JOURNAL OF Evolutionary Biology

doi: 10.1111/jeb.13033

Heterogeneity and concordance in locus-specific differentiation and introgression between species of towhees

S. E. KINGSTON*†‡, T. L. PARCHMAN§, Z. GOMPERT¶, C. A. BUERKLE** & M. J. BRAUN*†

*Program in Behavior, Ecology, Evolution, and Systematics, University of Maryland, College Park, MD, USA
†National Museum of Natural History, Vertebrate Zoology, Smithsonian Institution, Washington, DC, USA
‡Department of Biology and Coastal Studies Center, Bowdoin College, Brunswick, ME, USA
§Department of Biology, University of Nevada, Reno, NV, USA
¶Department of Biology, Utah State University, Logan, UT, USA
**Department of Botany, University of Wyoming, Laramie, WY, USA

Keywords:

divergence;
introgression;
population genomics;
reproductive isolation.

Abstract

The maintenance or breakdown of reproductive isolation is an observable outcome of secondary contact between species. In cases where hybrids beyond the F1 are formed, the representation of each species' ancestry can vary dramatically among genomic regions. This genomic heterogeneity in ancestry and introgression can offer insight into evolutionary processes, particularly if introgression is compared in multiple hybrid zones. Similarly, considerable heterogeneity exists across the genome in the extent to which populations and species have diverged, reflecting the combined effects of different evolutionary processes on genetic variation. We studied hybridization across two hybrid zones of two phenotypically well-differentiated bird species in Mexico (*Pipilo maculatus* and *P. ocai*), to investigate genomic heterogeneity in differentiation and introgression. Using genotyping-by-sequencing (GBS) and hierarchical Bayesian models, we genotyped 460 birds at over 41 000 single nucleotide polymorphism (SNP) loci. We identified loci exhibiting extreme introgression relative to the genome-wide expectation using a Bayesian genomic cline model. We also estimated locus-specific F_{ST} and identified loci with exceptionally high genetic divergence between the parental species. We found some concordance of locus-specific introgression in the two independent hybrid zones (6–20% of extreme loci shared across zones), reflecting areas of the genome that experience similar gene flow when the species interact. Additionally, heterogeneity in introgression and divergence across the genome revealed another subset of loci under the influence of locally specific factors. These results are consistent with a history in which reproductive isolation has been influenced by a common set of loci in both hybrid zones, but where local environmental and stochastic factors also lead to genomic differentiation.

Introduction

Lineages that evolved in allopatry and hybridize in secondary contact offer a setting in which to study the components and dynamics of reproductive isolation in nature. Contemporary isolation of these lineages results from selection against cross-species migrants as well as from poor fitness of hybrids. However, fitness of hybrids is expected to vary among first-generation hybrids and recombinant, advanced generation hybrids, and may additionally vary among environments. The genetic composition of hybrids observed following secondary contact will reflect the effects of selection, meiotic recombination and chance variation in allele frequencies in finite populations (Barton & Hewitt, 1985, 1989; Gompert *et al.*, 2012b; Nosil & Feder, 2012).

Correspondence: Sarah E. Kingston, Bowdoin College Department of Biology, 6500 College Station, Brunswick, ME 04011, USA.
Tel.: 207-798-7128; fax: 207-725-3405; e-mail: skingsto@bowdoin.edu

Although the influence of divergent selection on reproductive isolation is a long-standing question of interest in evolution, only now are we able to obtain population genomic data and begin to investigate the causes of genomic variation. Consequently, we have much to learn about the patterns of genomic differentiation between species, the genetic architecture of hybrid fitness and isolation between species and how these two factors relate. In hybrid zones, variation in strength of selection should cause heterogeneous patterns of differential introgression, on both geographic and genomic scales (Barton & Hewitt, 1989; Sattler & Braun, 2000; Yuri *et al.*, 2009; Payseur, 2010). Locus to locus variability in levels and patterns of introgression is regularly observed in hybrid zones (Teeter *et al.*, 2008; Nolte *et al.*, 2009; Yuri *et al.*, 2009; Payseur, 2010; Gompert *et al.*, 2012a,b; Kingston *et al.*, 2012; Parchman *et al.*, 2013). The potential of locus-specific analyses for mapping the genetic architecture of reproductive isolation is much greater now than even in the recent past due to the power provided by population genomic scale data sets (Hohenlohe *et al.*, 2010; Gompert *et al.*, 2012a). However, population genomic analyses of differential introgression and isolation across hybrid zones are still limited to a modest group of examples (e.g. Gompert *et al.*, 2012a; Janoušek *et al.*, 2012, 2015; Larson *et al.*, 2013; Parchman *et al.*, 2013).

To assess genome variation associated with reproductive isolation and divergence, we utilize two geographically separated hybrid zones between the same pair of species. Although genome-scale data sets are becoming more common, we are leveraging genomic-level variation in individuals across the dual hybrid zones. Using this powerful dual-zone approach, we investigate the patterns of differentiation and characterize loci that behave consistently despite differences in geographic and population demographic situations. Loci that introgress freely across the hybrid zones will likely reflect a signal similar to that of the multilocus genomic background. Loci reflecting a signal of directional or reduced introgression are potentially involved in selection and isolation. For example, loci associated with incompatibilities or underdominance may also differ similarly from the background genomic signal (Gompert *et al.*, 2012a,b; Parchman *et al.*, 2013; Lindtke & Buerkle, 2015).

Pipilo maculatus (Swainson, 1827) and *P. ocai* (Lawrence, 1867) are two towhee species that come into secondary contact and hybridize in the high-elevation pine/oak scrub habitat of Mexico. Although *P. maculatus* is widely distributed throughout western North America, *P. ocai* is a Mexican endemic. The two largest areas of hybridization between the species are along the mountains of the Sierra Madre Oriental (Teziutlán gradient) and the Transvolcanic Belt (Transvolcanic gradient) in central Mexico (Fig. 1a). These two large gradients offer a rare opportunity to study multiple instances of secondary contact and hybridization. These two species may have been in contact since the end of the last glaciation, and the hybrid zones have been observed as geographically stable for at least 150 years

