doi:10.1111/evo.13278

History cleans up messes: The impact of time in driving divergence and introgression in a tropical suture zone

Sonal Singhal[1,2,3] and Ke Bi[4,5]

[1]*Department of Ecology and Evolutionary Biology, University of Michigan, 830 North University, Ann Arbor, Michigan 48109*

[2]*Museum of Zoology, University of Michigan, 1109 Geddes Avenue, Ann Arbor, Michigan 48109*

[3]*E-mail: sonal.singhal1@gmail.edu*

[4]*Museum of Vertebrate Zoology, University of California, Berkeley, 3101 Valley Life Sciences Building, Berkeley, California 94720*

[5]*Computational Genomics Resource Laboratory, California Institute for Quantitative Biosciences, University of California, Berkeley, California 94720*

Received January 17, 2017
Accepted May 4, 2017

Contact zones provide an excellent arena in which to address questions about how genomic divergence evolves during lineage divergence. They allow us to both infer patterns of genomic divergence in allopatric populations isolated from introgression and to characterize patterns of introgression after lineages meet. Thusly motivated, we analyze genome-wide introgression data from four contact zones in three genera of lizards endemic to the Australian Wet Tropics. These contact zones all formed between morphologically cryptic lineage-pairs within morphologically defined species, and the lineage-pairs meeting in the contact zones diverged anywhere from 3.1 to 5.8 million years ago. By characterizing patterns of molecular divergence across an average of 11K genes and fitting geographic clines to an average of 7.5K variants, we characterize how patterns of genomic differentiation and introgression change through time. Across this range of divergences, we find that genome-wide differentiation increases but becomes no less heterogeneous. In contrast, we find that introgression heterogeneity decreases dramatically, suggesting that time helps isolated genomes "congeal." Thus, this work emphasizes the pivotal role that history plays in driving lineage divergence.

KEY WORDS: **Cryptic species, exome capture, hybridization, introgression, molecular evolution, speciation, skinks.**

As lineages diverge, mutation, recombination, selection, drift, and gene flow interact to shape patterns of genomic divergence. Understanding this rich interplay of processes can be explored through two primary, nonmutually exclusive approaches: we can (1) investigate patterns of genomic variation of lineages falling along a range of divergences and (2) determine what happens when differentiated lineages and genomes interact. In this study, we integrate both approaches to understand the genomic consequences of hybridization between lineages meeting in secondary contact. Theory predicts this interaction will be structured by the balance between selection and recombination (Barton 1983). During hybridization, two differentiated genomes meet, and the resulting admixed genome is subject to selection, whether due to environmentally dependent selection against resulting hybrid phenotypes (Schluter 2001) or intrinsic selection due to genetic incompatibilities (Dobzhansky 1934; Muller 1942). When selection against hybrids is strong, disequilibrium remains high across the genome, and loci cannot introgress even if they have no effect on hybrid fitness (Barton 1983). However, when selection on hybrids is weak, recombination dissociates disequilibrium between loci, thus allowing loci to introgress at a rate and extent that reflect the selective effects of that locus and closely linked loci (Baird 1995). Here, the hybrid zone functions as a sieve (Martinsen *et al*. 2001), reflecting the emerging consensus that most barriers are

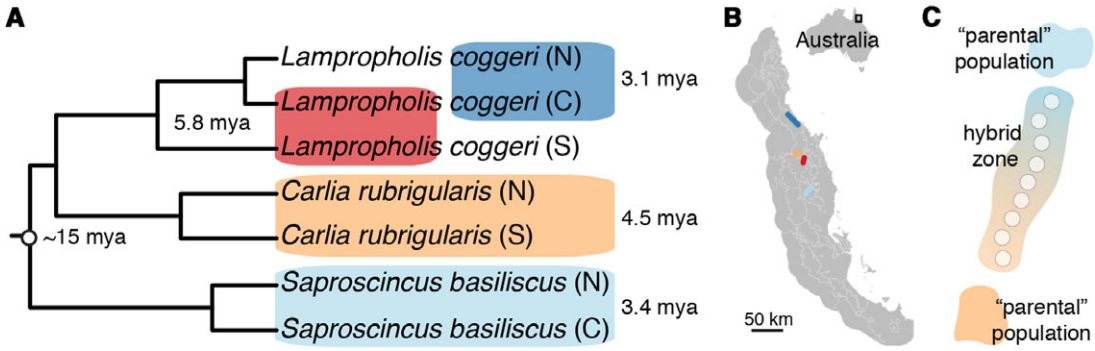

Figure 1. **(A) Phylogeny of lineages used in this study; boxes indicate lineage-pairs meeting in hybrid zones. Boxes are labeled with divergence time estimates for the lineage-pair (Singhal and Moritz 2013). (B) Map of the Australian Wet Tropics; contact zone locations labeled following colors in A. (C) Basic sampling approach used in study. Transcriptome data from two allopatric "parental" populations were used to infer patterns of genomic divergence; pooled target capture data from nine hybrid zone populations were used to infer patterns of introgression.**

semi-permeable (Key 1968; Bazykin 1969; Harrison and Larson 2016). As such, hybrid zones both represent the build-up of reproductive isolation between lineages and the potential for heterogeneous break-down of differentiation across the genome, and dynamics at a hybrid zone are likely to change as lineage-pairs become more phenotypically and genetically differentiated.

