PATTERNS OF INTROGRESSION ACROSS CLINAL HYBRID ZONES

A Dissertation

by

GASTON IGNACIO JOFRE RODRIGUEZ

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,	James J. Cai
	Kira E. Delmore
	Kirk O. Winemiller
Head of Department,	Thomas D. McKnight

May 2020

Major Subject: Biology

ABSTRACT

A common goal in studies of natural hybridization is to understand how adaptation and speciation create genome-wide variation over spatial gradients. The speciation process generates genomic signatures that can be detected by tracking introgression along the hybrid zone. Geographic cline analysis is a framework that integrates spatial data to detect genome-wide patterns of divergence. In this dissertation I focus on geographic cline analysis to detect and compare patterns of genomic introgression in two natural independent hybrid zones formed by the freshwater swordtail fish *Xiphophorus malinche- Xiphophorus birchmanni*. Geographic cline analysis is widely applied to compare the cline shapes of multiple genomic regions to reveal signatures of selection. However, I show how hybrid zones in migration/drift disequilibrium can generate the detection of neutral regions as candidate targets of selection and can fail to detect regions under selection. Recent population sizes and migration rate estimates suggest that the *X. malinche- X. birchmanni* swordtail hybrid zones are amenable for detecting patterns of introgression. Using dense spatial sampling of swordtails along two independent hybrid zones, and geographic cline analysis on genome-wide markers of ancestry, I show that both hybrid zones have reduced amounts of genomic introgression, with steep genome-wide clines. The cline centers considerably differ in elevation, suggesting that geographic barriers and intrinsic selection are mainly shaping the hybrid zone structure. Additionally, I show that the centers of both hybrid zones have different hybrid index distributions, with one showing high admixture, and the other one showing population structure. Finally, I show that even with high amounts

of selection against hybrids, some sexually dimorphic traits show transgressive expression while others show reduced introgression. Specifically, the *X. malinche* species specific sexual ornament, the sword extension, shows introgression of the "sword less" phenotype from *X. birchmanni* to *X. malinche* skewed hybrid localities. Together, my dissertation argues how geographic cline analysis is a powerful tool to detect introgression and selection in young hybrid zones, sheds light in how physical barriers and assortative mating can influence hybrid genomic and phenotypic distributions.

DEDICATION

Para mis abuelos Delia Navarro González y Encarnación Rodríguez Hernández.

ACKNOWLEDGEMENTS

I want to thank my advisor Gil Rosenthal, for his encouragement, mentorship, support, and friendship, even before I joined Texas A&M University and I was an undergraduate in Mexico. I want to thank my committee members Drs. James Cai, Kirk Winemiller, and (former member) Adam Jones, to challenge me to become a better researcher with their advice and, specifically, their questions during my candidacy exam. I am grateful with Dr. Kira Delmore, to have agreed to be added to my committee. Her experience, and advice was indispensable for my work. I am grateful with Drs. Ira Greenbaum, Mark Zoran, and Tim Scott, for mentoring me to become an instructor. I want to thank Drs. Charles Criscione for providing constructive ideas, and Dr. Heath Blackmon for his stimulating discussions, sharing his experience with modeling and simulations, and letting me use his computational resources. I am grateful with Dr. Peter Andolfatto, for allowing me to use the resources in his laboratory. I am in debt with Dr. Alisa Sedgifar and Dr. Molly Schumer for tutoring me the molecular protocol for my work, providing me essential knowledge about cline theory, and teaching me computational skills. And, to the Hernández-Perez family, specially Clemente and Osvaldo to help me collect the overwhelming number of specimens for my project.

My life in College Station was significantly marked by multiple people. Thank you Jamie Alfieri, Shishir Basant, Amanda Beckman, Stephen Bovio, Dr. Luke Bower, Alyson Brokaw, Paola Canales, Kristina Chyn, Allison Collins, Dr. Ray Cui, Dr. Zach Culumber, Dr. Pablo Delclós, Kristie DeLaGarza, Anais Dion, Owen Dorsey, Megan Exnicios, Paola Fascinetto, Dr. Chris Figgener, Dr. Sarah Flanagan, Melanie Florkowski,

Mateo García, Pablo Gesundheit, Janelle Goeke, Zach Hancock, Natalie Hamilton, Faith Hardin, Jessica Havens, Marina Hawkins, Dr. Chris Holland, Nick Holloway, Dr. Brad Johnson, Dr. Abby Kimmitt, Emma Lehmberg, Dr. Alejandra Maldonado, Jorge Medina, Jake Mitchell, Maria Mora, Brie Myre, Natalie Olmos, Dr. Dan Powell, Dr. Whitney Preisser, Janae Rapp, Mary Reed, Liliana Robledo, Jordan Rogan, John Rogers, Sherdina Romney, Sarah Ruckman, David Saenz, Dr. Molly Schumer, Dr. CarlyWill Sloan, Victoria Smith, Mattie Squire, Dr. Rhonda Struminger, Oona Takano, Samantha Trent, Dr. Katie Wedemeyer, Jeffery White, Dr. Allison Wilkes, Lilliana Wolf, Dr. Cynthia Wright, Sarah Wuerfele, Baruc Zago, and many others. Your moral support, friendship, and encouragement is reflected in the finalization of my Ph.D.

One of the hardest parts of doing a Ph.D. away from your home country is leaving a life behind. Nevertheless, my hometown friends never actually stopped encouraging me. Muchísimas gracias a ustedes Jorge, Caro, Mafer, Marisol, Jaci (Sunshine), Fernando, Diana, Gaby, Fermo, Andy, Papo, Carlos, Pavel, Yoshi, y Sensei. Los extraño montones, y ahora me les voy aún mas lejos.

The support from my family has been of paramount importance in my career. I would like to thank all my cousins, all my aunts, and uncles, my grandfather Encarnacion, and grandmothers Delia and Emma, who have always shown interest in my biology career, and most importantly, my mother Maria Rodríguez Navarro, and my father Gastón Enrique Jofre Ramírez. Amá, apá, se necesita de mucha dedicación y fortaleza para hacer un doctorado, y ojalá yo tuviera la mitad de la dedicación y fortaleza que ustedes tienen. Muchas gracias por su cariño interminable.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Dr. Gil G. Rosenthal, and Dr. Kira E. Delmore of the Department of Biology, Dr. James J. Cai of the Department of Veterinary Integrative Biosciences, and Dr. Kirk O. Winemiller of the Department of Wildlife and Fisheries Sciences.

The genomic data for Chapter 3 was generated with reagents and laboratory resources from Dr. Peter Andolfatto and Dr. Gil Rosenthal. Dr. Molly Schumer and Dr. Alisa Sedgifar taught me the genotyping pipeline and aided me in the generation of the sequencing libraries. The morphology analysis for chapter 4 was done with help from Marina Hawkins, Amy Chang, Will Ledbetter, and Jefferey White. All other work conducted for the thesis (or) dissertation was completed by the student independently.

Funding Sources

Graduate study was supported by a fellowship from Texas A&M University and a dissertation research fellowship from CONACyT (Consejo Nacional de Ciencia y Tecnología).

This work was also made possible in part by the National Science Foundation under Grants Number 1354172 and 1755327. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Science Foundation.

NOMENCLATURE

BDMI	Bateson- Dobzhansky-Muller Incompatibility
IQR	Interquartile range
FPR	False Positive Rate
FNR	False Negative Rate
hzar	Hybrid Zone Analysis in R
SL	Standard Length
BD	Body Depth
DW	Dorsal fin Width
DL	Dorsal fin Length
SE	Sword Extension
FGS	False Gravid Spot
SC	Spotted Caudal
SUP	Sword Upper Pigment
SLP	Sword Lower Pigment

TABLE OF CONTENTS

Page

ABSTRACT ii

DEDICATION iv

ACKNOWLEDGEMENTS v

CONTRIBUTORS AND FUNDING SOURCES vii

NOMENCLATURE viii

TABLE OF CONTENTS ix

LIST OF FIGURES xi

LIST OF TABLES xiii

1. INTRODUCTION 1

2. OUTLIER ANALYSIS IN GENOMIC STUDIES OF ADMIXED POPULATIONS: A SIMULATION-BASED EVALUATION 5

2.1. Abstract 5
2.2. Introduction 6
2.3. Materials and Methods 8
2.3.1. Simulation and modeling 8
2.3.2. Types of hybrid zones: 10
2.3.3. Patterns of divergence using F_{ST} 12
2.3.4. Geographic cline analysis 12
2.3.5. Outlier detection analysis 13
2.3.6. Confusion matrix and common performance metrics 13
2.4. Results 15
2.4.1. Effects of population size in patterns of divergence. 15
2.4.2. Effect of population size on false positive detection 16
2.4.3. Effect of population size on false negative detection 18
2.5. Discussion 20

3. DETECTING GENOMIC INTROGRESSION IN *XIPHOPHORUS MALINCHE-BIRCHMANNI* NATURAL HYBRID ZONES 26

3.1. Abstract 26
3.2. Introduction 27
3.3. Methods 32
3.3.1. Sample collection 32
3.3.2. Tissue processing and library preparation 33
3.3.3. Ancestry assignment and filtering 34
3.3.4. Cline Analysis 35
3.4. Results 36
3.4.1. Hybrid index changes considerably along the elevation gradient 36
3.4.2. Both hybrid zones have steep clines with different centers 37
3.4.3. Genome-wide cline centers and widths 40
3.4.4. Comparing BDM incompatibilities clines between hybrid zones 41
3.5. Discussion 42

4. DIFFERENTIAL INTROGRESSION OF MALE TRAITS IN *XIPHOPHORUS MALINCHE- BIRCHMANNI* NATURAL HYBRID ZONES 47

4.1. Abstract 47
4.2. Introduction 48
4.3. Methods 50
4.3.1. Sample collection 50
4.3.2. Measurements 50
4.3.3. Principal component analysis on morphology 52
4.3.4. Geographic Cline analysis 52
4.4. Results 53
4.4.1. Quantitative phenotypes along the elevation gradient 53
4.4.2. Principal component analysis on morphology 54
4.4.3. Cline analysis distributions 58
4.5. Discussion 60

5. CONCLUSIONS 65

REFERENCES 67

LIST OF FIGURES

Page

Figure 1.1. Simple Bateson- Dobzhansky-Muller model of hybrid incompatibility. 3

Figure 2.1. The stepping-stone model. Two non-admixed infinite size populations flank 19 admixed demes (i.e. H1, H2, …, H19). At generation 1 each admixed deme has a source population 2 ancestry proportion of *p.* Individuals symmetrically migrate with a migration rate of *m* to adjacent admixed demes, but not to source populations. 9

Figure 2.2. Selection coefficients for hybrid zones with Directional (A and B) and Underdominance (C) selection, and hybrid zones with BDMI epistatic interactions (D). 11

Figure 2.3. Per-locus F_{ST} distribution from 100 loci, after 100 generations of admixture, in different demographic histories. From left to right: Grey - loci under No Selection, Yellow – loci under Directional Selection, Orange – loci under Underdominant Selection, and Blue – pair of loci involved in BDM incompatibility. 14

Figure 2.4. A) False Positive Rate in neutrally admixed hybrid zones with different demographic histories. B) False Negative Rate in hybrid zones with loci under directional and under dominant selection, and with different demographic histories. 17

Figure 2.5.Clines of a loci without selection (grey) and average hybrid index clines (black), from 100 replicates in different demographic histories, and after 100 generations of admixture. 19

Sup. Figure 2.6. Per-locus F_{ST} distribution from 100 loci, after 1000 generations of admixture, in different demographic histories. From left to right: Grey - loci under No Selection, Yellow – loci under Directional Selection, Orange – loci under Underdominant Selection, and Blue – pair of loci involved in BDM incompatibility. 24

Figure 3.1. Map showing sampling locations for each population. Each locality displays individuals with a given *X. malinche* ancestry proportion (https://fatmap.com/). 31

Figure 3.2. Differences in the change in the frequency distribution of hybrid index from high elevation to low elevation between the two hybrid zones. Localities at the center of both clines are enclosed. 38

Figure 3.3. Hybrid index distribution at the center of the clines with representative hybrid male individuals. Upper Rio Huazalingo, lower Rio Pochula. 39

Figure 3.4.Maximum likelihood whole genome hybrid index clines from the Huazalingo hybrid zone (Blue), and Pochula hybrid zone (Yellow), vertical lines indicate the location of the respective cline centers. 39

Figure 3.5.Maximum likelihood distribution of cline centers and widths. Bold vertical and horizontal lines indicate hybrid index cline center and width. Dashed lines indicate the hybrid index cline 95% CI. Black dots indicate loci from the complete data set. Grey dots indicate loci from the BDM incompatibilities dataset. Orange dots, loci that are in the BDM incompatibilities dataset that are width outliers. Triangles and numbers indicate the point at the elevation gradient of collection localities (see Table 3.1). 41

Sup. Figure 3.6. Maximum likelihood of cline centers and widths of loci classified as center outliers (A & B) and width outliers (C & D). Bold vertical and horizontal lines indicate hybrid index cline center and width. Dashed lines indicate the hybrid index cline 95% CI. 46

Figure 4.1.Continuous phenotypes used for morphology analysis: Standard Length (SL), Body Depth (BD), Sword Extension (SE), Dorsal fin Width (DW), Dorsal fin Extension (DE). And discrete phenotypes scored for comparative analysis: False Gravid Spot (FGS), Sword Upper edge Pigment (SUP), Sword Lower edge Pigment (SLP), and Spotted Caudal (SP).....51

Figure 4.2. Distribution of continuous phenotypes in both hybrid zones. Bolded localities represent the center of the hybrid zone. 55

Figure 4.3.Frequency distributions for Standardized BD, DW, and SE, between the localities located at the center of the hybrid zones in Huazalingo (TCP2) and Pochula (TULA). 57

Figure 4.4.Relationship between mean ancestry and Sd. Sword Extension (panel A), and Sd. Body Depth (panel B) at the center of each hybrid zones. Huazalingo (blue dots) and Pochula (yellow dots). 58

Figure 4.5.Comparison between maximum likelihood clines for whole genome hybrid index (black), phenotype frequency of FGS (purple), SC (blue), SUP (green), SLP (yellow), and sword extension (red). See Table 4.2. for the cline parameters. 60

Sup. Figure 4.6. PC1 and PC2 from the Principal Component Analysis, showing the loadings from four standardized measurements. 64

LIST OF TABLES

Page

Table 2.1.Combination of demographic history parameters used in the hybrid zone simulations. 10

Sup. Table 2.2. Welch's *t*-test comparison between F_{ST} values derived from 100 generations and 1000 generations of admixture. (* $p<0.05$, ** $p<0.01$). 25

Table 3.1.Summary of sample locations 33

Table 4.1. The principal component (PC) scores from the morphometric distances 56

Table 4.2. Maximum likelihood cline parameters from the discrete phenotypes analyzed, and sword extension. Bold and underlined values did not overlap with the average genome hybrid cline, and arrows show direction of introgression. Colors match Figure 4.5..... 59

Sup. Table 4.3. Distribution of continuous phenotypes in both hybrid zones. Standard Length SL (mm), Body depth (BD, Sword extension (SE), Dorsal fin width (DW), and Dorsal fin length (DL). Bolded localities represent the center of the hybrid zone. 63

1. INTRODUCTION

It is not surprising that hybrid zones are regarded as natural laboratories for evolutionary studies (Hewitt, 1988). The novel genomic combinations generated by hybridization are tested by natural selection to determine the degree of introgression between species. Introgression, the movement of alleles from one species to another through natural hybridization (Abbott et al., 2013; Harrison & Larson, 2014; Seehausen et al., 2014), is a product of the amount of interspecific gene flow and natural selection (Abbott et al., 2013; Barton & Hewitt, 1985). When two species hybridize, alleles from one species can be selected against and removed from the gene pool (Mallet, 2005), reflecting patterns of reduced introgression. Alleles from one species can be advantageous, be selected for and spread in the hybrid populations (S. H. Martin & Jiggins, 2017), reflecting patterns of increased introgression. However, drift alone can also produce similar patterns of introgression, confounding the effects of selection in natural hybrid zones (Felsenstein, 1975; Slatkin & Maruyama, 1975). Therefore, the detection of increased or reduced introgression across the genome, taking into account the effects of drift, provides considerable information about the ongoing speciation process in natural hybrid zones.

The natural hybrid zones between the northern swordtail species, *Xiphophorus malinche* and *X. birchmanni* are an excellent model to characterize introgression. *X. malinche* and *X. birchmanni* are sister species endemic to the Rio Pánuco drainage of the Sierra Madre Oriental in northern Hidalgo state, México (Rauchenberger, Kallman, & Moritzot, 1990), and are members of the northern Rio Pánuco swordtail clade

(Rauchenberger et al., 1990). *X. malinche* is restricted to head water first order streams at high elevations (Culumber et al., 2011; Culumber, Shepard, Coleman, Rosenthal, & Tobler, 2012). X. *birchmanni* inhabits streams of higher order at lower elevations (Culumber et al., 2011; Culumber et al., 2012). They have formed natural hybrid zones along elevation gradients in at intermediate locations in at least seven rivers (Culumber et al., 2011; Rosenthal et al., 2003).

These hybrid zones are under substantial selection against hybrids genotypes (Powell et al., in press; Schumer & Brandvain, 2016; Schumer et al., 2014). Specifically, Schumer et al. (2014) reported hundreds of Bateson- Dobzhansky-Muller (BDM) incompatibilities in this system (See Figure 1.1). BDM incompatibilities are one of the multiple causes of reproductive barriers that maintain reproductive isolation between species(Seehausen et al., 2014), and multiple studies have found patterns of low introgression in these incompatible regions (Gompert, Lucas, et al., 2012; Hermansen et al., 2014; Rafati et al., 2018; Trier, Hermansen, Saetre, & Bailey, 2014; Walsh, Shriver, Olsen, & Kovach, 2016).

Interestingly, *X. birchmanni* females have reduced preferences for large dorsal fins (Fisher et al., 2009), and large caudal fin extensions (sword) (Wong & Rosenthal, 2006). In *X. malinche* the sword is a fixed in all males, and females lack a strong preference for the sword (Wong & Rosenthal, 2006). Intersexual selection against these traits could generate patterns of low introgression across the hybrid zones in this system.

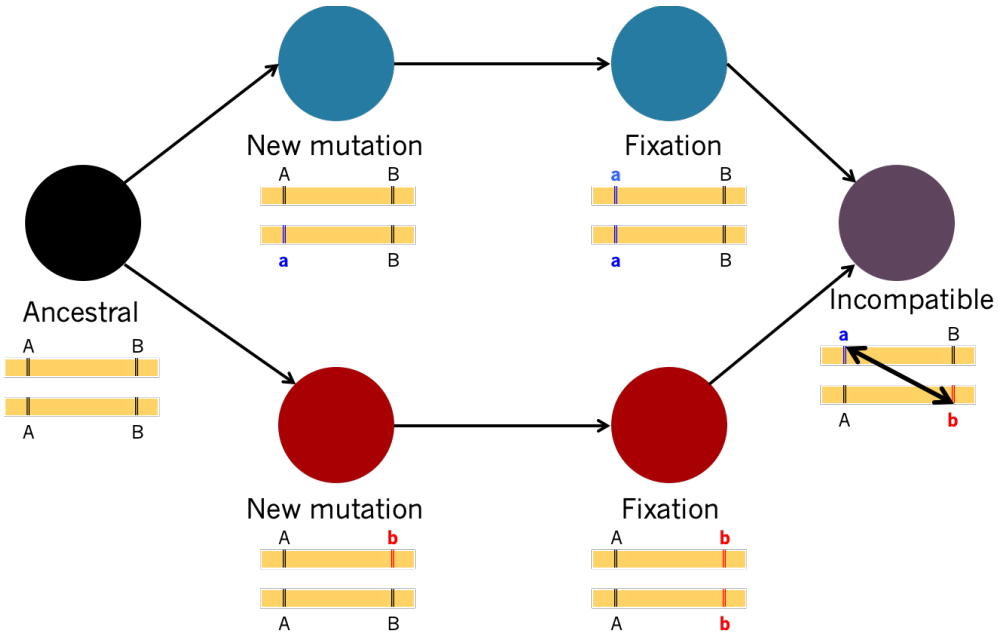


Figure 1.1. Simple Bateson- Dobzhansky-Muller model of hybrid incompatibility.

The detection of genomic regions, and phenotypes, with patterns of introgression that deviate from the genome wide average can be used to identify barriers to hybridization. There are multiple analyses in the literature used to detect these patterns (Gompert, Mandeville, & Buerkle, 2017; Payseur & Rieseberg, 2016). Specifically, Geographic cline analysis has been widely used to detect introgression, by comparing the cline shapes between multiple loci along the genome (Szymura & Barton, 1986), and between multiple traits (Gay, Crochet, Bell, & Lenormand, 2008; Lipshutz et al., 2019). The analysis determines two parameters, the center of the hybrid zone, and the width of the cline, to determine the amount and direction of introgression (Barton & Hewitt, 1985; Szymura & Barton, 1986). Traits and loci with low rates of introgression, show clines with reduced widths (Payseur, 2010; Sciuchetti, Dufresnes, Cavoto, Brelsford, &

Perrin, 2018); and traits and loci that indicate high rate of introgression can have centers shifted towards either end of the cline (Barton & Hewitt, 1985).

