference for the sword is correlated with sword length (Supporting Information v). Independent contrasts reveal a marginally non-significant positive correlation between male sword length and female preference (r=0.55, d.f.=9, p=0.083). Female preference significantly decreased with node height (r= -0.71, d.f.=8, p=0.022), suggesting an evolutionary trend for decreased sword preference. Sword length, however, was not correlated with node height in the species tree (r= -0.27, d.f.=23, p=0.185).

Discussion

Using sequence information generated from one paired-end Illumina lane, I resolve phylogenetic relationships in a group of species with a highly contested evolutionary history. I show that there are high levels of hybridization between many species, potentially due to weak postzygotic isolation in *Xiphophorus* (Schartl 2008; Rosenthal and Ryan 2011). Extensive admixture has occurred in lineages within the platyfish, northern, and southern swordtail clades, while cases of introgression have occurred within all major clades. A high-confidence molecular phylogeny of *Xiphophorus* allows me to address questions about the role of sexual selection and hybridization in trait evolution.

Phylogenetic relationships in *Xiphophorus*

I find strong support for monophyly of the three previously identified *Xiphophorus* clades: the southern swordtails, the northern swordtails, and the platyfishes. The rooting of my phylogeny (most closely corresponding to Figure 14 B) confirms results from previous nuclear datasets that northern swordtails and southern swordtails are not monophyletic (e.g. Meyer et al. 2006). This rooting conflicts with early studies using morphology (Rosen 1979) and mitochondrial DNA (Meyer et al. 1994). I also find major differences in intraclade relationships compared with previous studies, particularly in the northern swordtails (Rauchenberger et al. 1990; Morris et al. 2001; Kang et al. 2013).

Reticulate evolution in *Xiphophorus*

Based on my analysis of nuclear sequences, I find evidence of gene flow between multiple species within platyfishes, northern swordtails, and southern swordtails (Table 1, Figure 16 & Figure 17). Only 8 of the species surveyed show no evidence for discordance of nuclear gene trees. My results demonstrate that hybridization has been extensive within clades, even between non-sister species, but there is little evidence of hybridization between the major clades. The extent of historical hybridization in *Xiphophorus* is especially interesting given that contemporary natural hybrids are relatively rare. Premating isolation via sexual selection is the major barrier to hybridization in this genus (Schartl 2008; Rosenthal and Ryan 2011) but reproductive isolation via mating preferences may be particularly vulnerable to perturbation due to demographic fluctuations (Willis et al. 2011) or environmental disruption of sexual

communication (Fisher et al. 2006). The vulnerability of female preference as the major barrier to reproductive isolation is apparent in my results, which indicate rampant genetic exchange throughout the *Xiphophorus* phylogeny. Hybridization is likely to be even more widespread than I report, since my method will only detect gene flow when it is extensive enough to cause discordance >10%. For example, historical hybridization between the widely sympatric *X. maculatus* and *X. hellerii* has been recently reported (Schumer et al. 2013) but the level of shared ancestry (~6%) is lower than the detection threshold used here.

Mitochondrial DNA relationships have historically been used for phylogenetic studies and as a method of detecting past hybridization, but accumulating evidence casts doubt on the utility of mitochondrial genes in studying evolutionary relationships (Ballard and Rand 2005). My mtDNA results differ strikingly from relationships suggested by nuclear sequences. In the mtDNA phylogeny, the southern swordtails are grouped as sister to the platyfishes, while nuclear sequences group platyfishes and northern swordtails (Figure 15). Past research in *Xiphophorus* has inferred potential cases of hybridization based on mitonuclear incongruence (Meyer et al. 1994; Meyer et al. 2006; Kang et al. 2013). However, the mtDNA topology does not support the same hybridization patterns detected in my nuclear gene tree analysis. Instead, I find mtDNA incongruence can occur even when the nuclear genome shows no evidence of hybridization. For example, I did not find evidence of gene flow between platyfish and *X. monticolus* as recently suggested (Kang et al. 2013). Similarly, a previous study (Schumer et al. 2013) using whole genome sequencing of *X. clemenciae* did not support

extensive gene flow with *X. maculatus*. In northern swordtails, even though the sister relationship between the *X. pygmaeus* clade and *X. birchmanni* is highly supported by mtDNA, there is no evidence for gene flow between these species based on nuclear gene trees. Other factors such as incomplete lineage sorting or selection could be an alternate explanation of the observed mtDNA patterns (e.g. Schumer et al. 2013).

Evidence for extensive admixture in all major groups

I find evidence based on genome-wide data that *X. nezahualcoyotl*, *X. xiphidium*, and multiple species in the *X. hellerii* clade, have significant proportions of their genomes derived from hybridization. These species have levels of discordance ranging from 19-42% based on AU tests (Table 1), suggestive of extensive historical admixture or even hybrid speciation. Given current species distributions, hybridization between *X. xiphidium—X. andersi* and *X. nezahualcoyotl—X. montezumae* has not been previously suspected, suggesting that historical species distributions or migration allowed hybridization between these species.

Though all three of these groups show high levels of hybridization, extensive admixture in *X. nezahualcoyotl* is also consistent with morphological data and previous research. Based on such high levels of nuclear gene tree discordance (~30%) it is possible that *X. nezahualcoyotl* is the product of admixture between its sister species *X. cortezi* and *X. montezumae.* The sword of *X. nezahualcoyotl* is of intermediate length of *X. montezumae* and *X. cortezi* (sword index: 0.48, *X. montezumae* 1.0, *X. cortezi* 0.37; see Cui et al. 2013, Table S4 & Supporting Information v), and *X. nezahualcoyotl* shares morphological characteristics with both species (pigment patterns shared with *X. cortezi*

and lateral stripes shared with *X. montezumae*). Previous phylogenetic analyses could not resolve relationships between *X. nezahualcoyotl, X. montezumae,* and *X. cortezi* (Rauchenberger et al. 1990; Morris et al. 2001). I suggest that these difficulties stem from the extensive hybridization in the evolutionary history of these species. A more complete survey throughout the range of *X. nezahualcoyotl*, and whole genome sequencing of both species and *X. cortezi* is required to further evaluate this prediction.

Implications for the evolution of sexually selected traits

Early research proposed that platyfishes were basal to the genus and the evolution of the sword was driven by latent female preferences (Basolo 1995b, a). Based on the rooting of the genus in my phylogeny, I find that the ability to produce a sword was likely present in the common ancestor of *Xiphophorus* and secondarily lost in some platyfishes (as observed in Meyer et al. 1994; Fig. 5; Meyer et al. 2006). However, preferences for swords in outgroup taxa suggest that the sword could have evolved in response to a preexisting bias prior to the diversification of *Xiphophorus* (as suggested by Basolo 1995b).

My phylogeny changes the placement of *X. xiphidium*, a sworded platyfish, in comparison to previous studies. Because *X. maculatus* and *X. milleri* both have the capacity to produce a short sword if induced by androgens (Gordon et al. 1943; Dzwillo 1963, 1964; Zander and Dzwillo 1969; Offen 2008), and *X. andersi* and *X. xiphidium* are both sworded, my phylogeny demonstrates that the ability to produce a short sword was present in the common ancestor of platyfishes, and thus a synapomorphy of the genus. Based on my results, the genetic pathway for sword production was not completely lost

in platyfishes until after the diversification of *X. milleri* (Figure 18). This finding has important implications for research into the genetic basis of the sword; since this trait arose only once it is likely that the production of the sword is regulated by the same mechanisms in all *Xiphophorus* species. Despite the likely presence of the sword in the common ancestor of *Xiphophorus*, sword preference is negatively correlated with node height, suggesting an evolutionary trend for a reduction in sword preference in *Xiphophorus*.

Gene flow and sexually selected traits

Only two platyfish species have a short sword, *X. xiphidium* and *X. andersi*. The platyfish sword is significantly reduced compared to sword ornaments in northern and southern swordtails, and lacks pigmentation. An alternate explanation for the presence of short swords in *X. andersi* and *X. xiphidium* is a single loss of the trait in platyfishes followed by introgression. I consider this unlikely because I do not find evidence of swordtail ancestry in *X. xiphidium* or *X. andersi*, although I caution that lower levels of hybridization (less than my 10% detection threshold) could have occurred between swordtails and these two platyfishes. On the other hand, I find evidence of hybridization between *X. andersi* and *X. xiphidium*, raising the possibility that loci underlying the short sword could be shared in these two species. Identifying the genetic basis of the sword is a crucial next step in determining whether hybridization has played a role in the phylogenetic distribution of this trait (e.g. Heliconius Genome Consortium 2012).

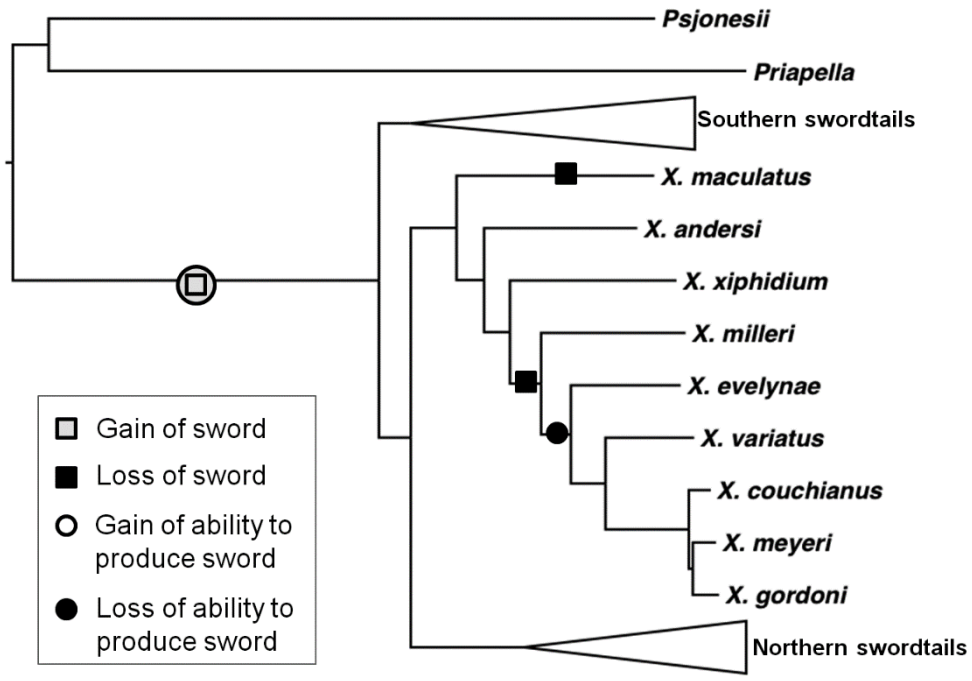


Figure 18. Sword mapped on tree. Sword traits (presence of a sword and ability to produce a sword following androgen induction) mapped onto the total evidence phylogeny with southern and northern swordtails collapsed. This topology shows a single loss of the ability to produce a sword in the northern platyfishes and multiple losses of the natural expression of sword.

Conclusions

Previous investigations of the relationships among *Xiphophorus* species have been fraught with inconsistencies, likely as a result of gene flow among taxa. My approach highlights the utility of genome-scale data in resolving patterns of gene flow in addition to determining species trees with high confidence. My results demonstrate that evolution of the sword predated the diversification of *Xiphophorus* and that the ability to produce the sword was not completely lost until after the diversification of northern platyfishes. In addition, I find that hybridization in *Xiphophorus* has been historically

common, and I document evidence of extensive admixture, potentially indicative of hybrid speciation, in the northern swordtail *X. nezahualcoyotl*, southern swordtails in the *X. hellerii* clade, and the platyfish *X. xiphidium*. My results suggest that species boundaries primarily maintained by behavioral pre-mating isolation may be particularly porous and that gene flow can be widespread without resulting in species collapse. Hybridization is likely to play a major role in the evolutionary history of many species without strong post-mating isolation.

CHAPTER V

CONCLUSIONS

Micro-evolutionary processes generates macro-evolutionary patterns I observe today (Hansen and Martins 1996), be it biogeographical, morphological, genetic or biochemical. However, despite the increasing ease of obtaining genome-wide data at an affordable budget, retrospective inference of evolutionary processes from genomic data alone has been difficult and error-prone due to confounding effects of distinct processes (Sunnåker et al. 2013). For example, whether human genomic patterns suggest ancestral population structure or intraspecific hybridization has been repeatedly debated (Lohse and Frantz 2014). Therefore, it is extremely valuable to understand mechanisms that drive evolutionary processes, such that *a priori* information can be provided to aid ruling out competing hypotheses.

Understanding the dynamics during early stage of divergence is the key to understanding the origin of biodiversity. Quite often, pre-zygotic, pre-mating mechanisms are the first to evolve to limit gene flow between diverging clades, especially in sympatric species (Coyne and Orr 1989; Coyne and Orr 1997) and also in allopatric species (Mendelson 2003). The underlying genetic loci contributing to these barriers are often under local adaptation in allopatry, or disruptive selection when in sympatry. Incompatibility between these loci in hybrids is likely acting on an ecological level, and the fitness cost is often small and is only expressed as relative fitness when in competition with parental forms. Therefore, one would expect that these initial barriers to be porous, despite the canalizing effects they may have on lineage divergence.

Behavioral mechanisms are one such isolating mechanisms. Very often the mechanistic details of how interacting signals and receiver coevolve into barriers are not made explicit in modeling work. Many assume that females have genetic loci that intrinsically target a certain male trait value (Boake 1991). The degree of isolating effects is then simply a function of divergence because the genetic values of the signal-receiver loci co-evolve, possibly in a run-away process (Fisher 1915). However, recent findings show that indirect genetic effects, in particular learned mate choice, can drastically affect the outcome of mate choice depending on the social environment (Verzijden et al. 2005; Bailey and Moore 2012). Because females now don't necessarily innately favor a particular trait value, the evolutionary dynamics and the strength of the reproductive barrier will be different from the more simplistic assumptions.

In chapters 1 & 2, I showed that theoretically, disdain for familiar male phenotypes can evolve under overdominance of loci linked to the male signaling loci under different modes of imprinting (peer, parental, oblique), and that opposing effects of learning in mate choice can sometimes create population structure (i.e. serve as barrier to gene flow) or promote gene flow (i.e. create pores in the barrier). A previous puzzling genetic pattern was found in several *Xiphophorus birchmanni* x *X. malinche* hybrid zones (Culumber et al. 2011). Certain zones have uniform distribution of hybrid index while some have bimodal distribution, suggesting a difference in population structure. Two competing hypotheses may explain such pattern: 1) non-random mating; 2) migration. By examining learned mate choice in *X. malinche*, and combined with previous behavioral result from *X. birchmanni*, I suggest that hypothesis 1) may explain

population genetic patterns in the wild. Future experiments on more hybrid zones will allow me to determine whether the prediction of positive correlation between overall *X. birchmanni* ancestry and population structure is valid.

In the Chapter 3, I further looked into molecular mechanisms of olfactory learning in *X. malinche*. Interestingly, olfactory learning in this species alters genes related to neural plasticity, which not only differ between presence or absence of adult males, but also between the model species. I identified a potential marker *fosb* that I can apply in future studies to quantify differentially activated brain regions using *in situ* hybridization. Finally, I provided preliminary evidence of differential expression of odorant receptors at the periphery in response to early learning.

Given that reproductive barrier in *Xiphophorus* is likely affected by learning and post-zygotic isolation is weak, I expect that patterns of permissive gene flow should occur throughout the phylogeny. Indeed, in Chapter 4, I use RNA-seq technique to reconstruct a reticulate phylogeny for 24 of the 26 *Xiphophorus* species and found that species boundary in this group is indeed porous. Gene flow is also more extensive between closer related lineages than between distantly related lineages. For example, although previous studies suggested possible introgression between *X. clemenciae* and platyfishes, a previous whole genome analysis (Schumer et al. 2013) as well as Chapter 3 reject this hypothesis. Instead, extensive gene flow usually occurs between more closely related groups, such as *X. cortezi* and *X. montezumae*. This pattern is consistent with two mutually non-exclusive hypotheses: 1) innate preference becomes dominant

compared to learned mate choice after long divergence; 2) there's innate bias towards learning, a phenomenon called the "instinct to learn" (Marler 1991).

In conclusion, understanding mechanisms of individual behavior will inform our understanding of long-term evolutionary patterns and aid interpretation of the vast amount of genomic data to be collected in the up-coming era.

REFERENCES

Abbott, R. J., M. J. Hegarty, S. J. Hiscock, and A. C. Brennan. 2010. Homoploid hybrid speciation in action. Taxon 59:1375-1386.

Aboim, M. A., J. Mavárez, L. Bernatchez, and M. M. Coelho. 2010. Introgressive hybridization between two Iberian endemic cyprinid fish: a comparison between two independent hybrid zones. J. Evol. Biol. 23:817-828.

Agrawal, A. F. 2001. The evolutionary consequences of mate copying on male traits. Behav. Ecol. Sociobiol. 51:33-40.

Amadou, C., R. M. Younger, S. Sims, L. H. Matthews, J. Rogers, A. Kumanovics, A. Ziegler, S. Beck, and K. F. Lindahl. 2003. Co-duplication of olfactory receptor and MHC class I genes in the mouse major histocompatibility complex. Hum. Mol. Genet. 12:3025-3040.

Andersson, M. B. 1994. Sexual selection. Princeton University Press, Princeton, NJ.

Andolfatto, P., D. Davison, D. Erezyilmaz, T. T. Hu, J. Mast, T. Sunayama-Morita, and D. L. Stern. 2011. Multiplexed shotgun genotyping for rapid and efficient genetic mapping. Genome Res. 21:610-617.

Ané, C., B. Larget, D. A. Baum, S. D. Smith, and A. Rokas. 2007. Bayesian estimation of concordance among gene trees. Mol. Biol. Evol. 24:412-426.

Aoki, K., M. W. Feldman, and B. Kerr. 2001. Models of sexual selection on a quantitative genetic trait when preference is acquired by sexual imprinting. Evolution 55:25-32.

Arnold, M. L. and N. H. Martin. 2010. Hybrid fitness across time and habitats. Trends Ecol. Evol. 25:530-536.

Bailey, N. W. and A. J. Moore. 2012. Runaway sexual selection without genetic correlations: social environments and flexible mate choice initiate and enhance the Fisher process. Evolution 66:2674-2684.

Ballard, J. W. O. and D. M. Rand. 2005. The population biology of mitochondrial DNA and its phylogenetic implications. Annu. Rev. Ecol. Evol. Syst. 36:621-642.

Basolo, A. L. 1990. Female preference predates the evolution of the sword in swordtail fish. Science 250:808-810.

Basolo, A. L. 1995a. A further examination of a pre-existing bias favouring a sword in the genus *Xiphophorus*. Anim. Behav. 50:365-375.

Basolo, A. L. 1995b. Phylogenetic evidence for the role of a pre-existing bias in sexual selection. Proc. R. Soc. Lond. B Biol. Sci. 259:307-311.

Basolo, A. L. 2006. Genetic linkage and color polymorphism in the southern platyfish (*Xiphophorus maculatus*): a model system for studies of color pattern evolution. Zebrafish 3:65-83.

Baz áes, A., J. Olivares, and O. Schmachtenberg. 2013. Properties, projections, and tuning of teleost olfactory receptor neurons. J. Chem. Ecol. 39:451-464.

Beaumont, M. A., W. Zhang, and D. J. Balding. 2002. Approximate Bayesian computation in population genetics. Genetics 162:2025-2035.

Boake, C. R. 1991. Coevolution of senders and receivers of sexual signals: genetic coupling and genetic correlations. Trends Ecol. Evol. 6:225-227.

Bolhuis, J. J., S. M. Gobes, N. J. Terpstra, A. M. den Boer-Visser, and M. A. Zandbergen. 2012. Learning-related neuronal activation in the zebra finch song system nucleus HVC in response to the bird's own song. PloS One 7:e41556.

Bourne, A., G. Mohan, M. Stone, M. Pham, C. Schultz, J. Meyerhoff, and L. Lumley. 2013. Olfactory cues increase avoidance behavior and induce Fos expression in the amygdala, hippocampus and prefrontal cortex of socially defeated mice. Behav. Brain Res. 256:188-196.

Burley, N., G. Krantzberg, and P. Radman. 1982. Influence of colour-banding on the conspecific preferences of zebra finches. Anim. Behav. 30:444-455.

Butlin, R. 1987. Speciation by reinforcement. Trends Ecol. Evol. 2:8-13.

Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, and T. Madden. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421.

Canavan, S. V., L. C. Mayes, and H. B. Treloar. 2011. Changes in maternal gene expression in olfactory circuits in the immediate postpartum period. Frontiers in Psychiatry 2:40.

Cate, C. and D. R. Vos. 1999. Sexual imprinting and evolutionary processes in birds: a reassessment. Adv. Study Behav. 28:1-31.

Chamero, P., T. F. Marton, D. W. Logan, K. Flanagan, J. R. Cruz, A. Saghatelian, B. F. Cravatt, and L. Stowers. 2007. Identification of protein pheromones that promote aggressive behaviour. Nature 450:899-902.

Chung, Y. and C. Ané. 2011. Comparing two Bayesian methods for gene tree/species tree reconstruction: simulations with incomplete lineage sorting and horizontal gene transfer. Syst. Biol. 60:261-275.

Coyne, J. A. and H. A. Orr. 1989. Patterns of speciation in *Drosophila*. Evolution 43:362-381.

Coyne, J. A. and H. A. Orr. 1997. "Patterns of speciation in *Drosophila*" revisited. Evolution 51:295-303.

Cranston, K. A., B. Hurwitz, D. Ware, L. Stein, and R. A. Wing. 2009. Species trees from highly incongruent gene trees in rice. Syst. Biol. 58:489-500.

Crapon de Caprona, M.-D. and M. J. Ryan. 1990. Conspecific mate recognition in swordtails, *Xiphophorus nigrensis* and *X. pygmaeus* (Poeciliidae): olfactory and visual cues. Anim. Behav. 39:290-296.

Cui, R., M. Schumer, K. Kruesi, R. Walter, P. Andolfatto, and G. G. Rosenthal. 2013. Phylogenomics reveals extensive reticulate evolution in *Xiphophorus* fishes. Evolution 67:2166-2179.

Culumber, Z. W., H. S. Fisher, M. Tobler, M. Mateos, P. H. Barber, M. D. Sorenson, and G. G. Rosenthal. 2011. Replicated hybrid zones of *Xiphophorus* swordtails along an elevational gradient. Mol. Ecol. 20:342-356.

Culumber, Z. W. 2013. Pigmentation in *Xiphophorus*: An emerging system in ecological and evolutionary genetics. Zebrafish 11:57-70.

Cummings, M. and D. Mollaghan. 2006. Repeatability and consistency of female preference behaviours in a northern swordtail, *Xiphophorus nigrensis*. Anim. Behav. 72:217-224.

Darwin, C. 1859. On the origin of species. John Murray, London.

Decker, J. E., J. C. Pires, G. C. Conant, S. D. McKay, M. P. Heaton, K. Chen, A. Cooper, J. Vilkki, C. M. Seabury, A. R. Caetano, G. S. Johnson, R. A. Brenneman, O. Hanotte, L. S. Eggert, P. Wiener, J.-J. Kim, K. S. Kim, T. S. Sonstegard, C. P. Van Tassell, H. L. Neibergs, J. C. McEwan, R. Brauning, L. L. Coutinho, M. E. Babar, G. A. Wilson, M. C. McClure, M. M. Rolf, J. Kim, R. D. Schnabel, and J. F. Taylor. 2009. Resolving the evolution of extant and extinct ruminants with high-throughput phylogenomics. PNAS 106:18644-18649.

Degnan, J. H. and N. A. Rosenberg. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends Ecol. Evol. 24:332-340.

dos Reis, M., J. Inoue, M. Hasegawa, R. J. Asher, P. C. J. Donoghue, and Z. Yang. 2012. Phylogenomic datasets provide both precision and accuracy in estimating the timescale of placental mammal phylogeny. Proc. R. Soc. B 279:3491-3500.

Dulac, C., R. Olson, and P. J. Bjorkman. 2006. MHC homologs in the nervous system - they haven't lost their groove. Curr. Opin. Neurobiol. 16:351-357.

Dulka, J., N. Stacey, P. Sorensen, and G. Van Der Kraak. 1987. A steroid sex pheromone synchronizes male–female spawning readiness in goldfish. Nature 325:251-253.

Durand, E. Y., N. Patterson, D. Reich, and M. Slatkin. 2011. Testing for ancient admixture between closely related populations. Mol. Biol. Evol. 28:2239-2252.

Dzwillo, M. 1963. Einfluß von Methyltestosteron auf die Aktivierung sekundärer Geschlechtsmerkmale Über den arttypischen Ausbildungsgrad hinaus. Ver. Dent. Zool. Gesell. 1962:152-159.

Dzwillo, M. 1964. Sekundäre Geschlechtsmerkmale an der Caudalflosse einiger Xiphophorini unter dem Einfiuß von Methyltestosteron. Mitt. Ham. Zool. Mus. 61:15-22.

Endler, J. A. and A. L. Basolo. 1998. Sensory ecology, receiver biases and sexual selection. Trends Ecol. Evol. 13:415-420.

Ewing, G. and J. Hermisson. 2010. MSMS: a coalescent simulation program including recombination, demographic structure and selection at a single locus. Bioinformatics 26:2064-2065.

Fadao, T., W. Tingzheng, and Z. Yajun. 2000. Inbreeding avoidance and mate choice in the mandarin vole (*Microtus mandarinus*). Can. J. Zool. 78:2119-2125.

Falcon, S. and R. Gentleman. 2007. Using GOstats to test gene lists for GO term association. Bioinformatics 23:257-258.

Felsenstein, J. 1985. Phylogenies and the comparative method. University of Chicago Press, Chicago, IL.

Fisher, H. S., S. J. Mascuch, and G. G. Rosenthal. 2009. Multivariate male traits misalign with multivariate female preferences in the swordtail fish, *Xiphophorus birchmanni*. Anim. Behav. 78:265-269.

Fisher, H. S., B. B. Wong, and G. G. Rosenthal. 2006. Alteration of the chemical environment disrupts communication in a freshwater fish. Proc. R. Soc. B 273:1187-1193.

Fisher, R. A. 1915. The evolution of sexual preference. The Eugenics Review 7:184.

Fitzpatrick, M. J., E. Feder, L. Rowe, and M. B. Sokolowski. 2007. Maintaining a behaviour polymorphism by frequency-dependent selection on a single gene. Nature 447:210-212.

Fletcher, M. L. and D. A. Wilson. 2003. Olfactory bulb mitral-tufted cell plasticity: odorant-specific tuning reflects previous odorant exposure. J. Neurosci. 23:6946-6955.

Gerlach, G. and N. Lysiak. 2006. Kin recognition and inbreeding avoidance in zebrafish, *Danio rerio*, is based on phenotype matching. Anim. Behav. 71:1371-1377.

Gordon, M., H. Cohen, and R. F. Nigrelli. 1943. A hormone-produced taxonomic character in *Platypoecilus maculatus* diagnostic of wild *P. xiphidium*. Am. Nat. 77:569-572.

Grant, B. R. and P. R. Grant. 2008. Fission and fusion of Darwin's finches populations. Phil. Trans. R. Soc. B 363:2821-2829.

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

Green, R. E., J. Krause, A. W. Briggs, T. Maricic, U. Stenzel, M. Kircher, N. Patterson, H. Li, W. Zhai, and M. H. Y. Fritz. 2010. A draft sequence of the Neandertal genome. Science 328:710-722.

Griffiths, S. W. 2003. Learned recognition of conspecifics by fishes. Fish. Fish. 4:256-268.

Grob, B., L. Knapp, R. Martin, and G. Anzenberger. 1998. The major histocompatibility complex and mate choice: inbreeding avoidance and selection of good genes. Exp. Clin. Immunogenet. 15:119-129.

Guillaume, F. and J. Rougemont. 2006. Nemo: an evolutionary and population genetics programming framework. Bioinformatics 22:2556-2557.

Haas, V. 1992. *Xiphophorus* phylogeny, reviewed on the basis of courtship behavior. Pp. 279-288 *in* J. H. Schroder, J. Bauer, and M. Schartl, eds. Trends in ichthyology. Blackwell Scientific, London, UK.

Hailer, F., V. E. Kutschera, B. M. Hallström, D. Klassert, S. R. Fain, J. A. Leonard, U. Arnason, and A. Janke. 2012. Nuclear genomic sequences reveal that Polar Bears are an old and distinct bear lineage. Science 336:344-347.

Hankison, S. J. and M. R. Morris. 2003. Avoiding a compromise between sexual selection and species recognition: female swordtail fish assess multiple species-specific cues. Behav. Ecol. 14:282-287.