We apply this dual perspective on hybrid zones to a natural laboratory for comparative analyses of speciation: the suture zone of the Australian Wet Tropics (AWT). This suture zone, a geographically restricted area consisting of multiple overlapping contact zones (Remington 1968), occurs in a narrow strip of rainforest in northeastern Australia (Moritz *et al*. 2009). The repeated glacial cycles of the Pliocene and Pleistocene led to changing rainforest distributions, and concordantly, populations of over 20 rainforest species became isolated in refugia and diverged. Since the Last Glacial Maximum, these populations have expanded out of the refugia and are now meeting in secondary contacts that comprise the suture zone (Moritz *et al*. 2009). Almost all the lineage-pairs are morphologically cryptic; however, the lineages span a wide range of genetic divergences allowing us to address questions about speciation along a continuum. In this work, we focus on four contact zones that have formed between phylogeographic lineages in a clade of ecologically similar skinks that diverged about 15 mya. These lineage-pairs are *Lampropholis coggeri* N/C, *Saproscincus basiliscus* N/C, *Carlia rubrigularis* N/S, and *L. coggeri* C/S, where N, C, and S designate the Northern, Central, and Southern locations of these lineages in the AWT. Coalescent modeling based on genomic data showed that these lineage-pairs have divergence times that span from 3.1 million years ago (mya) to 5.8 mya (Fig. 1A) and all meet in hybrid zones (Fig. 1B) (Singhal and Moritz 2013).

For each lineage-pair, we use transcriptome data to estimate patterns of divergence across the genome and exome capture data to infer geographic clines, and thus introgression, across the contact zone. In doing so, we address a number of questions. First, as lineage-pairs become older, how do patterns of genomic divergence change? Both verbal models and empirical data predict that, particularly in relatively young lineages that diverged with gene flow, there should be considerable heterogeneity in divergence across the genome (Wu 2001). As lineages age, divergence across the genome should become more homogeneous as mutations fix due to drift and selection (Roux *et al*. 2016). Although these lineages likely diverged without gene flow (Singhal and Moritz 2013), we predict a similar outcome, given heterogeneity in selection strength and recombination rate across the genome. Second, how does increased genome-wide differentiation affect introgression patterns? We predict that in more divergent lineage-pairs, selection against hybrids will be greater, and thus, the extent of introgression, both in terms of proportion of genome and spatial range, will be more limited than in less divergent lineage-pairs (Barton 1983; Barton and Gale 1993). Further, in more divergent versus less divergent lineage-pairs, linkage disequilibrium will be more extensive because recombination will be less effective (Barton and Bengtsson 1986; Pool and Nielsen 2009). Importantly, in making these predictions, we assume that these contact zones formed concurrently and that recombination rates are similar across the lineages, both reasonable assumptions given the lineages' shared history and similar biologies (Moritz *et al*. 2009; Williams *et al*. 2010). Third, how predictable are introgression patterns across loci in different lineage-pairs? Because introgression is a function of selection and recombination (Fisher 1950; Endler 1977; Slatkin 1973), we should be able to use proxies of these processes (here, the molecular evolution at a given locus) to predict introgression. Further, if selection and recombination patterns across the genome are correlated across lineages (Janousek *et al*. 2012; Singhal *et al*. 2015; Van Doren *et al*. 2017; Vijay *et al*.

2017), we might predict that introgression patterns across contact zones should be correlated. Here, we test these predictions by comparing patterns of introgression to patterns of genomic differentiation within contact zones and patterns of introgression across homologous loci between contact zones. Through answering these three questions, we simultaneously compare introgression across lineage-pairs with a common biogeographic setting yet different demographic histories and across genes with variable histories within lineage-pairs.

Methods

SAMPLING

Our approach sampled transcriptome data from two allopatric populations geographically isolated from the hybrid zone ("parental" populations) to design exome capture arrays and infer patterns of genomic divergence ((Singhal 2013; Singhal and Moritz 2013); Fig. 1C, S1, Table S1). To characterize patterns of introgression, we sampled exome capture data from nine populations through each of the four hybrid zone transects. Six of the populations occurred approximately 10, 2.5, and 1 km north and south of the hybrid zone center, two occurred at the 20 and 80% tails of the average cline in the hybrid zone, and a final occurred at the average cline center (Singhal and Moritz 2013). In total, we sampled an average of 133 individuals per contact zone, and each population consisted of an average of 14.8 individuals ($N = 8 - 17$). But for individuals comprising the 10 km populations for the *L. coggeri* C/S contact zone, all other individuals were included in previous studies (Phillips *et al*. 2004; Singhal and Moritz 2012, 2013).