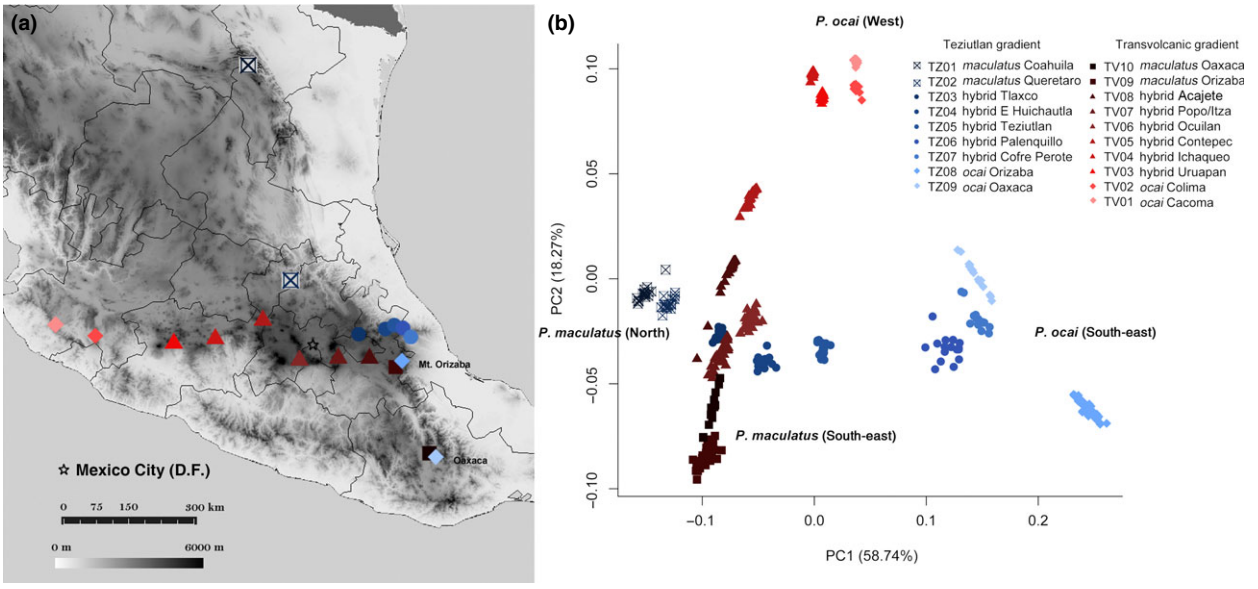

Fig. 1 (a) Map of sampling locations for the two hybrid zones in Mexico. Teziutlán gradient is coded by the blue colour ramp and Transvolcanic gradient by the red colour ramp. (b) PCA (first and second components plotted) performed on genotype probability covariance matrix for both Teziutlán and Transvolcanic gradients.

(Sibley, 1950, 1954; Sibley & West, 1958; Sibley & Sibley, 1964). Although they are congeners, *Pipilo maculatus* and *P. ocai* are not sister taxa in mitochondrial phylogenies but are each other's closest relative among the towhee species with ranges in Mexico. The pair's interspecies mtDNA sequence divergence is ~5% (Zink *et al.*, 1998; DaCosta *et al.*, 2009; Kingston *et al.*, 2014).

The Teziutlán gradient stretches along a north–south axis, transitioning from *P. maculatus* in the north through a ribbon of montane habitat to *P. ocai* in the south. The Transvolcanic gradient traverses high volcanic peaks that create a series of island-like patches of mesic highland habitat on an east–west axis; the birds here start as *P. maculatus* in the east and transition to *P. ocai* in the west. The species complex also includes a unique situation in the southern portion of the range where the two large gradients converge on Mt. Orizaba and south into Oaxaca. Here, the parental form of *P. maculatus* exists in sympatry with *P. ocai* populations, in some areas just tens of kilometres from the hybrid swarm populations (Sibley, 1954; Sibley & Sibley, 1964; Braun, 1983; Kingston *et al.*, 2012) (Table 1 and Fig. 1a, population numbers TZ08, TZ09, TV09 and TV10). The two species are highly morphologically distinct and there is no evidence of F1 hybridization in these southern, sympatric, populations (Sibley, 1954; Sibley & Sibley, 1964; Braun, 1983; Kingston *et al.*, 2014).

On a continental scale, *P. maculatus* also forms hybrid zones with *P. erythropthalmus* across the Great Plains. In such a complex, large-scale, multispecies hybrid interaction, 'pure' parental populations are more creatures of theory and less of reality, especially given the confounding factor of recent common ancestry among all three hybridizing species. Our hybrid zones, despite the complexity of shared phylogenetic history and the potential for long-distance introgression, are geographically disparate and the parental populations are divergent (Sibley & West, 1959; Kingston *et al.*, 2014). Although these natural replicates cannot be truly independent, there is still capacity for comparisons across the multiple areas of secondary contact and hybridization offered to us by the complex topography and climatic history of the towhee habitat.

This unique area of contact and hybridization offers the opportunity to harness the power of a large multilocus data set and an ancestry model-based framework to quantify patterns of differential introgression and reproductive isolation between hybridizing species, on both geographic and genomic scales. Comparison between zones allows us to observe both concordant and heterogeneous genetic architecture across multiple areas of secondary contact. We identify loci that differ significantly from the background levels of introgression across the hybrid zone and loci that are outliers in

Table 1 Sampling locations, sizes and dates across the two hybrid zones; plumage index scores range from 0 (*Pipilo ocai* plumage) to 24 (*Pipilo maculatus* plumage). To maintain consistency of plumage scores between transects (collected at different date ranges), all samples were rescored by one observer for this study. This resulted in some minor differences in plumage index scores for the Teziutlán transect from those reported by Kingston *et al.* (2012), but the patterns of plumage variation across the hybrid gradient are unchanged.

Site	Locality	Mean Plumage Index	Zone	Latitude	Longitude	*N*	Year(s)
Teziutlán gradient							
TZ01	Mesa de las Tablas, Coahuila	22.69	Teziutlán	25.25	−100.45	16	1981
TZ02	San Javier, Quéretaro	21.50	Teziutlán	20.783	−99.567	17	1981
TZ03	Tlaxco, Puebla	19.82	Teziutlán	19.667	−98.167	17	1979, 1981
TZ04	E. Huichautla, Puebla	15.75	Teziutlán	19.783	−97.6	14	1979
TZ05	Teziutlán, Puebla	14.73	Teziutlán	19.817	−97.367	15	1979
TZ06	R. Palenquillo, Veracruz	5.63	Teziutlán	19.65	−97.117	16	1979
TZ07	Cofre de Perote, Veracruz	4.77	Teziutlán	19.567	−97.1	15	1979
TZ08	Mt. Orizaba, Puebla	2.49	Teziutlán	19.05	−97.301	47	1981, 2009
TZ09	La Cumbre, Oaxaca	0.42	Teziutlán	17.165	−96.628	15	1981
Transvolcanic gradient							
TV01	Cacoma, Jalisco	1.55	Transvolcanic	19.851	−104.453	30	2008
TV02	Colima, Jalisco	1.88	Transvolcanic	19.633	−103.621	30	2008
TV03	Uruapan, Michoacán	6.45	Transvolcanic	19.491	−102.006	21	2008
TV04	Ichaqueo, Michoacán	9.23	Transvolcanic	19.578	−101.147	31	2008
TV05	Contepec, Michoacán	16.86	Transvolcanic	19.965	−100.159	29	2008
TV06	Ocuilan, Mexico	15.58	Transvolcanic	18.982	−99.377	29	2009
TV07	Popocatépetl–Iztaccíhuatl massif, Mexico	19.62	Transvolcanic	19.101	−98.594	31	2009
TV08	Acajete, Puebla	20.00	Transvolcanic	19.149	−97.926	29	2008
TV09	Mt. Orizaba, Puebla	19.48	Transvolcanic	19.075	−97.331	43	1981, 2009
TV10	La Cumbre, Oaxaca	20.33	Transvolcanic	17.167	−96.633	15	1981
	Total					460	

levels of genetic differentiation between the two species, as well as overlap between these two characterizations. We expect that variation observed among loci within each individual zone should be driven by a number of both adaptive and nonadaptive forces: both endogenous and exogenous selection, drift, dispersal and linkage disequilibrium (Buerkle *et al.*, 2011). Many of these factors may differ to some extent in the two geographically disparate zones. The two hybrid zones differ in width, and, in addition, there is geographic population differentiation among the same-species parental populations that gave rise to each zone (Kingston *et al.*, 2014). Given these differences, we expect to observe both heterogeneity between the hybrid zones and overlap in locus-specific patterns between the two zones. Concordance in underlying genetic patterns between the two parental species in different spatial settings may be consistent with a shared architecture of potentially adaptive differentiation (and isolation). The dual hybrid zone design and genomic scale of our study offer novel opportunities for insight into the underlying genetic architecture of reproductive isolation.