Hansen, T. F. and E. P. Martins. 1996. Translating between microevolutionary process and macroevolutionary patterns: the correlation structure of interspecific data. Evolution 50:1404-1417.

Hashiguchi, Y., Y. Furuta, and M. Nishida. 2008. Evolutionary patterns and selective pressures of odorant/pheromone receptor gene families in teleost fishes. PLoS One 3:e4083.

Hedges, V., S. Chakravarty, E. Nestler, and R. Meisel. 2009. Δ FosB overexpression in the nucleus accumbens enhances sexual reward in female Syrian hamsters. Genes Brain Behav. 8:442-449.

Heliconius Genome Consortium. 2012. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. Nature 487:94-98.

Hellenthal, G. and M. Stephens. 2007. msHOT: modifying Hudson's ms simulator to incorporate crossover and gene conversion hotspots. Bioinformatics 23:520-521.

Hittinger, C. T., M. Johnston, J. T. Tossberg, and A. Rokas. 2010. Leveraging skewed transcript abundance by RNA-Seq to increase the genomic depth of the tree of life. PNAS 107:1476-1481.

Hoban, S., G. Bertorelle, and O. E. Gaggiotti. 2012. Computer simulations: tools for population and evolutionary genetics. Nat. Rev. Genet. 13:110-122.

Ihara, Y., K. Aoki, and M. W. Feldman. 2003. Runaway sexual selection with paternal transmission of the male trait and gene-culture determination of the female preference. Theor. Popul. Biol. 63:53-62.

Irwin, D. E. and T. Price. 1999. Sexual imprinting, learning and speciation. Heredity 82:347-354.

Jiggins, C. D. and J. Mallet. 2000. Bimodal hybrid zones and speciation. Trends Ecol. Evol. 15:250-255.

Kallman, K. D. and S. Kazianis. 2006. The genus *Xiphophorus* in Mexico and Central America. Zebrafish 3:271-285.

Kaneshiro, K. Y. 1980. Sexual isolation, speciation and the direction of evolution. Evolution 34:437-444.

Kang, J. H., M. Schartl, R. Walter, and A. Meyer. 2013. Comprehensive phylogenetic analysis of all species of swordtails and platies (Pisces: Genus *Xiphophorus*) uncovers a hybrid origin of a swordtail fish, *Xiphophorus monticolus*, and demonstrates that the sexually selected sword originated in the ancestral lineage of the genus, but was lost again secondarily. BMC Evol. Biol. 13:25.

Kelz, M. B., J. Chen, W. A. Carlezon, K. Whisler, L. Gilden, A. M. Beckmann, C. Steffen, Y.-J. Zhang, L. Marotti, D. W. Self, T. Tkatch, G. Baranauskas, D. J. Surmeier, R. L. Neve, R. S. Duman, M. R. Picciotto, and E. J. Nestler. 1999. Expression of the transcription factor ΔFosB in the brain controls sensitivity to cocaine. Nature 401:272-276.

Kermen, F., L. M. Franco, C. Wyatt, and E. Yaksi. 2013. Neural circuits mediating olfactory-driven behavior in fish. Frontiers in Neural Circuits. 7:62.

Kirkpatrick, M. and L. A. Dugatkin. 1994. Sexual selection and the evolutionary effects of copying mate choice. Behav. Ecol. Sociobiol. 34:443-449.

Kirkpatrick, M. and M. J. Ryan. 1991. The evolution of mating preferences and the paradox of the lek. Nature 350:33-38.

Kocot, K. M., J. T. Cannon, C. Todt, M. R. Citarella, A. B. Kohn, A. Meyer, S. R. Santos, C. Schander, L. L. Moroz, B. Lieb, and K. M. Halanych. 2011. Phylogenomics reveals deep molluscan relationships. Nature 477:452-456.

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. R. Soc. B. 278:2604-2610.

Kubatko, L. S. 2009. Identifying hybridization events in the presence of coalescence via model selection. Syst. Biol. 58:478-488.

Laland, K. N. 1994a. On the evolutionary consequences of sexual imprinting. Evolution 48:477-489.

Laland, K. N. 1994b. Sexual selection with a culturally transmitted mating preference. Theor. Popul. Biol. 45:1-15.

Larget, B. R., S. K. Kotha, C. N. Dewey, and C. Ané. 2010. BUCKy: Gene tree/species tree reconciliation with Bayesian concordance analysis. Bioinformatics 26:2910-2911.

Lassance, J.-M. and C. Löfstedt. 2013. Chemical communication: A jewel sheds light on signal evolution. Curr. Biol. 23:R346-R348.

Lawniczak, M. K. and D. J. Begun. 2004. A genome-wide analysis of courting and mating responses in *Drosophila melanogaster* females. Genome 47:900-910.

Lazarini, F. and P.-M. Lledo. 2011. Is adult neurogenesis essential for olfaction? Trends Neurosci. 34:20-30.

Lehmann, L. and N. Perrin. 2003. Inbreeding avoidance through kin recognition: choosy females boost male dispersal. Am. Nat. 162:638-652.

Leinders-Zufall, T., P. Brennan, P. Widmayer, P. Chandramani, A. Maul-Pavicic, M. Jager, X. H. Li, H. Breer, F. Zufall, and T. Boehm. 2004. MHC class I peptides as chemosensory signals in the vomeronasal organ. Science 306:1033-1037.

Lewcock, J. W. and R. R. Reed. 2004. A feedback mechanism regulates monoallelic odorant receptor expression. PNAS 101:1069-1074.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25:2078-2079.

Liu, L. 2008. BEST: Bayesian estimation of species trees under the coalescent model. Bioinformatics 24:2542-2543.

Lohse, K. and L. A. Frantz. 2014. Neandertal admixture in Eurasia confirmed by maximum likelihood analysis of three genomes. Genetics genetics-114.162396.

Luikart, G., F. Allendorf, J.-M. Cornuet, and W. Sherwin. 1998. Distortion of allele frequency distributions provides a test for recent population bottlenecks. J. Hered. 89:238-247.

Lunter, G. and M. Goodson. 2011. Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. Genome Res. 21:936-939.

Maddison, W. and D. Maddison. 2011. Mesquite: a modular system for evolutionary analysis. Version 2.75 http://mesquiteproject.org

Mallet, J. 2005. Hybridization as an invasion of the genome. Trends Ecol. Evol. 20:229-237.

Mallet, J. 2007. Hybrid speciation. Nature 446:279-283.

Mallet, J. 2008. Hybridization, ecological races and the nature of species: empirical evidence for the ease of speciation. Philos. Trans. R. Soc. Lond. B Biol. Sci. 363:2971-2986.

Marcillac, F., Y. Grosjean, and J.-F. Ferveur. 2005. A single mutation alters production and discrimination of *Drosophila* sex pheromones. Proc. R. Soc. B 272:303-309.

Marcus, J. M. and A. R. McCune. 1999. Ontogeny and phylogeny in the northern swordtail clade of *Xiphophorus*. Syst. Biol. 48:491-522.

Marler, P. 1991. The instinct to learn. Pp. 37-66 *In* S. Carey and R. Gelman eds. The epigenesis of mind: essays on biology and cognition. Lawrence Erlbaum Associates, Inc., Hillsdale, NJ.

Masly, J. P. and D. C. Presgraves. 2007. High-resolution genome-wide dissection of the two rules of speciation in *Drosophila*. PLoS Biol. 5:e243.

Mateos, M., O. I. Sanjur, and R. C. Vrijenhoek. 2002. Historical biogeography of the livebearing fish genus *Poeciliopsis* (Poeciliidae: Cyprinodontiformes). Evolution 56:972-984.

McLennan, D. A. and M. J. Ryan. 1999. Interspecific recognition and discrimination based upon olfactory cues in northern swordtails. Evolution 53:880-888.

Meierjohann, S. and M. Schartl. 2006. From Mendelian to molecular genetics: the *Xiphophorus* melanoma model. Trends Genet. 22:654-661.

Mendelson, T. C. 2003. Sexual isolation evolves faster than hybrid inviability in a diverse and sexually dimorphic genus of fish (Percidae: Etheostoma). Evolution 57:317-327.

Meng, C. and L. S. Kubatko. 2009. Detecting hybrid speciation in the presence of incomplete lineage sorting using gene tree incongruence: a model. Theor. Popul. Biol. 75:35-45.

Meyer, A., J. M. Morrissey, and M. Schartl. 1994. Recurrent origin of a sexually selected trait in *Xiphophorus* fishes inferred from a molecular phylogeny. Nature 368:539-542.

Meyer, A., W. Salzburger, and M. Schartl. 2006. Hybrid origin of a swordtail species (Teleostei : *Xiphophorus clemenciae*) driven by sexual selection. Mol. Ecol. 15:721-730.

Midford, P. E., T. G. Jr., and W. P. Maddison. 2011. PDAP: PDTREE package for Mesquite. Version 1.16 http://mesquiteproject.org.

Morizot, D. C., S. A. Slaugenhaupt, K. D. Kallman, and A. Chakravarti. 1991. Genetic linkage map of fishes of the genus *Xiphophorus* (Teleostei: Poeciliidae). Genetics 127:399-410.

Morris, M. R., K. de Queiroz, D. C. Morizot, and J. D. McEachran. 2001. Phylogenetic relationships among populations of northern swordtails (*Xiphophorus*) as inferred from allozyme data. Copeia 2001:65-81.

Nabholz, B., A. Künstner, R. Wang, E. D. Jarvis, and H. Ellegren. 2011. Dynamic evolution of base composition: causes and consequences in avian phylogenomics. Mol. Biol. Evol. 28:2197-2210.

Nestler, E. J., M. Barrot, and D. W. Self. 2001. ΔFosB: A sustained molecular switch for addiction. PNAS 98:11042-11046.

Nevitt, G. A., A. H. Dittman, T. P. Quinn, and W. J. Moody. 1994. Evidence for a peripheral olfactory memory in imprinted salmon. PNAS 91:4288-4292.

Niehuis, O., J. Buellesbach, J. D. Gibson, D. Pothmann, C. Hanner, N. S. Mutti, A. K. Judson, J. Gadau, J. Ruther, and T. Schmitt. 2013. Behavioural and genetic analyses of *Nasonia* shed light on the evolution of sex pheromones. Nature 494:345-348.

Offen, N. 2008. The molecular basis of development of the sword, a sexual selected trait in the genus *Xiphophorus*. PhD. thesis, Mathematisch-naturwissenschaftliche Sektion Fachbereich Biologie. Universität Konstanz, Konstanz, Germany.

Owens, I. P. F., C. Rowe, and A. L. R. Thomas. 1999. Sexual selection, speciation and imprinting: separating the sheep from the goats. Trends Ecol. Evol. 14:131-132.

Parry, J. W., K. L. Carleton, T. Spady, A. Carboo, D. M. Hunt, and J. K. Bowmaker. 2005. Mix and match color vision: tuning spectral sensitivity by differential opsin gene expression in Lake Malawi cichlids. Curr. Biol. 15:1734-1739.

Partan, S. R. and P. Marler. 2005. Issues in the classification of multimodal communication signals. Am. Nat. 166:231-245.

Peng, B. and C. I. Amos. 2008. Forward-time simulations of non-random mating populations using simuPOP. Bioinformatics 24:1408-1409.

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

Pfister, P. and I. Rodriguez. 2005. Olfactory expression of a single and highly variable V1r pheromone receptor-like gene in fish species. PNAS 102:5489-5494.

Pickrell, J. K. and J. K. Pritchard. 2012. Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genet 8:e1002967.

Pilastro, A., S. Benetton, and A. Bisazza. 2003. Female aggregation and male competition reduce costs of sexual harassment in the mosquitofish *Gambusia holbrooki*. Anim. Behav. 65:1161-1167.

Pitchers, K. K., K. S. Frohmader, V. Vialou, E. Mouzon, E. J. Nestler, M. N. Lehman, and L. M. Coolen. 2010. ΔFosB in the nucleus accumbens is critical for reinforcing effects of sexual reward. Genes Brain Behav. 9:831-840.

Plenderleith, M., C. van Oosterhout, R. L. Robinson, and G. F. Turner. 2005. Female preference for conspecific males based on olfactory cues in a Lake Malawi cichlid fish. Biol. Lett. 1:411-414.

Pollard, D. A., V. N. Iyer, A. M. Moses, and M. B. Eisen. 2006. Widespread discordance of gene trees with species tree in *Drosophila*: evidence for incomplete lineage sorting. PLoS Genetics 2:e173.

Presgraves, D. C., L. Balagopalan, S. M. Abmayr, and H. A. Orr. 2003. Adaptive evolution drives divergence of a hybrid inviability gene between two species of *Drosophila*. Nature 423:715-719.

Pusey, A. E. 1980. Inbreeding avoidance in chimpanzees. Anim. Behav. 28:543-552.

Qvarnström, A., V. Blomgren, C. Wiley, and N. Svedin. 2004. Female collared flycatchers learn to prefer males with an artificial novel ornament. Behav. Ecol. 15:543-548.

R Development Core Team. 2010. R version 2.15.2. R project for statistical computing Vienna, Austria.

Rauchenberger, M., K. D. Kallman, and D. C. Morizot. 1990. Monophyly and geography of the Río Pánuco basin swordtails (genus *Xiphophorus*) with descriptions of four new species. Am. Mus. Novit. no. 2975:1-41.

Rieseberg, L. H., O. Raymond, D. M. Rosenthal, Z. Lai, K. Livingstone, T. Nakazato, J. L. Durphy, A. E. Schwarzbach, L. A. Donovan, and C. Lexer. 2003. Major ecological transitions in wild sunflowers facilitated by hybridization. Science 301:1211-1216.

Robinson, M. and A. Oshlack. 2010. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biology 11:R25.

Robinson, M. and G. Smyth. 2007. Moderated statistical tests for assessing differences in tag abundance. Bioinformatics 23:2881 - 2887.

Robinson, M. and G. Smyth. 2008. Small-sample estimation of negative binomial dispersion, with applications to SAGE data. Biostatistics 9:321 - 332.

Rodriguez, I., C. A. Greer, M. Y. Mok, and P. Mombaerts. 2000. A putative pheromone receptor gene expressed in human olfactory mucosa. Nat. Genet. 26:18-19.

Rosen, D. E. 1960. Middle-American poeciliid fishes of the genus *Xiphophorus*. Bull. Fla. Mus. Nat. Hist. 5:58-242.

Rosen, D. E. 1979. Fishes from the uplands and intermontane basins of Guatemala : revisionary studies and comparative geography. Bull. Am. Mus. Nat. Hist. 162:267-376.

Rosenthal, G. and M. Ryan. 2000. Visual and acoustic communication in non-human animals: a comparison. J. Biosci. 25:285-290.

Rosenthal, G. G., X. F. de la Rosa Reyna, S. Kazianis, M. J. Stephens, D. C. Morizot, M. J. Ryan, F. J. García de León, and R. E. Wood. 2003. Dissolution of sexual signal complexes in a hybrid zone between the swordtails *Xiphophorus birchmanni* and *Xiphophorus malinche* (Poeciliidae). Copeia 2003:299-307.

Rosenthal, G. G. and C. S. Evans. 1998. Female preference for swords in *Xiphophorus helleri* reflects a bias for large apparent size. PNAS 95:4431-4436.

Rosenthal, G. G., J. N. Fitzsimmons, K. U. Woods, G. Gerlach, and H. S. Fisher. 2011. Tactical release of a sexually-selected pheromone in a swordtail fish. PLoS One 6:e16994.

Rosenthal, G. G. and M. J. Ryan. 2011. Conflicting preferences within females: sexual selection versus species recognition. Biol. Lett. 7:525-527.

Rowe, C. 1999. Receiver psychology and the evolution of multicomponent signals. Anim. Behav. 58:921-931.

Ryan, M. J. 1990. Sexual selection, sensory systems and sensory exploitation. Oxf. Surv. Evol. Biol. 7:157-195.

Ryan, M. J. 1998. Sexual selection, receiver biases, and the evolution of sex differences. Science 281:1999-2003.

Ryan, M. J., D. K. Hews, and W. E. Wagner Jr. 1990. Sexual selection on alleles that determine body size in the swordtail *Xiphophorus nigrensis*. Behav. Ecol. Sociobiol. 26:231-237.

Ryan, M. J. and W. Wilczynski. 1988. Coevolution of sender and receiver: effect on local mate preference in cricket frogs. Science 240:1786-1788.

Schartl, M. 1995. Platyfish and swordtails: a genetic system for the analysis of molecular mechanisms in tumor formation. Trends Genet. 11:185-189.

Schartl, M. 2008. Evolution of Xmrk: an oncogene, but also a speciation gene? Bioessays 30:822-832.

Schartl, M., R. B. Walter, Y. Shen, T. Garcia, J. Catchen, A. Amores, I. Braasch, D. Chalopin, J.-N. Volff, and K.-P. Lesch. 2013. The genome of the platyfish, *Xiphophorus maculatus*, provides insights into evolutionary adaptation and several complex traits. Nat. Genet. 45:567-572

Schlicker, A., F. Domingues, J. Rahnenfuhrer, and T. Lengauer. 2006. A new measure for functional similarity of gene products based on Gene Ontology. BMC Bioinformatics 7:302.

Schmidt, H. A. 2009. Testing tree topologies. Pp. 381-395 *in* P. Lemey, M. Salemi, and A.-M. Vandamme, eds. The phylogenetic handbook - a practical approach to

phylogenetic analysis and hypothesis testing. Cambridge University Press, Cambridge, UK.

Schulz, M. H., D. R. Zerbino, M. Vingron, and E. Birney. 2012. Oases: robust *de novo* RNA-seq assembly across the dynamic range of expression levels. Bioinformatics 28:1086-1092.

Schumer, M., R. Cui, B. Boussau, R. Walter, G. Rosenthal, and P. Andolfatto. 2013. An evaluation of the hybrid speciation hypothesis for *Xiphophorus clemenciae* based on whole genome sequences. Evolution 67:1155-1168.

Schumer, M., R. Cui, D. Powell, R. Dresner, G. Rosenthal, and P. Andolfatto. In review. High-resolution mapping reveals hundreds of genetic incompatibilities in hybridizing fish species.

Schwarzer, J., B. Misof, and U. K. Schliewen. 2012. Speciation within genomic networks: a case study based on *Steatocranus* cichlids of the lower Congo rapids. J. Evol. Biol. 25:138-148.

Seehausen, O. 2004. Hybridization and adaptive radiation. Trends Ecol. Evol. 19:198-207.

Seehausen, O., Y. Terai, I. S. Magalhaes, K. L. Carleton, H. D. J. Mrosso, R. Miyagi, I. van der Sluijs, M. V. Schneider, M. E. Maan, H. Tachida, H. Imai, and N. Okada. 2008. Speciation through sensory drive in cichlid fish. Nature 455:620-626.

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

Servedio, M. R., S. A. Saether, and G. P. Saetre. 2009. Reinforcement and learning. Evol. Ecol. 23:109-123.

Shimodaira, H. 2002. An approximately unbiased test of phylogenetic tree selection. Syst. Biol. 51:492-508.

Shimodaira, H. and M. Hasegawa. 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics 17:1246-1247.

Shykind, B. M. 2005. Regulation of odorant receptors: one allele at a time. Hum. Mol. Genet. 14:R33-R39.

Silva, A. J., J. H. Kogan, P. W. Frankland, and S. Kida. 1998. CREB and memory. Annu. Rev. Neurosci. 21:127-148.

Smith, S. A., N. G. Wilson, F. E. Goetz, C. Feehery, S. C. S. Andrade, G. W. Rouse, G. Giribet, and C. W. Dunn. 2011. Resolving the evolutionary relationships of molluscs with phylogenomic tools. Nature 480:364-367.

Sorensen, P. W., J. M. Fine, V. Dvornikovs, C. S. Jeffrey, F. Shao, J. Z. Wang, L. A. Vrieze, K. R. Anderson, and T. R. Hoye. 2005. Mixture of new sulfated steroids functions as a migratory pheromone in the sea lamprey. Nat. Chem. Biol. 1:324-328.

Sorensen, P. W., T. J. Hara, N. E. Stacey, and J. G. Dulka. 1990. Extreme olfactory specificity of male goldfish to the preovulatory steroidal pheromone 17α,20β-dihydroxy-4-pregnen-3-one. J. Comp. Physiol. A 166:373-383.

Sorensen, P. W., H. M. Levesque, D. Scaffidi, and C. N. Polkinghorne. 2011. A multi-component species identifying pheromone in the goldfish. J. Chem. Ecol. 37:219-227.

Speca, D. J., D. M. Lin, P. W. Sorensen, E. Y. Isacoff, J. Ngai, and A. H. Dittman. 1999. Functional identification of a goldfish odorant receptor. Neuron 23:487-498.

Stamatakis, A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688-2690.

Stamatakis, A., P. Hoover, and J. Rougemont. 2008. A rapid bootstrap algorithm for the RAxML web servers. Syst. Biol. 57:758-771.

Sunnåker, M., A. G. Busetto, E. Numminen, J. Corander, M. Foll, and C. Dessimoz. 2013. Approximate bayesian computation. PLoS Computational Biology 9:e1002803.

Thornton, T., H. Tang, T. J. Hoffmann, H. M. Ochs-Balcom, B. J. Caan, and N. Risch. 2012. Estimating kinship in admixed populations. Am. J. Hum. Gen. 91(1):122-138.

Tramm, N. A. and M. R. Servedio. 2008. Evolution of mate-choice imprinting: competing strategies. Evolution 62:1991-2003.

Tudor, M. S. and M. R. Morris. 2009. Experience plays a role in female preference for symmetry in the swordtail fish *Xiphophorus malinche*. Ethology 115:812-822.

Ulery, P. G., G. Rudenko, and E. J. Nestler. 2006. Regulation of ΔFosB stability by phosphorylation. J. Neurosci. 26:5131-5142.

Verzijden, M. N., Z. W. Culumber, and G. G. Rosenthal. 2012a. Opposite effects of learning cause asymmetric mate preferences in hybridizing species. Behav. Ecol. 23:1133-1139.

Verzijden, M. N., R. E. M. Korthof, and C. ten Cate. 2008. Females learn from mothers and males learn from others. The effect of mother and siblings on the development of female mate preferences and male aggression biases in Lake Victoria cichlids, genus *Mbipia*. Behav. Ecol. Sociobiol. 62:1359-1368.

Verzijden, M. N., R. F. Lachlan, and M. R. Servedio. 2005. Female mate-choice behavior and sympatric speciation. Evolution 59:2097-2108.

Verzijden, M. N. and G. G. Rosenthal. 2011. Effects of sensory modality on learned mate preferences in female swordtails. Anim. Behav. 82:557-562.

Verzijden, M. N. and C. ten Cate. 2007. Early learning influences species assortative mating preferences in Lake Victoria cichlid fish. Biol. Lett. 3:134-136.

Verzijden, M. N., C. Ten Cate, M. R. Servedio, G. M. Kozak, J. W. Boughman, and E. I. Svensson. 2012b. The impact of learning on sexual selection and speciation. Trends Ecol. Evol. 27:511-519.

Vitti, J. J., S. R. Grossman, and P. C. Sabeti. 2013. Detecting natural selection in genomic data. Annu. Rev. Genet. 47:97-120.

Volff, J.-N. and M. Schartl. 2001. Variability of genetic sex determination in poeciliid fishes. Genetica 111:101-110.

Vos, D. R. 1995. Sexual imprinting in zebra-finch females: do females develop a preference for males that look like their father? Ethology 99:252-262.

Walling, C., N. Royle, J. Lindström, and N. Metcalfe. 2010. Do female association preferences predict the likelihood of reproduction? Behav. Ecol. Sociobiol. 64:541-548.

Walter, R., J. Rains, J. Russell, T. Guerra, C. Daniels, D. A. Johnston, J. Kumar, A. Wheeler, K. Kelnar, and V. Khanolkar. 2004a. A microsatellite genetic linkage map for *Xiphophorus*. Genetics 168:363-372.

Walter, R. B., J. D. Rains, J. E. Russell, T. M. Guerra, C. Daniels, D. A. Johnston, J. Kumar, A. Wheeler, K. Kelnar, V. A. Khanolkar, E. L. Williams, J. L. Hornecker, L. Hollek, M. M. Mamerow, A. Pedroza, and S. Kazianis. 2004b. A microsatellite genetic linkage map for *Xiphophorus*. Genetics 168:363-372.

Weisfeld, G. E., T. Czilli, K. A. Phillips, J. A. Gall, and C. M. Lichtman. 2003. Possible olfaction-based mechanisms in human kin recognition and inbreeding avoidance. J. Exp. Child Psychol. 85:279-295.

Westerman, E. L., A. Hodgins-Davis, A. Dinwiddie, and A. Monteiro. 2012. Biased learning affects mate choice in a butterfly. PNAS 109:10948-10953.

Whitlock, K. E., M. V. Harden, L. A. Newton, and R. C. Lloyd. 2006. Olfactory imprinting is correlated with changes in gene expression in the olfactory epithelia of the zebrafish. J. Neurobiol. 66:1452-1466.

Willis, P. M., G. G. Rosenthal, and M. J. Ryan. 2012. An indirect cue of predation risk counteracts female preference for conspecifics in a naturally hybridizing fish *Xiphophorus birchmanni*. PloS One 7:e34802.

Willis, P. M., M. J. Ryan, and G. G. Rosenthal. 2011. Encounter rates with conspecific males influence female mate choice in a naturally hybridizing fish. Behav. Ecol. 22:1234-1240.

Wilson, D. A. and R. J. Stevenson. 2003. The fundamental role of memory in olfactory perception. Trends Neurosci. 26:243-247.

Witte, K. and N. Sawka. 2003. Sexual imprinting on a novel trait in the dimorphic zebra finch: sexes differ. Anim. Behav. 65:195-203.

Wolf, J. B., E. D. Brodie III, J. M. Cheverud, A. J. Moore, and M. J. Wade. 1998. Evolutionary consequences of indirect genetic effects. Trends Ecol. Evol. 13:64-69.

Wong, Bob B. M. and Gil G. Rosenthal. 2006. Female disdain for swords in a swordtail fish. Am. Nat. 167:136-140.

Wood, M. A., M. A. Attner, A. M. M. Oliveira, P. K. Brindle, and T. Abel. 2006. A transcription factor-binding domain of the coactivator CBP is essential for long-term memory and the expression of specific target genes. Learn. Memory 13:609-617.

Wyatt, T. D. 2003. Pheromones and animal behaviour: communication by smell and taste. Cambridge University Press, Cambridge, United Kingdom.

Yang, H., P. Shi, Y.-p. Zhang, and J. Zhang. 2005. Composition and evolution of the V2r vomeronasal receptor gene repertoire in mice and rats. Genomics 86:306-315.

Zander, C. D. and M. Dzwillo. 1969. Untersuchungen zur Entwicklung und Vererbung des Caudalfortsatzes der *Xiphophorus*-Arten (Pisces). Zeitschr. Wissensch. Zool. 178:275-315.

Zerbino, D. and E. Birney. 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. Genome Res. 18:821-829.

APPENDIX A

ADMIXSIMUL: A FLEXIBLE FRAMEWORK FOR FORWARD-TIME INDIVIDUAL-BASED POPULATION SIMULATION WITH REALISTIC MATE CHOICE

i. Introduction

Population genetic simulations are invaluable to empiricists analyzing population-genetic data, since they can be used to test alternative hypotheses, novel algorithms, and recently, in approximated Bayesian computations (Beaumont et al. 2002). A recent review showed that no simulators currently allow for complex selection regimes that simultaneously incorporate linkage, epistasis and multi-phenotype selection for an unlimited number of loci (Hoban et al. 2012). For instance, a commonly adopted simulation package, msHOT (Hellenthal and Stephens 2007), does not allow for selection; while another version, msms (Ewing and Hermisson 2010), only allows selection on a single locus. Some simulators, while allowing for complex demographic history, life cycle and natural selection (Guillaume and Rougemont 2006), either do not allow for sexual selection or indirect genetic effects (genotype of other individuals affect phenotype of the focal individual; Verzijden et al. 2012b) or require further programming by the user (Peng and Amos 2008).

