ABSTRACT

Title of Dissertation: GENOME-WIDE ANALYSIS OF DIVERGENCE AND INTROGRESSION IN TOWHEE HYBRID ZONES

Sarah Elizabeth Kingston, Doctor of Philosophy, 2012

Dissertation directed by: William F. Fagan, PhD, Department of Biology
Michael J. Braun, PhD, Smithsonian Institution

Hybrid zones offer a natural laboratory in which investigation of the evolutionary forces involved with reproductive isolation and differentiation is possible. Highly multilocus population genomics is a powerful and feasible new tool with which to address such questions of evolutionary interest. I utilize a unique spatial setting that incorporates two hybrid gradients of the towhess *Pipilo maculatus* and *P. ocai*. These species likely diverged in allopatry and are in secondary contact. I utilize genome-scale multilocus techniques to address questions regarding the architecture of differentiation and introgression across these hybrid gradients and the influence of location specific environmental factors on isolation. The multilocus analysis reveals cross-genomic variation in selective constraints on gene flow and locus-specific flexibility in the permeability of the interspecies membrane. Maintenance of historical divergence is acting in a cohesive manner, but local environmental and stochastic factors are also important driving forces. Habitat corridors for dispersal potential indicate hotspots of connectivity where the two transects meet. Both habitat connectivity and genetic

differentiation between geographically disparate parental types appear to influence the dynamics of gene flow across the hybrid gradient. Environmentally-mediated gene flow in the context of secondary contact and hybridization is an important force influencing evolutionary trajectory.

GENOME-WIDE ANALYSIS OF DIVERGENCE AND INTROGRESSION IN TOWHEE HYBRID ZONES

by

Sarah Elizabeth Kingston

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2012

Advisory Committee:
Professor William F. Fagan, Co-Chair
Professor Michael J. Braun, Co-Chair
Professor Matthew P. Hare
Professor David J. Hawthorne
Professor Robert W. Jernigan
Professor Kevin E. Omland
Professor Gerald S. Wilkinson

This dissertation is dedicated to my stalwart support system,

Steven Allen

my ever-encouraging parents,

Thomas and Susan Kingston

and in loving memory of my original academic inspiration,

Emily Elizabeth Isaacs

TABLE OF CONTENTS

Introduction ……1

Chapter 1: Genomic variation in cline shape across a hybrid zone

Introduction ……5

Materials and Methods……10

Results ……15

Discussion ……..18

Tables ……24

Figures ……..……..27

Chapter 2: Genetic differentiation and habitat connectivity in hybrid zones

Introduction ……32

Materials and Methods ……...……34

Results ……38

Discussion ……..……42

Tables ……..……50

Figures ……53

Chapter 3: Genome-wide analysis and comparison of two hybrid transects

Introduction ……61

Materials and Methods ……...……65

Results ……71

Discussion ……..……75

Tables ……..……80

Figures ……83

Supporting Material (Ch 1) ……...……..…90

Supporting Material (Ch 2) ……102

Voucher and Specimen Numbers ……...……106

Literature Cited . .……117

LIST OF TABLES

Table 1.1 sampling locations ordered from north to south through the Teziutlán hybrid gradient 24

Table 1.2. Adapters and primers used in the AFLP assay 25

Table 1.3. Pearson correlation coefficients 26

Table 2.1. Sampling locations across two towhee hybrid transects in Mexico 50

Table 2.2. Population structure schemes tested using the mtDNA ND2 locus in Migrate-n 51

Table 2.3. Evanno method to differentiate among number of populations inferred 52

Table 3.1. Sampling locations across the two hybrid transects 80

Table 3.2. Pairwise population estimates of F_{ST} across all loci 81

Table 3.3. Loci with extreme parameter values and outlier F_{ST} values for each transect 82

LIST OF FIGURES

Figure 1.1. Sampling locations of towhee populations along the Teziutlán hybrid gradient 27

Figure 1.2. Cline models fit to empirical data for all 61 clinal loci using Analyse 1.3 28

Figure 1.3. Distributions of cline centers and widths 29

Figure 1.4. Principal co-ordinates in two dimensions 30

Figure 2.1. Sampling locations across both Teziutlán and Transvolcanic gradients 53

Figure 2.2. Median joining network of the mtDNA ND2 haplotypes 54

Figure 2.3. Structure results showing percentage of ancestry for each individual across both hybrid transects 55

Figure 2.4. Estimates of percent ancestry pooled by sampling location 57

Figure 2.5. Cumulative current map depicting habitat corridors among sampling sites across the Teziutlán and Transvolcanic gradients 58

Figure 2.6. *Circuitscape* resistance distance plotted against pairwise population F_{ST} 59

Figure 3.1. Map of sampling locations for the two hybrid transects in Mexico 83

Figure 3.2. PCA derived from genotype probability matrix 84

Figure 3.3. Frequency distributions of multi-locus hybrid indices 85

Figure 3.4. Frequency distributions and quantiles for genomic cline parameters α and β across both transects. 86

Figure 3.5. F_{ST} estimates between parental population and quantiles for both transects 88

Figure 3.6. The cline graph shows the relationship between the hybrid index and the probability of *P. ocai* ancestry (ϕ) as predicted by each single locus 89

Introduction

Interspecies hybridization offers a powerful natural setting in which we can investigate the interaction of divergent genomes. Lineages can swamp each other or parental species can persist, driven by both selective and stochastic forces (Barton 2001). Introgression of alleles from one species to the other can offer new options to be tested on the adaptive landscape. Alternatively, genomic incompatibilities accumulated during divergence can bind gene flow through reduction of hybrid fitness (Orr & Turelli 2001). Both selective and stochastic forces during historical divergence and the period of hybrid contact can influence the architecture and maintenance of reproductive isolation (Barton 2001; Orr & Turelli 2001). In an area of secondary contact, the selective and stochastic forces associated with initial divergence are at least partially separated in time and possibly space from the conditions during reconnection and hybridization (Gompert *et al.* 2012). There is much interest in processes of incipient speciation, divergence with gene flow, and divergent selection driving sympatric speciation (Danley *et al.* 2000; Emelianov *et al.* 2003; Niemiller *et al.* 2008; Nosil 2008). Due to this partial decoupling of forces, lineages in secondary contact can also offer a powerful natural hybrid laboratory setting in which to observe the evolutionary dynamics of reproductive isolation.

Two species of towhee, *Pipilo maculatus* (spotted towhee) and *P. ocai* (collared towhee) come into contact throughout the high elevation pine/oak scrub habitat in Mexico. There are two major axes of hybridization: the Teziutlán gradient through ribbon-like montane habitat along the ridge of the Sierra Madre Oriental from north to south, and the Transvolcanic gradient across the island-like volcanic peaks that transverse

the central axis of Mexico from west to east (Sibley 1954; Sibley & West 1958). These two axes act as replicate areas contact; while they are not completely independent, each gradient exhibits parental types at each end that are geographically separated from the other gradient's corresponding parental types. A unique aspect of the axes is the area of contact in the southern portion of the range (Mt. Orizaba and south into Oaxaca) where the southern parental form of *P. maculatus* occurs in sympatry with *P. ocai* populations with very low occurrence of hybridization (Sibley 1954; Sibley & Sibley 1964; Braun 1983). The nature of the towhee hybrid contact has a rich history of study and the areas of hybridization have been stable for at least 150 years (and likely much longer) (Sibley 1950, 1954; Sibley & West 1958; Sibley & Sibley 1964; Braun 1983).

I aim to address a suite of evolutionary questions utilizing the power of molecular genomic technology and the unique geographic characteristics of the towhee hybrid contact. I will employ geographic cline (tension zone model) analyses to assess the genomic distribution of cline parameters across the hybrid interface and the variation in signature of the effects of selection and introgression. In addition, I will test the extent and strength of a cohesive cross-locus signal by attempting to pinpoint natural groupings among loci. Using multi-locus data and population genetic analyses, I will test what genetic differentiation is being maintained in the face of current hybridization and how effective current gene flow is at promoting introgression. In addition, I use measures of genetic differentiation and habitat connectivity models to ask: how do the genetic signals of local differentiation and gene flow between populations relate to habitat availability and connectivity? Finally, the power of highly multi-locus genome scale data and population genomic tools allows much more powerful inference from locus-specific

analyses than previously possible even in the recent past. I will use a genome-scale data set to evaluate the level of concordance of both adaptive and non-adaptive forces on the architecture of isolation across both hybrid transects. By isolating loci with extreme parameters in two major models, F-model for locus-specific differentiation and the genomic clines that describe rate and nature of introgression across loci, I am testing overlap in outlier loci patterns between the two transects; concordance should reflect the signal of divergence that maintains core parental species identity in the face of gene flow, discordance will reveal the influence of selective and stochastic forces differing in each environment.

Chapter 1: Genomic variation in cline shape across a hybrid zone

Abstract

Hybrid zones are unique biological interfaces that reveal both population level and species level evolutionary processes. A genome-scale approach to assess gene flow across hybrid zones is vital, and now possible. In Mexican towhees (genus *Pipilo*), several morphological hybrid gradients exist. We completed a genome survey across one such gradient (9 populations, 140 birds) using mitochondrial DNA, 28 isozyme, and 377 AFLP markers. To assess variation in introgression among loci, cline parameters (i.e. width, center) for the 61 clinally varying loci were estimated and compiled into genomic distributions for tests against three empirical models spanning the range of observed cline shape. No single model accounts for observed variation in cline shape among loci. Numerous backcross individuals near the gradient center confirm a hybrid origin for these populations, contrary to a previous hypothesis based on social mimicry and character displacement. In addition, the observed variation does not bin into well-defined categories of locus types (e.g., neutral v. highly selected). Our multi-locus analysis reveals cross-genomic variation in selective constraints on gene flow and locus-specific flexibility in the permeability of the interspecies membrane.

Key Words: gene flow, introgression, *Pipilo maculatus*, *Pipilo ocai*, towhee, tension zone model, hybridization, AFLP, isozyme, mitochondrial DNA

Introduction

The phenomenon of divergence in the face of gene flow and the role of hybridization in the evolutionary process are of collective interest (Barton 2001; Doebeli & Dieckmann 2003; Grahame *et al.* 2006; Strasburg & Rieseberg 2008). Hybridization between closely related species can inform our understanding of evolutionary processes on population through phylogenetic levels. Steep genetic clines across stable areas of hybridization are the classic examples of the "tension zone" model of hybrid zones, where the transition between parental forms is maintained by a balance between endogenous selection and dispersal, arbitrated by linkage disequilibrium (Barton & Hewitt 1985, 1989). In this framework, hybrids will, on average, be less fit than the parental types. Such fitness differences arise, in part, because the new recombinant genotypes of hybrid individuals have never been tested on the opposite genotypic backgrounds even though parental lineages have been subject to generations of selective pressure. 'Speciation genes' can deepen this valley of hybrid fitness through effects on reproductive success (Wu & Ting 2004). On the other hand, some recombinant genotypes may be advantageous in alternative environments, in which case even strong selection against early generation hybrids will not prevent the movement of these alleles across many hybrid zones (Moore 1977; Barton & Bengtsson 1986; Pialek & Barton 1997; Barton 2001; Strasburg & Rieseberg 2008). If hybridization is frequent and ongoing, neutral alleles may also leak across a hybrid zone; however, sufficiently strong linkage disequilibrium can oppose such allelic escape (Gavrilets 1997). Thus, the relative freedom of any single allele to introgress is related to the strength of selection on, and linkage disequilibrium to, favored or disfavored alleles.

Research focusing on divergent parental species and diagnostic markers accentuating these differences may underestimate levels of interspecies introgression (Brumfield *et al.* 2001). The classic assertion that avian hybrid zones, in particular, 'swallow' the interspecies genetic mixing via strong selection against hybrid offspring may arise because many of the well-studied hybrid zones are between highly differentiated species exhibiting steep clinal variation (Moore 1977). Although the conscious choice of diagnostic markers facilitates identification of parental types and hybrids, it may also perpetuate a focus on markers and traits under strong purifying selection (Sattler & Braun 2000; Brumfield *et al.* 2001). This level of selection may not be representative of the rest of the genome, which may experience greater gene flow. To provide a more comprehensive portrait of introgression across hybrid zones, a dense genome survey is required (Rieseberg *et al.* 2003). Such a survey, which allows inference of cline shape parameters from spatial changes in allele frequency across hybrid transects for many loci, can provide estimates of the degree to which introgression varies among loci across the genome.

Divergence despite gene flow between incipient or young species may be an important part of the speciation process (Hey 2006). Empirical evidence of introgression due to hybridization is growing. Allelic introgression between species has been demonstrated in several systems including *Drosophila,* butterflies, mice and reef fish (Wang *et al.* 1997; Crow *et al.* 2007; Mallet *et al.* 2007; Teeter *et al.* 2008; Gompert *et al.* 2010a; Gompert *et al.* 2010b). Among birds, mitochondrial and nuclear markers suggest that cryptic introgression between golden-winged and blue-winged warblers is common (Vallender *et al.* 2007). Introgression of plumage and microsatellite markers

has occurred across a Panamanian manakin hybrid zone (Brumfield *et al.* 2001; Yuri *et al.* 2009). Several avian hybrid zones are moving spatially where one species is encroaching upon the other through hybridization and introgression. These include black-capped and Carolina chickadees, Townsend's and hermit warblers, lazuli and indigo buntings, and blue-winged and golden-winged warblers (Sattler & Braun 2000; Rohwer *et al.* 2001; Owen-Ashley & Butler 2004; Bronson *et al.* 2005; Vallender *et al.* 2007; Carling & Zuckerberg 2011).

In contrast to the moving hybrid zones, stable regions of hybridization between two species of towhee, *Pipilo maculatus* (spotted towhee) and *P. ocai* (collared towhee), in Mexico have been intensively delineated using morphological characters (Sibley 1950, 1954; Sibley & West 1958; Sibley & Sibley 1964; Braun 1983) and isozyme loci (Braun 1983). Characterization of specimens dating back to the nineteenth century affirms that morphological intermediacy within these towhee populations has been geographically stable over a period of at least 150 years (Sibley 1950; Braun 1983).

P. maculatus and *P. ocai* are strikingly different in plumage and have substantial mitochondrial DNA (mtDNA) sequence divergence (~5.4% in cytochrome *b*, ND 2, and the control region) (Zink *et al.* 1998). In general, *P. ocai* prefer higher elevation (~3000m – 3700m), cooler, moister, coniferous (fir/pine forest) habitat, whereas *P. maculatus* prefer a lower elevation (~2100m – 3300m), warmer, drier, more open brushy pine/oak forest mixture. Sibley describes several major axes of hybridization: 1) the Teziutlán gradient running north – south down the Sierra Madre Oriental, 2) the Transplateau gradient running west – east across the transvolcanic belt of central Mexico, and 3) a smaller gradient running north – south along the southern edge of the Sierra

Madre Occidental (Sibley 1950). These areas of hybridization are not simply intrusions of parental types into shared habitat, but rather extensive hybrid swarms where each population in the gradient demonstrates intermediate plumage, song, and behavior (Sibley 1950, 1954; Sibley & Sibley 1964; Braun 1983). Remarkably, in addition to the three hybrid axes, there are sympatric populations of *P. ocai* and *P. maculatus* on Mt. Orizaba and south into Oaxaca, just tens of kilometers away from two of the gradients, which bear little morphological evidence of hybridization (Sibley 1950, 1954).

The towhee species may have diverged in allopatry due to different Pleistocene glacial refugia, and come back into contact sometime in the last 10,000 years. The re-established contact may predate the most recent glacial retreat and the two lineages may have been in and out of contact in a cyclical fashion. Although anthropogenic impact on forest habitat may have contributed to reestablishing contact between the two species, the area of species contact has been under anthropogenic influence for at least 3000 – 5000 years (Sibley & Sibley 1964). Thus, the contact and subsequent hybridization between *P. maculatus* and *P. ocai* is likely at least 3000 years old, and probably older. Given the several stable areas of hybridization, why hasn't gene flow swamped the species differences? Is selection acting in a classic tension zone manner to constrain the interspecies exchange of genes, or are some alleles free to flow between species despite the retention of other genetic differences? The towhee hybrid contact offers a great system in which to address such evolutionary questions, as these hybrid gradients are often proffered as a classic case of the breakdown of reproductive isolation (Mayr 1963).

Given the presence of sympatric towhee populations near the hybrid gradients, it has also been theorized that the morphological variation demonstrated among these

towhee populations is not related to hybridization at all. The "character displacement/social mimicry hypothesis" posits plumage character displacement causing enhanced differences among sympatric populations of *P. maculatus* and *P. ocai,* while allopatric populations are free from such selective divergence and the plumage differences are weakened or lost entirely (Brown & Wilson 1956). In these allopatric populations, social mimicry is hypothesized between towhees and the *ocai*-like chestnut-capped brush finch (*Arremon brunneinucha)*, causing convergence of plumage patterns between either *Pipilo* species and *A. brunneinucha* in order to increase interspecies territoriality and reduce competition for resources between these ecologically similar birds (Cody & Brown 1970). Thus, this hypothesis predicts that *P. maculatus* populations in allopatry with *P. ocai* are not only free to drift to a more *ocai*-like appearance due to release from character displacement, but driven to a more *ocai*-like appearance by territorial niche-partitioning with *A. brunneinucha,* consequently resembling hybrids between *maculatus* and *ocai.* In this case, rather than mixed genetic signatures in a gradient across the morphological areas of apparent hybridization, we would expect to see *P. maculatus*-like genotypes even as the appearance of towhee populations shift towards *ocai*-like.

Here, we assess genome-wide variation in permeability across the interspecies genetic membrane. We will test the contrasting predictions expected under the interspecies hybridization and character displacement/social mimicry hypotheses. We also aim to assess the genomic distribution of cline parameters across the hybrid interface and the variation in signature of the effects of selection and introgression. We hypothesize there is not one universal genome-wide clinal model, but diversity in cline

extent and shape demonstrating substantial variance in permeability of the species boundary. To evaluate the heterogeneity of underlying factors influencing the system, we will test the extent and strength of a cohesive cross-locus signal by attempting to pinpoint natural groupings among loci.

Methods

Sampling

Nine populations (140 individual birds, 15.56 mean sampled per population) were sampled across the 1156 km Teziutlán transect through a ribbon of Sierra Madre Oriental montane forest (Figure 1.1, Table 1.1). Collection sites ranged from 1956m to 3781m in elevation, exhibiting pine/oak scrub habitat at lower elevations and pine/fir habitat at higher elevations. Birds were collected using mist nets or shotgun, tissue was collected for cryopreservation, and vouchered museum skin specimens were prepared (Braun 1983). The morphological variation in these samples encompasses the full range from pure *P. maculatus* to pure *P. ocai* (Braun 1983) based on Sibley's six character hybrid index (Sibley 1950, 1954). Tissues were deposited at the US National Museum of Natural History and vouchers were deposited at Louisiana State University Museum of Natural Science (Supporting Information Table 1.1).

Molecular analyses

DNA was extracted from tissue using a standard proteinase K/phenol-chloroform protocol similar to that of Sattler and Braun (2000). DNA concentration was quantified using a Nanodrop ND-1000 spectrophotometer, and chain length was assessed using agarose gel electrophoresis.

Mitochondrial DNA

Mitochondrial DNA haplotypes were scored from Southern blots (Southern 1975) as described by Sattler and Braun (2000). Briefly, total genomic DNA was digested with restriction enzymes and the fragments separated by size in agarose gels. The digests were transferred to nylon membranes and mtDNA fragments for each individual visualized by hybridization with radiolabelled mtDNA from House Finch (*Carpodacus mexicanus*). An initial survey of 31 restriction enzymes revealed that 12 enzymes showed diagnostic differences between parental populations of *P. ocai* and *P. maculatus*. Three diagnostic enzymes (*Bcl I*, *Bgl I*, and *Hae III*) were then used to score all individuals.

Isozymes

Protein electrophoresis was completed on homogenized muscle, heart, and liver tissue according to Braun (1983) yielding 39 isozyme loci. When loci exhibited more than two alleles, clinal patterns were visualized using the allele demonstrating the greatest variation in frequency and/or the greatest frequency in population 1.

AFLP

The AFLP protocol was run according to Vos *et al*. (1995) with modifications according to Kingston and Rosel (2004). The restriction enzymes used were *Taq*I and *Eco*RI. Digestions were loaded with 200ng of genomic DNA (50uL total digestion reaction volume). Ten selective primer combinations were utilized (Table 1.2). Selective *Eco*RI primers were labeled with fluorescent dye and fragment sizes were assessed on an ABI 3100 genetic analyzer via capillary electrophoresis with an internal size standard. Electropherograms were visualized and fragment presence/ absence scored in ABI

GeneMapper 4.1 software. To avoid artifacts, the Kingston and Rosel (2004) scoring protocol was utilized.

Analyses of cline position and shape

Population allele frequencies of AFLP markers were inferred from the dominant binary data using the assumptions of Hardy-Weinberg equilibrium (HWE; frequency of the null state, 0, represents q^2). A simulation using the co-dominant isozyme data was completed to assess the least-biased allele frequency estimation method for dominant markers in the towhee system. Simulated dominant markers were re-sampled based on the empirical population allele frequencies of the 28 polymorphic isozyme loci. Allele frequencies were estimated using the HWE square root method and Bayesian inference through the program Hickory (Holsinger & Lewis 2003). These allele frequencies were then compared to the co-dominant data from which the dominant scores were derived. HWE square root method provided the least biased estimates (mean [± s.d.] difference between co-dominant allele frequency and frequency estimated from dominant scores was 0.015 [± 0.032]). In order to measure differentiation between parental species, locus-specific F_{ST} estimates involving the parental populations only (pop 1, pop 2, pop 8, pop 9) were calculated for the mtDNA locus (FSTAT 2.9.3.2) and AFLP loci (DFDIST). In order to assess population differentiation across the entire hybrid transect, F_{ST} for each locus was calculated (across all populations) in FSTAT 2.9 (mtDNA and isozyme loci) and DFDIST (AFLP loci).

We defined as clinal those loci that demonstrated a 20% or greater allele frequency change across all populations. Cline parameters were estimated according to a likelihood model (Szymura & Barton 1986) using the program Analyse 1.3 (Barton &

Baird 1998). The program utilizes a sigmoid central curve and independent exponential decay curve tails to infer eight cline parameters of interest: center, width, splice points for the decay of each tail, the rate of decay of each tail, and maximum and minimum allele frequencies at each end of the transect (Brumfield *et al.* 2001). The fit of each locus to the inferred model is assessed via likelihood. Each locus was run through a 10,000 step Metropolis-Hastings fit optimization process 10 separate times to assure convergence of parameter estimates. Model equations and likelihood calculations are as described in Brumfield *et al*. (2001), although there was a typographical error in their Equation (3), corrected here in Supporting Information.

Histograms of cline widths and centers estimated from the mtDNA, isozyme, and AFLP loci were compiled. Means and variances of the isozyme and AFLP parameter distributions were compared using ANOVA and F-tests, respectively. The sampling distribution of the cline parameters was explored with bootstrap model fits of three representative loci (mtDNA, TRI2, and AFLP 09_322); 10,000 additional Metropolis-Hastings iterations were run from which all likelihood values within two units of the maximized fit were sampled to create parameter confidence limits for each model. The representative loci were chosen as exemplary models due to the characteristics of their model parameters: the mtDNA locus demonstrates a steep, sharp cline much like one that would result from intense selective pressure against hybrids. In fact, because the heterogametic sex in birds is female, if any locus were to reflect Haldane's rule, it would be the maternally inherited mitochondrial locus. The AFLP locus 09_322 was chosen to represent a 'median model' because it exemplifies the observed median value for width among all loci and is very close to the median value for center. The TRI2 locus was

selected as a 'wide-shifted' exemplar. The sampling distribution of parameters for such different models might vary significantly from each other (and should cover much of the variation in the data), hence the choice of three, rather than a single exemplary model. The bootstrap resampling of the representative loci does not infer that one particular type of selection is responsible for any pattern; it instead provides a statistical method for assessing the variation associated with the characteristic cline shapes in order to test whether one type of model can account for all observed variation in the data.

Pearson's correlation coefficients were calculated to assess the relationship between cline width, cline center shift, and two types of locus-specific F_{ST} (calculated across parental populations and across all populations). Amplitude of cline center shift in either north or south direction from the mean was quantified using absolute value of [mean of all cline centers – locus cline center]. If loci that are highly differentiated between parental populations are more likely to exhibit reproductive incompatibilities between species, we should expect loci with large F_{ST} between parental populations to exhibit certain cline parameters (like narrow width). The second transect-wide estimate of F_{ST} across all populations should better reflect the pattern of transition through the hybrid transect; cline width and transect-wide F_{ST} should be measuring the same underlying structure at each locus across the hybrid zone.

To test the universality of cline parameters, pair-wise likelihood fits of each locus' empirical allele frequency data were fit to each of the cline models inferred for other loci. Each locus was used as both model and empirical data, resulting in a full matrix of log likelihood fits. A full matrix of pair-wise log-likelihood fits for all loci was compiled (pair-wise fits calculated in R, http://www.r-project.org/). A principal co-

ordinates analysis was performed on the asymmetrical pair-wise similarity matrix to assess clusters of similar loci; a custom R script which averages the off-diagonal components before executing the principal co-ordinates analysis was used (Gower 1966). The first two principal co-ordinate axes were plotted to identify any clustering in two-dimensional Euclidean distance space.

Results

mtDNA

The mitochondrial locus is diagnostic in parental populations, with each individual bearing a fixed *maculatus* or *ocai* type pattern. This locus exhibits a prototypical steep clinal pattern, with cline width of 14.66 km and cline center at 841.54 km between populations 5 and 6, near the morphologically defined center of the hybrid zone (Braun 1983).

Isozymes

Of the 39 isozyme loci, 28 are polymorphic and nine qualify as clinal, with none showing fixed differences. Mean cline width is broader than that for mtDNA at 197.29 km (range 40.33 km – 450.57 km) while mean cline center is at 841.47 km (range 728.00 km – 976.34 km).

AFLP

Ten primer combinations render 377 polymorphic loci. Fifty-one of the AFLP loci qualify as clinal, with six showing fixed differences between at least two of the four parental populations. Mean cline width is 298.47 km (range 3.46 km – 3303.57 km) and mean cline center is at 821.04 km (range 567.84 km – 1145.39 km). Many of the

individual birds demonstrate genetic intermediacy and backcross genotypes among the fixed markers, consistent with hybridization underlying the observed morphological variation rather than character displacement theory (Supporting Information, Figure 1.1).

Analyses of cline position and shape

Model fits to each of the loci reveal extensive variation in cline position and shape (Figure 1.2). Histograms of the inferred parameters cline width and cline center for all 61 clinal loci (mtDNA locus + 9 isozyme loci +51 AFLP loci) demonstrate this variation quantitatively (Figure 1.3). Analysis of variance reveals no significant difference between isozyme and AFLP loci for either mean cline width or mean cline center. However, the variances of the cline width distributions differ significantly between isozyme and AFLP data due to the longer tail in the AFLP distribution consisting of loci corresponding to wide clines (F-test, $p = 0.0008$). The variances of the center distributions do not differ significantly. The pooled cline center distribution is normal (Shapiro-Wilk W p=0.08, D'Agostino Kurtosis p=0.08, D'Agostino Skewness p=0.67).

Differentiation among parental populations (parental F_{ST}) is not significantly correlated with either cline center shift or width. Cline center shift and cline width are significantly positively correlated; cline width and transect-wide locus-specific F_{ST} are significantly negatively correlated (Table 1.3). Wider clines tend to be more highly shifted from the hybrid zone center and tend to have lower cross-transect F_{ST} values; the lower levels of population differentiation and wider clines are reflecting the same underlying structure across the hybrid transect. The great majority of the measured loci do not exhibit fixed differences among species. Many clines that exhibit relatively low change in empirical allele frequency (Δp) also exhibit quite steep clines (Supporting

Figure 1.3). Although it is inherent that F_{ST} be correlated with Δp, it is not inherent that cline width should be correlated with Δp (and indeed it is not significantly correlated, Pearson's correlation coefficient= -0.16, p = 0.21). Because Δp and cline width are not confounded in this data set of mostly non-fixed loci, and the cline width parameter is independent of the inferred p_{max} and p_{min} parameters (Supporting Information), a system-wide correction for Δp is not warranted.