Materials and methods

Sampling

Population samples were collected along roughly linear transects designed to track the two major hybrid gradients (Table 1, Fig. 1a) creating two zone transects across the observed hybrid variation. The Teziutlán transect (~1150 km in length, TZ) comprised nine populations and 172 individuals, whereas the Transvolcanic transect (~760 km in length, TV) included 10 populations and 288 individuals. Birds were collected via mist net or shotgun, tissue samples (muscle, heart, liver) 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 (LSUMZ), the US National Museum of Natural History (USNM) and the Museo de Zoología (Facultad de Ciencias, UNAM; MZFC). Populations TZ08 and TV09 were sampled at Orizaba in 1981 and again 2009 (Table 1). The temporal subsamples were not genetically differentiated in a previous study (Kingston *et al.*, 2014), so were combined for the present analyses. Morphological intermediacy was assessed using a semiquantitative index composed of six plumage characters, each scored on a scale ranging from 0 to 4; the cumulative plumage index thus ranges from 0 (pure *P. ocai*) to 24 (pure *P. maculatus*) (Sibley, 1950; Braun, 1983). Our sampling covered a broad range of variation including pure parental types in each transect. DNA was extracted using a proteinase/phenol–chloroform protocol (Sambrook *et al.*, 1989) via manual methods or AutoGen extraction robot. DNA chain length was assessed via gel electrophoresis, and DNA was quantified using a NanoDrop ND-1000 spectrophotometer.

Reduced complexity library preparation and DNA sequencing

We collected genome-scale DNA sequence data from the 460 individual samples by creating 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 targeted reduction of the genome, results in SNP data from tens of thousands of loci and is one of several genotyping-by-sequencing (GBS) methods (Elshire *et al.*, 2011; Narum *et al.*, 2013). Briefly, 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 identification of sequences for each individual. We used 460 unique indexed adaptors, which allowed us to pool all individuals in one library to run across three Illumina Hiseq sequencing lanes. We amplified index-tagged fragments via PCR using two replicate reactions for each individual. Amplicons for the 460 individuals with unique indices were then pooled. We size-selected the pooled amplified library via electrophoresis on a 2% agarose gel and excision of fragments between approximately 350 and 450 bp in length. We purified these fragments using the Qiaquick Gel Extraction Kit (Qiagen Inc., Hilden, Germany). Concentration of the size-selected 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) on an Illumina HiSeq platform; 100-base pair single-end sequencing reads were generated in each of three sequencing lanes.

We removed indices and restriction sites from the *Eco*RI end of all sequences, replacing the sequence IDs with the individual IDs, and corrected indices that were off by up to two bases due to errors in sequencing or oligonucleotide synthesis. We used SeqMan NGen 3.0.4 (DNASTAR, Madison, WI, USA) to perform an initial *de novo* assembly for a subset of sequences (this subset was representative of both species and inclusive of geographic variation). After removing low-quality contigs and those with consensus sequences shorter than 82 bases in length or >88 bases in length, we created a GBS reference sequence set from the contig consensus sequences. We then aligned the full set of sequences onto the GBS reference using SeqMan 1.0.3.3 (DNASTAR). We used custom Perl scripts 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. In order for a locus to be included, 40% of individuals had to be represented by at least one read. We used a full prior for variant calling; the threshold probability for identifying a variant site was set at $P = 0.05$. We discarded loci where individuals appeared to have more than two alleles as well as any loci where the observed allele counts from putatively heterozygous individuals were very unlikely given a binomial distribution with a parameter of 0.5 (in concordance with the expectations of Mendelian inheritance (Parchman *et al.*, 2012). A final set of SNP loci was retained for analysis after removal of SNPs with minor allele frequency <0.1. The assembly and SNP calling processes were performed on a single, pooled data set for all individuals sequenced.

Population genetic analyses

We used a hierarchical Bayesian model to coincidentally estimate genotype probabilities and population allele frequencies for each of 41 324 SNPs based on the observed sequence data (as described in Gompert *et al.*, 2012b) at each locality. We treated both the individual genotype at a locus and the population allele frequency as unknown model parameters, estimating genotypes and allele frequencies separately for each of the 19 sampled populations. We used Markov chain Monte Carlo (MCMC) to sample from parameter posterior distributions. Each analysis utilized a single chain iterated for 20 000 steps where samples were recorded every fourth step. To summarize genetic variation across the sampled populations and individuals, we applied principal component analysis to the genetic covariances between all individuals (PCA, R prcomp function) (R_Core_Team, 2013). Covariances were calculated based on the point estimates of each individual genotype from the model above, which were centred on the mean genotype at a locus before calculating covariances. Additionally, we applied PCA to only those individuals that had been identified *a priori* as birds from the parental species (TZ01, TZ02, TZ08, TZ09, TV01, TV02, TV09, TV10), rather than hybrids, so as to evaluate the effect of sampling of individuals on the partitioning of variation in the PCA.

A hierarchical Bayesian F-model was used to quantify genetic differentiation (Gompert *et al.*, 2012b). We estimated posterior probabilities of F_{ST} at each individual locus (between parental populations) and a genome-wide metric of F_{ST} (between all population pairs, 171 pairwise comparisons) using 25 000 MCMC steps; every 10th step was saved and the first 1000 steps were discarded as burn-in. The F-model incorporates uncertainty in both allele frequency (uncertainty arising from finite samples from a true population) and genotype estimation (uncertainty arising from stochastic sequencing of individuals and loci in the library) simultaneously, and F_{ST} estimates integrate over this uncertainty. Quantiles of the posterior distribution of genome-wide F_{ST} allow for identification of loci with improbable, outlier estimates of F_{ST} relative to the underlying parametric distribution.

We used a Bayesian genomic cline model (Gompert & Buerkle, 2011) to assess locus-specific variation in introgression among the admixed *Pipilo* populations. This model compares each individual locus to a hybrid index derived from the composite multilocus 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 (centre parameter) and β (rate parameter). When α and β for a particular locus equal zero, the hybrid index completely predicts the probability of *P. ocai* ancestry at that locus. It is important to note that ancestry is not equivalent to allelic state. The zero value for both parameter estimates means that particular locus is aligned with the genomic background. As α shifts in one direction, the probability of *P. ocai* ancestry for the focal locus proportionally shifts away from the one-to-one line predicted by the hybrid index. A positive α parameter shifts the probability towards *P. ocai*, a negative one towards *P. maculatus*. Similarly, as β departs from zero, the rate of transition across the admixture gradient shifts away from the slope predicted by the hybrid index; a positive β parameter results in a steeper transition, and a negative parameter results in a wider transition (Gompert *et al.*, 2012b). A crucial strength of this approach is that 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 precise population-level parameter estimation even from lower sequencing coverage (Gompert *et al.*, 2012a,b).