Random genetic drift has substantial effects on clines. Specifically, in hybrid zones with small population sizes, and low migration rates, drift distorts the cline shape, reducing the width and shifting the center (Felsenstein, 1975; Polechova & Barton, 2011; Slatkin & Maruyama, 1975). Hybrid zones with these demographic histories can complicate the geographic cline analysis, detecting outliers in genomic regions in which selection is not responsible for differential introgression.

I begin my dissertation by analyzing the effects of drift in the generation of clines with shapes characteristic of regions under selection, using simulated hybrid zones under different demographic histories (Chapter 2). Then, I explore the variation of ancestry at multiple genomic regions along elevational gradients, to detect geographic areas with reduced introgression along two natural hybrid zones between *X. malinche* and *X. birchmanni* (Chapter 3). I finalize comparing the patterns of genomic introgression with phenotypic introgression of sexually dimorphic traits along these two natural hybrid zones (Chapter 4). My dissertation incorporates genomic, phenotypic, and spatial data to assist in the identification of loci involved in reproductive isolation, and traits under sexual selection; emphasizing the importance of demographic history and geographic variation in shaping the distribution of natural hybrid zones.

2. OUTLIER ANALYSIS IN GENOMIC STUDIES OF ADMIXED POPULATIONS: A SIMULATION-BASED EVALUATION

2.1. Abstract

Using genomic data to scan for regions with rates of differentiation and disproportionate amount introgression provides paramount information about natural hybridization and the genetic basis of speciation. Multiple analyses are used in the literature to detect these signatures. Specifically, using geographic cline analysis to compare the cline shapes of multiple genomic regions can reveal signatures of selection. However, hybrid zones with some demographic histories can generate type one and type two errors; and can generate the detection of neutral regions as candidate targets of selection and can fail to detect regions under selection. In this study I aim to evaluate the effects of migration/drift disequilibrium on genetic differentiation, and the rate of false positive generation. I modeled neutrally evolving stepping stone hybrid zones under different demographic histories, and with some genomic regions under selection and with negative epistatic interactions. I found that in young hybrid zones migration/drift disequilibrium increases genomic differentiation in neutral genomic regions, generating patterns similar to genomic regions under directional selection, and underdominance, and involved in epistatic interactions. My simulations show that hybrid zones in migration/drift disequilibrium generate non coincident clines with high discordance in neutral regions. These hybrid zones also show high rates of false positives. Additionally, migration/drift disequilibrium distorts the clines from regions under selection, masking them as neutral, increasing the false negative rate. I discuss

how geographic cline analysis is a powerful framework for studying genome-wide patterns of introgression, but I extreme caution when the model system is under migration/drift disequilibrium.

2.2. Introduction

Hybridization occurs when two genetically distinct populations reproduce and generate viable offspring of mixed ancestry(Abbott et al., 2013; Barton & Hewitt, 1985). The narrow spatial range of overlap between hybridizing populations, i.e. hybrid zone (Barton & Hewitt, 1985), acts as a natural laboratory for evolutionary questions. Within the hybrid zone, existing reproductive barriers can decrease introgression and promote divergence (Mallet, 2005). Alternatively, alleles favored by selection can introgress differently across the hybrid zone (S. H. Martin & Jiggins, 2017). These different types of selection can leave signatures that can be identified using outlier identification methods.

There are multiple analyses in the literature used to detect these signatures (See (Gompert et al., 2017; Payseur & Rieseberg, 2016)). Cline theory studies the change of allele frequencies across the hybrid zone environmental gradient (Endler, 1977; Haldane, 1948), and geographic cline analysis describes and compares the cline shapes to detect loci potentially under selection (Szymura & Barton, 1986). The analysis determines two parameters, the center of the hybrid zone, where the geographic cline is the steepest, and the width of the cline, determined by the dispersal of parental alleles (Barton & Hewitt, 1985; Szymura & Barton, 1986). Neutral alleles are expected to pass easily across the cline (neutral diffusion) and their change of frequencies represents the null expectation

of no selection (Barton & Hewitt, 1985; Gay et al., 2008; Szymura & Barton, 1986). Comparing the centers and the widths of multiple loci with the null expectation of no selection through the geographic gradient, enhances the detection of outliers (Stankowski, Sobel, & Streisfeld, 2017). Specifically, outlier loci that are involved in reproductive isolation between the parental species can have reduced widths (Payseur, 2010; Sciuchetti et al., 2018); and outlier loci that indicate high rate of introgression can have centers shifted towards either end of the cline (Barton & Hewitt, 1985).

The cost reduction and increase in sequence output of next-generation sequencing over the years (Goodwin, McPherson, & McCombie, 2016; van Dijk, Auger, Jaszczyszyn, & Thermes, 2014), keeps revolutionizing the analysis of hybrid zones. Authors have continued to developed thousands of markers for geographic cline analysis using, for example, RADseq (Ryan et al., 2018; Ryan et al., 2017; Stankowski et al., 2017), or GBS (Daniel T. Baldassarre, White, Karubian, & Webster, 2014; Lipshutz et al., 2019; Pulido-Santacruz, Aleixo, & Weir, 2018; Taylor et al., 2014). The breath of genomic data available makes geographic cline analysis a powerful tool to detect introgression and selection in wild hybrid zones. However, hybrid zones with some demographic histories can generate type one and type two errors; i.e. detecting neutral regions as candidate targets of selection and failing to detect regions under selection.

Random genetic drift distorts the cline shape, reducing the width and shifting the center (Felsenstein, 1975; Polechova & Barton, 2011; Slatkin & Maruyama, 1975). (Polechova & Barton, 2011) show that at small deme sizes, random genetic drift causes fixation at local points through the cline, even with high migration rates (m=0.5).

However, deme size was not modified. Low nearest neighbor migration rates but large deme population sizes could still have similar cline shape distortions.

Hybrid zones with these demographic histories can cause the detection of outliers using geographic cline analysis, in which selection is not responsible. In this study I aim to evaluate how often random genetic drift cause neutral genomic regions to have cline shapes characteristic of regions under selection. Specifically, I aim to calculate the rate of false positive detection and the rate of false negative detection in geographic clines using simulations in multiple hybrid zones with different demographic histories. I model a neutrally evolving stepping stone hybrid zone, with some genomic regions under selection and with negative epistatic interactions. Hybrid zones with migration and drift disequilibrium could generate cline shape distortion, and will have high rate of false positives. Additionally, random genetic drift could have high rates of false negatives, causing genomic regions under selection, become cryptic outliers, masked like neutral clines. Finally, even in the presence of selection against a given allele, drift could shift the center of the cline to the side of the gradient where that given allele is selected against.

2.3. Materials and Methods

2.3.1. Simulation and modeling

To model the effects of different natural hybrid zones on the detection of outliers, I used Admix'em (Cui, Schumer, & Rosenthal, 2016) (github: https://github.com/melop/admixem). This program generates realistic simulations of admixed diploid populations (Massey, 2017; Schumer & Brandvain, 2016; Schumer et

al., 2014; Schumer et al., 2018). For all simulations, I modeled admixture in a cline with 19 demes and two infinitely large source populations, each at one extreme of the cline (Figure 2.1). The genomes consisted of ten chromosomes, each with one recombination event per arm per meiosis. I tracked ancestry in 11 markers randomly distributed in each chromosome, 110 markers in total. I allowed random mating for 1000 generations, and sampled in the second, hundredth, and thousandth generation. Generations did not overlap, at the end of each generation parental individuals were removed after mating.

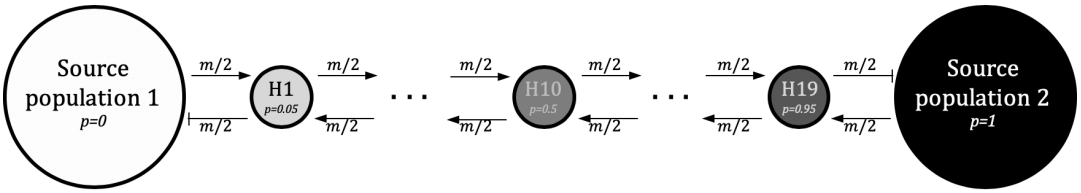


Figure 2.1. The stepping-stone model. Two non-admixed infinite size populations flank 19 admixed demes (i.e. H1, H2, …, H19). At generation 1 each admixed deme has a source population 2 ancestry proportion of *p*. Individuals symmetrically migrate with a migration rate of *m* to adjacent admixed demes, but not to source populations.

I varied the effect of genetic drift by modifying the population size and migration rate. I restricted the population sizes to 100, 500, and 1000 individuals. Migration rate was symmetrical and restricted to 0.1 and 0.01. Recombination rate was set at one recombination event per arm per meiosis. I performed 100 replicated simulations with each parameter combination (see Table 2.1). These combinations were performed for each type of selection with randomized uniform recombination breakpoints.

Table 2.1.Combination of demographic history parameters used in the hybrid zone simulations.

Population size of each deme	100				500				1000			
Migration rate	0.1		0.01		0.1		0.01		0.1		0.01	
Number of generations in admixture	100	1000	100	1000	100	1000	100	1000	100	1000	100	1000
Number of replicates	100	100	100	100	100	100	100	100	100	100	100	100

2.3.2. Types of hybrid zones:

2.3.2.1. A) No selection:

I ran the simulations as described above with the null expectation of no selection with respect to ancestry.

2.3.2.2. B) Directional selection and underdominance selection

To determine the effects of natural selection, I modeled the effects of directional selection (with incomplete dominance) and underdominance. I modeled directional selection on one locus at the middle of chromosomes three, denoted as 3.5 (Figure 2.2, A) and another on chromosome seven, denoted as 7.5 (Figure 2.2, B). I modeled underdominance on loci in the middle of chromosomes five and nine, denoted as 5.5 and 9.5. (see Figure 2.2, C). I assumed additive fitness, where relative fitness is $f = 1 - x_{3.5} - x_{5.5} - x_{7.5} - x_{9.5}$, where x is a function of an individual's genotype at the specific focal loci (see the selection coefficients in Figure 2.2, A-C).

2.3.2.3. Bateson–Dobzhansky–Muller incompatibility

I modeled the effects of epistatic interactions by simulating a Bateson–Dobzhansky–Muller incompatibility (BDMI , see Figure 1.1)(Dobzhansky, 1937; Muller, 1940). I considered two autosomal locus pairs, in the middle of chromosomes three (3.5) and seven (7.5). I assumed codominance in the incompatible alleles and applied additive fitness, where relative fitness is $f = 1 - x_{3.5} - x_{7.5}$ (see selection coefficients in Figure 2.2, D).

A)

	A	a
A	AA s=0.025	Aa s=0.005
a	Aa s=0.005	aa s=0

B)

	A	a
A	AA s=0	Aa s=0.005
a	Aa s=0.005	aa s=0.025

C)

	A	a
A	AA s=0	Aa s=0.025
a	Aa s=0.025	aa s=0

D)

	BB	Bb	bb
AA	AABB s=0	AABb s=0	AAbb s=0.2
Aa	AaBB s=0	AaBb s=0	Aabb s=0
aa	aaBB s=0.2	aaBb s=0	aabb s=0

Figure 2.2. Selection coefficients for hybrid zones with Directional (A and B) and Underdominance (C) selection, and hybrid zones with BDMI epistatic interactions (D).

2.3.3. Patterns of divergence using F_{ST}

To look at the variation in the patterns of divergence along the genome, from each replicate in each hybrid zone, I calculated per-locus F_{ST} following Weir and Cockerman (Weir & Cockerham, 1984). To calculate F_{ST}, I used only the 19 admixed demes and ignored the source populations. I used the original deme population size of each hybrid zone. I used the R package *pegas* (Paradis, 2010) to calculate the per-locus F_{ST} values in all 100 replicates of each hybrid zone. To observe the variation in divergence, I obtained the mean, median, range, and interquartile range of F_{ST} values of 100 neutral loci, 100 loci under directional selection, 100 loci under underdominant selection, and 100 pair of loci in epistatic interaction (BDMI).

2.3.4. Geographic cline analysis

I sampled 30 random individuals from each deme in generations 100, and 1000. Admix'em generates an output file with ancestries from the defined markers in the input files. I then used custom scripts to generate the input files necessary for the geographic cline analysis. I used the Metropolis–Hastings algorithm in the R package *hzar* (Derryberry, Derryberry, Maley, & Brumfield, 2014), to fit clines to allele frequencies for all the 110 markers in each generation. To fit a cline, I used the following five parameters in hzar: cline center $\mathcal{y}$, cline width $\mathcal{w}$, the ends of the cline P_{min} and P_{max} , the rate in which the cline tail decays $\theta_{P1}\ and\ \theta_{P2}$, and the size of the cline tails $\mathcal{B}_{P1}\ and\ \mathcal{B}_{P2}$. For computational speed, I fitted three models. Model I estimated only center $\mathcal{y}$ and width $\mathcal{w}$ from the data, assumed no tails (θ and $\mathcal{B}$ are one and zero respectively), and included fixed ends (P_{min} and P_{max} fixed to zero and one). Model II

estimated $\mathcal{y}$, $\mathcal{w}$, and P_{min} and P_{max} from the data and assumed no tails. Model III was the same as Model II, but with tail estimates θ and $\mathcal{B}$ allowed to vary. Each model parameter was estimated using three independent chain runs using 500,000 MCMC steps after a burn-in of 100,000 steps. The model with the lowest Akaike criterion was chosen for each marker.

2.3.5. Outlier detection analysis

For each replicate in each hybrid zone, I obtained the cline center $\mathcal{y}$, cline width $\mathcal{w}$, P_{min} and P_{max} , and the 95% confidence intervals for all loci, and the whole genome hybrid index. Centers and widths of any two loci were considered coincident (same $\mathcal{y}$) and concordant (same $\mathcal{w}$) if their parameters overlapped with each other's 95% confidence intervals. Loci were considered outliers if they did not overlap with the 95% confidence intervals from the whole genome hybrid index cline $\mathcal{y}$ and $\mathcal{w}$.

2.3.6. Confusion matrix and common performance metrics

I generated an error matrix for each hybrid zone. I counted the amount of observed null and outlier loci in each replicate. I obtained the average number of false null and false outlier loci, and true nulls and true outlier loci. I calculated the average false positive rate (FPR) from 100 replicates of all hybrid zones, and the average false negative rate (FNR) from 100 replicates of hybrid zones under directional and underdominant selection, and epistatic interactions.

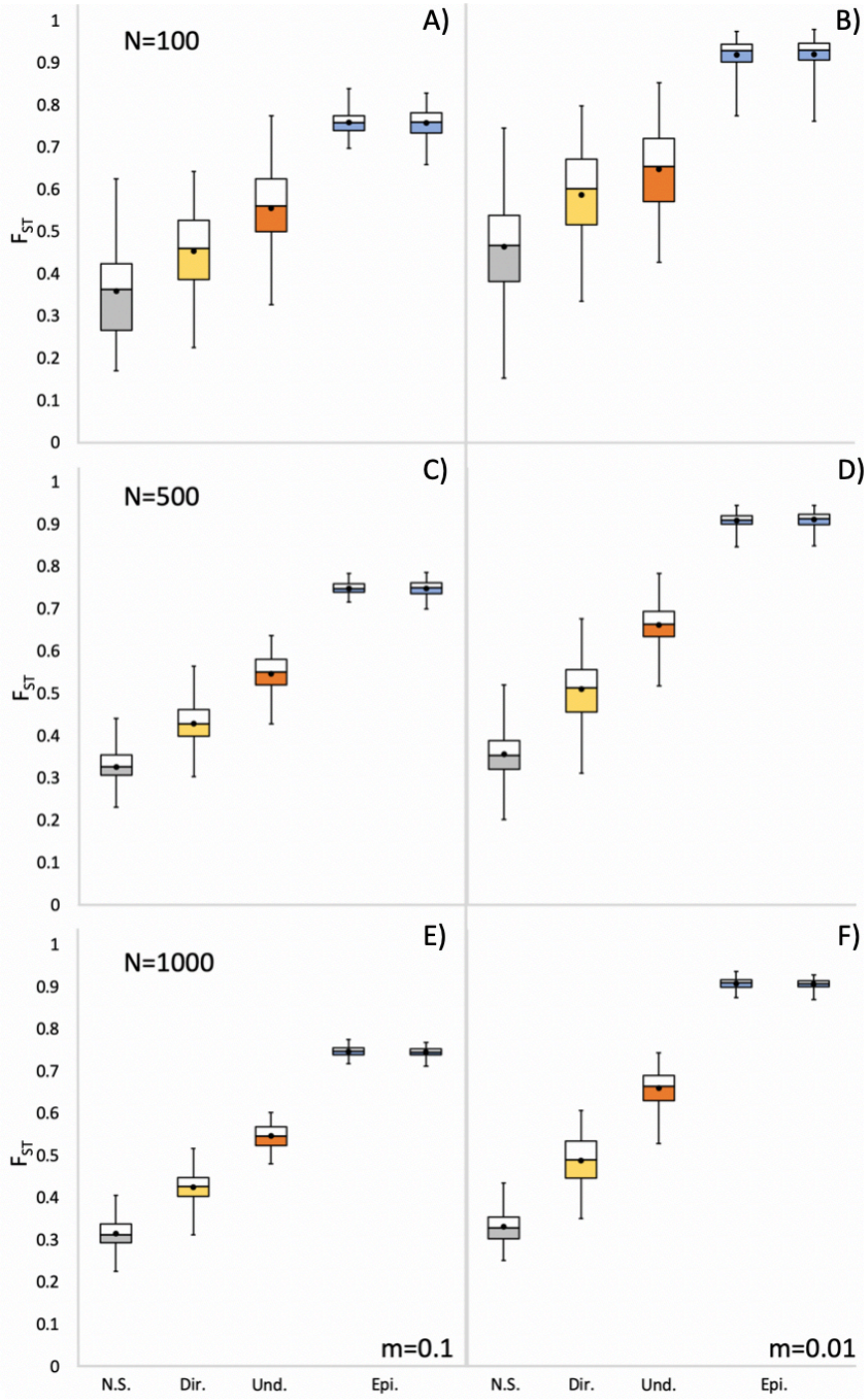


Figure 2.3. Per-locus F_{ST} distribution from 100 loci, after 100 generations of admixture, in different demographic histories. From left to right: Grey - loci under No Selection, Yellow – loci under Directional Selection, Orange – loci under Underdominant Selection, and Blue – pair of loci involved in BDM incompatibility.

2.4. Results

2.4.1. Effects of population size in patterns of divergence.

To determine the effects of population size on selected loci divergence, I calculated per-locus F_{ST} values from 100 replicates, after 100, and 1000 generations of admixture in different demographic histories. The variation in F_{ST} at generation 100 is shown in Figure 2.3. Small population sizes generated the highest range of F_{ST} values and having low migration rate (m=0.01) increased it further. In hybrid zones with no selection, F_{ST} had the highest value when N= 100 and m=0.01 (mean = 0.467, median=0.466, range 0.814, IQR = 0.15), and showed more population homogenization when N=1000 and m=0.1 (mean = 0.316, median=0.316, range 0.23, IQR = 0.42).

When N=100, the IQR of loci with directional selection overlapped with the IQR of loci from the hybrid zones with no selection. The IQR of loci with underdominant selection, and epistatic interaction, do not overlap with the IQR from the hybrid zones with no selection when N=100, 500, or 1000. The IQR of loci under directional selection, overlaps with the range of loci from the hybrid zones with no selection in all population sizes, and migration rates. The IQR of loci under underdominant selection overlaps with the range of loci from the hybrid zones with no selection only when N=100.

Thousand generations of admixture intensified genomic differentiation of neutral regions in hybrid zones with low population sizes and with low migration rate (Sup. Figure 2.6, B,D, and F). Reduced migration drives regions under directional selection to fixation in most demes along the gradient, homogenizing the population and generating

low F_{ST} values (Sup. Figure 2.6, B,D and F). Interestingly this pattern is present in few regions with underdominant selection (Sup. Figure 2.6, B). With increasing population sizes, but only with higher migration rate, neutral regions became more homogenized and regions under selection showed increased differentiation.

I used Welch's t-test to compare the FST distribution from 100 generations with 1000 generations of admixture. Neutral loci, in most hybrid zones with m=0.01 showed a significantly different FST distributions after 1000 generations of admixture (see Sup. Table 2.2). On the contrary, the variation of FST distribution of neutral loci in hybrid zones with m=0.1, after 1000 generations, showed mostly similar distributions (see Sup. Table 2.2), suggesting that hybrid zones under these demographic histories (specifically, N=500, and N=1000) are reaching an equilibrium between migration and drift.