Three exemplary models spanning the range of inferred parameters reveal that sampling error alone cannot explain the observed variation among loci in cline extent and shape. Confidence limits (95% CI) for the exemplary models for cline center (*c*) and width (*w*) are: 833.07 km < steep model *c* < 845.87 km, 0.60 km < steep model *w* < 34.56 km (4612 model fits within 2 log-likelihoods of the maximized fit); 755.47 km < median model *c* < 848.23 km, 6.92 km < median model *w* < 418.90 km (9147 model fits within 2 log-likelihoods of the maximized fit); 843.67 km < wide-shifted model *c* < 1261.03 km, 97.96 km < wide-shifted model *w* < 1135.43 km (6944 model fits within 2 log-likelihoods of the maximized fit). All sets of confidence limits are plotted in relation to all loci on Figure 1.3. Only 10% of all loci center parameter estimates (6 loci) fall inside the steep model center support limits, 37% of all loci center parameter estimates (22 loci) fall inside the median model center support limits, while 42% of all loci center estimates (25 loci) fall inside the wide-shifted model support limits. Thirty percent of all loci width estimates (18 loci) fall inside the steep model center support limits, 65% of all loci width estimates (39 loci) fall inside the median model width support limits, while only 43% of all loci width estimates (26 loci) fall within the wide-shifted model support limits.

The principal co-ordinates analysis of pair-wise model fits among all loci does not demonstrate any clearly defined groupings among the loci like one would expect if groups of loci were governed similarly by only a few sets of constraints. On the two-dimensional principal co-ordinates plot, PC1 accounts for 67.34% of the variance while PC2 accounts for 8.62% (Figure 1.4). In order to summarize empirical and model fit variation among loci, we plot graphs of each locus' empirical cline data (Supporting Information, Figure 1.2) as well as present model-inferred cline parameters for all loci (Supporting Information Table 1.2).

Discussion

Our multi-locus analysis demonstrates definitively that the morphological variation observed in this classic system is due to hybridization and not to release from character displacement or social mimcry between *P. maculatus* and *P. ocai*. If character displacement were driving the observed patterns, intermediate populations should be genetically like *maculatus* (Brown & Wilson 1956; Braun 1983). Instead, the populations in the middle of the Teziutlán gradient reveal extensive genetic admixture, as expected with hybridization (Supporting Figure 1.1). Clarification of the hybrid status of these populations allows us to focus on the remarkable features of this system, which has been viewed as a classic case of the breakdown of reproductive isolation (Mayr 1963).

With a broad sampling of loci in hand, it is clear that significant variation exists among the clinal loci across the towhee hybrid gradient. The archetypal steep cline model cannot describe change in gene frequency across space for all loci. Even a median or wide-shifted model cannot fully encompass alone the magnitude of variance observed

in these data. In addition, the heterogeneity observed lacks a discrete number of empirically evident groups, suggesting the underlying forces driving this variation do not have a strong cohesive cross-locus signal (Figure 1.4). We demonstrate not only significant variation in cline model parameters and corresponding clines shapes among loci across the genome, but also a negative correlation between wide clines and locus-specific differentiation (transect-wide F_{ST}) across the gradient. Similar lack of co-incidence among cline centers has been observed in the North American hybrid zone between glaucous-winged gulls and western gulls; however, the extent of variation in cline width we see here is not matched in the gull hybrid zone (Gay *et al.* 2008).

While the mitochondrial locus falls near the mean and mode in center and width parameters, we see substantial variance in both the isozyme and AFLP loci (Figures 1.2 and 1.3). Similar nuclear variation from the mitochondrial signal has also been demonstrated in a patchy butterfly hybrid zone (Gompert *et al.* 2008). While many towhee loci exhibit cline centers shifted from the mean, there is no bias in direction of shift as one would expect in a mobile rather than stable hybrid zone (Figure 1.3). The lack of correlation between parental F_{ST} and cline width suggests that the more highly differentiated loci are not solely representing highly constrained loci (likely to exhibit incompatibilities). This pattern may result from the observed loci existing in equilibrium state due to the age and stability of the hybrid contact; over time, breakdown of linkage disequilibrium may reduce the effect of incompatibilities. In addition, some loci highly differentiated among parentals may reflect patterns created by demographic-oriented processes across the hybrid transect, like isolation by distance.

The loci that exhibit the classic sharp cline morphology are likely constrained against gene flow due to the balance between dispersal and selection. Much like the mitochondrial locus, plumage characters in the towhee Teziutlán gradient exhibit steep transitions and are likely under strong selective pressure (Braun 1983). In the case of the mitochondrial locus, Haldane's rule could contribute to this strong transition, as avian females are the heterogametic sex (Haldane 1922). However, there is clear cross-genomic variation in selective constraint and locus-specific flexibility in the permeability of the interspecies membrane (Figure 1.3). Both exogenous and endogenous selective factors could be important driving factors for these loci; coupling between endogenous and exogenous barriers to gene flow even in the presence of genetic-environment association has been demonstrated theoretically (Bierne *et al.* 2011). Sex-linked loci are often focal in hybrid zone studies as they are likely to be under greater selective pressure (Payseur *et al.* 2004; Macholán *et al.* 2007; Carling & Brumfield 2008). While our analyses do not elucidate a tightly-associated group of loci with narrow, steep clines, the sex-linkage status of our anonymous loci warrants further exploration.

The lack of natural groupings of loci demonstrated by the principal co-ordinates plot suggests that while we have sampled a cross-section of among locus variation, loci do not naturally fall into easily defined categories based on similarities in cline extent and shape. Differentiation and gene flow across the genome is not uniform nor easily binned into a small number of categories. Even though the hybrid zone itself appears geographically stable, some alleles are free to introgress across species while others are highly constrained. This lack of uniformity in cline structures suggests that among-locus variation in introgression is high but not categorical, which is relevant to the 'islands of

differentiation' concept (Harr 2006; Nosil *et al.* 2008; Feder & Nosil 2010; Nadeau *et al.* 2012). If loci important to maintaining reproductive isolation between these two species were tightly linked and clustered into a small number of genomic islands, we would expect the observed variation to exhibit at least some natural groupings. The lack of natural groupings suggests this system does not exhibit a strong signal for a small number of islands of differentiation. Although evidence of similar levels of heterogeneity in differential introgression is prevalent in the hybrid literature (Macholán *et al.* 2007; Carling & Brumfield 2008; Teeter *et al.* 2008; Yuri *et al.* 2009; Carling & Zuckerberg 2011; Dufkova *et al.* 2011), searching for clusters of loci (or lack thereof) in the absence of a genome or linkage map is a unique aspect of this analysis.

Given the likely age of the secondary contact between *P. maculatus* and *P. ocai*, one might hypothesize that the less constrained alleles would sweep across the hybrid zone, erasing some portion of the original differentiation between these taxa. Although the movement would occur very quickly for advantageous alleles, gene flow could eventually bring across neutral alleles as well (Macholán *et al.* 2007). So why have not all the unconstrained alleles swept across the zone and further increased homogeneity between the two species? One possibility is that local variation in habitat and the island-like distribution of habitat patches along the hybrid gradient contribute to local differentiation. Drift and differential selection within semi-isolated populations complicates long distance gene flow across the species. Differentiation in hybrid populations is often observed in mosaic hybrid zones, but is less commonly emphasized in classic hybrid transects where greater landscape continuity may facilitate gene flow (Vines *et al.* 2003; Mallet *et al.* 2007; Yuri *et al.* 2009; Gompert *et al.* 2010b). Barriers

to habitat continuity show the potential to stifle introgression in a manakin hybrid zone in Panama (Brumfield 1999; Stein & Uy 2006). Habitat patchiness related to elevation and availability of scrubby pine/oak forest could influence local population isolation and corridors of gene flow in the Mexican towhee hybrid zone.

Clines within the *P. maculatus – P. ocai* Teziutlán hybrid gradient reveal cross-genomic heterogeneity indicating significant locus-to-locus variation in the porosity of the species membrane. The presence of loci in many phases of constrained and unconstrained gene flow suggests that even within an old, stable area of contact, differential introgression is an ongoing fluid and dynamic process. This potential for evolutionarily significant differential introgression across species boundaries bears importantly on our interpretation of species concepts. Incomplete reproductive isolation and divergence with gene flow may not just be relevant in the context of incipient speciation, but also for older lineages in and out of contact cyclically (Broyles 2002; Gay *et al.* 2008). While regular instances of introgressive hybridization may add a level of variance to the classic and beautiful model of bifurcating lineages, it also reaffirms our need for extensive multi-locus datasets for inference in both population and phylogenetic fields.

Acknowledgments

I thank Stephen Braun, Nicholas Lanson, Nimrod Funk, and George Oliver for their able assistance in the field. We thank our funding sources for this project: University of Maryland Program in Behavior, Ecology, Evolution, and Systematics, Smithsonian Ornithology, Smithsonian Institution National Museum of Natural History Vertebrate

Zoology. NSF grants DEB 0228675 and DEB0733029, and the NMNH Frontiers in Phylogenetics Program provided research assistantships. I thank Michael Braun, William Fagan, and Robert Jernigan for discussion and comments on the manuscript, as well as theoretical and analytical support. Appreciation goes to Gerald Wilkinson, Kevin Omland, David Hawthorne, and Matthew Hare for comments on the manuscript.

Table 1.1. Local area, plumage type, mean population hybrid index (plumage), coordinates, location elevation, length along the transect, and sample size of sampling locations ordered from north to south through the Teziutlán hybrid gradient.

Number	Location	Plumage Type	Hybrid Index	Lat	Long	Elevation (m)	Km from no. 1	Sample N
1	Coahuila	*Pipilo maculatus*	24.0	25.250	-100.450	2902	0	16
2	Queretaro	*Pipilo maculatus*	22.8	20.783	-99.567	3063	538	17
3	Tlaxco	hybrid	21.5	19.667	-98.167	2667	728	17
4	E. Huichautla	hybrid	17.8	19.783	-97.600	2355	793	14
5	Teziutlán	hybrid	14.9	19.817	-97.367	1956	818	15
6	R. Palenquillo	hybrid	4.4	19.650	-97.117	2383	846	16
7	Cofre Perote	hybrid	3.4	19.567	-97.100	3028	856	15
8	Orizaba	*Pipilo ocai*	0.2	19.050	-97.308	3781	917	16
9	Oaxaca	*Pipilo ocai*	0.0	17.167	-96.633	2760	1156	14
	total							140

Table 1.2. Adapters and primers used in the AFLP assay

oligo name	**Sequence**
EcoRIadp_F	CTCGTAGACTGCGTACC
EcoRIadp_R	AATTGGTACGCAGTCTAC
TaqIadp_F	GACGATGAGTCCTGAC
TaqIadp_R	CGGTCAGGACTCAT
EcoRIpresel	GACTGCGTACCAATTCA
TaqIpresel	GATGAGTCCTGACCGAA
EcoRI+AAC	GACTGCGTACCAATTCAAC
EcoRI+AAG	GACTGCGTACCAATTCAAG
TaqI+AAC	GATGAGTCCTGACCGAAAC
TaqI+AAG	GATGAGTCCTGACCGAAAG
TaqI+ACA	GATGAGTCCTGACCGAACA
TaqI+ACT	GATGAGTCCTGACCGAACT
TaqI+AGA	GATGAGTCCTGACCGAAGA

Table 1.3. Pearson correlation coefficients (lower diagonal) and associated p-values (upper diagonal) for pair-wise comparisons of cline center shift, width, and both parental and transect-wide F_{ST}.

	center shift	width	parental F_{ST}	transect-wide F_{ST}
center shift	1	*0.010*	*0.600*	*0.528*
width	0.329	1	*0.227*	*0.043*
parental F_{ST}	0.074	-0.170	1	*1.5 e-11*
transect- wide F_{ST}	-0.082	-0.260	0.775	1

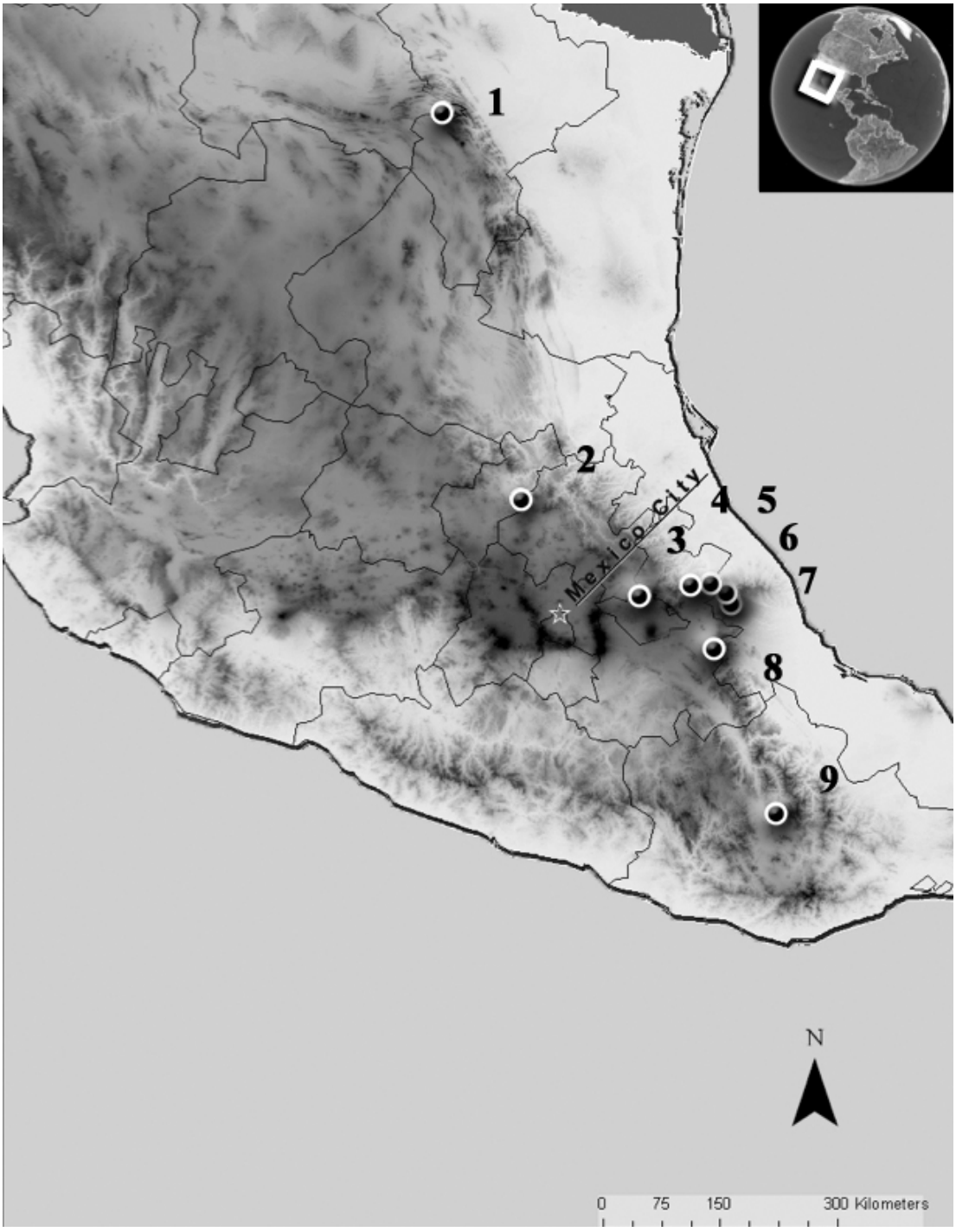


Figure 1.1. Sampling locations of towhee populations along the Teziutlán hybrid gradient in the Sierra Madre Oriental. Numbered locations are detailed in Table 1.1.

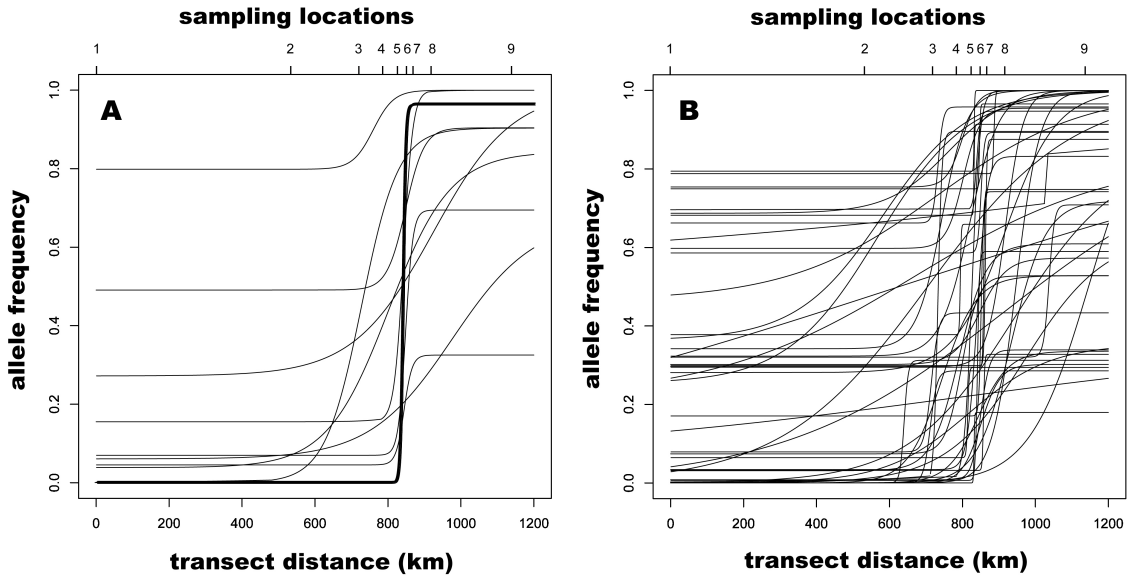


Figure 1.2. Cline models fit to empirical data for all 61 clinal loci using Analyse 1.3. Panel A shows the mtDNA (bold) and 9 isozyme loci, while panel B shows the 51 AFLP loci. Sampling locations are noted on the secondary X-axis starting with population 1 at 0 km transect distance.

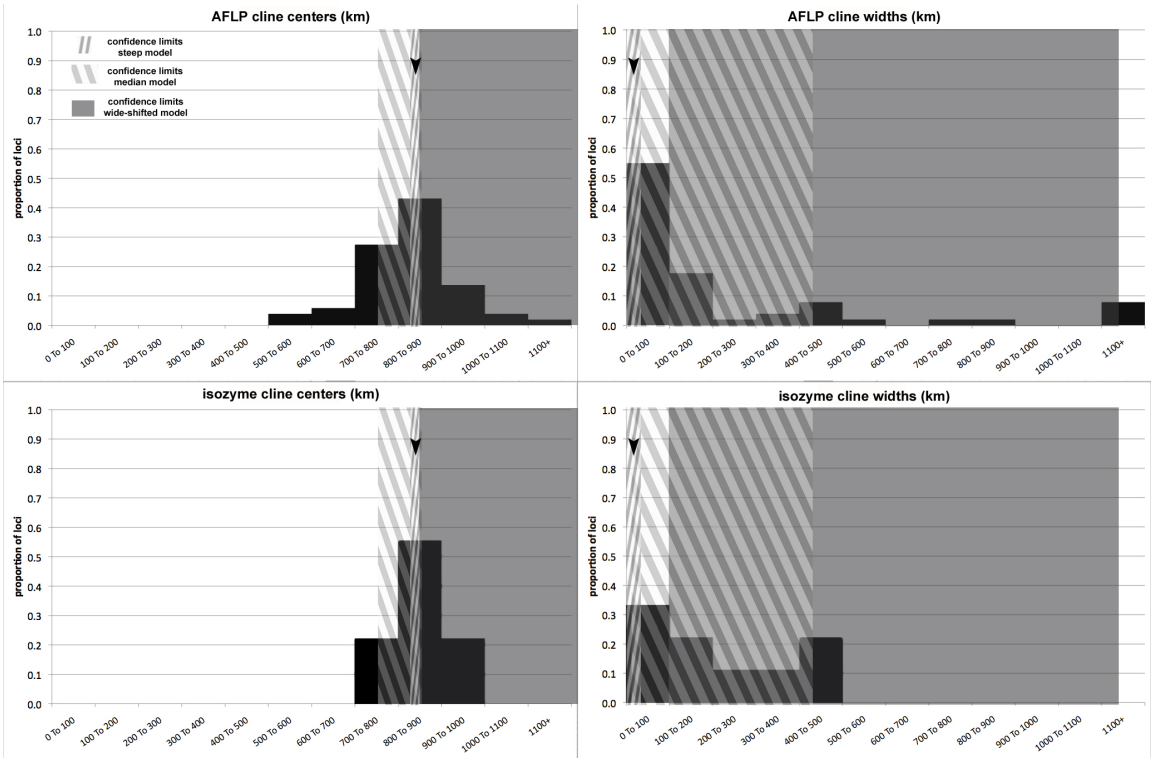


Figure 1.3. Distributions of cline centers and widths for AFLP (51) and isozyme (9) loci. Mitochondrial locus position is indicated by arrows. 95% confidence intervals for the bootstrap support on three empirical models (steep, median, and wide-shifted) are indicated by shading or cross-hatching.

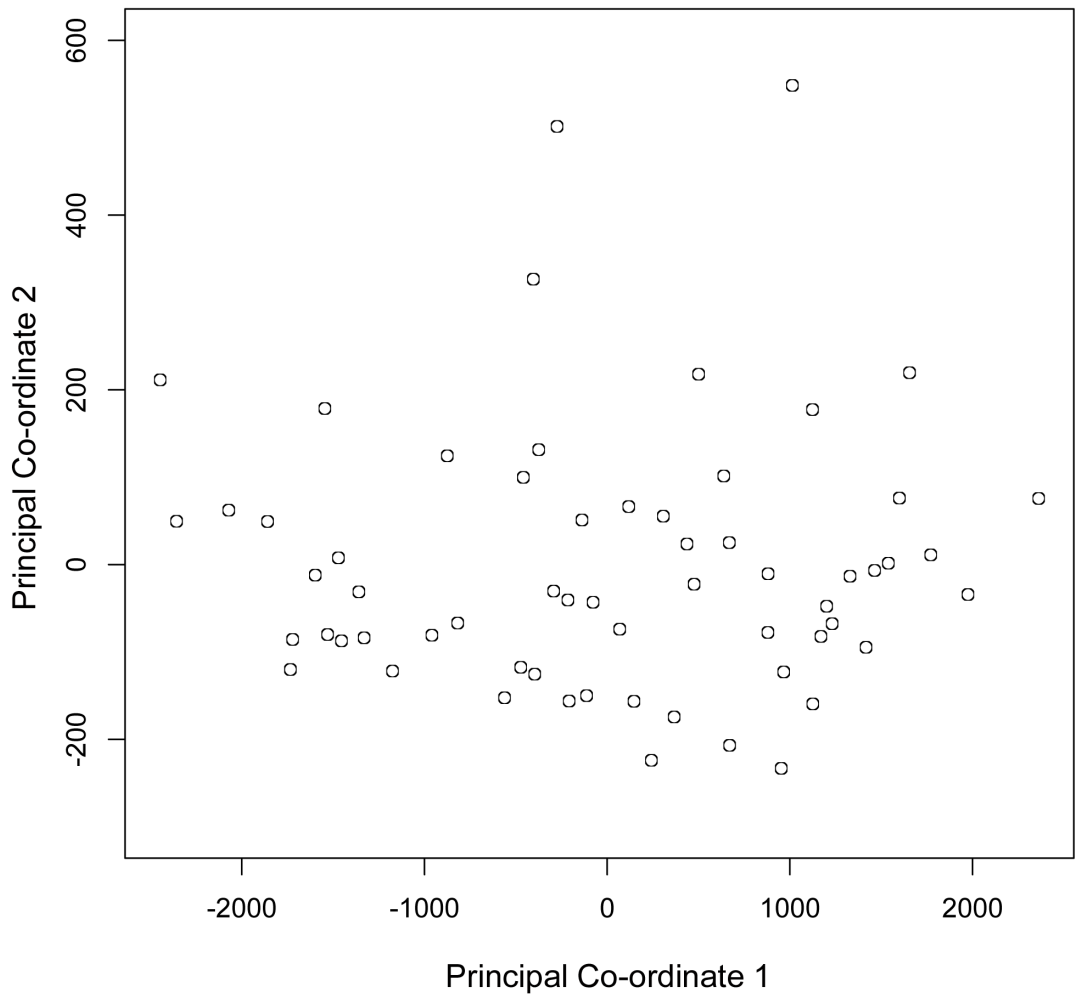


Figure 1.4. Principal co-ordinates in two dimensions performed on the asymmetrical log likelihood model fit matrix of the 61 different clinal loci.

Chapter 2: Genetic differentiation and habitat connectivity in hybrid zones

Abstract

Hybrid zones offer a natural experimental setting in which to assess the dynamics of reproductive isolation. Secondary contact zones, in particular, present a partial severance of the original divergence due to selection and/or genetic drift and the genetic architecture of reproductive isolation. In an area of hybridization between two Mexican towhee species, *Pipilo maculatus* and *P. ocai*, two major hybrid gradients and a region of sympatry provide natural replicates for comparison. Population genetic analyses indicate divergence between geographically disparate parental types (in both species) as well as divergence of populations within the hybrid zone. The two hybrid transects (Teziutlán and Transvolcanic) are distinct and evidence suggests allelic introgression across the species boundary and between the two transects. In the sympatric populations (Mt. Orizaba and Oaxaca) where morphological evidence for hybridization is scarce, introgression of opposing species' alleles appear to traverse through the hybrid zones rather than from cryptic local hybridization. Habitat corridors with increased dispersal potential indicate hotspots of connectivity where the two transects meet. Both habitat connectivity and genetic differentiation between geographically disparate parental types appear to influence the dynamics of gene flow across the hybrid gradient. Environmentally-mediated gene flow in the context of secondary contact and hybridization is an important force influencing evolutionary trajectory.

Keywords:

gene flow, habitat connectivity, hybrid zone, introgression, dispersal

Introduction

Hybridization offers a window into speciation processes and the mechanics of isolation (Arnold 1997; Rieseberg *et al.* 1999; Seehausen 2003; Mallet 2005; Arnold 2006; Gompert & Buerkle 2009). Hybrid zones are natural laboratories in which we may investigate the genetic basis of species divergence and analyze ecological factors that may contribute to that divergence. The concept of divergence with gene flow is not only important in the context of incipient speciation but also for lineages cyclically in and out of secondary contact (Nosil 2008; Nosil & Feder 2012; Kingston *et al.* in review). As lineages separate and reconnect due to climate change and fluctuations in habitat connectivity, it is important to understand what factors contribute to the maintenance, or demise, of genetic differences between closely related lineages. In fact, secondary contact hybrid zones may be a particularly powerful tools for investigating reproductive isolation due to the separation in time and space of the initial forces driving divergence and the current selective landscape (Gompert *et al.* 2012).

Spatially explicit tools recently adapted in the field of landscape genetics tender new power in determining correlations between genetic and environmental factors. Applying these novel techniques to natural systems such as hybrid zones can provide great insight into the genomic structure of hybridization and interspecies gene flow, as well as the critical relationship between habitat and diversification (Brumfield *et al.* 2001; Hanotte *et al.* 2002; Manel *et al.* 2003; Holderegger & Wagner 2006; Cicero & Johnson 2007; McRae & Beier 2007; McRae *et al.* 2008; Balkenhol *et al.* 2009). Determining the importance of current environmental factors on gene flow can help isolate dynamics confounded within the genomic and spatial landscapes (Andrew *et al.*

2012). Influence of extant selective and stochastic forces in a system can be layered on top of historical signals of isolation and divergence. Accounting for the present patterns associated with these environmental forces can help reveal underlying historical signals of isolation and divergence.

The extensive, stable regions of hybridization between two species of towhees, *Pipilo ocai* and *Pipilo maculatus*, in Mexico are a unique model system in which to explore questions concerning porous species boundaries and the interaction between ecological factors and hybridization. This system has a long, rich history of study (Sibley 1950, 1954; Sibley & West 1958; Sibley & Sibley 1964; Braun 1983). Two large axes of hybridization exist: a ribbon of montane habitat extending North-South along the Sierra Madre Oriental (Teziutlán gradient) and another patchier swath of habitat extending West-East along the island-like volcanic peaks of the Transvolcanic region (Transvolcanic gradient) (Sibley 1954). The two gradients offer comparative differences in altitude, habitat, seasonality, precipitation, and landscape fragmentation. Local populations in most areas of contact appear morphologically as panmictic hybrid swarms, not simple F_1 mixtures of dispersing parental individuals (Sibley 1954; Sibley & West 1958; Sibley & Sibley 1964). However, at Mt. Orizaba and south into Oaxaca the parental plumage type of *P. maculatus* re-appears and exists in sympatry with *P. ocai*. In both gradients, the species prefer high elevation pine/oak scrub habitat, and are frequently observed in agricultural edge habitat. The collared towhee, *P. ocai*, prefers higher elevations (2000 – 3000+m) and wetter, pine-centric habitat while the spotted towhee, *P. maculatus*, prefers more intermediate elevations (1500 – 3000m) and drier, oak-centric habitat (Sibley 1954; Sibley & West 1958; Sibley & Sibley 1964). The

species overlap significantly in habitat preference, however, and Sibley (Sibley 1954) noted that in places such as Mt. Orizaba they often exist "in the same thicket."