Each hybrid zone was analysed independently using the genomic cline model. Parental populations for the Teziutlán gradient (TZ) 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 populations for the Transvolcanic (TV) gradient; the remaining populations represented the admixed portion of the gradient. Previous studies of these admixed populations with molecular and plumage markers suggested high and variable levels of hybrid ancestry (Braun, 1983; Kingston *et al.*, 2012, 2014). We estimated marginal posterior probability distributions for hybrid indexes and cline parameters α (centre 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 combined the output of the five chains after examining the MCMC output to assess chain convergence. We considered genomic cline parameter estimates as outliers in relation to the genomic background signal if the upper and lower 99% credible intervals of the parameter estimate (α or β) did not overlap zero. 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 (centre parameter) and β (rate parameter) using Pearson's product–moment correlation coefficient. We utilized the two-hybrid-zone sampling design to evaluate concordance in α (centre), β (rate) and F_{ST} outlier loci across the two hybrid zones. Previous simulations indicate that spurious correlations between cline parameters and F_{ST} loci can arise because SNPs with low F_{ST} values contain little information about local ancestry and will thus have lower point estimates for cline parameters, but risk of this phenomenon is avoided when most loci are differentiated at >0.1 F_{ST} (Gompert *et al.*, 2012a).

We tested significance of the overlap in extreme loci between transects with a null model created by randomizing outlier status for each parameter among loci (while holding number of outlier loci for each transect to the observed number). Outlier overlap between transects was calculated after each of 1000 iterations. An enrichment ratio was calculated by comparing number of observed overlapping outlier loci for each parameter to the mean number of overlapping loci generated by the null. *P*-values were assessed by tallying the frequency of iterations where number of overlapping loci in the null exceeded the observed level.

Results

Sequencing, assembly and SNP calling

After parsing indices and filtering reads, 282 895 590 reads were retained for analysis. The *de novo* assembly of 40 million sequences resulted in a partial GBS reference genome (408 347 contigs), which was used for a reference-based assembly of all reads from each individual. The full reference-based assembly placed 180 627 551 reads uniquely, for an average coverage depth of 1220× per sequenced region (1.4× per individual per sequenced region). A set of 41 324 polymorphic loci (SNPs) met our criteria (see Methods) and was used for most analyses.

Population genetic analyses

The PCA of genetic covariances results in a first principal axis that differentiates the parental species, as expected (PC1 explains 58.74% of variation among all individuals, 68.7% of variation for only *a priori* parental individuals; Fig. 1b, see also Fig. S1). Secondarily, the PCA provides evidence for substantial genetic differentiation between western and south-eastern forms of *P. ocai*, with individuals from the western localities intermediate on PC1 between southern *P. ocai* and all parental *P. maculatus*, while differentiated from both on PC2 (Fig. 1b). Furthermore, individuals from localities on each of the geographic transects of the hybrid zones are genetically intermediate between the putative parental species. In addition, for each transect, putative hybrids are genetically intermediate between *P. maculatus* and the specific *P. ocai* localities in that transect, so that the first two axes of PCA are consistent with the geography of sampling locations. PC 1 largely separates the populations of the Teziutlán transect (connecting *maculatus* in the north with *ocai* in the south-east), whereas populations of the Transvolcanic gradient are distributed along PC 2 (connecting *ocai* in the west with *maculatus* in the south-east). Individuals from the same locality are more genetically similar to one another than individuals from other localities. The tightness of the clustering is a combination of information from individual genotypes and from using the estimated population allele frequency as a prior for loci and individuals with relatively low coverage.

The pairwise genome-wide population F_{ST} estimates support differentiation between the species, with F_{ST} = 0.139 (0.137–0.141, 95% credible interval) between the species along the Teziutlán transect (TZ01/TZ02 vs. TZ08/TZ09) and F_{ST} = 0.101 (0.099–0.103) between the species along the Transvolcanic gradient (TV01/TV02 vs. TV09/TV10). Additionally, pairwise F_{ST} estimates are relatively large between geographically separated populations within each of the parental species. The eastern and western *P. ocai* types are slightly more differentiated (average pairwise population between east and west F_{ST} = 0.109) than the northern and southern *P. maculatus* (average pairwise population F_{ST} between north and south = 0.084).

Hybrid individuals in each zone occupied different ranges of admixture (TZ 0.30–0.88, TV 0.09–0.67), and the distribution of hybrid index among individuals in both zones was somewhat bimodal (Fig. 2). The transition in ancestry between the parental species was strongly associated with sampling locations, with nearly stepwise transitions in mean admixture proportion across populations in both transects (Fig. 2).

The exceptional genomic cline parameter estimates for the Teziutlán gradient include 4538 total α and β outlier loci (Fig. 3, Table 2). The number of F_{ST} outlier loci (beyond the 95% quantile of the parametric distribution) across the Teziutlán gradient is 356. Loci with extreme genomic clines that also exhibit outlier F_{ST} estimates number 269 loci. F_{ST} estimates are correlated with genomic cline parameter estimates (α and F_{ST}: $r_{41,322}$ = 0.190, P < 2.2e-16; β and F_{ST}: $r_{41,322}$ = 0.190,

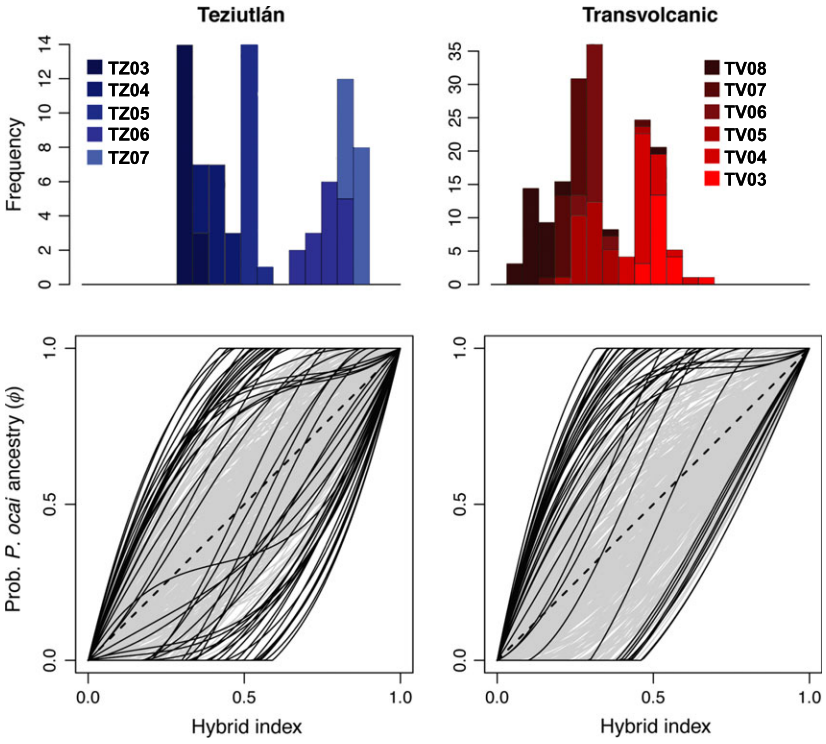

Fig. 2 Frequency distributions of multilocus hybrid indices across the admixed populations from both hybrid zones and cline plots showing the relationship between the hybrid index and the probability of *Pipilo ocai* ancestry (ϕ) as predicted by each single locus. The hashed line demonstrates where the 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 black 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 zone appear at the top of the graph.