The aim of Admixsimul is to provide a general, configuration-based application that allows simulations to be conducted using biologically realistic parameters. Admixsimul differs from similar programs in that it takes an individual-centric view, such that experimental parameters obtained on individual levels, such as mate preference (Kirkpatrick and Ryan 1991), can be incorporated in the model. Admixsimul deals with complex epistatic interactions as well as highly complex selection regimes. In addition to biallelic functional loci, neutral biallelic SNP markers can also be generated throughout the genome, connected by user-defined, sex-specific linkage maps. Individual genotypes can then be exported to mapping or cline analysis software. The program is implemented in C++ with an object-oriented structure, facilitating extension of functionalities. Compilation was tested on win32, win64, Linux x64 platforms with OpenMP.
Under the current implementation, admixsimul makes the following assumptions: 1) non-overlapping generations; 2) diploid sexual individuals; 3) only one sex (assumed to be female) exercises mate choice.

ii. Algorithms

Recombination

Admixsimul tracks ancestry and allelic states of all markers and genes for each living individual. Instead of simulating full nucleotide sequences, admixsimul places candidate breakpoints (CB) on the genome, such that at least one CB is present between user-defined markers or genes. The accumulative recombination probability can then be used to determine the stochastic selection of CBs during recombination. When the number of pre-defined CB is fixed, the algorithm scales at a linear time O(*M*) with respect to the number of markers *M* and recombination fractions *R* (Suppl. i). Admixsimul is able to simulate recombination with biologically realistic genome sizes (~2000cM), marker numbers (~10,000) and population sizes (1,000~10,000) within a tractable time period.

Mutation

Currently Admixsimul supports perturbation of allelic values of functional loci with locus-wise mutation probabilities. Users can set the new allelic value using a mathematical formula and limit the values within lower and upper bounds. This capability allows full control over mutational bias and mutational rate.

Phenotypes

The program supports an unlimited number of phenotypes as arbitrary functions of allelic states with optional random noise. This flexibility makes possible any complex additive, dominance or epistatic effects, including sex-limited traits. Sex is modeled as simply another quantitative trait, which allows the program to model a range of sex-determination mechanisms (Volff and Schartl 2001).

Natural and sexual selection

Admixsimul allows for complex selection regimes dependent not only on the individual's phenotypes, but also on interactions with other conspecifics. For each population, users can specify arbitrary fitness or mating probability functions that evaluate to the range [0 , 1]. For example, stabilizing selection can be expressed with a Gaussian function $e^{-(T-T_{opt})^2/2\sigma^2}$, where T is the phenotype of the focal individual, T_{opt} is the optimum phenotype and σ is the phenotypic standard deviation. Available gamete numbers (with variation) of females can be defined such that females are able to produce more offspring than the carrying capacity.

Variables summarized from population statistics and individual phenotypes are dynamically supplied during run-time. This flexibility facilitates implementation of important models including indirect genetic effects (Wolf et al. 1998) such as learned mating preferences (paternal, maternal and peer imprinting, Verzijden et al. 2012b) as well as frequency-dependent natural and sexual selection (Fitzpatrick et al. 2007). For example, the mate preference function under paternal imprinting can be modeled as $e^{-(S-S_{father})^2/2\sigma^2}$, where S is the signaling trait of a courting male and S_{father} the trait of the female's father. Currently, mate searching is conducted randomly with a user-defined

number of mates to assess. When the allowed number of assessed mates is equal to the male population size, females are guaranteed to mate with their most-preferred male.

Population

Currently Admixsimul accommodates two parental populations and an unlimited number of admixing populations. Pair-wise migration rates between parental and admixed populations can be supplied by the user. When certain rates are set to 0, the model can be reduced to the stepping stone, or island model. Migration rates can be set differently for any specific generation for any pair-wise populations. Carrying capacity can differ between populations. While this value cannot vary by generation in the current version, a bottleneck effect can be simulated simply by setting a reduced generation-specific fitness probability in the natural selection file.

iii. Outputs

Admixsimul outputs states of functional and neutral loci, as well as phenotypic values for all individuals for user-defined ranges of generations. These outputs can then be converted to other file formats for analyses. Since the program tracks parent-offspring relationships, it is straightforward to reconstruct full pedigrees from the simulated data. Currently scripts are supplied to convert SNP data to Admixmap for mapping analysis, and to calculate and visualize F_{IS} statistics, ancestry distribution and phenotype distributions. When neutral markers are turned off, Admixsimul can be used as a general-purpose simulator for evolutionary models that involve quantitative trait loci.

iv. Parallelization

Admixsimul is parallelized with OpenMP on a multi-core, shared-memory computer and performance scales linearly with number of CPUs. Under this implementation, simulations are not repeatable due to randomness in the execution order. Extension to OpenMPI is planned in future versions.

v. Performance evaluation of recombination algorithm

We simulated random mating of individuals with recombination in a single population for 10 generations to evaluate the speed and scaling of the recombination algorithm. Because population size fluctuates around the carrying capacity of 1000, we normalize the CPU time to 1000 individuals by dividing observed CPU time by the observed population size multiplied by 1000.

First, we evaluated how an increase in recombination rate (number of break points per meiosis) affects speed. We put 10,000 predefined breakpoints (CBs) and markers on 10 chromosomes, totaled 100,000 CBs and markers. Note that in real applications, the number of CBs and markers are likely much lower. We started from a 1x rate, which assumes that on average two breakpoints occur per chromosome per meiosis. This assumption is very close or higher than most realistic scenarios. We then scale up the recombination rate to 10x, 50x and 100x to obtain a linear regression. The

result shows that processing time increases linearly, but slowly, with recombination rate, (Time = 0.2691*Recombination_Rate + 13.596). It shows that time increases slowly compared to the baseline time required for recombination. For example, increasing recombination rate to 2x adds 0.5s to the baseline 13.6s. Because biologically realistic recombination rates seldom exceeds our 1x condition, the effects of increase in recombination rate are largely negligible.

Second we evaluated the effect of marker number on speed. We started from the same setup as above, and thinned the number of markers (while keeping the CB numbers as before) to 10,000, 5,000 and 3,333. In real applications, the number of CB can reduce with the markers without any adverse effects. Results show that computation time scales linearly with number of simulated markers. On average, adding 10,000 markers results in 1 added second in computation time per 1000 individuals.

All simulations were performed on a computer node with 8-core Intel Xeon E5420 CPU at 2.5 GHz, 32GB RAM running CentOS 5.7 x86_64.

vi. Application example

Here we present a case study using admixsimul to simulate both natural and sexual selection acting in concert on a genomic cline.

Genomic architecture

We simulated 10 chromosomes each 87cM long, with and 10 neutral markers evenly placed on each (we set the number of CB to 100,000, a number that well exceeds marker numbers). Chromosomes 1-8 each contain an additive locus coding for the male trait, which the females based their preference on. We then modeled the female preference as a learned preference. A locus on Chromosome 9 determines whether the females prefer or avoid the phenotypes they learned during sexual imprinting (Bailey and Moore 2012). For traits under thermal selection, on chromosome 1 we place a locus for cold tolerance, and on chromosome 6 we place a locus for heat tolerance. Both loci are 5.5Mb (14.5cM) away from the male trait locus. Chromosome 10 harbors a single hemizygous locus for sex determination, which results in a 1:1 sex ratio.

Natural selection

We use a threshold gamma function to simulate the effects of heat and cold tolerance effects conferred by a combination of these two loci given the highest and lowest temperatures of the habitat on the cline using the following algorithm:

```
min(
        if (ColdTol>0,
                1 - 1/(1+exp((lowtmp-0)*0.5)) ,
                1 - 1/(1+exp((lowtmp-8)*0.5))
        ) ,
        if (HeatTol>0,
                1/(1+exp((hightmp-38)*0.5)),
                1/(1+exp((temp-25)*0.5))
        )
)
```

Values of ColdTol and HeatTol indicate whether the individual harbors the cold and heat tolerant alleles. When ColdTol locus is >0, the individual can tolerate a temperature as low as 0, otherwise 8. When HeatTol locus is >0, the individual can tolerate up to 38 degrees, but only 25 degrees otherwise.
We vary the temperatures linearly based on a cline of the 10 populations by a 0.5 or 1 degree increment (lowtemp starting from 2, hightemp starting from 10). We also seasonally varied the average temperature. In addition, we simulated global warming by adding 0.1 degrees per year to the average temperature.

Sexual selection

We simulate mate preference based on sexual imprinting.

```
if (Psi==0 ,
        1,
         if(Psi>0,
                  exp( -pow( PrevGenCurrPopCourter_Avg_Signal  - Courter_Signal , 2)
                 /(2*pow(2 , 2))) ,
                 1 - exp( -pow( PrevGenCurrPopCourter_Avg_Signal - Courter_Signal ,
                 2)
                 /(2*pow(6, 2)))
        )
)
```

The mate preference depends on the phenotype Psi, which promotes preference for the population average male trait from the previous generation and avoids such preference when smaller than 0.

The cline

We simulated a cline by placing 8 intermediate hybrid populations between two extreme, parental populations and setting appropriate per-generation migration rules, such that gene flow is symmetrical from both directions. Parental population 1 has highest signaling trait values and negative Psi, while parental population 2 has the opposite. We let the cline reach equilibrium with only sexual selection for 100 generations, and then we start applying natural selection rules at generation 101 for 20 generations. We then visually examine the simulation results by plotting trait values (cold tolerance and male trait) against generations and elevation in a 3D surface plot.

Results

Admixsimul is able to simultaneously simulate sexual and natural selection with phenotypes determined by both complex genetic architectures and indirect genetic effects. Before natural selection is in place, sexual selection due to difference in learning behavior generates a concave cline compared to random mating expectations on both thermal tolerance (Figure A-3) and signaling traits (Figure A-4). After natural selection occurs, the cline for thermal tolerance quickly becomes convex but due to weak physical linkage, signaling trait cline remains concave.

Figures

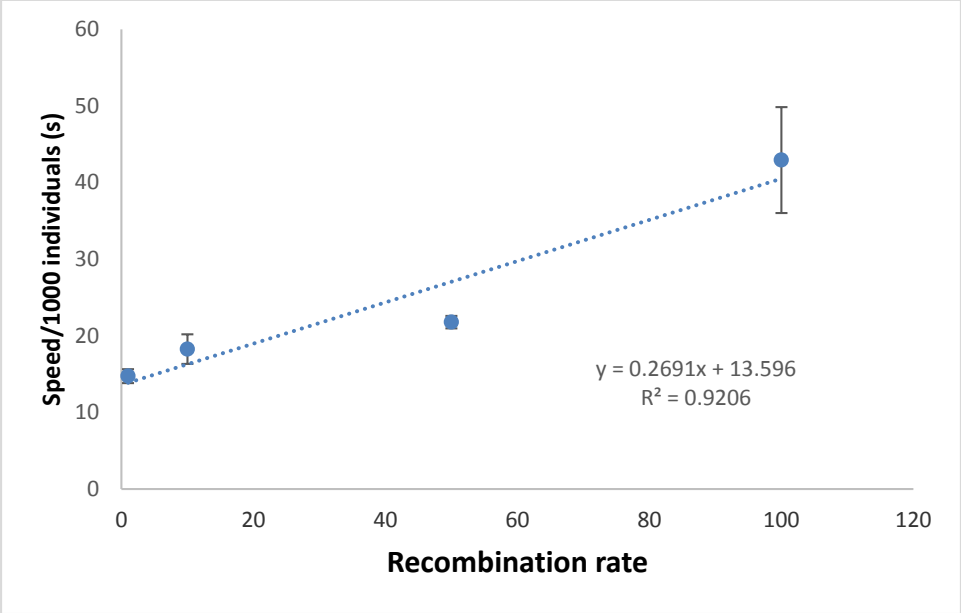


Figure A-1. Linear regression of computational time (seconds per 1000 individuals) on recombination rate, where 1 unit = 2 break points/chromosome/generation. Computation performed on a computer node with 8-core Intel Xeon E5420 CPU at 2.5 GHz, 32GB RAM.

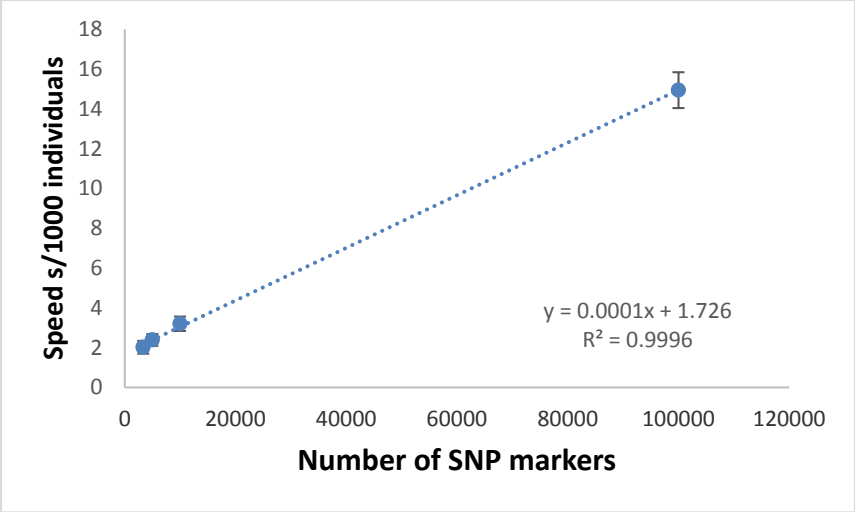


Figure A-2. Linear regression of computational time (seconds per 1000 individuals) on number of SNP markers simulated (number of predefined breakpoints fixed at 100,000). Computation performed on a computer node with 8-core Intel Xeon E5420 CPU at 2.5 GHz, 32GB RAM.

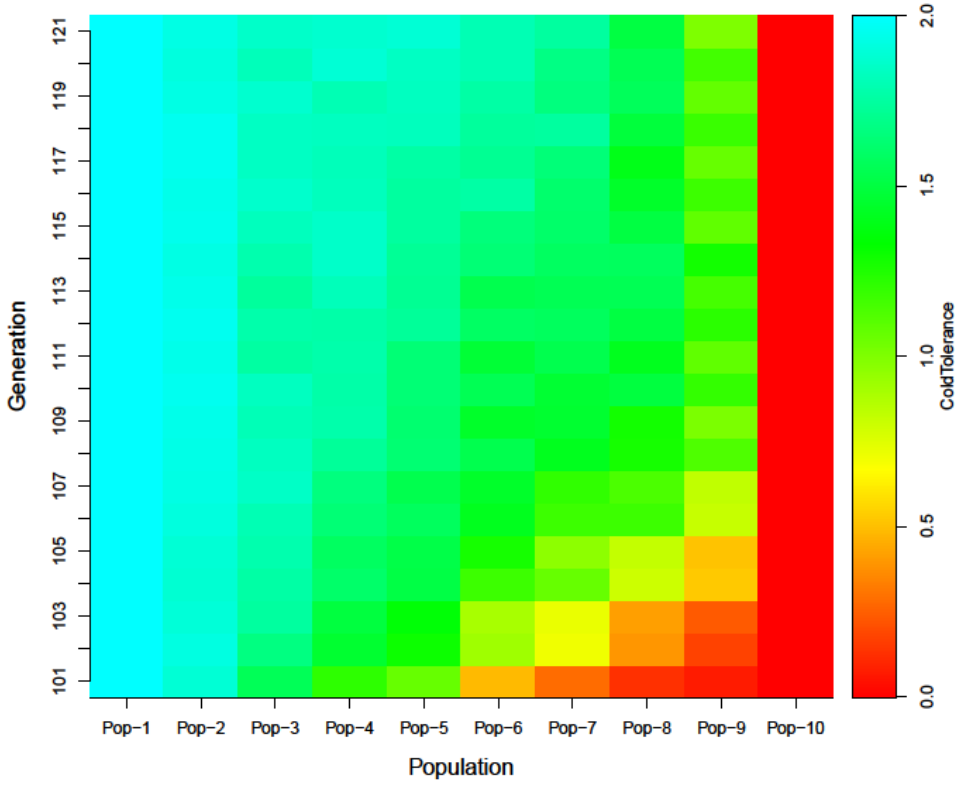


Figure A-3. Heat map of population average cold tolerance trait on a genomic cline. Population was initially equilibrated under only sexual selection for 100 generations. Starting from generation 101, linear temperature cline was applied with seasonal fluctuations. The average temperature also increases by 0.1 degree per year.

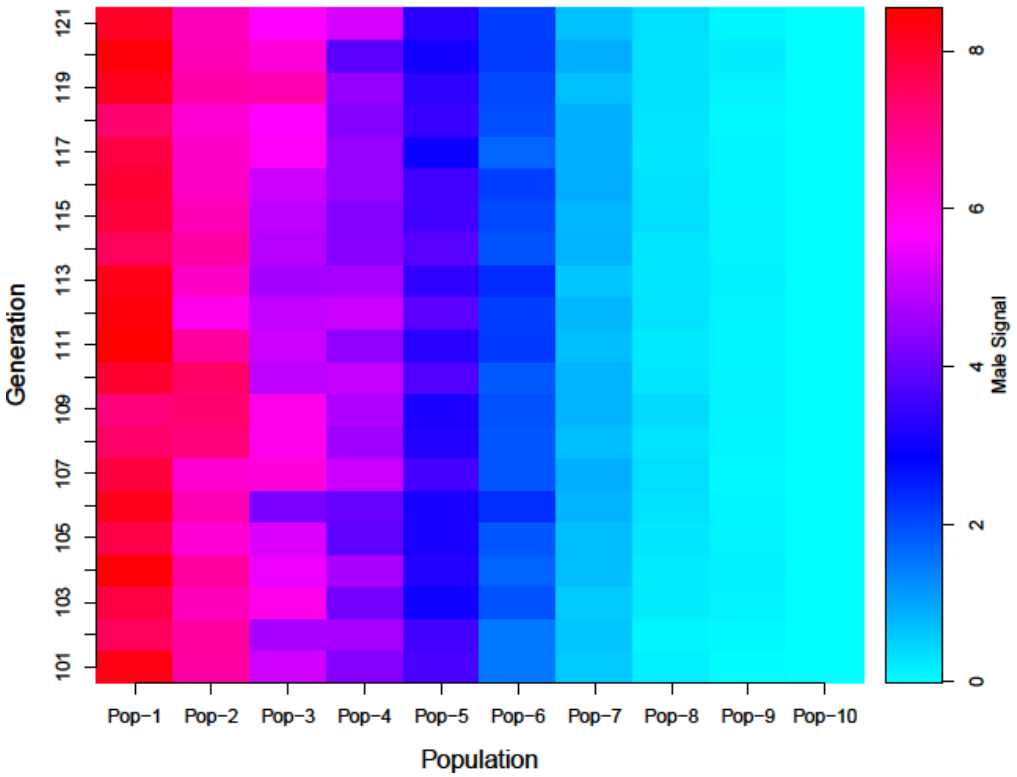


Figure A-4. Heat map of population average male signal trait on a genomic cline. Population was initially equilibrated under only sexual selection for 100 generations. Starting from generation 101, linear temperature cline was applied with seasonal fluctuations. Note that the male trait loci are in weak physical linkage to the temperature adaptation loci.

APPENDIX B

SUPPORTING INFORMATION FOR

EVOLUTION OF REVERSED HEDONIC VALUE IN LEARNED MATING

PREFERENCES

i. Evolution of Ψ assuming peer imprinting and paternal imprinting

We repeated simulation 1 with peer imprinting and paternal imprinting by substituting the average of male trait values of the current generation or paternal trait values in $\bar{s}$, for the H_{he}:H_{ho} proportions 8:0 and 0:8. Results were qualitatively similar with oblique imprinting: assuming peer imprinting, Ψ evolved to negative values (-0.1295 $\pm$0.0421) when H_{he}:H_{ho} =8:0 and positive (0.2057 $\pm$0.0492) when H_{he}:H_{ho} =0:8 (Figure B-4A); assuming paternal imprinting, Ψ evolved to less extreme negative values (-0.0577 $\pm$0.0227) when H_{he}:H_{ho} =8:0 and positive (0.2100 $\pm$0.0227) when H_{he}:H_{ho} =0:8 (Figure B-4B).

ii. Evolution of Ψ with different baseline frequencies of the rare alleles

In simulation 1 we used a 1:10 gene flow to maintain a baseline polymorphism of the 8 male trait loci (baseline frequency of the rare allele: 9.1%). To explore the effect of the baseline frequency of the rare allele on the evolution of Ψ, we ran simulation 1 using 1:20 and 5:6 gene flow proportions (baseline rare allele frequencies 4.7% and 45.5%), with H_{he}:H_{ho} = 0:8 (Figure B-5). When the baseline rare allele frequency was lowered to 4.7%, Ψ evolves to more negative values and the averages across runs no longer overlaps with zero (-0.2050 $\pm$0.0401). When baseline rare allele frequency is increased to 45.5%, Ψ becomes more positive and frequently exceeds zero (-0.0378 $\pm$0.0509). This is an intuitive result because the effect of negative Ψ is to hold the rare allele at a high frequency. If this effect is provided by other means (in our case, gene flow), then a more negative Ψ need not evolve.

iii. Admixing lineages with divergent Ψ's assuming peer imprinting

We repeated simulation 2 with peer imprinting by substituting the average of male trait values of the current generation in $\bar{s}$, without linkage of Ψ and male signal. As with oblique imprinting, there exists a critical threshold where gene flow no longer counteracts the effect of hybridization in breaking up their LD. Compared to oblique imprinting, peer imprinting does not result in F_1 hybrids during early generations, allowing reproductive isolation to be maintained at a higher admixture proportion, up to 32:68 per generation (pop size=2000) ,before extensive admixture occurs. Past this threshold and below 50:50, complete reproductive isolation was briefly maintained for 50-200 generations prior to extensive admixture, where similar fluctuations in male trait values were observed.

iv. Admixing lineages with divergent Ψ's assuming unimodal preference functions

Simulation 2 was repeated assuming oblique imprinting with a modified unimodal preference function, replacing the directional function we used in the main analyses. This unimodal function is more simplistic (Figure B-2B, Figure B-3B):

$$p = \begin{cases} 1 & \Psi = 0 \\ \exp\left[\dfrac{-(\bar{s}-s)^2}{2(1/\Psi)^2}\right] & \Psi > 0 \\ 1-\exp\left[\dfrac{-(\bar{s}-s)^2}{2(-1/\Psi)^2}\right] & \Psi < 0 \end{cases}$$

It is only conditioned on the mean male trait value in the imprinted population, and females use this mean value as the preference optimum when Ψ>0. When Ψ<0, females flip their preference function by subtracting the probability from 1. This approximation is only valid when male trait distribution is close to unimodal. When male traits are bimodally distributed, the mean trait value may fall in an interval where no male traits are observed. If we assume Ψ>0 in this case, then females would prefer a trait that they have never experienced, and vice versa for Ψ<0.

The admixture result is qualitatively similar to results obtained with directional functions (Figure B-9). Elevated F_{IS} is observable up to 30:70 admixture proportion between Ψ<0 and Ψ>0 lineages. The isolation, however, is incomplete on a genome-wide level compared to directional functions. When the proportion exceeds 50:50, negative F_{IS} is also less pronounced than the directional function, but nevertheless F_{IS} never on average significantly exceeds 0. Twelve out of 100 loci show significant negative F_{IS} at an admixture proportion of 90:10 and 8 out of 100 show negative F_{IS} at 80:20.

v. Hybrid zone genetic patterns

Genomic resources for *X. malinche*, *X. birchmanni* and two independent hybrid zones were reported elsewhere (Cui et al. 2013; Schumer et al. 2013; Schumer et al. In review) and further analyzed herein to test predictions from simulation 2.

Detail methods and method validations are described in (Schumer et al. In review). Briefly, reference genomes at 35X coverage of *X. birchmanni* and *X. malinche* were assembled by mapping paired-end reads to the *X. maculatus* reference genome (Schartl et al. 2013). Loci that are polymorphic within species (determined by re-sequencing 60 individuals from each parental species using multiplexed shotgun sequencing (MSG; Andolfatto et al. (2011)) were masked for subsequent analyses. Random samples (all males) from two hybrid zones were collected and the ancestry of 20,609 loci on all 24 chromosomes were imputed by MSG, with an estimated accuracy of 98.1%. Because these loci are reciprocally fixed in parental species, the ancestry

states were treated as their allelic states. These two hybrid zones were known to differ in their ancestry bias and their population structures. Low resolution markers previously suggested that Calnali-Mid (CALM) , a *X. birchmanni*-biased population have a higher population structure than Tlatemaco (TLMC), which is a *X. malinche*-biased population (Culumber et al. 2011).

Hybrid index was calculated as the percentage of loci of *X. malinche* ancestry. As reported previously, TLMC shows a unimodal distribution of hybrid indices while CALM shows a bimodal distribution (Figure B-12).

F_{IS} was calculated for each locus respectively for TLMC and for the *X. birchmanni*-like subpopulation of CALM, excluding missing data. Because CALM as a whole have bimodal distribution of hybrid index, the lumped F_{IS} is significantly >0 as expected by simulation (not shown). Within each sub-population in CALM, however, random mating is expected according to our assumption of directional mate preference function (see Figure B-2A), thus we use it to obtain a null distribution of F_{IS} values. We tested the two distributions' deviation from 0 with a Student's t-test. As predicted from simulation 2, the F_{IS} of the *X. birchmanni*-like subpopulation from CALM does not significantly deviate from 0, while TLMC has a mean F_{IS} significantly more negative than CALM (Figure 10S).

Kinship coefficient between individuals in the sample correlates with level of assortative mating in the population, because mating with similar individuals increases the probability of mating with kin. Simulation 2 predicts that TLMC should have on average lower relatedness than CALM due to disassortative mating. We thinned the markers so that they are spaced by at least 100kb, we then inferred kinship coefficient using REAP (Thornton et al. 2012), a software designed for admixed populations. The samples from CALM (both subpopulations included, 0.09860) are more related than the samples (t=4.385, d.f.=6411, p=0.00001) from TLMC (0.09305). Further thinning did not alter our conclusions.

One possibility causing the distinct population structure is non-equilibrium state in CALM. To rule out the hypothesis that structure in CALM is due to recent hybridization, we estimated numbers of generations since initial hybridization through ancestry block size. The number of generations after initial admixture in these two hybrid zones were inferred by the median ancestry block size of the minor parent using:

$$T_{admix} = \frac{1}{LP}$$

T_{admix} - number of generations past

L - block size of the minor parent in morgans, assuming a *Xiphophorus*–specific recombination rate at 1 cM/378 kb (Walter et al. 2004a).

P - proportion ancestry of the major parent.

We bootstrapped the block sizes to obtain a confidence interval of the estimate. TLMC was inferred to have admixed for 378 (95% CI 362-394) generations, CALM for 276 (95% CI 260-296) generations. These estimates suggest that the genetic patterns in these hybrid zones are unlikely a result of non-equilibrium states due to recent admixture dynamics.

Tables

Table B-1. General linear model of factors influencing olfactory and visual preference indices (net preference for *X. birchmanni* male stimulus plus 300) in *X. birchmanni* and *X. malinche* females (Focal female species) reared with exposure to *X. birchmanni* or *X. malinche* adults (Exposure models). * $P < 0.05$. Dispersion parameters for quasi-poisson distribution: olfactory =135.72, visual=61.12.