2.4.2. Effect of population size on false positive detection

I first determined the number of false positives in a neutrally admixed hybrid zone. Migration/drift disequilibrium increased rates of introgression in neutral alleles (Figure 2.5) and inflates the false positive rate (FPR, Figure 2.4). Specifically, I found that population size has the most drastic effect on the generation of false positives (see Figure 2.4, A; Figure 2.5, A & B). After 100 generations of admixture, and a migration rate of *m*=0.1, hybrid zones with a population size of N=100 had the highest FPR, compared to hybrid zones with N=500, and N=1000. FPR increased significantly in hybrid zones with low migration rate, *m*=0.01. Hybrid zones with N=500 and N= 1000 and m=0.1 showed the most coincident clines and lowest FPR. However, after 1000

generations of admixture, FPR significantly increased in all hybrid zones showing the maximum value with low migration rates.

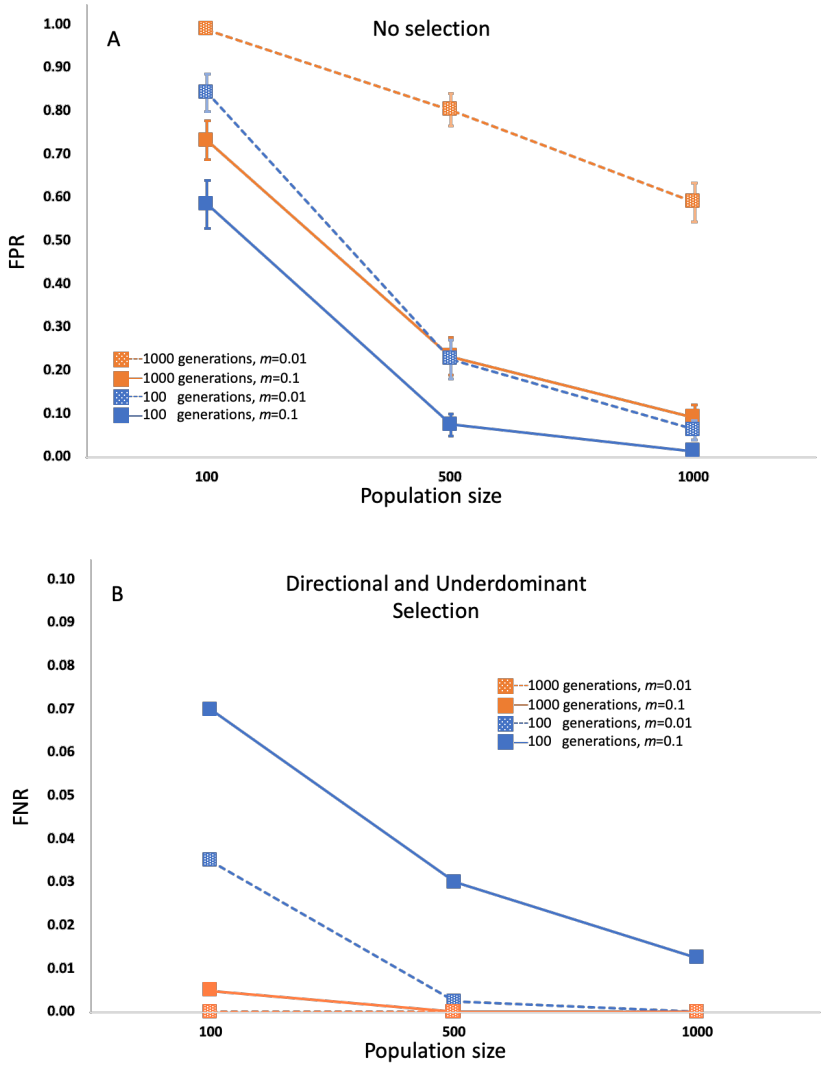


Figure 2.4. A) False Positive Rate in neutrally admixed hybrid zones with different demographic histories. B) False Negative Rate in hybrid zones with loci under directional and under dominant selection, and with different demographic histories.

I then determined the number of false positives in hybrid zones with directional and underdominant selection, and with BDM incompatibilities. I found that hybrid zones under selection, and with epistatic interactions, showed a similar FPR distribution as neutrally admixed hybrid zones.

2.4.3. Effect of population size on false negative detection

I then determined the number of false negatives in hybrid zones under directional and underdominant selection and with BDM incompatibilities. I measured the effect of population size and migration rate on the false negative rate (FNR). In hybrid zones under directional and underdominant selection showed a general low FNR. High values of FNR were found in hybrid zones high migration rates (m=0.1) (see Figure 2.4, B). FNR decreased with higher population sizes. Having a low migration rate reduced the FNR, and after 1000 generations FNR became zero. Hybrid zones with BDM incompatibilities, showed an FNR of zero in all circumstances.

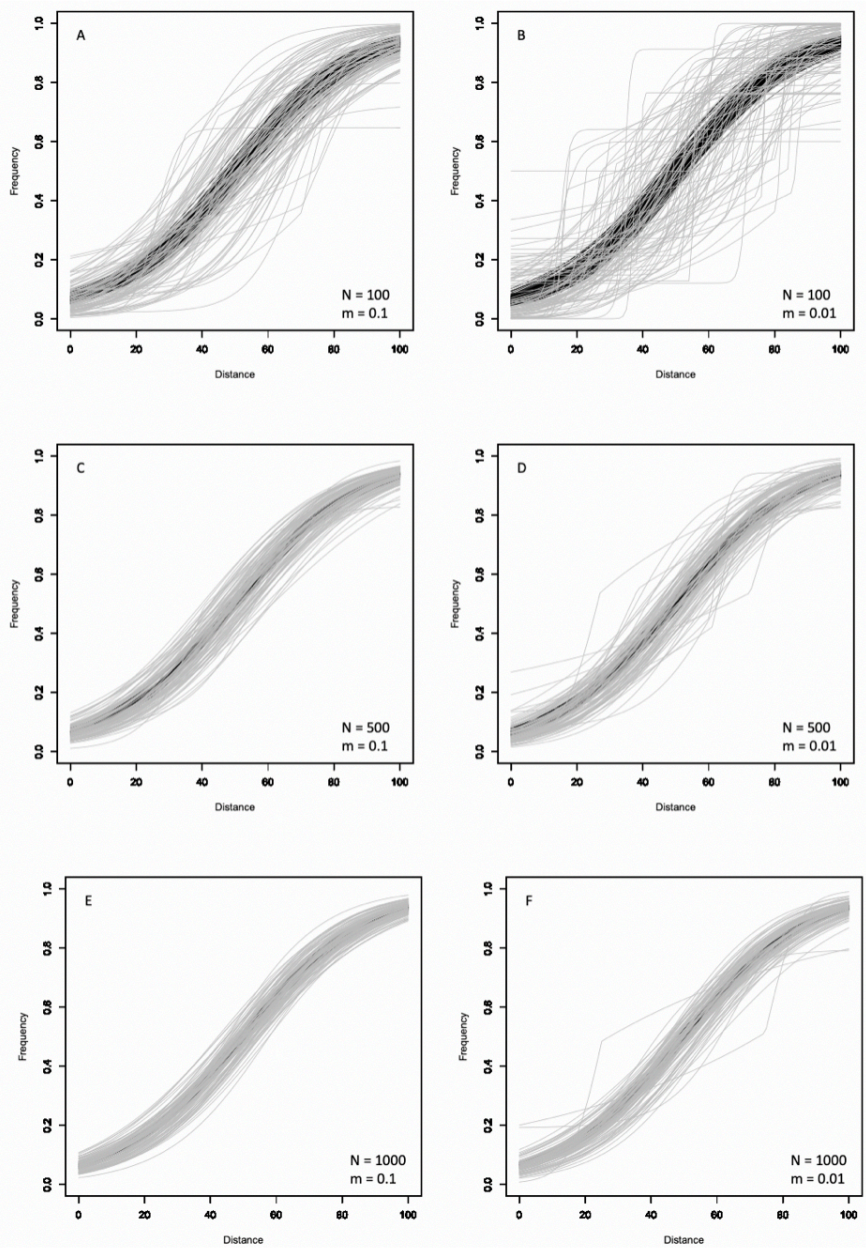


Figure 2.5.Clines of a loci without selection (grey) and average hybrid index clines (black), from 100 replicates in different demographic histories, and after 100 generations of admixture.

2.5. Discussion

Outlier analysis rely on the detection of substantial differences of allele frequencies across the genome, to detect regions that deviate from a neutrality (Seehausen et al., 2014), and that can potentially be under selection (e.g. F_{ST} scans (Beaumont & Balding, 2004), genomic clines (Fitzpatrick, 2013; Gompert & Buerkle, 2011), and geographic clines(Barton & Hewitt, 1985; Stankowski et al., 2017)). Specifically, geographic cline analysis has been widely used to detect the genetic differences between hybridizing taxa (e.g. McKenzie, Dhillon, & Schulte, 2015; Rosser, Dasmahapatra, & Mallet, 2014; Stankowski et al., 2017; Wang et al., 2011), by measuring levels of introgression along a spatial gradient (Barton & Hewitt, 1985; Gay et al., 2008). By comparing introgression of multiple genomic regions, and with the whole genomic ancestry, one can make inferences of how selection is acting on them. In my study I aim to compare patterns of introgression when hybridizing populations are in migration/drift disequilibrium. My results show that migration/drift disequilibrium can generate strong patterns of reduced and increased introgression, even in the absence of selection, that could be taken as candidates of selection. I extend the caution when inferring targets of selection in hybrid zones with small population sizes, and restricted migration, since migration/drift disequilibrium can generate spurious patterns, or even obscure the effects of selection.

My per-locus F_{ST} comparisons show that in young hybrid zones migration/drift disequilibrium increases genomic differentiation in neutral genomic regions. On the most drastic scenario, with low population sizes and low migration rate, neutral genomic regions can generate higher F_{ST} values than genomic regions under directional selection,

and underdominance, and involved in epistatic interactions. Increasing migration rate does not substantially decrease differentiation in neutral genomic regions, and some neutral regions in small hybrid zones could still be detected as outliers. Interestingly, regions under directional and underdominant selection could show signs of homogenization, having reduced F_{ST} values similar to neutral regions, having a risk of not being detected in F_{ST} scans. This pattern is more drastic after 1000 generations, since migration/drift disequilibrium generates fixation (Sup. Figure 2.6, B, C & F). Hybrid zones with larger population sizes show more population homogenization, and with increased migration rate these hybrid zones showed patterns of migration/drift equilibrium. However, some neutral genomic regions can still show signatures of selection even in large population sizes.

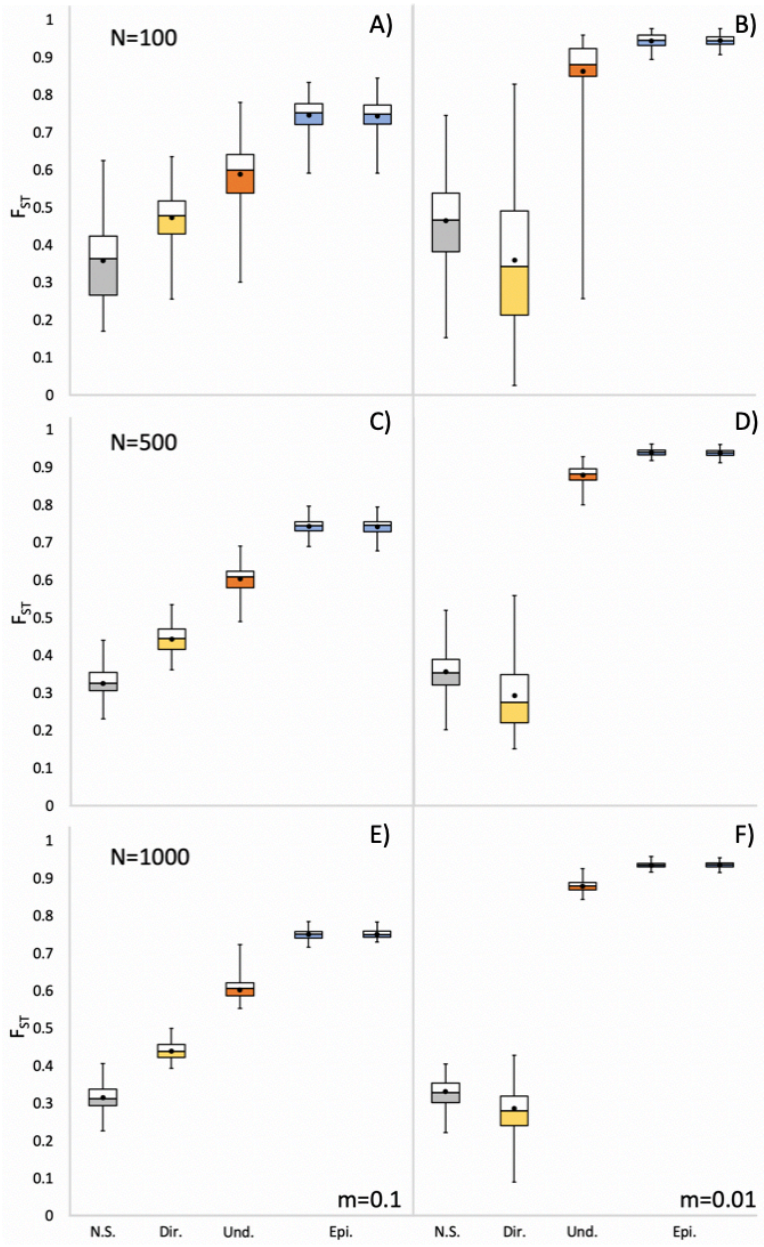
My simulations show that hybrid zones in migration/drift disequilibrium show the highest rates of false positives. The high levels of differentiation in neutral regions generate clines with reduced width, and the local risk of fixation, caused by reduced migration between demes, shifts the centers away from the expected hybrid index cline (see Figure 2.5). Increasing the population size does decrease migration/drift disequilibrium, reducing significantly the rate of false positives, and increasing cline coincidence. However, large population sizes with low migration rate, even with big population sizes, have a risk of generating false positives. Additionally, after a thousand generations of admixture, even with high population sizes, hybrid zones in migration/drift disequilibrium have multiple genomic regions fixed along the cline, causing the significant increase of false positives.

As noted above, drift alone distorts cline shapes, which has also been documented before (Polechova & Barton, 2011). Additionally, my results show that migration/drift disequilibrium can obscure genomic regions under selection. Even though the false negative rates were generally low, hybrid zones with small population sizes have higher risk of masking outliers as neutral. Additionally, the distortion caused by drift can also generate clines with parameters characteristic of a different type of selection. For example, with underdominance selection some clines present drastic shifted centers, which could be wrongly considered as being under directional selection. This shift in centers was also present in genomic regions in negative epistatic interactions, to a lesser degree, however. Additionally, in some minor instances, genomic regions under directional selection, instead of having centers shifted towards the adaptive side of the gradient, showed centers shifted to opposite (less fit) side of the gradient.

Knowing the demographic history of the model organism a priori can reduce the risk of misidentifying neutral regions as candidate regions under selection in natural populations. If the natural hybrid zone is suspected to be in drift-migration disequilibrium, one option is to ignore outlier loci with cline parameters close to the confidence interval. By only considering strong outliers, the probability of misidentification can decrease. Additionally, using multiple replicated hybrid zones showing the same strong outlier loci can reduce the risk of misidentification.

Selection can generate signatures that can be detected by additional methods, For example genomic clines (Gompert & Buerkle, 2011), ancestry track lengths (Sedghifar, Brandvain, & Ralph, 2016), and ancestry junctions (Hvala, Frayer, & Payseur, 2018).

Indeed Drift-migration disequilibrium could also generate false positives, (e.g. increasing ancestry track sizes, reduce junction densities, and generating excess ancestry). Geographic cline analysis still is a powerful framework for studying genome-wide patterns of introgression. When facing a hybrid zone in migration/drift disequilibrium, using multiple methods in conjunction with geographic cline analysis, and over multiple independent transects, can reduce the risk of misidentifying candidate targets of selection.



Sup. Figure 2.6. Per-locus F_{ST} distribution from 100 loci, after 1000 generations of admixture, in different demographic histories. From left to right: Grey - loci under No Selection, Yellow – loci under Directional Selection, Orange – loci under Underdominant Selection, and Blue – pair of loci involved in BDM incompatibility.

Sup. Table 2.2. Welch's *t*-test comparison between F_{ST} values derived from 100 generations and 1000 generations of admixture. (* p<0.05, ** p<0.01).

N	Simulation	Migration rate	t	df	p value
100	neutral	m=0.1	-2.0587	197.98	0.04083*
	directional	m=0.1	-1.5096	187.81	0.1328
	underdominance	m=0.1	-2.5435	197.75	0.01174*
	Epistasis 1	m=0.1	2.2844	173.3	0.02356*
	Epistasis 2	m=0.1	2.2486	191.04	0.02568*
	neutral	m=0.01	-5.3157	167.64	3.35E-07**
	directional	m=0.01	9.9398	155.18	2.20E-16**
	underdominance	m=0.01	-14.952	197.64	2.20E-16**
	Epistasis 1	m=0.01	-5.8967	144.62	2.51E-08**
	Epistasis 2	m=0.01	-5.6153	127.24	1.18E-07**
500	neutral	m=0.1	0.37091	188.43	0.7111
	directional	m=0.1	-2.384	180.74	0.01816*
	underdominance	m=0.1	-10.294	191.67	2.20E-16**
	Epistasis 1	m=0.1	2.5393	186.52	0.01193*
	Epistasis 2	m=0.1	2.4228	195.9	0.01631*
	neutral	m=0.01	-1.7889	130.43	0.07594
	directional	m=0.01	18.649	193.08	2.20E-16**
	underdominance	m=0.01	-34.342	133.79	2.20E-16**
	Epistasis 1	m=0.01	-16.048	154.77	2.20E-16**
	Epistasis 2	m=0.01	-14.131	153.51	2.20E-16**
1000	neutral	m=0.1	-0.010541	197.92	0.9916
	directional	m=0.1	-3.103	158.64	0.002268**
	underdominance	m=0.1	-14.153	197.99	2.20E-16**
	Epistasis 1	m=0.1	-1.8673	187.2	0.06342
	Epistasis 2	m=0.1	-3.153	197.17	0.001868**
	neutral	m=0.01	-2.3767	129.45	0.01893*
	directional	m=0.01	22.182	192.74	2.20E-16**
	underdominance	m=0.01	-48.291	122.87	2.20E-16**
	Epistasis 1	m=0.01	-18.738	161.78	2.20E-16**
	Epistasis 2	m=0.01	-20.702	165.43	2.20E-16**

3. DETECTING GENOMIC INTROGRESSION IN *XIPHOPHORUS MALINCHE- BIRCHMANNI* NATURAL HYBRID ZONES

3.1. Abstract

Natural hybrid zones, in which two divergent species come into secondary contact and interbreed, are maintained in a balance between selection and the dispersal of parental individuals. Two freshwater fish species, the swordtails *Xiphophorus malinche* and *X. birchmanni,* have formed independent hybrid zones along elevation gradients in distinct stream reaches of the Rio Pánuco drainage of the Sierra Madre Oriental. I conducted dense spatial sampling of swordtails along two independent hybrid zones (12 localities along 40 km in each stream). I performed geographic cline analysis on genome-wide markers of ancestry at a resolution of ~10 kb. Both hybrid zones showed steep, genome-wide clines, but cline centers differed considerably in elevation (788.12 m vs 424.12 m). These data suggest that geographic barriers and intrinsic selection play a primary role in shaping hybrid zone structure. Near the cline center, one hybrid zone showed a unimodal distribution of ancestry genotypes, while the other showed bimodal structure consistent with strong premating barriers. Consistent with intrinsic selection, clines for previously identified genetic incompatibility loci overwhelmingly fell into two categories: 87% were steep clines coincident with geographic barriers (high density of riffles and cascades), while 12% were shallow, as expected when incompatibilities are resolved. Outlier steep clines co-localize with physical barriers at elevations higher than the whole genome cline centers, suggesting that these genomic regions are under strong selection against hybrid genotypes, restrict introgression and isolate upstream localities.

3.2. Introduction

Hybridization, or the viable reproduction between members of two genetically distinct populations (Barton & Hewitt, 1985), creates a "genomic collision" (Buerkle & Lexer, 2008), generating individuals with mixed ancestry and a combination of phenotypes from both parental species (Abbott et al., 2013). Hybridization is ubiquitous in nature (Payseur & Rieseberg, 2016), and with the natural combination of genomes, hybrid zones are regarded as natural laboratories to understand speciation (Hewitt, 1988). Two main goals of hybrid zones studies are to characterize the degree and direction of introgression and detect barriers to gene flow between hybridizing populations (Payseur & Rieseberg, 2016).