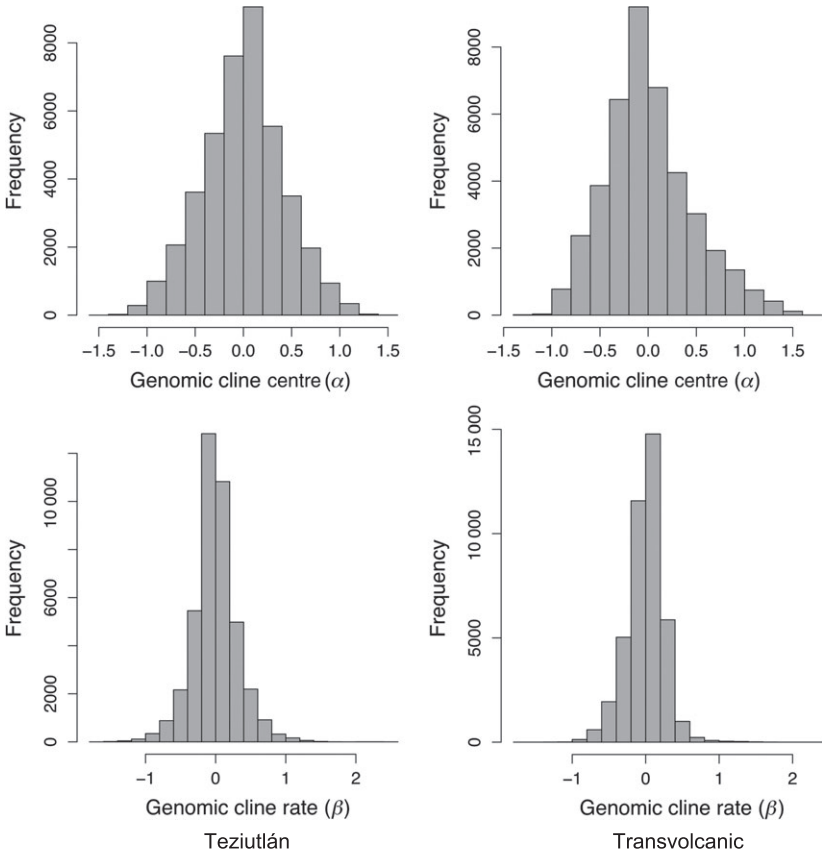

Fig. 3 Frequency distributions for genomic cline parameters β and α across both hybrid zones.

Table 2 Number of loci with extreme parameter values and outlier F_{ST} values for each zone and number of loci exhibiting extreme values in both hybrid zones.

	Hybrid zone				
Type of outlier	Teziutlán	Transvolcanic	Both zones	Enrichment ratio	*P*-value
α outlier	4251	3146	332	0.99	0.58
β outlier	287	80	17	30.25	<0.001
F_{ST} outlier	356	440	152	39.81	<0.001
α outlier + F_{ST} outlier	170	185	33	42.64	<0.001
β outlier + F_{ST} outlier	99	61	26	152.94	<0.001

$P < 2.2e\text{-}16$), a pattern consistent with adaptation and reproductive isolation sharing some genetic basis (although overall allele frequency differential between the parental species can also influence the observed correlation (Gompert *et al.*, 2012a)).

The outlier genomic cline parameter estimates for the Transvolcanic gradient reveal 3226 total α and β outlier loci (Fig. 3, Table 2). There are 440 F_{ST} outlier loci (outside 95% quantile) in the Transvolcanic gradient. There are 246 total loci with extreme genomic cline parameters that also have outlier F_{ST} estimates (Table 2). Again, F_{ST} and cline parameters in the Transvolcanic are significantly correlated (α and F_{ST}: $r_{41,322} = 0.291$ $P < 2.2e\text{-}16$, β and F_{ST}: $r_{41,322} = 0.368$ $P < 2.2e\text{-}16$), revealing a core genetic architecture. The majority of F_{ST} values analysed across both zones are greater than the threshold where spurious correlations are likely occur (i.e. >0.08–0.10) (Gompert *et al.*, 2012a).

Hybrid zone comparison

Comparing loci across both hybrid zones offers the opportunity to identify loci exhibiting consistent signal despite demographic and stochastic differences between the zones. Some loci overlap as outliers in the genomic cline analyses across both zones: 332 loci with exceptional α estimates in both zones, 17 loci with exceptional β estimates in both zones and 152 loci with exceptional F_{ST} estimates in both zones. All but the overlapping extreme α loci significantly exceed expectations of overlap by chance (Table 2). The positive tail of the α parameter distribution in the Transvolcanic is longer than that observed in the Teziutlán gradient – in the Transvolcanic, a greater proportion of loci deviate from the hybrid index background signal towards *P. ocai* ancestry (Fig. 3). The Transvolcanic gradient also exhibits a distribution of β parameters that are compressed around zero, suggesting less overall variation in rate in comparison with the Teziutlán gradient. Both cline parameter and differentiation estimates are significantly correlated across both zones, although the magnitude of correlation among cline parameters is small: α $r_{41,322} = 0.065$ $P < 2.2e\text{-}16$; β $r_{41,322} = 0.079$ $P < 2.2e\text{-}16$; F_{ST} $r_{41,322} = 0.233$ $P < 2.2e\text{-}16$. The overlapping extreme loci are a subset of the extreme loci in each zone, as reflected by the small magnitude of the correlations. Despite the differences in a variety of factors between the two hybrid zones, the overlapping extreme loci are consistent with shared genetic architecture underlying isolation and adaptation. Alternatively, the discordance in outlier loci between the zones points to portions of genetic architecture that differ between these two zones and a role for stochastic processes.

Discussion

Our dual hybrid zone approach allows us to detect shared patterns in introgression between multiple instances of secondary contact despite pronounced differences in geographic and population demographic contexts. Our study utilizes the strength of model-based inference to ascertain genome-scale patterns of divergence and introgression across hybrid zones, as well as variance and overlap in these patterns, all in a non-model system. The genomic scale of our data and the multiple hybrid zones allow a fine-scale comparison of loci that deviate from background genomic patterns.

This particular set of towhee hybrid zones is rich in historical data in the literature (Sibley, 1950, 1954; Sibley & West, 1958; Sibley & Sibley, 1964; Braun, 1983). The secondary contact between species is likely influenced heavily by climatic cycles, and the two species have likely been subject to repeated separation and reconnection since the lineages initially diverged (Sibley, 1950, 1954). Extant habitat for both species in the contact areas exists in montane islands (TV) or ribbons (TZ) of high-elevation pine/oak forest surrounded by a matrix of dry plateau, a pattern common to mosaic hybrid zones (Harrison, 1993). These habitat-mediated contact areas are important to both differentiation and introgression between towhee species (Kingston *et al.*, 2014).