DATA COLLECTION

To capture a uniform subset of the genome across populations, we designed exome capture arrays based on transcriptome data from populations goeographically isolated from the contact zone (Fig. S1; (Singhal 2013)). Each array targeted a random set of exons and genes with putatively interesting biological functions and with outlier patterns of molecular evolution. To help validate our anonymous pooling approach (see below), we both optimized experimental design through simulations (Figs. S2– S4) and included loci that we had previously genotyped for the same individuals (Singhal and Moritz 2013). In total, we targeted an average of 3082 loci and 1.83 Mbp of sequence, and the four capture arrays had 1120 loci in common. Further details on the capture array design are available in the Supporting Information.

We used anonymous population pooling (Pool-Seq) to make libraries because it is an efficient way to generate population genomic data (Schlötterer *et al*. 2014). However, pooling prevents us from calling genotypes and therefore conducting individual-based analyses such as inferring genomic clines (Gompert and Buerkle 2012) and estimating levels of linkage disequilibrium beyond a sequencing read or read-pair (Feder *et al*. 2012). For each individual, we extracted DNA (Aljanabi and Martinez 1997) and measured DNA concentration using a Qubit dsDNA BR Assay kit. We then pooled equimolar amounts of DNA per individual across a population, sonicated the population pools to 150–600 bp using a Bioruptor ultrasonicator (Diagenode), and then prepared uniquely barcoded libraries (Meyer and Kircher 2010). After measuring library concentrations using a Qubit, we pooled libraries by contact to obtain a total of 20μg for exome capture. We performed capture reactions following the protocol published in (Hodges *et al*. 2009) with modifications by (Bi *et al*. 2012). To improve capture efficiency, we isolated COT-1 DNA from *L. coggeri* (Trifonov *et al*. 2009) and used a 50/50 mix of *L. coggeri* and commercially purchased chicken COT-1 DNA as our blocking reagent (Hybloc; Applied Genetics Lab). Following qPCR validation, each library was sequenced across one 100-bp paired-end lane on an Illumina HiSeq 2000 at the Vincent Coates Genome Sequencing Laboratory at University of California, Berkeley.

DATA FILTRATION, ASSEMBLY, AND VARIANT DISCOVERY

To characterize patterns of genomic divergence in populations geographically-isolated from the contact zone, we used previously published variant data from transcriptomes (Singhal and Moritz 2013). To analyze the exome capture data, we followed the pipeline for data filtration and assembly published by Bi *et al.*, 2012. Briefly, raw sequencing reads were trimmed and merged using the programs `cutadapt`, `trimmomatic`, `cope`, and `flash` (Martin 2011; Magoč and Salzberg 2011; Liu *et al*. 2012; Lohse *et al*. 2012). For each lineage-pair, we assembled and annotated these reads to generate a pseudo-reference genome (PRG) via `ABYSS`, `cap3`, and `blat` (Huang and Madan 1999; Kent 2002; Biron *et al*. 2009). We restricted all downstream analyses to only those contigs that matched our targets.

To identify variants segregating in the hybrid zones, we aligned reads from our pooled libraries to the PRG using `bowtie2` and then called variants using `SAMtools` (Langmead *et al*. 2009; Li *et al*. 2009). For this putative set of variable sites, we used `SAMtools mpileup` and `bcftools` to calculate allele frequencies for each population in the hybrid zone transect (Li *et al*. 2009) and to call genotypes for each individual from the geographically isolated populations. The pipeline outlined here is summarized in Fig. S5.

ANALYSIS

Evaluating success of the experiment

We measured the efficacy of our anonymous pooling strategy and our overall exome capture experiment. To do so, we compared allele frequencies estimated from pooled data to those estimated

from individual genotypes at 10 loci (Singhal and Moritz 2013), looked at variance in estimated allele frequencies across SNPs within the nonrecombining mtDNA, and calculated standard measures of exome capture efficacy (Parla *et al*. 2011) (see Supplemental Information).

Inferring patterns of genomic divergence

Using transcriptome data for the populations geographically distant from the contact zone, we inferred patterns of genome-wide molecular evolution and differentiation. For the coding sequence for each transcript assembled and for each exon targeted on the array, we calculated four indices: d_{xy} (Nei and Li 1979), d_a, F_{ST} (Weir and Cockerham 1984), and $\frac{dN}{dS}$ (Yang 2007). All metrics were calculated across all variants in a given contig, whether a full transcript or exon.

Inferring patterns of introgression

We used the variant data from the contact zone populations to infer introgression patterns at each SNP. We first filtered our allele frequency data by setting as missing any estimate based on less than 50× coverage for a given population (Fig. S2). We then removed any SNPs that had missing data at >2 of the transect populations. We then filtered any SNPs where the difference between the lowest and greatest allele frequencies across the parental or 10-km populations was $p_{diff} \leq 0.50$. These lowly differentiated SNPs were filtered because most did not fit clinal patterns. For the remaining SNPs, we fit sigmoidal clines to allele frequency data for the central seven populations in each transect. We omitted the 10-km populations because they were typically off the linear hybrid zone transect.