In hybrid populations patterns of introgression are expected to be heterogeneous across their genome. When a small number of loci are under divergent selection, e.g. reproductive barriers, these genomic regions, including linked loci, are likely to show low levels of gene flow (Mallet, 2005). Distantly linked and unlinked genomic regions (neutral loci) are able to permeate through the hybrid zone. In this case, genomic patterns of introgression will be less heterogeneous, and neutral diffusion will be high along the genome. With an increased number of loci under divergence selection, or even few loci with larger effects (Gompert, Parchman, & Buerkle, 2012), patterns of introgression become more heterogeneous (Gompert, Parchman, et al., 2012; Ryan et al., 2017), restricting the amount of genomic regions to permeate through the hybrid zone.

The detection of genomic regions with patterns of introgression that deviate from the genome wide average can be used to identify barriers to hybridization. Multiple

situations that can cause differential introgression at a given locus. Genomic areas with low recombination, such as centromeres, inversions and sex chromosomes, often have high levels of divergence (S. H. Martin & Jiggins, 2017), since introgressed variants can be rapidly removed by natural selection (S. H. Martin, Davey, Salazar, & Jiggins, 2019). Intrinsic selection also can cause loci to have low rates of introgression. Specifically, underdominant selection and negative epistatic interactions can also reduce the amount of introgression (Gompert, Parchman, et al., 2012).

Recombination between hybridizing genomes generates negative epistatic interactions, e.g. BDM incompatibilities (Dobzhansky, 1937; Muller, 1940; Figure 1.1), between genes with different evolutionary history (Coyne & Orr, 2004). Reduced fitness happens in a subset of hybrid genotypes, depending on the combination of alleles at two interactive genes. BDM incompatibilities are one of the multiple causes of reproductive barriers that maintain reproductive isolation between species (Seehausen et al., 2014), and multiple studies have found low introgression in these incompatible regions (Gompert, Lucas, et al., 2012; Hermansen et al., 2014; Rafati et al., 2018; Trier et al., 2014; Walsh et al., 2016). Additionally, genomic regions that showed high level high levels of divergence, and had low introgression, co-localized at similar points in the geographic gradient of the hybrid zones.

Two mathematical models aim to explain the variation in the level of divergence across the genome, and the shape and locations of hybrid zones. The Tension Zone model (Barton & Hewitt, 1985, 1989) assumes that the main barriers to gene flow are caused by interactions within the genome. Hybrid zones are maintained in balance

between the degree of intrinsic selection (e.g. the situations mentioned above) and the dispersal of parental individuals (Arnold, 1997; Barton & Hewitt, 1985). Strong selection and weak dispersal cause narrow clinal hybrid zones, and their location can move in space (Barton, 1979a; Buggs, 2007). The geographic selection gradient model (GSGM, Endler, 1977) implies that selection in hybrids is mainly exogenous, and environmental variables are the main factor restricting introgression and causing divergence in hybridizing species (Endler, 1977; Slatkin, 1973). The location and extension of ecotones determines the location of hybrid zones (Moore & Price, 1993). The ecological variables decrease hybrid fitness ecotones can cause similar narrow clinal hybrid zones as in tension zones. In most hybrid zones however, the hybrid genome experiences a mix endogenous and exogenous selection, decreasing hybrid fitness, inhibiting introgression, and maintaining species boundaries (Arnold, 1997; Ryan et al., 2017).

Exploring the variation in ancestry at multiple genomic regions along a geographic gradient is often used to detect differential introgression in hybrid zones (see Harrison and Larson (2014); Payseur and Rieseberg (2016) for examples). Specifically, geographic cline analysis plots the gradual change of allele frequencies, genotypes, or phenotypes, from one parental species to another, over a spatial gradient modeling a sigmoid function (Barton & Hewitt, 1985; Szymura & Barton, 1986). The analysis assumes that the distribution of parental species and hybrid individuals follow a one dimensional transect, with parental species at each extreme, and hybrids genotypes gradually changing in the middle (Gay et al., 2008; Stankowski et al., 2017; Szymura &

Barton, 1986). To fit a sigmoid cline the analysis uses four parameters, cline center, where the geographic cline is the steepest, and cline width, determined by the amount of dispersal of parental alleles, and two extreme tails that describe the frequency of allele on the parental species ends of the cline (Gay et al., 2008; Stankowski et al., 2017; Szymura & Barton, 1986).

Comparing the centers and the widths of multiple loci, along with the average genome ancestry through the geographic gradient, allows the detection of regions with differential introgression (Stankowski et al., 2017). Genomic regions involved in BDM incompatibilities are expected to be independent from environmental transitions (Barton, 1979b), and asymmetries in fitness from the allelic combination of interactive genes, generates cline coincidence at different locations (Abbott et al., 2013). Due to the effect of reduced hybrid fitness, BDM incompatibilities are expected to generate steep clines with reduced widths (Carling & Brumfield, 2008; Sciuchetti et al., 2018).

Two northern swordtail species, *Xiphophorus malinche* and *X. birchmanni*, have formed at least seven natural hybrid zones along elevation gradients (Culumber et al., 2011; Rosenthal et al., 2003) in the Rio Pánuco drainage of the Sierra Madre Oriental in northern Hidalgo state, México (Rauchenberger et al., 1990). *X. malinche* and *X. birchmanni* are members of the Northern Rio Pánuco swordtails clade (Rauchenberger et al., 1990). *X. malinche* is restricted to head water first order streams at high elevations, ranging from 658 m to 1499 m (Culumber et al., 2011; Culumber et al., 2012). *X. birchmanni* inhabits streams of higher order at lower elevations, ranging from 146 m to 556 m (Culumber et al., 2011; Culumber et al., 2012). Hybrids individuals inhabit

intermediate locations, registered from 173 m to 1168 m (Culumber et al., 2011; Culumber et al., 2012) (See Figure 3.1).

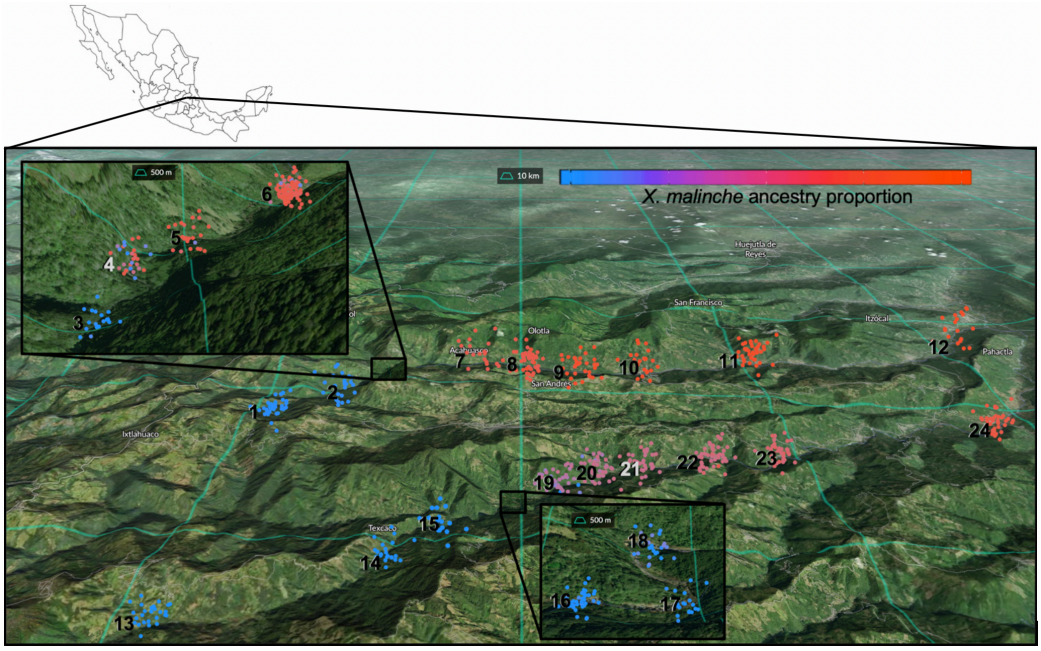


Figure 3.1. Map showing sampling locations for each population. Each locality displays individuals with a given *X. malinche* ancestry proportion (https://fatmap.com/).

River networks are complex dendritic landscapes in which elevation creates unidirectional migration (Brown & Swan, 2010; E. H. C. Grant, Lowe, & Fagan, 2007; Morrissey & de Kerckhove, 2009). At high altitudes in first stream orders physical factors reinforce this asymmetry (Brown & Swan, 2010; Limdi, Perez-Escudero, Li, & Gore, 2018; Vannote, Minshall, Cummins, Sedell, & Cushing, 1980); for example, riffles and cascades. The linear pattern of migration has an impact on the genetic structure of populations along river networks (Morrissey & de Kerckhove, 2009), creating a reduction in heterozygosity, and allelic diversity in headwater streams (Crispo,

Bentzen, Reznick, Kinnison, & Hendry, 2006; Hänfling & Weetman, 2006). The asymmetric migration of the *Xiphophorus malinche* and *X. birchmanni* hybrid system in these river networks has formed multiple elevational clines (Culumber et al., 2011). In this Chapter I aim to describe the differential patterns of genomic introgression in two independent hybrid zones between *X. malinche* and *X. birchmanni* following an altitudinal gradient. The genus *Xiphophorus* is an established model in evolutionary genetics (Cui et al., 2013; Jones, Fan, Franchini, Schartl, & Meyer, 2013; Schartl et al., 2013; Schumer et al., 2013; Schumer, Cui, Powell, Rosenthal, & Andolfatto, 2016), with substantial work focused on this hybrid system (Culumber et al., 2011; Culumber, Ochoa, & Rosenthal, 2014; Culumber et al., 2012; Powell et al., in press; Schumer et al., 2014; Schumer et al., 2017; Schumer et al., 2018). Specifically, I aim to use geographic clines analysis to characterize the amount of introgression of regions involved in hybrid incompatibilities and detect genomic regions with outstanding patterns of introgression.

3.3. Methods

3.3.1. Sample collection

A total of 788 fish were collected during 2013 to 2017, from 12 localities along altitudinal gradients in the Rio Huazalingo (N= 410) and Rio Pochula (N=378), in the state of Hidalgo, Mexico (see Table 3.1). In each locality specimens were lightly anesthetized using tricane methansesulfate. I then photographed each specimen on both sides using a digital camera (Nikon D90 Nikon, Tokyo, Japan). The camera was mounted on a tripod with the lens parallel to a measurement grid. I then took a 5 x 5 mm

tissue sample from the upper corner of the caudal fin and preserved it in 95% ethanol. Specimens were allowed to recover before returning them to the river.

Table 3.1.Summary of sample locations

Huazalingo

Abbreviation	Latitude	Longitude	Elevation (Km)	Distance (Km)	Hybrid index	Samples
1 VNNO	20°54'13.21"N	98°39'39.16"W	1180	40.53	0.991	32
2 CNSD	20°55'37.13"N	98°38'35.54"W	900	36.34	0.987	27
3 CSCD	20°56'1.01"N	98°38'18.31"W	880	34.94	0.991	19
4 TCP2	20°56'17.57"N	98°38'13.62"W	769	34.41	0.549	30
5 TCP1	20°56'24.83"N	98°38'7.83"W	751	34.07	0.279	31
6 TOTO	20°56'40.15"N	98°37'51.08"W	720	33.37	0.289	89
7 AHLW	20°57'21.09"N	98°35'24.54"W	588	28.06	0.259	19
8 ACUA	20°57'22.71"N	98°34'17.40"W	450	25.35	0.238	51
9 PILC	20°57'5.04"N	98°33'1.99"W	431	22.8	0.139	32
10 SNPD	20°57'1.37"N	98°31'31.44"W	384	19.83	0.128	22
11 TZCO	20°57'53.36"N	98°28'29.30"W	304	13.63	0.119	38
12 ACHI	20°59'20.15"N	98°22'12.15"W	186	0	0.113	19

Pochula

Abbreviation	Latitude	Longitude	Elevation (Km)	Distance (Km)	Hybrid index	Samples
13 CULH	20°48'23.72"N	98°40'24.98"W	1400	42.65	0.998	26
14 TEXC	20°50'14.75"N	98°36'35.82"W	860	33.68	0.998	23
15 CSRL	20°50'38.49"N	98°35'52.79"W	765	30.89	0.998	16
16 POCH	20°51'26.39"N	98°34'42.44"W	552	27.72	0.994	39
17 PHLW	20°51'27.61"N	98°34'33.09"W	542	27.33	0.995	19
18 OCOT	20°51'40.20"N	98°34'34.35"W	522	26.64	0.954	26
19 ISLT	20°52'2.31"N	98°33'35.65"W	442	24.42	0.661	36
20 PICO	20°52'9.95"N	98°33'3.59"	438	23.39	0.556	39
21 TULA	20°52'45.36"N	98°32'18.45"W	405	21.02	0.512	33
22 VCHO	20°53'12.54"N	98°30'23.88"W	375	16.68	0.465	45
23 ATMP	20°53'8.34"N	98°29'15.44"W	330	13.63	0.386	37
24 AREN	20°54'46.92"N	98°23'34.06"W	221	0	0.227	39

3.3.2. Tissue processing and library preparation

I extracted DNA from all the specimen tissues with Agencourt bead-based purification method (Beckman Coulter Inc., Brea, CA) or DNeasy Tissue Kit (Qiagen, Hilden, Germany). Tissues were incubated overnight at 55 °C in an

incubator (100 rpm) overnight in a 94 ul lysis solution buffer with 3.25 uL 40 mg/mL proteinase K and 2.5 ul DTT. I then followed the bead and spin column extraction and purification. I quantified the genomic samples using a Qubit 4 fluorometer (Invitrogen, Carlsbad, CA) and diluted to 10 ng/ul.

3.3.3. Ancestry assignment and filtering

As in previous *Xiphophorus* studies (Schumer et al. 2014, 2017, 2018) I used the multiplexed shotgun genotyping (MSG) protocol developed by Andolfatto et al. (2011) to generate libraries with custom barcodes. From each specimen, 10 ng of DNA was digested using Tn5 enzyme charged adapter sequences. After digestion, one µl of each sample was amplified with primers containing custom index sequences for 12 cycles. The Tn5 protocol generates high-quality sequencing libraries, with low costs, and high efficiency (Picelli et al., 2014).

Libraries where sequenced on an Illumina Hiseq2500 platform. Raw reads where parsed by index and trimmed to remove low quality (Phred quality score < 20) using Cutadapt and Trim Galore (Li et al., 2012; Martin, 2011). Reads with < 30 bp where discarded. I checked for duplicate reads using Picard tools and discarded any samples with more than 20% duplication. The number of reads per individual ranged from 0.9 to 1.5 million. The MSG pipeline returns the posterior probability of ancestry for each parental species. I used the following parameters following (Schumer et al., 2017): recRate = 420; rfac = 1; *X. birchmanni* error (delta- par1) = 0.05; *X. malinche* error (deltapar2) = 0.05, expected recombination rate : 0.0018 cM/Mb. For each individual I parsed all loci into linkage groups. I then selected all the loci that were

present in all localities. I called posterior ancestry probabilities < 0.95 as ambiguous, >= 0.95 for each parental species as homozygote, and >= 0.95 for both species as heterozygotes. I then filtered out loci with > 50 % of ambiguous calls in all individuals (i.e. > 197 in Pochula and, > 209 in Huazalingo). Adjacent loci with more than 2.5% similarity were considered identical and combined. I used a total of 110510 loci in Huazalingo, and 70253 in Pochula for geographic cline analysis (lower amount due to fewer individuals collected, and more regions removed after ambiguity filtering).

3.3.4. Cline Analysis

I used the Metropolis–Hastings algorithm in the R package hzar (Derryberry et al., 2014), to fit the whole genome hybrid index, and the allele frequencies of each locus to a geographic cline model (Szymura & Barton, 1986). To fit a cline, the model estimates five parameters: cline center $\mathcal{y}$, cline width $\mathcal{w}$, the ends of the cline P_{min} and P_{max} , the rate in which the cline tail decays $\theta_{P1}\ and\ \theta_{P2}$, and the size of the cline tails $\mathcal{B}_{P1}\ and\ \mathcal{B}_{P2}$. For computational speed I fitted three models following Brumfield, Jernigan, McDonald, and Braun (2001). Model I estimated only center $\mathcal{y}$ and width $\mathcal{w}$ from the data, assumed no tails (θ and $\mathcal{B}$ are one and zero respectively), and included fixed ends (P_{min} and P_{max} fixed to zero and one). Model II estimated $\mathcal{y}$, $\mathcal{w}$, and P_{min} and P_{max} from the data and assumed no tails. Model III was the same as Model II with tail estimates θ and $\mathcal{B}$ allowed to vary. Each model parameter was estimated using three independent chain runs using 500,000 MCMC steps after a burn-in of 100,000 steps. Model selection was based on the lowest Akaike information criterion (AIC).

With hzar I obtained the cline center $\mathcal{y}$, cline width $\mathcal{w}$, P_{min} and P_{max} , and the 95% confidence intervals for all loci, and the whole genome hybrid index from the Huazalingo and Pochula rivers. Then I calculated an average value of center $\bar{\mathcal{y}}$, cline width $\bar{\mathcal{w}}$, $\bar{P}_{min}$ and $\bar{P}_{max}$, and the 95% confidence intervals from both rivers. Centers and widths of any two loci were considered coincident (same $\mathcal{y}$) and concordant (same $\mathcal{w}$) if their parameters overlap with the 95% confidence intervals. Loci where considered outliers if they did not overlap with the 95% confidence intervals from $\bar{\mathcal{y}}$, and $\bar{\mathcal{w}}$, and the whole genome hybrid index cline $\mathcal{y}$ and $\mathcal{w}$. I used the function *foverlaps* in the R package data.table (https://rdatatable.gitlab.io/data.table/), to obtain common center and width outlier regions in the Huazalingo cline and the Pochula cline.

3.4. Results

3.4.1. Hybrid index changes considerably along the elevation gradient

The change in proportion of *X. malinche* ancestry (hybrid index) in each locality changes drastically differently between hybrid zones (see Figure 3.2). In the Huazalingo hybrid zone, the change from high *X. malinche* ancestry proportion to high *X. birchmanni* proportion happens at a higher altitude, and within a shorter distance. In the Pochula hybrid zone, this change in ancestry proportions happens at a lower elevation and over a longer distance. The localities at their respective cline centers have also different ancestry proportion distributions (see Figure 3.3). The cline center at the Huazalingo hybrid zone shows a bimodal distribution, with 43% of individuals being a *X. birchmanni* skewed, 37% being *X. malinche* skewed , and only 20% being admixed

individuals. The center at the Pochula hybrid zone shows a unimodal distribution, with only admixed individuals.

3.4.2. Both hybrid zones have steep clines with different centers

Both Huazalingo and Pochula rivers show a steep hybrid index cline (See Figure 3.4). The Huazalingo cline width (w=37.9, 95% CI=12.7-120.8) was lower than the Pochula cline (w=164.1, 95% CI=114.15-237.15). The center of both clines did not coincide. In Huazalingo the cline center (y=788.12, 95%CI= 764.78-797.78) is at higher elevation than in Pochula (y=424.12, 95%CI= 405.41-441.8). Both clines present similar values of P_{max} and P_{min} , however Huazalingo shows more variation (Huazalingo P_{max}= 0.993, 95%CI= 0.949-0.999, $P_{min} = 0.222$, 95%CI= 0.171-0.267; Pochula P_{max}=0.997, $P_{min} = 0.226$).