The hybrid transects of *Pipilo maculatus* and *P. ocai* enable investigations regarding levels of gene flow and differentiation across species boundaries and how the genetic components relate to heterogeneity in habitat connectivity across the areas of hybridization. Utilizing powerful multi-locus genetic markers, we aim to determine what genetic differentiation is being maintained in the face of current hybridization. How effective is current gene flow at promoting introgression? How do the genetic signals of local differentiation and gene flow between populations relate to habitat availability and connectivity?

Highly multi-locus approaches to identifying and quantifying gene flow across hybrid zones are essential (Rieseberg *et al.* 2003). This study utilizes a genome survey technique, AFLP, as well as more traditional DNA sequencing analyses. The natural 'treatment' levels of environmental variation found in the towhee system render these hybrid zones useful models in which to test the relationship between divergence, gene flow and habitat. What we learn from the towhee hybrid zone can inform analyses and questions on the origins and maintenance of biodiversity at larger taxonomic, geographic, and genomic scales.

Materials and Methods

Sampling

Both the Teziutlán (~1150km in length, 172 individuals) and Transvolcanic gradients (~760km in length, 288 individuals) were sampled to incorporate the two

largest axes of hybrid towhee variation in Mexico (Figure 2.1, Table 2.1, collected 1979, 1981, 2008, 2009). Birds were collected via mist net or shotgun, tissue samples were frozen and archived, and voucher museum specimens were prepared for each individual collected. Tissues and vouchers were deposited at Louisiana State University Museum of Natural Science and the US National Museum of Natural History (Supporting Table 2.2). Sampling location collection size was 24.21 ± 10.41 (mean ± standard deviation, target sample size for 1979 and 1981 was 15 individuals per site; target sample size for 2008-2009 was 30 individuals per site). Although the majority of the Teziutlán gradient was sampled twenty-five years prior to the Transvolcanic gradient, populations of both species at Mt. Orizaba were sampled during both time periods to assess stability of population genetic signature over time.

DNA extraction

DNA was extracted using phenol-chloroform preparation (Sambrook *et al.* 1989). The method was executed either manually or via AutoGen extraction robot. DNA was quantified using a Nanodrop ND-1000 spectrophotometer and DNA quality assessed via gel electrophoresis.

mtDNA sequencing

The mitochondrial ND2 locus was targeted using the primers L5215 and H6313 (Hackett 1996; Sorenson *et al.* 1999). The gene was amplified using Biolase *Taq* and buffer, 1.5mM $MgCl_2$ concentration, and the following cycling profile: 94ºC 30s, 50ºC 30s, 72ºC 60s, 30 cycles total with a single 94ºC initial soak step and a final 10 minute 72ºC extension step. Amplicon bands of 850bp in size were visualized on agarose gels. PCR product was purified with ExoSAP-IT (single step hydrolytic enzyme clean-up) and

cycle sequenced with ABI BigDye 3.0. Original amplification primers and additional internal primers were used for cycle sequencing to assure maximization of double stranded data. Sequences were generated on an ABI 3100XL Genetic Analyzer. Chromatograms were vetted, contigs assembled, and consensus sequences created in Sequencher 4.10.1.

Sequences were aligned by eye in Se-al 2.0. Haplotype redundancy was calculated in MacClade 4.08. A median joining network was calculated using Network 4.610 (for the network haplotypes were considered identical if resolution of ambiguous bases or missing data could render them so; proportion of missing data and ambiguous bases was < 0.04%).

AFLP

The AFLP assay was run according to Vos *et al*. (1999) with modifications by Kingston and Rosel (2004). Twelve selective primer pairs were utilized. Fragments were visualized via capillary electrophoresis on an ABI 3130 Genetic Analyzer and scored across all populations in GeneMapper 4.0. Global F_{ST} values across all AFLP loci for pairwise population comparisons were calculated in AFLP-SURV using Bayesian estimation of allele frequencies; locus-specific F_{ST} values were calculated in DFDIST.

Population genetic and sequence analyses

Mitchondrial sequences were analyzed in a coalescent context using Migrate-n to infer mitchondrial gene exchange across the hybrid transects (Hey 2004; Beerli 2006). Three scenarios of population subdivision were compared using Bayes Factors in Migrate-n (Table 2.2, calculated with Bezier Thermodynamic estimate of marginal likelihood) (Beerli & Palczewski 2010). These three scenarios include both transects and

all sampling sites and range from 4-8 subdivisions. For each population scheme, three replicate runs were completed through 1,000,000 MCMC iterations using 4 heated chains at 4 different temperatures with the following options selected: Bayesian inference, haploid sequence data, bidirectional migration estimates for all population pairs inferred, theta and migration priors derived from F_{ST}, heated chain swapping enabled, heated chain using default temperatures. Population schemes were selected to cover a range of biologically realistic parameters that were also computationally feasible. Pooling of populations was done based on *P. maculatus* and *P. ocai* haplotype group frequencies (Supporting Figure 2.1).

Multi-locus population structure across the hybrid transects (without *a priori* population designations and assumptions) was inferred using the hierarchical Bayesian modeling construct in the program *structure* (Pritchard *et al.* 2000; Falush *et al.* 2007). Rather than aiming to optimize results for a particular number of populations (denoted by the parameter K), analysis parameters were optimized on preliminary data (16 populations, 342 individuals, 477 markers) and a large window of K values were investigated (K = 2 – 12 with the following options for *structure*: burn-in 100,000 iterations, run 1,000,000 iterations, 10 replicates per K value, admixture model, infer lambda, infer alpha). To maximize the biological relevance of our efforts, we employed the Evanno method (Evanno *et al.* 2005) to pinpoint the K value with the largest rate of change in likelihood and also present results from a range of K values surrounding this 'optimal' K value. We adopted this approach because methods designed to pinpoint an 'optimal' K have been criticized as *ad hoc*, and we feel presenting results as a function of K is biologically meaningful (Evanno *et al.* 2005; Falush *et al.* 2007).

Habitat connectivity analyses

Habitat corridors were inferred through categorized land use maps of Mexico and the landscape ecology analysis employed in *Circuitscape* 2.0 (Shah & McRae 2008). *Circuitscape* utilizes electronic circuit theory to model connectivity through heterogeneous landscape. Preferred habitat (pine/oak high elevation forests and edge agricultural habitat) was coded as having high current/ low resistance in the *Circuitscape* framework, while matrix habitat (all other land-use types) was coded as having low current/ high resistance. Full pairwise models (each population sampling location acts as a ground and a source versus every other population) and the all-to-one model (each population sampling location acts as a current source iteratively while all other locations act as grounds) were implemented. Resistance distances between sampling locations were compared to pairwise measures of genetic population differentiation (F_{ST} measure via the program AFLP-surv 1.0). Significance of this relationship was assessed using a Mantel Test (R package ade4, mantel.rtest, 9999 iterations).

Results

Sequence/Coalescent analyses of mtDNA Variation

The mtDNA ND2 locus is 1018 bp in length; sequence from 313 individuals (9-32 individuals per sampling site) encompassing the complete sampling regime reveals 63 unique haplotypes. Distinct *P. maculatus* and *P. ocai* haplotype groups are separated by 39 bp changes (Figure 2.2). All morphological parental types exhibit a mtDNA haplotype matched to plumage type with one exception in each species, both at the Orizaba location. These two mismatched individuals were sampled 28 years apart, run in

different extraction and sequencing plates, both re-amplified and re-sequenced, so the mismatches likely represent true haplotype introgression rather than laboratory error. Neither individual exhibits an unusual signature of admixture in relation to the rest of their populations in the AFLP data. Each species also demonstrates a geographically isolated haplotype group: the Oaxaca site (TZ09) for *P. maculatus* and the western (Jalisco, TV01 and TV02) populations for *P. ocai*. All individuals sampled at these locations exhibit haplotypes unique to these groups, and these unique haplotypes are also not observed elsewhere across the sampling locations (Supporting Figure 2.1).

Probabilities calculated using Bayes Factors derived from the Migrate-n analyses (Beerli & Palczewski 2010) indicate that the mtDNA data better represent the five-population model (scheme 2, Table 2.2). Bi-directional estimates of migration between these pooled population groups indicate that the greatest mtDNA gene flow occurs between the northern half of the Teziutlán gradient and the central and eastern portions of the Transvolcanic gradient (Supporting Table 2.2). These populations include the areas where the two transects meet. However, the gene flow appears to be asymmetrical, running more heavily from the Teziutlán into the Transvolcanic. Secondarily, smaller migration estimates indicate low levels of asymmetric gene flow from the northern Teziutlán into the southern Teziutlán and from southern Teziutlán into the central/eastern Transvolcanic gradient. These last two estimates of gene flow cross species haplotype groups (Supporting Table 2.2).

Population genetic analyses

The AFLP assay renders 561 total polymorphic loci across 12 primer combinations. Only three of these loci exhibit fixed differences among pairs of parental

populations in either transect; none are fixed in both transects. Locus-specific F_{ST} across all populations varies from -0.02 – 0.85 (0.18 ± 0.25, mean ± standard deviation). Locus-specific F_{ST} comparing only cross-species parental populations is correlated across transects (Pearson's correlation coefficient = 0.52 p=0.001).

Pairwise population comparisons of global F_{ST} derived from all AFLP loci vary from 0.01 – 0.18 (0.07 ± 0.03, mean ± standard deviation). Average pairwise differentiation between *P. ocai* and *P. maculatus* populations was F_{ST} = 0.12. Intraspecific differentiation was also strong: Average pairwise F_{ST} between western and eastern *P. ocai* parental populations = 0.099 (TV01/TV02 v TZ08/TZ09) while northern and southern *P. maculatus* parental populations exhibit slightly lower levels of differentiation (average pairwise F_{ST} = 0.074, TZ01/TZ02 v TV09/TV10).

The Orizaba populations (both *P. maculatus* and *P. ocai*) sampled in 1981 and 2009 show no significant divergence (F_{ST} 95% confidence intervals for Orizaba comparisons are: *P. maculatus*, $-0.0163 < F_{ST} < 0.0010$; *P. ocai*, $-0.0131 < F_{ST} < 0.0029$; AFLP-surv, Bayesian estimation of allele frequencies, 10,000 iterations for significance test). The temporal samples are therefore pooled into single *P. maculatus* Orizaba and *P. ocai* Orizaba localities for the purposes of analyses.

Three types of differentiation (namely, marked transitions across each hybrid transect, population differentiation among geographically disparate within-species populations, and differentiation within each of the hybrid gradients) are all evident in the K=3 through K=7 *structure* analyses (Figure 2.3 and Figure 2.4). In the simplest model (K=3), species-level differentiation between *P. maculatus* and *P. ocai* as well as population-level differentiation between eastern and western forms of *P. ocai* are

apparent in the population signatures (Figure 2.3 and Figure 2.4). As the models move from simple to more complicated (K=4 through K=7), population differentiation between northern and southern *P. maculatus* and gradient-specific within hybrid population differentiation both appear (Figure 2.3 and Figure 2.4). All population lineage signatures span the full spatial domain of both hybrid gradients, although relative proportions vary greatly. The delta K values from the Evanno method (Evanno *et al.* 2005; Earl & vonHoldt 2011) reveal the largest rate of change in likelihood at K=3, but also a smaller peak in likelihood rate of change at K=7 (Table 2.3).

Habitat connectivity analyses

From the *Circuitscape* analysis of habitat connectivity, a high current signature along the high elevation pine/oak habitat suggests dispersal between populations may be funneled into narrow corridors constrained by habitat (Figure 2.5). The greater continuity of high current areas in the Sierra Madre Oriental suggests that the eastern Teziutlán gradient passes through more continuous ribbon-like habitat whereas the Transvolcanic gradient, though still connected, is patchier and more island-like. However, due to the elevational drop off at the eastern edge of the Sierra Madre Oriental toward the Gulf Slope, the Sierra Madre Oriental/ Teziutlán corridor is compressed laterally into a smaller width, while the Transvolcanic habitat offers a more diverse selection of possible dispersal paths, even if each shows slightly greater individual resistance. The resistance distances inferred along these corridors between populations significantly correlate with pairwise measures of population differentiation (F_{ST}) inferred from the genomic AFLP data (Figure 2.6, Mantel test r = 0.32, p-value = 0.03). When comparisons involving parental populations are removed, the correlation between resistance distances and F_{ST}

improves (Mantel test r = 0.57, p-value = 0.0017). The northernmost population (TZ01) exhibits high overall resistance distance values despite a large range of F_{ST} values. The southern *P. ocai* populations (TZ08 and TZ09) reveal an over-representation of high F_{ST} values whereas the western *P. ocai* parental type (TV01) shows a slight excess of lower F_{ST} values given the extent of high resistance values (Supporting Figure 2.2). The northernmost population (TZ01) is remote but not completely isolated, as it demonstrates high but not infinite resistance values (concurrently, low but non-zero current values) in relation to the other populations.

Discussion

Population genetic and spatially explicit analyses assist in unraveling the architecture of this hybrid complex on a genomic level and can offer insight into the broader implications of interspecies gene flow. The simplest application of the *structure* population differentiation model (K=3) reveals significant population structure for at least *P. ocai* parentals, and the more complicated models not only mirror population structure between geographically disparate *P. maculatus* locations but also suggest differentiation within each of the hybrid gradients (Figures 2.2 and 2.3). This pattern suggests the strongest signal of differentiation after the *P. ocai – P. maculatus* divergence is the population differentiation between the geographically separated *P. ocai*, also supported by the pairwise population F_{ST} values. Unique population signals within a hybrid zone, often marked by private alleles in hybrid populations, are common across taxa (Barton *et al.* 1983; Woodruff & Gould 1987; Guiller *et al.* 1996). Although these signatures of differentiation are apparent, there is also no one 'pure' parental population, suggesting

differential introgression across each of the hybrid gradients as well as admixture between the gradients where they meet near Mt. Orizaba. The mitochondrial locus reveals not only divergence between the two species but population differentiation and asymmetrical gene flow into the hybrid gradients. Although it may initially seem incongruent with the *structure* results that the mitochondrial coalescent analysis reveals a 5-population model as most appropriate, it is important to note that this coalescent analysis is based on a single locus rather than the signal from many loci. In addition, mitochondrial analyses alone can often be misleading, especially in the context of hybridization (Jacobsen & Omland 2011). The fact that the general signals of gene flow and differentiation among the mitochondrial and AFLP loci are somewhat concordant is notable. In addition, the individuals in the hybrid zones show a range in variance of admixture, confirming what the plumage suggests, that these hybrid populations are not simple cases of parental dispersal and F_1 or even F_2 hybridization events, but populations involving long, stable, hybrid swarms (Figure 2.3).

The pine/oak habitat and associated edge agricultural habitat corridors reveal narrow, ribbon like dispersal corridors among both Teziutlán and Transvolcanic populations. The two hybrid gradients connect near Mt. Orizaba, and show lower resistance/ higher conductivity areas where the two transects meet. Although the Teziutlán habitat may demonstrate greater linear conductance, the corridor is narrow. The Transvolcanic habitat may offer a greater variety of dispersal pathways, but some feature higher resistance values (Figure 2.5). The mean inter-site resistance distance is slightly greater across the Transvolcanic belt, but the variance among resistance distances is greater across the Teziutlán gradient; this increased among-site variance in the

Teziutlán gradient is also reflected in the genetic differentiation values. These differing connectivity characteristics of the respective hybrid gradients may reflect ecological differences driving the geographically differentiated population signatures observed in the genetic data.

Several studies have used resistance distances to explore questions regarding isolation by distance and habitat preference, but this is the first implementation of the approach in a hybrid zone with replicated transects (Smit *et al.* 2007; Shah & McRae 2008; Andrew *et al.* 2012). Although the relationship between resistance distance and genetic differentiation is statistically significant, this isolation by resistance distance signature cannot account for all of the observed variation because of the sympatry of parental types near Mt. Orizaba and in Oaxaca. However, many of the large values of both resistance distance and F_{ST} are accounted for by comparisons involving disparate parental types (Figure 2.6) and the resistance distance - F_{ST} correlation improves markedly when parental comparisons are removed from the analysis. Although the sympatric populations of *P. maculatus* and *P. ocai* are living and mating in heavily overlapping habitat, most introgression of alleles from the opposite species appears to be coming through the hybrid gradients rather than from frequent F_1 hybrids at these sympatric sites (Figure 2.3). Both species' Oaxaca populations show a unique admixture signal, although the more complicated *structure* models suggest this arises from population differentiation (Figures 2.3 and 2.4), a pattern echoed in the mtDNA for Oaxacan *maculatus* (Supporting Fig 1). In addition, mainly the southern *P. ocai* population signature 'leaks' into the southern *P. maculatus* populations, so the main source of hybrid introgression in the southeast may be the Teziutlán gradient, or at least

the swarm of admixture that surrounds the central Teziutlán gradient and the Tranvolcanic intersection (Figures 2.2 and 2.3). An asymmetry in introgression is reflected in the over-representation of larger F_{ST} values in the southern *P. ocai* populations compared to the general trend (Supporting Figure 2.2), which in turn is not observed in the southern *P. maculatus* populations. The pattern is less pronounced farther south in Oaxaca, perhaps again driven by the local population differentiation (Figure 2.4, Supporting Figure 2.2). The analyses including resistance distances not only help clarify the patterns driven by geographic isolation and fragmentation, but they also help isolate inter-site differences distinct from the general isolation by distance pattern.

Although the unique topographic and habitat characteristics of the Teziutlán and Transvolcanic gradients contribute to both differentiation and gene flow, the uniqueness of the parental types contributing to these same gradients is also reflected in the genetic signature: data from both hybrid transects suggests low levels of long distance introgression occurring together with local differentiation in hybrid populations and geographically disparate parental populations. As evidenced by both the *structure* analyses and the pairwise population F_{ST} values, the geographically disparate *P. ocai* populations (TV01/TV02 v TZ08/TZ09) exhibit a greater extent of divergence than the *P. maculatus* (TZ01/TZ02 v TV09/TV10). The dynamics of reproductive isolation in the face of gene flow can be important and influential in systems going through cycles of secondary contact, particularly those interacting in a complex spatial setting.

Hybrid zones often exemplify classical examples of endogenous selection where hybrid genotypes reveal incompatibilities on foreign genomic backgrounds. Researchers often utilize this powerful hybrid context to search for genes contributing to reproductive

isolation; the importance of sex chromosomes in reproductive isolation is frequently emphasized (Harr 2006; Macholán *et al.* 2007; Carling & Brumfield 2008; Teeter *et al.* 2008; Dufkova *et al.* 2011). Even in conjunction with endogenous selection, exogenous selective factors can be important in hybrid contexts, especially in relation to habitat patchiness and local differentiation (Panova *et al.* 2006; Carson *et al.* 2012). In the towhee hybrid zone setting, both hybridization and hybrid 'swarm' populations clearly persist. However, the identities of the parental species have been maintained in the face of hundreds of generations of gene flow. This resultant pattern of observed differentiation and introgression could be a net effect of genes important to reproductive isolation in conjunction with variation in exogenous selective pressures due to patchy habitat. The fact that the habitat differences between the two transects influence the nature and force of gene flow is evolutionarily relevant. The two transect comparison reveals that while narrow high conductance corridors are effective in maintaining gene flow, a number of slightly higher resistance pathways for migration may actually reduce heterogeneity among sites. Although these populations seem to have been stable for quite some time, habitat fragmentation or restriction due to climatic change could alter the differentiation-introgression balance currently observed. Researchers could utilize the resistance distance methodology and empirical levels of differentiation in a model framework to predict effects of habitat restriction or expansion due to factors like climate change or anthropomorphic land use change.

To tease apart the dynamics of a spatially complicated system, researchers must appropriately account for neutral demographic processes like isolation by distance before evaluating adaptive forces (Andrew *et al.* 2012). Our resistance distance analyses not

only reveal the instances in which genetic differentiation is acting in accordance with drift-migration equilibrium, but also site-specific instances in which other factors must be in effect. The replicate transect analysis also helps tease apart confounding factors by revealing increased genetic heterogeneity in association with increased among-site heterogeneity in connectivity. Because we see retention of parental species identity for both species in both transects, there must be a group of loci maintaining phenotypic identity despite observed gene flow. In addition to the retention of ancestral divergence, additional divergence both within the hybrid zones and among parental types demonstrates the continued influence of local pressures increasing differentiation.

Acknowledgements

I appreciate greatly Adolfo G. Navarro S., Blanca Hernández, Laura Villaseñor, Alex Gordillo, Erick A. García Trejo, Hernán Vázquez, Gala Cortez, Daniella Tovilla-Sierra, Ruben Ortega, Marco Ortiz, Mauricio Pérez-Vera, Laila Cascabelera, and César A. Ríos Muñoz for their invaluable guidance, leadership, and assistance in the field across the Transvolcanic Belt in Mexico. I thank David Braun, Stephen Braun, Nicholas Lanson, Nimrod Funk, and George Oliver for their able assistance in the field in the Teziutlán gradient. Thanks to Darrilyn Albright for labwork assistance. I thank our funding sources for this project: Smithsonian Research Endowment Fund, University of Maryland Program in Behavior, Ecology, Evolution, and Systematics, Smithsonian Ornithology, Smithsonian Institution National Museum of Natural History Vertebrate Zoology. NSF grants DEB 0228675 and DEB0733029, and the NMNH Frontiers in Phylogenetics Program provided research assistantships. I thank Michael Braun and

William Fagan for invaluable input and comments on the manuscript. Appreciation goes to Gerald Wilkinson, Kevin Omland, David Hawthorne, and Matthew Hare for comments on the manuscript.

Tables

Table 2.1. Sampling locations across two towhee hybrid transects in Mexico

Num.	Location	plumage type	transect	latitude	longitude	sample N	sampling date
TZ01	Coahuila	*Pipilo maculatus*	Teziutlán	25.250	-100.450	16	1981
TZ02	Queretaro	*Pipilo maculatus*	Teziutlán	20.783	-99.567	17	1981
TZ03	Tlaxco	hybrid	Teziutlán	19.667	-98.167	17	1979, 1981
TZ04	E. Huichautla	hybrid	Teziutlán	19.783	-97.600	14	1979
TZ05	Teziutlán	hybrid	Teziutlán	19.817	-97.367	15	1979
TZ06	R. Palenquillo	hybrid	Teziutlán	19.650	-97.117	16	1979
TZ07	Cofre Perote	hybrid	Teziutlán	19.567	-97.100	15	1979
TZ08	Orizaba	*Pipilo ocai*	Teziutlán	19.075	-97.331	47	1981, 2009
TZ09	Oaxaca	*Pipilo ocai*	Teziutlán	17.167	-96.633	15	1981
TV01	Cacoma	*Pipilo ocai*	Transvolcanic	19.851	-104.453	30	2008
TV02	Colima	*Pipilo ocai*	Transvolcanic	19.633	-103.621	30	2008
TV03	Uruapan	hybrid	Transvolcanic	19.491	-102.006	21	2008
TV04	Ichaqueo	hybrid	Transvolcanic	19.578	-101.147	31	2008
TV05	Contepec	hybrid	Transvolcanic	19.965	-100.159	29	2008
TV06	Ocuilan	hybrid	Transvolcanic	18.982	-99.377	29	2009
TV07	Popocatépetl - Iztaccíhuatl massif	hybrid	Transvolcanic	19.101	-98.594	31	2009
TV08	Acajete	hybrid	Transvolcanic	19.149	-97.926	29	2008
TV09	Orizaba	*Pipilo maculatus*	Transvolcanic	19.075	-97.331	44	1981, 2009
TV10	Oaxaca	*Pipilo maculatus*	Transvolcanic	17.167	-96.633	14	1981
	total					460	

Table 2.2. Population structure schemes tested using the mtDNA ND2 locus in Migrate-n

Transect orientation	sampling number	pop structure scheme 1	pop structure scheme 2	pop structure scheme 3
	model probability	1.0e-12	1.0	1.7e-36
Teziutlán North	TZ01	A	A	A
	TZ02	A	A	A
	TZ03	A	A	B
	TZ04	A	A	B
	TZ05	A	A	B
	TZ06	B	B	C
	TZ07	B	B	C
	TZ08	B	B	D
South Transvolcanic West	TZ09	B	B	D
	TV01	C	C	E
	TV02	C	C	E
	TV03	D	D	F
	TV04	D	D	F
	TV05	D	D	F
	TV06	D	D	G
	TV07	D	D	G
	TV08	D	D	G
	TV09	D	D	H
East	TV10	D	E	H

Table 2.3. Evanno method using delta K (rate of change in likelihood among models) to differentiate among number of populations inferred from *structure* analysis

K	Reps	Mean LnP(K)	Stdev LnP(K)	Ln'(K)	|Ln''(K)|	**Delta K**
2	10	-79490.60	346.13	NA	NA	NA
3	10	-77460.87	13.93	2029.73	985.31	**70.713**
4	10	-76416.45	140.66	1044.42	152.16	1.082
5	10	-75524.19	222.55	892.26	291.02	1.308
6	10	-74922.95	291.14	601.24	68.46	0.235
7	10	-74390.17	40.85	532.78	493.41	**12.079**
8	10	-74350.80	127.66	39.37	322.80	2.529
9	10	-73988.63	242.84	362.17	83.40	0.343
10	10	-73709.86	296.88	278.77	1431.35	4.821
11	10	-74862.44	3802.28	-1152.58	734.15	0.193
12	10	-76749.17	9780.71	-1886.73	NA	NA

Figures

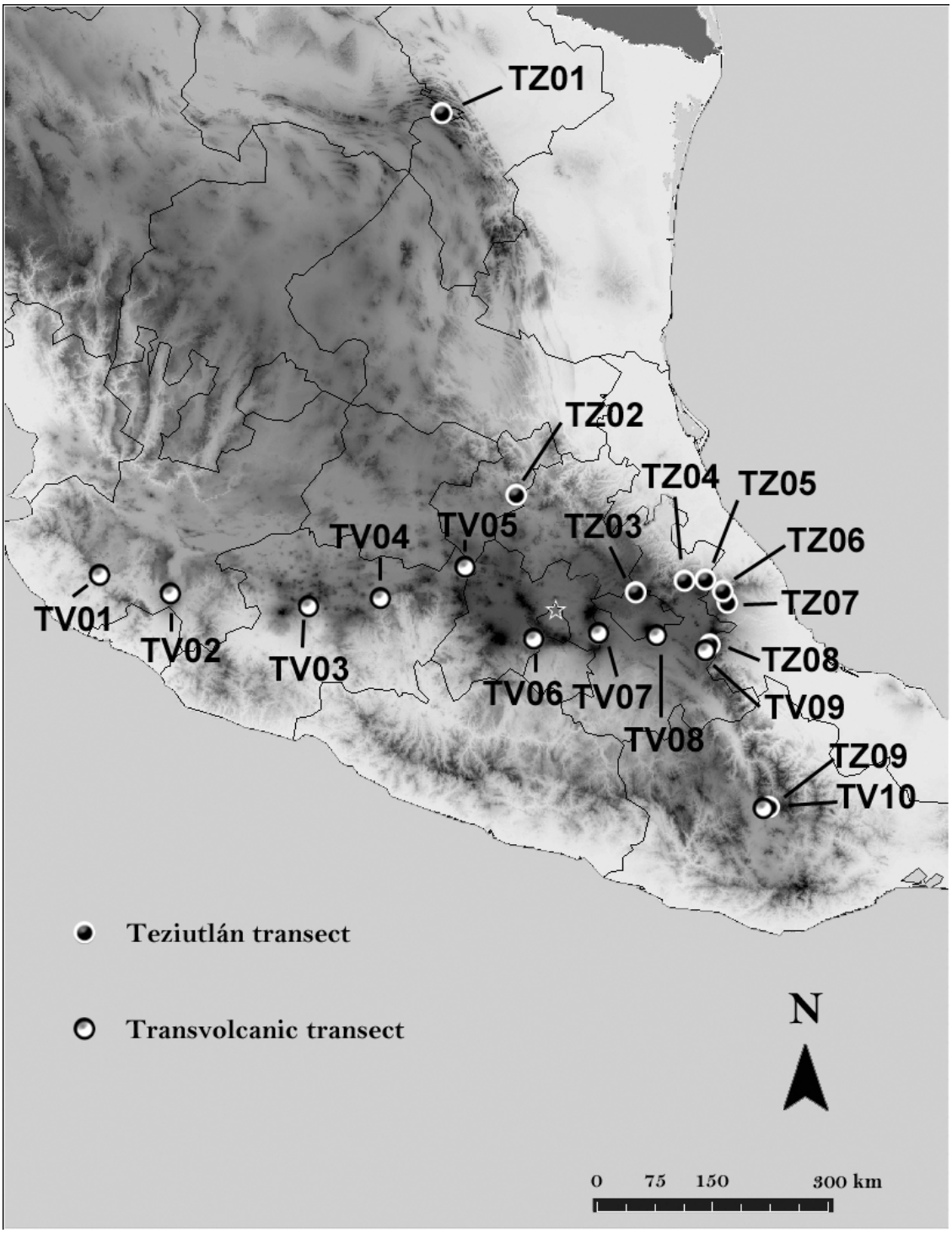


Figure 2.1. Sampling locations across both Teziutlán and Transvolcanic gradients depicted in Mexico.