We fit only a standard sigmoid cline model to our data (Barton and Gale 1993) because (1) more complex clinal models, such as those allowing patterns at the tails of clines to vary, require denser geographic sampling and (2) previous data from these hybrid zones have found that 95% of clines best fit a sigmoidal model (Phillips *et al*. 2004; Singhal and Moritz 2013). We fit this model using a maximum-likelihood function implemented in Python (Porter *et al*. 1997). We employed a brute force approach to explore a wide ranging parameter space for cline center and width (for both, 100–20,000 m). Further, initial explorations suggested that fitting p_{min} and p_{max} from the data generated poorly fitting clines. As such, we did not fit p_{min} and p_{max}; instead, we fixed them based on estimated minimum and maximum allele frequencies inferred across all sampled populations. Introgression at a SNP was categorized as showing a "sweep" pattern if the allele frequency difference between the parental populations was $p_{diff} \geq 0.5$ and allele frequencies at all populations in the contact zone–including the 10 km populations–were uniformly between $0 \leq p \leq 0.2$ or $0.8 \leq p \leq 1.0$. Such sweep loci could arise either due to demographic or selective processes. See Supplemental Information for further details.

To address our questions, we first profiled patterns of cline center and width. Second, as a proxy for linkage disequilibrium, we estimated Moran's I, a spatial auto-correlation measure that can be applied across genomic distances (Gompert *et al*. 2012b; Parchman *et al*. 2013). In systems where selection overwhelms recombination, clinal patterns will be constant across physical genome distances and will lead to high genomic Moran's I. Where selection is weak, recombination will break down linkage disequilibrium, leading to low genomic Moran's I. To measure Moran's I, we estimated the degree of genomic spatial correlation in cline width estimates, restricting our analysis to only those targeted exons for which we inferred more than one cline. Finally, we investigated the predictability of heterogeneous introgression patterns across loci across contacts. First, we tested if locus characteristics can predict locus-specific introgression. To do so, we calculated correlations between metrics of locus divergence (see *Inferring Patterns of Genomic Divergence*) and patterns of introgression across each contact zone. In addition, we compared patterns of introgression across functional classes (as determined using Gene Ontology terms) across contacts. Here, we used the R package GOstats and the Ensembl BioMart database for *A. carolinensis* to calculate the average cline width for genes in a given GO term (Falcon and Gentleman 2007). Second, we tested the prediction that, if locus characteristics are conserved across lineages, locus-specific introgression patterns should also be correlated across lineage-pairs. To test this, we calculated correlations in cline widths among homologous loci across each pair of lineage-pairs. For this final set of analyses, we averaged cline width across both exons and genes; we estimated a mean of 2.8 and 3.7 clines per exon and per gene, respectively.

The limitations of our cline estimation approach – that is, relying on pooled data, sampling fewer individuals and demes than is standard, fixing rather than inferring p_{min} and p_{max}, not accounting for Hardy–Weinberg disequilibrium, fitting only a simple sigmoidal model – all certainly increase the error of locus-specific estimates of cline width and center (Figs. S2, S3). Further, this approach likely means we have miscategorized some portion of the variants fitting a "sweep" pattern (Figs. S6, S7). Thus, across all analyses, we refrain from making arguments about locus-specific introgression patterns, instead focusing on general patterns across groups of clines and across contact zones.

Results

EFFICACY OF EXOME CAPTURE EXPERIMENT

The exome capture experiments for each contact zone were successful; briefly, we acquired high-coverage and high-quality data for our targets, extended our in-target assembly by 60% by

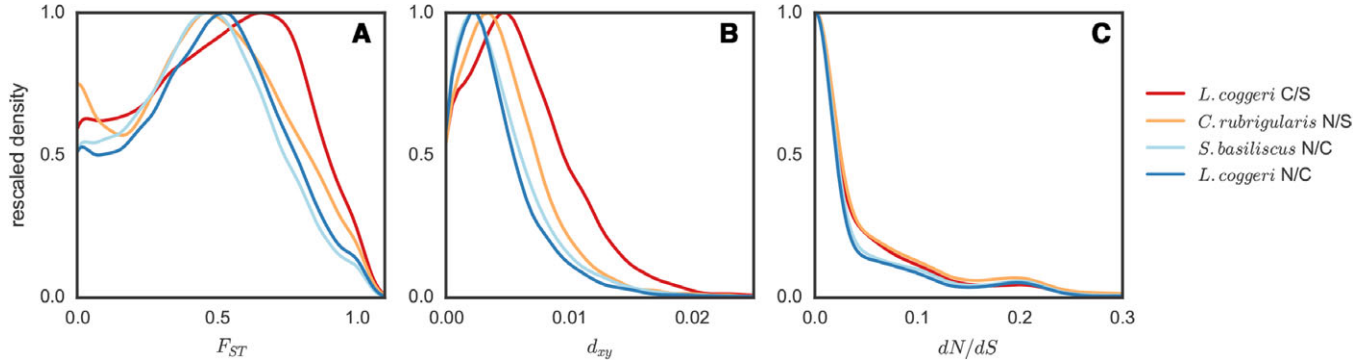

Figure 2. **Distributions for three measures of genetic divergence – (A) F_{ST}, (B) d_{xy}, and (C) *dN/dS* – calculated for the coding sequence of an average of 11K transcripts across each of the four lineage-pairs. Lineage-pairs are listed in the legend in order of most to least divergent. Rescaled density was calculated by dividing density estimates by the maximum density seen for the distribution. Across lineage-pair comparisons, the mean of the distribution changes, but the shape of the distribution remains similar.**