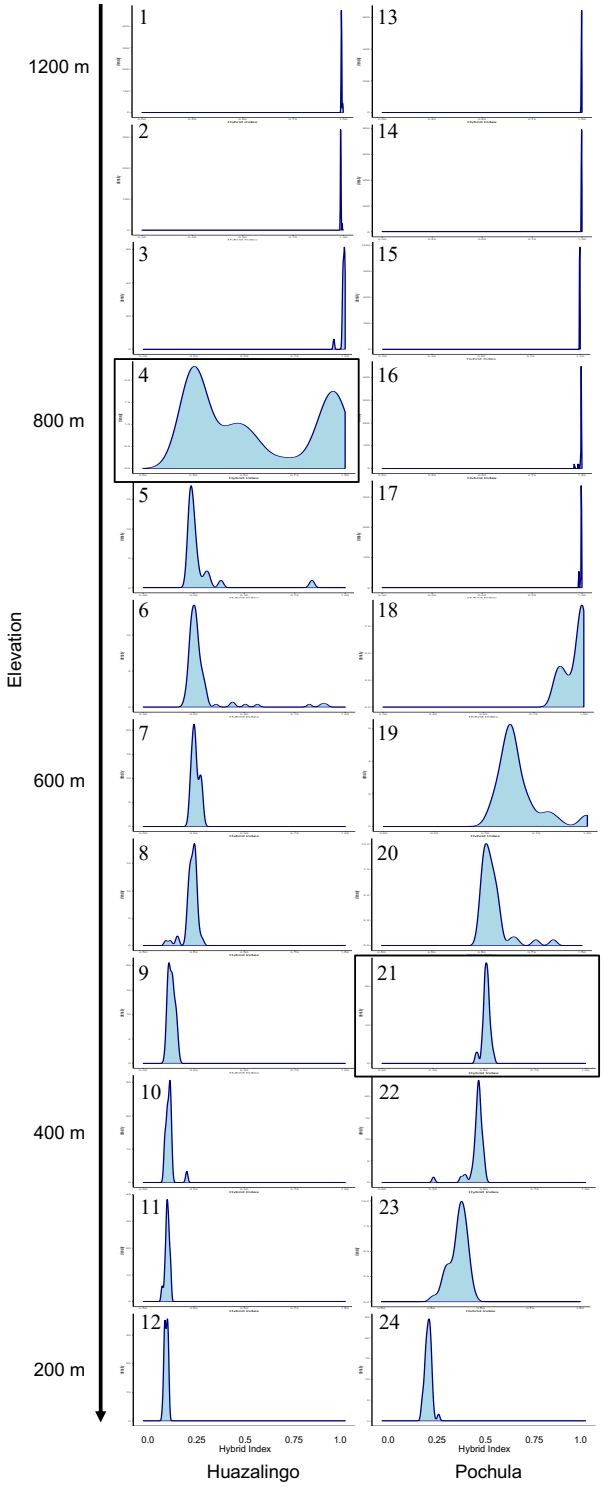


Figure 3.2. Differences in the change in the frequency distribution of hybrid index from high elevation to low elevation between the two hybrid zones. Localities at the center of both clines are enclosed.

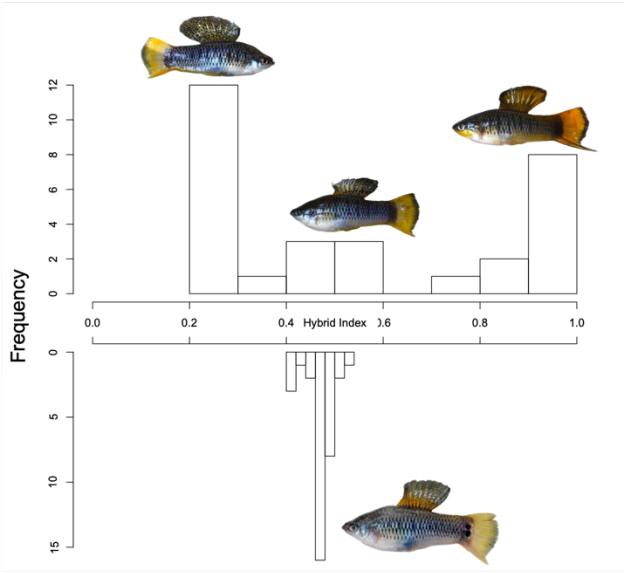


Figure 3.3. Hybrid index distribution at the center of the clines with representative hybrid male individuals. Upper Rio Huazalingo, lower Rio Pochula.

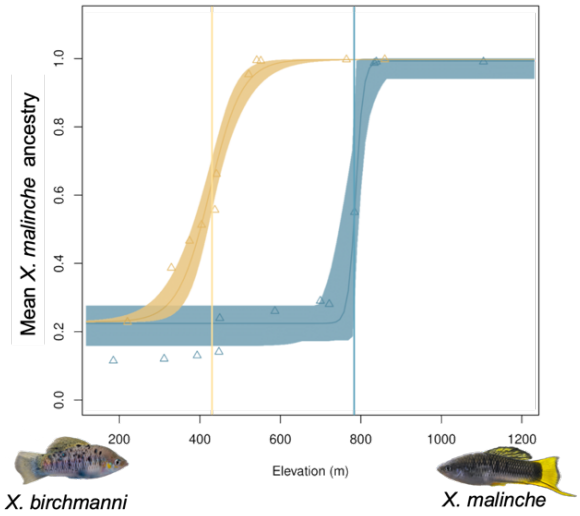


Figure 3.4.Maximum likelihood whole genome hybrid index clines from the Huazalingo hybrid zone (Blue), and Pochula hybrid zone (Yellow), vertical lines indicate the location of the respective cline centers.

3.4.3. Genome-wide cline centers and widths

From the 110510 loci analyzed in Huazalingo, the majority show coincident centers, with 5860 (5.3%) being non-coincident (center outliers) (See Figure 3.5; Sup. Figure 3.6. B). In Pochula, from the 70253 loci analyzed, most loci are also coincident, with 4080 (5.8%) being center outliers (See Figure 3.5; Sup. Figure 3.6, A). In both hybrid zones center outliers show a higher amount of introgression towards *X. birchmanni.* I found a higher number of these outliers (with low centers) in the Huazalingo hybrid zone (94.7%) compared to the Pochula hybrid zone (77.6%).

Most loci analyzed also show concordance. From the 110510 loci in Huazalingo, 5662 (5.12%) show non-concordant clines (width outliers) (See Figure 3.5; Sup. Figure 3.6, D), and from the 70253 loci analyzed in Pochula, 1582 (2.25%) are width outliers (See Figure 3.5; Sup. Figure 3.6, C). I found a larger number of shallow outliers in Huazalingo than Pochula. From 5662 of the width outliers in Huazalingo, 95.6% have large widths, compared to 32.4% from the 1582 outliers in Pochula.

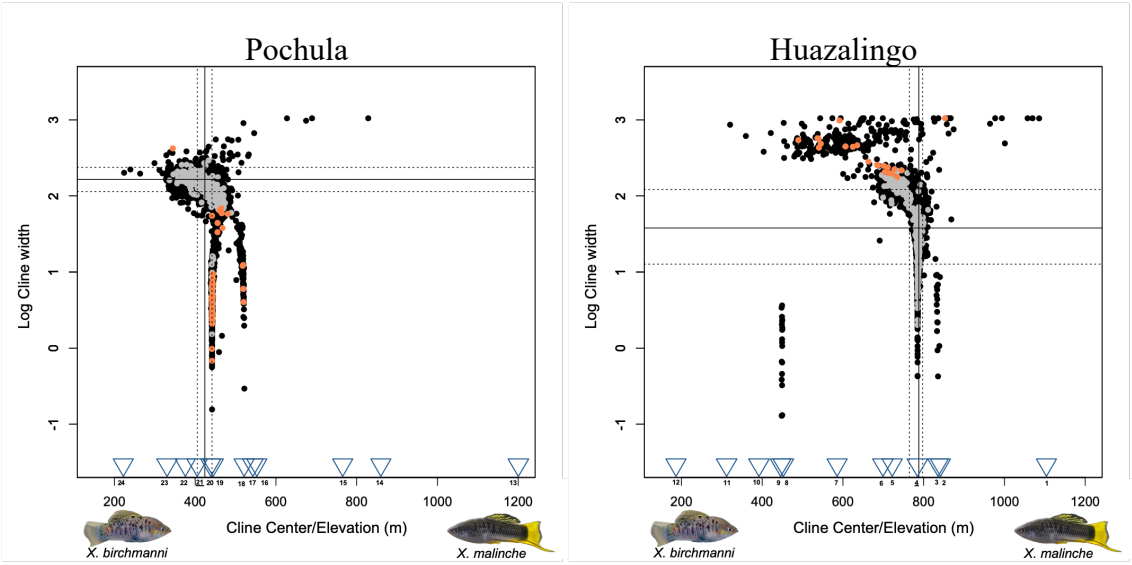


Figure 3.5.Maximum likelihood distribution of cline centers and widths. Bold vertical and horizontal lines indicate hybrid index cline center and width. Dashed lines indicate the hybrid index cline 95% CI. Black dots indicate loci from the complete data set. Grey dots indicate loci from the BDM incompatibilities dataset. Orange dots, loci that are in the BDM incompatibilities dataset that are width outliers. Triangles and numbers indicate the point at the elevation gradient of collection localities (see Table 3.1).

3.4.4. Comparing BDM incompatibilities clines between hybrid zones

I fit clines from a previously published database of Bateson-Dobzhansky-Muller (BDM) incompatibilities between *X. birchmanni* and *X. malinche* (Schumer & Brandvain, 2016; Schumer et al., 2014), and compare the center and width distribution of each pair of loci between hybrid zones. From 374 loci involved in BDM incompatibilities in the Huazalingo hybrid zone, 94.3% are concordant, having a width within the confidence intervals from the overall hybrid index cline (w=37.9, 95% CI=12.7-120.8), while only 5.7% have significant shallow widths. In the Pochula hybrid zone, however, from 363 loci involved in BDM incompatibilities, 86.2% are concordant, having a width within the confidence intervals (w=164.1, 95% CI=114.15-237.15), 13.4% have significant steep widths, and only 0.4% have shallow widths.

Most loci involved in BDM incompatibilities have coincident clines. In Huazalingo from 374 loci 83.9% coincide within the confidence intervals from the overall hybrid index cline (ψ=788.12, 95%CI= 764.78-797.78), and 16.9% are shifted downstream. In Pochula, from 363 loci 80.4% coincide within the confidence intervals (ψ=424.12, 95%CI= 405.41-441.8), 8.2% are shifted downstream and 11.4% are shifted upstream.

3.5. Discussion

In this study I used geographic cline analysis to compare two independent natural hybrid zones between the swordtails *Xiphophorus malinche* and *Xiphophorus birchmanni.* I scanned globally through the genome for outliers using geographic cline analysis. Over all my results show that both hybrid zones present different hybrid distributions along the elevation gradient. The whole genome cline centers are located at considerably different elevations. Hybrid zones under extrinsic selection, located at ecotones, are expected to coincide at similar locations along the environmental gradient (Moore & Price, 1993). Stream gradients have multiple variable gradients (temperature, precipitation, resources, and predation) (Vannote et al., 1980). Using correlations between variation in temperature and allele frequencies along an elevational gradient, Culumber et al. (2012) suggested that drastic and short term changes in temperature indeed play a role in maintaining the structure in these hybrid zones. However, Culumber et al. (2012) also showed that hybrid individuals have greater thermal tolerance than either parental species. The difference in hybrid index cline centers, in addition to Culumber et al. (2012), suggests that environmental selection is not the main pressure maintaining these hybrid zones.

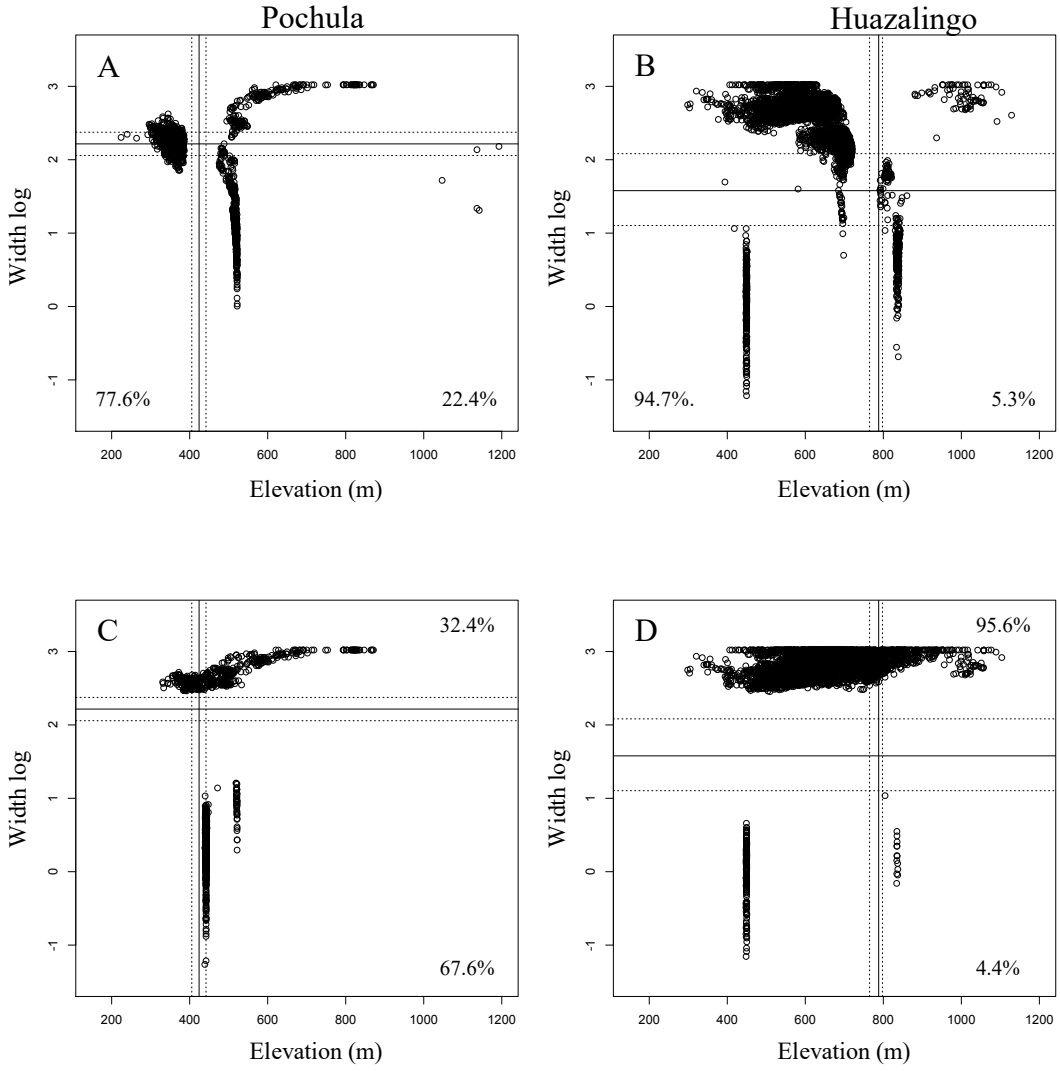
Both hybrid zones have reduced widths, with a drastic change in ancestry over a brief change of elevation. Most of the loci with steep widths colocalized together into specific regions along the elevational gradient in both hybrid zones (see Figure 3.5). In the Pochula hybrid zone, steep clines aggregate in two clusters, at ~440m and at ~540m. In the Huazalingo hybrid zones, they cluster at ~450m, at ~785m, and ~834m. The majority of BDM incompatibilities in Huazalingo and Pochula coincide at the elevation that corresponds to the cline center (~785m and ~440m, respectively). Interestingly, strong steep width outliers aggregate upstream from the cline centers (~834m in Huazalingo and ~540m in Pochula). Specifically, at 834m of elevation, the Huazalingo river has a ~5m waterfall between the localities CNSD and CSCD. In Pochula, the river follows a series of small waterfalls of about ~1-2m. The elevation between the localities POCH and PHLW changes ~10m over ~300m of distance. In both hybrid zones the *X. malinche* average ancestry proportions start to decline after these two specific points on the elevation gradient. Physical barriers impose a drastic effect on geneflow in guppies (Crispo et al., 2006; Shaw, Carvalho, Magurran, & Seghers, 1994), and even small size waterfalls can be strong barriers to larger fish species, as salmonids (Koizumi, Yamamoto, & Maekawa, 2006; Ovidio & Philippart, 2002). My results suggest that physical barriers are also playing an important in shaping the structure of these hybrid zones.

The Huazalingo hybrid zone shows a bimodal hybrid index distribution at the locality near the center of the cline, while the Pochula hybrid zone shows a unimodal distribution (Figure 3.2; Figure 3.3). The hybrid index distribution at the center of

Huazalingo shows two hybrid subpopulations, with some admixed individuals. This bimodality may be caused by assortative mating at the cline center of the Huazalingo hybrid zone. A recent study (Schumer et al., 2017) suggests that assortative mating in a locality (Aguazarca), at another natural hybrid zone between these two species, is causing population structure with two hybrid subpopulations. The results by Schumer et al. suggest that the hybrid index distribution in Aguazarca is caused by strong reproductive isolation between sympatric hybrid subpopulations. Historical sampling shows population structure with some admixed individuals ~25 generations ago. Indeed, my sample size at the center (n = 30) is lower than Schumer et al. (n = 642), however the similarity between both hybrid index distributions suggest that reproductive isolation between clusters may be causing this population structure, and that assortative mating may not be unusual in these natural hybrid zones. Further sampling at the center of the Huazalingo hybrid zone could elucidate these points.

My results suggest that these hybrid zones are tension zones, in which intrinsic selection is the primary role maintaining their distribution. The tension zone model assumes that intrinsic selection causes coincident and concordant clines, significant linkage disequilibrium, and lower fitness of hybrid individuals (Arnold, 1997; Barton & Hewitt, 1985). The fact that the majority of the genomic regions analyzed in both hybrid zones are coincident (94.7% in Huazalingo, 94.2% in Pochula) and showed concordance, having widths equal or lower than the 95% CI in both hybrid zones (~76.7% in Huazalingo and ~97.1% in Pochula) supports this idea. Additionally, the analyzed loci from a previously published database of BDM incompatibilities (Schumer & Brandvain,

2016; Schumer et al., 2014), show them as coincident and concordant as well. These hundreds incompatible loci in significant LD reported by Schumer et al. (2014) imply that strong reproductive isolation is generating the low widths found in my analyzed loci. Additionally, the reported results by Schumer et al. (2014) were analyzing two other independent hybrid zones (Calnali drainage and Rio Claro drainage) than the ones reported in here. The fact that these loci generate a majority of steep clines in my hybrid zones supports the idea of intrinsic selection against hybrids, generating tension zones.



Sup. Figure 3.6. Maximum likelihood of cline centers and widths of loci classified as center outliers (A & B) and width outliers (C & D). Bold vertical and horizontal lines indicate hybrid index cline center and width. Dashed lines indicate the hybrid index cline 95% CI.

4. DIFFERENTIAL INTROGRESSION OF MALE TRAITS IN *XIPHOPHORUS MALINCHE- BIRCHMANNI* NATURAL HYBRID ZONES

4.1. Abstract

Hybridization generates novel genomic combinations in admixed individuals. Generally, these act to reduce fitness caused by epistatic incompatibilities. However, genomic novelty combinations in hybrids can also generate transgressive phenotypes, and traits that increase reproductive success can introgress between hybridizing populations. In this study I aim to characterize phenotypic introgression of sexually-dimorphic traits in two natural hybrid zones between the swordtails *Xiphophorus malinche* and *X. birchmanni.* I measured and compared phenotypic variation along the two elevation gradients. I then performed geographic cline analysis on sexually dimorphic traits, including sword extension. My results show that body size was higher in admixed localities, suggesting transgressive expression of interspecific allele combinations. Sword extension and sword pigment frequency clines showed a considerable reduction despite high *malinche* ancestry localities in the Pochula hybrid zone. These results coincide with previous work showing antipathy for swords by *X. birchmanni* females. In contrast, at Huazalingo sword extension and sword pigment frequency clines did not show this reduction, with the lower sword pigment showing introgression into *X. birchmanni*. The drastic difference between hybrid zones is consistent with a role for assortative mating in maintaining the sword across the gradient in the Huazalingo hybrid zone.

4.2. Introduction

The genomic combinations caused by hybridization are amenable to test hypotheses of the effects of selection on sexually dimorphic traits. On one hand, the admixed genomes in hybrid individuals generate disadvantageous combinations of coadapted genes that impact hybrid fitness (BDM incompatibilities) (Cutter, 2012; Turner & Harr, 2014). Selection against hybrids creates barriers to geneflow, generating narrow clines through the hybrid zone. On the other hand, alleles under positive selection can cross through these barriers, and spread along the hybrid zone (Barton, 1979b). This effect can happen with adaptive traits, being able to introgress despite barriers to gene flow (Parsons, Olson, & Braun, 1993).

Adaptive introgression has been reported in multiple hybrid zones. For example in the sunflower between *Helianthus annuus texanus* and *H. debilis*, one parental species (*H. debilis)* has higher seed resistance to midges, and the formation of a stable hybrid zone has allowed the introgression of seed resistance traits from *H. debilis* to *H. annuus* (Whitney, Randell, & Rieseberg, 2006). In Iris hybrids between *Iris fulva* and *I. brevicaulis*, flood resistance is introgressing from *I. fulva* to *I. brevicaulis* (N. H. Martin, Bouck, & Arnold, 2006). In an avian sex-role reversed hybrid zone between *Jacana spinosa* and *J. jacana*, large female body mass is associated with higher reproductive success (Lipshutz et al., 2019). The asymmetrical introgression of body weight from *J. spinosa* to *J. jacana* across the hybrid zone could be caused by intrasexual selection (Lipshutz et al., 2019). Phenotypic introgression can be greatly augmented by intersexual selection, for example yellow plumage is highly preferred by females in the

manakins *Manacus vitelinus*. Hybrid individuals between *M. vitellinus* and *M. candei* having the preferred phenotypes have higher mating success, generating asymmetric introgression of yellow plumage along the hybrid zone (D. T. Baldassarre & Webster, 2013; Brumfield et al., 2001).