	Factor	*d.f.*	*t*	*P*
Olfactory	Intercept	67	41.5	<2e-16
	Focal female species	67	-3.709	0.0004*
	Exposure model	67	-2.875	0.0055*
	Focal female species * Exposure model	67	3.980	0.0002*
Visual	Intercept	57	49.2	<2e-16
	Focal female species	57	0.581	0.564
	Exposure model	57	-2.523	0.0146*
	Focal female species * Exposure model	57	1.613	0.113

Figures

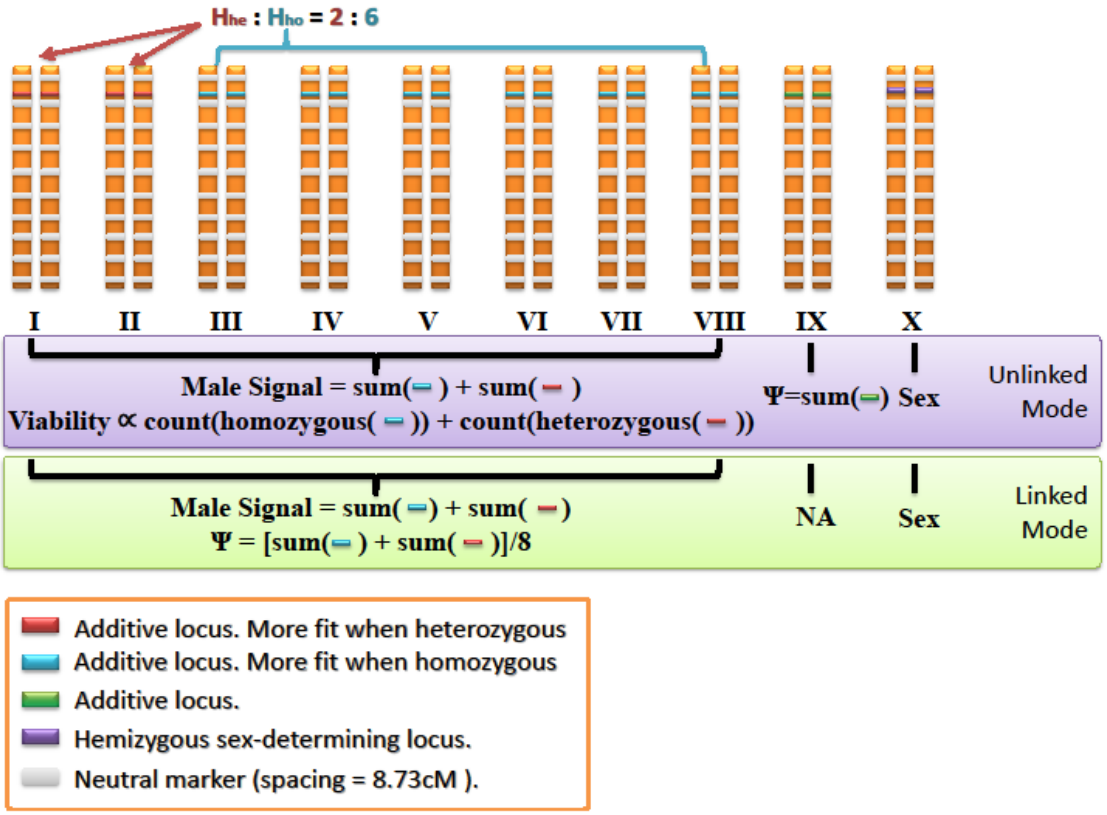
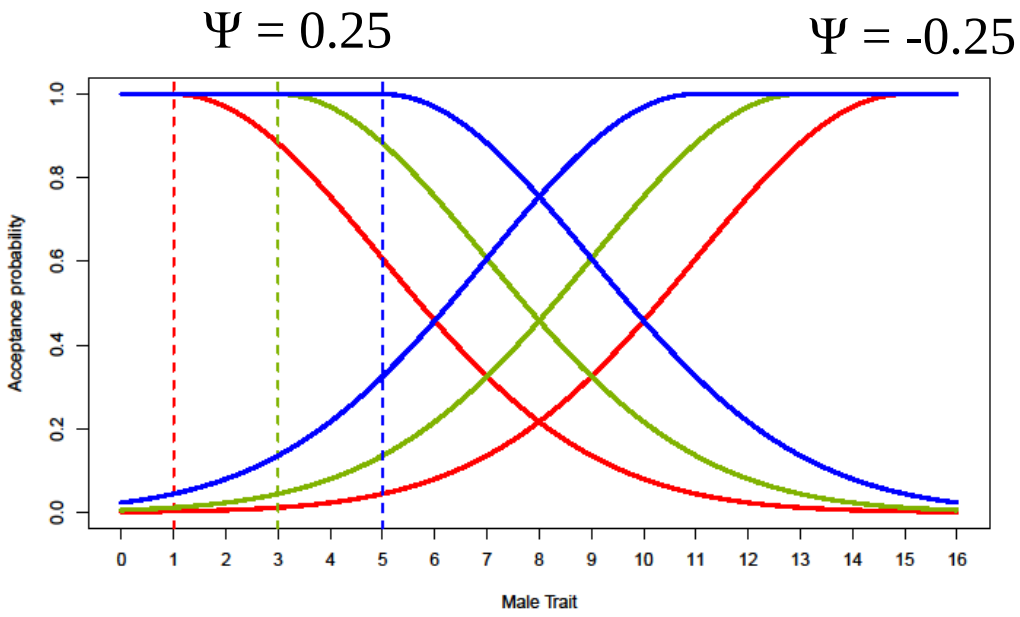


Figure B-1. Diploid genome used in simulations.
All 10 chromosomes are 3.3 × 107 bp long (87.30cM assuming Xiphophorus recombination rate of 378kb/cM). The unlinked mode was used in simulation 1 (evolution of Ψ), while the linked mode was also explored in simulations 2 (admixture) in addition to the unlinked mode. Note that "Viability" is only relevant in simulation 1, and neutral markers are only relevant in simulation 2. The proportions of red vs. light blue loci (Hhe: Hho) were varied in simulation 1.

A



B

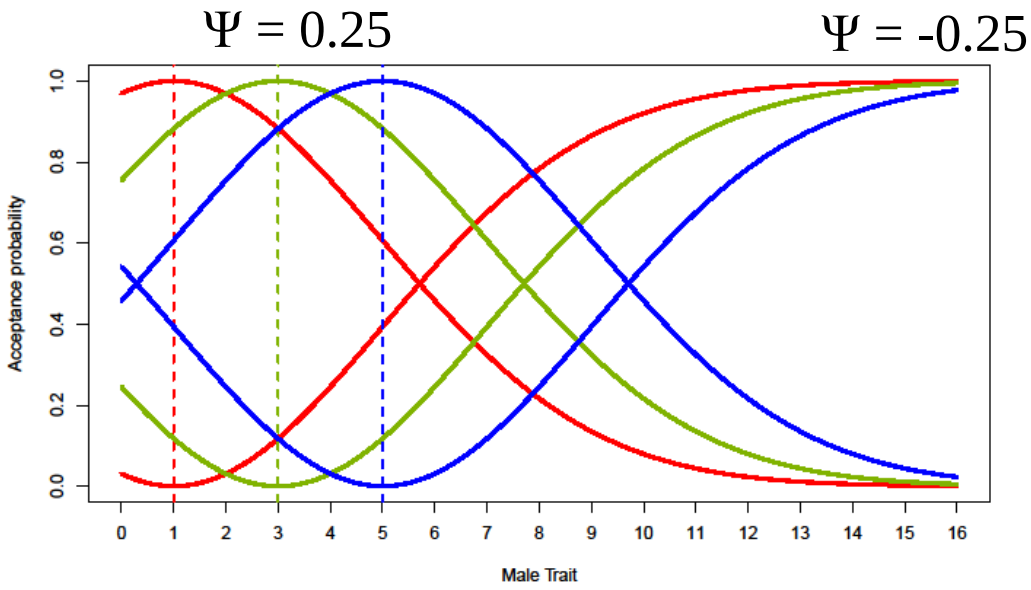
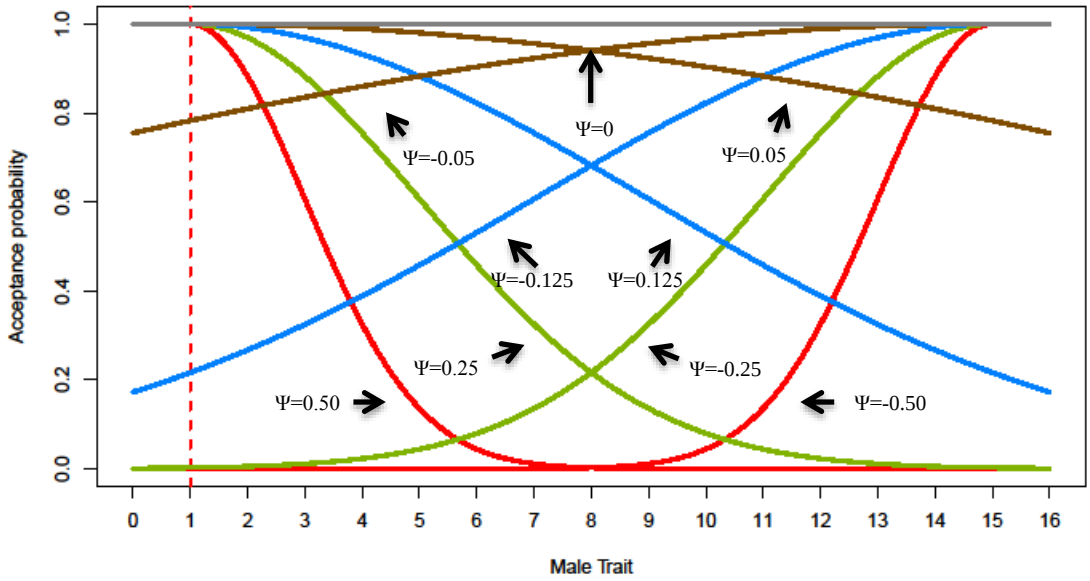


Figure B-2 Preference functions in response to different mean male traits.
A) Directional preference functions used in main analyses with varying mean population male signals (dotted lines) and Ψ=0.25 (curves on left) or Ψ=-0.25 (curves right). B) Unimodal preference functions used in Appendix B iv with varying means of male signals (colors and Ψ's same as A). Note that under this preference function, when male trait distribution is bimodal, Ψ>0 results in an unrealistic scenario where females prefer male traits that are infrequent or even unobserved in the population, and vice versa for Ψ<0.

A



B

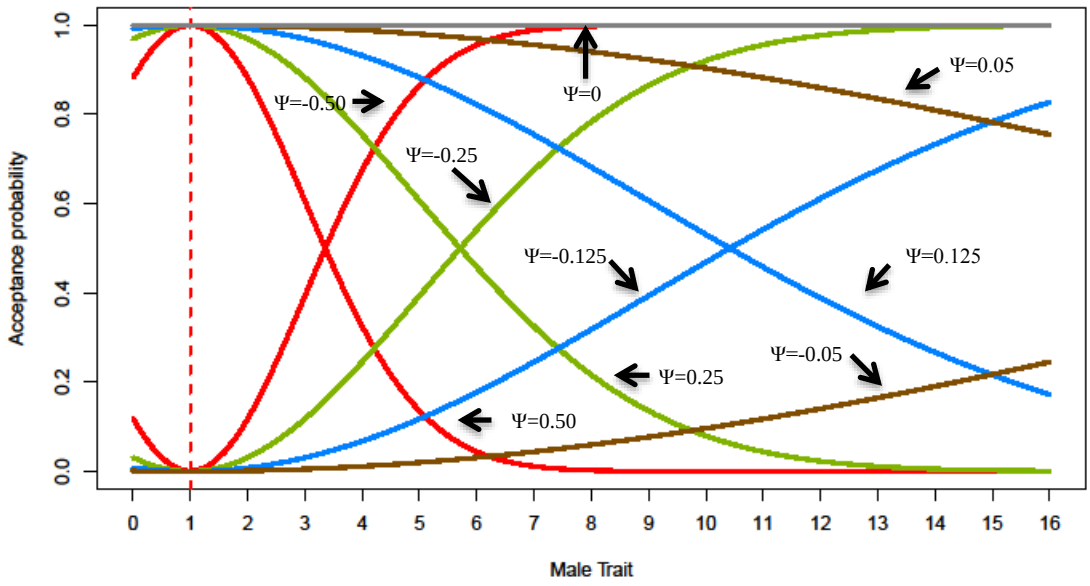


Figure B-3. Preference functions with different values of Ψ.
A) Directional preference functions with varying Ψ's. Mean male signal = 1 (red dotted line). Curves on left have Ψ>0, curves on right have Ψ < 0. **B)** Unimodal preference functions with varying Ψ's. Mean male signal = 1 (red dotted line).

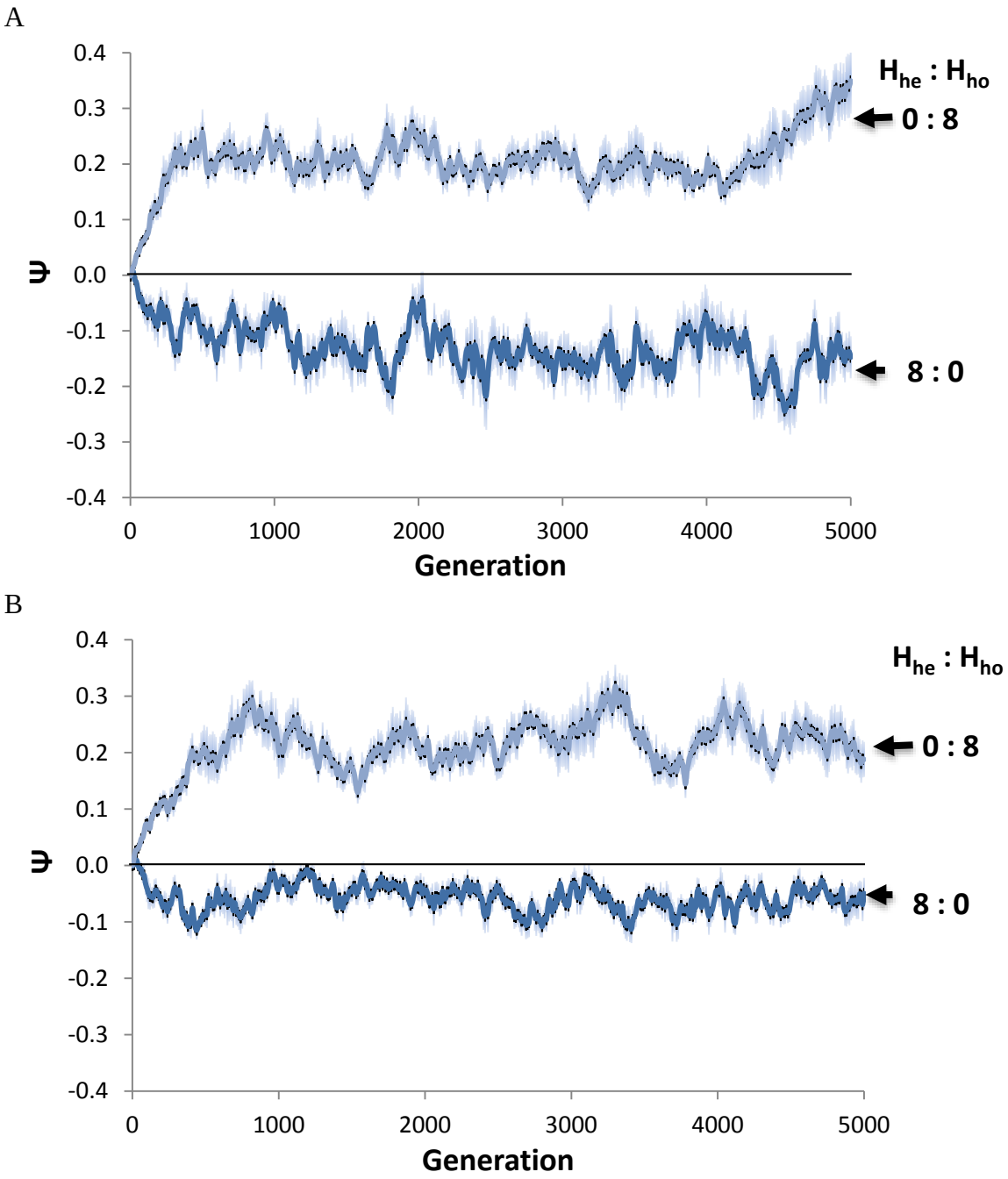


Figure B-4. Evolution of Ψ under different modes of imprinting.
Mean and S.E.M. of population average Ψ across 10 independent runs plotted against generations for each proportion of loci. A) Peer imprinting: females imprint on peer males in their cohort; B) Paternal imprinting: females imprint on fathers.

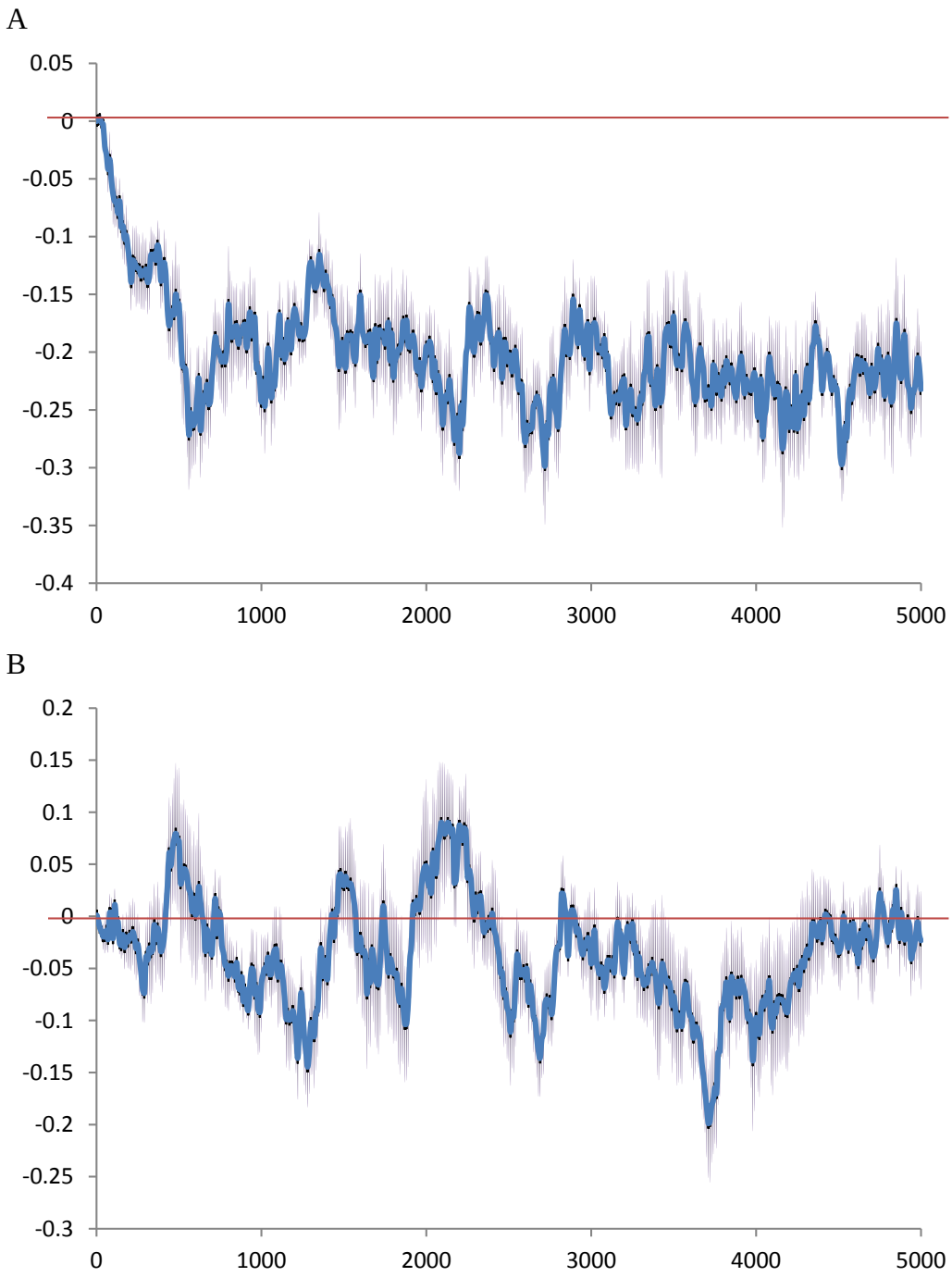


Figure B-5. Evolution of Ψ under different frequencies of the rare male trait alleles. H_{he} : H_{ho} = 8:0, A) baseline rare allele frequency provided by gene flow at 4.7% and B) baseline rare allele frequency provided by gene flow at 45.5%. Note that the negative Ψ will eventually drive the rare allele frequency close to 50%.

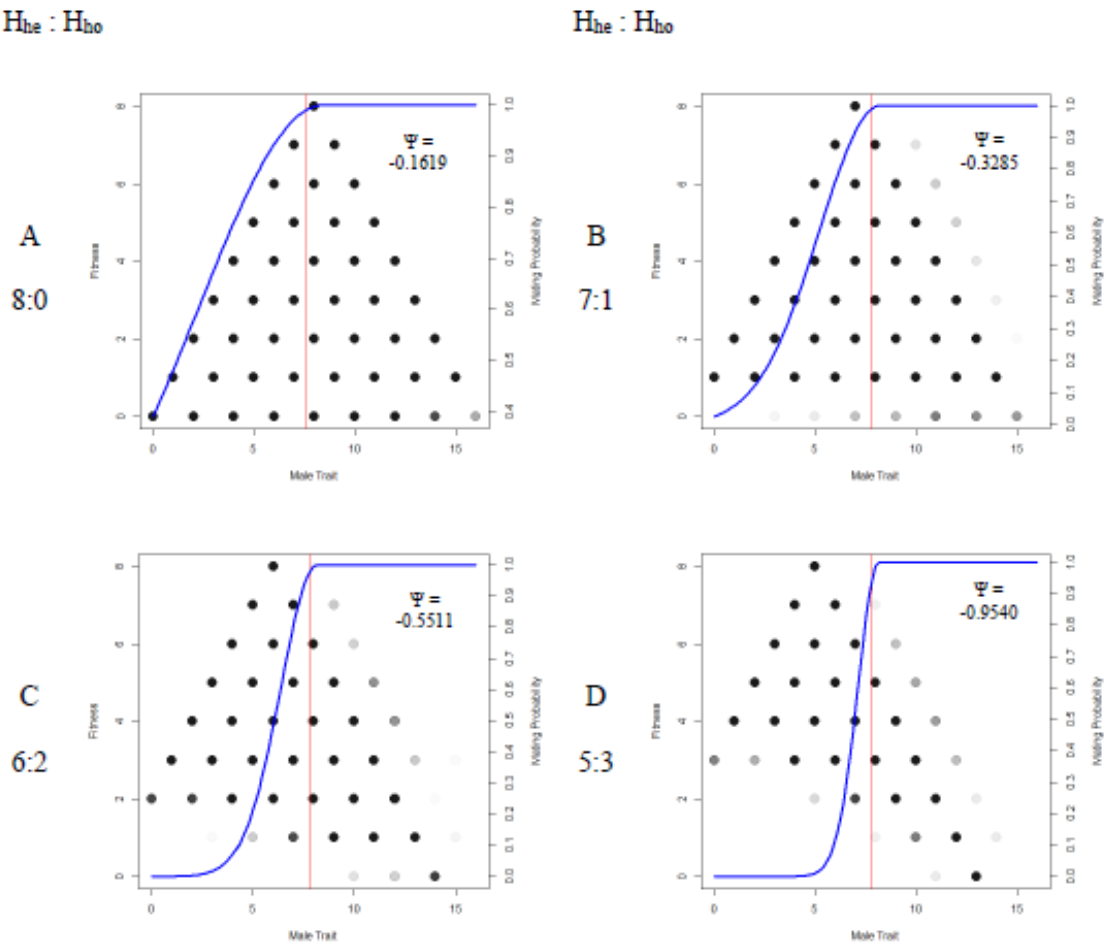


Figure B-6. Scatter plots of male trait against fitness showing mean preference function (blue line) under different H_{he} : H_{ho} proportions in simulation 1 (Figure 1).
Data are pooled from 10 runs in simulation 1 for generations 4000-5000. Darkness of dot represents number of individuals in the sample. Red vertical line is the mean male trait in the population for those pooled 1000 generations; blue line is the average mate preference function using the Ψ estimates for the corresponding simulations in generations 4000-5000. The slope of preference function increases when the proportions decreased from 8:0 to 5:3. Switching of the optimal mating strategy occurs between 5:3 and 4:4, where Ψ changed from negative to positive.

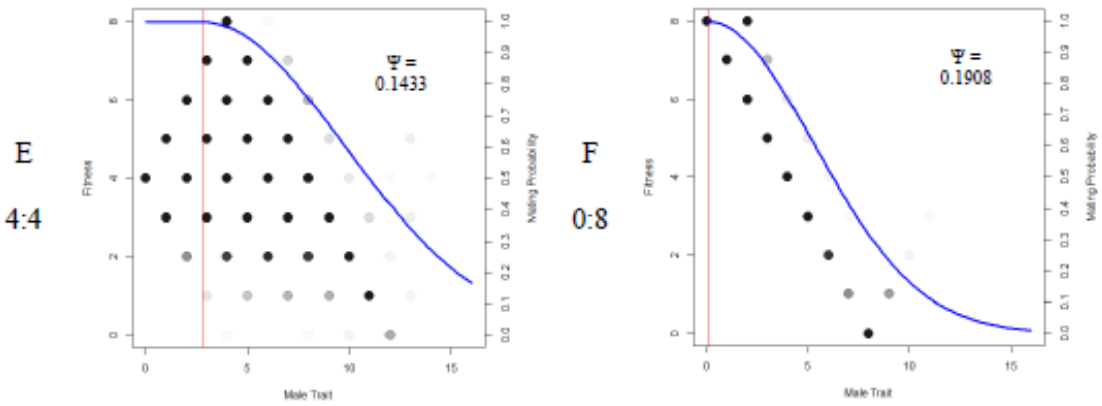


Figure B-6. Continued.

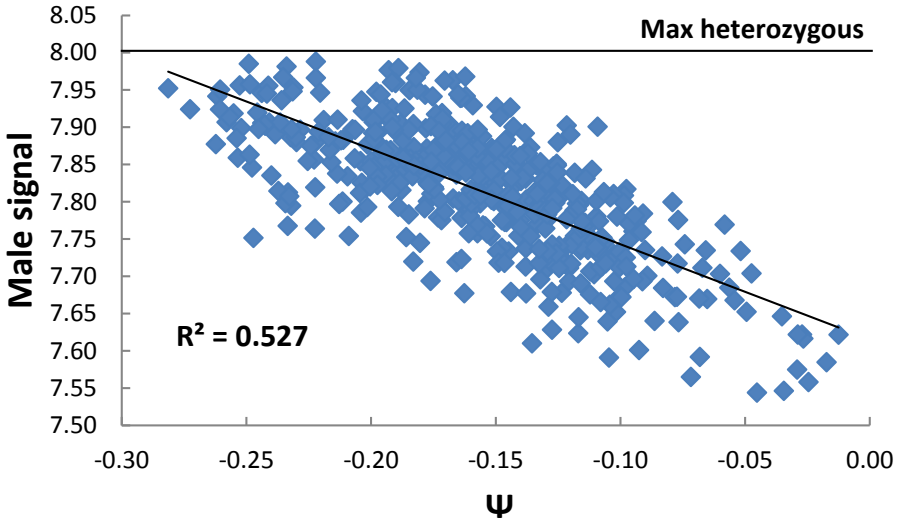


Figure B-7. Negative correlation between Ψ and male signal (*s*).
H_{he} : H_{ho} = 8:0. Male signaling trait takes values in the range [0,16], *s* = 8 in this case suggest maximal heterozygosity (more likely to be fit during natural selection, also see Figure B-6A).

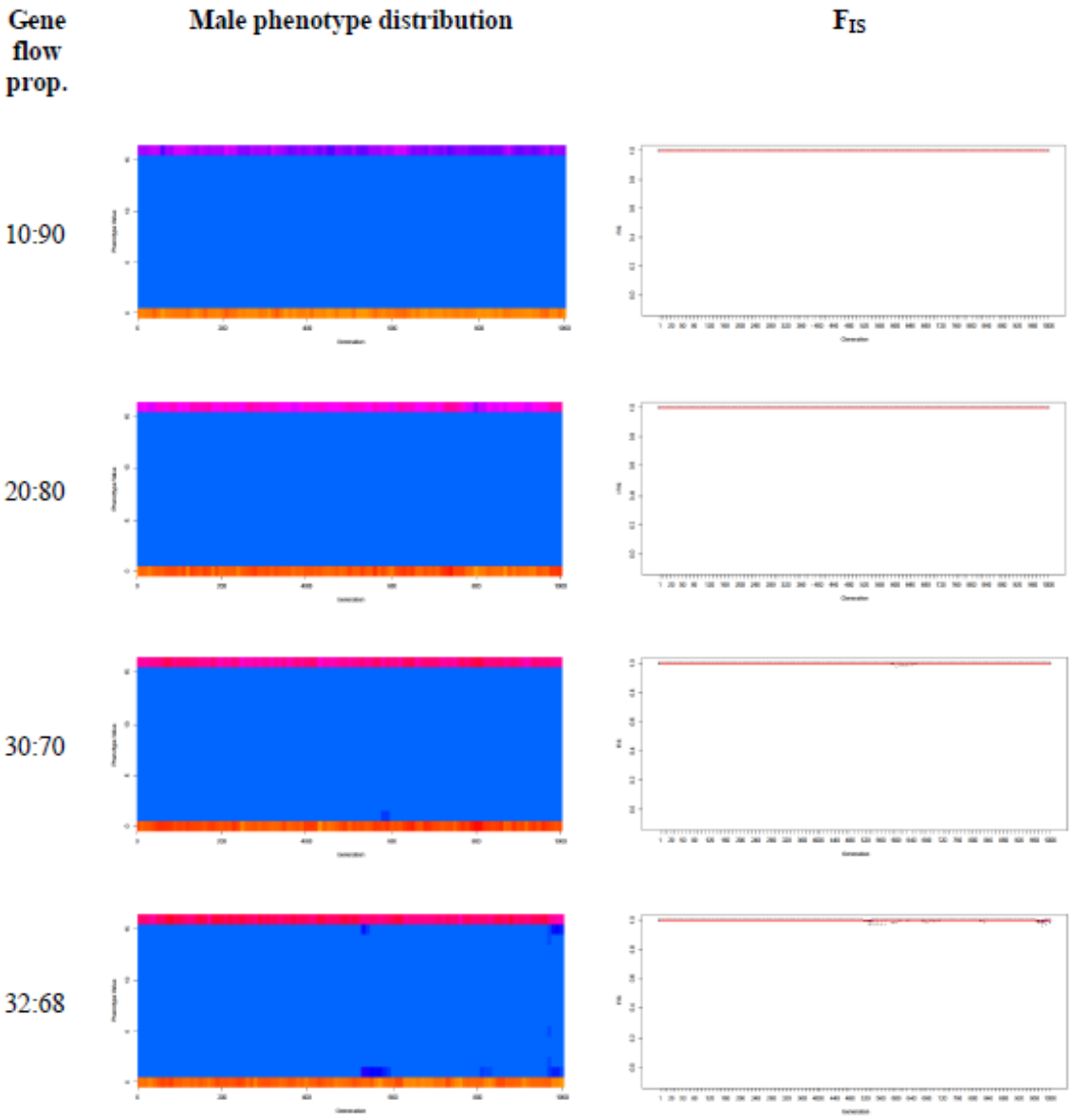


Figure B-8. Reproductive isolation by mixing lineages with divergent Ψ's can be achieved assuming peer imprinting. The ratio of migrants from the lineage with $\Psi < 0$ relative to the lineage with $\Psi > 0$ is less than 32:68 individuals per generation, in a hybrid population with carrying capacity of 2000.