assembling our reads de novo, and recovered high and consistent specificity (≈65%) (see Supplementary Material for further details; Figs. S8– S14; Tables S2 and S3).

We also evaluated the success of our anonymous pooling strategy two ways. First, we compared allele frequencies estimated from pooled data to those estimated from individual-level genotyping assays (Fig. S15; (Singhal and Moritz 2013)). We find substantial and significant correlation between estimated and known allele frequencies (average $r = 0.97$), suggesting that sampling drift due to anonymous pooling was minimal. Second, we estimated variance in estimated allele frequencies at highly differentiated SNPs in the mitochondrial genome. Because the mitochondrial genome is non-recombining, we expect that all SNPs that are highly differentiated between parental populations should have similar allele frequencies within any population along the transect. We find this is true across most populations and contacts (Fig. S16).

The approach effectively discovered variation for downstream analyses; we identified an average of 57K SNPs after filtering for low coverage and high missing data (Table S4), and we fit clines at 1.5K to 13.4K of these SNPs (Fig. S17).

INFERRING PATTERNS OF DIVERGENCE AND INTROGRESSION

Genome-wide divergence

Comparing patterns of genome-wide differentiation between allopatric populations for all four lineage-pairs results in two clear patterns. First, across all three metrics and across all lineage-pair comparisons, patterns of genomic divergence are significantly correlated across genes (r=0.28–0.54; Table S5). Second, as we showed previously with ten loci (Singhal and Moritz 2013), mean values of genomic divergence (measured here as d_{xy}) are highly correlated with divergence time (r=0.99; P=0.01). Although the mean of these metrics increases as the lineage-pairs become more diverged, the shape of the distributions themselves remains similar (Fig. 2). These results suggest the heterogeneity in genomic divergence does not change dramatically across the range of divergence histories sampled here.

Introgression across contacts

We tested the influence of divergence history on introgression patterns by comparing results across contacts. First, we were able to infer introgression patterns (cline or sweep) at 25–30% of the filtered SNPs in *C. rubrigularis* N/S and *L. coggeri* C/S and in 7–10% in *S. basiliscus* N/C and *L. coggeri* N/C (Fig. S17). This difference across contact zones is partially because many SNPs in *S. basiliscus* N/C and *L. coggeri* N/C are too undifferentiated across transect populations to allow cline fitting. Additionally, we found that about 5% of the variants fit in *L. coggeri* C/S, 8% in *C. rubrigularis* N/S, 16% in *L. coggeri* N/C and 64% in *S. basiliscus* N/C show a "sweep" pattern (Fig. S17).

The distributions of cline widths across contact zones shows a clear pattern; both the average spatial extent of introgression and the variance across loci in introgression extent are reduced as divergence time between lineage-pairs increases (Fig. 3A, S18, S19). We see a similar pattern in cline centers (Fig. 3B, S18); the distribution of cline centers is significantly narrower in the two more divergent lineage-pairs than the two less divergent lineage-pairs. Despite the limitations of our cline estimation approach, cline estimates for these data concord quantitatively with those estimated previously but for *S. basiliscus* N/C, for which original data were limited (Table S6). Finally, the two less divergent lineage-pairs have near zero values of genomic Moran's I beyond 100 base pairs, suggesting very limited spatial autocorrelation and, thus, limited linkage disequilibrium (Fig. 3C, S18). The two more divergent lineage-pairs have extensive autocorrelation that extends for at least 1 kb, with only moderate declines over distance.