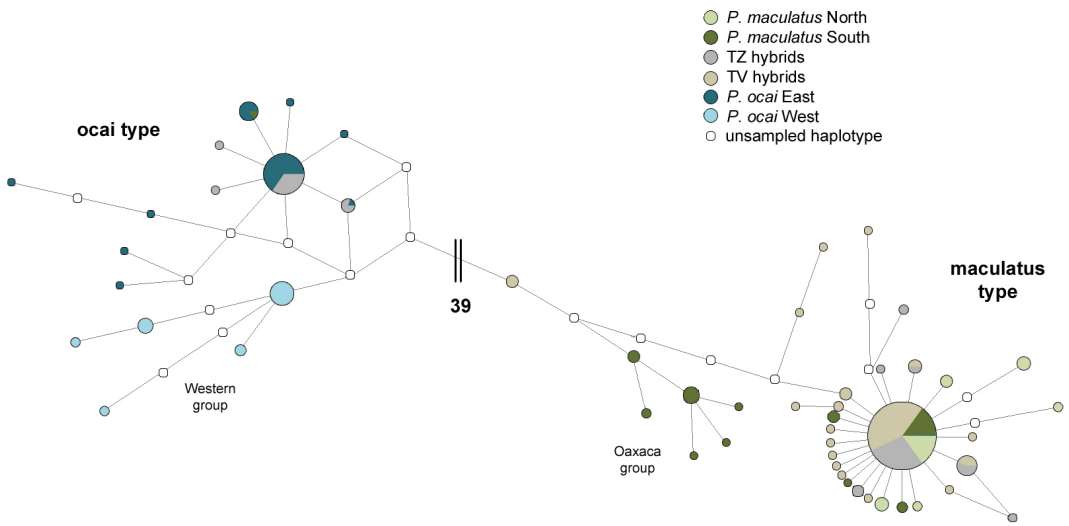


Figure 2.2. Median joining network of the mtDNA ND2 haplotypes surveyed across the parental species and both hybrid transects. Haplotypes are represented by circles at the nodes and joined by edges representing single base-pair changes; the size of the node is proportional to the observed frequency of the haplotype. Additionally, nodes are color-coded by parental and hybrid types. The lengths of the edges are proportional to the number of changes, with the single large break between species groups noted.

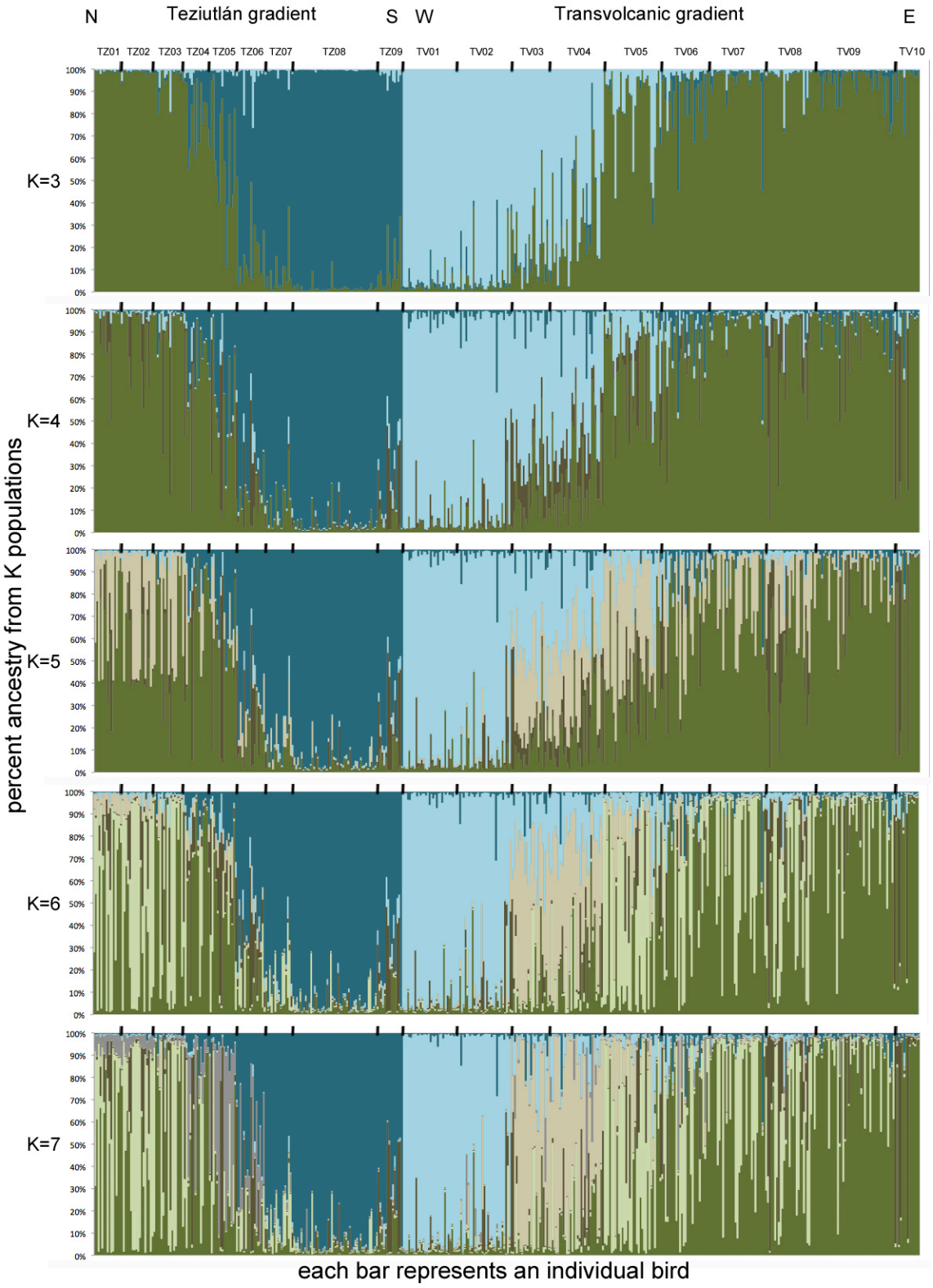

N
Teziutlán gradient
S
W
Transvolcanic gradient
E
TZ01 TZ02 TZ03 TZ04 TZ05 TZ06 TZ07 TZ08 TZ09 TV01 TV02 TV03 TV04 TV05 TV06 TV07 TV08 TV09 TV10
K=3
K=4
K=5
K=6
K=7
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
percent ancestry from K populations
each bar represents an individual bird

Figure 2.3. Structure results showing percentage of ancestry for each individual across both hybrid transects for all K levels from 3 to 7. Each color represents a population signature (number of different colors per panel = K). Colors listed by sampling locations in which the corresponding signature appears at the greatest proportion: dark green, parental *maculatus*; dark blue, parental *ocai* (southern); light blue, parental *ocai* (western); light green, northern *maculatus*; beige, TV hybrids; grey, TZ hybrids; dark brown, southeastern hybrids (both transects) and Oaxaca populations (both species).

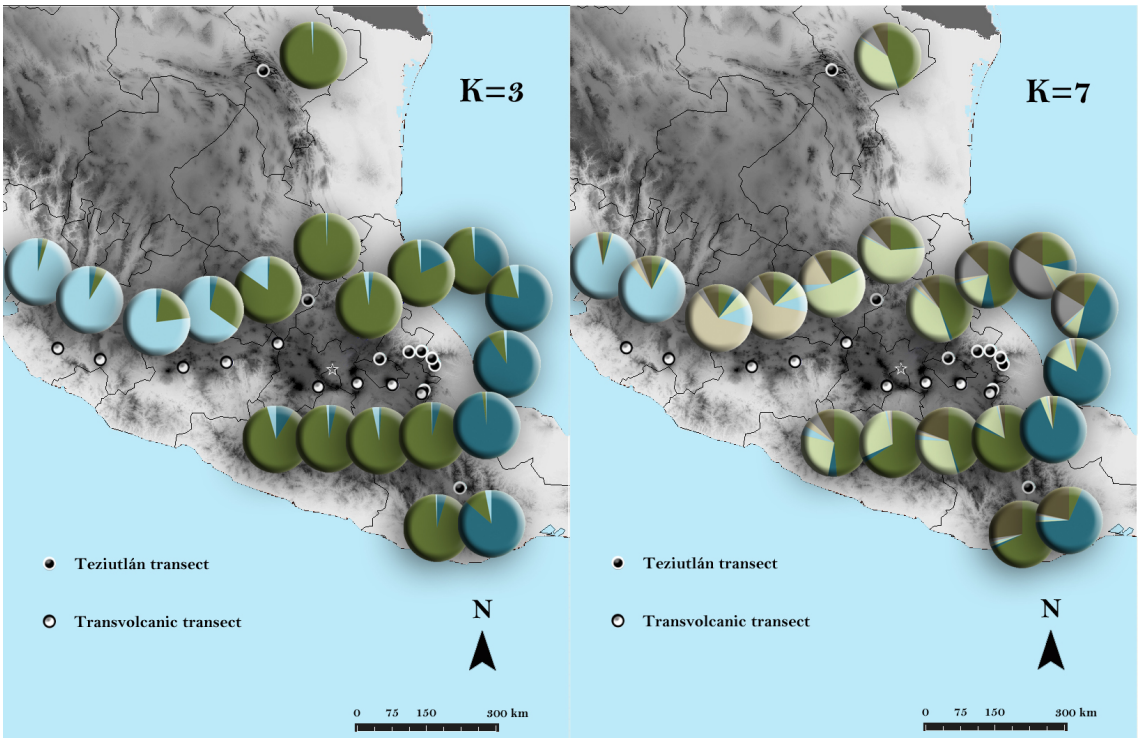


Figure 2.4. Estimates of percent ancestry pooled by sampling location for both K=3 and K=7. Each color represents a population signature. Colors listed by sampling locations in which the corresponding signature appears at the greatest proportion: dark green, parental *maculatus*; dark blue, parental *ocai* (southern); light blue, parental *ocai* (western); light green, northern *maculatus*; beige, TV hybrids; grey, TZ hybrids; dark brown, southeastern hybrids (both transects) and Oaxaca populations (both species).

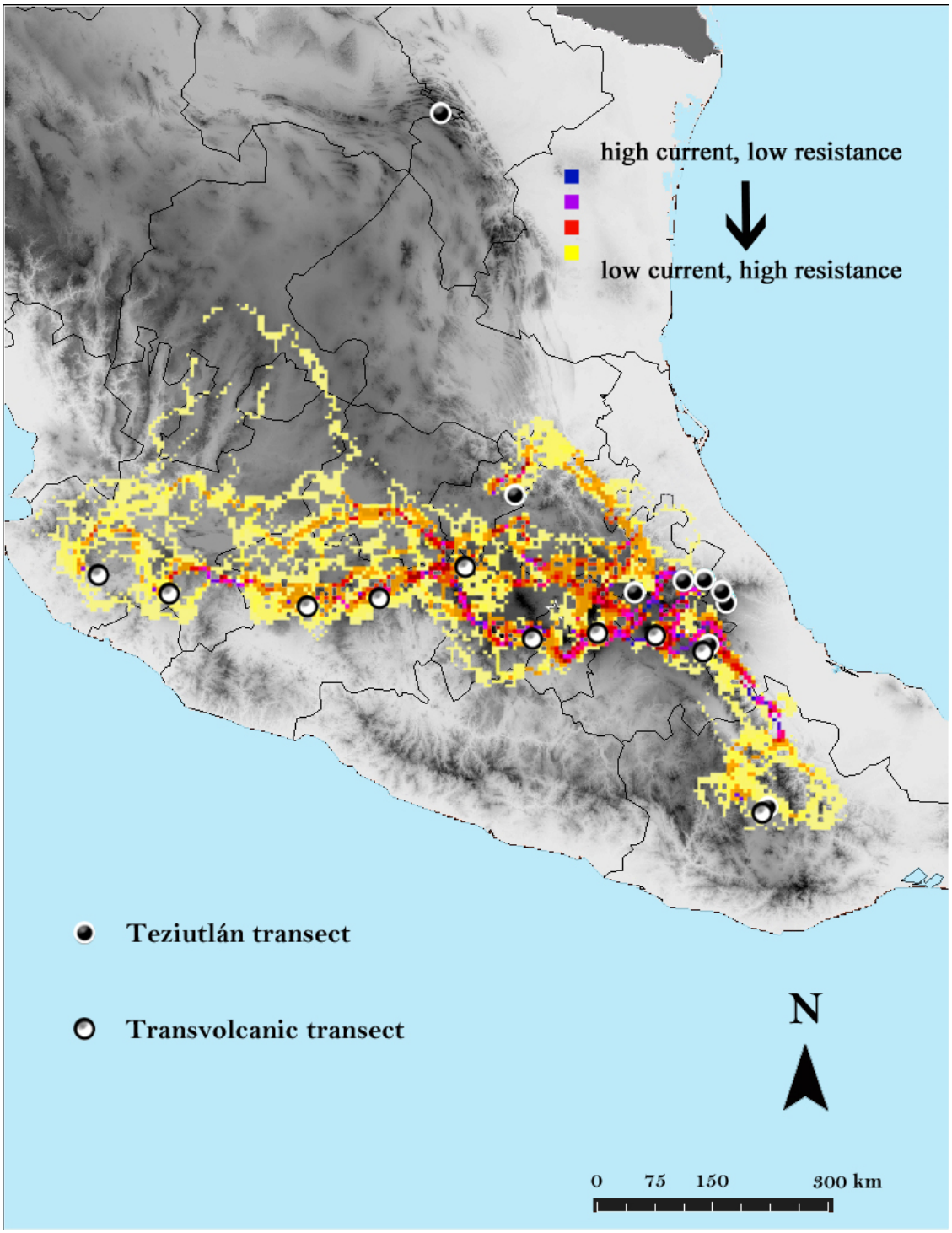


Figure 2.5. Cumulative current map based on all-to-one analysis depicting habitat corridors among sampling sites across the Teziutlán and Transvolcanic gradients. Some non-zero but very low current values exist that are not represented in the color ramp in order to avoid obscuring relevant topographical data.

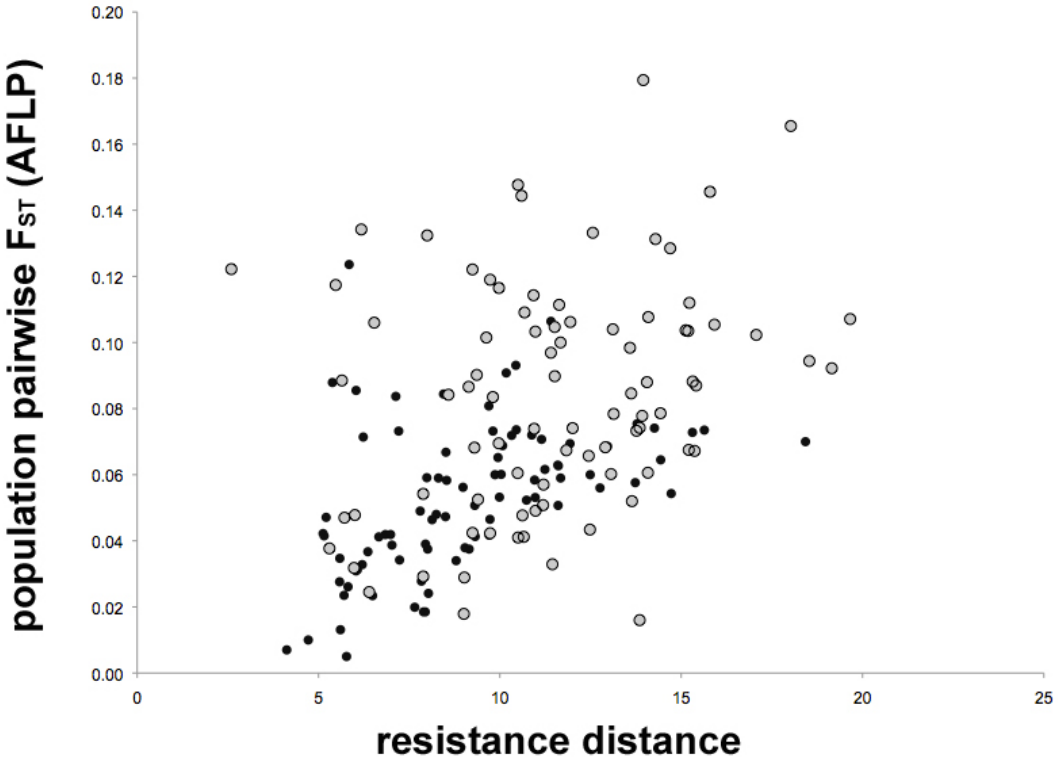


Figure 2.6. *Circuitscape* resistance distance (X-axis) plotted against pairwise population F_{ST} derived from all AFLP loci (Y-axis); pairwise comparisons involving the parental type populations (TZ01, TZ02, TZ08, TZ09, TV01, TV09, and TV10) account for nearly all the variation on the upper spectra of each axis (grey shaded points). Mantel test using all comparisons, r = 0.32, p-value = 0.03; Mantel test with parental populations removed, r = 0.57, p-value = 0.0017)

Chapter 3: Genome-wide analysis and comparison of two hybrid transects

Abstract

Divergence, gene flow, and reproductive isolation are topics of general evolutionary interest that can be explored in the context of inter-species hybrid zones. When two genomes interact, genomic patterns of heterogeneity can offer insight into selective as well as stochastic evolutionary processes. We utilize two transects across areas of hybridization in Mexico between the towhee species *Pipilo maculatus* and *P. ocai*. Using high-throughput sequencing techniques and a hierarchical Bayesian population model, we collected data on nearly 28,000 single nucleotide polymorphisms (SNPs). SNP loci exhibiting extreme patterns of introgression in relation to the genome-wide expectation were detected using a Bayesian genomic cline model. Loci with similar extreme patterns across both hybrid transects may reflect areas of the genome associated with global maintenance of divergence, while those extreme loci that differ between transects are more likely indicative of genome areas under the influence of locally specific factors (both adaptive and stochastic). We indeed find a small but significant level of outlier overlap between the two transects, as well as a heterogeneous signal of introgression and adaptation across the genome. Maintenance of historical divergence is acting in a cohesive manner on a core set of loci across transects, but local environmental and stochastic factors are also important driving forces.

Keywords:

Divergence, introgression, population genomics

Introduction

Species concepts and theory regarding modes of speciation are constantly evolving (Dobzhansky 1940; Barton & Hewitt 1989; Barton & Whitlock 1997; deQueiroz 2005; Nosil & Feder 2012). Interaction between divergent or diverging genomes allows observation and delimitation of important reproductive isolating mechanisms. Hybrid zones, in which divergent genomes come into contact, offer natural laboratories in which to assess important factors in speciation processes *in situ*.

A number of evolutionary forces can create the patterns of variation we see across areas of interspecies hybridization (Barton 1986; Rieseberg *et al.* 1999; Sattler & Braun 2000; Martinsen *et al.* 2001; Payseur *et al.* 2004; Yuri *et al.* 2009). Mechanisms governing flow of alleles across a species boundary can range from neutral introgression to a selective sweep of globally advantageous alleles. For selected traits, extent and rate of allelic introgression is related to the strength of selection (Barton & Hewitt 1989). In addition, introgression via 'hitchhiking' of alleles adjacent to selected traits is also related to strength of selection and level of linkage disequilibrium. The combination of divergent selection and linkage disequilibrium hitchhiking is often offered as an explanatory factor for observed islands of differentiation (Harr 2006; Nosil *et al.* 2008; Feder & Nosil 2010; Nadeau *et al.* 2012). When divergent lineages reconnect, several types of forces can act on hybrids to maintain parental species identity. Hybrids can be less fit than parental types because novel genotypes have not been tested on the adaptive landscape or because combinations of loci that arose in separate lineages are inherently incompatible (Barton 2001; Orr & Turelli 2001). Even if hybridization exposes advantageous allelic combinations previously unavailable for testing by selection, these

combinations can be bounded by habitat or environmental gradients in the selective pressure itself (Moore 1977).

Much emphasis has been placed on the processes of incipient speciation, divergence with gene flow, and divergent selection driving sympatric speciation (Danley *et al.* 2000; Emelianov *et al.* 2003; Niemiller *et al.* 2008; Nosil 2008). However, lineages that evolved in allopatry and have come into secondary contact can offer a different kind of natural hybrid laboratory setting in which to observe the evolutionary dynamics of reproductive isolation. Gompert *et al*. (2012) point out that these types of secondary contact hybrid zones offer at least a partial separation of historical divergent selective forces and mechanisms of reproductive isolation. Environmental and selective factors that change over time may differ between the original habitat and period of divergence and the current area of contact and hybridization. Much of the literature on hybrids shows that fitness effects can be environmentally variable, and that exogenous selective pressures can play an important role in hybrid zones (Teeter *et al.* 2008; Nolte *et al.* 2009; Carson *et al.* 2012). Due to the novel recombinant genotypes that arise in a secondary contact hybrid zone, measures of introgression across the genomic landscape can act as proxies for assessing reproductive isolation between the two lineages (Gompert *et al.* 2012b). Non-adaptive processes like hitchhiking and drift can also come into play as divergent alleles interact not only across geographic space but also as genomes intersect.

Because divergent selection among lineages increases differentiation and decreases within lineage variation, loci on the extremes of the divergence scale are likely to be involved with the genetic architecture of reproductive isolation. In hybrid zones,

variation in selective pressure will be reflected in heterogeneous patterns of differential introgression, both on geographic and genomic scales (Gompert *et al.* 2012b). Heterogeneity in the porosity of the species membrane has been observed in hybrid zone settings (Rieseberg *et al.* 1999; Rieseberg *et al.* 2003; Payseur *et al.* 2004; Kingston *et al.* in review). Using simulations and the genomic cline model, Gompert *et al*. (2012) demonstrate heavily decreased introgression rates in association with clustered loci of large effect on reproductive isolation (a pattern empirically observed in the European house mouse hybrid zone (Harr 2006; Macholán *et al.* 2007; Teeter *et al.* 2008)). Two types of analyses can lend insight into the architecture of reproductive isolation: those that assess differentiation and those that assess rate and nature of introgression at individual loci across a hybrid gradient (Gompert *et al.* 2012a; Gompert *et al.* 2012b; Parchman *et al.* 2012). Due to the tremendous increase in power provided by population genomic scale data sets, the inferences we can make from locus-specific analyses is much greater now than even in the recent past. To assess the variation in loci associated with divergence, and thus potentially associated with reproductive isolation across a species boundary, we utilize two geographically disparate hybrid transects between the same pair of species. With this approach we can analyze concordance in the pattern of isolation in two different areas of hybridization by isolating loci that differ from the genomic background in patterns of differentiation or patterns of introgression.

The target species for this study are two towhees, *Pipilo maculatus* and *P. ocai*, which come into secondary contact and hybridize in the high elevation pine-oak scrub habitat of Mexico. The two largest gradients of hybridization are along the Sierra Madre Oriental (Teziutlán gradient) and the Transvolcanic Belt (Transvolcanic gradient) in

central Mexico (Figure 3.1). These two gradients can serve as replicate transects. In addition to natural replicate gradients, the area of contact also includes a unique situation in the southern portion of the range (Mt. Orizaba and south into Oaxaca) where the parental form of *P. maculatus* re-appears in sympatry with *P. ocai* populations, in some areas just tens of kilometers from the hybrid swarm populations (Sibley 1954; Sibley & Sibley 1964; Braun 1983; Kingston *et al.* in review) (Table 3.1 and Figure 3.1, population numbers TZ08, TZ09, TV09, and TV10). These two species have likely been in contact since the end of the last glaciation, and have been historically observed as stable hybrid populations for at least 150 years (Sibley 1950, 1954; Sibley & West 1958; Sibley & Sibley 1964). This unique area of contact and hybridization offers us an opportunity to investigate the nature and extent of variation in the genetic architecture of maintenance of divergence and reproductive isolation between two species.

We aim to use the comparison between transects to test whether maintenance of divergence is acting in concordance in both transects. We expect that patterns among outlier loci within each individual transect could be driven by a number of both adaptive and non-adaptive forces: both endogenous and exogenous selection, drift, linkage disequilibrium, gene duplication, and gene conversion. Many of these forces likely differ to some extent in the two geographically disparate replicates. In addition, there is geographic population differentiation among the same-species parental types used in each transect (Kingston *et al.* in prep). Given these differences, overlap in outlier loci patterns between the two transects should reflect the signal of divergence that maintains core parental species identity in the face of gene flow.

Materials and Methods

Sampling

Both the Teziutlán (~1150km in length, 172 individuals) and Transvolcanic gradients (~760km in length, 288 individuals) were sampled creating replicate transects across the observed hybrid variation (Table 3.1, Figure 3.1). While these natural replicates are not completely independent nor identical, they do represent two hybrid interfaces where geographically disparate populations of the same parental species interact. Birds were collected via mist net or shotgun, tissue samples were frozen in liquid nitrogen in the field and voucher museum specimens were prepared for each individual collected (sampling efforts in 1979, 1981, 2008, and 2009). Tissues and vouchers were deposited at Louisiana State University Museum of Natural Science and the US National Museum of Natural History.

DNA was extracted using phenol-chloroform preparation (Braun 1983; Sambrook *et al.* 1989) via manual methods or AutoGen extraction robot. DNA chain length assessed via gel electrophoresis and DNA was quantified using a Nanodrop ND-1000 spectrophotometer.

DNA Sequencing

We collected genome-scale DNA sequence data from the 460 individual samples by generating reduced genomic complexity libraries for each individual via a restriction fragment based procedure (Gompert *et al.* 2012a; Parchman *et al.* 2012). The method is heavily multiplexed, utilizes restriction enzymes for target reduction of the genome, and results in highly multilocus data; our method is similar to the genotype-by-sequencing (GBS) approach (Elshire *et al.* 2011). We digested genomic DNA with the restriction

endonucleases *Eco*RI and *Mse*I and ligated double-stranded adaptor oligonucleotides to the resultant fragments. These adaptors consisted of the Illumina sequencing priming sites followed by eight, nine, or ten bp index sequences that allow for the unique identification of sequences for each individual. We used a total of 576 unique indexed adaptors, which allowed us to pool many individuals in the same Illumina Hiseq sequencing lane. We used PCR to amplify index-tagged fragments using two replicate reactions for each individual. Amplicons for sets of up to 300 individuals with unique indices were then pooled. We size-selected pooled amplified libraries via electrophoresis on a 2% agarose gel and excised fragments between approximately 350 and 450 bp in length. We purified these fragments using the Qiaquick Gel Extraction Kit (Qiagen Inc.). A detailed description of this protocol can be found in Parchman *et al*. (2012). Concentration of the pooled library was measured on an Agilent BioAnalyzer, and suitability as a sequencing template was verified using qPCR. Sequencing of the libraries was completed by the National Center for Genome Research (Santa Fe, NM, USA) using the Illumina HiSeq platform; 100 base sequencing reads were generated in each of three sequencing lanes.