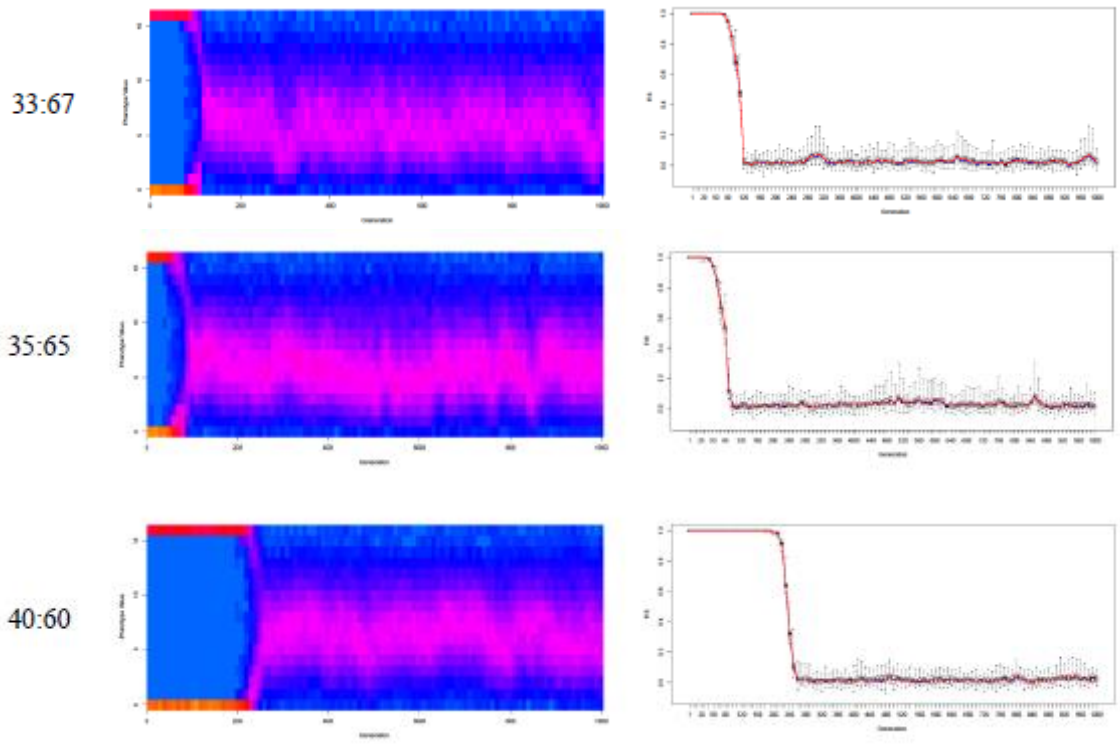


Figure B-8 Continued.

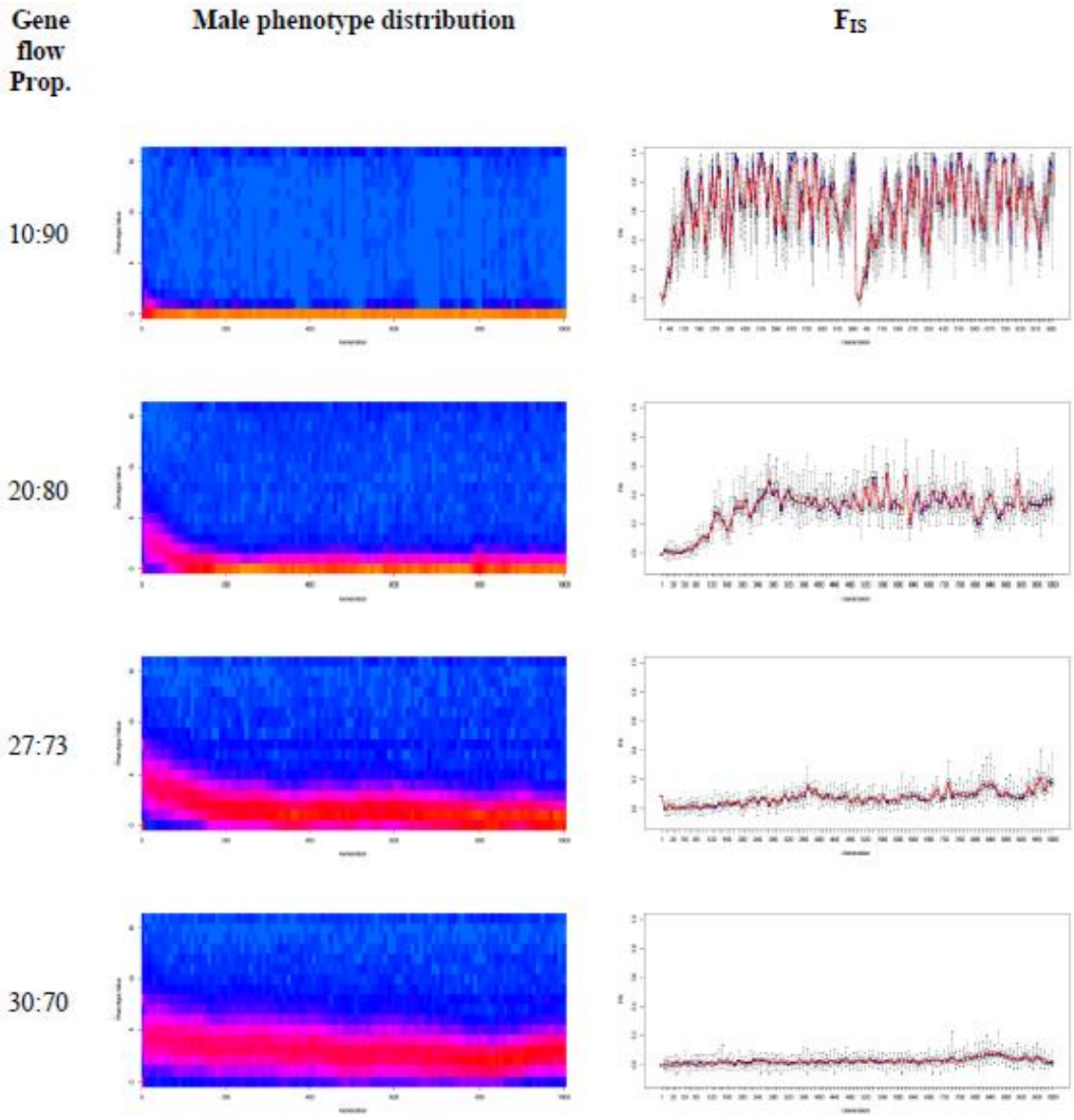


Figure B-9. Elevated reproductive isolation by mixing lineages with divergent Ψ's can be achieved assuming oblique imprinting and unimodal preference functions. The gene flow from the lineage with $\Psi < 0$: $\Psi > 0$ is less than 27:73 individuals per generation, in a hybrid population with carrying capacity of 2000.

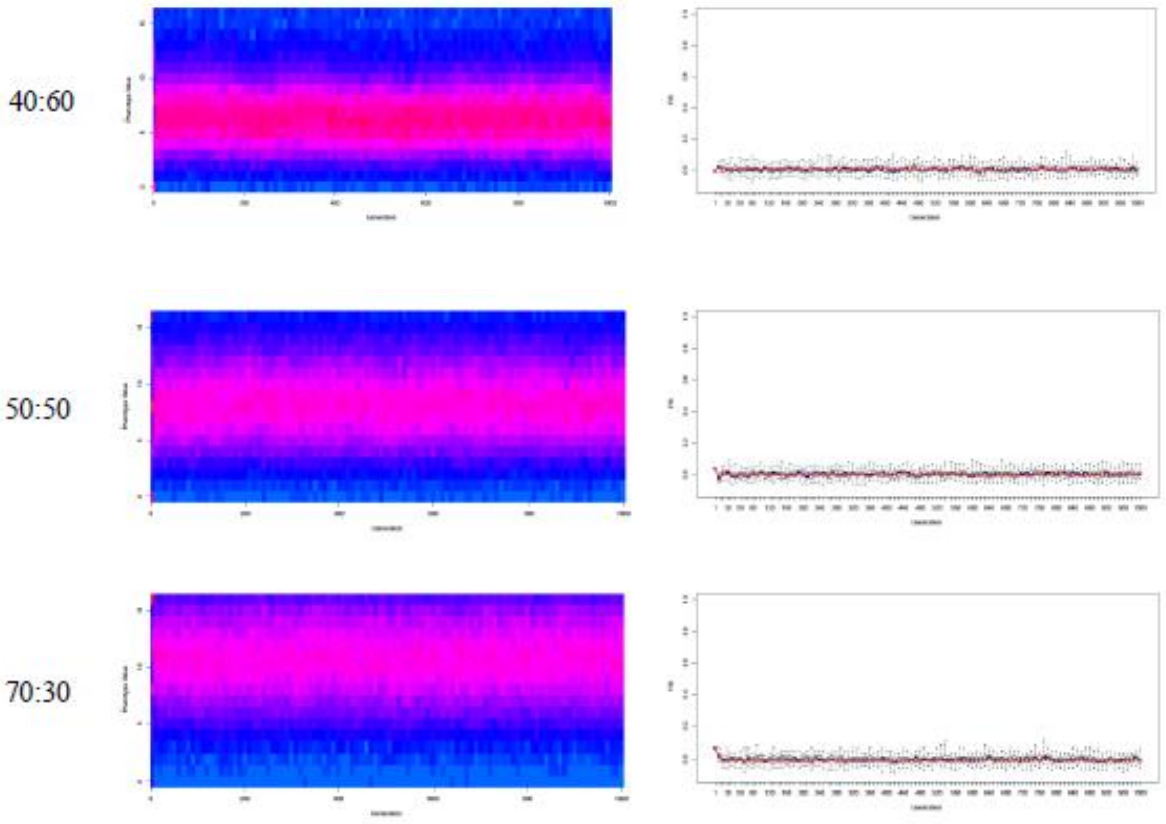


Figure B-9 Continued.

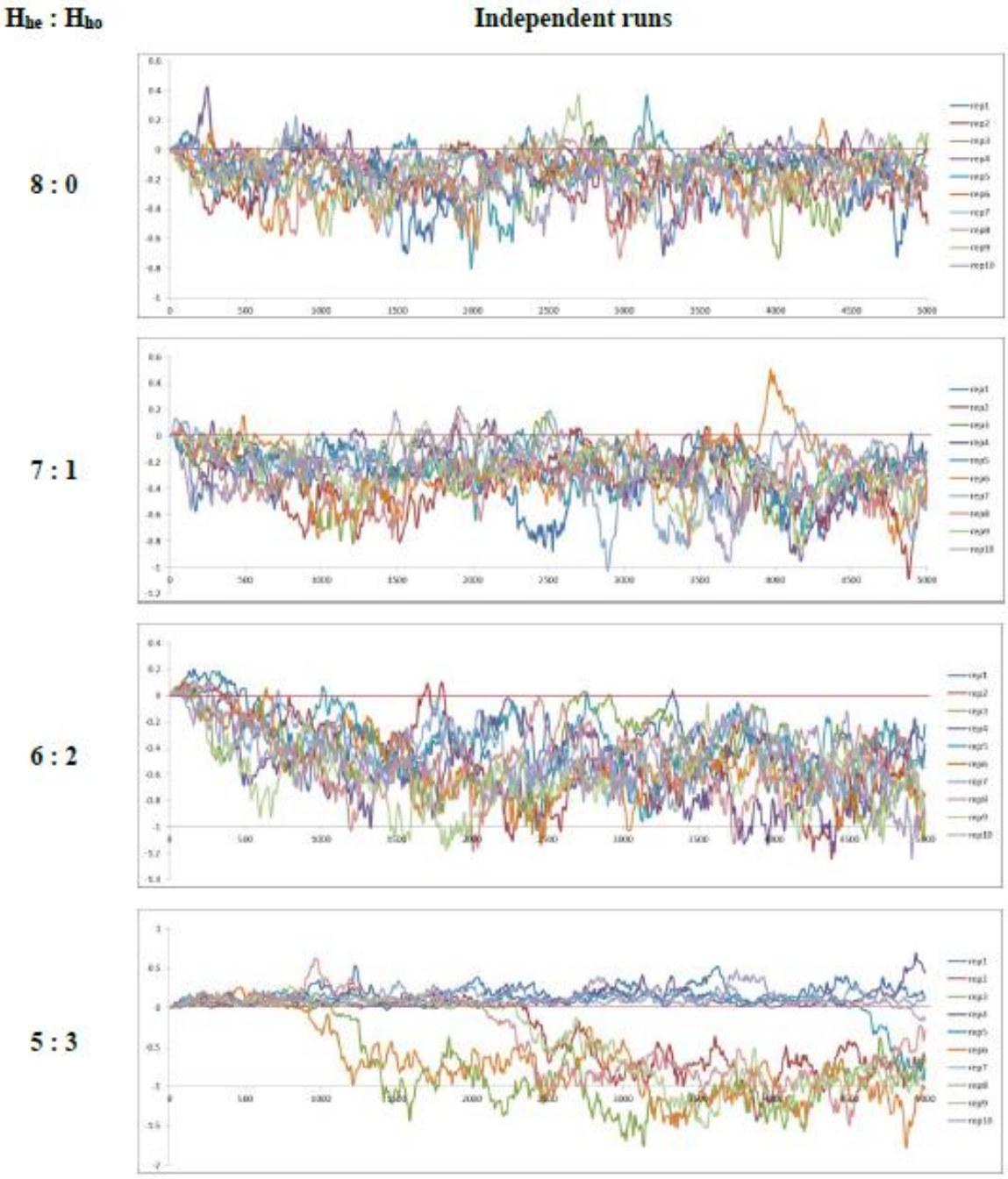
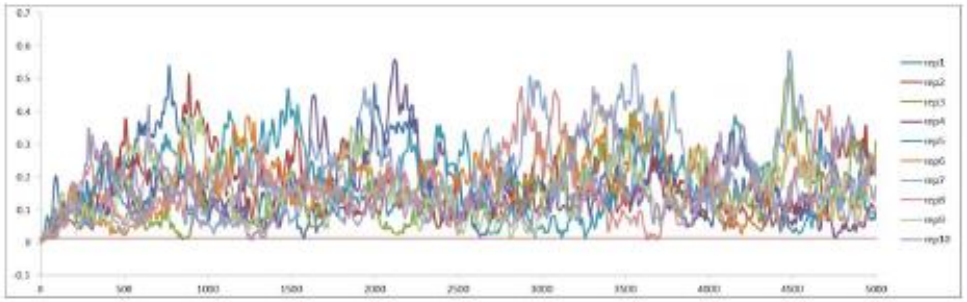


Figure B-10. Plots of independent runs of the evolution of Ψ in simulation 1 (Figure 1).

4 : 4

2 : 6



0 : 8

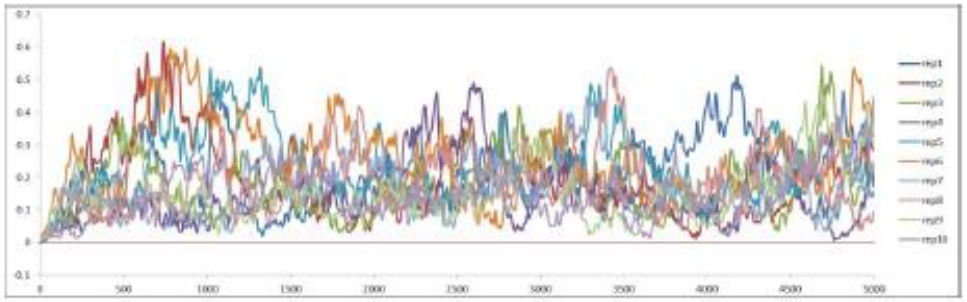


Figure B-10 Continued.

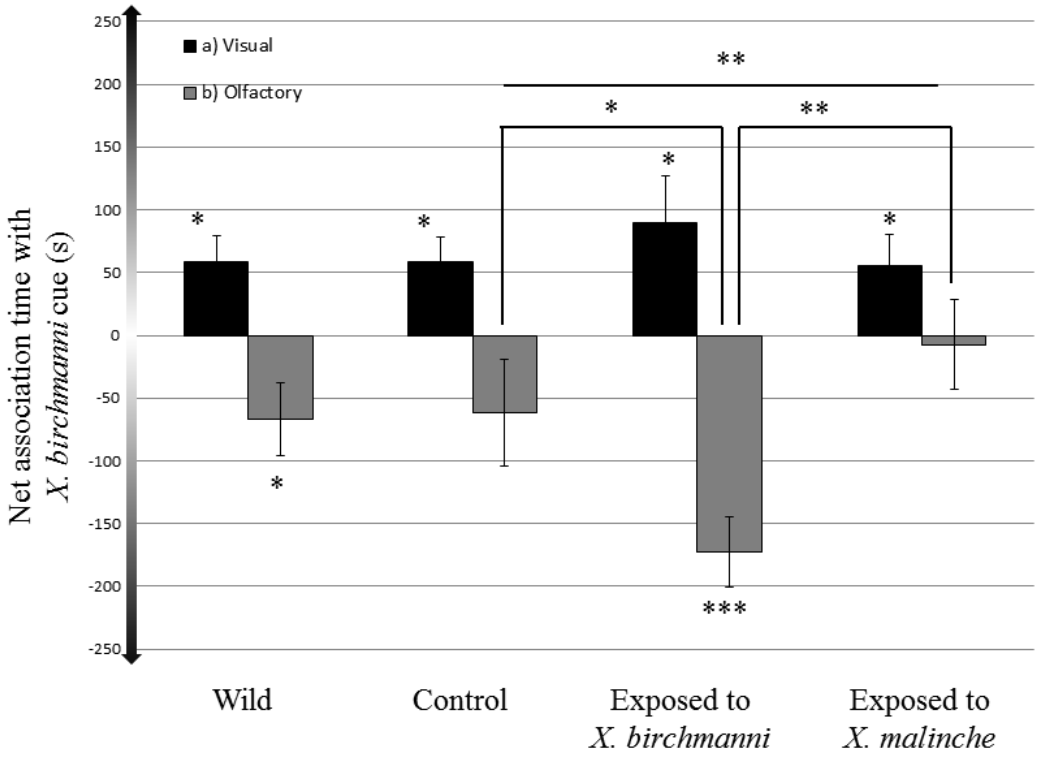
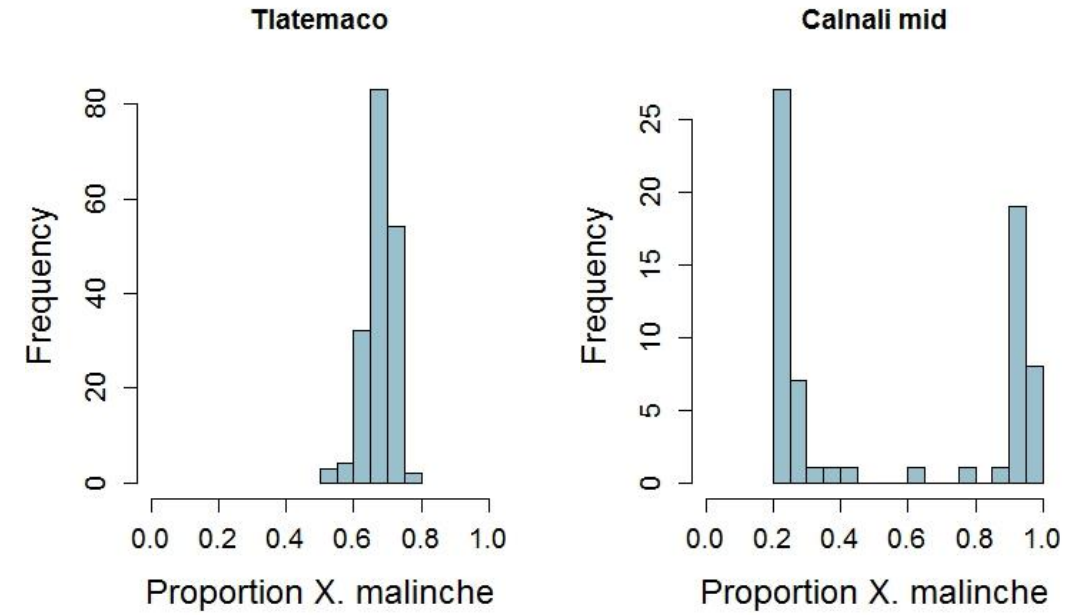
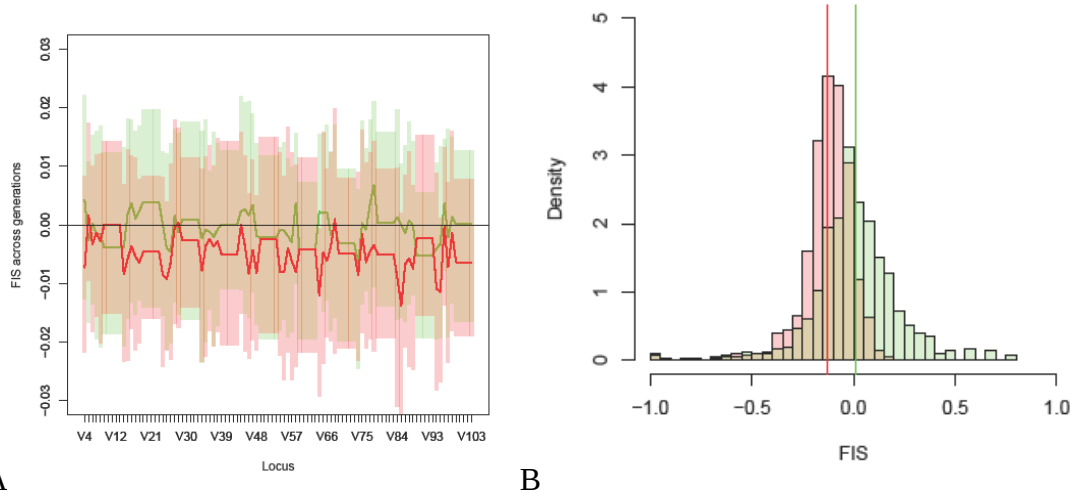


Figure B-11. Preferences of female *Xiphophorus malinche* varying in early social experience for different stimulus pairs.
a) Visual cues: computer animations representing mean male *X. birchmanni* and *X. malinche* phenotypes; b) olfactory cues: water containing male pheromones of each species. Bar height represents net association time with the heterospecific *X. birchmanni* ± SEM. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.



A B

Figure B-12. Distribution of hybrid indices sampled from two natural hybrid populations between *X. birchmanni* and *X. malinche*.
A) Tlatemaco consists of mostly *X. malinche*-biased individuals showing even admixture, while **B)** Calnali-Mid contains mainly *X. birchmanni*-biased individuals showing highly bimodal distribution of hybrid indices. In both populations hybridization is inferred to have occurred for > 100 generations from high-throughput genotyping data (Schumer et al. In review).



A B

Figure B-13. Populations with ancestry biased towards the parental species and Ψ < 0 show an excess of heterozygosity (F_{IS} < 0).
A) Simulated F_{IS} summarized for each neutral locus across generations 100-1000 with admixture proportion Ψ < 0 : Ψ > 0 = 80 : 20, assuming random mating (green) and mate preference affected by oblique imprinting (red). Boxes are the lower and upper quartiles. Genome-wide average F_{IS} significantly more negative than random mating across generations (paired t-test, t=9.246, p<<0.001, d.f.=99), especially when markers are physically closely linked to the signaling or Ψ loci. B) Observed excess of heterozygosity in Tlatemaco (*X. malinche* ancestry ~72%, F_{IS} in red bars, mean as red line) significantly more evident (t-test, t = -59.41, df = 20608, p-value << 0.001) compared to the balanced distribution of F_{IS} around 0 in the *X. birchmanni*-like subpopulation of Calnali-mid (green bars, mean as green line). Note that F_{IS} will be significantly larger than 0 if all individuals are included in the analysis of Calnali-mid, due to the apparent population structure (
Figure B-12B). Within subpopulations, however, random mating is expected based on our model assumption of directional preference functions, because the preference functions evaluate to uniform in these trait ranges (see Figure B-2A).

APPENDIX C

SUPPORTING INFORMATION FOR

EARLY LEARNING TRIGGERS EXPRESSION CHANGES IN NEURAL

PLASTICITY GENES AND ODORANT RECEPTORS

IN *XIPHOPHORUS MALINCHE*

i. Comparisons of different alignment methods

Methods of aligning reads to the genome and stringency of alignment parameters can potentially have major effects on genes detected as differentially expressed. To investigate the possible impact of different aligners and different alignment parameters on the genes detected in our analyses, we repeated alignment and all downstream analysis steps with STAMPY (Lunter and Goodson 2011) with an expected divergence of 0.02. STAMPY does not use an explicit intron-exon model in mapping and also maps more divergent reads than other aligners. In addition, we repeated TopHat analysis using a more stringent mismatch parameter (2 allowed mismatches instead of 3). The results of these three analyses were strikingly similar (Figure C-2), suggesting that our major results are not very sensitive to the aligner or alignment parameters used.

Figures

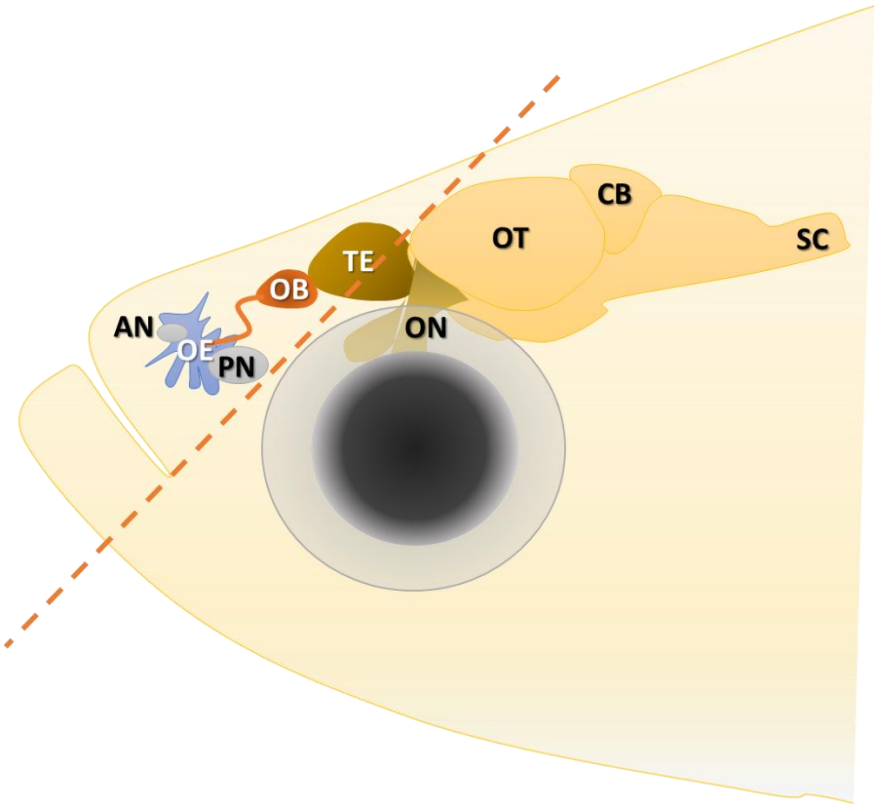


Figure C-1. Sampling method of pooled sensory and brain tissue. Brain structure drawing modified from (Kermen et al. 2013). A single 45 degree cut was made in front of the orbits, taking tissues including lips, olfactory epithelium, olfactory bulb and dorsal part of telencephalon. Abbreviations: AN – anterior nostril, PN – posterior nostril, OE – olfactory epithelium, OB - olfactory bulb, TE – telencephalon, ON – optic nerve, OT – optic tectum, CB- cerebellum, SC – spinal cord.