Variation between the hybrid zones' environmental and geographic compositions could contribute to discordance between the zones via differential selection pressure as well as differential rates of dispersal and gene flow. Similar patterns are observed in other avian hybrid zones: environmental heterogeneity is an influential factor for patterns of introgression in a secondary contact hybrid zone between Lazuli (*Passerina cyanea*, Linnaeus, 1766) and Indigo (*P. amoena*, Say, 1823) buntings (Carling & Thomassen, 2012). In a moving chickadee hybrid zone, genetic introgression of loci (at least those likely under selection or involved in isolation) is limited by shifting climate factors (Taylor *et al.*, 2014). The observed heterogeneity between the towhee

zones suggests there is a significant influence of geographically specific environmental factors, as has been demonstrated across taxa in other hybrid zones (Teeter *et al.*, 2008, 2010; Nolte *et al.*, 2009; Carson *et al.*, 2012). The importance of habitat corridors in relation to gene flow in this system has been previously documented (Kingston *et al.*, 2014), and the genomic signal observed here is consistent with that result. Environmental heterogeneity – both across patchy available habitat and through time as climate and secondary contact fluctuated – has left a signature on the genetic architecture of the towhee hybrid zones.

Differentiation patterns among our two hybrid zones reflect a general stepping stone-like pattern across the geographic orientation of the populations (Fig. 1). The phenomenon of genetic signal reflecting geography in multilocus data is well known (Novembre *et al.*, 2008). Although the Oaxaca populations are most geographically isolated from the remainder of the transect, the Oaxaca populations of both *P. maculatus* and *P. ocai* parental populations appear more similar to their respective hybrid zones than either Orizaba population (Fig. 1b). This two-dimensional pattern may be misleading, as at least the *P. maculatus* Oaxaca population is differentiated further from the hybrid zones along the PC3 axis (Fig. S2). The first three PC axes account for 81.22% of the observed variation, indicating substantial genetic differentiation among the parental species and a strong geographic effect on population structure (Novembre *et al.*, 2008; Parchman *et al.*, 2012, 2013). Although the geographic signal is prevalent, remaining variation could represent other stochastic, demographic and selective forces playing an important role in structuring genomic variation among these populations.

A significant concordance exists between the two hybrid zones in a core set of outlier F_{ST} and cline parameter loci. The overlap between hybrid zones of outlier F_{ST} loci is higher than the overlap of loci with extreme cline parameters (Table 2). We observe high levels of population differentiation and different distributions of hybrid composition across the two zones, yet still observe a significant concordance among outlier loci. Despite the stochastic and environmental factors that can influence reproductive isolation upon secondary contact, the set of concordant loci between zones points to a shared genetic pattern. The concordance in introgression and genetic differentiation across multiple hybrid zones is consistent with a set of loci that may tag genetic regions important to isolation and adaptation, and also suggests that these loci likely tag genetic regions that play a more general, rather than zone-specific, role in isolation. This level of concordance is notable given that the divergence between the geographically separated *P. ocai* parental populations is nearly equal to the divergence between the species.

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 (Dobzhansky, 1940; Howard, 1993). During the speciation process, reproductive isolation has been shown to arise as a product of divergent selection in many systems (Dodd, 1989; Danley *et al.*, 2000; Bolnick *et al.*, 2006; Funk *et al.*, 2006; Nosil & Feder, 2012). However, high rates of hybridization and relatively high fitness of hybrid individuals are not always in conflict with maintenance of species identities and continued divergence of lineages (Lindtke *et al.*, 2014). In addition, there can be disagreement between ecological forces and reproductive isolation; Dobzhansky–Muller incompatibilities can arise through stochastic processes, 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 often be underestimated (Buerkle *et al.*, 2011).

It is through comparing the overlapping extreme loci in the two hybrid zones that we can begin to tease out which loci may be associated with global isolation and which loci differ from the genomic background due to zone-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, the concordant set of outlier loci could represent not only 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). In contrast, the discordant set of outlier loci is consistent with environment-specific exogenous selection or even drift within zones. This dichotomous signal between subsets of outlier loci indicates the legacy of adaptation and divergence in allopatry is influencing maintenance of isolation after secondary contact similarly in both hybrid zones, but also suggests local environmental and stochastic factors contribute to heterogeneity between the hybrid zones. This type of heterogeneity is also observed across multiple crow hybrid zones (Vijay *et al.*, 2016).

Our two zones also differ in the proportion of ancestry in hybrids across the gradient (Fig. 2). The observed bimodality in hybrid indices is unlikely driven by sampling scheme, given the close geographic proximity of Teziutlán gradient admixed populations and the diversity of hybrid indices found in the TV08 population in the Transvolcanic. The observed bimodality likely reflects dynamics of stable, old hybrid populations structured by stepping stone-like dispersal of parental genes into each hybrid gradient. Both variation in hybrid ancestry and population differentiation within

each parental species across the two zones may influence patterns of introgression.

The Transvolcanic gradient appears to exhibit a slight asymmetry in introgression patterns (long positive tail on the α (centre) distribution) as well as a lower variance in rate of change (compressed β (rate) distribution, Fig. 3). The variance estimates of the parameter distributions reflect the differences: variance Transvolcanic α (0.165) >Teziutlán α (0.136), and variance Transvolcanic β (0.023) <Teziutlán β (0.074). These differences in introgression patterns are likely tied to the different hybrid index distributions – the Teziutlán indices are shifted towards *ocai* whereas the Transvolcanic indices are shifted towards *maculatus* (Fig. 2). The heterogeneity of the hybrids between the zones could contribute to the asymmetry of introgressing *ocai* alleles in the more *maculatus*-like Transvolcanic zone (Fig. 3). Although it is common to associate asymmetry in cline parameters with geographic movement in geographic clines, our measures are assessing locus-specific deviation from a genomic background and the parameters are not inherently geographic in nature. However, it is possible that the zone-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-zone differences point to the complexity of how ecological and evolutionary contexts structure introgression and the maintenance of isolation, and that generalizations about admixture and the architecture of isolation across different hybrid zones should be made with caution (Cutter, 2015; Mandeville *et al.*, 2015).

Introgression from periodic pulses of gene flow resulting from cycles of secondary contact and hybridization may heavily influence isolation and introgression and its resulting genomic architecture (Rheindt & Edwards, 2011). Loci with extreme differentiation values (F_{ST}) between parental populations may reside in genomic regions influenced by divergent selection (Beaumont & Nichols, 1996; Beaumont & Balding, 2004); it has also been demonstrated that extreme F_{ST} values can arise from low intrapopulation diversity and may mark loci involved in local adaptation instead (Cruickshank & Hahn, 2014). Our analyses reveal that many loci with extreme F_{ST} estimates are also characterized by extreme genomic cline parameters estimates (α and β). The loci with dual extreme values (Table 2) likely track portions of the genome experiencing selection although we cannot discern via this association whether the selective force arose pre- or post-speciation.