We used a custom Perl script to remove indices and restriction sites from the *Eco*RI end of all sequences, replacing the sequence IDs with the individual IDs. This script also corrected indices that were off by up to two bases due to sequencing error or oligonucleotide synthesis. After parsing indices and removing reads containing contaminant sequences, 282,895,590 reads were retained for analysis. We used SeqMan NGen 3.0.4 (DNASTAR) to perform an initial *de novo* assembly for a subset of 40 million sequences, which placed 19,143,814 reads into 408,347 contigs. After removing

low-quality contigs and those with consensus sequences less than 82 bases in length or greater than 90 bases in length, we generated an artificial partial reference genome from the contig consensus sequences from the *de novo* assembly (366,607 contigs). We then assembled the full set of 282,895,590 sequences onto the reference using SeqMan 1.0.3.3 (DNASTAR). This assembly placed 180,627,551 reads onto the reference, for an average coverage depth of 1,220x per sequenced region (1.4x per individual per sequenced region). We used custom Perl scripts together with samtools and bcftools (Li *et al.* 2009) to identify variant sites in the assembled sequence data and determine the number of reads supporting each alternative nucleotide state for each individual and locus. The data from the called variant sites were placed in a file containing counts for the number of reads for each SNP in each individual. We discarded loci where individuals appeared to have more than two haplotypes. We also discarded any variable sites where the observed allele counts from putative heterozygous individuals were very unlikely given a binomial distribution with $p = 0.5$ (in concordance with the expectations of Mendelian inheritance) (Parchman *et al.* 2012). A final set of 27,968 SNP loci were retained for most analyses after removal of SNPs with low frequency variants (global minor allele frequency < 0.1).

Population genetic analyses

We used a Bayesian model to estimate population allele frequencies for each of 27,968 variable nucleotides based on the observed sequence data (the model is described in Gompert *et al*. 2012). We treated both the individual genotype at a locus and the population allele frequency as unknown model parameters, estimating genotype probabilities and allele frequencies separately for each of the 19 sampled populations.

We used Markov chain Monte Carlo (MCMC) to acquire parameter posterior distributions. Each analysis utilized a single chain iterated for 20,000 steps where samples were recorded every fourth step. The main strength of this approach is the Bayesian model allows for simultaneous estimation of cline parameters while accounting for uncertainty associated with variable (and often low) sequencing coverage (uncertainty arising not from sequencing error, but from uneven coverage of heterozygous alleles and uneven population sampling among loci). The hierarchical nature of the uncertainty estimation allows for very precise population level parameter estimation even from lower sequence coverage (average per site per individual 0.5 -1.5x) (Gompert *et al.* 2012a; Gompert *et al.* 2012b; Parchman *et al.* 2012; Parchman *et al.* in prep). To summarize population genetic structure across all of the sampled population and individuals, we employed principal component analysis (PCA). We used the model-estimated genotype probabilities for two of the three possible genotypes (the heterozygous genotype and one homozygous genotype) for each individual at each locus as variables for PCA. We performed the PCA in R using the prcomp function.

A Bayesian execution of the F-model was also used to quantify genome-wide genetic differentiation for pairs of population samples (Gompert *et al*. 2012). The F-model estimates a parameter of genetic differentiation that is equivalent to F_{ST} under several neutral evolutionary models (Balding & Nichols 1995; Nicholson *et al.* 2002; Falush *et al.* 2003); in this system we equate the F-model parameter with F_{ST}. We performed these analyses using all loci with a minor allele frequency ≥ 0.1 across all populations pooled. We estimated posterior probabilities of F_{ST} at each individual locus (between parental populations) and a genome-wide global metric of F_{ST} (among all

population pairs) F_{ST} using 25,000 MCMC steps; every 10^{th} step was saved and the first 1,000 steps were discarded as burnin. The F-model incorporates uncertainty in both allele frequency (population level sampling uncertainty) and genotype estimation (allele level uncertainty) simultaneously. The estimation of genome-wide quantiles allow for isolation of loci with outlier F_{ST} estimates.

Genomic clines

We used a modified version of the Bayesian genomic cline model (Gompert & Buerkle 2011) to assess genome-wide variation in introgression among the admixed *Pipilo* populations. This model compares each individual locus to a global genetic hybrid index derived from the composite multi-locus signal (the probability of an admixed individual's ancestry from parental species *P. ocai*). This probability of ancestry is estimated using the likelihood of the observed data across admixed individuals, loci, alleles, and populations. Locus specific deviations from the genomic background are quantified with two parameters: α (center parameter) and β (rate parameter). When α and β equal zero, the hybrid index completely predicts the probability of *P. ocai* ancestry at that single locus. As α shifts the probability of *P. ocai* ancestry also shifts with relation to the hybrid index prediction in either a positive or negative direction. As β shifts, the rate of transition across the admixture gradient shifts away from the one-to-one line predicted by the hybrid index in either a positive or negative direction (Gompert *et al.* 2012b).

Each transect (TZ and TV) was analyzed independently. Parental populations for the Teziutlán gradient were designated as TZ01, TZ02 (*P. maculatus*) and TZ08, TZ09 (*P. ocai*); the remaining populations were used as the admixed representatives.

Populations TV01, TV02 (*P. ocai*) and TV09, TV10 (*P. maculatus*) were designated as parental types for the Transvolcanic gradient; the remaining populations represented the admixed portion of the gradient. These admixed populations have been evaluated with other types of molecular and plumage markers and show high and variable levels of hybrid intermediacy (Braun 1983; Kingston *et al.* in prep; Kingston *et al.* in review). We estimated marginal posterior probability distributions for hybrid indices and cline parameters α (center parameter) and β (rate parameter) using MCMC. We ran five independent chains, each for 50,000 steps; following a 30,000 step burn-in, we recorded samples from the posterior distribution every 20th step. We considered genomic cline parameter estimates as outliers in relation to the genomic background signal if the upper credible bound of α did not overlap zero, if the lower credible bound of α did not overlap zero, if the upper credible bound of β did not overlap zero, or if the lower credible bound of β did not overlap zero. We combined the output of the five chains only after examining the MCMC output to assess convergence. In order to quantify any association between the measure of differentiation and measures of introgression, we tested for associations between F_{ST} and genomic cline parameters α (center parameter) and β (rate parameter) using Pearson's product moment correlation coefficient. We utilized the two transect sampling design to compare concordance in α (center), β (rate), and F_{ST} outlier loci across the two replicates. Previous simulations indicate that spurious (non-selectively driven) correlations between cline parameters and F_{ST} loci arise most often in the context of loci differentiated at $F_{ST} < 0.08 - 0.1$ (Gompert *et al.* 2012a); spurious correlations generally can be avoided by utilizing loci differentiated above this threshold.

Results

Population genetic analyses

The PCA reveals tight groupings within population samples and a perpendicular pattern intersecting the two hybrid transects (Figure 3.2). This pattern roughly mirrors the geographic orientation of the two hybrid transects, as well as highlights differentiation between geographically disparate parental types. PC 1 accounts for 13.60 percent of the observed variation across the 27,968 variable loci and PC 2 accounts for 6.98 percent. The level of resolution in population differentiation across both transects is remarkable and likely due to the power associated with the enormous number of SNPs utilized. Although we've previously shown with separate data that the populations at Orizaba sampled at different points in time are not genetically differentiated (Kingston *et al.* in prep), there is also no differentiation between 1981 and 2009 samples among the PCA coordinates. PC 1 largely separates the populations of the Tezuitlan gradient, while populations of the Transvolcanic gradient are distributed along PC 2. We observe differentiation between geographically disparate populations of both parental types, which is greater for *P. ocai* than for *P. maculatus*. The pairwise population F_{ST} estimates corroborate this general pattern where the eastern and western *P. ocai* types are slightly more differentiated (average pairwise population F_{ST} between east and west = 0.104) than the northern and southern *P. maculatus* (average pairwise population F_{ST} between north and south = 0.084) (Table 3.2). In addition, the Oaxaca populations of both *P. maculatus* and *P. ocai* parental types seem to appear more similar to their respective hybrid transects than either Orizaba population, even though the Oaxaca populations are much more geographically isolated. This initial pattern may be misleading, however, as PC3

accounts for half as much variation as PC2 (3.55%), and both Oaxaca populations are differentiated further from the hybrid transects along the PC3 axis. The first three PC axes account for 24.24% of the observed variation, which is quite a substantial proportion of variation given the nature of a highly multilocus genetic data set (Novembre *et al.* 2008; Parchman *et al.* 2012; Parchman *et al.* in prep). The phenomenon of genetic signal reflecting geography has been previously noted (Novembre *et al.* 2008); although the geographic signal is prevalent, the unaccounted for variation could suggest other demographic and selective forces play an important role in the dynamics of this hybrid zone.

Genomic clines

Genetic hybrid indices in each transect reveal bimodal distributions (each sampled individual bears a corresponding hybrid index); the two transects exhibit different ranges of hybrid index values across the gradient of admixture (Figure 3.3). While this transition is stepwise through populations in the Teziutlán gradient, variation in hybrid index seems slightly more disperse across populations within the Transvolcanic. The bimodality is unlikely driven by sampling scheme, given the close geographic proximity of Teziutlán gradient admixed populations (all sampled within 130km) and the diversity of hybrid indices found in the TV08 population in the Transvolcanic. Instead, the bimodality likely reflects the dynamics of gene flow and reproductive isolation across the hybrid gradients.

The exceptional genomic cline parameter estimates for the Teziutlán gradient include 967 loci with an α lower credible interval greater than zero, 1276 loci with an α upper credible interval less than zero, 81 loci with a β lower credible interval greater than

zero, and 32 loci with a β upper credible interval less than zero (2356 total outlier loci, Figure 3.4, Table 3.3). These are loci that likely tag genetic regions involved in maintenance of divergence. Genome-wide F_{ST} between the parental species (TZ01/TZ02 and TZ08/TZ09) is 0.138 (0.135 – 0.141, 95% quantile). The number of F_{ST} outlier loci (outside 95% quantile) across the Teziutlán gradient is 231 (Figure 3.5). The loci with extreme genomic clines that are also outlier F_{ST} estimates are: 18 loci with lower credible interval α greater than 0 and F_{ST} outlier, 68 loci with upper credible interval α less than 0 and F_{ST} outlier, 43 loci with lower credible interval β greater than 0 and F_{ST} outlier, and zero loci with upper credible interval β less than 0 and F_{ST} outlier. The correlation between F_{ST} and the absolute value of α is 0.212 (Pearson's product-moment correlation, p-value < 2.2e-16). The correlation between F_{ST} and the absolute value of β is 0.449 (Pearson's product-moment correlation, p-value < 2.2e-16). The vast majority of F_{ST} values analyzed across this transect are greater than the threshold where spurious correlations can occur (Gompert *et al.* 2012a) (Figure 3.5).

The exceptional genomic cline parameter estimates for the Transvolcanic gradient include 1514 loci with an α lower credible interval greater than zero, 728 loci with an α upper credible interval less than zero, 728 loci with a β lower credible interval greater than zero, and 23 loci with a β upper credible interval less than zero (2993 total outlier loci, Figure 3.4, Table 3.3). Genome wide F_{ST} between the parental species populations across the Transvolcanic gradient (TV01/TV02 and TV09/TV10) is 0.105 (0.103 - 0.107, 95% quantile). The number of F_{ST} outlier loci (outside 95% quantile) in the Transvolcanic gradient is 292 (Figure 3.5). Extreme genomic clines that also have outlier F_{ST} estimates are: 102 loci with lower credible interval α greater than 0 and F_{ST} outlier,

40 loci with upper credible interval α less than 0 and F_{ST} outlier, 17 loci with lower credible interval β greater than 0 and F_{ST} outlier, zero loci with upper credible interval β less than 0 and F_{ST} outlier. In the Transvolcanic gradient, the correlation between F_{ST} and the absolute value of α is 0.288 (Pearson's product-moment correlation, p-value < 2.2e-16), while the correlation between F_{ST} and the absolute value of β is 0.368 (Pearson's product-moment correlation, p-value < 2.2e-16). As with the Teziutlán gradient, the vast majority of F_{ST} values analyzed across the Transvolcanic transect are greater than the threshold where spurious correlations can occur (Gompert *et al.* 2012a) (Figure 3.5).

Transect Comparison

Some loci overlap as outliers in both transect genome cline analyses: 233 loci with exceptional α estimates in both transects, 9 loci with exceptional β estimates in both transects, 108 loci with exceptional F_{ST} estimates in both transects. Differences in shape between the transect outlier loci clines can be visualized in Figure 3.6. The right-hand tail of the α parameter distribution in the Transvolcanic appears to be longer than that in that observed in the Teziutlán gradient suggesting the Transvolcanic is slightly asymmetrical in introgression pattern (Figure 3.4 and Figure 3.6). The correlation between α estimates from both transects is 0.069 (Pearson's product-moment correlation, p-value < 2.2e-16). The Transvolcanic gradient also appears to exhibit a distribution of β parameters that are compressed around zero suggesting less overall variation in rate in comparison to the Teziutlán. The correlation between β estimates from both transects is 0.077 (Pearson's product-moment correlation, p-value < 2.2e-16). The correlation between F_{ST} estimates from both transects is 0.259 (Pearson's product-moment

correlation, p-value < 2.2e-16). The overlapping extreme loci are a subset the extreme loci in each transect, as reflected by the correlations between cross transect α and β outliers. The partial concordance in these parameter estimates illustrates both similarities and differences in the architecture of isolation in the two hybrid zones.

Discussion

We observe significant correlation between both the center (α) and rate (β) cline parameters and a measure of genome-wide differentiation (F_{ST}) suggesting that the more extreme clinal loci are more differentiated than the background genomic signal. Those loci with extreme differentiation values (F_{ST}) between parental types are likely to be loci influenced by divergent selective processes (Gompert *et al.* 2012b); in turn these same loci are associated with extreme cline parameters (α and β). These loci with dual extreme values likely track portions of the genome driven by selection and are presumably important to the maintenance of divergence between the species. The observed heterogeneity among cline parameters across the genome is high within each transect and varies between transects, as well (Figure 3.6). This pattern points to complex genetic architecture governing the interaction between the divergent genomes (Gompert *et al.* 2012b) and differences between the hybrid zones point to complex historical and environmental variation in structuring reproductive isolation.

A significant but not complete concordance exists between a core set of outlier cline parameter loci in the two hybrid transects. The overlap of outlier F_{ST} loci is slightly higher than the overlap of loci with extreme cline parameters. We observe population differentiation and different hybrid index distributions across the two transects, yet still

observe a significant concordance among outlier loci. Despite the stochastic and environmental factors that can influence the process of isolation and secondary contact, the significant core set of concordant loci between transects points to a shared genetic pattern of isolation. This core set of concordant outlier loci likely tag genetic regions that play a major role in maintenance of divergence between these species.

In addition to the core concordant loci, the variability observed suggests there is a significant influence of geographically specific environmental heterogeneity, as has been demonstrated across taxa in hybrid zones (Teeter *et al.* 2008; Nolte *et al.* 2009; Carson *et al.* 2012). We have previously demonstrated the importance of habitat corridors in relation to gene flow in this same towhee hybrid zone system (Kingston *et al.* in prep), and the genomic signal observed here reinforces that result. Not only do the two transects differ in composition of hybrid indices across the gradient (Figure 3.3), they also differ in locus-specific patterns of introgression. The Transvolcanic gradient appears to exhibit a slight asymmetry in introgression patterns (long positive tail on the center (α) distribution) as well as a lower variance in rate of change (compressed rate (β) distribution, Figure 3.4). These differences in introgression patterns are likely tied to the differences observed in hybrid index composition. It is possible that the transect-specific habitat connectivity differences help drive the divergent portions of introgression patterns. In addition, locality specific selective pressures could also influence the observed patterns. These cross-transect differences highlight the fact that the complexity of both ecological and evolutionary contexts can structure introgression and the evolution of isolation.

Historical divergent selection is often relevant in the context of hybrid fitness; it is in the context of hybridization in which the completeness of reproductive isolation is naturally tested (Bolnick *et al.* 2006). Therefore, the empirical observation of genome-wide heterogeneity in introgression or divergence is interpreted as evidence of selective influence on a system (Nosil *et al.* 2009). However, there can be disagreement between ecological forces and genomic signal; Dobzhansky-Muller incompatibilities can arise through neutral processes as well as selective ones, and an observed signal of post-zygotic reproductive incompatibilities and linkage disequilibrium in the genome does not preclude either scenario (Orr & Turelli 2001; Gavrilets 2003). In fact, the influence of stochastic factors on differentiation and reproductive isolation may be underestimated (Buerkle *et al.* 2011). It is through comparing the overlapping extreme loci in the two hybrid transects that we can begin to tease out which loci are associated with global maintenance of divergence and which loci differ from the genomic background due to transect-specific selective or stochastic factors. We would expect those loci associated with historical divergence and isolation to show consistent cross-comparison signal and those influenced by local selective and stochastic factors to differ in the comparison. Therefore, concordant set of outlier loci could not only represent the classic tension zone reproductive isolating mechanisms (those loci incompatible on foreign genomic backgrounds acting through endogenous selection) but also those associated with historical divergence and isolation (Barton & Bengtsson 1986; Barton 2001). The discordant set of loci should represent those loci involved with divergence or gene flow governed by environment-specific exogenous selection or even transect-specific drift-related processes. The ability to preliminarily parse out loci into these categories is

particularly relevant in a candidate gene search; if researchers are interested in locating particular genes associated with adaptive divergence the cross-comparison concordance and discordance can inform the search. Both types of loci, those associated with maintenance of divergence, and those associated with environmentally specific forces, are of general evolutionary interest.

The utilization of the unique structure of the replicated hybrid transects in this hybrid system allows us to compare signals of divergence and introgression across the same species' genomic architecture in two separate geographic settings. The comparison reveals a core of concordant outlier loci likely associated with divergent selection and globally associated with maintenance of isolation. The remaining extreme loci that vary between transects reveal heterogeneity in differential introgression between the two settings. Locally specific habitat and environmental factors, as well as neutral processes, are also influential in the dynamics of this system. Because secondary contact may partially decouple the effects of the divergent selective processes that drove these two lineages in different evolutionary directions, further investigation of the genetic architecture of this hybrid system should offer insight into both adaptive and non-adaptive forces relevant to speciation and divergence with gene flow.

Acknowledgements

I appreciate greatly Adolfo G. Navarro S., Blanca Hernández, Laura Villaseñor, Alex Gordillo, Erick A. García Trejo, Hernán Vázquez, Gala Cortez, Daniella Tovilla-Sierra, Ruben Ortega, Marco Ortiz, Mauricio Pérez-Vera, Laila Cascabelera, and César A. Ríos Muñoz for their invaluable guidance, leadership, and assistance in the field

across the Transvolcanic Belt in Mexico. I thank David Braun, Stephen Braun, Nicholas Lanson, Nimrod Funk, and George Oliver for their able assistance in the field in the Teziutlán gradient. I thank funding sources for this project: Smithsonian Research Endowment Fund, University of Maryland Program in Behavior, Ecology, Evolution, and Systematics, Smithsonian Ornithology, Smithsonian Institution National Museum of Natural History Vertebrate Zoology and Don Wilson. NSF grants DEB 0228675 and DEB0733029, and the NMNH Frontiers in Phylogenetics Program provided research assistantships. I thank Michael Braun and William Fagan for advisorial support and comments on the manuscript. I thank Thomas Parchman, Zachary Gompert, and C. Alex Buerkle for analytical guidance and comments on the manuscript. Appreciation goes to Gerald Wilkinson, Kevin Omland, David Hawthorne, and Matthew Hare for comments on the manuscript.

Tables

Table 3.1. Sampling locations across the two hybrid transects

Num.	Location	plumage type	transect	latitude	longitude	sample N	sampling date
TZ01	Coahuila	*Pipilo maculatus*	Teziutlán	25.250	-100.450	16	1981
TZ02	Queretaro	*Pipilo maculatus*	Teziutlán	20.783	-99.567	17	1981
TZ03	Tlaxco	hybrid	Teziutlán	19.667	-98.167	17	1979, 1981
TZ04	E. Huichautla	hybrid	Teziutlán	19.783	-97.600	14	1979
TZ05	Teziutlán	hybrid	Teziutlán	19.817	-97.367	15	1979
TZ06	R. Palenquillo	hybrid	Teziutlán	19.650	-97.117	16	1979
TZ07	Cofre Perote	hybrid	Teziutlán	19.567	-97.100	15	1979
TZ08	Orizaba	*Pipilo ocai*	Teziutlán	19.075	-97.331	47	1981, 2009
TZ09	Oaxaca	*Pipilo ocai*	Teziutlán	17.167	-96.633	15	1981
TV01	Cacoma	*Pipilo ocai*	Transvolcanic	19.851	-104.453	30	2008
TV02	Colima	*Pipilo ocai*	Transvolcanic	19.633	-103.621	30	2008
TV03	Uruapan	hybrid	Transvolcanic	19.491	-102.006	21	2008
TV04	Ichaqueo	hybrid	Transvolcanic	19.578	-101.147	31	2008
TV05	Contepec	hybrid	Transvolcanic	19.965	-100.159	29	2008
TV06	Ocuilan	hybrid	Transvolcanic	18.982	-99.377	29	2009
TV07	Popocatépetl - Iztaccíhuatl massif	hybrid	Transvolcanic	19.101	-98.594	31	2009
TV08	Acajete	hybrid	Transvolcanic	19.149	-97.926	29	2008
TV09	Orizaba	*Pipilo maculatus*	Transvolcanic	19.075	-97.331	44	1981, 2009
TV10	Oaxaca	*Pipilo maculatus*	Transvolcanic	17.167	-96.633	14	1981
	total					460	

Table 3.2. Pairwise population estimates of F_{ST} across all loci

	TZ03	TZ04	TZ05	TZ06	TZ07	TV03	TV04	TV05	TV06	TV07	TV08	TZ01	TZ02	TV09	TV10	TZ08	TZ09	TV01	TV02
TZ03	-																		
TZ04	0.0420	-																	
TZ05	0.0468	0.0409	-																
TZ06	0.0686	0.0563	0.0456	-															
TZ07	0.0888	0.0756	0.0612	0.0358	-														
TV03	0.1059	0.1091	0.1066	0.1082	0.1080	-													
TV04	0.1016	0.1074	0.1052	0.1110	0.1110	0.0558	-												
TV05	0.0607	0.0652	0.0730	0.0839	0.0962	0.0784	0.0718	-											
TV06	0.0437	0.0473	0.0545	0.0690	0.0914	0.1102	0.1058	0.0681	-										
TV07	0.0391	0.0456	0.0530	0.0744	0.0984	0.1177	0.1125	0.0714	0.0255	-									
TV08	0.0605	0.0680	0.0733	0.0896	0.1015	0.0931	0.0853	0.0617	0.0843	0.0771	-								
TZ01	0.0637	0.0756	0.0887	0.1215	0.1403	0.1331	0.1296	0.0731	0.0741	0.0740	0.0847	-							
TZ02	0.0482	0.0572	0.0679	0.0942	0.1137	0.1172	0.1120	0.0613	0.0518	0.0538	0.0731	0.0403	-						
TV09	0.0396	0.0429	0.0505	0.0723	0.0983	0.1352	0.1313	0.0809	0.0359	0.0303	0.0693	0.0831	0.0587	-					
TV10	0.0710	0.0734	0.0800	0.1008	0.1154	0.1369	0.1302	0.0881	0.0925	0.0879	0.0671	0.1068	0.0890	0.0684	-				
TZ08	0.0915	0.0772	0.0616	0.0317	0.0358	0.1204	0.1232	0.1033	0.0815	0.0910	0.1134	0.1570	0.1261	0.0955	0.1273	-			
TZ09	0.1267	0.1183	0.1067	0.0775	0.0735	0.1172	0.1232	0.1186	0.1232	0.1333	0.1308	0.1780	0.1523	0.1408	0.1520	0.0618	-		
TV01	0.1225	0.1250	0.1187	0.1156	0.1126	0.0705	0.0711	0.0929	0.1225	0.1325	0.1072	0.1557	0.1352	0.1494	0.1552	0.1180	0.1170	-	
TV02	0.1044	0.1050	0.1007	0.0951	0.0932	0.0638	0.0664	0.0756	0.1047	0.1149	0.0955	0.1345	0.1154	0.1305	0.1370	0.0984	0.1021	0.0562	-

Table 3.3. Loci with extreme parameter values and outlier F_{ST} values for each transect

	Teziutlán	Transvolcanic
α lower credible interval greater than zero	967	1514
α upper credible interval less than zero	1276	728
β lower credible interval greater than zero	81	728
β upper credible interval less than zero	32	23
F_{ST} outliers	231	292
α lower credible interval greater than zero + F_{ST} outlier	18	102
α upper credible interval less than zero + F_{ST} outlier	68	40
β lower credible interval greater than zero + F_{ST} outlier	43	17
β upper credible interval less than zero + F_{ST} outlier	0	0

Figures

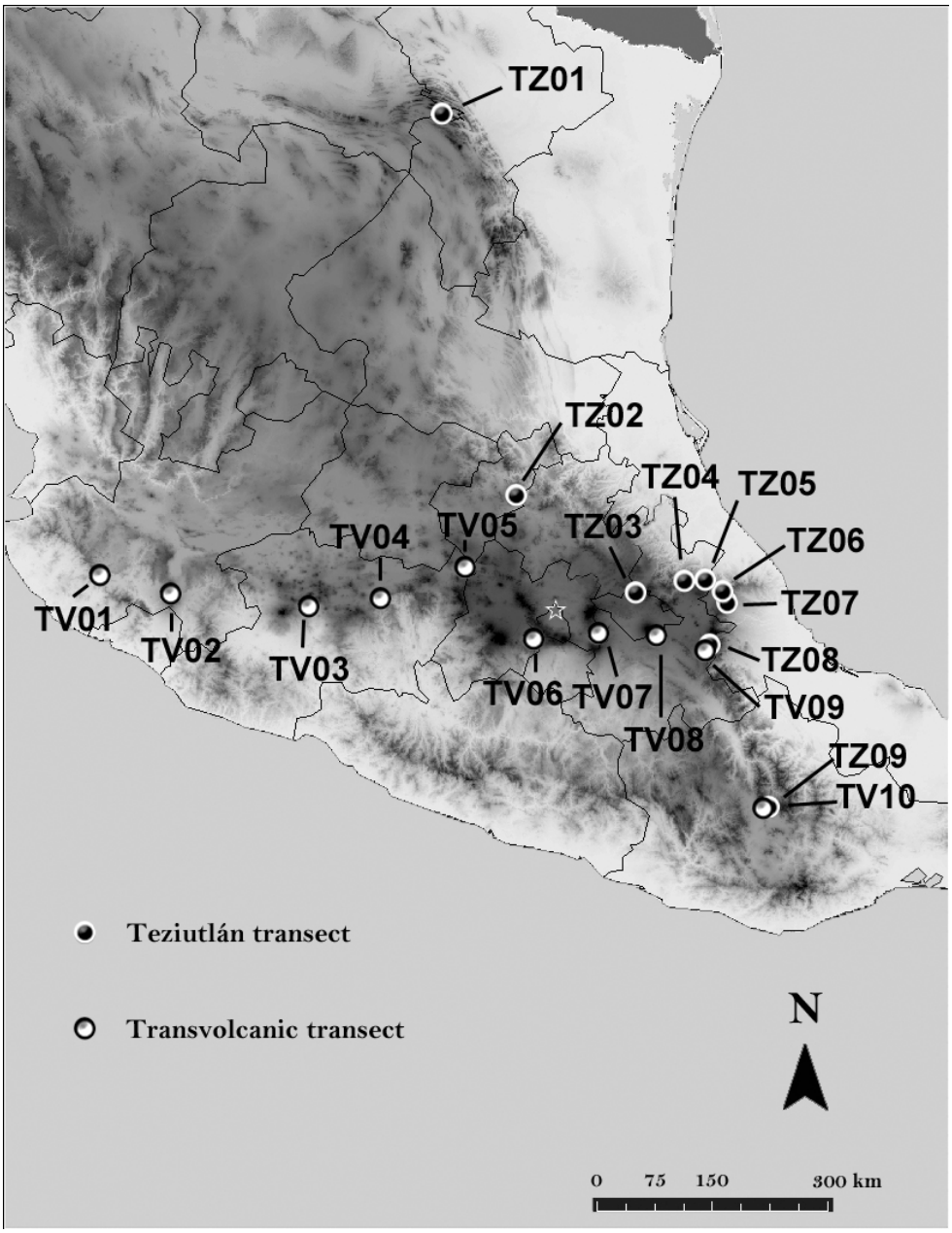


Figure 3.1. Map of sampling locations for the two hybrid transects in Mexico

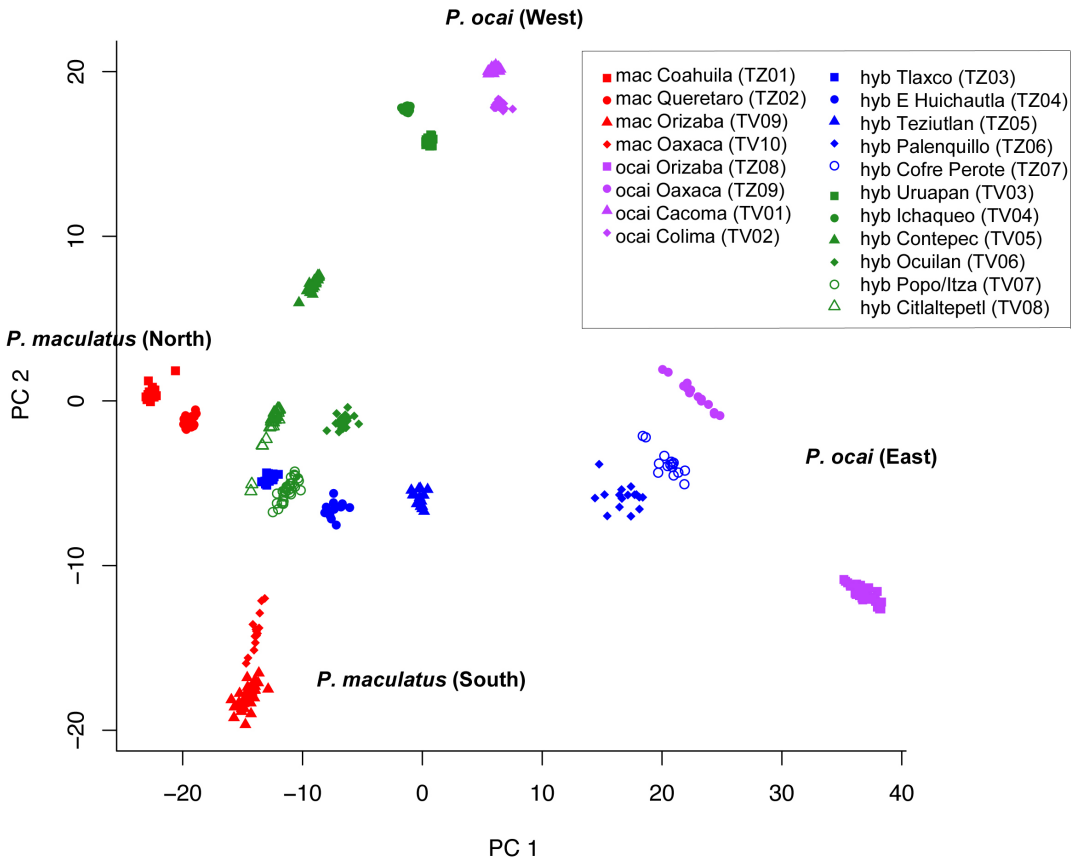


Figure 3.2. PCA derived from genotype probability matrix for 27,968 loci and all individuals from both Teziutlán and Transvolcanic gradients.