The *Xiphophorus malinche* and *X. birchmanni* hybrid system is an excellent model to test the evolutionary effects of natural and sexual selection on phenotype introgression. Male *X. malinche* have a pigmented elongation of the lower rays of the caudal fin (sword), and a moderately sized dorsal fin. *X. birchmanni*, have reduced or an absence of swords, and have moderate or elongated dorsal fins (Rauchenberger et al., 1990; Rosenthal et al., 2003). Both species have similar body depths and body lengths, however *X. birchmanni* males show higher variance (Johnson, Macedo, Passow, & Rosenthal, 2014; Rosenthal et al., 2003). *X. malinche* and *X. birchmanni* male hybrids display multivariate trait combinations, often outside of the range of parental species (Fisher, Wong, & Rosenthal, 2006; Rosenthal et al., 2003). Fisher et al. (Fisher, Mascuch, & Rosenthal, 2009) reported that *X. birchmanni* females preferer males with trait combinations that can occur in hybrid males. Specifically, *X. birchmanni* females have reduced preferences for large dorsal fins (Fisher et al., 2009), and large caudal fin extensions (sword) (Wong & Rosenthal, 2006). Interestingly, *X. malinche* females lack a strong preference for swords (Wong & Rosenthal, 2006). Intersexual selection against these traits could generate patterns of low introgression across the hybrid zones in this system.

Geographic cline analysis allows one to compare the levels of introgression between phenotypes and genomic regions (Gay et al., 2008; Szymura & Barton, 1986). Specifically, traits known to be involved in reproductive isolation are expected to show reduced patterns of introgression (Gay et al., 2008). Assortative mating can also generate reduced introgression, which can affect traits involved in mating success contributing to strong phenotypic differences (See Billerman, Cicero, Bowie, and Carling (2019), also discussed in Chapter 2). The opposite can be seen when traits that increase mating success, are preferred by the choosers from both species, being able to introgress asymmetrically (Lipshutz et al., 2019; Parsons et al., 1993). In this Chapter I aim to compare the distribution and amount of introgression of multiple sexually dimorphic male traits in two natural hybrid zones between *X. malinche* and *X. birchmanni.*

4.3. Methods

4.3.1. Sample collection

Fish were collected during 2013 to 2017, from 12 localities along an elevation gradient in the Rio Huazalingo and Rio Pochula respectively, in the state of Hidalgo, Mexico (see Sup. Table 4.3). In each locality specimens were lightly anesthetized using tricane methansesulfate. Then I photographed each specimen on both sides using a digital camera (Nikon D90 Nikon, Tokyo, Japan). The camera was mounted on a tripod with the lens parallel to a measurement grid.

4.3.2. Measurements

I identified adult males by the presence of a fully developed gonopodium. performed classical morphometrics in all adult males (See Figure 4.1). I measured

Standard Length (SL) (from the tip of the snout to the posterior end of the mid-lateral portion of the hypural plate), and Body Depth (BD) (from the insertion of the most proximate ray of the dorsal fin to the insertion of the most proximate ray of the anal fin). I also measured Sword Extension (SE, the extension of the ventral caudal fin rays), Dorsal fin Width DW (from the insertion of the most proximate ray of the dorsal fin to the insertion of the most distal ray of the dorsal fin), and Dorsal fin Extension DE (from the insertion of the most distal ray of the dorsal fin to the tip of the most distal ray of the dorsal fin). Then, measurements (BD, SE, DW, and DL) were standardized by the SL to control for allometric effects. Additionally, I recorded the presence or absence of the following discrete phenotypes: a False Gravid Spot (FGS), Sword Upper edge Pigment (SUP) , Sword Lower edge Pigment (SLP), and Spotted Caudal (SC) (Rauchenberger et al., 1990).

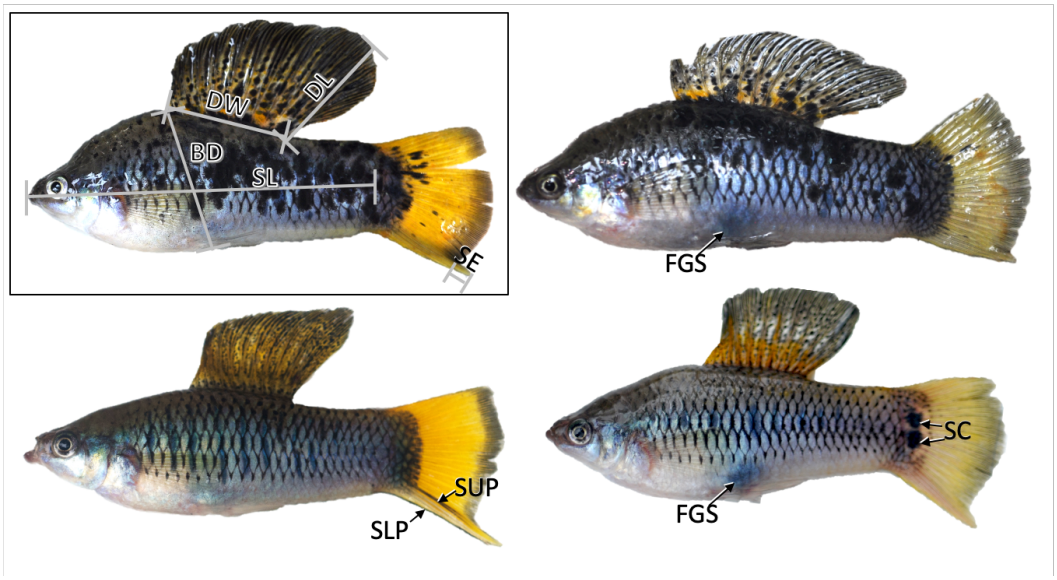


Figure 4.1.Continuous phenotypes used for morphology analysis: Standard Length (SL), Body Depth (BD), Sword Extension (SE), Dorsal fin Width (DW), Dorsal fin Extension (DE). And discrete phenotypes scored for comparative analysis: False Gravid Spot (FGS), Sword Upper edge Pigment (SUP), Sword Lower edge Pigment (SLP), and Spotted Caudal (SP).

4.3.3. Principal component analysis on morphology

All morphometric measurements were log transformed, checked for normality and outstanding outliers were removed. To compare phenotypic variation across the sampled populations and individuals in both hybrid zones, I applied principal component analysis (PCA) to the morphometric measurements divided by Standard length between all adult males using the function prcomp in R (R Core Team, 2019).

4.3.4. Geographic Cline analysis

I fit clines using the Metropolis–Hasting algorithm in the R package hzar (Derryberry et al., 2014), to a geographic cline model (Szymura & Barton, 1986). Specifically, I fitted clines for the sword extension (SE) and, the frequencies of each discrete phenotype (FGS, SUP, SLP, SC) per locality. To fit a cline, the model estimates five parameters: cline center $\mathcal{y}$, cline width $\mathcal{w}$, the ends of the cline P_{min} and P_{max} , the rate in which the cline tail decays $\theta_{P1}\ and\ \theta_{P2}$, and the size of the cline tails $\mathcal{B}_{P1}\ and\ \mathcal{B}_{P2}$. For computational speed I fitted three models following Brumfield et al. (2001). Model I estimated only center $\mathcal{y}$ and width $\mathcal{w}$ from the data, assumed no tails (θ and $\mathcal{B}$ are one and zero respectively), and included fixed ends (P_{min} and P_{max} fixed to zero and one). Model II estimated $\mathcal{y}$, $\mathcal{w}$, and P_{min} and P_{max} from the data, and assumed no tails. Model III was the same as Model II with tail estimates θ and $\mathcal{B}$ allowed to vary. Each model parameter was estimated using three independent chain runs using 500,000 MCMC steps after a burn-in of 100,000 steps. Model selection was based on the lowest Akaike information criterion (AIC).

To estimate phenotypic introgression, I used hzar to obtain the cline center $\mathcal{y}$, cline width $\mathcal{w}$, P_{min} and P_{max} , and the 95% confidence intervals for the sword index (SI), and compared it between the Huazalingo and Pochula rivers, and with the whole genome hybrid index from the Huazalingo and Pochula rivers (See Chapter 3). The sword index Center and width of were considered coincident (same $\mathcal{y}$) and concordant (same $\mathcal{w}$) if their parameters overlap with the 95% confidence intervals of the whole genome hybrid index.

4.4. Results

4.4.1. Quantitative phenotypes along the elevation gradient

Sexually-dimorphic traits either showed shallower clines than the genome-wide average or transgressive phenotypes in hybrids. The distributions of SL, and the standardized BD, SE, DL, and DW (See Figure 4.1 Enclosed) are displayed in Figure 4.2, and Sup. Table 4.3. In both hybrid zones, species typical traits did not always covary with ancestry, with localities with highly admixed individuals having the largest individuals (SL and BD). Highly admixed individuals with larger sizes than either *X. malinche* and *X. birchmanni* skewed populations suggests transgressive phenotypic segregation in these hybrid zones, specifically BD (See figure 4.2). Dorsal fin length showed similar distributions between localities in Huazalingo, with the highest values in VNNO and CNSD. Body size is plastic and also can covary with habitat. The locality of VNNO had environmental perturbation, with local people using calcium hydroxide in the river. In Pochula, DL declined in localities near the center of the hybrid zone and increased at downstream populations with higher *X. birchmanni* ancestry. The DW

distribution showed also a reduction near the center of both hybrid zones. Sword Extension (SE) however, had a maximum value in localities with high *X. malinche* ancestry declining along the elevation, in both hybrid zones.

4.4.2. Principal component analysis on morphology

The first principal component (PC1) explained 49.7% of the total morphometric variation (See Sup. Figure 4.6) and loaded mainly on distances explained by standardized dorsal fin width (Table 4.1). The second principal component (PC2) explained 32.3% of the morphometric variation and loaded mainly on distances explained by standardized sword extension (See Table 4.1). Male adult individuals from both hybrid zones show similar morphometric distributions. However, males along the Pochula hybrid zone appear to have an abrupt reduction in sword extension, showing two groups, one with large swords and another with reduced swords, whereas males along from Huazalingo show a less abrupt reduction in sword extension (See Sup. Figure 4.6, and Figure 4.2).

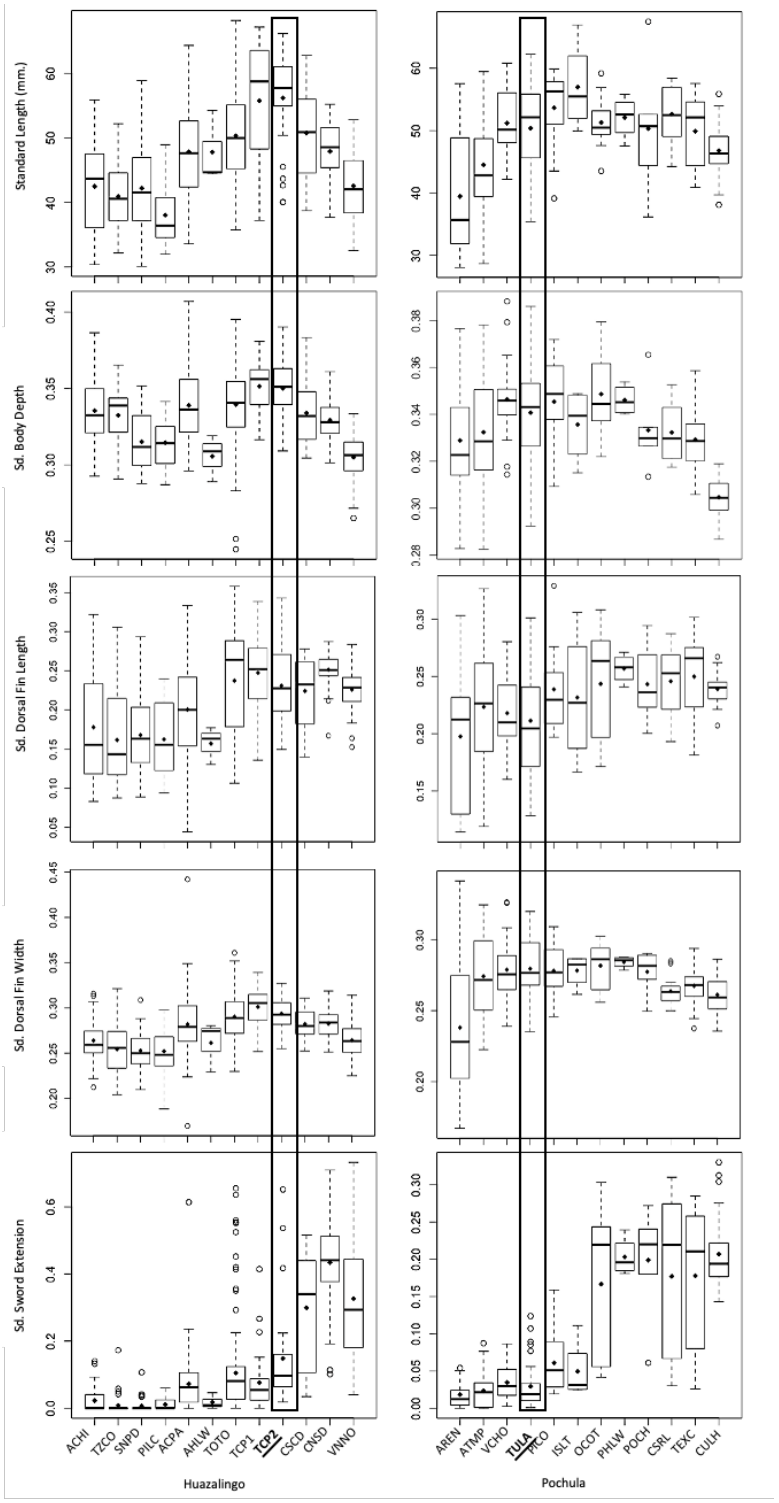


Figure 4.2. Distribution of continuous phenotypes in both hybrid zones. Bolded localities represent the center of the hybrid zone.

Table 4.1. The principal component (PC) scores from the morphometric distances

Phenotypes	PC1	PC2
Sd. Body Depth	0.5640178	-0.3381778
Sd. Sword Extension	0.0330049	**0.8194528**
Sd. Dorsal fin Width	**0.6342987**	-0.1148568
Sd. Dorsal fin Length	0.5276929	0.4482642

The localities at the center of the hybrid zones (TCP2 at Huazalingo, and TULA at Pochula) show unimodal frequency distributions of standardized BD, and DW (See Figure 4.3, A-D). However, the standardized SE distributions were heavily skewed towards reduced swords in both localities (Figure 4.3, E & F). The relationships between the average *X. malinche* ancestry and the sword extension of each individual at the hybrid center are contrastingly different (Figure 4.4, A). The center in Huazalingo shows a bimodal distribution of ancestry, with a correlation with Sd. Sword Extension (Spearman Rank R=0.41 p=0.021). Body depth does not have a correlation with ancestry at the Huazalingo hybrid zone center (Spearman Rank R=0.038 p=0.84 at Huazalingo).

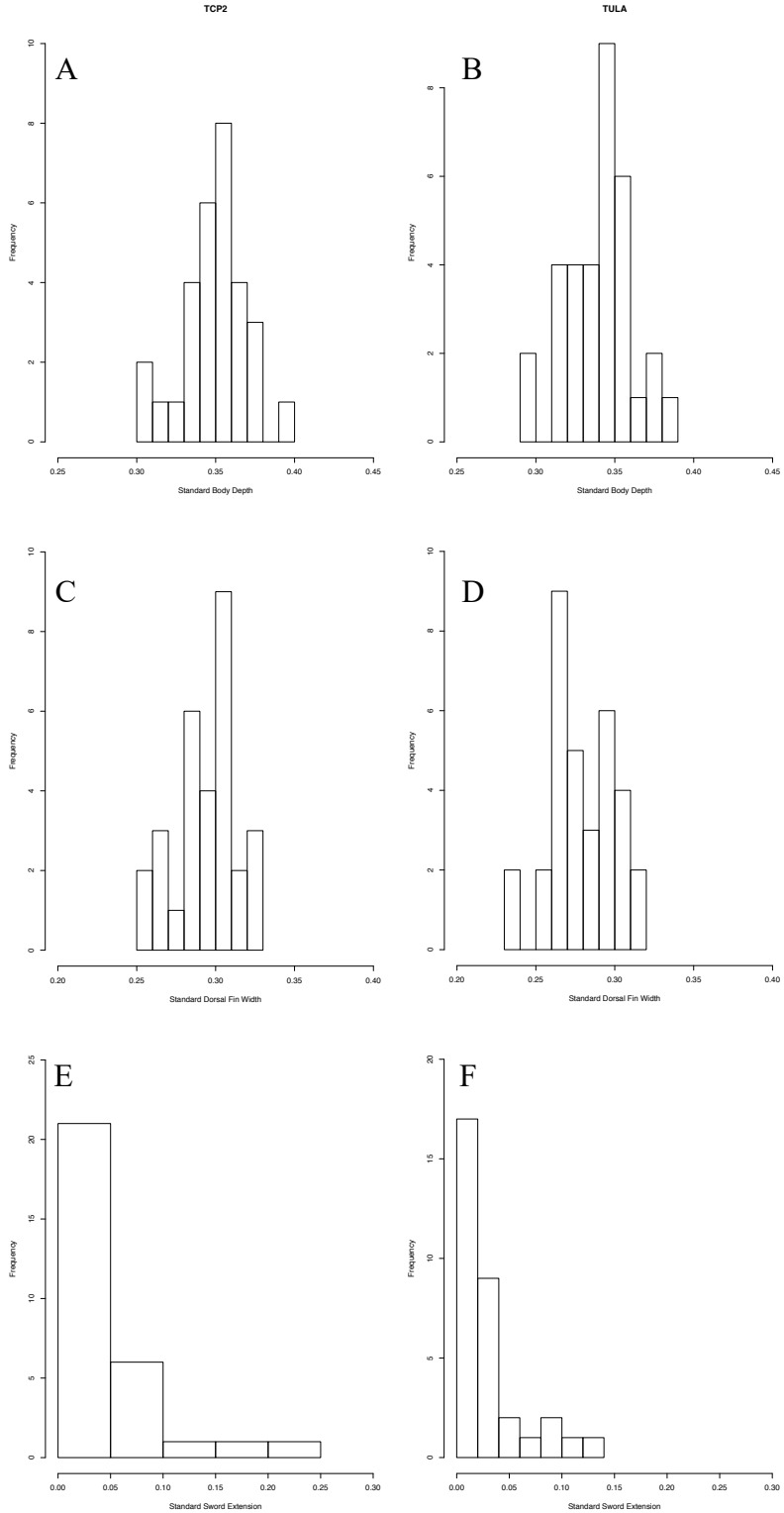


Figure 4.3.Frequency distributions for Standardized BD, DW, and SE, between the localities located at the center of the hybrid zones in Huazalingo (TCP2) and Pochula (TULA).

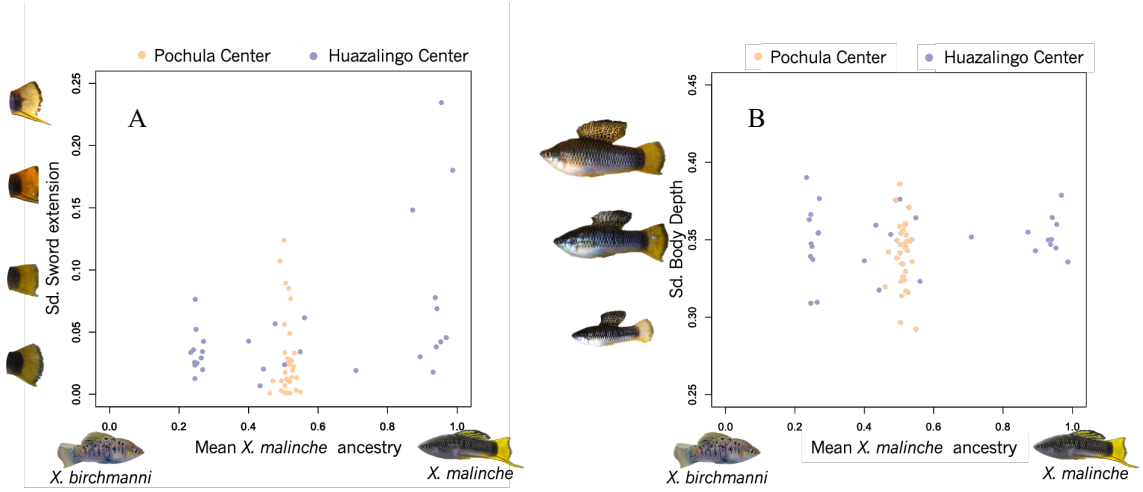


Figure 4.4.Relationship between mean ancestry and Sd. Sword Extension (panel A), and Sd. Body Depth (panel B) at the center of each hybrid zones. Huazalingo (blue dots) and Pochula (yellow dots).