Conclusions

Our genome-scale analyses of two towhee hybrid zones suggest that whereas divergent selection likely contributes to isolation, the demographic and environmental context of fitness upon secondary contact and hybridization contributes to variation in patterns of introgression. Genome-scale data for other avian species have also elucidated the importance of selectively divergent genomic regions in maintenance of isolation upon secondary contact, as well as the contributions of complex demographic histories during speciation (Nadachowska-Brzyska *et al.*, 2013; Poelstra *et al.*, 2014). Our comparison of two geographically disparate hybrid zones reveals a core of concordant outlier loci likely associated with historical divergence and maintenance of isolation. The remaining extreme loci that vary between hybrid zones reveal heterogeneity in differential introgression between the two geographic and demographic settings. Thus, locally specific habitat and environmental factors, as well as neutral processes, are also likely influential in the dynamics of this system. A future goal will be to more definitively link genomic patterns of variation with the evolutionary processes of drift, recombination, adaptive divergence of populations and selection on phenotypes that gives rise to reproductive isolation.

Acknowledgments

We appreciate greatly Adolfo G. Navarro S. and Blanca Hernández for extending us collecting permits in Mexico. We are also extremely grateful to Adolfo G. Navarro S. for making our 2008–2009 fieldwork in Mexico possible as well as for many fruitful discussions of avian phylogeography and natural history. We thank Laura Villaseñor, Sarahy Contreras, Erick A. García Trejo, Hernán Vázquez, Alejandro Gordillo, Gala Cortés, Daniela Tovilla-Sierra, Rubén Ortega, Marco Ortiz, Mauricio Pérez-Vera, Laila Yunes-Jiménez and César A. Ríos-Muñoz for their invaluable guidance, leadership and assistance in the field across the Transvolcanic Belt in Mexico. We thank David Braun, Stephen Braun, Nicholas Lanson, Nimrod Funk and George Oliver for their able assistance in the field in the Teziutlán gradient. We thank our funding sources for this project: Smithsonian Research Endowment Fund, University of Maryland Program in Behavior, Ecology, Evolution and Systematics, Smithsonian Ornithology, the Department of Vertebrate Zoology, National Museum of Natural History (NMNH) and Don Wilson. The work was made possible by research assistantships from the NMNH Frontiers in Phylogenetics Program and NSF grants DEB0228675 and DEB0733029.

Data Sharing and Archival

The data obtained from this study are archived and accessible in DRYAD (doi:10.5061/dryad.3vd6k) as final assembly file (.bam), a genotype file with SNP genotype probabilities for each locus and individual, and a plumage index and metadata file with scores across the six

phenotypic characters for each individual and associated metadata.

References

Barton, N.H. 2001. The role of hybridization in evolution. *Mol. Ecol.* **10**: 551–568.

Barton, N.H. & Bengtsson, B.O. 1986. The barrier to genetic exchange between hybridizing populations. *Heredity* **57**: 357–376.

Barton, N.H. & Hewitt, G.M. 1985. Analysis of hybrid zones. *Annu. Rev. Ecol. Syst.* **16**: 113–148.

Barton, N.H. & Hewitt, G.M. 1989. Adaptation, speciation and hybrid zones. *Nature* **341**: 497–503.

Beaumont, M.A. & Balding, D.J. 2004. Identifying adaptive genetic divergence among populations from genome scans. *Mol. Ecol.* **13**: 969–980.

Beaumont, M.A. & Nichols, R.A. 1996. Evaluating loci for use in the genetic analysis of population structure. *Proc. R. Soc. Biol. Sci. B* **263**: 1619–1626.

Bolnick, D.I., Near, T.J. & Wainwright, P.C. 2006. Body size divergence promotes post-zygotic reproductive isolation in centrarchids. *Evol. Ecol. Res.* **8**: 903–913.

Braun, M.J. 1983. *Molecular Versus Morphological and Behavioral Differentiation Across Contact Zones Between Closely Related Avian Species*. Department of Biochemistry, Louisiana State University Medical Center, New Orleans, LA. PhD: 190.

Buerkle, C.A., Gompert, Z. & Parchman, T.L. 2011. The n = 1 constraint in population genomics. *Mol. Ecol.* **20**: 1575–1581.

Carling, M.D. & Thomassen, H.A. 2012. The role of environmental heterogeneity in maintaining reproductive isolation between hybridizing Passerina (Aves: Cardinalidae) buntings. *Int. J. Ecol.* **2012**: 1–11.

Carson, E.W., Tobler, M., Minckley, W.L., Ainsworth, R.J. & Dowling, T.E. 2012. Relationships between spatio-temporal environmental and genetic variation reveal an important influence of exogenous selection in a pupfish hybrid zone. *Mol. Ecol.* **21**: 1209–1222.

Cruickshank, T.E. & Hahn, M.W. 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Mol. Ecol.* **23**: 3133–3157.

Cutter, A.D. 2015. Repeatability, ephemerality and inconvenient truths in the speciation process. *Mol. Ecol.* **24**: 1643–1644.

DaCosta, J.M., Spellman, G.M., Escalante, P. & Klicka, J. 2009. A molecular systematic revision of two historically problematic songbird clades: *Aimophila* and *Pipilo*. *J. Avian Biol.* **40**: 206–216.

Danley, P.D., Markert, J.A., Arnegard, M.E. & Kocher, T.D. 2000. Divergence with gene flow in the rock-dwelling cichlids of Lake Malawi. *Evolution* **54**: 1725–1737.

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

Dodd, D.M.B. 1989. Reproductive isolation as a consequence of adaptive divergence in *Drosophila pseudoobscura*. *Evolution* **43**: 1308–1311.

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S. *et al.* 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* **6**: PMC3087801.

Funk, D.J., Nosil, P. & Etges, W.J. 2006. Ecological divergence exhibits consistently positive associations with reproductive isolation across disparate taxa. *Proc. Natl. Acad. Sci. USA* **103**: 3209–3213.

Gavrilets, S. 2003. Perspective: models of speciation: what have we learned in 40 years? *Evolution* **57**: 2197–2215.

Gompert, Z. & Buerkle, C.A. 2011. Bayesian estimation of genomic clines. *Mol. Ecol.* **20**: 2111–2127.

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

Gompert, Z., Parchman, T.L. & Buerkle, C.A. 2012b. Genomics of isolation in hybrids. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **367**: 439–450.

Harrison, R.G. 1993. *Hybrid Zones and the Evolutionary Process*. Oxford University Press, New York, NY.

Hohenlohe, P.A., Bassham, S., Etter, P.D., Stiffler, N., Johnson, E.A. & Cresko, W.A. 2010. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genet.* **6**: e1000862.

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

Janoušek, V., Wang, L., Luzynski, K., Dufková, P., Vyskocilová, M.M., Nachman, M.W. *et al.* 2012. Genome-wide architecture of reproductive isolation in a naturally occurring hybrid zone between *Mus musculus musculus* and *M. m. domesticus*. *Mol. Ecol.* **21**: 3032–3047.

Janoušek, V., Munclinger, P., Wang, L., Teeter, K.C. & Tucker, P.K. 2015. Functional organization of the genome may shape the species boundary in the house mouse. *Mol. Biol. Evol.* **32**: 1208–1220.