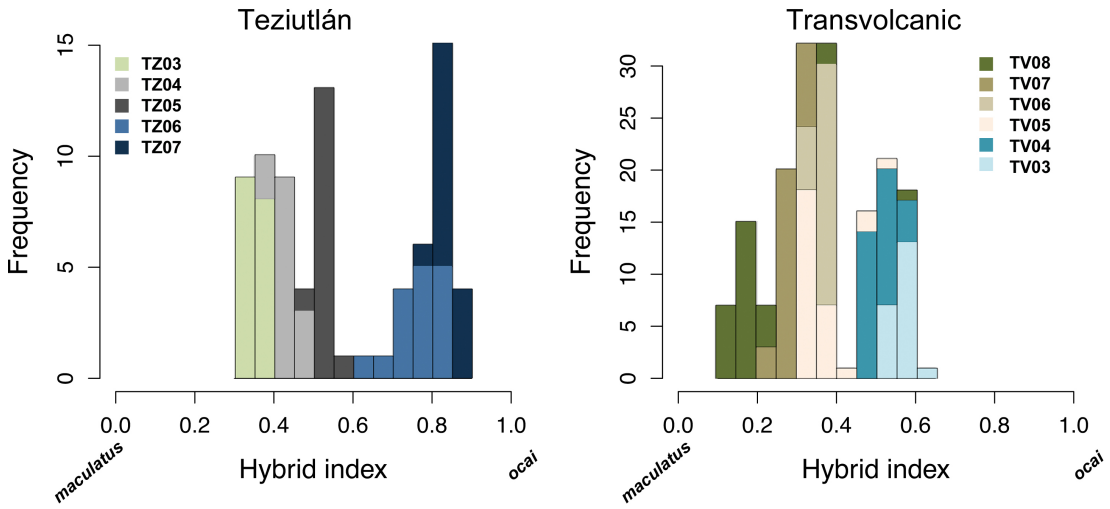


Figure 3.3. Frequency distributions of multi-locus hybrid indices across the admixed populations from both transects.

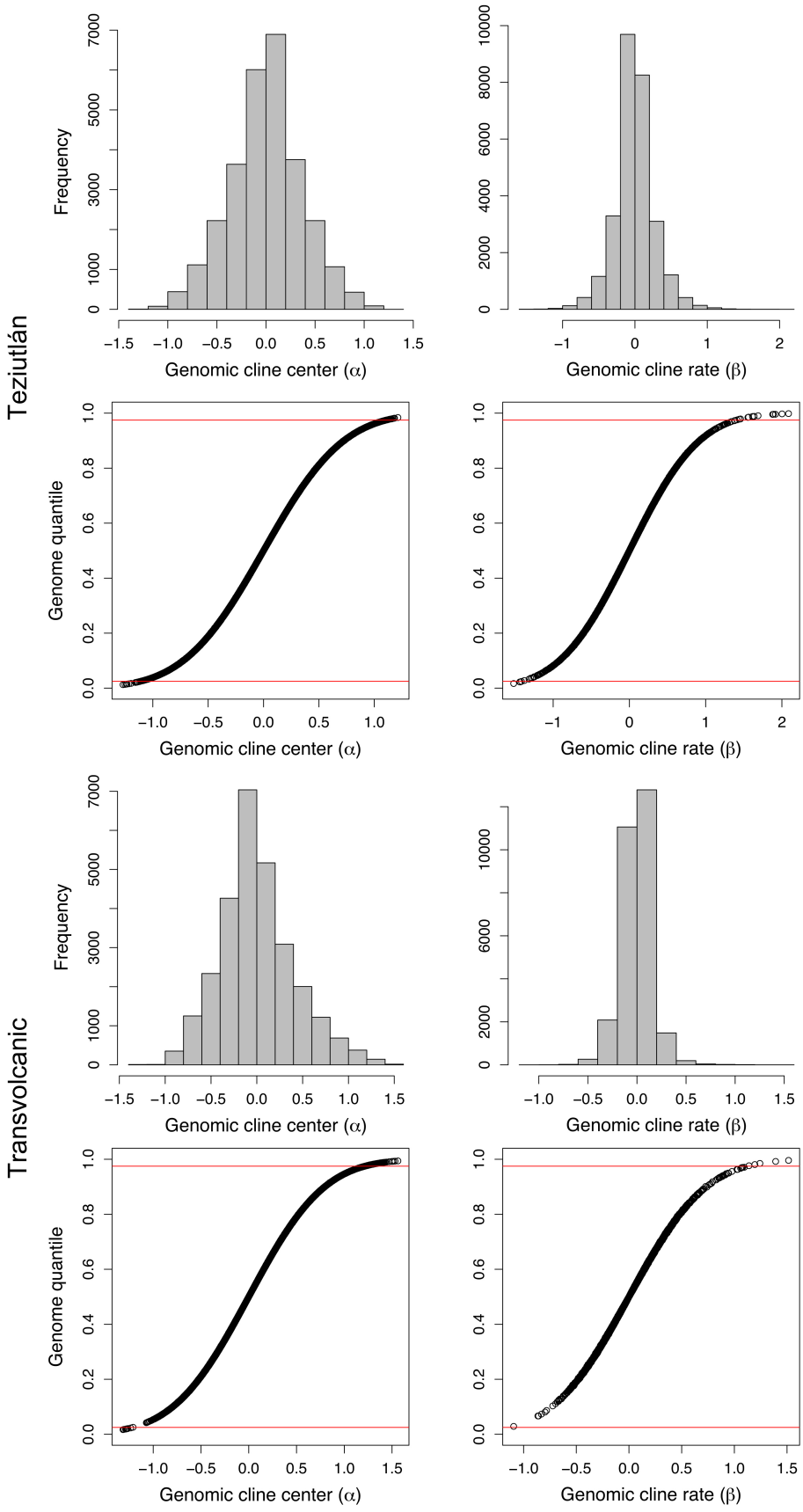

Teziutlán
Transvolcanic
Frequency
Genome quantile
Genomic cline center (α)
Genomic cline rate (β)

Figure 3.4. Frequency distributions and quantiles for genomic cline parameters α and β across both transects. Red lines denote 2.5% and 97.5% quantiles.

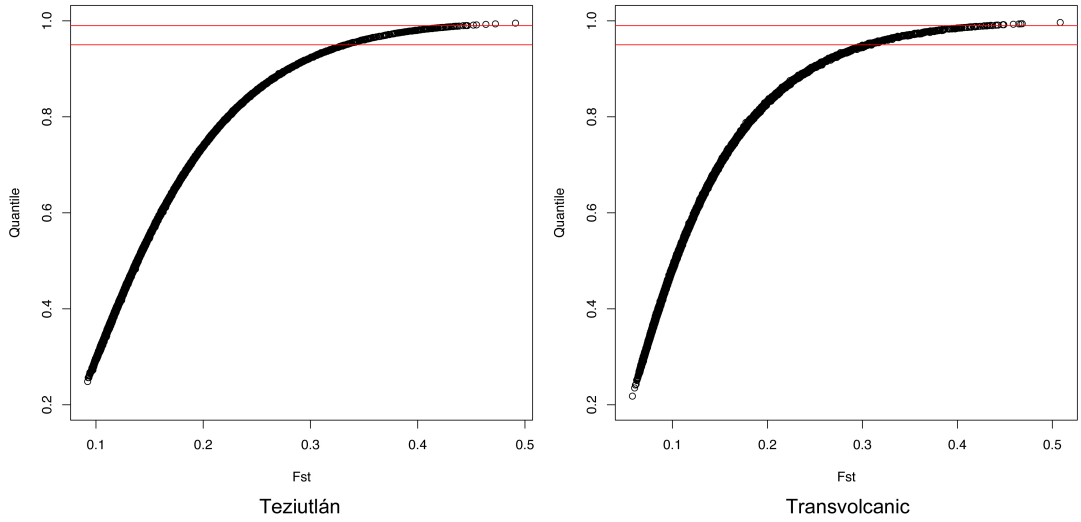


Figure 3.5. F_{ST} estimates between parental population and quantiles for both transects. Red lines denote 95% and 99% quantiles.

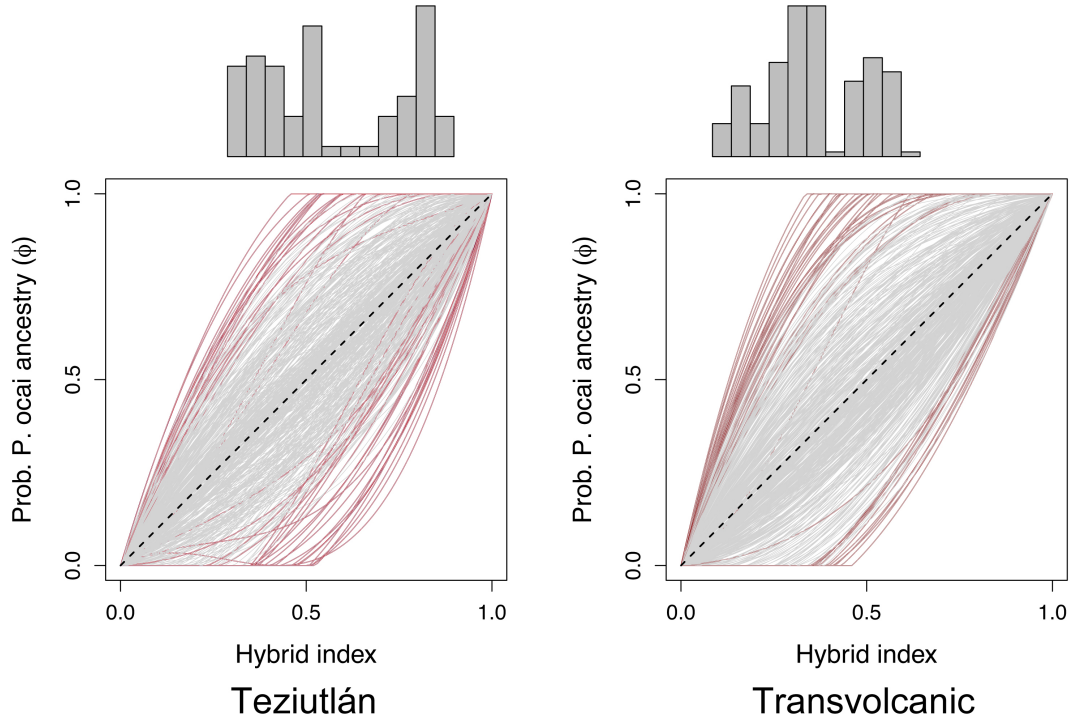


Figure 3.6. The cline graph shows the relationship between the hybrid index and the probability of *P. ocai* ancestry (ϕ) as predicted by each single locus. The hashed line demonstrates where the hybrid index completely predicts the probability of *P. ocai* ancestry (ϕ). Representative loci are plotted; loci with outlier values for α or β are highlighted in red against loci in grey which do not exhibit extreme parameters for α or β. The frequency distributions of the observed hybrid indices across the admixed populations for each transect appear at the top of the graph.

Supporting Information - Chapter 1, Genomic variation in cline shape across a hybrid zone

Here we present the equations for the tension zone-based cline model that can accommodate spatial shifts in cline location as well as changes in changes in cline shape. Model equations and likelihood calculations are from Brumfield et al. (2001), but are re-presented here to correct a typographical error in Equation (3) of that paper. Along a sampled transect, the estimated frequency of the modeled allele, p_{est}, is described piecewise by a three-part function (Szymura and Barton 1986, 1991) :

$$\begin{aligned} p_{est} &= e^{X(decay_N)}\text{intercept}_\text{N} && \text{(Northern tail)} \\ &= \frac{1}{(1+e^{-4X})} && \text{(Sigmoid center)} \\ &= 1-e^{-X(decay_S)}\text{intercept}_\text{S} && \text{(Southern tail)} \end{aligned} \qquad (1)$$

where $decay_N = 2\sqrt{\theta_N}$, $decay_S = 2\sqrt{\theta_S}$, $intercept_N = \frac{t_S}{(t_N + t_S + t_N t_S)}$,

$intercept_S = \frac{t_N}{(t_N + t_S + t_N t_S)}$, $t_N = \beta_N decay_N$, and $t_S = \beta_S decay_S$. Note that the position x along the transect (measured in km from the northernmost population) gets rescaled for use in Eq. 1 according to $X = \frac{(x-c)}{w}$, where the cline width, $w = \frac{1}{\max \text{slope}}$ and the cline center is *c*. We are interested in a total of eight cline parameters. In addition to *w* and *c*, we seek θ_N and θ_S (the northern and southern tail splice parameters, respectively), β_N and β_S (shape parameters governing the decay of the northern and southern tails, respectively), plus p_{min} and p_{max} (the lowest, and highest modeled allele

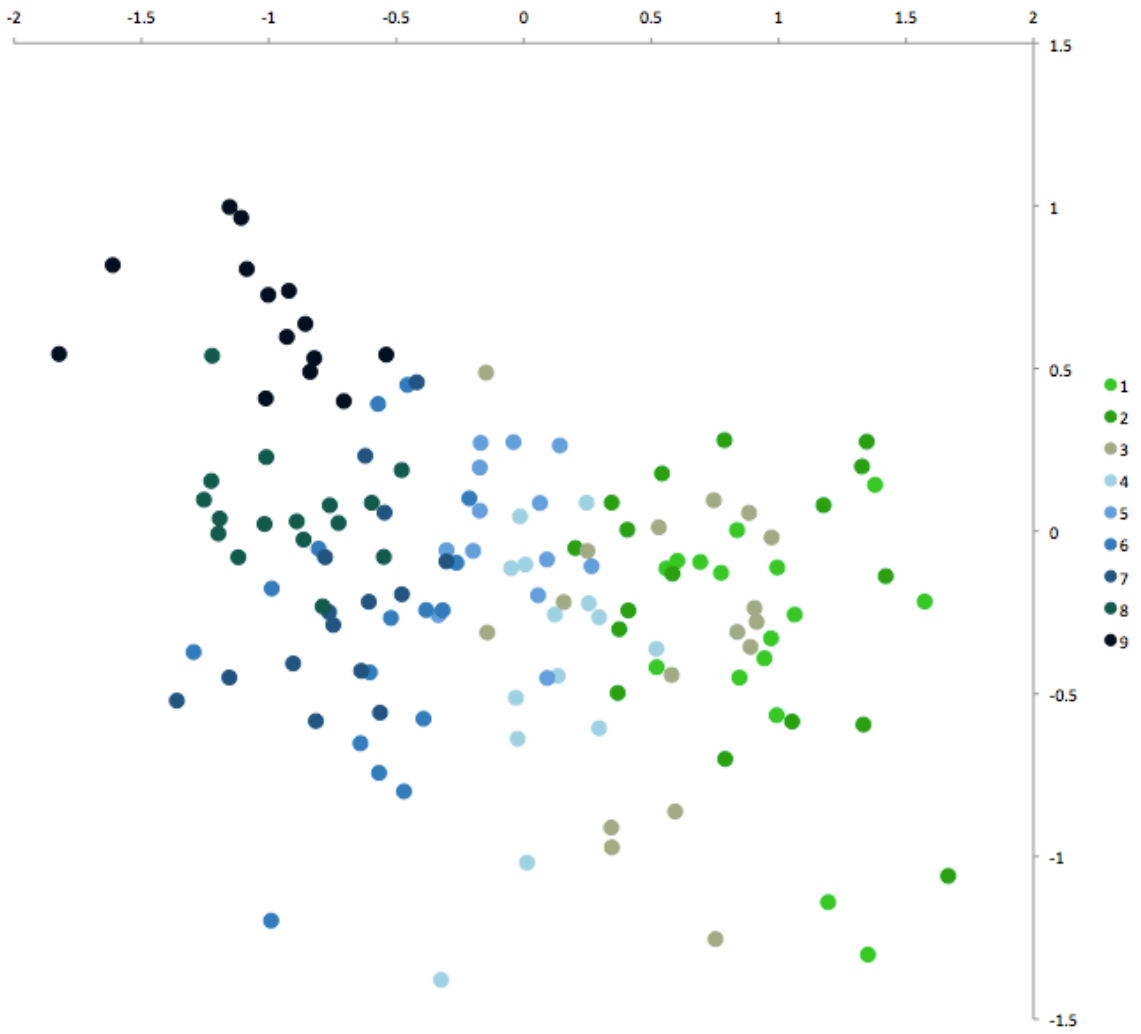
frequencies, respectively). To obtain maximum likelihood estimates of the parameters, we rescale the allele frequencies according to

$$p_{est'} = p_{min} + p_{est}(p_{max} - p_{min}) \tag{2}$$

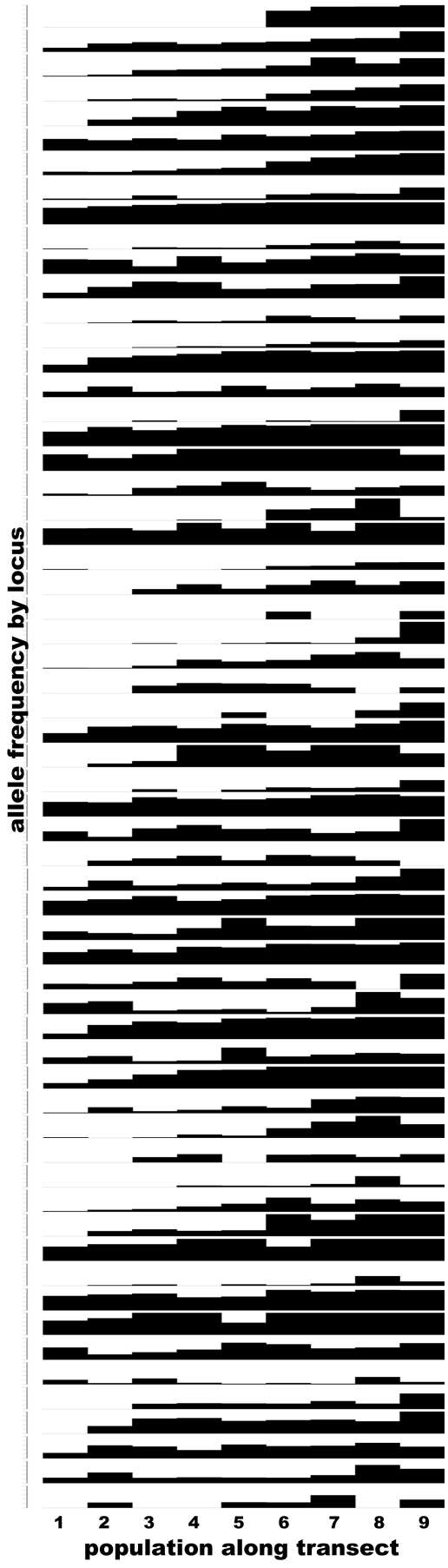
and calculate

$$-\ln L = -n\left[p_{obs}\ln\left(\frac{p_{obs}}{p_{est'}}\right) + (1 - p_{obs})\ln\left(\frac{1 - p_{obs}}{1 - p_{est'}}\right)\right] \tag{3}$$

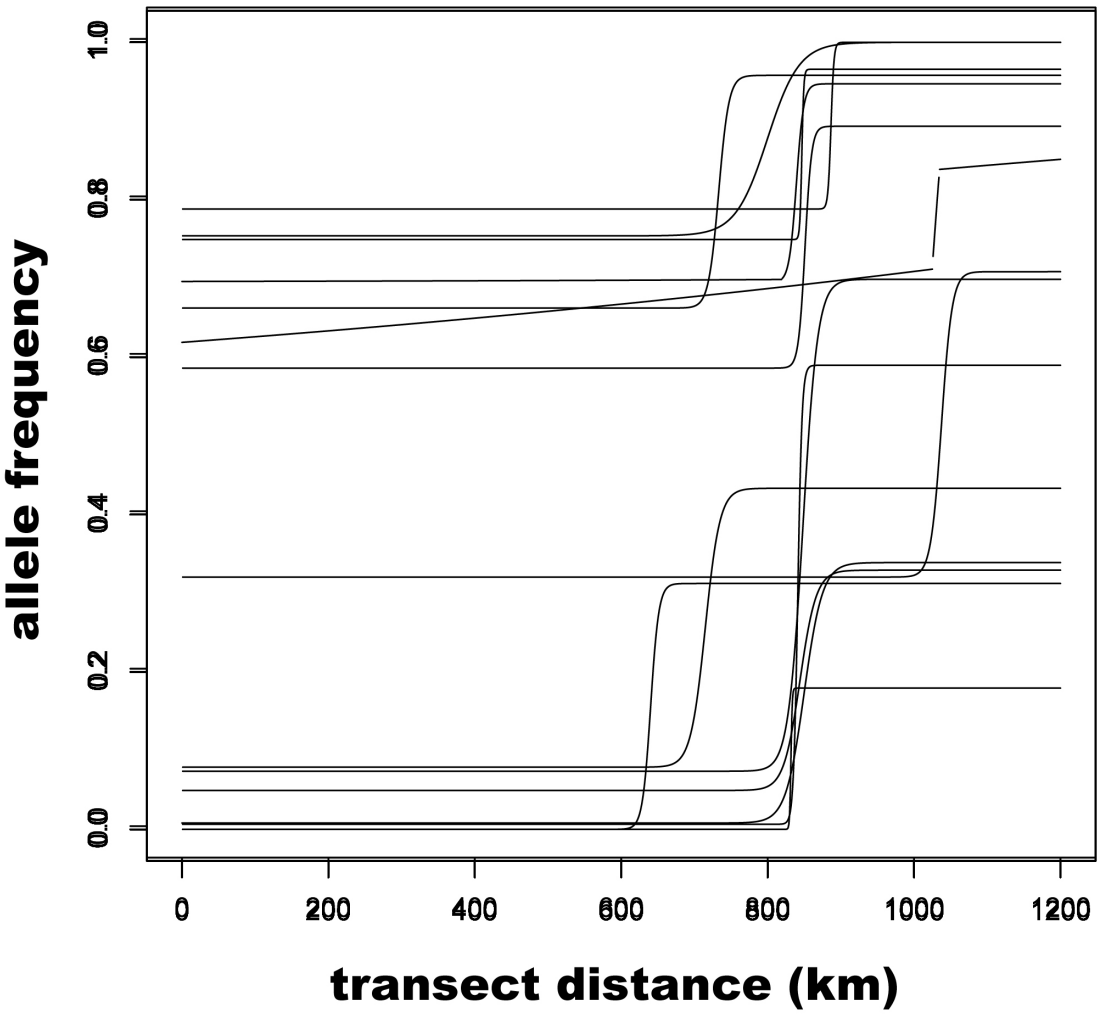
where n is the number of alleles sampled in the population and p_{obs} is the observed frequency of the modeled allele at position X. The total model support is then the sum of population-level log-likelihoods.



Supporting Figure 1.1. Non-metric multidimensional scaling analysis in two dimensions derived from a Jaccard similarity matrix based on the binary AFLP scores at 377 polymorphic loci for all individuals. Individuals are color-coded by sampling population.



Supporting Figure 1.2. Empirical data for all 61 clinal loci displayed in vertically adjacent plots. Populations extend across the X-axis; each pane's Y-axis represents allele frequency.



Supporting Figure 1.3. Exemplary cline fits where change in allele frequency over the transect is relatively small, but cline width is narrow.

Supporting Table 1.1. Tissue and voucher specimen numbers listed by population. Tissue samples were deposited at the US National Museum of Natural History and vouchers were deposited at Louisiana State University Museum of Natural Science.