4.4.3. Cline analysis distributions

In both hybrid zones sword extension shows steep clines (see Figure 4.5, red cline). The sword extension cline center at Pochula (y= 468.31, 95%CI= 451.49 - 492.69) did not coincide with the overall Hybrid index cline center and is located at a higher elevation (y=424.12, 95%CI= 405.41-441.8). The sword extension cline center (y= 783.83, 95% CI= 770.9 - 836.03) in Huazalingo overlapped with the Hybrid index cline center (y=788.12, 95%CI= 764.78-797.78). This suggests that the *X. birchmanni* phenotype (loss of sword) is introgressing in localities with high *X. malinche* ancestry at Pochula. This introgression is not present in Huazalingo hybrid zone.

The cline parameters of the discrete phenotypes (See Figure 4.1.) is shown in Table 4.2, and visualized in Figure 4.5. In both hybrid zones the False gravid spot (FGS) and Spotted Caudal (SC) phenotypic clines coincide with the average hybrid index cline center. The Lower Sword Pigment (SLP) phenotypic cline did not coincide with the

hybrid index cline. In Huazalingo the SLP center was located at a lower elevation, suggesting that the presence of this trait is introgressing towards *X. birchmanni* genome backgrounds. Interestingly in Pochula, I saw the opposite. The SLP center was located at a higher elevation from the hybrid index center, suggesting that the absence of this trait is introgressing into *X. malinche* genomic backgrounds. Additionally, the upper sword edge pigment (SUP) also showed an upstream introgression only in Pochula. In Huazalingo this trait coincided with the hybrid index cline.

Table 4.2. Maximum likelihood cline parameters from the discrete phenotypes analyzed, and sword extension. Bold and underlined values did not overlap with the average genome hybrid cline, and arrows show direction of introgression. Colors match Figure 4.5.

Huazalingo

Phenotype	Center	Width	pMin	pMax
Whole genome hybrid index	788.12	37.9	0.22	0.99
False Gravid Spot (FGS)	785.70	31.86	0.00	0.35
Sword Upper Pigment (SUP)	798.55	76.84	0.00	1.00
Sword Lower Pigment (SLP)	↓ **686.27**	**279.77**	0.04	1.00
Spotted Caudal (SC)	780.49	21.29	0.00	0.49
Sword Extension (SE)	783.83	42.6	1.2	9.51

Pochula

Phenotype	Center	Width	pMin	pMax
Whole genome hybrid index	424.12	164.1	0.22	0.99
False Gravid Spot (FGS)	447.56	212.46	0.00	0.62
Sword Upper Pigment (SUP)	↑ **520.12**	8.08	0.17	0.89
Sword Lower Pigment (SLP)	↑ **522.68**	9.26	0.35	0.95
Spotted Caudal (SC)	441.93	1.78	0.00	0.54
Sword Extension (SE)	↑ **468.31**	95.3	0.09	5.34

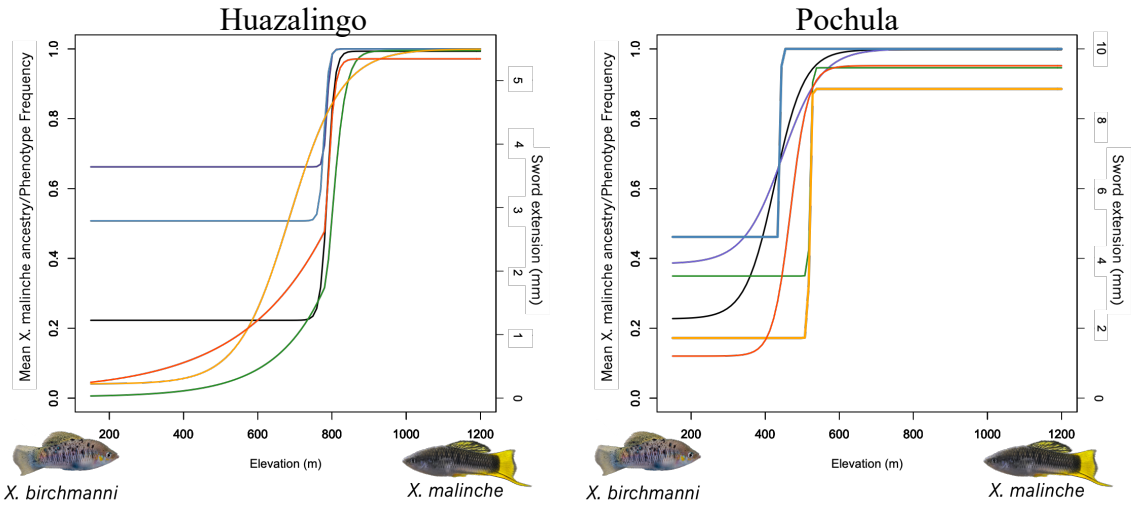


Figure 4.5.Comparison between maximum likelihood clines for whole genome hybrid index (black), phenotype frequency of FGS (purple), SC (blue), SUP (green), SLP (yellow), and sword extension (red). See Table 4.2. **for the cline parameters.**

4.5. Discussion

Hybridization generates new combination of genomic regions in admixed individuals from recently diverged taxa (P. R. Grant & Grant, 2010). Generally, the main consequence is reduced fitness caused by incompatibilities in these new combinations (Seehausen et al., 2014). However advantageous alleles that increase fitness can move between hybridizing populations (Anderson, 1949), crossing reproductive barriers by recombining into a favorable genomic background (Barton, 1979b). Adaptive introgression has been widely reported in hybridizing taxa (as in radiations (P. R. Grant & Grant, 2008; Seehausen, 2004), hybrid speciation (Gompert, Fordyce, Forister, Shapiro, & Nice, 2006; Mavarez et al., 2006; Rieseberg, Archer, & Wayne, 1999), and phenotypic introgression (Greig, Baldassarre, & Webster, 2015; N. H. Martin et al., 2006; Parsons et al., 1993; Whitney et al., 2006). Specifically, phenotypic introgression

can be greatly augmented by sexual preferences, when hybrid individuals having a preferred species specific phenotypes outcompetes other individuals lacking them (Brumfield et al., 2001; Greig et al., 2015). Additionally new genomic combinations in hybrid individuals can generate extreme transgressive phenotypes (Rieseberg et al., 1999; Seehausen, 2004) not seen in either parental species (Harrison & Larson, 2014) . In this study I compared the distribution and assessed introgression of multiple sexually dimorphic traits, in two natural hybrid zones between the swordtails *X. malinche* and *X. birchmanni.*

My results show that hybrids between *X. malinche* and *X. birchmanni* show transgressive expression of body size in locations with high admixture. These results coincide with Rosenthal et al. (Rosenthal et al., 2003) who also found transgressive expression in body depth in another hybrid zone (Rio Canali) independent from the two reported in this study. Transgressive hybrid individuals have an advantage since female preference is correlated with larger male body sizes in *X. birchmanni* (Fisher et al., 2009).

My results show a drastic reduction in sword extension at admixed localities. The sexually dimorphic sword shows a sharp decline on the elevation gradient compared to the hybrid index in the Pochula hybrid zone. This signal is weak in Huazalingo, the sword cline coincides with the hybrid index cline. Additionally, the frequency of SUP and SLP also declines in areas with high *X. malinche* ancestry in Pochula (the SUP decline is also weak in Huazalingo). In Huazalingo, SLP shows a shallow cline, introgressing from *X. malinche* to *X. birchmanni* backgrounds. The pattern of upstream

introgression of reduced sword extension, and the reduced sword pigment frequencies at Pochula, suggest that sexual selection could be driving the reduction of this trait. Even though females from the genus *Xiphophorus* have an ancestral preference for individuals with swords (Basolo, 1990, 1995), previous work shows that in *X. birchmanni* females have an antipathy for sworded conspecific males (Fisher et al., 2009; Wong & Rosenthal, 2006). Large and unsworded hybrid individuals could be outcompeting sworded males by being more attractive to hybrid females. Hybrid females could vary in their preferences based on their genomic ancestry, however Paczolt et al. (2015) found no variation in mating patterns between hybrid females (with ~ 0.12% *X. malinche* ancestry) and *X. birchmanni* females. Future studies measuring female preferences along these geographic gradients could elucidate the loss of reduction of the sword in these hybrid zones.

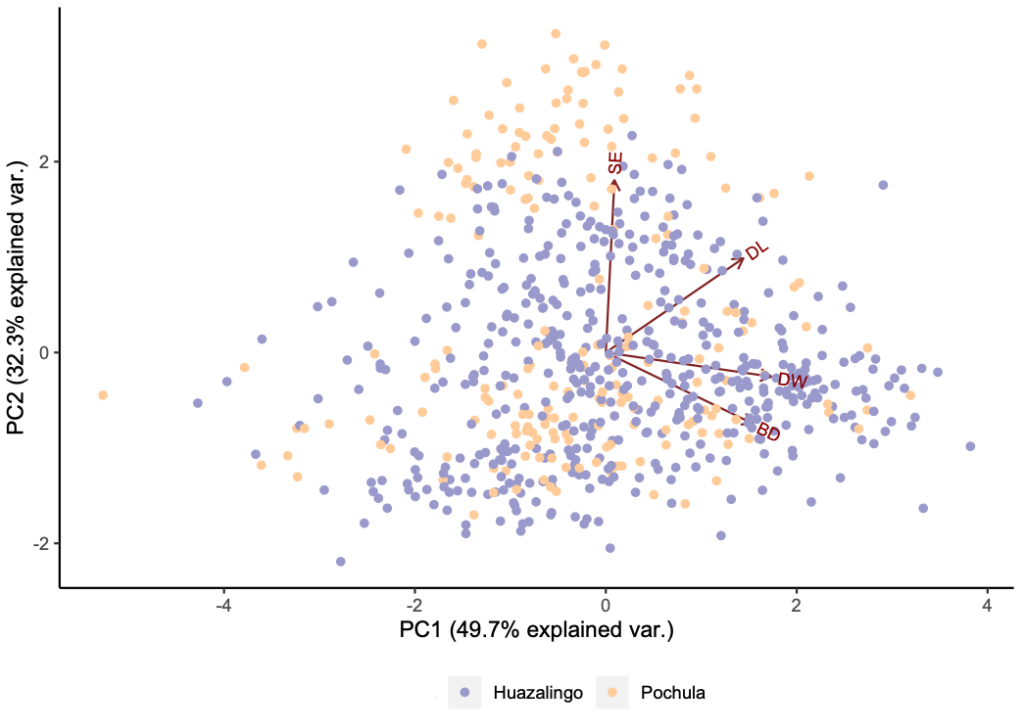
The differences in the patterns of sword extension and sword pigment introgression between Pochula and Huazalingo can be explained by the population structure seen at the center of the Huazalingo hybrid zone. If there is indeed population structure with two subpopulations, and if assortative mating is driving that ancestry distribution, as it has been reported another independent stream by Schumer et al. (2017), then assortative mating could increase the frequency of *X. malinche* traits in Huazalingo, but not in Pochula. Hybrid females at TCP2 could mate more frequently within their ancestry group, with hybrid males with similar ancestry, increasing the frequency of sword phenotypes, specifically, the upper and lower sword pigment. At

Pochula, hyrbid females could discriminate against those sword phenotypes, increasing the reduction in frequency along the hybrid zone.

Sup. Table 4.3. Distribution of continuous phenotypes in both hybrid zones. Standard Length SL (mm), Body depth (BD, Sword extension (SE), Dorsal fin width (DW), and Dorsal fin length (DL). Bolded localities represent the center of the hybrid zone.

Huazalingo							SL		BD		SE		DW		DL	
Abbreviation	Latitude	Longitude	Elevation(m)	Distance	Hybrid index	Samples	mean	SD	mean	SD	mean	SD	mean	SD	mean	SD
VNNO	20°54'13.21"N	98°39'39.16"W	1180	40.53	0.991	56	42.40	5.25	0.31	0.01	0.10	0.06	0.26	0.02	0.23	0.03
CNSD	20°55'37.13"N	98°38'35.54"W	900	36.34	0.987	57	47.82	4.29	0.33	0.01	**0.14**	0.04	0.28	0.01	0.25	0.02
CSCD	20°56'1.01"N	98°38'18.31"W	880	34.94	0.991	38	50.65	6.47	0.33	0.02	0.10	0.05	0.28	0.02	0.22	0.04
TCP2	20°56'17.57"N	98°38'13.62"W	769	34.41	0.549	31	**56.03**	7.12	0.35	0.02	0.05	0.05	0.29	0.02	0.23	0.05
TCP1	20°56'24.83"N	98°38'7.83"W	751	34.07	0.279	30	55.61	9.01	**0.35**	0.02	**0.03**	0.03	**0.30**	0.02	**0.25**	0.05
TOTO	20°56'40.15"N	98°37'51.08"W	720	33.37	0.289	172	50.19	6.93	0.34	0.02	0.04	0.04	0.29	0.02	0.24	0.06
AHLW	20°57'21.09"N	98°35'24.54"W	588	28.06	0.259	3	47.75	5.75	0.31	0.02	0.01	0.01	0.26	0.03	0.16	0.02
ACUA	20°57'22.71"N	98°34'17.40"W	450	25.35	0.238	119	47.81	6.78	0.34	0.02	0.03	0.03	0.28	0.03	0.20	0.06
PILC	20°57'5.04"N	98°33'1.99"W	431	22.8	0.139	42	38.08	4.92	0.31	0.01	0.00	0.01	0.25	0.03	0.16	0.05
SNPD	20°57'1.37"N	98°31'31.44"W	384	19.83	0.128	32	42.22	7.35	0.32	0.02	0.00	0.01	0.25	0.02	0.17	0.05
TZCO	20°57'53.36"N	98°28'29.30"W	304	13.63	0.119	38	40.93	5.41	0.33	0.02	0.00	0.01	0.25	0.03	0.16	0.06
ACHI	20°59'20.15"N	98°22'12.15"W	186	0	0.113	35	42.50	7.19	0.34	0.02	0.01	0.01	0.26	0.03	0.18	0.07

Pochula							SL		BD		SE		DW		DL	
Abbreviation	Latitude	Longitude	elevation	Distance	Hybrid index	Samples	mean	SD	mean	SD	mean	SD	mean	SD	mean	SD
CULH	20°48'23.72"N	98°40'24.98"W	1400	42.65	0.998	32	46.77	3.83	0.30	0.01	0.21	0.05	0.26	0.01	0.24	0.01
TEXC	20°50'14.75"N	98°36'35.82"W	860	33.68	0.998	23	49.89	5.87	0.33	0.01	0.18	0.09	0.27	0.01	0.25	0.04
CSRL	20°50'38.49"N	98°35'52.79"W	765	30.89	0.998	16	52.62	4.52	0.33	0.01	0.18	0.10	0.26	0.01	0.25	0.03
POCH	20°51'26.39"N	98°34'42.44"W	552	27.72	0.994	6	50.32	10.36	0.33	0.02	0.20	0.07	0.28	0.02	0.24	0.03
PHLW	20°51'27.61"N	98°34'33.09"W	542	27.33	0.995	4	**52.10**	3.46	**0.35**	0.01	**0.20**	0.03	**0.28**	0.00	**0.26**	0.01
OCOT	20°51'40.20"N	98°34'34.35"W	522	26.64	0.954	12	51.28	4.14	0.35	0.02	0.17	0.10	0.28	0.02	0.24	0.05
ISLT	20°52'2.31"N	98°33'35.65"W	442	24.42	0.661	4	56.96	7.26	0.34	0.02	0.05	0.04	0.28	0.01	0.23	0.06
PICO	20°52'9.95"N	98°33'3.59"	438	23.39	0.556	13	53.65	6.21	0.35	0.02	0.06	0.04	0.28	0.02	0.24	0.04
TULA	20°52'45.36"N	98°32'18.45"W	405	21.02	0.512	33	50.38	7.24	0.34	0.02	0.03	0.03	0.28	0.02	0.21	0.05
VCHO	20°53'12.54"N	98°30'23.88"W	375	16.68	0.465	26	51.19	5.00	0.35	0.02	0.03	0.02	0.28	0.02	0.22	0.03
ATMP	20°53'8.34"N	98°29'15.44"W	330	13.63	0.386	27	44.50	7.76	0.33	0.02	0.02	0.02	0.27	0.03	0.22	0.05
AREN	20°54'46.92"N	98°23'34.06"W	221	0	0.227	13	39.45	10.59	0.33	0.03	0.02	0.02	0.24	0.05	0.20	0.06



Sup. Figure 4.6. PC1 and PC2 from the Principal Component Analysis, showing the loadings from four standardized measurements.

5. CONCLUSIONS

Geographic cline analysis can be a powerful tool to detect introgression and selection in young hybrid zones. In Chapter 2, I evaluated the performance of geographic cline analysis at detecting introgression in hybrid zones with different demographic histories. My estimation of False Positive Rates in Chapter 2, and the estimated population sizes and migration rates by Culumber et al. (2014), and the estimated hybrid zone age by Schumer et al. (2017), supports the idea that the *Xiphophorus malinche-X. birchmanni* is an amenable model for detecting and comparing levels of introgression using cline theory.

I demonstrate in Chapter 3 that in the *X. malinche-X. birchmanni* hybrid system, selection against incompatible genotypes generates narrow hybrid zones. I demonstrate that the distribution of hybrids indices along two independent rivers do not coincide along the elevation gradient. I argue in Chapter 3 suggest that these two hybrid zones are tension zones, with physical barriers and intrinsic selection shaping the structure along the two gradients.

My results in Chapter three also shows population structure at the center the Huazalingo hybrid zone. Assortative mating has generated similar population structure, with two hybrid subpopulations, at another independent natural hybrid zone (Schumer et al., 2017). The increased number of BDM incompatibility loci with shallow widths at Huazalingo, suggests that assortative mating could facilitate introgression of these incompatibility alleles. Finally, Chapter 4 explores the introgression patterns of sexually dimorphic traits. I found that, even with strong selection reducing genomic introgression

(see Chapter 3), sexually dimorphic traits introgress broadly across the hybrid zones, with body size showing transgressive expression in highly admixed localities. Interestingly the sword extension, and the sword upper and lower pigments showed contrasting introgression patterns in both hybrid zones. Along the Huazalingo hybrid zone these phenotypes are maintained, possibly by assortative mating. Along the Pochula these phenotypes are lost, possibly by female preference for absence of swords.

How does selection shape the genomic and phenotypic introgression along spatial gradients in hybridizing taxa? What determines the spatial location of limited introgression in hybrid zones? My dissertation sheds light in these questions and opens up new avenues to explore. For example, how independent are hybrid zones distributions from environmental and physical variables? What is the effect of assortative mating in shaping genome wide introgression? How do female mate preferences vary along the hybrid zone? And, how does mate choice affect patterns of phenotypic introgression? Addressing these questions will enhance our understanding of the evolutionary consequences of natural hybridization.

REFERENCES

Abbott, R., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J. E., Bierne, N., . . . Zinner, D. (2013). Hybridization and speciation. *Journal of Evolutionary Biology, 26*(2), 229-246.

Anderson, E. (1949). *Introgressive hybridization*: John Wiley & Sons, Inc., New York.

Andolfatto, P., Davison, D., Erezyilmaz, D., Hu, T. T., Mast, J., Sunayama-Morita, T., & Stern, D. L. (2011). Multiplexed shotgun genotyping for rapid and efficient genetic mapping. *Genome Research, 21*(4), 610-617.

Arnold, M. L. (1997). Natural hybridization and evolution *Science, 278*(5336), 235-236.

Baldassarre, D. T., & Webster, M. S. (2013). Experimental evidence that extra-pair mating drives asymmetrical introgression of a sexual trait. *Proceedings of the Royal Society B-Biological Sciences, 280*(1771).

Baldassarre, D. T., White, T. A., Karubian, J., & Webster, M. S. (2014). Genomic and morphological analysis of a semipermeable avian hybrid zone suggests asymmetrical introgression of a sexual signal. *Evolution, 68*(9), 2644-2657.

Barton, N. H. (1979a). The dynamics of hybrid zones. *Heredity, 43*(3), 341-359. doi:10.1038/hdy.1979.87

Barton, N. H. (1979b). Gene flow past a cline. *Heredity, 43*(Dec), 333-339. doi:Doi 10.1038/Hdy.1979.86

Barton, N. H., & Hewitt, G. M. (1985). Analysis of hybrid zones. *Annual Review of Ecology and Systematics, 16*, 113-148.

Barton, N. H., & Hewitt, G. M. (1989). Adaptation, speciation and hybrid zones. *Nature, 341*(6242), 497-503.

Basolo, A. L. (1990). Female preference predates the evolution of the sword in swordtail fish. *Science, 250*(4982), 808-810.