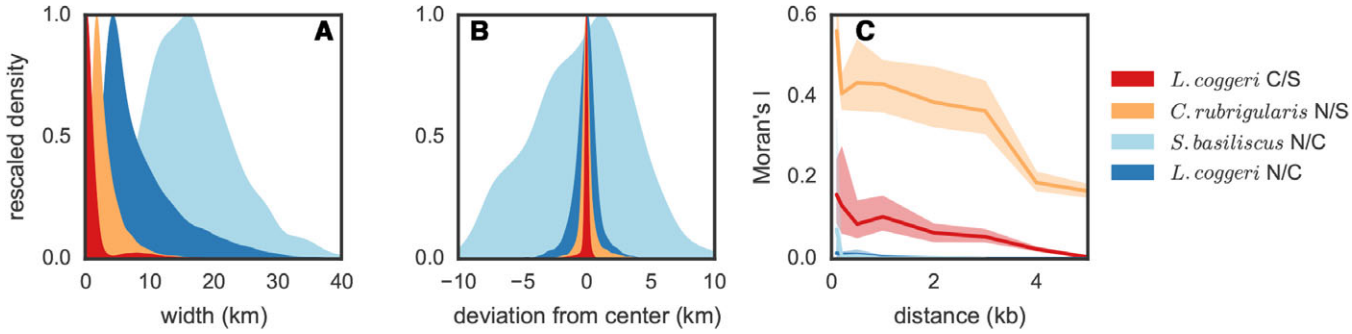

Figure 3. Distributions for (A). cline width and (B). cline center across an average of 7.5K clines at contact zones between each of the four lineage-pairs. C. Moran's I, a measure of spatial auto-correlation applied to genomic distance, for cline width at each of contact across an average of 5.3K comparisons for distances ≤500 bp and 810 comparisons for distances >500 bp. Uncertainty in Moran's I was estimated by drawing 100 bootstrap samples and recalculating means. Rescaled density was calculated by dividing density estimates by the maximum density seen for the distribution. A version of this figure showing data from only the 1120 exons shared across all four contact zones is available at Figure S18. More-divergent lineage pairs show narrower, more coincident clines than less-divergent lineage pairs.

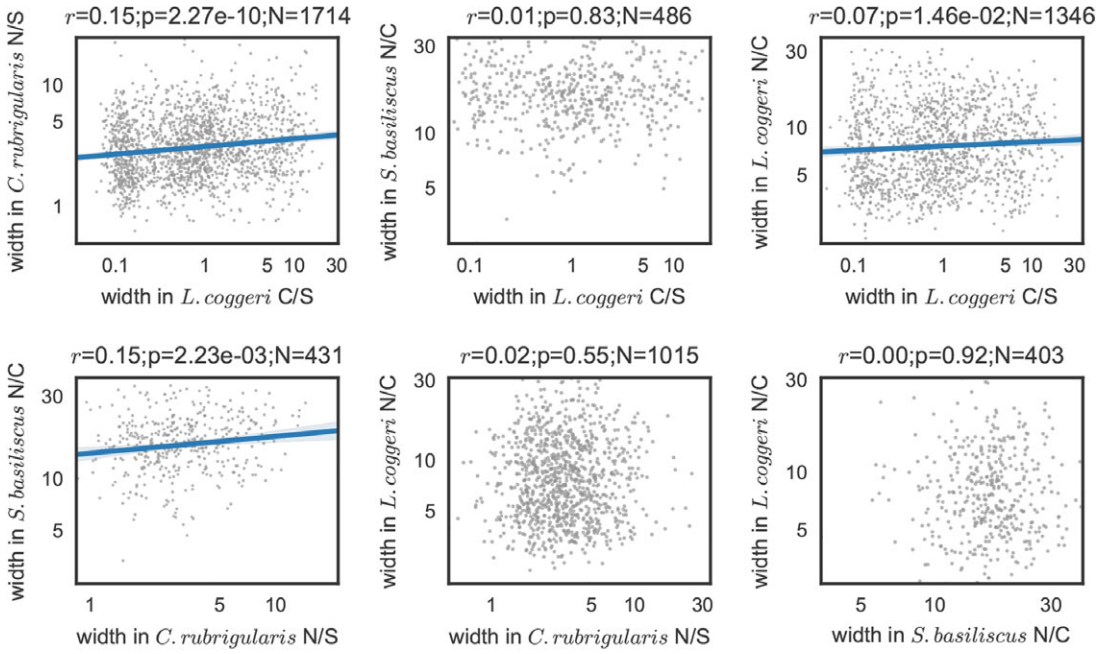

Figure 4. Pearson correlations in cline widths between all pairwise comparisons of the four sampled contacts. For each hybrid zone, we averaged cline widths for all variants within a gene to get a gene-wide estimate of cline width, took the natural log of the mean cline width to normalize data, and then compared widths between homologous loci across hybrid zones. For significant correlations, we show the linear regression with uncertainty. We see modest but significant correlations in locus-specific introgression extent in three of the six comparisons.

Introgression across the genome

To test how predictable introgression is within a contact and among contacts, we first calculated correlations between metrics of differentiation at a locus – as measured from allopatric populations – and that locus's cline width – as measured across the hybrid zone transect (Table S7). Across three of the four lineage-pairs, we find a negative correlation between average cline width and F_{ST} per exon. These correlations range from $r=-0.09$ to $r=-0.23$ and are stronger in the more divergent contacts. We recover a similar pattern for two contact zones for another relative measure of divergence, d_a. We also recover a weak and unexpectedly positive relationship between divergence and cline width in *L. coggeri* C/S only, and we find no evidence for correlations between $\frac{dN}{dS}$ and cline width. Second, we find no evidence that putative gene function influences introgression patterns either in terms of cline width or "sweep" versus clinal pattern within contact zones or across

contact zones (Table S8-10). Finally, comparing introgression patterns gene-by-gene recovers weak but significant correlations in three out of six possible comparisons (Fig. 4).