USNM Tissue No.	Voucher No.	Population
B00625	LSUMZ101112	Coahuila
B00626	LSUMZ101109	Coahuila
B00627	LSUMZ101110	Coahuila
B00628	LSUMZ101113	Coahuila
B00629	LSUMZ101114	Coahuila
B00630	LSUMZ101119	Coahuila
B00631	LSUMZ101121	Coahuila
B00632	LSUMZ101115	Coahuila
B00633	LSUMZ101120	Coahuila
B00634	LSUMZ101122	Coahuila
B00635	LSUMZ101111	Coahuila
B00636	LSUMZ101116	Coahuila
B00637	LSUMZ101117	Coahuila
B00638	LSUMZ101123	Coahuila
B00639	LSUMZ101118	Coahuila
B00640	LSUMZ101146	Coahuila
B00641	LSUMZ101124	Queretaro
B00642	LSUMZ101128	Queretaro
B00643	LSUMZ101076	Queretaro
B00644	LSUMZ101127	Queretaro
B00645	LSUMZ101125	Queretaro
B00646	LSUMZ101126	Queretaro
B00647	LSUMZ101147	Queretaro
B00648	LSUMZ101132	Queretaro
B00649	LSUMZ101129	Queretaro
B00650	LSUMZ101131	Queretaro
B00651	LSUMZ101133	Queretaro
B00652	LSUMZ101130	Queretaro
B00653	LSUMZ101134	Queretaro
B00654	LSUMZ101137	Queretaro
B00655	LSUMZ101136	Queretaro
B00656	LSUMZ101135	Queretaro
B00657	LSUMZ101138	Queretaro
B00067	LSUMZ91346	Tlaxco
B00068	LSUMZ91347	Tlaxco
B00069	LSUMZ91348	Tlaxco
B00070	LSUMZ91349	Tlaxco
B00071	LSUMZ91350	Tlaxco

USNM Tissue No.	Voucher No.	Population
B00072	LSUMZ91351	Tlaxco
B00658	LSUMZ101077	Tlaxco
B00659	LSUMZ101081	Tlaxco
B00660	LSUMZ101082	Tlaxco
B00661	LSUMZ101083	Tlaxco
B00662	LSUMZ101079	Tlaxco
B00663	LSUMZ101080	Tlaxco
B00664	LSUMZ101084	Tlaxco
B00665	LSUMZ101078	Tlaxco
B00666	LSUMZ101148	Tlaxco
B00667	LSUMZ101086	Tlaxco
B00668	LSUMZ101085	Tlaxco
B00001	LSUMZ91332	E. Huichautla
B00002	LSUMZ91333	E. Huichautla
B00006	LSUMZ91334	E. Huichautla
B00007	LSUMZ91335	E. Huichautla
B00008	LSUMZ91336	E. Huichautla
B00009	LSUMZ91337	E. Huichautla
B00010	LSUMZ91338	E. Huichautla
B00011	LSUMZ91339	E. Huichautla
B00032	LSUMZ91340	E. Huichautla
B00033	LSUMZ91341	E. Huichautla
B00034	LSUMZ91342	E. Huichautla
B00035	LSUMZ91343	E. Huichautla
B00036	LSUMZ91344	E. Huichautla
B00037	LSUMZ91345	E. Huichautla
B00012	LSUMZ91325	Teziutlán
B00019	LSUMZ91326	Teziutlán
B00020	LSUMZ91327	Teziutlán
B00021	LSUMZ91328	Teziutlán
B00038	LSUMZ91317	Teziutlán
B00039	LSUMZ91318	Teziutlán
B00040	LSUMZ91319	Teziutlán
B00041	LSUMZ91320	Teziutlán
B00042	LSUMZ91321	Teziutlán
B00046	LSUMZ91329	Teziutlán
B00047	LSUMZ91330	Teziutlán
B00049	LSUMZ91331	Teziutlán
B00050	LSUMZ91322	Teziutlán
B00051	LSUMZ91323	Teziutlán
B00052	LSUMZ91324	Teziutlán
B00004	LSUMZ91316	R. Palenquillo
B00014	LSUMZ91304	R. Palenquillo
B00015	LSUMZ91305	R. Palenquillo
B00017	LSUMZ91306	R. Palenquillo
B00018	LSUMZ91307	R. Palenquillo

USNM Tissue No.	Voucher No.	Population
B00059	LSUMZ91304	R. Palenquillo
B00060	LSUMZ91305	R. Palenquillo
B00061	LSUMZ91310	R. Palenquillo
B00062	LSUMZ91311	R. Palenquillo
B00063	LSUMZ91312	R. Palenquillo
B00064	LSUMZ91313	R. Palenquillo
B00065	LSUMZ91314	R. Palenquillo
B00066	LSUMZ91315	R. Palenquillo
B00669	LSUMZ101058	R. Palenquillo
B00670	LSUMZ101075	R. Palenquillo
B00671	LSUMZ101074	R. Palenquillo
B00003	LSUMZ91289	Cofre Perote
B00005	LSUMZ91290	Cofre Perote
B00024	LSUMZ91291	Cofre Perote
B00025	LSUMZ91292	Cofre Perote
B00026	LSUMZ91293	Cofre Perote
B00027	LSUMZ91294	Cofre Perote
B00028	LSUMZ91295	Cofre Perote
B00029	LSUMZ91296	Cofre Perote
B00030	LSUMZ91297	Cofre Perote
B00053	LSUMZ91298	Cofre Perote
B00054	LSUMZ91299	Cofre Perote
B00055	LSUMZ91300	Cofre Perote
B00056	LSUMZ91301	Cofre Perote
B00057	LSUMZ91302	Cofre Perote
B00058	LSUMZ91303	Cofre Perote
B00676	LSUMZ101043	Orizaba
B00679	LSUMZ101044	Orizaba
B00683	LSUMZ101045	Orizaba
B00684	LSUMZ101046	Orizaba
B00688	LSUMZ101051	Orizaba
B00689	LSUMZ101048	Orizaba
B00690	LSUMZ101049	Orizaba
B00691	LSUMZ101050	Orizaba
B00692	LSUMZ101047	Orizaba
B00695	LSUMZ101055	Orizaba
B00696	LSUMZ101052	Orizaba
B00697	LSUMZ101056	Orizaba
B00698	LSUMZ101054	Orizaba
B00699	LSUMZ101073	Orizaba
B00700	LSUMZ101053	Orizaba
B00701	LSUMZ101057	Orizaba
B00703	LSUMZ101060	Oaxaca
B00704	LSUMZ101067	Oaxaca
B00705	LSUMZ101063	Oaxaca
B00706	LSUMZ101065	Oaxaca

USNM Tissue No.	Voucher No.	Population
B00707	LSUMZ101064	Oaxaca
B00709	LSUMZ101066	Oaxaca
B00710	LSUMZ101061	Oaxaca
B00711	LSUMZ101059	Oaxaca
B00712	LSUMZ101068	Oaxaca
B00714	LSUMZ101071	Oaxaca
B00716	LSUMZ101070	Oaxaca
B00717	LSUMZ101069	Oaxaca
B00718	LSUMZ101480	Oaxaca
B00723	LSUMZ101072	Oaxaca

Supporting Table 1.2. Inferred cline parameters for all loci and cross-transect locus-specific F_{ST}

locus	Type	center	width	theta N	theta S	BN	BS	pmin	pmax	transect-wide F_{ST}
mtDNA	mtDNA RFLP	841.54	14.66	0.0145	0.7567	25014685735	117314165095	0.0001	0.9642	0.859
DIP1	isozyme	923.00	436.11	0.0683	0.1255	2629207686972	1471177871213	0.2719	0.9999	0.122
GOT1	isozyme	802.45	337.90	0.4025	0.1805	997679893923	3070931114324	0.0384	0.8434	0.283
EDH	isozyme	848.29	40.33	0.7872	9305269248	215854781821	7	0.0698	0.6947	0.379
TF	isozyme	728.00	200.74	0.1175	0.1948	85	568090617618	0.0001	0.9035	0.424
EST2	isozyme	848.36	102.49	0.3629	0.8888	88	500401715077	0.4908	0.9041	0.119
GPI	isozyme	842.77	51.05	0.2269	0.3218	40	36	0.1551	0.9999	0.332
TRI2	isozyme	976.34	450.57	0.9697	0.5055	4214766170223	3826153990072	0.0606	0.6721	0.155
6PGD	isozyme	759.81	112.04	0.2564	0.6586	181969208310	927608439613	0.7985	0.9999	0.021
DIP2	isozyme	844.21	44.40	0.5625	0.7993	338662616126	374616181516	0.0453	0.3252	0.094
01_114	AFLP	851.07	17.83	0.2084	0.8584	19361033857	65119090017	0.5861	0.8933	0.293
01_141	AFLP	917.00	283.80	0.1611	0.9907	256	2110742974933	0.2809	0.9999	0.147
01_148	AFLP	816.44	449.71	0.0144	0.7560	2160840609	980268095	0.0001	0.3536	0.101
01_203	AFLP	842.79	169.82	0.3931	0.8680	298204932220	752872214961	0.0001	0.3357	0.124
01_220	AFLP	582.39	472.52	0.9702	0.1638	4156905602604	1624572537086	0.3641	0.9999	0.277
01_409	AFLP	666.91	1102.91	0.9770	0.9999	6052193617971	21	0.2164	0.8339	0.040
02_215	AFLP	1145.39	322.91	1.0000	0.0049	343	3975	0.0001	0.9982	0.301
02_333	AFLP	799.60	86.96	0.4281	0.4179	312944186829	286059191063	0.7541	0.9999	0.134
02_342	AFLP	733.17	27.81	0.3149	0.9636	107965818220	15167009022	0.6623	0.9580	0.221
03_140	AFLP	714.84	40.00	0.9928	0.9974	3606	1211	0.0790	0.4334	0.089
03_402	AFLP	841.16	11.52	0.9617	0.7706	48538338639	90299182601	0.0065	0.5897	0.649
03_407	AFLP	886.53	8.00	0.4778	0.4957	34119456739	19377420290	0.7881	0.9999	0.177
04_125	AFLP	849.91	50.06	0.9398	0.9738	84	480	0.0081	0.3390	0.166
04_130	AFLP	743.66	189.04	0.3543	0.6820	1532682836951	1888792059802	0.0001	0.5282	0.276
04_218	AFLP	831.83	3.46	0.0043	0.1881	647462468	31618008613	0.0001	0.1795	0.321
04_428	AFLP	946.40	130.77	0.8888	0.3629	638482121986	113	0.0001	0.9999	0.551

locus	Type	center	width	theta N	theta S	BN	BS	pmin	pmax	transect-wide F_{ST}
05_120	AFLP	790.75	149.75	0.2173	0.7940	1098557461678	642515851281	0.0312	0.6093	0.264
05_187	AFLP	640.17	22.37	0.6656	0.1259	75479880620	139826931125	0.0001	0.3124	0.226
05_190	AFLP	934.30	188.84	4619326976	2573597440	109	62	0.0001	0.7186	0.458
05_216	AFLP	730.53	809.00	0.6320	0.6660	4667265447321	5180115450923	0.4648	0.9999	0.104
05_225	AFLP	732.34	22.50	0.0026	0.3166	442	84055321352	0.0001	0.8956	0.546
05_240	AFLP	973.32	439.78	0.1502	0.3162	753744322942	3745930469394	0.0001	0.6349	0.191
05_258	AFLP	846.70	5.22	0.3801	0.7685	49917466010	23985803914	0.7494	0.9657	0.070
09_165	AFLP	1026.90	40.06	0.0000	0.0000	27	39	0.4796	0.9999	0.141
09_215	AFLP	917.00	1217.73	1.0000	1.0000	1218	1218	0.0001	0.8751	0.167
09_217	AFLP	917.31	99.39	0.9468	0.9212	99	99	0.2964	0.9999	0.247
09_232	AFLP	839.29	19.78	0.0002	0.4609	2333	41245963	0.6950	0.9472	0.066
09_287	AFLP	819.92	160.98	0.7992	0.8365	1444386498490	1079119584783	0.3232	0.9999	0.297
09_322	AFLP	802.08	92.95	0.8371	0.4432	918523525571	374496772682	0.5977	0.9550	0.171
10_192	AFLP	1037.47	29.26	0.9357	0.0870	175867	1098	0.3206	0.7086	0.219
10_194	AFLP	863.04	7.95	0.4494	0.1182	1290623850	67133132194	0.2990	0.8753	0.469
10_216	AFLP	567.84	497.88	0.2305	0.2612	3461912019282	1112863523894	0.2532	0.9999	0.294
10_288	AFLP	848.61	144.91	0.9997	0.9891	324	160401	0.3017	0.5729	0.146
10_332	AFLP	743.05	138.43	0.2958	0.6210	364450853661	591988341230	0.3420	0.9999	0.396
10_395	AFLP	851.55	12.18	0.3856	0.6984	54882868803	93274107327	0.1707	0.7420	0.256
11_140	AFLP	840.48	52.59	0.1078	0.4331	138382876630	426085388837	0.0340	0.8320	0.591
11_177	AFLP	711.86	37.38	0.8118	0.4175	134376572510	102622223582	0.0001	0.2857	0.127
11_216	AFLP	850.95	105.88	0.0055	0.9851	916037054300	677177499743	0.0001	0.3019	0.245
11_347	AFLP	798.67	90.14	0.0534	0.8992	54881956778	533210250414	0.0741	0.5279	0.152
12_109	AFLP	826.34	43.08	0.0027	0.5789	509	41	0.0001	0.9999	0.501
12_196	AFLP	760.81	332.31	0.3152	0.0527	2133099620383	2655234030043	0.6871	0.9999	0.177
12_207	AFLP	857.84	6.64	0.3748	0.7631	4911262285	48105947299	0.0320	0.3277	0.146
12_277	AFLP	831.77	3.71	0.6796	0.3690	24469771899	2651050797	0.6822	0.9137	0.134
13_137	AFLP	832.10	3.69	0.4217	0.7657	8048501307	6139812471	0.7943	0.9999	0.292
13_159	AFLP	794.09	5.23	0.1118	0.1921	4054573917	43662187994	0.3779	0.6591	0.127
13_194	AFLP	768.74	2382.19	0.0853	1.0000	3466152673965	54	0.0688	0.3624	0.076

locus	Type	center	width	theta N	theta S	BN	BS	pmin	pmax	transect-wide F_{ST}
13_226	AFLP	942.93	580.00	0.3654	0.2041	5272962061005	4647989098076	0.0001	0.8420	0.234
13_229	AFLP	704.25	796.18	0.1880	0.3205	796	796	0.0001	0.9999	0.249
13_243	AFLP	627.87	3303.57	1.0000	0.9999	3304	3304	0.0014	0.9999	0.099
13_264	AFLP	858.30	5.41	0.5714	0.6347	23342906074	53933760037	0.2950	0.7478	0.140
13_295	AFLP	814.66	7.94	0.3207	0.5834	26182761727	57761518464	0.0644	0.2943	0.270

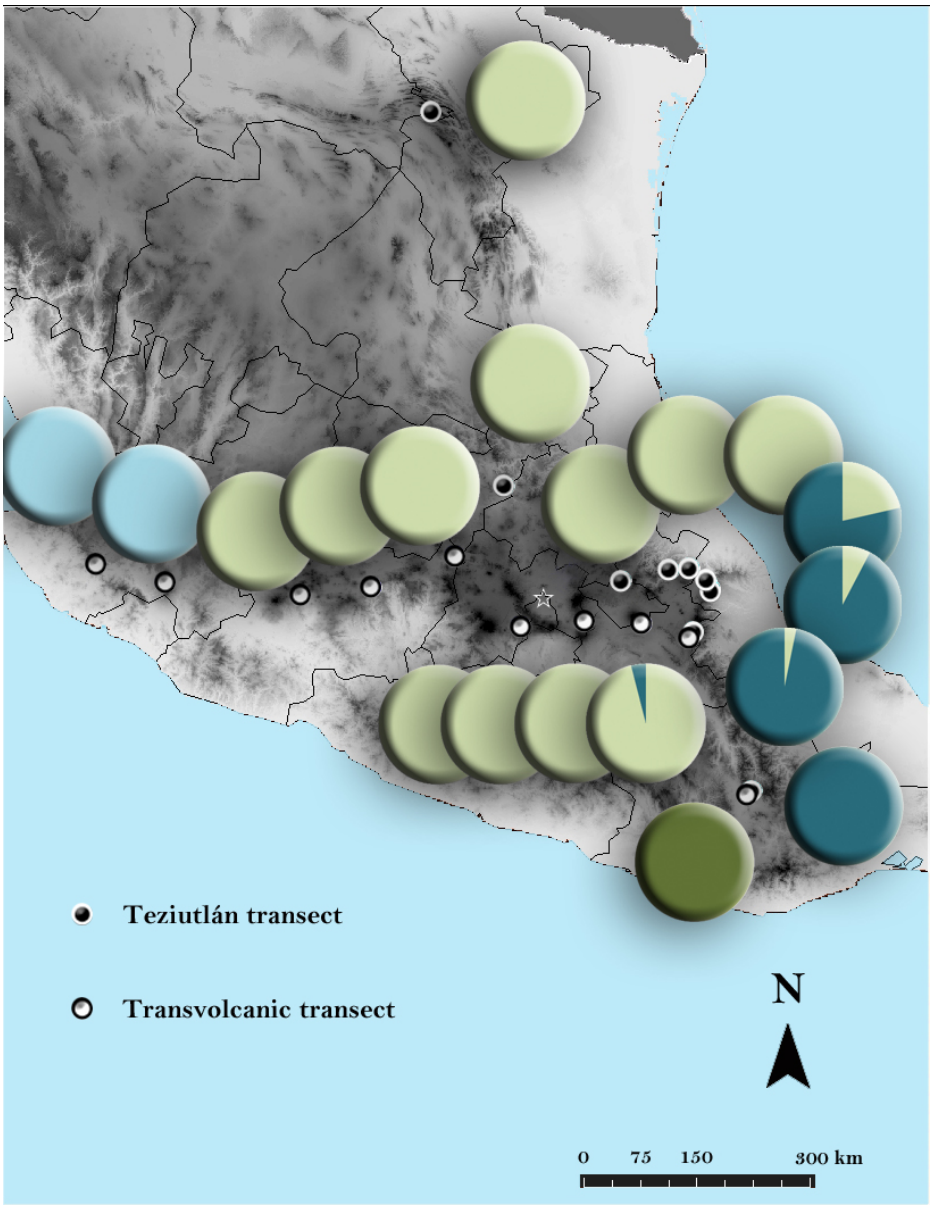
Supporting Information – Chapter 2, Genetic differentiation and habitat connectivity in hybrid zones

Supporting Table 2.1. Tissue and voucher specimen catalog numbers for individuals by collection population. "ENT" numbers are archived at the Universidad Nacional Autónoma de México, Museo de Zoología (tissues also archived in parallel at the US National Museum of Natural History), "LSUMZ" numbers are archived at the Louisiana State University Museum of Natural Science, and "USNM" numbers are archived at the US National Museum of Natural History.

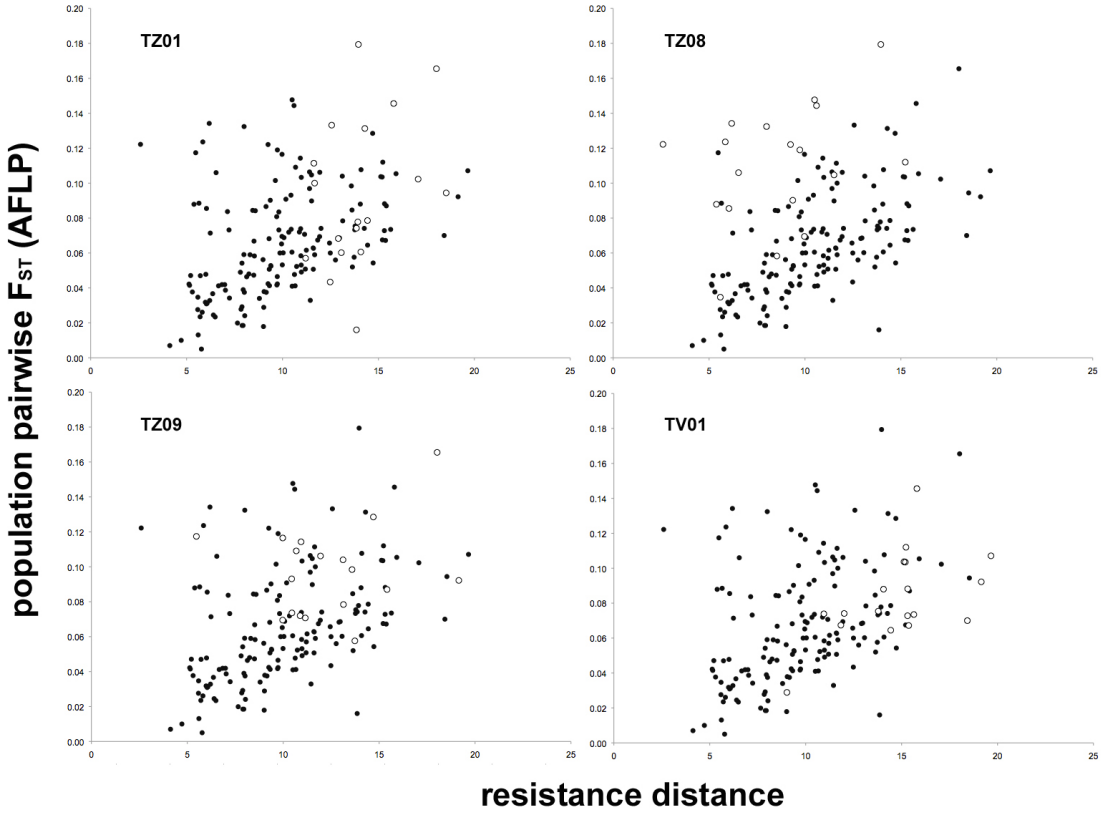
(See attached file Voucher Specimen Numbers)

Supporting Table 2.2. Migrate-n parameter estimated for population scheme 2. A (TZ01 – TZ05, northern Teziutlán gradient); B (TZ06-TZ09 southern Teziutlán gradient); C (TV01-TV01 western Transvolcanic *ocai*); D (TV03-TV09 central and eastern Transvolcanic gradient); E (TV10 Oaxaca *maculatus*).

Parameter	2.50%	Mode	Median	97.50%
theta A	0.0002	0.0026	0.0027	0.0050
theta B	0.0000	0.0022	0.0024	0.0045
theta C	0.0000	0.0016	0.0018	0.0040
theta D	0.0025	0.0061	0.0067	0.0121
theta E	0.0000	0.0012	0.0015	0.0035
migration				
B -> A	0.0000	0.3000	71.0000	332.7000
C -> A	0.0000	0.3000	69.0000	306.7000
D -> A	0.0000	0.3000	266.3000	882.0000
E -> A	0.0000	0.3000	76.3000	358.7000
A -> B	0.0000	17.7000	193.7000	722.0000
C -> B	0.0000	0.3000	105.7000	489.3000
D -> B	0.0000	0.3000	226.3000	803.3000
E -> B	0.0000	0.3000	102.3000	507.3000
A-> C	0.0000	0.3000	133.7000	644.7000
B-> C	0.0000	0.3000	173.7000	715.3000
C-> C	0.0000	0.3000	139.7000	653.3000
E -> C	0.0000	0.3000	155.0000	700.7000
A-> D	474.7000	980.3000	872.3000	1000.0000
B-> D	0.0000	70.3000	130.3000	414.0000
D -> D	0.0000	0.3000	53.0000	240.0000
E -> D	0.0000	0.3000	61.7000	273.3000
A-> E	0.0000	0.3000	271.0000	855.3000
B-> E	0.0000	0.3000	267.0000	856.7000
C-> E	0.0000	0.3000	239.7000	832.0000
D-> E	0.0000	0.3000	270.3000	861.3000



Supporting Figure 2.1. Mitochondrial haplotype group frequency (parental *maculatus* – light green, Oaxaca *maculatus* – dark green, eastern *ocai* – dark blue, western *ocai* – light blue) plotted by sampling population.



Supporting Figure 2.2. *Circuitscape* resistance distance (X-axis) plotted against pairwise population F_{ST} derived from all AFLP loci (Y-axis); pairwise comparisons involving the parental type populations TZ01 (northern *maculatus*), TZ08 (Orizaba *ocai*), TZ09 (Oaxaca *ocai*), and TV01 (western *ocai*) are denoted in each panel respectively with open circles for comparison to the general pattern (black circles).

Tissue No.	Voucher No.	Population
B00625	LSUMZ101112	TZ01
B00626	LSUMZ101109	TZ01
B00627	LSUMZ101110	TZ01
B00628	LSUMZ101113	TZ01
B00629	LSUMZ101114	TZ01
B00630	LSUMZ101119	TZ01
B00631	LSUMZ101121	TZ01
B00632	LSUMZ101115	TZ01
B00633	LSUMZ101120	TZ01
B00634	LSUMZ101122	TZ01
B00635	LSUMZ101111	TZ01
B00636	LSUMZ101116	TZ01
B00637	LSUMZ101117	TZ01
B00638	LSUMZ101123	TZ01
B00639	LSUMZ101118	TZ01
B00640	LSUMZ101146	TZ01
B00641	LSUMZ101124	TZ02
B00642	LSUMZ101128	TZ02
B00643	LSUMZ101076	TZ02
B00644	LSUMZ101127	TZ02
B00645	LSUMZ101125	TZ02
B00646	LSUMZ101126	TZ02
B00647	LSUMZ101147	TZ02
B00648	LSUMZ101132	TZ02
B00649	LSUMZ101129	TZ02
B00650	LSUMZ101131	TZ02
B00651	LSUMZ101133	TZ02
B00652	LSUMZ101130	TZ02
B00653	LSUMZ101134	TZ02
B00654	LSUMZ101137	TZ02
B00655	LSUMZ101136	TZ02
B00656	LSUMZ101135	TZ02
B00657	LSUMZ101138	TZ02
B00067	LSUMZ91346	TZ03
B00068	LSUMZ91347	TZ03
B00069	LSUMZ91348	TZ03
B00070	LSUMZ91349	TZ03
B00071	LSUMZ91350	TZ03
B00072	LSUMZ91351	TZ03
B00658	LSUMZ101077	TZ03
B00659	LSUMZ101081	TZ03
B00660	LSUMZ101082	TZ03

Tissue No.	Voucher No.	Population
B00661	LSUMZ101083	TZ03
B00662	LSUMZ101079	TZ03
B00663	LSUMZ101080	TZ03
B00664	LSUMZ101084	TZ03
B00665	LSUMZ101078	TZ03
B00666	LSUMZ101148	TZ03
B00667	LSUMZ101086	TZ03
B00668	LSUMZ101085	TZ03
B00001	LSUMZ91332	TZ04
B00002	LSUMZ91333	TZ04
B00006	LSUMZ91334	TZ04
B00007	LSUMZ91335	TZ04
B00008	LSUMZ91336	TZ04
B00009	LSUMZ91337	TZ04
B00010	LSUMZ91338	TZ04
B00011	LSUMZ91339	TZ04
B00032	LSUMZ91340	TZ04
B00033	LSUMZ91341	TZ04
B00034	LSUMZ91342	TZ04
B00035	LSUMZ91343	TZ04
B00036	LSUMZ91344	TZ04
B00037	LSUMZ91345	TZ04
B00012	LSUMZ91325	TZ05
B00019	LSUMZ91326	TZ05
B00020	LSUMZ91327	TZ05
B00021	LSUMZ91328	TZ05
B00038	LSUMZ91317	TZ05
B00039	LSUMZ91318	TZ05
B00040	LSUMZ91319	TZ05
B00041	LSUMZ91320	TZ05
B00042	LSUMZ91321	TZ05
B00046	LSUMZ91329	TZ05
B00047	LSUMZ91330	TZ05
B00049	LSUMZ91331	TZ05
B00050	LSUMZ91322	TZ05
B00051	LSUMZ91323	TZ05
B00052	LSUMZ91324	TZ05
B00004	LSUMZ91316	TZ06
B00014	LSUMZ91304	TZ06
B00015	LSUMZ91305	TZ06
B00017	LSUMZ91306	TZ06
B00018	LSUMZ91307	TZ06

Voucher Specimen Numbers

Tissue No.	Voucher No.	Population
B00059	LSUMZ91304	TZ06
B00060	LSUMZ91305	TZ06
B00061	LSUMZ91310	TZ06
B00062	LSUMZ91311	TZ06
B00063	LSUMZ91312	TZ06
B00064	LSUMZ91313	TZ06
B00065	LSUMZ91314	TZ06
B00066	LSUMZ91315	TZ06
B00669	LSUMZ101058	TZ06
B00670	LSUMZ101075	TZ06
B00671	LSUMZ101074	TZ06
B00003	LSUMZ91289	TZ07
B00005	LSUMZ91290	TZ07
B00024	LSUMZ91291	TZ07
B00025	LSUMZ91292	TZ07
B00026	LSUMZ91293	TZ07
B00027	LSUMZ91294	TZ07
B00028	LSUMZ91295	TZ07
B00029	LSUMZ91296	TZ07
B00030	LSUMZ91297	TZ07
B00053	LSUMZ91298	TZ07
B00054	LSUMZ91299	TZ07
B00055	LSUMZ91300	TZ07
B00056	LSUMZ91301	TZ07
B00057	LSUMZ91302	TZ07
B00058	LSUMZ91303	TZ07
B00676	LSUMZ101043	TZ08
B00679	LSUMZ101044	TZ08
B00683	LSUMZ101045	TZ08
B00684	LSUMZ101046	TZ08
B00688	LSUMZ101051	TZ08
B00689	LSUMZ101048	TZ08
B00690	LSUMZ101049	TZ08
B00691	LSUMZ101050	TZ08
B00692	LSUMZ101047	TZ08
B00695	LSUMZ101055	TZ08
B00696	LSUMZ101052	TZ08
B00697	LSUMZ101056	TZ08
B00698	LSUMZ101054	TZ08
B00699	LSUMZ101073	TZ08
B00700	LSUMZ101053	TZ08
B00701	LSUMZ101057	TZ08

Tissue No.	Voucher No.	Population
B27897		TZ08
B27902		TZ08
B27903		TZ08
B27904		TZ08
B27906		TZ08
B27907		TZ08
B27908		TZ08
B27909		TZ08
B27910		TZ08
B27912		TZ08
B27914		TZ08
B27918		TZ08
B27919		TZ08
B27920		TZ08
B27921		TZ08
B27922		TZ08
B27923		TZ08
B27924		TZ08
ENT09-017	ENT09-017	TZ08
ENT09-021	ENT09-021	TZ08
ENT09-024	ENT09-024	TZ08
ENT09-029	ENT09-029	TZ08
ENT09-030	ENT09-030	TZ08
ENT09-031	ENT09-031	TZ08
ENT09-035	ENT09-035	TZ08
ENT09-036	ENT09-036	TZ08
ENT09-038	ENT09-038	TZ08
ENT09-041	ENT09-041	TZ08
ENT09-042	ENT09-042	TZ08
ENT09-044	ENT09-044	TZ08
ENT09-045	ENT09-045	TZ08
ENT09-046	ENT09-046	TZ08
B00703	LSUMZ101060	TZ09
B00704	LSUMZ101067	TZ09
B00705	LSUMZ101063	TZ09
B00706	LSUMZ101065	TZ09
B00707	LSUMZ101064	TZ09
B00709	LSUMZ101066	TZ09
B00710	LSUMZ101061	TZ09
B00711	LSUMZ101059	TZ09
B00712	LSUMZ101068	TZ09
B00714	LSUMZ101071	TZ09