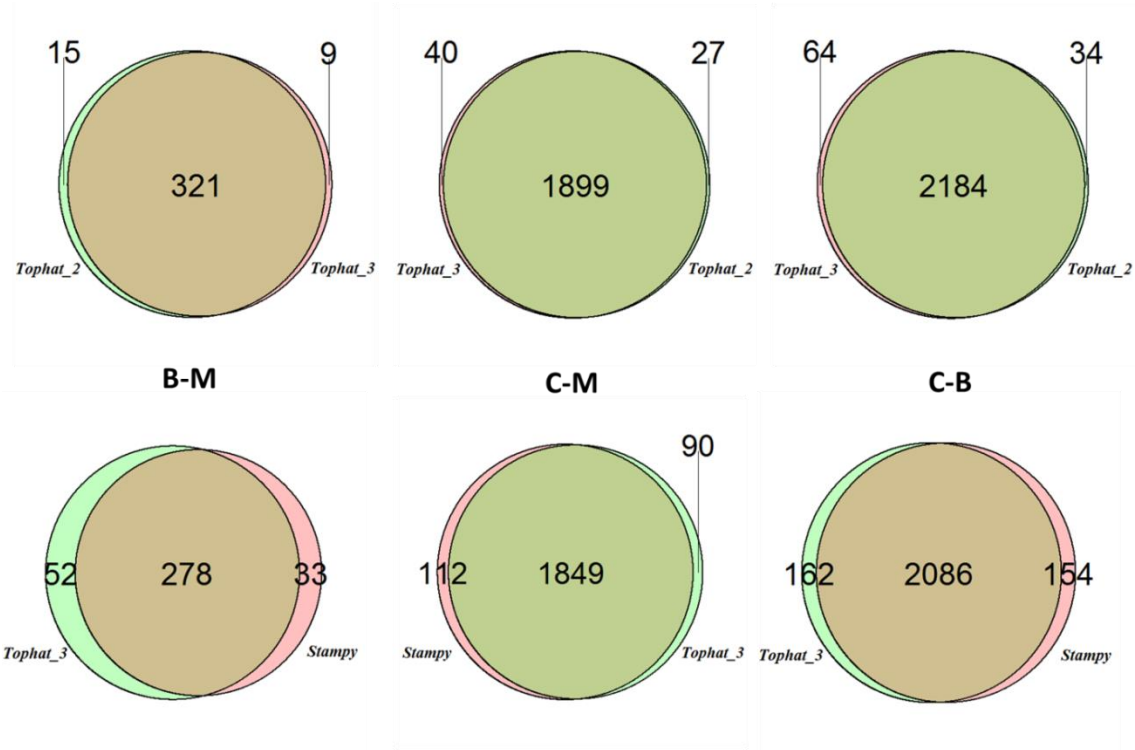


Figure C-2. Repeatability of differential expression analysis with different read mappers. Tophat_2: allowing for two mismatches; Tophat_3: allowing for three mismatches; Stampy: expected divergence 0.02. B-M: Differentially expressed (DE) genes between *birchmanni-malinche* exposure treatments; C-M: between control group and malinche exposed group; C-B: between control and *birchmanni*-exposed group. Comparisons are performed using an FDR cutoff of 0.05.

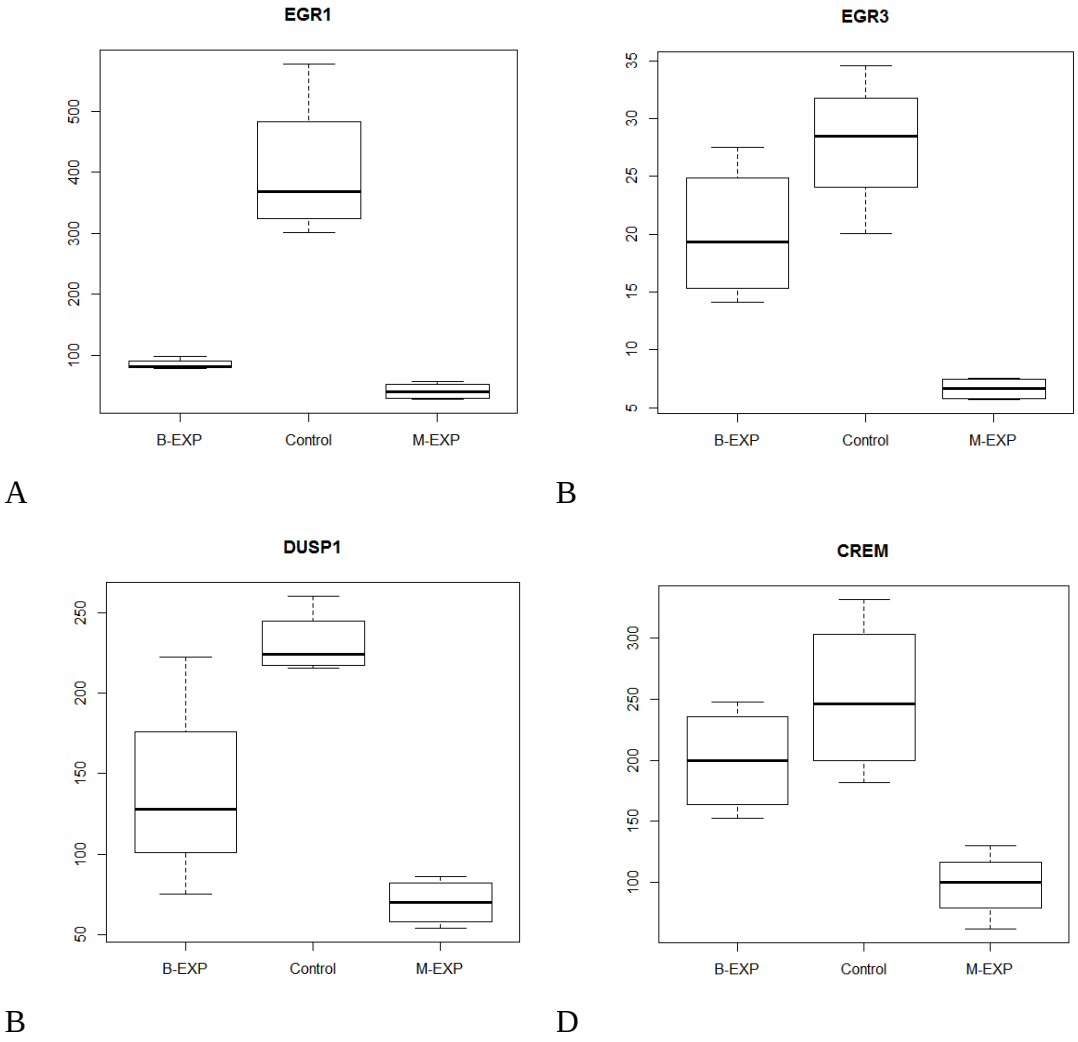


Figure C-3. Expression levels of genes regulated by CREB. Y-axis: reads per million. These genes have the expression pattern Control > *birchmanni*-exposed > *malinche*-exposed.

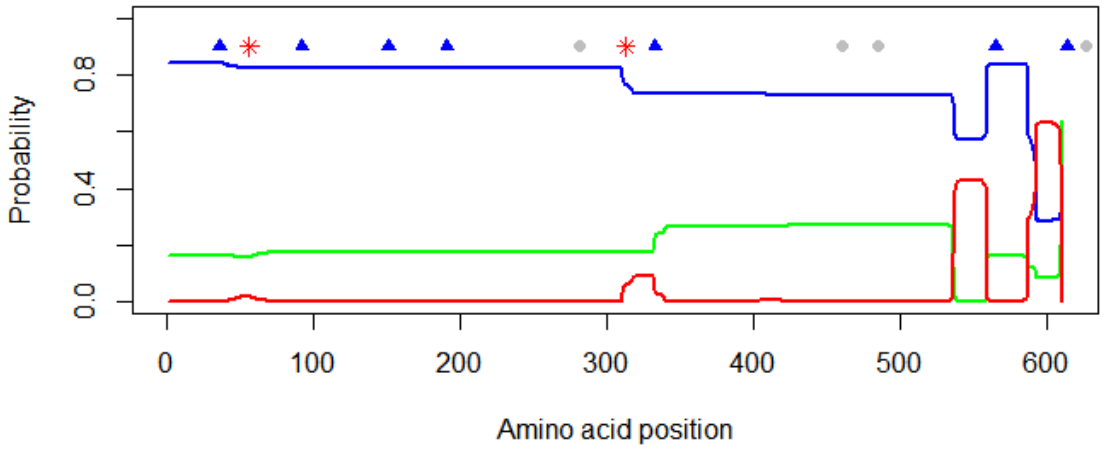


Figure C-4. Transmembrane regions of V2R 6 from *X. malinche* inferred by TMHMM-2.0. Blue line- extracellular domain, red line – transmembrane domain, green line – intracellular domain. Amino acid differences between *X. birchmanni* and *X. malinche* are shown: red asterisk – mutation unique to this species pair; blue triangles – mutation also found in one or more of *X. clemenciae*, *X. hellerii* and *X. maculatus*; grey rounded dots – ambiguous due to missing data in another species. Putative ligand binding region in V2R locates on the extracellular N-terminus (closer to position 0).

Tables

Table C-1. Alignment statistics for the samples used in this study with tophat, allowing for 3 mismatches.

Sample	**Number of reads**	**Percent Mapped**	**Percent multiple alignments**
Control female 1	21828760	83.1	1.3
Control female 2	26403288	81.0	2.1
Control female 3	25922662	82.0	1.9
Control female 4	24819697	81.8	2.2
malinche treated 1	23865638	81.9	1.9
malinche treated 2	25122442	82.0	1.9
malinche treated 3	22644001	81.8	1.9
malinche treated 4	25581923	82.1	1.8
birchmanni treated 1	24367400	82.1	1.8
birchmanni treated 2	25207335	80.8	1.9
birchmanni treated 3	23926948	74.7	2.4
birchmanni treated 4	21840338	81.5	1.7

Table C-2. Differentially expressed genes shared between at least two comparisons. FDR p values for BM: *birchmanni*-exposed vs. *malinche*-exposed; CM: control vs. *malinche*-exposed; CB: control vs. *birchmanni*-exposed. Only genes annotated with gene symbols are shown, for full table, refer to file diff_exp_supptable.xlsx.

GeneName	GeneID	BM	CM	CB
FOSB	ENSXMAG00000018663	0.000	0.000	0.000
DUSP5	ENSXMAG00000007046	0.000	0.000	0.000
SERPINE1	ENSXMAG00000011952	0.000	0.000	0.037
HSD3B7	ENSXMAG00000008008	0.000	0.000	0.000
PTGS2	ENSXMAG00000002418	0.000	0.000	0.000
LAMC2	ENSXMAG00000011284	0.000	0.000	0.000
ZBTB7C	ENSXMAG00000004232	0.000	0.000	0.001
EGR1	ENSXMAG00000008996	0.001	0.000	0.000
HGFAC	ENSXMAG00000015841	0.001	0.046	0.000
KLF4	ENSXMAG00000014131	0.002	0.000	0.000
GLDC	ENSXMAG00000003032	0.003	0.000	0.000
EEF2K	ENSXMAG00000015752	0.004	0.000	0.000
HMGA1	ENSXMAG00000016571	0.005	0.000	0.000
DUSP1	ENSXMAG00000009651	0.006	0.000	0.021
LAMB3	ENSXMAG00000018790	0.006	0.000	0.000
ATF3	ENSXMAG00000004443	0.006	0.000	0.000
LAMA3	ENSXMAG00000016504	0.006	0.001	0.000
DUSP8	ENSXMAG00000019232	0.012	0.000	0.000
HAS1	ENSXMAG00000010604	0.017	0.000	0.001
FOS	ENSXMAG00000004017	0.018	0.000	0.000
BNC1	ENSXMAG00000000835	0.019	0.000	0.000
KLF2	ENSXMAG00000000640	0.023	0.000	0.000
FOSL1	ENSXMAG00000013157	0.000	> 0.05	0.000
MMP1	ENSXMAG00000013976	0.000	> 0.05	0.000
SLC16A12	ENSXMAG00000017203	0.000	> 0.05	0.000
RFX2	ENSXMAG00000002530	0.000	> 0.05	0.000
EGR3	ENSXMAG00000001144	0.000	0.000	> 0.05
MT1F	ENSXMAG00000020384	0.000	0.000	> 0.05
CCKAR	ENSXMAG00000009902	0.000	0.000	> 0.05
USP30	ENSXMAG00000018244	0.001	0.000	> 0.05
CPNE2	ENSXMAG00000005853	0.001	> 0.05	0.024
SP140L	ENSXMAG00000001258	0.001	0.018	> 0.05
PRDM8	ENSXMAG00000009993	0.001	0.010	> 0.05
GRIN1	ENSXMAG00000003092	0.001	0.008	> 0.05
NPAS4	ENSXMAG00000013170	0.001	> 0.05	0.045
NRN1	ENSXMAG00000012826	0.001	> 0.05	0.029
CD44	ENSXMAG00000000801	0.001	> 0.05	0.000
PHLDA2	ENSXMAG00000019767	0.001	0.015	> 0.05

Table C-2**.** Continued

GeneName	GeneID	BM	CM	CB
ZNFX1	ENSXMAG00000005859	0.001	> 0.05	0.000
NPY	ENSXMAG00000000838	0.001	> 0.05	0.001
NEDD9	ENSXMAG00000016120	0.002	0.002	> 0.05
NRSN2	ENSXMAG00000013627	0.002	> 0.05	0.001
SOX11	ENSXMAG00000020179	0.002	> 0.05	0.000
NECAB2	ENSXMAG00000013000	0.002	> 0.05	0.007
CEBPB	ENSXMAG00000019665	0.002	0.000	> 0.05
TMEM240	ENSXMAG00000012610	0.002	> 0.05	0.027
CREM	ENSXMAG00000006361	0.002	0.000	> 0.05
RAMP3	ENSXMAG00000006290	0.003	0.017	> 0.05
SLC6A15	ENSXMAG00000002814	0.003	> 0.05	0.003
DLK1	ENSXMAG00000007730	0.004	> 0.05	0.006
NAB2	ENSXMAG00000003591	0.005	0.020	> 0.05
SEZ6L2	ENSXMAG00000011571	0.005	> 0.05	0.024
RSAD1	ENSXMAG00000012923	0.005	0.000	> 0.05
SYBU	ENSXMAG00000007087	0.006	> 0.05	0.027
SLC6A11	ENSXMAG00000017478	0.006	> 0.05	0.029
GABRD	ENSXMAG00000019394	0.006	> 0.05	0.028
PADI3	ENSXMAG00000017865	0.006	0.016	> 0.05
MAFK	ENSXMAG00000009331	0.006	0.000	> 0.05
PRRT1	ENSXMAG00000014134	0.012	> 0.05	0.011
KIFC2	ENSXMAG00000014088	0.013	> 0.05	0.019
SKIL	ENSXMAG00000000477	0.016	> 0.05	0.000
FZD10	ENSXMAG00000020088	0.017	> 0.05	0.001
AHNAK	ENSXMAG00000010087	0.019	> 0.05	0.000
AK5	ENSXMAG00000016357	0.020	> 0.05	0.015
ELAVL2	ENSXMAG00000004750	0.022	> 0.05	0.043
SNCA	ENSXMAG00000014135	0.022	> 0.05	0.031
RASA3	ENSXMAG00000019239	0.022	> 0.05	0.000
TIMP3	ENSXMAG00000010147	0.023	> 0.05	0.007
INPP1	ENSXMAG00000003597	0.024	0.000	> 0.05
TPPP	ENSXMAG00000014084	0.024	0.006	> 0.05
STAT2	ENSXMAG00000002056	0.024	> 0.05	0.000
DNER	ENSXMAG00000010820	0.027	> 0.05	0.018
NAV2	ENSXMAG00000010189	0.031	> 0.05	0.000
RABGAP1	ENSXMAG00000010078	0.031	0.009	> 0.05
MAP3K5	ENSXMAG00000016278	0.037	0.000	> 0.05
GPR152	ENSXMAG00000019184	0.038	0.000	> 0.05
ESRP1	ENSXMAG00000013126	0.039	0.029	> 0.05
RASSF6	ENSXMAG00000004159	0.039	> 0.05	0.004
HBEGF	ENSXMAG00000010006	0.041	0.001	> 0.05
ZFPM2	ENSXMAG00000003403	0.041	> 0.05	0.041

Table C-2**.** Continued

GeneName	GeneID	BM	CM	CB
NDRG4	ENSXMAG00000015838	0.041	> 0.05	0.010
TNMD	ENSXMAG00000013726	0.045	> 0.05	0.000
ACTG1	ENSXMAG00000015551	0.045	> 0.05	0.002
AMER2	ENSXMAG00000000702	0.046	> 0.05	0.008
ABCA12	ENSXMAG00000015970	> 0.05	0.000	0.043
ABCB7	ENSXMAG00000009070	> 0.05	0.007	0.028
ABCE1	ENSXMAG00000001778	> 0.05	0.004	0.024
ABCF2	ENSXMAG00000001927	> 0.05	0.000	0.000
ABCF3	ENSXMAG00000014038	> 0.05	0.004	0.001
ABHD14B	ENSXMAG00000016772	> 0.05	0.047	0.044
ABHD15	ENSXMAG00000008472	> 0.05	0.000	0.000
ABHD17B	ENSXMAG00000001288	> 0.05	0.000	0.000
ABTB1	ENSXMAG00000015862	> 0.05	0.006	0.005
ACBD3	ENSXMAG00000007475	> 0.05	0.000	0.000
ACER2	ENSXMAG00000005820	> 0.05	0.038	0.004
ACOT7	ENSXMAG00000018646	> 0.05	0.000	0.000
ACOX1	ENSXMAG00000006043	> 0.05	0.027	0.006
ADA	ENSXMAG00000001249	> 0.05	0.000	0.000
ADAM19	ENSXMAG00000010315	> 0.05	0.000	0.000
ADAMTS18	ENSXMAG00000014796	> 0.05	0.032	0.034
AFTPH	ENSXMAG00000015767	> 0.05	0.037	0.007
AGO3	ENSXMAG00000007141	> 0.05	0.000	0.000
AGTR1	ENSXMAG00000020155	> 0.05	0.000	0.000
AGXT2L1	ENSXMAG00000014966	> 0.05	0.000	0.000
AGXT2L2	ENSXMAG00000008961	> 0.05	0.018	0.003
ALPK1	ENSXMAG00000011532	> 0.05	0.005	0.035
AMMECR1	ENSXMAG00000005004	> 0.05	0.001	0.003
AMT	ENSXMAG00000019322	> 0.05	0.013	0.013
ANGPTL5	ENSXMAG00000010840	> 0.05	0.002	0.000
ANKFN1	ENSXMAG00000016934	> 0.05	0.007	0.041
ANKRD13B	ENSXMAG00000012073	> 0.05	0.004	0.029
ANKRD32	ENSXMAG00000019402	> 0.05	0.002	0.037
ANKRD50	ENSXMAG00000018353	> 0.05	0.009	0.000
ANO6	ENSXMAG00000004021	> 0.05	0.008	0.008
AP1G1	ENSXMAG00000005983	> 0.05	0.006	0.005
APLP1	ENSXMAG00000010499	> 0.05	0.006	0.000
APPL2	ENSXMAG00000017453	> 0.05	0.002	0.008
AQP3	ENSXMAG00000019407	> 0.05	0.000	0.000
ARFGAP2	ENSXMAG00000014012	> 0.05	0.025	0.031
ARG2	ENSXMAG00000014792	> 0.05	0.000	0.000
ARHGEF26	ENSXMAG00000001477	> 0.05	0.000	0.000
ARHGEF33	ENSXMAG00000017168	> 0.05	0.001	0.004

Table C-2**.** Continued

GeneName	GeneID	BM	CM	CB
ARID5B	ENSXMAG00000018209	> 0.05	0.000	0.000
ARL14EP	ENSXMAG00000000016	> 0.05	0.013	0.003
ARNTL	ENSXMAG00000015619	> 0.05	0.000	0.001
ARTN	ENSXMAG00000003941	> 0.05	0.031	0.006
ASTE1	ENSXMAG00000004808	> 0.05	0.043	0.016
ATAD3A	ENSXMAG00000016601	> 0.05	0.000	0.000
ATL2	ENSXMAG00000018291	> 0.05	0.004	0.015
ATP10B	ENSXMAG00000009322	> 0.05	0.000	0.039
ATP11C	ENSXMAG00000013304	> 0.05	0.000	0.002
ATP13A1	ENSXMAG00000013346	> 0.05	0.002	0.002
ATP8B1	ENSXMAG00000002723	> 0.05	0.000	0.004
ATXN1L	ENSXMAG00000005978	> 0.05	0.009	0.023
ATXN2L	ENSXMAG00000010133	> 0.05	0.003	0.000
AVPI1	ENSXMAG00000020097	> 0.05	0.000	0.004
BANK1	ENSXMAG00000008559	> 0.05	0.000	0.024
BARD1	ENSXMAG00000015997	> 0.05	0.012	0.032
BARX1	ENSXMAG00000007527	> 0.05	0.003	0.044
BARX2	ENSXMAG00000000962	> 0.05	0.000	0.000
BAZ1A	ENSXMAG00000004820	> 0.05	0.005	0.004
BCKDK	ENSXMAG00000008794	> 0.05	0.000	0.000
BMP4	ENSXMAG00000000766	> 0.05	0.034	0.024
BSPRY	ENSXMAG00000005501	> 0.05	0.000	0.000
BTBD1	ENSXMAG00000015211	> 0.05	0.028	0.047
C10ORF118	ENSXMAG00000008110	> 0.05	0.001	0.000
C10ORF54	ENSXMAG00000004819	> 0.05	0.010	0.010
C12ORF23	ENSXMAG00000001688	> 0.05	0.003	0.001
C2CD4A	ENSXMAG00000020324	> 0.05	0.000	0.000
CABLES1	ENSXMAG00000016420	> 0.05	0.003	0.002
CAD	ENSXMAG00000005369	> 0.05	0.007	0.006
CAPN8	ENSXMAG00000006017	> 0.05	0.000	0.005
CAPN9	ENSXMAG00000016015	> 0.05	0.000	0.005
CARD11	ENSXMAG00000001085	> 0.05	0.011	0.001
CARM1	ENSXMAG00000002542	> 0.05	0.000	0.000
CASC3	ENSXMAG00000013937	> 0.05	0.000	0.000
CASZ1	ENSXMAG00000011322	> 0.05	0.000	0.000
CAT	ENSXMAG00000019152	> 0.05	0.004	0.022
CCBE1	ENSXMAG00000004414	> 0.05	0.004	0.001
CCDC85B	ENSXMAG00000019513	> 0.05	0.005	0.000
CCDC89	ENSXMAG00000004854	> 0.05	0.005	0.022
CCNT2	ENSXMAG00000002028	> 0.05	0.000	0.000
CCPG1	ENSXMAG00000014277	> 0.05	0.004	0.040
CDC42BPB	ENSXMAG00000013537	> 0.05	0.000	0.000

Table C-2**.** Continued

GeneName	GeneID	BM	CM	CB
CDCA7	ENSXMAG00000005311	> 0.05	0.002	0.012
CDK12	ENSXMAG00000008716	> 0.05	0.000	0.000
CDS1	ENSXMAG00000010049	> 0.05	0.000	0.010
CDV3	ENSXMAG00000001835	> 0.05	0.006	0.008
CHMP2B	ENSXMAG00000008780	> 0.05	0.000	0.001
CHPF2	ENSXMAG00000001943	> 0.05	0.011	0.000
CKAP4	ENSXMAG00000001319	> 0.05	0.030	0.001
CLDN23	ENSXMAG00000020077	> 0.05	0.000	0.000
CLIC3	ENSXMAG00000001265	> 0.05	0.000	0.000
CLPTM1	ENSXMAG00000004298	> 0.05	0.001	0.000
CLTC	ENSXMAG00000000159	> 0.05	0.002	0.035
CMIP	ENSXMAG00000006193	> 0.05	0.001	0.000
CNKSR1	ENSXMAG00000013641	> 0.05	0.030	0.002
CNPPD1	ENSXMAG00000003024	> 0.05	0.000	0.014
COASY	ENSXMAG00000015601	> 0.05	0.041	0.013
COQ10A	ENSXMAG00000011840	> 0.05	0.000	0.000
CPEB2	ENSXMAG00000002128	> 0.05	0.009	0.012
CPEB3	ENSXMAG00000007652	> 0.05	0.011	0.044
CRY1	ENSXMAG00000009278	> 0.05	0.002	0.008
CSNK1D	ENSXMAG00000003430	> 0.05	0.002	0.000
CXADR	ENSXMAG00000001502	> 0.05	0.008	0.022
CYB5B	ENSXMAG00000003284	> 0.05	0.005	0.000
CYP1A2	ENSXMAG00000016761	> 0.05	0.002	0.005
CYP24A1	ENSXMAG00000010218	> 0.05	0.000	0.000
CYP26C1	ENSXMAG00000004889	> 0.05	0.000	0.022
CYP27B1	ENSXMAG00000005176	> 0.05	0.000	0.022
CYSLTR1	ENSXMAG00000017569	> 0.05	0.000	0.000
DAO	ENSXMAG00000018217	> 0.05	0.035	0.039
DAPL1	ENSXMAG00000011382	> 0.05	0.007	0.037
DCUN1D5	ENSXMAG00000004471	> 0.05	0.020	0.000
DDB2	ENSXMAG00000003694	> 0.05	0.001	0.003
DDC	ENSXMAG00000000755	> 0.05	0.047	0.002
DDHD1	ENSXMAG00000009680	> 0.05	0.002	0.017
DDIT3	ENSXMAG00000017918	> 0.05	0.014	0.025
DDX47	ENSXMAG00000008268	> 0.05	0.027	0.034
DGKG	ENSXMAG00000005254	> 0.05	0.007	0.036
DGKQ	ENSXMAG00000015446	> 0.05	0.017	0.015
DHDDS	ENSXMAG00000013272	> 0.05	0.001	0.000
DHDH	ENSXMAG00000000038	> 0.05	0.034	0.014
DHRS1	ENSXMAG00000009929	> 0.05	0.001	0.025
DHX30	ENSXMAG00000016336	> 0.05	0.001	0.003
DNAJC4	ENSXMAG00000003217	> 0.05	0.000	0.003