Discussion

SYSTEM-SPECIFIC PATTERNS

As shown previously (Singhal and Moritz 2013), the extent of introgression becomes more limited as these lineage-pairs become more divergent. Our work builds on these results in a few novel ways. First, earlier work from the *L. coggeri* C/S hybrid zone found that clines across ten loci were all exceptionally narrow and largely concordant (Singhal and Moritz 2012). Theory predicts that, unless selection is extremely strong, some clines will diffuse neutrally, leading to wider and nonconcordant clines (Barton 1983). Because we originally failed to recover this pattern, we hypothesized that this hybrid zone might not be at equilibrium (Singhal and Moritz 2012). In this system, neutral diffusion should result in cline widths on the order of 6–10 km. By sampling more loci, we found anywhere from 4.6 to 9% of all clines were this wide or wider, suggesting neutral diffusion is occurring, albeit rarely. Thus, in contrast to our proposed hypothesis, these data suggest that this hybrid zone is at tension zone equilibrium; however, most of the genome is subject to direct or correlated selection, leading to narrow, and coincident clines.

Second, the clines are much narrower for *L. coggeri* C/S than *C. rubrigularis* N/S (Fig. 3), and given the two species have similar dispersal rates (Phillips *et al*. 2004; Singhal and Moritz 2012), this difference suggests selection against hybrids is stronger for *L. coggeri* C/S. However, *C. rubrigularis* N/S has higher values of genomic Moran's I (Fig. 3), a result that suggests linkage disequilibrium across the genome is higher in this contact zone. While this pattern could certainly emerge due to stochastic demographic forces, it also could result from differences in the genomic structure of segments under selection, such that the individual locus effect is stronger in *C. rubrigularis* N/S, leading to more extensive spatial auto-correlation (or, linkage disequilibrium) than in *L. coggeri* C/S.

Third, earlier work showed evidence for asymmetric hybridization in both the *S. basiliscus* N/C and *L. coggeri* N/C contact zones (Singhal and Moritz 2013). For both of these hybrid zones, it was unclear if the patterns we recovered were from stochastic or selective processes. However, with a larger data set, we find that many loci show a "sweep" pattern, indicative of asymmetric introgression (Fig. S17). Given the large number of loci recovered showing this pattern, we hypothesize that this asymmetry is likely due to demographic effects, such as differences in population density between lineages or patterns of lineage range expansion (Dasmahapatra *et al*. 2002; Currat *et al*. 2008; Dufkova *et al*. 2011). Population modeling of the hybrid zone might help further characterize how demography is impacting introgression (Fitzpatrick *et al*. 2010).

IMPLICATIONS FOR SPECIATION

In contrast to expectations from lineages that diverged with gene flow (Feder *et al*. 2014), we see no evidence that genome differentiation becomes less heterogeneous as lineage-pairs become more divergent (Fig. 2). Most models for genome divergence during speciation describe how the spatial extent of differentiation across the genome changes through time; we cannot speak to these patterns because we lack an appropriate genome to scaffold our variation. Instead, we can compare the shape of the distributions through time (Fig. 2). The mean and median of divergence increase with older lineage-pairs, as expected. However, the distributions are markedly similar across lineage-pairs, despite their nearly 2× range in divergence times. This likely reflects an intermediate stage in lineage divergence where differentiation remains relatively heterogeneous for some time (Roux *et al*. 2016). Further, across lineage-pairs, we see significant correlations in patterns of divergence across genes (Table S5), a pattern that has been recovered across closely related lineages (Nadeau *et al*. 2012; Renaut *et al*. 2014) and more distant comparisons (Van Doren *et al*. 2017; Vijay *et al*. 2017). Because our lineage-pairs are sampled across three different genera spanning more than 15 million years of evolution, these correlations are unlikely to result from sorting of standing ancestral variation or introgression across species borders. Instead, these shared patterns of genomic differentiation suggest conservation in the strength of linked selection across species genomes. More generally, this pattern indicates these species show a certain predictability in their genetic divergence through time (Stern and Orgogozo 2009; Burri *et al*. 2015; Vijay *et al*. 2017).

While the increasing age of these lineage-pairs has only modest effects on the heterogeneity of genome divergence (Fig. 2), it impacts heterogeneity in genome-wide introgression more substantially. By correlating divergence history with introgression, we are implicitly testing the strength of selection on hybrids, which acts at the level of the individual and thus can affect a large portion of the genome. Strong selection against hybrids leads to extensive linkage disequilibrium in hybrid zones, preventing introgression even at loci that neither have, nor are linked to loci with, selective effect. We see this pattern in the highly divergent lineage-pairs *L. coggeri* C/S and *C. rubrigularis* N/S, both as the narrow and limited range of introgression and as the high spatial auto-correlation in introgression across genomic space (Fig. 3). These results follow nicely from population genetic descriptions of speciation as the accumulation of linkage disequilibrium (Felsenstein 1981; Kirkpatrick and Ravigne 2002). In the lesser divergent lineage-pairs *L. coggeri* N/C and *S. basiliscus* N/C, selection against hybrids appears to be weaker, and

accordingly, the extent of introgression is broader and spatial correlation is limited. This leads to the observation that history cleans up messes, or that time, and the divergence that typically accumulates with time, leads to patterns across the genome congealing such that two divergent genomes eventually act as completely isolated units (Turner 1967).