Basolo, A. L. (1995). A further examination of a preexisting bias favoring a sword in the genus xiphophorus. *Animal Behaviour, 50*, 365-375.

Beaumont, M. A., & Balding, D. J. (2004). Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology, 13*(4), 969-980.

Billerman, S. M., Cicero, C., Bowie, R. C. K., & Carling, M. D. (2019). Phenotypic and genetic introgression across a moving woodpecker hybrid zone. *Molecular Ecology, 28*(7), 1692-1708.

Brown, B. L., & Swan, C. M. (2010). Dendritic network structure constrains metacommunity properties in riverine ecosystems. *Journal of Animal Ecology, 79*(3), 571-580.

Brumfield, R. T., Jernigan, R. W., McDonald, D. B., & Braun, M. J. (2001). Evolutionary implications of divergent clines in an avian (Manacus : Aves) hybrid zone. *Evolution, 55*(10), 2070-2087.

Buerkle, C. A., & Lexer, C. (2008). Admixture as the basis for genetic mapping. *Trends in Ecology & Evolution, 23*(12), 686-694.

Buggs, R. J. A. (2007). Empirical study of hybrid zone movement. *Heredity, 99*(3), 301-312.

Carling, M. D., & Brumfield, R. T. (2008). Haldane's rule in an avian system: using cline theory and divergence population genetics to test for differential introgression of mitochondrial, autosomal, and sex-linked loci across the passerina bunting hybrid zone. *Evolution, 62*(10), 2600-2615.

Coyne, J. A., & Orr, H. A. (2004). *Speciation*: Sunderland, MA: Sinaeur Associates.

Crispo, E., Bentzen, P., Reznick, D. N., Kinnison, M. T., & Hendry, A. P. (2006). The relative influence of natural selection and geography on gene flow in guppies. *Molecular Ecology, 15*(1), 49-62.

Cui, R. F., Schumer, M., Kruesi, K., Walter, R., Andolfatto, P., & Rosenthal, G. G. (2013). Phylogenomics reveals extensive reticulate evolution in xiphophorus fishes. *Evolution, 67*(8), 2166-2179.

Cui, R. F., Schumer, M., & Rosenthal, G. G. (2016). Admix'em: a flexible framework for forward-time simulations of hybrid populations with selection and mate choice. *Bioinformatics, 32*(7), 1103-1105.

Culumber, Z. W., Fisher, H. S., Tobler, M., Mateos, M., Barber, P. H., Sorenson, M. D., & Rosenthal, G. G. (2011). Replicated hybrid zones of Xiphophorus swordtails along an elevational gradient. *Molecular Ecology, 20*(2), 342-356.

Culumber, Z. W., Ochoa, O. M., & Rosenthal, G. G. (2014). Assortative mating and the maintenance of population structure in a natural hybrid zone. *American Naturalist, 184*(2), 225-232.

Culumber, Z. W., Shepard, D. B., Coleman, S. W., Rosenthal, G. G., & Tobler, M. (2012). Physiological adaptation along environmental gradients and replicated hybrid zone structure in swordtails (Teleostei: Xiphophorus). *Journal of Evolutionary Biology, 25*(9), 1800-1814.

Cutter, A. D. (2012). The polymorphic prelude to Bateson-Dobzhansky-Muller incompatibilities. *Trends in Ecology & Evolution, 27*(4), 209-218.

Derryberry, E. P., Derryberry, G. E., Maley, J. M., & Brumfield, R. T. (2014). hzar: hybrid zone analysis using an R software package. *Molecular Ecology Resources, 14*(3), 652-663.

Dobzhansky, T. (1937). *Genetics and the Origin of Species* (Vol. 11): Columbia University Press.

Endler, J. A. (1977). Geographic variation, speciation, and clines. *Monographs in population biology, 10*, 1-246.

Felsenstein, J. (1975). Genetic drift in clines which are maintained by migration and natural-selection. *Genetics, 81*(1), 191-207.

Fisher, H. S., Mascuch, S. J., & Rosenthal, G. G. (2009). Multivariate male traits misalign with multivariate female preferences in the swordtail fish, Xiphophorus birchmanni. *Animal Behaviour, 78*(2), 265-269.

Fisher, H. S., Wong, B. B. M., & Rosenthal, G. G. (2006). Alteration of the chemical environment disrupts communication in a freshwater fish. *Proceedings of the Royal Society B-Biological Sciences, 273*(1591), 1187-1193.

Fitzpatrick, B. M. (2013). Alternative forms for genomic clines. *Ecology and Evolution, 3*(7), 1951-1966.

Gay, L., Crochet, P. A., Bell, D. A., & Lenormand, T. (2008). Comparing clines on molecular and phenotypic traits in hybrid zones: a window on tension zone models. *Evolution, 62*(11), 2789-2806.

Gompert, Z., & Buerkle, C. A. (2011). Bayesian estimation of genomic clines. *Molecular Ecology, 20*(10), 2111-2127.

Gompert, Z., Fordyce, J. A., Forister, M. L., Shapiro, A. M., & Nice, C. C. (2006). Homoploid hybrid speciation in an extreme habitat. *Science, 314*(5807), 1923-1925.

Gompert, Z., Lucas, L. K., Nice, C. C., Fordyce, J. A., Forister, M. L., & Buerkle, C. A. (2012). Genomic regions with a history of divergent selection affect fitness of hybrids between two butterfly species. *Evolution, 66*(7), 2167-2181.

Gompert, Z., Mandeville, E. G., & Buerkle, C. A. (2017). Analysis of Population Genomic Data from Hybrid Zones. *Annual Review of Ecology, Evolution, and Systematics, Vol 48, 48*, 207-229.

Gompert, Z., Parchman, T. L., & Buerkle, C. A. (2012). Genomics of isolation in hybrids. *Philosophical Transactions of the Royal Society B-Biological Sciences, 367*(1587), 439-450.

Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics, 17*(6), 333-351.

Grant, E. H. C., Lowe, W. H., & Fagan, W. F. (2007). Living in the branches: population dynamics and ecological processes in dendritic networks. *Ecology Letters, 10*(2), 165-175.

Grant, P. R., & Grant, B. R. (2008). *How and why species multiply: the radiation of Darwin's finches.* Princeton, NJ: Princeton University Press.

Grant, P. R., & Grant, B. R. (2010). Conspecific versus heterospecific gene exchange between populations of Darwin's finches. *Philosophical Transactions of the Royal Society B-Biological Sciences, 365*(1543), 1065-1076.

Greig, E. I., Baldassarre, D. T., & Webster, M. S. (2015). Differential rates of phenotypic introgression are associated with male behavioral responses to multiple signals. *Evolution, 69*(10), 2602-2612.

Haldane, J. B. S. (1948). The Theory of a Cline. *Journal of Genetics, 48*(3), 277-284.

Hänfling, B., & Weetman, D. (2006). Concordant genetic estimators of migration reveal anthropogenically enhanced source-sink population structure in the river sculpin, *Cottus gobio*. *Genetics. 173*(3), 1487-1501.

Harrison, R. G., & Larson, E. L. (2014). Hybridization, introgression, and the nature of species boundaries. *Journal of Heredity, 105*, 795-809.

Hermansen, J. S., Haas, F., Trier, C. N., Bailey, R. I., Nederbragt, A. J., Marzal, A., & Saetre, G.-P. (2014). Hybrid speciation through sorting of parental incompatibilities in Italian sparrows. *Molecular Ecology, 23*(23), 5831-5842.

Hewitt, G. M. (1988). Hybrid zones - Natural laboratories for evolutionary studies. *Trends in Ecology & Evolution, 3*(7), 158-167.

Hvala, J. A., Frayer, M. E., & Payseur, B. A. (2018). Signatures of hybridization and speciation in genomic patterns of ancestry. *Evolution, 72*(8), 1540-1552.

Johnson, J. B., Macedo, D. C., Passow, C. N., & Rosenthal, G. G. (2014). Sexual ornaments, body morphology, and swimming performance in naturally hybridizing swordtails (Teleostei: *Xiphophorus*). *Plos One, 9*(10).

Jones, J. C., Fan, S. H., Franchini, P., Schartl, M., & Meyer, A. (2013). The evolutionary history of Xiphophorus fish and their sexually selected sword: a genome-wide approach using restriction site-associated DNA sequencing. *Molecular Ecology, 22*(11), 2986-3001.

Koizumi, I., Yamamoto, S., & Maekawa, K. (2006). Decomposed pairwise regression analysis of genetic and geographic distances reveals a metapopulation structure of stream-dwelling Dolly Varden charr. *Molecular Ecology, 15*(11), 3175-3189.

Li J-W, Robison K, Martin M, et al. (2012). The SEQanswers wiki: a wiki database of tools for high-throughput sequencing analysis. *Nucleic Acids Research*, *40*, D1313-D1317.

Limdi, A., Perez-Escudero, A., Li, A. M., & Gore, J. (2018). Asymmetric migration decreases stability but increases resilience in a heterogeneous metapopulation. *Nature Communications, 9*, 2969.

Lipshutz, S. E., Meier, J. I., Derryberry, G. E., Miller, M. J., Seehausen, O., & Derryberry, E. P. (2019). Differential introgression of a female competitive trait in a hybrid zone between sex-role reversed species. *Evolution, 73*(2), 188-201.

Mallet, J. (2005). Hybridization as an invasion of the genome. *Trends in Ecology & Evolution, 20*(5), 229-237.

Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, *17*(1), 10-12.

Martin, N. H., Bouck, A. C., & Arnold, M. L. (2006). Detecting adaptive trait introgression between Iris fulva and I. brevicaulis in highly selective field conditions. *Genetics, 172*(4), 2481-2489.

Martin, S. H., Davey, J. W., Salazar, C., & Jiggins, C. D. (2019). Recombination rate variation shapes barriers to introgression across butterfly genomes. *PLOS Biology, 17*(2), e2006288.

Martin, S. H., & Jiggins, C. D. (2017). Interpreting the genomic landscape of introgression. *Current Opinion in Genetics & Development, 47*, 69-74.

Massey, S. E. (2017). Strong amerindian mitonuclear discordance in puerto rican genomes suggests amerindian mitochondrial benefit. *81*(2), 59-77.

Mavarez, J., Salazar, C. A., Bermingham, E., Salcedo, C., Jiggins, C. D., & Linares, M. (2006). Speciation by hybridization in Heliconius butterflies. *Nature, 441*(7095), 868-871.

McKenzie, J. L., Dhillon, R. S., & Schulte, P. M. (2015). Evidence for a bimodal distribution of hybrid indices in a hybrid zone with high admixture. *Royal Society Open Science, 2*(12), 150285.

Moore, W. S., & Price, J. T. (1993). Nature of selection in the northern flicker hybrid zone and its implications for speciation theory. In: Harrison RG (ed). *Hybrid Zones and the Evolutionary Process*. Oxford University Press: Oxford. 196–225.

Morrissey, M. B., & de Kerckhove, D. T. (2009). The maintenance of genetic variation due to asymmetric gene flow in dendritic metapopulations. *American Naturalist, 174*(6), 875-889.

Muller, H. J. (1940). Bearing of the Drosophila work on systematics. *The new systematics*, 185-268.

Ovidio, M., & Philippart, J. C. (2002). The impact of small physical obstacles on upstream movements of six species of fish - Synthesis of a 5-year telemetry study in the River Meuse basin. *Hydrobiologia, 483*(1-3), 55-69.

Paczolt, K. A., Passow, C. N., Delclos, P. J., Kindsvater, H. K., Jones, A. M. G., & Rosenthal, G. G. (2015). Multiple mating and reproductive skew in parental and introgressed females of the live-bearing fish *Xiphophorus birchmanni*. *Journal of Heredity, 106*(1), 57-66.

Paradis, E. (2010). pegas: an R package for population genetics with an integrated-modular approach. *Bioinformatics, 26*(3), 419-420.

Parsons, T. J., Olson, S. L., & Braun, M. J. (1993). Unidirectional spread of secondary sexual plumage traits across an avian hybrid zone. *Science, 260*(5114), 1643-1646.

Payseur, B. A. (2010). Using differential introgression in hybrid zones to identify genomic regions involved in speciation. *Molecular Ecology Resources, 10*(5), 806-820.

Payseur, B. A., & Rieseberg, L. H. (2016). A genomic perspective on hybridization and speciation. *Molecular Ecology, 25*(11), 2337-2360.

Picelli, S., Bjorklund, A. K., Reinius, B., Sagasser, S., Winberg, G., & Sandberg, R. (2014). Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Research, 24*(12), 2033-2040.

Polechova, J., & Barton, N. (2011). Genetic drift widens the expected cline but narrows the expected cline width. *Genetics, 189*(1), 227-U905.

Powell, D. L., Garcia, M., Keegan, M., Reilly, P., Du, K., Díaz-Loyo, A. P., . . . Schumer, M. (in press). Natural hybridization reveals incompatible alleles that cause melanoma in swordtail fish. *Science.*

Pulido-Santacruz, P., Aleixo, A., & Weir, J. T. (2018). Morphologically cryptic Amazonian bird species pairs exhibit strong postzygotic reproductive isolation. *Proceedings of the Royal Society B: Biological Sciences, 285*(1874).

R Core Team. (2019). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Rafati, N., Blanco-Aguiar, J. A., Rubin, C. J., Sayyab, S., Sabatino, S. J., Afonso, S., . . . Carneiro, M. (2018). A genomic map of clinal variation across the European rabbit hybrid zone. *Molecular Ecology, 27*(6), 1457-1478.

Rauchenberger, M., Kallman, K. D., & Moritzot, D. C. (1990). Monophyly and geography of the Rio Pánuco basin swordtails (genus *Xiphophorus*) with descriptions of four new species. *American Museum Novitates, 2975*, 1–41.

Rieseberg, L. H., Archer, M. A., & Wayne, R. K. (1999). Transgressive segregation, adaptation and speciation. *Heredity, 83*, 363-372. doi:DOI

Rosenthal, G. G., De La Rosa Reyna, X. F., Kazianis, S., Stephens, M. J., Morizot, D. C., Ryan, M. J., & De Leon, F. J. G. (2003). Dissolution of sexual signal complexes in a hybrid zone between the swordtails *Xiphophorus birchmanni* and *Xiphophorus malinche* (Poeciliidae). *Copeia,* (2), 299-307.

Rosser, N., Dasmahapatra, K. K., & Mallet, J. (2014). Stable Heliconius butterfly hybrid zones are correlated with a local rainfall peak at the edge of the Amazon basin. *Evolution, 68*(12), 3470-3484.

Ryan, S. F., Deines, J. M., Scriber, J. M., Pfrender, M. E., Jones, S. E., Emrich, S. J., & Hellmann, J. J. (2018). Climate-mediated hybrid zone movement revealed with genomics, museum collection, and simulation modeling. *Proceedings of the National Academy of Sciences of the United States of America, 115*(10), E2284-E2291.

Ryan, S. F., Fontaine, M. C., Scriber, J. M., Pfrender, M. E., O'Neil, S. T., & Hellmann, J. J. (2017). Patterns of divergence across the geographic and genomic landscape of a butterfly hybrid zone associated with a climatic gradient. *Molecular Ecology, 26*(18), 4725-4742.

Schartl, M., Walter, R. B., Shen, Y., Garcia, T., Catchen, J., Amores, A., . . . Warren, W. C. (2013). The genome of the platyfish, *Xiphophorus maculatus*, provides insights into evolutionary adaptation and several complex traits. *Nature Genetics, 45*(5), 567–572.

Schumer, M., & Brandvain, Y. (2016). Determining epistatic selection in admixed populations. *Molecular Ecology, 25*(11), 2577-2591.

Schumer, M., Cui, R., Powell, D. L., Dresner, R., Rosenthal, G. G., & Andolfatto, P. (2014). High-resolution Mapping Reveals Hundreds of Genetic Incompatibilities in Hybridizing Fish Species. *Elife, 3*.

Schumer, M., Cui, R. F., Boussau, B., Walter, R., Rosenthal, G., & Andolfatto, P. (2013). An evaluation of the hybrid speciation hypothesis for *Xiphophorus clemenciae* based on whole genome sequences. *Evolution, 67*(4), 1155-1168.

Schumer, M., Cui, R. F., Powell, D. L., Rosenthal, G. G., & Andolfatto, P. (2016). Ancient hybridization and genomic stabilization in a swordtail fish. *Molecular Ecology, 25*(11), 2661-2679.

Schumer, M., Powell, D. L., Delclos, P. J., Squire, M., Cui, R. F., Andolfatto, P., & Rosenthal, G. G. (2017). Assortative mating and persistent reproductive isolation in hybrids. *Proceedings of the National Academy of Sciences of the United States of America, 114*(41), 10936-10941.

Schumer, M., Xu, C. L., Powell, D. L., Durvasula, A., Skov, L., Holland, C., . . . Przeworski, M. (2018). Natural selection interacts with recombination to shape the evolution of hybrid genomes. *Science, 360*(6389), 656-659.

Sciuchetti, L., Dufresnes, C., Cavoto, E., Brelsford, A., & Perrin, N. (2018). Dobzhansky-Muller incompatibilities, dominance drive, and sex-chromosome introgression at secondary contact zones: A simulation study. *Evolution, 72*(7), 1350-1361.

Sedghifar, A., Brandvain, Y., & Ralph, P. (2016). Beyond clines: lineages and haplotype blocks in hybrid zones. *Molecular Ecology, 25*(11), 2559-2576.

Seehausen, O. (2004). Hybridization and adaptive radiation. *Trends in Ecology & Evolution, 19*(4), 198-207.

Seehausen, O., Butlin, R. K., Keller, I., Wagner, C. E., Boughman, J. W., Hohenlohe, P. A., . . . Widmer, A. (2014). Genomics and the origin of species. *Nature Reviews Genetics, 15*(3), 176-192.

Shaw, P. W., Carvalho, G. R., Magurran, A. E., & Seghers, B. H. (1994). Factors affecting the distribution of genetic-variability in the guppy, *Poecilia reticulata*. *Journal of Fish Biology, 45*(5), 875-888.

Slatkin, M. (1973). Gene flow and selection in a cline. *Genetics, 75*(4), 733-756.

Slatkin, M., & Maruyama, T. (1975). Genetic drift in a cline. *Genetics, 81*(1), 209-222.

Stankowski, S., Sobel, J. M., & Streisfeld, M. A. (2017). Geographic cline analysis as a tool for studying genome-wide variation: a case study of pollinator-mediated divergence in a monkeyflower. *Molecular Ecology, 26*(1), 107-122.

Szymura, J. M., & Barton, N. H. (1986). Genetic-analysis of a hybrid zone between the fire-bellied toads, *Bombina bombina* and *Bombina variegata*, near Cracow in southern Poland. *Evolution, 40*(6), 1141-1159.

Taylor, S. A., White, T. A., Hochachka, W. M., Ferretti, V., Curry, R. L., & Lovette, I. (2014). Climate-mediated movement of an avian hybrid zone. *Current Biology, 24*(6), 671-676.

Trier, C. N., Hermansen, J. S., Saetre, G.-P., & Bailey, R. I. (2014). Evidence for mito-nuclear and sex-linked reproductive barriers between the hybrid italian sparrow and its parent species. *Plos Genetics, 10*(1).

Turner, L. M., & Harr, B. (2014). Genome-wide mapping in a house mouse hybrid zone reveals hybrid sterility loci and Dobzhansky-Muller interactions. *Elife, 3*.

van Dijk, E. L., Auger, H., Jaszczyszyn, Y., & Thermes, C. (2014). Ten years of next-generation sequencing technology. *Trends in Genetics, 30*(9), 418-426.

Vannote, R. L., Minshall, G. W., Cummins, K. W., Sedell, J. R., & Cushing, C. E. (1980). River continuum concept. *Canadian Journal of Fisheries and Aquatic Sciences, 37*(1), 130-137.

Walsh, J., Shriver, W. G., Olsen, B. J., & Kovach, A. I. (2016). Differential introgression and the maintenance of species boundaries in an advanced generation avian hybrid zone. *Bmc Evolutionary Biology, 16*.

Wang, L. Y., Luzynski, K., Pool, J. E., Janousek, V., Dufkova, P., Vyskocilova, M. M., . . . Tucker, P. K. (2011). Measures of linkage disequilibrium among neighbouring SNPs indicate asymmetries across the house mouse hybrid zone. *Molecular Ecology, 20*(14), 2985-3000.

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-Statistics for the Analysis of Population-Structure. *Evolution, 38*(6), 1358-1370.

Whitney, K. D., Randell, R. A., & Rieseberg, L. H. (2006). Adaptive introgression of herbivore resistance traits in the weedy sunflower Helianthus annuus. *American Naturalist, 167*(6), 794-807.

Wong, B. B. M., & Rosenthal, G. G. (2006). Female disdain for swords in a swordtail fish. *American Naturalist, 167*(1), 136-140.