Kingston, S.E., Jernigan, R.W., Fagan, W.F., Braun, D. & Braun, M.J. 2012. Genomic variation in cline shape across a hybrid zone. *Ecol. Evol.* **2**: 2737–2748.

Kingston, S.E., Navarro-Sigüenza, A.G., García-Trejo, E.A., Vázquez-Miranda, H., Fagan, W.F. & Braun, M.J. 2014. Genetic differentiation and habitat connectivity across towhee hybrid zones in Mexico. *Evol. Ecol.* **28**: 277–297.

Larson, E.L., Andres, J.A., Bogdanowicz, S.M. & Harrison, R.G. 2013. Differential introgression in a mosaic hybrid zone reveals candidate barrier genes. *Evolution* **67**: 3653–3661.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N. *et al.* 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**: 2078–2079.

Lindtke, D. & Buerkle, C.A. 2015. The genetic architecture of hybrid incompatibilities and their effect on barriers to introgression in secondary contact. *Evolution* **69**: 1987–2004.

Lindtke, D., Gompert, Z., Lexer, C. & Buerkle, C.A. 2014. Unexpected ancestry of Populus seedlings from a hybrid zone implies a large role for postzygotic selection in the maintenance of species. *Mol. Ecol.* **23**: 4316–4330.

Mandeville, E.G., Parchman, T.L., McDonald, D.B. & Buerkle, C.A. 2015. Highly variable reproductive isolation among pairs of *Catostomus* species. *Mol. Ecol.* **24**: 1856–1872.

Nadachowska-Brzyska, K., Burri, R., Olason, P.I., Kawakami, T., Smeds, L. & Ellegren, H. 2013. Demographic divergence history of pied flycatcher and collared flycatcher inferred from whole-genome re-sequencing data. *PLoS Genet.* **9**: e1003942.

Narum, S.R., Buerkle, C.A., Davey, J.W., Miller, M.R. & Hohenlohe, P.A. 2013. Genotyping-by-sequencing in ecological and conservation genomics. *Mol. Ecol.* **22**: 2841–2847.

Nolte, A.W., Gompert, Z. & Buerkle, C.A. 2009. Variable patterns of introgression in two sculpin hybrid zones suggest that genomic isolation differs among populations. *Mol. Ecol.* **18**: 2615–2627.

Nosil, P. & Feder, J.L. 2012. Genomic divergence during speciation: causes and consequences. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **367**: 332–342.

Novembre, J., Johnson, T., Bryc, K., Kutalik, Z.N., Boyko, A.R., Auton, A. *et al.* 2008. Genes mirror geography within Europe. *Nature* **456**: 98–103.

Orr, H.A. & Turelli, M. 2001. The evolution of postzygotic isolation: accumulating Dobzhansky-Muller incompatibilities. *Evolution* **55**: 1085–1094.

Parchman, T.L., Gompert, Z., Mudge, J., Schilkey, F.D., Benkman, C.W. & Buerkle, C.A. 2012. Genome-wide association mapping of an adaptive trait in lodgepole pine. *Mol. Ecol.* **21**: 2991–3005.

Parchman, T.L., Gompert, Z., Braun, M.J., Brumfield, R., McDonald, D.B., Uy, J.A.C. *et al.* 2013. The genomic consequences of adaptive divergence and reproductive isolation between species of manakins. *Mol. Ecol.* **22**: 3304–3317.

Payseur, B.A. 2010. Using differential introgression in hybrid zones to identify genomic regions involved in speciation. *Mol. Ecol. Resour.* **10**: 806–820.

Poelstra, J.W., Vijay, N., Bossu, C.M., Lantz, H., Ryll, B., Müller, I. *et al.* 2014. The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. *Science* **344**: 1410–1414.

R_Core_Team 2013. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rheindt, F.E. & Edwards, S.V. 2011. Genetic introgression: an integral but neglected component of speciation in birds. *Auk* **128**: 620–632.

Sambrook, J., Fritsch, E.F. & Maniatis, T. 1989. *Molecular Cloning: A Laboratory Manual*, 2 edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sattler, G.D. & Braun, M.J. 2000. Morphometric variation as an indicator of genetic interactions between black-capped and Carolina chickadees at a contact zone in the Appalachian mountains. *Auk* **117**: 427–444.

Sibley, C.G. 1950. Species formation in the red-eyed towhees of Mexico. *Univ. Calif. Publ. Zool.* **50**: 109–194.

Sibley, C.G. 1954. Hybridization in the red-eyed towhees of Mexico. *Evolution* **8**: 252–290.

Sibley, C.G. & Sibley, F.C. 1964. Hybridization in the red-eyed towhees of Mexico: the populations of the southeastern plateau region. *Auk* **81**: 479–504.

Sibley, C.G. & West, D.A. 1958. Hybridization in the red-eyed towhees of Mexico: the eastern plateau populations. *Condor* **60**: 85–104.

Sibley, C.G. & West, D.A. 1959. Hybridization in the Rufous-sided towhees of the great plains. *Auk* **76**: 326–338.

Taylor, S.A., White, T.A., Hochachka, W.M., Ferretti, V., Curry, R.L. & Lovette, I. 2014. Climate-mediated movement of an avian hybrid zone. *Curr. Biol.* **24**: 671–676.

Teeter, K.C., Payseur, B.A., Harris, L.W., Bakewell, M.A., Thibodeau, L.M., O'Brien, J.E. *et al.* 2008. Genome-wide patterns of gene flow across a house mouse hybrid zone. *Genome Res.* **18**: 67–76.

Teeter, K.C., Thibodeau, L.M., Gompert, Z., Buerkle, C.A., Nachman, M.W. & Tucker, P.K. 2010. The variable genomic architecture of isolation between hybridizing species of house mice. *Evolution* **64**: 472–485.

Vijay, N., Bossu, C.M., Poelstra, J.W., Weissensteiner, M.H., Suh, A., Kryukov, A.P. *et al.* 2016. Evolution of heterogeneous genome differentiation across multiple contact zones in a crow species complex. *Nat. Commun.* **7**: 13195.

Yuri, T., Jernigan, R.W., Brumfield, R.T., Bhagabati, N.K. & Braun, M.J. 2009. The effect of marker choice on estimated levels of introgression across an avian (Pipridae: *Manacus*) hybrid zone. *Mol. Ecol.* **18**: 4888–4903.

Zink, R.M., Weller, S.J. & Blackwell, R.C. 1998. Molecular phylogenetics of the avian genus *Pipilo* and a biogeography argument for taxonomic uncertainty. *Mol. Phylogenet. Evol.* **10**: 191–201.

Supporting information

Additional Supporting Information may be found online in the supporting information tab for this article:

Figure S1 PCA performed on the genotype probability covariance matrix for *a priori* parental species in both gradients (first three PCs plotted).

Figure S2 PCA (third PC plotted against both first and second) performed on the genotype probability covariance matrix for both Teziutlán and Transvolcanic gradients.

Figure S3 *Pipilo* specimens showing variations in plumage traits from *ocai*-like on the left side to *maculatus*-like on the right side.

Data deposited at Dryad: doi: 10.5061/dryad.3vd6k

Received 28 August 2016; revised 13 November 2016; accepted 19 December 2016