Voucher Specimen Numbers

Tissue No.	Voucher No.	Population
B00716	LSUMZ101070	TZ09
B00717	LSUMZ101069	TZ09
B00718	LSUMZ101480	TZ09
B00723	LSUMZ101072	TZ09
B25805		TV01
B25806		TV01
B25807		TV01
B25808		TV01
B25809		TV01
B25810		TV01
B25811		TV01
B25812		TV01
B25813		TV01
B25814		TV01
B25815		TV01
B25816		TV01
ENT032	ENT032	TV01
ENT034	ENT034	TV01
ENT037	ENT037	TV01
ENT040	ENT040	TV01
ENT044	ENT044	TV01
ENT052	ENT052	TV01
ENT053	ENT053	TV01
ENT053b	ENT053b	TV01
ENT054	ENT054	TV01
ENT055	ENT055	TV01
ENT062	ENT062	TV01
ENT063	ENT063	TV01
ENT064	ENT064	TV01
ENT065	ENT065	TV01
ENT069	ENT069	TV01
ENT070	ENT070	TV01
ENT071	ENT071	TV01
ENT072	ENT072	TV01
B25818		TV02
B25819		TV02
B25821		TV02
B25822		TV02
B25823		TV02
B25824		TV02
B25826		TV02
B25827		TV02

Voucher Specimen Numbers

Tissue No.	Voucher No.	Population
B25828		TV02
B25829		TV02
B25830		TV02
B25833		TV02
B25834		TV02
B25835		TV02
B25836		TV02
ENT080	ENT080	TV02
ENT081	ENT081	TV02
ENT082	ENT082	TV02
ENT083	ENT083	TV02
ENT085	ENT085	TV02
ENT086	ENT086	TV02
ENT088	ENT088	TV02
ENT089	ENT089	TV02
ENT090	ENT090	TV02
ENT091	ENT091	TV02
ENT092	ENT092	TV02
ENT093	ENT093	TV02
ENT101	ENT101	TV02
ENT102	ENT102	TV02
ENT103	ENT103	TV02
B25837		TV03
B25838		TV03
B25839		TV03
B25840		TV03
B25841		TV03
B25842		TV03
B25843		TV03
B25844		TV03
B25845		TV03
B25846		TV03
B25847		TV03
B25850		TV03
ENT108	ENT108	TV03
ENT109	ENT109	TV03
ENT111	ENT111	TV03
ENT114	ENT114	TV03
ENT120	ENT120	TV03
ENT121	ENT121	TV03
ENT122	ENT122	TV03
ENT125	ENT125	TV03

Tissue No.	Voucher No.	Population
ENT126	ENT126	TV03
B25852		TV04
B25853		TV04
B25854		TV04
B25855		TV04
B25856		TV04
B25857		TV04
B25858		TV04
B25859		TV04
B25860		TV04
B25861		TV04
B25862		TV04
B25863		TV04
B25864		TV04
B25865		TV04
B25866		TV04
ENT127	ENT127	TV04
ENT128	ENT128	TV04
ENT129	ENT129	TV04
ENT130	ENT130	TV04
ENT131	ENT131	TV04
ENT132	ENT132	TV04
ENT133	ENT133	TV04
ENT134	ENT134	TV04
ENT135	ENT135	TV04
ENT136	ENT136	TV04
ENT137	ENT137	TV04
ENT138	ENT138	TV04
ENT139	ENT139	TV04
ENT140	ENT140	TV04
ENT141	ENT141	TV04
ENT142	ENT142	TV04
B25867		TV05
B25868		TV05
B25869		TV05
B25870		TV05
B25871		TV05
B25872		TV05
B25873		TV05
B25874		TV05
B25875		TV05
B25876		TV05

Voucher Specimen Numbers

Tissue No.	Voucher No.	Population
B25877		TV05
B25878		TV05
B25879		TV05
B25880		TV05
B25882		TV05
B25883		TV05
B25884		TV05
B25885		TV05
ENT144	ENT144	TV05
ENT145	ENT145	TV05
ENT152	ENT152	TV05
ENT156	ENT156	TV05
ENT158	ENT158	TV05
ENT159	ENT159	TV05
ENT160	ENT160	TV05
ENT162	ENT162	TV05
ENT163	ENT163	TV05
ENT164	ENT164	TV05
ENT168	ENT168	TV05
B27876		TV06
B27877		TV06
B27878		TV06
B27879		TV06
B27880		TV06
B27881		TV06
B27882		TV06
B27883		TV06
B27884		TV06
B27885		TV06
B27886		TV06
B27887		TV06
B27888		TV06
B27889		TV06
B27890		TV06
B27891		TV06
B27892		TV06
ENT09-001	ENT09-001	TV06
ENT09-003	ENT09-003	TV06
ENT09-004	ENT09-004	TV06
ENT09-005	ENT09-005	TV06
ENT09-006	ENT09-006	TV06
ENT09-007	ENT09-007	TV06

Voucher Specimen Numbers

Tissue No.	Voucher No.	Population
ENT09-008	ENT09-008	TV06
ENT09-010	ENT09-010	TV06
ENT09-011	ENT09-011	TV06
ENT09-012	ENT09-012	TV06
ENT09-013	ENT09-013	TV06
ENT09-014	ENT09-014	TV06
B27925		TV07
B27926		TV07
B27927		TV07
B27928		TV07
B27929		TV07
B27931		TV07
B27932		TV07
B27933		TV07
B27934		TV07
B27935		TV07
B27936		TV07
B27937		TV07
B27938		TV07
B27939		TV07
B27940		TV07
B27941		TV07
ENT09-048	ENT09-048	TV07
ENT09-049	ENT09-049	TV07
ENT09-050	ENT09-050	TV07
ENT09-051	ENT09-051	TV07
ENT09-053	ENT09-053	TV07
ENT09-054	ENT09-054	TV07
ENT09-055	ENT09-055	TV07
ENT09-056	ENT09-056	TV07
ENT09-058	ENT09-058	TV07
ENT09-059	ENT09-059	TV07
ENT09-060	ENT09-060	TV07
ENT09-061	ENT09-061	TV07
ENT09-062	ENT09-062	TV07
ENT09-063	ENT09-063	TV07
ENT09-064	ENT09-064	TV07
B25793		TV08
B25794		TV08
B25795		TV08
B25796		TV08
B25797		TV08

Voucher Specimen Numbers

Tissue No.	Voucher No.	Population
B25798		TV08
B25799		TV08
B25800		TV08
B25801		TV08
B25802		TV08
B25803		TV08
ENT005	ENT005	TV08
ENT006	ENT006	TV08
ENT007	ENT007	TV08
ENT008	ENT008	TV08
ENT009	ENT009	TV08
ENT010	ENT010	TV08
ENT011	ENT011	TV08
ENT013	ENT013	TV08
ENT014	ENT014	TV08
ENT015	ENT015	TV08
ENT020	ENT020	TV08
ENT021	ENT021	TV08
ENT022	ENT022	TV08
ENT027	ENT027	TV08
ENT028	ENT028	TV08
ENT029	ENT029	TV08
ENT030	ENT030	TV08
B00672	LSUMZ101087	TV09
B00673	LSUMZ101091	TV09
B00674	LSUMZ101095	TV09
B00675	LSUMZ101089	TV09
B00677	LSUMZ101092	TV09
B00678	LSUMZ101088	TV09
B00680	LSUMZ101090	TV09
B00681	LSUMZ101093	TV09
B00682	LSUMZ101149	TV09
B00685	LSUMZ101094	TV09
B00686	LSUMZ101097	TV09
B00687	LSUMZ101099	TV09
B00693	LSUMZ101096	TV09
B00702	LSUMZ101139	TV09
B27893		TV09
B27894		TV09
B27895		TV09
B27896		TV09
B27898		TV09

Voucher Specimen Numbers

Tissue No.	Voucher No.	Population
B27899		TV09
B27900		TV09
B27901		TV09
B27905		TV09
B27911		TV09
B27913		TV09
B27916		TV09
B27917		TV09
ENT09-016	ENT09-016	TV09
ENT09-018	ENT09-018	TV09
ENT09-019	ENT09-019	TV09
ENT09-020	ENT09-020	TV09
ENT09-022	ENT09-022	TV09
ENT09-023	ENT09-023	TV09
ENT09-025	ENT09-025	TV09
ENT09-026	ENT09-026	TV09
ENT09-027	ENT09-027	TV09
ENT09-028	ENT09-028	TV09
ENT09-032	ENT09-032	TV09
ENT09-033	ENT09-033	TV09
ENT09-034	ENT09-034	TV09
ENT09-037	ENT09-037	TV09
ENT09-039	ENT09-039	TV09
ENT09-040	ENT09-040	TV09
ENT09-043	ENT09-043	TV09
B00713	LSUMZ101100	TV10
B00715	LSUMZ101107	TV10
B00719	LSUMZ101140	TV10
B00720	LSUMZ101102	TV10
B00721	LSUMZ101103	TV10
B00722	LSUMZ101101	TV10
B00724	LSUMZ101108	TV10
B00725	LSUMZ101106	TV10
B00726	LSUMZ101145	TV10
B00727	LSUMZ101143	TV10
B00728	LSUMZ101144	TV10
B00729	LSUMZ101142	TV10
B00730	LSUMZ101141	TV10
B00731	LSUMZ101104	TV10
B00732	LSUMZ101105	TV10

Literature Cited

Andrew RL, Ostevik KL, Ebert DP, Riesberg LH (2012) Adaptation with gene flow across the landscape in a dune sunflower. *Molecular Ecology* **21**, 2078–2091.

Arnold ML (1997) *Natural Hybridization and Evolution* Oxford University Press, Oxford.

Arnold ML (2006) *Evolution Through Genetic Exchange* Oxford University Press, Oxford.

Balding DJ, Nichols RA (1995) A method for quantifying differentiation between populations at multi-allelic loci and its implications for investigating identity and paternity. *Genetica* **96**, 3–12.

Balkenhol N, Gugerli F, Cushman SA*, et al.* (2009) Identifying future research needs in landscape genetics: where to from here? *Landscape Ecology* **24**, 455-463.

Barton NH (2001) The role of hybridization in evolution. *Molecular Ecology* **10**, 551-568.

Barton NH, Baird SJE (1998) Analyse http://helios.bto.ed.ac.uk/evolgen/Mac/Analyse/, Edinburgh.

Barton NH, Bengtsson BO (1986) The barrier to genetic exchange between hybridising populations. *Heredity* **56**, 357-376.

Barton NH, Halliday RB, Hewitt GM (1983) Rare electrophoretic variants in a hybrid zone. *Heredity* **50**, 139-149.

Barton NH, Hewitt GM (1985) Analysis of hybrid zones. *Annual Review of Ecology and Systematics* **16**, 113-148.

Barton NH, Hewitt GM (1989) Adaptation, speciation and hybrid zones. *Nature* **341**, 497-503.

Barton NH, Whitlock MC (1997) The evolution of metapopulations. In: *Metapopulation Biology: Ecology, Genetics, and Evolution* (eds. Hanski I, Gilpin ME), pp. 183-210. Academic Press, San Diego.

Barton NHB, B. O. (1986) The barrier to genetic exchange between hybridizing populations. *Heredity* **57**, 357-376.

Beerli P (2006) Comparison of Bayesian and maximum-likelihood inference of population genetic parameters. *Bioinformatics* **22**, 341-345.

Beerli P, Palczewski M (2010) Unified framework to evaluate panmixia and migration direction among multiple sampling locations. *Genetics* **185**, 313-326.

Bierne N, Welch J, Loire E, Bonhomme F, David P (2011) The coupling hypothesis: why genome scans may fail to map local adaptation genes. *Molecular Ecology* **20**, 2044-2072.

Bolnick DI, Near TJ, Wainwright PC (2006) Body size divergence promotes post-zygotic reproductive isolation in centrarchids. *Evolutionary Ecology Research* **8**, 903-913.

Braun MJ (1983) *Molecular versus morphological and behavioral differentiation across contact zones between closely related avian species*, Louisiana State University.

Bronson CL, Grubb TCJ, Sattler GD, Braun MJ (2005) Reproductive success across the black-capped chickadee (*Poecile atricapillus*) and Carolina chickadee (*P. carolinensis*) hybrid zone in Ohio. *The Auk* **122**, 759-772.

Brown WL, Wilson EO (1956) Character displacement. *Systematic Zoology* **5**, 49-64.

Broyles SB (2002) Hybrid bridges to gene flow: a case study in milkweeds (*Asciepias*). *Evolution* **56**, 1943–1953.

Brumfield RT (1999) *Evolution of brilliant male plumage traits in Manacus: hybrid zones, molecular systematics, and riverine barriers*, University of Maryland.

Brumfield RT, Jernigan RW, McDonald DB, Braun MJ (2001) Evolutionary implications of divergent clines in an avian (*Manacus*: Aves) hybrid zone. *Evolution* **55**, 2070-2087.

Buerkle CA, Gompert Z, Parchman TL (2011) The n=1 constraint in population genomics. *Molecular Ecology* **20**, 1575-1581.

Carling MD, Brumfield RT (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**, 2600-2615.

Carling MD, Zuckerberg B (2011) Spatio-temporal changes in the genetic structure of the *Passerina* bunting hybrid zone. *Molecular Ecology* **20**, 1166-1175.

Carson EW, Tobler M, Minckley WL, Ainsworth RJ, Dowling TE (2012) Relationships between spatio-temporal environmental and genetic variation reveal an important influence of exogenous selection in a pupfish hybrid zone. *Molecular Ecology* **21**, 1209-1222.

Cicero C, Johnson NK (2007) Narrow contact of desert Sage Sparrows (*Amphispiza belli nevadensis* and *A. b. canescens*) in Owens Valley, eastern California: evidence from mitochondrial DNA, morphology, and GIS-based nice models *Ornithological Monographs* **2007**, 87-95.

Cody ML, Brown JH (1970) Character convergence in Mexican finches. *Evolution* **24**, 304-310.

Crow KD, Munehara H, Kanamoto Z*, et al.* (2007) Maintenance of species boundaries despite rampant hybridization between three species of reef fishes (Hexagrammidae): implications for the role of selection. *Biological Journal of the Linnean Society* **91**, 135–147.

Danley PD, Markert JA, Arnegard ME, Kocher TD (2000) Divergence with gene flow in the rock-dwelling cichlids of Lake Malawi. *Evolution* **54**, 1725–1737. .

deQueiroz K (2005) Ernst Mayr and the modern concept of species. *Proceedings of the National Academy of Sciences, USA* **102**, 6600-6607.

Dobzhansky T (1940) Speciation as a stage in evolutionary divergence. *American Naturalist* **74**, 312-321.

Doebeli M, Dieckmann U (2003) Speciation along environmental gradients. *Nature* **421**, 259-264.

Dufkova P, Macholan M, Pialek J (2011) Inference of selection and stochastic effects in the house mouse hybrid zone. *Evolution* **65**, 993–1010.

Earl DA, vonHoldt BM (2011) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* **doi: 10.1007**, s12686-12011-19548-12687.

Elshire RJ, Glaubitz JC, Sun Q*, et al.* (2011) A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLoS One* **6**, PMC3087801.

Emelianov I, František M, Mallet J (2003) Genomic evidence for divergence with gene flow in host races of the larch budmoth. *Proceedings of the Royal Society of London B* **271**, 97-105.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14**, 2611-2620.

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotypes data: linked loci and correled allele frequencies. *Genetics* **164**, 1567-1587.

Falush D, Stephens M, Pritchard JK (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Molecular Ecology Notes* **7**, 574 - 578.

Feder JL, Nosil P (2010) The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. *Evolution* **64**, 1729-1747.

Gavrilets S (1997) Hybrid Zones with Dobzhansky-type epistatic selection. *Evolution* **51**, 1027-1035.

Gavrilets S (2003) Perspective: Models of speciation: what have we learned in 40 years? *Evolution* **57**.

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

Gompert Z, Buerkle CA (2009) A powerful regression-based method for admixture mapping of isolation across the genome of hybrids. *Molecular Ecology* **18**, 1207-1224.

Gompert Z, Buerkle CA (2011) Bayesian estimation of genomic clines. *Molecular Ecology* **20**.

Gompert Z, Fordyce JA, Forister ML, Nice CC (2008) Widespread mito-nuclear discordance with evidence for introgressive hybridization and selective sweeps in *Lycaeides*. *Molecular Ecology* **17**, 5231-5244.

Gompert Z, Forister ML, Fordyce JA*, et al.* (2010a) Bayesian analysis of molecular variance in pyrosequences quantifies population genetic structure across the genome of *Lycaeides* butterflies. *Molecular Ecology* **19**, 2455-2473.

Gompert Z, Lucas LK, Fordyce JA, Forister ML, Nice CC (2010b) Secondary contact between *Lycaeides idas* and *L. melissa* in the Rocky Mountains: extensive admixture and a patchy hybrid zone. *Molecular Ecology* **19**, 3171-3192.

Gompert Z, Lucas LK, Nice CC*, et al.* (2012a) Genomic regions with a history of divergent selection affect fitness of hybrids between two butterfly species. *Evolution*.

Gompert Z, Parchman TL, Buerkle CA (2012b) Genomics of isolation in hybrids. *Philosophical Transactions of the Royal Society B* **367**, 439-450.

Gower JC (1966) Some distance properties of latent roots and vector methods used in multivariate analysis. *Biometrika* **53**, 325-338.

Grahame JW, Wilding CS, Butlin RK (2006) Adaptation to a steep environmental gradient and an associated barrier to gene exchange in *Littorina saxitilis*. *Evolution* **60**, 268-278.

Guiller A, Coutellec-Vreto M-A, Madec L (1996) Genetic relationships among suspected contact zone populations of *Helix aspersa* (Gastropoda: Pulmonata) in Algeria. *Heredity* **77**, 113-129.

Hackett SJ (1996) Molecular phylogenetics and biogeography of tanagers in the genus *Ramphocelus* (Aves). *Molecular Phylogenetics and Evolution* **5**, 368-382.

Haldane JBS (1922) Sex ratio and unisexual sterility in hybrid animals. *Journal of Genetics* **12**, 101.

Hanotte O, Bradley DG, Ochieng JW*, et al.* (2002) African pastoralism: genetic imprints of origins and migrations. *Science* **296**, 336 - 339.

Harr B (2006) Genomic islands of differentiation between house mouse subspecies. *Genome Research* **16**, 730-737.s

Hey J (2004) Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics* **167**, 747-760.

Hey J (2006) Recent advances in assessing gene flow between diverging populations and species. *Current Opinion in Genetics & Development* **16**, 592 - 596.

Holderegger R, Wagner HH (2006) A brief guide to Landscape Genetics *Landscape Ecology* **21**, 793-796.

Holsinger KE, Lewis P (2003) Hickory, software for analysis of geographic structure in genetic data, Storrs.

Jacobsen F, Omland KE (2011) Species tree inference in a recent radiation of orioles (genus Icterus): multiple markers and methods reveal cytonuclear discordance in the northern oriole group. *Molecular Phylogenetics and Evolution* **61**, 460-469.

Kingston SE, Rosel PE (2004) Genetic differentiation among recently diverged Delphinid taxa determined using AFLP markers. *Journal of Heredity* **95**, 1-10.

Kingston SE, Jernigan RW, Fagan WF, Braun D, Braun MJ (in review) Genomic variation in cline shape across a hybrid zone. *Ecology and Evolution*.

Kingston SE, Navarro S AG, García-Trejo EA*, et al.* (in prep) Genetic differentiation and habitat connectivity in hybrid zones.

Li H, Handsaker B, Wysoker A*, et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079.

Macholán M, Munclinger P, Šugerková M*, et al.* (2007) Genetic analysis of autosomal and X-linked markers across a mouse hybrid zone. *Evolution* **61**, 746–771.

Mallet J (2005) Hybridization as an invasion of the genome. *TRENDS in Ecology and Evolution* **20**, 229-227.

Mallet J, Beltran M, Neukirchen W, Linares M (2007) Natural hybridization in heliconiine butterflies: the species boundary as a continuum. *BMC Evolutionary Biology* **7**.

Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology and Evolution* **18**, 189-197.

Martinsen GD, Whitnam TG, Turek RJ, Keim P (2001) Hybrid populations selectively filter gene introgression between species. *Evolution* **55**, 1325-1335.

Mayr E (1963) *Animal species and evolution* Harvard University Press, Cambridge, MA.

McRae B, Beier P (2007) Circuit theory predicts gene flow in plant and animal populations. *Proceedings of the National Academy of Sciences* **104**, 19885-19890.

McRae B, Dickson BG, Keitt TH, Shah VB (2008) Using circuit theory to model connectivity in ecology, evolution, and conservation. *Ecology* **89**, 2712-2724.

Moore WS (1977) An evaluation of narrow hybrid zones in vertebrates. *Quarterly Review of Biology* **52**, 263-278.

Nadeau NJ, Whibley A, Jones RT*, et al.* (2012) Genomic islands of divergence in hybridizing *Heliconius* butterflies identified by large-scale targeted sequencing. *Philosophical Transactions of the Royal Society of London B Biological Sciences* **367**, 343-353.

Nicholson G, Smith AV, Jónsson F*, et al.* (2002) Assessing population differentiation and isolation from singlenucleotide polymorphism data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64**, 695–715.

Niemiller ML, Fitzpatrick BM, Miller BT (2008) Recent divergence with gene flow in Tennessee cave salamanders (Plethodontidae: Gyrinophilus) inferred from gene genealogies. *Molecular Ecology* **17**, 2258–2275.

Nolte AW, Gompert Z, Buerkle CA (2009) Variable patterns of introgression in two sculpin hybrid

zones suggest that genomic isolation differs among

populations. *Molecular Ecology* **18**, 2615-2627.

Nosil P (2008) Speciation with gene flow could be common. *Molecular Ecology* **17**, 2103–2106.

Nosil P, Egan SP, Funk DJ (2008) Heterogeneous genomic differentiation between walking-stick ecotypes: "Isolation by adaptation" and multiple roles for divergent selection. *Evolution* **62**, 316-336.

Nosil P, Feder JL (2012) Genomic divergence during speciation: causes and consequences. *Philosophical Transactions of the Royal Society B* **367**, 332-342.

Nosil P, Funk DJ, Ortiz-Barrientos D (2009) Divergent selection and heterogenous genomic divergence. *Molecular Ecology* **18**, 375-402.

Novembre J, Johnson T, Bryc K*, et al.* (2008) Genes mirror geography within Europe. *Nature* **456**, 98-103.

Orr HA, Turelli M (2001) The evolution of postzygotic isolation: Accumulating Dobzhansky-Muller incompatabilities. *Evolution* **55**, 1085-1094.

Owen-Ashley NT, Butler LK (2004) Androgens, interspecific competition and species replacement in hybridizing warblers. *Proceedings of the Royal Society of London B (Suppl.)* **271**, S498-S500.

Panova M, Hollander J, Johannesson K (2006) Site-specific genetic divergence in parallel hybrid zones suggests nonallopatric evolution of reproductive barriers. *Molecular Ecology* **15**, 4021-4031.

Parchman T, Braun M, Gompert Z, Buerkle CA (in prep) Genomic differentiation and isolation across a manakin hybrid zone.

Parchman TL, Gompert Z, Mudge J*, et al.* (2012) Genome-wide association mapping of an adaptive trait in lodgepole pine. *Molecular Ecology*.

Payseur BA, Krenz JG, Nachman MW (2004) Differential patterns of introgression across the X chromosome in a hybrid zone between two species of house mice. *Evolution* **58**, 2064-2078.

Pialek J, Barton NH (1997) The spread of an advantageous allele across a barrier: the effects of random drift and selection against heterozygotes. *Genetics* **145**.

Pritchard JK, Stephens M, Donnely P (2000) Inference of population structure using multilocus genotype data. *Genetics* **155**, 945-959.

Rieseberg LH, Whitton J, Gardner K (1999) Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. *Genetics* **152**, 713-727.

Rieseberg LH, Widmer A, Arntz AM, Burke JM (2003) The genetic architecture necessary for transgressive segregation is common in both natural and domesticated populations. *Philosphical Transactions of the Royal Society, Biological Sciences* **358**, 1141-1147.

Rohwer S, Bermingham E, Wood C (2001) Plumage and mitochondrial DNA halpotype variation across a moving hybrid zone. *Evolution* **55**, 405-422.

Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning: A Laboratory Manual*, 2 edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Sattler GD, Braun MJ (2000) Morhpometric variation as an indicator of genetic interactions between black-capped and carolina chickadees at a contact zone in the Appalachian mountains. *The Auk* **117**, 427-444.

Seehausen O (2003) Hybridization and adaptive radiation. *Trends in Ecology and Evolution* **19**, 198-207.

Shah VB, McRae B (2008) Circuitscape: a tool for landscape ecology. *Proceedings of the 7th Python in Science Conference* **2008**, 62-66.

Sibley CG (1950) Species formation in the red-eyed towhees of Mexico. *University of California Publications in Zoology* **50**, 109-194.

Sibley CG (1954) Hybridization in the red-eyed towhees of Mexico. *Evolution* **8**, 252-290.

Sibley CG, Sibley FC (1964) Hybridization in the red-eyed towhees of Mexico: the populations of the southeastern plateau region. *The Auk* **81**, 479-504.

Sibley CG, West DA (1958) Hybridization in the red-eyed towhees of Mexico: the eastern plateau populations. *The Condor* **60**, 85-104.

Smit HA, Robinson TJ, VanVuuren BJ (2007) Coalescence methods reveal the impact of vicariance on the spatial genetic structure of *Elephantulus edwardii* (Afrotheria, Macroscelidea). *Molecular Ecology* **16**, 2680–2692.

Sorenson MD, Ast JC, Dimcheff DE, Yuri T, Mindell DP (1999) Primers for a PCR-based approach to mitochondrial genome sequencing in birds and other vertebrates. *Molecular Phylogenetics and Evolution* **12**, 105-114.

Southern EM (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *Journal of Molecular Biology* **98**, 503–517.

Stein AC, Uy AC (2006) Unidirectional introgression of a sexually selected trait across an avian hybrid zone: a role for female choice? *Evolution* **60**, 1476-1485.

Strasburg JL, Rieseberg LH (2008) Molecular demographic history of the annual sunflowers *Helianthus annuus* and *H. petiolaris*—large effective population sizes and rates of long-term gene flow. *Evolution* **6**, 1936-1950.

Szymura JM, Barton NH (1986) Genetic Analysis of a hybrid zone between the fire-bellied toads, *Bombina bombina* and *B. variegata*, near Cracow in southern Poland. *Evolution* **40**, 1141-1159.

Teeter KC, Payseur BA, Harris LW*, et al.* (2008) Genome-wide patterns of gene flow across a house mouse hybrid zone *Genome Research* **18**, 67-76.

Vallender R, Robertson J, Friesen VL, Lovette J (2007) Complex hybridization dynamics between golden-winged and blue-winged warblers (Vermivora chrysoptera and Vermivora pinus) revealed by AFLP, microsatellite, intron and mtDNA markers. *Molecular Ecology* **16**, 2017-2029.

Vines TH, Köhler SC, Thiel M*, et al.* (2003) The maintenance of reproductive isolation in a mosaic hybrid zone between the fire-bellied toads *Bombina bomina* and *B. variegata*. *Evolution* **57**, 1876-1888.

Vos P, Hogers R, Bleeker M*, et al.* (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research* **23**, 4407-4414.

Wang RL, Wakeley J, Hey J (1997) Gene flow and natural selection in the origin of *Drosophila pseudorobscura* and close relatives. *Genetics* **147**, 1091-1106.

Woodruff DS, Gould SJ (1987) Fifty years of interspecific hybridization: genetics and morphometrics of a controlled experiment on the landsnail *Cerion* in Florida Keys. *Evolution* **41**, 1022-1045.

Wu C-I, Ting C-T (2004) Genes and Speciation. *Nature Reviews* **5**, 114-122.

Yuri T, Jernigan RW, Brumfield RT, Bhagabati NK, Braun MJ (2009) The effect of marker choice on estimated levels of introgression across an avian (Pipridae: *Manacus*) hybrid zone. *Molecular Ecology* **18**, 4888–4903.

Zink RM, Weller SJ, Blackwell RC (1998) Molecular phylogenetics of the avian genus *Pipilo* and a biogeography argument for taxonomic uncertainty. *Molecular Phylogenetics and Evolution* **10**, 191-201.