Table C-2**.** Continued

GeneName	GeneID	BM	CM	CB
DNAJC9	ENSXMAG00000018555	> 0.05	0.025	0.022
DNMT3B	ENSXMAG00000009115	> 0.05	0.000	0.025
DNTTIP2	ENSXMAG00000003347	> 0.05	0.001	0.000
DOM3Z	ENSXMAG00000001092	> 0.05	0.033	0.036
DOT1L	ENSXMAG00000012854	> 0.05	0.000	0.000
DPAGT1	ENSXMAG00000005959	> 0.05	0.007	0.030
DSC1	ENSXMAG00000016712	> 0.05	0.000	0.000
DTWD2	ENSXMAG00000018696	> 0.05	0.023	0.016
DTX1	ENSXMAG00000015100	> 0.05	0.003	0.016
DUSP2	ENSXMAG00000014933	> 0.05	0.000	0.000
ECE1	ENSXMAG00000017683	> 0.05	0.000	0.000
EDA	ENSXMAG00000009620	> 0.05	0.002	0.049
EDAR	ENSXMAG00000016298	> 0.05	0.000	0.000
EGFL6	ENSXMAG00000004416	> 0.05	0.047	0.017
EGFL8	ENSXMAG00000005711	> 0.05	0.032	0.008
EGLN3	ENSXMAG00000000068	> 0.05	0.000	0.000
EHHADH	ENSXMAG00000015547	> 0.05	0.017	0.012
ELANE	ENSXMAG00000012492	> 0.05	0.017	0.008
ELF2	ENSXMAG00000017764	> 0.05	0.000	0.011
ELF4	ENSXMAG00000013721	> 0.05	0.001	0.003
ELOVL5	ENSXMAG00000010261	> 0.05	0.000	0.000
ELOVL7	ENSXMAG00000015573	> 0.05	0.000	0.000
ELTD1	ENSXMAG00000017894	> 0.05	0.049	0.001
EMC4	ENSXMAG00000001312	> 0.05	0.006	0.015
ENTPD3	ENSXMAG00000000746	> 0.05	0.000	0.000
ENTPD7	ENSXMAG00000005683	> 0.05	0.000	0.000
EPGN	ENSXMAG00000010272	> 0.05	0.000	0.013
EPS15L1	ENSXMAG00000000632	> 0.05	0.019	0.020
EPT1	ENSXMAG00000010844	> 0.05	0.000	0.000
ERBB2	ENSXMAG00000018390	> 0.05	0.000	0.000
ERGIC1	ENSXMAG00000009666	> 0.05	0.028	0.002
ETF1	ENSXMAG00000004958	> 0.05	0.001	0.000
ETNK2	ENSXMAG00000008864	> 0.05	0.000	0.000
ETV7	ENSXMAG00000008241	> 0.05	0.036	0.027
EYA2	ENSXMAG00000019373	> 0.05	0.014	0.000
EZH1	ENSXMAG00000009130	> 0.05	0.002	0.015
EZH2	ENSXMAG00000012115	> 0.05	0.015	0.006
FAHD2B	ENSXMAG00000014921	> 0.05	0.003	0.022
FAM101A	ENSXMAG00000011724	> 0.05	0.000	0.000
FAM110C	ENSXMAG00000007737	> 0.05	0.002	0.001
FAM120C	ENSXMAG00000009559	> 0.05	0.000	0.000
FAM13A	ENSXMAG00000009862	> 0.05	0.001	0.011

Table C-2**.** Continued

GeneName	GeneID	BM	CM	CB
FAM175B	ENSXMAG00000012550	> 0.05	0.024	0.037
FAM208B	ENSXMAG00000016889	> 0.05	0.001	0.020
FAM214A	ENSXMAG00000018302	> 0.05	0.005	0.004
FAM60A	ENSXMAG00000009068	> 0.05	0.000	0.011
FAM65C	ENSXMAG00000012157	> 0.05	0.005	0.023
FAM83C	ENSXMAG00000017094	> 0.05	0.039	0.001
FBXO10	ENSXMAG00000012473	> 0.05	0.045	0.001
FBXO30	ENSXMAG00000015378	> 0.05	0.000	0.002
FBXW7	ENSXMAG00000008053	> 0.05	0.000	0.000
FCGBP	ENSXMAG00000010439	> 0.05	0.000	0.001
FDFT1	ENSXMAG00000003786	> 0.05	0.031	0.045
FECH	ENSXMAG00000001624	> 0.05	0.000	0.000
FERMT2	ENSXMAG00000009635	> 0.05	0.003	0.001
FGF16	ENSXMAG00000009107	> 0.05	0.000	0.001
FGF18	ENSXMAG00000008345	> 0.05	0.000	0.000
FGF2	ENSXMAG00000018373	> 0.05	0.034	0.000
FGFR2	ENSXMAG00000009512	> 0.05	0.004	0.025
FILIP1L	ENSXMAG00000009955	> 0.05	0.000	0.000
FKBP5	ENSXMAG00000015416	> 0.05	0.000	0.000
FKBP8	ENSXMAG00000012879	> 0.05	0.049	0.050
FOXN1	ENSXMAG00000011786	> 0.05	0.000	0.000
FOXN3	ENSXMAG00000001667	> 0.05	0.037	0.031
FOXO4	ENSXMAG00000013124	> 0.05	0.000	0.000
FOXO5	ENSXMAG00000003269	> 0.05	0.000	0.000
FOXRED1	ENSXMAG00000011302	> 0.05	0.001	0.033
G3BP1	ENSXMAG00000019255	> 0.05	0.000	0.000
G3BP2	ENSXMAG00000010011	> 0.05	0.000	0.000
GALNT3	ENSXMAG00000011705	> 0.05	0.000	0.002
GAN	ENSXMAG00000006216	> 0.05	0.050	0.015
GAS2L1	ENSXMAG00000017929	> 0.05	0.002	0.000
GCLC	ENSXMAG00000010230	> 0.05	0.006	0.012
GDPD2	ENSXMAG00000013255	> 0.05	0.005	0.002
GDPD3	ENSXMAG00000011975	> 0.05	0.000	0.000
GFM1	ENSXMAG00000000261	> 0.05	0.000	0.000
GFOD1	ENSXMAG00000005406	> 0.05	0.026	0.017
GFOD2	ENSXMAG00000014083	> 0.05	0.009	0.032
GFPT2	ENSXMAG00000008775	> 0.05	0.041	0.044
GLCE	ENSXMAG00000000463	> 0.05	0.020	0.002
GMPPB	ENSXMAG00000013349	> 0.05	0.000	0.001
GNA15	ENSXMAG00000013441	> 0.05	0.000	0.000
GNE	ENSXMAG00000006581	> 0.05	0.000	0.000
GOPC	ENSXMAG00000005257	> 0.05	0.014	0.037

Table C-2**.** Continued

GeneName	GeneID	BM	CM	CB
GPR124	ENSXMAG00000001857	> 0.05	0.020	0.009
GPR132	ENSXMAG00000019758	> 0.05	0.004	0.001
GRAP2	ENSXMAG00000000013	> 0.05	0.003	0.011
GRASP	ENSXMAG00000011711	> 0.05	0.000	0.003
GRB7	ENSXMAG00000016919	> 0.05	0.029	0.001
GRHL3	ENSXMAG00000013634	> 0.05	0.002	0.001
H1F0	ENSXMAG00000020306	> 0.05	0.000	0.000
HACL1	ENSXMAG00000013857	> 0.05	0.001	0.013
HEATR6	ENSXMAG00000010799	> 0.05	0.000	0.000
HERPUD1	ENSXMAG00000015722	> 0.05	0.032	0.028
HEYL	ENSXMAG00000006950	> 0.05	0.014	0.008
HIF3A	ENSXMAG00000003186	> 0.05	0.000	0.000
HIGD1C	ENSXMAG00000005817	> 0.05	0.000	0.000
HMGXB3	ENSXMAG00000008628	> 0.05	0.009	0.014
HNRNPU	ENSXMAG00000010356	> 0.05	0.030	0.042
HOMER2	ENSXMAG00000002059	> 0.05	0.000	0.000
HOMEZ	ENSXMAG00000020098	> 0.05	0.005	0.001
HOOK2	ENSXMAG00000009113	> 0.05	0.000	0.012
HPRT1	ENSXMAG00000017356	> 0.05	0.002	0.006
HPS1	ENSXMAG00000002534	> 0.05	0.009	0.021
HSD17B3	ENSXMAG00000015794	> 0.05	0.000	0.002
HSPA12B	ENSXMAG00000008789	> 0.05	0.008	0.038
HSPA13	ENSXMAG00000010122	> 0.05	0.003	0.000
HYAL4	ENSXMAG00000014418	> 0.05	0.004	0.005
HYOU1	ENSXMAG00000001626	> 0.05	0.000	0.004
ICMT	ENSXMAG00000004596	> 0.05	0.004	0.005
ID2	ENSXMAG00000008031	> 0.05	0.000	0.003
IDE	ENSXMAG00000007681	> 0.05	0.000	0.013
IER5L	ENSXMAG00000000145	> 0.05	0.000	0.000
IFFO2	ENSXMAG00000017316	> 0.05	0.000	0.000
IFRD1	ENSXMAG00000009043	> 0.05	0.000	0.000
IGSF8	ENSXMAG00000017842	> 0.05	0.018	0.000
IKBIP	ENSXMAG00000001321	> 0.05	0.034	0.033
IL17C	ENSXMAG00000006238	> 0.05	0.000	0.003
IL17RE	ENSXMAG00000015725	> 0.05	0.008	0.000
IL1RAP	ENSXMAG00000014781	> 0.05	0.041	0.000
IL20RA	ENSXMAG00000007111	> 0.05	0.000	0.011
ILVBL	ENSXMAG00000001360	> 0.05	0.008	0.013
IMP4	ENSXMAG00000004915	> 0.05	0.007	0.010
INPP5F	ENSXMAG00000007830	> 0.05	0.022	0.011
IPMK	ENSXMAG00000014549	> 0.05	0.000	0.000
IPO4	ENSXMAG00000006041	> 0.05	0.007	0.005

Table C-2**.** Continued

GeneName	GeneID	BM	CM	CB
IPO5	ENSXMAG00000004307	> 0.05	0.000	0.000
IPO7	ENSXMAG00000009410	> 0.05	0.005	0.012
IPPK	ENSXMAG00000004787	> 0.05	0.000	0.001
IRF7	ENSXMAG00000015063	> 0.05	0.000	0.047
IRS4	ENSXMAG00000014720	> 0.05	0.021	0.015
IRX3	ENSXMAG00000016177	> 0.05	0.000	0.011
ISCA1	ENSXMAG00000002782	> 0.05	0.000	0.000
ITGA4	ENSXMAG00000001540	> 0.05	0.000	0.000
ITGB4	ENSXMAG00000012430	> 0.05	0.000	0.000
ITSN1	ENSXMAG00000013547	> 0.05	0.000	0.000
JAG1	ENSXMAG00000006308	> 0.05	0.011	0.000
JHDM1D	ENSXMAG00000004648	> 0.05	0.000	0.000
JMJD1C	ENSXMAG00000014154	> 0.05	0.000	0.002
JUP	ENSXMAG00000000209	> 0.05	0.000	0.005
KAT2B	ENSXMAG00000003253	> 0.05	0.035	0.004
KCNQ5	ENSXMAG00000012035	> 0.05	0.001	0.006
KCTD3	ENSXMAG00000017642	> 0.05	0.000	0.010
KDELC1	ENSXMAG00000011560	> 0.05	0.000	0.004
KEAP1	ENSXMAG00000007183	> 0.05	0.000	0.000
KIAA0196	ENSXMAG00000008620	> 0.05	0.017	0.047
KIAA1324	ENSXMAG00000010362	> 0.05	0.004	0.013
KIF7	ENSXMAG00000013974	> 0.05	0.004	0.000
KLC1	ENSXMAG00000018455	> 0.05	0.005	0.004
KLF12	ENSXMAG00000011014	> 0.05	0.030	0.001
KLF13	ENSXMAG00000003301	> 0.05	0.000	0.000
KLF8	ENSXMAG00000005126	> 0.05	0.018	0.041
KLF9	ENSXMAG00000001428	> 0.05	0.000	0.000
KLHL12	ENSXMAG00000011484	> 0.05	0.001	0.000
KLHL9	ENSXMAG00000012602	> 0.05	0.037	0.000
KPNA3	ENSXMAG00000011285	> 0.05	0.000	0.000
KPNA6	ENSXMAG00000000350	> 0.05	0.003	0.040
KREMEN1	ENSXMAG00000002678	> 0.05	0.000	0.000
LACTB2	ENSXMAG00000006526	> 0.05	0.042	0.043
LEPR	ENSXMAG00000002246	> 0.05	0.000	0.000
LEPREL2	ENSXMAG00000011089	> 0.05	0.023	0.035
LETM1	ENSXMAG00000008107	> 0.05	0.013	0.009
LGR6	ENSXMAG00000008206	> 0.05	0.005	0.000
LHPP	ENSXMAG00000016538	> 0.05	0.043	0.016
LIMK2	ENSXMAG00000018076	> 0.05	0.000	0.000
LMTK2	ENSXMAG00000014669	> 0.05	0.018	0.034
LPL	ENSXMAG00000003404	> 0.05	0.005	0.001
LRCH3	ENSXMAG00000014055	> 0.05	0.008	0.015

Table C-2**.** Continued

GeneName	GeneID	BM	CM	CB
LRIG3	ENSXMAG00000010178	> 0.05	0.004	0.000
LRR1	ENSXMAG00000010950	> 0.05	0.000	0.000
LRRC1	ENSXMAG00000010214	> 0.05	0.000	0.000
LRRC45	ENSXMAG00000014712	> 0.05	0.032	0.015
LRRC58	ENSXMAG00000004050	> 0.05	0.024	0.005
LRRCC1	ENSXMAG00000007622	> 0.05	0.001	0.033
LSM12	ENSXMAG00000017589	> 0.05	0.047	0.027
LSR	ENSXMAG00000006612	> 0.05	0.008	0.002
MACC1	ENSXMAG00000010861	> 0.05	0.003	0.000
MAP3K19	ENSXMAG00000002035	> 0.05	0.000	0.006
MAP3K4	ENSXMAG00000009609	> 0.05	0.019	0.000
MAPK7	ENSXMAG00000013778	> 0.05	0.003	0.031
MAPRE1	ENSXMAG00000009094	> 0.05	0.017	0.011
MARK2	ENSXMAG00000008732	> 0.05	0.003	0.030
MARS	ENSXMAG00000017920	> 0.05	0.020	0.005
MBD3L2	ENSXMAG00000003932	> 0.05	0.013	0.000
MBD4	ENSXMAG00000017444	> 0.05	0.022	0.036
MBOAT1	ENSXMAG00000009682	> 0.05	0.000	0.000
MDC1	ENSXMAG00000010974	> 0.05	0.018	0.011
MEIS2	ENSXMAG00000009557	> 0.05	0.000	0.001
MEX3D	ENSXMAG00000019311	> 0.05	0.001	0.001
MFSD12	ENSXMAG00000018868	> 0.05	0.030	0.009
MFSD6	ENSXMAG00000010414	> 0.05	0.000	0.000
MFSD6L	ENSXMAG00000020358	> 0.05	0.001	0.002
MFSD9	ENSXMAG00000011576	> 0.05	0.001	0.000
MGLL	ENSXMAG00000009052	> 0.05	0.050	0.007
MGST1	ENSXMAG00000015719	> 0.05	0.006	0.000
MIDN	ENSXMAG00000012420	> 0.05	0.000	0.000
MKNK2	ENSXMAG00000002666	> 0.05	0.002	0.000
MLEC	ENSXMAG00000002971	> 0.05	0.000	0.000
MORC3	ENSXMAG00000004110	> 0.05	0.013	0.004
MPV17L2	ENSXMAG00000002525	> 0.05	0.002	0.000
MSX1	ENSXMAG00000018444	> 0.05	0.012	0.006
MTHFR	ENSXMAG00000012737	> 0.05	0.023	0.025
MXD1	ENSXMAG00000004566	> 0.05	0.000	0.003
MYCN	ENSXMAG00000003386	> 0.05	0.001	0.010
MYH14	ENSXMAG00000012405	> 0.05	0.000	0.000
MYO1E	ENSXMAG00000006657	> 0.05	0.000	0.002
MYO5B	ENSXMAG00000013502	> 0.05	0.000	0.003
MYSM1	ENSXMAG00000011201	> 0.05	0.046	0.039
MYZAP	ENSXMAG00000014372	> 0.05	0.000	0.003
NAA25	ENSXMAG00000005930	> 0.05	0.039	0.038

Table C-2**.** Continued

GeneName	GeneID	BM	CM	CB
NAA30	ENSXMAG00000008386	> 0.05	0.004	0.032
NCLN	ENSXMAG00000018708	> 0.05	0.000	0.000
NEDD4	ENSXMAG00000011211	> 0.05	0.000	0.000
NEIL1	ENSXMAG00000015912	> 0.05	0.000	0.000
NEK1	ENSXMAG00000011778	> 0.05	0.039	0.044
NFE2L3	ENSXMAG00000001015	> 0.05	0.000	0.000
NIPA2	ENSXMAG00000016124	> 0.05	0.000	0.000
NIPAL3	ENSXMAG00000007686	> 0.05	0.001	0.000
NKRF	ENSXMAG00000013716	> 0.05	0.020	0.009
NOD1	ENSXMAG00000001359	> 0.05	0.004	0.029
NOS1	ENSXMAG00000002504	> 0.05	0.000	0.000
NOXO1	ENSXMAG00000007295	> 0.05	0.021	0.021
NPC1	ENSXMAG00000016426	> 0.05	0.000	0.000
NR1I2	ENSXMAG00000003990	> 0.05	0.038	0.005
NT5C	ENSXMAG00000002512	> 0.05	0.013	0.044
NT5DC2	ENSXMAG00000016109	> 0.05	0.000	0.000
NTHL1	ENSXMAG00000012778	> 0.05	0.006	0.026
NUFIP2	ENSXMAG00000012404	> 0.05	0.004	0.008
NUMB	ENSXMAG00000003561	> 0.05	0.000	0.000
NUP153	ENSXMAG00000003437	> 0.05	0.001	0.002
OAF	ENSXMAG00000008884	> 0.05	0.002	0.000
ODC1	ENSXMAG00000018490	> 0.05	0.000	0.000
OFD1	ENSXMAG00000004394	> 0.05	0.037	0.002
OSR1	ENSXMAG00000015661	> 0.05	0.000	0.000
OSTC	ENSXMAG00000015007	> 0.05	0.001	0.000
P2RY6	ENSXMAG00000018087	> 0.05	0.007	0.042
PACS1	ENSXMAG00000008449	> 0.05	0.030	0.039
PAQR5	ENSXMAG00000000461	> 0.05	0.000	0.000
PARD6A	ENSXMAG00000016111	> 0.05	0.000	0.003
PCGF2	ENSXMAG00000018149	> 0.05	0.011	0.001
PCK1	ENSXMAG00000010445	> 0.05	0.000	0.001
PDCD4	ENSXMAG00000006928	> 0.05	0.002	0.021
PDE4D	ENSXMAG00000015552	> 0.05	0.047	0.007
PDE8A	ENSXMAG00000002082	> 0.05	0.020	0.040
PDGFC	ENSXMAG00000002233	> 0.05	0.007	0.006
PDK2	ENSXMAG00000014954	> 0.05	0.000	0.000
PDK3	ENSXMAG00000019098	> 0.05	0.008	0.026
PDP1	ENSXMAG00000013155	> 0.05	0.000	0.000
PDPK1	ENSXMAG00000018875	> 0.05	0.002	0.003
PDS5B	ENSXMAG00000008480	> 0.05	0.003	0.000
PDZD11	ENSXMAG00000013263	> 0.05	0.000	0.000
PEX1	ENSXMAG00000002844	> 0.05	0.011	0.004

Table C-2**.** Continued

GeneName	GeneID	BM	CM	CB
PEX5	ENSXMAG00000011133	> 0.05	0.018	0.006
PFAS	ENSXMAG00000011539	> 0.05	0.030	0.038
PGBD4	ENSXMAG00000020313	> 0.05	0.041	0.003
PGP	ENSXMAG00000011747	> 0.05	0.024	0.000
PHKG2	ENSXMAG00000018776	> 0.05	0.000	0.000
PHYHD1	ENSXMAG00000005677	> 0.05	0.007	0.002
PIAS4	ENSXMAG00000012759	> 0.05	0.022	0.020
PIGO	ENSXMAG00000004993	> 0.05	0.024	0.012
PIM1	ENSXMAG00000016388	> 0.05	0.002	0.000
PITPNB	ENSXMAG00000001916	> 0.05	0.000	0.000
PITX1	ENSXMAG00000008331	> 0.05	0.000	0.043
PLCB4	ENSXMAG00000008481	> 0.05	0.011	0.015
PLCG1	ENSXMAG00000015379	> 0.05	0.000	0.003
PLEKHG7	ENSXMAG00000001345	> 0.05	0.000	0.000
PLIN4	ENSXMAG00000013240	> 0.05	0.000	0.031
PLSCR3	ENSXMAG00000015802	> 0.05	0.032	0.033
PMM1	ENSXMAG00000006944	> 0.05	0.000	0.003
PNKP	ENSXMAG00000002791	> 0.05	0.008	0.008
PNPLA2	ENSXMAG00000003849	> 0.05	0.008	0.000
POLG	ENSXMAG00000000917	> 0.05	0.002	0.002
POLR2A	ENSXMAG00000015687	> 0.05	0.002	0.027
POLR3B	ENSXMAG00000001568	> 0.05	0.006	0.015
POLR3C	ENSXMAG00000016215	> 0.05	0.010	0.030
POLRMT	ENSXMAG00000012464	> 0.05	0.030	0.003
PPARGC1B	ENSXMAG00000019179	> 0.05	0.014	0.000
PPL	ENSXMAG00000014943	> 0.05	0.000	0.000
PPP2R2D	ENSXMAG00000012743	> 0.05	0.002	0.019
PPP4C	ENSXMAG00000002510	> 0.05	0.002	0.000
PRDM15	ENSXMAG00000013849	> 0.05	0.012	0.002
PRKAA1	ENSXMAG00000016837	> 0.05	0.008	0.007
PRMT5	ENSXMAG00000000469	> 0.05	0.000	0.000
PRPF4B	ENSXMAG00000001762	> 0.05	0.029	0.011
PRRG4	ENSXMAG00000014189	> 0.05	0.026	0.000
PRSS12	ENSXMAG00000017835	> 0.05	0.013	0.018
PTDSS2	ENSXMAG00000000545	> 0.05	0.000	0.000
PTGES	ENSXMAG00000005893	> 0.05	0.046	0.018
PTGFRN	ENSXMAG00000004148	> 0.05	0.000	0.000
PTK6	ENSXMAG00000009161	> 0.05	0.001	0.035
PTPN21	ENSXMAG00000010368	> 0.05	0.000	0.000
PTPRJ	ENSXMAG00000018596	> 0.05	0.002	0.026
PTPRQ	ENSXMAG00000003867	> 0.05	0.029	0.047
QTRT1	ENSXMAG00000013159	> 0.05	0.024	0.004

Table C-2**.** Continued

GeneName	GeneID	BM	CM	CB
R3HDM2	ENSXMAG00000017978	> 0.05	0.039	0.006
RANBP2	ENSXMAG00000003519	> 0.05	0.000	0.000
RAPGEFL1	ENSXMAG00000012565	> 0.05	0.000	0.000
RASSF9	ENSXMAG00000002858	> 0.05	0.004	0.000
REEP3	ENSXMAG00000014205	> 0.05	0.001	0.000
REG1A	ENSXMAG00000008221	> 0.05	0.001	0.008
RELA	ENSXMAG00000011436	> 0.05	0.000	0.000
RER1	ENSXMAG00000017376	> 0.05	0.021	0.023
REXO1L10P	ENSXMAG00000002681	> 0.05	0.013	0.004
RFC1	ENSXMAG00000009566	> 0.05	0.001	0.003
RGS3	ENSXMAG00000005690	> 0.05	0.029	0.035
RHBDF2	ENSXMAG00000006018	> 0.05	0.000	0.000
RILP	ENSXMAG00000008504	> 0.05	0.004	0.003
RIPK2	ENSXMAG00000004540	> 0.05	0.022	0.035
RNF111	ENSXMAG00000014476	> 0.05	0.006	0.033
RNF13	ENSXMAG00000005738	> 0.05	0.003	0.001
RNPEPL1	ENSXMAG00000014126	> 0.05	0.000	0.000
RPRD1A	ENSXMAG00000003172	> 0.05	0.017	0.001
RPS6KB2	ENSXMAG00000004706	> 0.05	0.009	0.011
RPS6KC1	ENSXMAG00000003715	> 0.05	0.001	0.015
RRAD	ENSXMAG00000019054	> 0.05	0.000	0.034
RRP12	ENSXMAG00000013498	> 0.05	0.022	0.025
RSL1D1	ENSXMAG00000012616	> 0.05	0.008	0.004
SAR1B	ENSXMAG00000008294	> 0.05	0.027	0.021
SART3	ENSXMAG00000018160	> 0.05	0.034	0.031
SBNO2	ENSXMAG00000012091	> 0.05	0.001	0.000
SCAP	ENSXMAG00000019115	> 0.05	0.000	0.034
SCEL	ENSXMAG00000011150	> 0.05	0.000	0.000
SCXA	ENSXMAG00000008863	> 0.05	0.043	0.017
SCYL2	ENSXMAG00000018567	> 0.05	0.010	0.047
SDR42E2	ENSXMAG00000000979	> 0.05	0.046	0.037
SEC23B	ENSXMAG00000004757	> 0.05	0.000	0.023
SEC24A	ENSXMAG00000008298	> 0.05	0.008	0.015
SEC24C	ENSXMAG00000015879	> 0.05	0.002	0.015
SESN1	ENSXMAG00000003264	> 0.05	0.017	0.000
SETDB1	ENSXMAG00000016600	> 0.05	0.003	0.004
SFXN3	ENSXMAG00000002825	> 0.05	0.005	0.005
SH3RF1	ENSXMAG00000014129	> 0.05	0.029	0.000
SH3YL1	ENSXMAG00000001042	> 0.05	0.004	0.048
SHC1	ENSXMAG00000007234	> 0.05	0.002	0.000
SIK1	ENSXMAG00000012114	> 0.05	0.000	0.000
SIK2	ENSXMAG00000003840	> 0.05	0.000	0.000

Table C-2**.** Continued

GeneName	GeneID	BM	CM	CB
SIK3	ENSXMAG00000002893	> 0.05	0.002	0.000
SIX2	ENSXMAG00000016855	> 0.05	0.046	0.002
SLC12A6	ENSXMAG00000001221	> 0.05	0.000	0.000
SLC16A4	ENSXMAG00000010114	> 0.05	0.001	0.000
SLC16A7	ENSXMAG00000010168	> 0.05	0.000	0.003
SLC17A3	ENSXMAG00000016257	> 0.05	0.039	0.006
SLC23A2	ENSXMAG00000014856	> 0.05	0.007	0.016
SLC25A22	ENSXMAG00000003828	> 0.05	0.011	0.000
SLC25A30	ENSXMAG00000011434	> 0.05	0.000	0.000
SLC25A35	ENSXMAG00000012090	> 0.05	0.018	0.024
SLC25A37	ENSXMAG00000001218	> 0.05	0.013	0.004
SLC25A42	ENSXMAG00000019341	> 0.05	0.000	0.002
SLC27A4	ENSXMAG00000001495	> 0.05	0.000	0.001
SLC30A1	ENSXMAG00000017578	> 0.05	0.000	0.000
SLC35E1	ENSXMAG00000002620	> 0.05	0.000	0.000
SLC38A9	ENSXMAG00000007655	> 0.05	0.002	0.002
SLC44A4	ENSXMAG00000014065	> 0.05	0.000	0.024
SLC45A3	ENSXMAG00000016517	> 0.05	0.000	0.000
SLC6A8	ENSXMAG00000013815	> 0.05	0.000	0.000
SLC7A11	ENSXMAG00000010874	> 0.05	0.011	0.000
SLC7A3	ENSXMAG00000000319	> 0.05	0.000	0.000
SLC7A5	ENSXMAG00000016047	> 0.05	0.000	0.000
SLC7A8	ENSXMAG00000001695	> 0.05	0.000	0.001
SMAD7	ENSXMAG00000004239	> 0.05	0.001	0.015
SMARCA5	ENSXMAG00000009678	> 0.05	0.021	0.005
SMC4	ENSXMAG00000018289	> 0.05	0.014	0.043
SMC6	ENSXMAG00000015209	> 0.05	0.011	0.010
SMOX	ENSXMAG00000005505	> 0.05	0.000	0.000
SMURF1	ENSXMAG00000001013	> 0.05	0.000	0.000
SNRK	ENSXMAG00000002760	> 0.05	0.039	0.000
SOCS5	ENSXMAG00000019827	> 0.05	0.021	0.016
SOLH	ENSXMAG00000019044	> 0.05	0.000	0.000
SORBS3	ENSXMAG00000001865	> 0.05	0.004	0.000
SOX2	ENSXMAG00000020128	> 0.05	0.000	0.000
SOX3	ENSXMAG00000019515	> 0.05	0.000	0.001
SPATA2	ENSXMAG00000015528	> 0.05	0.000	0.000
SPDEF	ENSXMAG00000017942	> 0.05	0.005	0.012
SPECC1	ENSXMAG00000005351	> 0.05	0.030	0.044
SPRY4	ENSXMAG00000001556	> 0.05	0.002	0.000
SRMS	ENSXMAG00000009154	> 0.05	0.000	0.007
SRP68	ENSXMAG00000006101	> 0.05	0.005	0.002
SRPK2	ENSXMAG00000012620	> 0.05	0.009	0.002

Table C-2**.** Continued

GeneName	GeneID	BM	CM	CB
SRPR	ENSXMAG00000001315	> 0.05	0.004	0.001
SRPX2	ENSXMAG00000013538	> 0.05	0.040	0.005
ST6GAL2	ENSXMAG00000011642	> 0.05	0.005	0.001
STK35	ENSXMAG00000012847	> 0.05	0.000	0.000
STK39	ENSXMAG00000011767	> 0.05	0.002	0.005
STK40	ENSXMAG00000014777	> 0.05	0.000	0.000
SWI5	ENSXMAG00000001550	> 0.05	0.026	0.034
SYTL3	ENSXMAG00000017431	> 0.05	0.001	0.034
TAS1R3	ENSXMAG00000018681	> 0.05	0.000	0.000
TAT	ENSXMAG00000009040	> 0.05	0.006	0.009
TBC1D14	ENSXMAG00000011398	> 0.05	0.000	0.029
TBC1D17	ENSXMAG00000010584	> 0.05	0.030	0.002
TCEA3	ENSXMAG00000010685	> 0.05	0.000	0.000
TCF7L2	ENSXMAG00000014710	> 0.05	0.000	0.004
TCOF1	ENSXMAG00000005588	> 0.05	0.001	0.006
TDP1	ENSXMAG00000011680	> 0.05	0.018	0.011
TET2	ENSXMAG00000005438	> 0.05	0.002	0.002
TFAP2A	ENSXMAG00000012878	> 0.05	0.003	0.000
TFB2M	ENSXMAG00000017710	> 0.05	0.009	0.003
TGIF2-C20ORF24	ENSXMAG00000010500	> 0.05	0.000	0.000
TICAM1	ENSXMAG00000008608	> 0.05	0.001	0.027
TIMM17A	ENSXMAG00000005996	> 0.05	0.000	0.000
TINAGL1	ENSXMAG00000003194	> 0.05	0.000	0.000
TJP2	ENSXMAG00000005032	> 0.05	0.006	0.016
TLCD1	ENSXMAG00000011894	> 0.05	0.000	0.000
TLCD2	ENSXMAG00000008589	> 0.05	0.001	0.000
TLE1	ENSXMAG00000015157	> 0.05	0.005	0.000
TMED10	ENSXMAG00000004020	> 0.05	0.009	0.017
TMEM159	ENSXMAG00000000939	> 0.05	0.009	0.025
TMEM38B	ENSXMAG00000011387	> 0.05	0.019	0.010
TMEM41B	ENSXMAG00000009444	> 0.05	0.000	0.000
TMOD1	ENSXMAG00000006094	> 0.05	0.000	0.004
TMTC4	ENSXMAG00000013374	> 0.05	0.000	0.000
TNFAIP3	ENSXMAG00000018214	> 0.05	0.046	0.044
TNFRSF19	ENSXMAG00000001169	> 0.05	0.000	0.000
TNFSF11	ENSXMAG00000000329	> 0.05	0.001	0.034
TOPBP1	ENSXMAG00000001818	> 0.05	0.009	0.003
TP53INP1	ENSXMAG00000013571	> 0.05	0.001	0.000
TP73	ENSXMAG00000017360	> 0.05	0.004	0.000
TPP2	ENSXMAG00000019214	> 0.05	0.000	0.001
TRABD	ENSXMAG00000019199	> 0.05	0.006	0.022

Table C-2**.** Continued

GeneName	GeneID	BM	CM	CB
TRAM1L1	ENSXMAG00000002721	> 0.05	0.006	0.007
TRPM5	ENSXMAG00000011779	> 0.05	0.000	0.035
TSEN54	ENSXMAG00000018818	> 0.05	0.022	0.031
TSPAN19	ENSXMAG00000006666	> 0.05	0.042	0.014
TSPAN33	ENSXMAG00000011021	> 0.05	0.010	0.023
TSPAN8	ENSXMAG00000006565	> 0.05	0.031	0.021
TTC22	ENSXMAG00000002256	> 0.05	0.000	0.001
TTC39A	ENSXMAG00000003295	> 0.05	0.001	0.006
TUBGCP2	ENSXMAG00000014666	> 0.05	0.050	0.037
TXNRD3	ENSXMAG00000009544	> 0.05	0.000	0.000
UBA6	ENSXMAG00000012548	> 0.05	0.000	0.000
UBL3	ENSXMAG00000008399	> 0.05	0.000	0.000
UBQLN4	ENSXMAG00000003208	> 0.05	0.024	0.023
UEVLD	ENSXMAG00000015346	> 0.05	0.001	0.006
UPP1	ENSXMAG00000014219	> 0.05	0.000	0.000
UROS	ENSXMAG00000016520	> 0.05	0.044	0.033
USP4	ENSXMAG00000011719	> 0.05	0.000	0.000
USP44	ENSXMAG00000015410	> 0.05	0.003	0.007
UST	ENSXMAG00000007576	> 0.05	0.047	0.029
UTP15	ENSXMAG00000014206	> 0.05	0.001	0.001
VSIG10	ENSXMAG00000002588	> 0.05	0.000	0.002
WARS	ENSXMAG00000007718	> 0.05	0.009	0.000
WDR11	ENSXMAG00000007153	> 0.05	0.018	0.021
WNK3	ENSXMAG00000010810	> 0.05	0.000	0.003
WNT10A	ENSXMAG00000001801	> 0.05	0.000	0.000
WNT5A	ENSXMAG00000002614	> 0.05	0.000	0.000
WNT9A	ENSXMAG00000005793	> 0.05	0.011	0.008
WNT9B	ENSXMAG00000009198	> 0.05	0.000	0.000
XYLT2	ENSXMAG00000012122	> 0.05	0.008	0.012
YBX3	ENSXMAG00000015755	> 0.05	0.000	0.000
YEATS2	ENSXMAG00000015052	> 0.05	0.043	0.011
YPEL2	ENSXMAG00000010954	> 0.05	0.028	0.006
YRDC	ENSXMAG00000014606	> 0.05	0.001	0.000
YTHDF2	ENSXMAG00000018758	> 0.05	0.034	0.006
ZBED4	ENSXMAG00000019580	> 0.05	0.000	0.001
ZBTB1	ENSXMAG00000020178	> 0.05	0.004	0.010
ZBTB34	ENSXMAG00000000202	> 0.05	0.007	0.018
ZBTB48	ENSXMAG00000012617	> 0.05	0.019	0.015
ZBTB7A	ENSXMAG00000012740	> 0.05	0.000	0.005
ZBTB7B	ENSXMAG00000012652	> 0.05	0.000	0.000
ZC3H12D	ENSXMAG00000005023	> 0.05	0.008	0.003
ZCRB1	ENSXMAG00000009359	> 0.05	0.047	0.037

Table C-2**.** Continued

GeneName	**GeneID**	**BM**	**CM**	**CB**
ZDHHC24	ENSXMAG00000009789	> 0.05	0.028	0.021
ZFP36	ENSXMAG00000018206	> 0.05	0.000	0.002
ZFYVE28	ENSXMAG00000013540	> 0.05	0.001	0.001
ZMYND19	ENSXMAG00000005494	> 0.05	0.049	0.000
ZNF259	ENSXMAG00000019854	> 0.05	0.001	0.000
ZNF598	ENSXMAG00000001469	> 0.05	0.000	0.000
ZNF740	ENSXMAG00000008098	> 0.05	0.001	0.001
ZNF862	ENSXMAG00000006891	> 0.05	0.038	0.000
ZZEF1	ENSXMAG00000005834	> 0.05	0.000	0.008
ATF6	ENSXMAG00000013231	> 0.05	0.000	0.000
BOC	ENSXMAG00000010032	> 0.05	0.007	0.000
CCNT1	ENSXMAG00000009357	> 0.05	0.003	0.000
CECR2	ENSXMAG00000005401	> 0.05	0.000	0.000

Table C-3. Biological process gene ontology results (p<0.01) for M-B. *X. malinche* treated versus *X. birchmanni* treated differentially expressed genes at FDR <0.05.

Process ID	Pvalue	ExpCount	Count	Process name
GO:0042221	6.44E-06	2.69052	19.164	response to chemical stimulus
GO:0050896	6.46E-06	2.59944	42.634	response to stimulus
GO:0065007	1.47E-05	2.76842	55.49	biological regulation
GO:0051716	2.93E-05	2.36117	34.434	cellular response to stimulus
GO:0032501	4.52E-05	2.31008	33.961	multicellular organismal process
GO:0007275	5.80E-05	2.31757	26.09	multicellular organismal development
GO:0080090	6.28E-05	2.28245	29.716	regulation of primary metabolic process
GO:0070887	7.38E-05	2.6307	12.663	cellular response to chemical stimulus
GO:0048731	7.57E-05	2.32472	22.924	system development
GO:0009888	7.80E-05	2.75622	10.564	tissue development
GO:0050794	0.00016	2.26837	50.554	regulation of cellular process
GO:0050789	0.00017	2.2994	52.773	regulation of biological process
GO:0032502	0.00021	2.14662	29.219	developmental process
GO:0048856	0.0003	2.12436	26.09	anatomical structure development
GO:0031323	0.00032	2.08959	30.626	regulation of cellular metabolic process
GO:0060255	0.00035	2.08763	28.916	regulation of macromolecule metabolic process
GO:0007154	0.00036	2.07109	31.693	cell communication
GO:0007610	0.00036	2.98039	6.1737	behavior
GO:0023052	0.00038	2.06854	30.832	signaling
GO:0010033	0.00051	2.29392	14.203	response to organic substance
GO:0006950	0.00057	2.10252	21.602	response to stress
GO:0019222	0.00061	2.00444	33.282	regulation of metabolic process
GO:0048513	0.00095	2.12054	17.175	organ development
GO:0040011	0.00115	2.34995	10.382	locomotion
GO:0009611	0.00164	2.31945	9.9337	response to wounding

Table C-3**.** Continued.

Process ID	Pvalue	ExpCount	Count	Process name
GO:0009719	0.00171	2.35706	9.2423	response to endogenous stimulus
GO:0009653	0.00172	2.08713	15.355	anatomical structure morphogenesis
GO:0007165	0.0018	1.89615	28.249	signal transduction
GO:0048583	0.00223	2.00417	17.26	regulation of response to stimulus
GO:0051239	0.00235	2.10576	13.293	regulation of multicellular organismal process
GO:0006952	0.00246	2.32426	8.8057	defense response
GO:0006979	0.00252	2.69704	5.3732	response to oxidative stress
GO:0009605	0.00252	2.18695	11.05	response to external stimulus
GO:0010035	0.00254	2.59318	6.0403	response to inorganic substance
GO:0009607	0.0026	2.50211	6.7316	response to biotic stimulus
GO:0050877	0.00275	2.24927	9.6305	neurological system process
GO:0003008	0.00303	2.11611	11.983	system process
GO:0071310	0.00364	2.14701	10.625	cellular response to organic substance
GO:0007399	0.00392	2.01133	13.815	nervous system development
GO:0006928	0.00401	2.16502	9.9579	cellular component movement
GO:0051171	0.00415	1.82501	24.003	regulation of nitrogen compound metabolic process
GO:0051246	0.00453	2.00918	13.172	regulation of protein metabolic process
GO:0010468	0.00469	1.81617	23.288	regulation of gene expression
GO:0019219	0.00494	1.80852	23.361	regulation of nucleobase-containing compound metabolic process
GO:0080134	0.00507	2.25256	7.9445	regulation of response to stress
GO:0051707	0.00527	2.37215	6.5497	response to other organism
GO:0007166	0.00583	1.8807	16.714	cell surface receptor signaling pathway
GO:0019226	0.00584	2.27398	7.3381	transmission of nerve impulse

Table C-3**.** Continued.

Process ID	**Pvalue**	**ExpCount**	**Count**	**Process name**
GO:0035637	0.00602	2.26565	7.3623	multicellular organismal signaling
GO:0050890	0.00652	2.58771	4.6575	cognition
GO:0030154	0.00676	1.82647	18.582	cell differentiation
GO:0016477	0.00695	2.17076	8.2114	cell migration
GO:0008283	0.0071	1.94814	12.857	cell proliferation
GO:0008284	0.00751	2.20497	7.5443	positive regulation of cell proliferation
GO:0042127	0.00777	2.01777	10.589	regulation of cell proliferation
GO:0007267	0.00789	2.09151	9.0725	cell-cell signaling
GO:0007268	0.00816	2.24341	6.8893	synaptic transmission
GO:0009889	0.00862	1.73609	23.263	regulation of biosynthetic process
GO:0048518	0.00955	1.70644	25.204	positive regulation of biological process
GO:0010605	0.00964	1.91254	12.396	negative regulation of macromolecule metabolic process

Table C-4. Biological process gene ontology results (p<0.01) for M-C. *X. malinche* treated versus *control* differentially expressed genes at FDR <0.01

Process ID	Pvalue	ExpCount	Count	Process name
GO:0065007	4.80E-05	305.7871	348	biological regulation
GO:0050896	0.000156	234.9381	275	response to stimulus
GO:0050789	0.000356	290.8152	328	regulation of biological process
GO:0032502	0.000419	161.0145	196	developmental process
GO:0051716	0.000577	189.7551	225	cellular response to stimulus
GO:0050794	0.00069	278.5837	314	regulation of cellular process
GO:0070887	0.000751	69.77961	95	cellular response to chemical stimulus
GO:0060255	0.000887	159.3435	192	regulation of macromolecule metabolic process
GO:0010468	0.000914	128.3303	159	regulation of gene expression
GO:0007275	0.001067	143.7701	175	multicellular organismal development
GO:2000112	0.001151	121.5128	151	regulation of cellular macromolecule biosynthetic process
GO:0009889	0.001222	128.1966	158	regulation of biosynthetic process
GO:0031326	0.001281	127.3946	157	regulation of cellular biosynthetic process
GO:0010556	0.001435	123.0501	152	regulation of macromolecule biosynthetic process
GO:0045595	0.001518	52.06735	73	regulation of cell differentiation
GO:0031323	0.001596	168.7677	200	regulation of cellular metabolic process
GO:0019222	0.001693	183.4054	215	regulation of metabolic process
GO:0006355	0.001813	112.5564	140	regulation of transcription, DNA-dependent
GO:0050793	0.001937	65.63561	88	regulation of developmental process
GO:0048513	0.002162	94.64361	120	organ development
GO:2001141	0.002306	113.2248	140	regulation of RNA biosynthetic process
GO:0007165	0.002373	155.6674	185	signal transduction
GO:0051252	0.002429	115.2299	142	regulation of RNA metabolic process
GO:0051239	0.002433	73.25523	96	regulation of multicellular organismal process
GO:0030154	0.003528	102.3969	127	cell differentiation
GO:0080090	0.003631	163.7548	192	regulation of primary metabolic process

Table C-4. Continued.

Process ID	Pvalue	ExpCount	Count	Process name
GO:0009888	0.003646	58.21652	78	tissue development
GO:2000026	0.004064	54.94142	74	regulation of multicellular organismal development
GO:0042221	0.004104	105.6052	130	response to chemical stimulus
GO:0048583	0.005036	95.11148	118	regulation of response to stimulus
GO:0032501	0.006615	187.1484	214	multicellular organismal process
GO:0048731	0.007782	126.3252	150	system development

Table C-5. Biological process gene ontology results (p<0.01) for B-C. *X. birchmanni* treated versus *control* differentially expressed genes at FDR <0.01.

Process ID	Pvalue	ExpCount	Count	Process name
GO:0007275	2.06E-05	151.2639	194	multicellular organismal development
GO:0032502	0.000226	169.4071	207	developmental process
GO:0048513	0.000417	99.57677	130	organ development
GO:0048731	0.00068	132.9097	165	system development
GO:0031327	0.000885	57.03161	80	negative regulation of cellular biosynthetic process
GO:0050896	0.001122	247.1839	282	response to stimulus
GO:0009890	0.001183	57.59419	80	negative regulation of biosynthetic process
GO:0009888	0.001299	61.25097	84	tissue development
GO:2000113	0.00176	54.92194	76	negative regulation of cellular macromolecule biosynthetic process
GO:0032501	0.001878	196.9032	229	multicellular organismal process
GO:0010558	0.002001	56.0471	77	negative regulation of macromolecule biosynthetic process
GO:2000026	0.00202	57.80516	79	regulation of multicellular organismal development
GO:0009892	0.002072	74.61226	98	negative regulation of metabolic process
GO:0031324	0.002111	71.09613	94	negative regulation of cellular metabolic process
GO:0048856	0.002128	151.2639	181	anatomical structure development
GO:0045892	0.002207	51.05419	71	negative regulation of transcription, DNA-dependent
GO:0051716	0.002377	199.6458	231	cellular response to stimulus
GO:0006357	0.002563	63.57161	85	regulation of transcription from RNA polymerase II promoter
GO:0010629	0.002638	54.00774	74	negative regulation of gene expression
GO:0051239	0.002746	77.07355	100	regulation of multicellular organismal process
GO:0051253	0.003589	52.03871	71	negative regulation of RNA metabolic process

Table C-5. Continued.

Process ID	Pvalue	ExpCount	Count	Process name
GO:0045934	0.003935	54.85161	74	negative regulation of nucleobase-containing compound metabolic process
GO:0051172	0.004477	55.1329	74	negative regulation of nitrogen compound metabolic process
GO:0042221	0.005677	111.1097	135	response to chemical stimulus
GO:0050793	0.005795	69.05677	89	regulation of developmental process
GO:0008283	0.005986	74.54194	95	cell proliferation
GO:0010605	0.006064	71.86968	92	negative regulation of macromolecule metabolic process
GO:0007165	0.006832	163.7813	190	signal transduction
GO:0007166	0.006983	96.90452	119	cell surface receptor signaling pathway
GO:0065007	0.00751	321.7258	349	biological regulation

Table C-6. Biological process gene ontology results ($p<0.01$) for 22 genes that are differentially expressed in all three pairwise comparisons. See Table C-2 for gene names.

Process ID	Pvalue	ExpCount	Count	Process name
GO:0070887	5.90E-06	2.963613	12	cellular response to chemical stimulus
GO:0009888	7.42E-06	2.472516	11	tissue development
GO:0051246	0.000364	3.082839	10	regulation of protein metabolic process
GO:0032502	0.000372	6.838452	15	developmental process
GO:0042221	0.000412	4.485161	12	response to chemical stimulus
GO:0080090	0.000457	6.954839	15	regulation of primary metabolic process
GO:0007275	0.000467	6.106065	14	multicellular organismal development
GO:0010033	0.000677	3.324129	10	response to organic substance
GO:0010605	0.001157	2.901161	9	negative regulation of macromolecule metabolic process
GO:0060255	0.00146	6.767484	14	regulation of macromolecule metabolic process
GO:0009892	0.00152	3.011871	9	negative regulation of metabolic process
GO:0019222	0.001759	7.789419	15	regulation of metabolic process
GO:0071310	0.001931	2.48671	8	cellular response to organic substance
GO:0048856	0.001993	6.106065	13	anatomical structure development
GO:0010035	0.002101	1.413677	6	response to inorganic substance
GO:0048519	0.002546	5.436129	12	negative regulation of biological process
GO:0051716	0.002606	8.059097	15	cellular response to stimulus
GO:0031323	0.002707	7.167742	14	regulation of cellular metabolic process
GO:0034097	0.003289	1.544258	6	response to cytokine stimulus
GO:0009719	0.003935	2.163097	7	response to endogenous stimulus
GO:0031324	0.004815	2.869935	8	negative regulation of cellular metabolic process
GO:0008544	0.00547	1.186581	5	epidermis development
GO:0048523	0.005607	5.123871	11	negative regulation of cellular process
GO:0050794	0.006055	11.83174	18	regulation of cellular process

Table C-6. Continued.

Process ID	Pvalue	ExpCount	Count	Process name
GO:0006979	0.006988	1.257548	5	response to oxidative stress
GO:0032501	0.007893	7.948387	14	multicellular organismal process
GO:0048731	0.008118	5.365161	11	system development
GO:0042127	0.008362	2.478194	7	regulation of cell proliferation
GO:0031399	0.008519	2.48671	7	regulation of protein modification process
GO:0009725	0.008568	1.876387	6	response to hormone stimulus
GO:0050896	0.008672	9.978065	16	response to stimulus
GO:0030334	0.009923	1.368258	5	regulation of cell migration

Table C-7. Detection of positive selection by model selection in Codeml, using 5 *Xiphophorus* species. Bolded font have p values <0.05.

OR ID	Codon Length	M8-M7	M8-M8a
V2R_6	641	**0.003155217**	**0.001308771**
V2R_2	116	**0.023974728**	**0.00665671**
V2R_40	895	**0.045077953**	**0.013569711**
V2R_29	864	**0.042785758**	**0.01379486**
V2R_52	859	0.081065809	**0.02539546**
V2R_20	377	0.103994424	**0.034717288**
V2R_28	806	0.101489493	**0.042614738**
V2R_17	341	0.175296755	0.073646196
V2R_61	845	0.174958583	0.077687327
V2R_50	634	0.292533526	0.144509567
V2R_3	221	0.358268706	0.151922371
V2R_30	533	0.376318658	0.170932275
V2R_4	399	0.391230648	0.19772083
V2R_53	928	0.446396816	0.204232225
V2R_59	856	0.358608505	0.205921068
V2R_13	479	0.426898927	0.205997618
V2R_1	186	0.559323646	0.311659508
V2R_45	732	0.608687061	0.329673993
V2R_14	217	0.635317757	0.340844597
V2R_55	325	0.657693014	0.370300454
V2R_71	644	0.62608571	0.37147874
V2R_69	271	0.708698564	0.410009902
V2R_65	855	0.706785538	0.416480874
V2R_46	722	0.715466143	0.44868011
V2R_15	887	0.739516267	0.457039964
V2R_54	528	0.861601855	0.58664647
V2R_7	553	0.87687748	0.655713883
V2R_58	854	1	0.660189075
V2R_44	811	0.926242685	0.695694227
V2R_70	609	0.926645688	0.699243534
V2R_57	856	0.938324914	0.725715557
V2R_18	448	0.930969279	0.747452695
V2R_48	877	0.891577423	0.750084562
V2R_64	813	0.98245376	0.905839636
V2R_22	214	0.994285391	0.916644525
V2R_51	854	0.978754925	0.92164922

Table C-7. Continued.

OR ID	Codon Length	M8-M7	M8-M8a
V2R_11	616	0.999230296	0.96143502
V2R_41	874	1	0.975489186
V2R_16	861	1	0.9864601
V2R_5	179	0.999982	0.992515289
V2R_19	31	1	0.992953366
V2R_8	194	1	1
V2R_21	518	1	1
V2R_23	872	1	1
V2R_26	302	1	1
V2R_38	711	1	1
V2R_39	913	1	1
V2R_42	9	1	1
V2R_43	191	0.999487132	1
V2R_47	550	1	1
V2R_49	642	1	1
V2R_56	856	1	1
V2R_62	552	1	1
V2R_63	301	1	1
V2R_66	860	1	1
V2R_67	847	1	1
V2R_68	504	1	1
OR_49	198	**0.000620485**	**0.00012505**
OR_114	203	0.050569179	**0.016014694**
OR_12	202	0.095620983	**0.032648215**
OR_30	151	0.11972953	**0.039649419**
OR_46	198	0.156902137	0.05557845
OR_17	201	0.160213977	0.065581733
OR_87	196	0.197885242	0.07202919
OR_85	202	0.245964515	0.095539758
OR_73	203	0.227990344	0.096694406
OR_80	202	0.25178317	0.098371812
OR_83	203	0.259074141	0.102189744
OR_42	202	0.352490009	0.148706841
OR_88	196	0.355090355	0.1510737
OR_5	198	0.344151951	0.155828464
OR_82	202	0.361755033	0.18013572
OR_100	200	0.428116896	0.192717751
OR_98	174	0.451397027	0.209835049

Table C-7. Continued.

OR ID	Codon Length	M8-M7	M8-M8a
OR_1	134	0.475935947	0.223002899
OR_24	201	0.530045726	0.259947604
OR_57	206	0.51169732	0.26826952
OR_25	202	0.595882368	0.337768168
OR_112	203	0.665447189	0.366825521
OR_101	201	0.656770261	0.370221173
OR_109	181	0.665974432	0.370670346
OR_9	196	0.744640297	0.442726318
OR_26	202	0.7795675	0.483703459
OR_15	178	0.831842632	0.543973935
OR_21	15	0.85537612	0.576194075
OR_35	195	0.911570813	0.676817893
OR_14	117	0.921332765	0.724472093
OR_93	196	0.943628244	0.73335871
OR_95	196	1	0.772201753
OR_81	202	1	0.888620492
OR_53	23	0.994325163	0.918329813
OR_2	198	0.999985	0.934656412
OR_11	202	1	0.975359745
OR_18	201	1	0.977350989
OR_36	199	0.999803019	0.984611914
OR_51	198	0.999985	0.988058803
OR_106	12	0.999992	0.994472138
OR_23	203	1	0.995486507
OR_50	198	0.999965001	0.996257603
OR_90	28	0.999975	0.996431764
OR_78	202	1	0.998871621
OR_104	201	1	0.998871621
OR_48	198	1	0.998871621
OR_41	204	1	1
OR_43	204	1	1
OR_58	204	1	1
OR_59	204	1	1
OR_60	204	1	1
OR_64	204	1	1
OR_74	204	0.999999	1
OR_76	204	1	1
OR_27	203	1	1

Table C-7. Continued.

OR ID	Codon Length	M8-M7	M8-M8a
OR_33	203	1	1
OR_34	203	1	1
OR_75	203	1	1
OR_103	203	1	1
OR_111	203	1	1
OR_113	203	1	1
OR_68	202	1	1
OR_69	202	1	1
OR_70	202	1	1
OR_71	202	1	1
OR_72	202	1	1
OR_77	202	1	1
OR_79	202	1	1
OR_84	202	0.999974	1
OR_86	202	1	1
OR_16	201	0.999992	1
OR_22	201	1	1
OR_107	201	1	1
OR_108	200	1	1
OR_110	200	1	1
OR_44	198	1	1
OR_45	198	1	1
OR_47	198	0.999736035	1
OR_52	198	0.999973	1
OR_96	197	1	1
OR_97	197	1	1
OR_10	196	1	1
OR_65	196	0.999735035	1
OR_66	196	1	1
OR_67	196	1	1
OR_91	196	1	1
OR_92	196	1	1
OR_94	196	0.999999	1
OR_105	188	1	1
OR_4	178	1	1
OR_3	165	1	1
OR_61	160	1	1
OR_19	151	0.999995	1

Table C-7. Continued.

OR ID	Codon Length	M8-M7	M8-M8a
OR_99	138	1	1
OR_56	118	1	1
OR_31	51	1	1
OR_28	38	1	1
OR_20	23	1	1
OR_7	22	1	1
OR_55	13	1	1
OR_37	8	1	1
OR_39	8	1	1
TAAR_40	329	**0.000341085**	**0.00006494**
TAAR_30	277	**0.008989321**	**0.00214229**
TAAR_46	316	**0.016470572**	**0.005374057**
TAAR_6	279	**0.026458363**	**0.007074759**
TAAR_45	329	**0.039232963**	**0.010932134**
TAAR_4	296	0.085847484	**0.026766527**
TAAR_49	322	0.279240461	0.110460821
TAAR_32	350	0.281022933	0.11233491
TAAR_52	325	0.323683854	0.139131373
TAAR_47	317	0.345148334	0.158431038
TAAR_51	322	0.500892887	0.240696517
TAAR_41	321	0.47516603	0.248999935
TAAR_38	249	0.62997509	0.336385058
TAAR_17	192	0.714768189	0.413422114
TAAR_36	277	0.96262245	0.782528587
TAAR_31	349	0.979682245	0.855177993
TAAR_11	221	1	0.949684414
TAAR_2	236	1	0.977833719
TAAR_20	318	0.99999	0.99680847
TAAR_42	326	0.999133376	0.998045592
TAAR_10	319	1	0.998871621
TAAR_15	5	1	0.998871621
TAAR_1	197	1	1
TAAR_7	279	1	1
TAAR_8	313	1	1
TAAR_9	244	1	1
TAAR_12	316	0.999999	1
TAAR_13	314	1	1
TAAR_18	326	1	1

Table C-7. Continued.

OR ID	Codon Length	M8-M7	M8-M8a
TAAR_19	318	1	1
TAAR_22	314	1	1
TAAR_23	305	1	1
TAAR_24	29	1	1
TAAR_25	311	1	1
TAAR_26	327	1	1
TAAR_33	333	1	1
TAAR_34	322	1	1
TAAR_39	309	1	1
TAAR_43	281	1	1
TAAR_44	326	1	1
TAAR_48	322	1	1
TAAR_50	323	1	1
TAAR_53	322	1	1
TAAR_54	323	1	1
TAAR_55	325	1	1
TAAR_57	219	1	1
TAAR_58	142	1	1
V1R_1	321	0.436005248	0.198456089
V1R_2	313	1	1