As predicted by (Barton 1983), this work shows that, as clines become more coincident, the barrier strength increases quickly, even if the average magnitude of locus-specific selection decreases. Thus, as we see here, even gradual increases in genomic divergence can dramatically decrease how permeable species boundaries are, and thus, how isolated hybridizing genomes remain. These findings nicely dovetail with results from a broader array of taxonomic lineage-pairs (Roux *et al*. 2016), in which lineages transition between acting as populations to acting as evolutionary-independent lineages over a "gray zone" of silent net genomic divergence that extends fourfold. Silent net divergence (d_a) between our four lineage-pairs ranges from 0.51% to 0.86%, placing them all within this "gray zone." We find this transition occurs even more quickly across this more narrow phylogenetic sampling. D_a for *C. rubrigularis* N/S, which shows both limited introgression and evidence for extensive disequilibrium in the hybrid zone, is just 1.15× greater than that for the two lesser divergent lineage pairs.

Although the scale of heterogeneity in introgression decreases with the increasing age of lineage-pairs, all contact zones show evidence for heterogeneous introgression. Introgression extent across loci varies both due to variance in selection against introgressing loci and in recombination rates across a genome (Barton and Bengtsson 1986). Here, we find evidence for some of the possible factors structuring this variance. Our predictive ability is modest, particularly given the striking differences in introgression extent across contact zones, and likely both reflects noise introduced by our approach to cline inference (Figs. S3 and S4) and the complexity of factors structuring introgression in hybrid zones. As shown by other studies relating genomic divergence to introgression (Gompert *et al*. 2012a; Nosil *et al*. 2012), we see decreased introgression at regions of the genome that are more highly differentiated among allopatric populations, although the strength of this correlation is weak (Table S7). Notably, we only see this correlation with our relative measures of divergence (F_{ST} and d_a) and not our absolute measure of divergence (d_{xy}) (Cruickshank and Hahn 2014). These correlations do not appear to be artifacts of how the data were filtered; we find the same pattern even when we restrict our analyses to just those clines fit at the most differentiated variants (i.e., those where $p_{max} - p_{min} > 0.9$). As has been show in numerous studies (Cruickshank and Hahn 2014), we find that regions of high relative differentiation occur in areas of reduced diversity (Table S11). Due to the effects of linked selection, low recombination rates can lead to lower levels of diversity (Charlesworth *et al*. 1993). Because we only recovered a correlation between introgession extent and relative differentiation but not absolute differentiation, this suggests that recombination is playing an important role in structuring introgression patterns. In areas of low recombination, linkage blocks should be physically larger and thus more likely to have a larger selective effect in foreign genomic backgrounds than a smaller linkage block. Thus, lower recombination rates can lead to reduced introgression (Slatkin 1975; Kruuk *et al*. 1999). We find some preliminary support for this hypothesis (Table S12), but testing this prediction properly will require a better characterization of these species' recombination landscapes and the genetic architecture of traits under selection in hybrids.

Given our modest ability to predict patterns of introgression across the genome based on proxies of selection and recombination (Table S7), and given that these proxies are correlated across species (Table S5), we might further expect correlated introgression patterns across contact zones. Indeed, we see low but significant correlations across genes in cline widths between three of the six possible comparisons between contacts (Fig. 4). At least for *C. rubrigularis* N/S - *L. coggeri* C/S, this correlation is greater than expected given that patterns of genomic differentiation are correlated across lineage-pairs and that patterns of genomic differentiation predict introgression (Table S13). In many studies, multiple transects across the same hybrid zone exhibit considerably different patterns of introgression (Teeter *et al*. 2010; Harrison and Larson 2014), which suggests that introgression patterns can be somewhat idiosyncratic depending on the demography and geography at a given transect. In contrast, these results indicate that common factors structure introgression across these contacts, even though these contacts occur in different genera. These factors could include (1) selection against introgression in highly differentiated loci, given that differentiation is correlated across contacts (Table S5), (2) similar effects of recombination if the recombination landscape is conserved across taxa (Janousek *et al*. 2012), and (3) although we find no evidence for such effects here (Tables S8, S9), selection against loci involved in species-specific traits – that is genes involved in gametic isolation. Together, these correlated patterns of both genomic divergence and introgression suggest there is modest predictability to how genomes diverge and how divergent genomes interact across the speciation process.

Conclusions

Most studies investigating lineage divergence through genomic divergence have focused on lineages exchanging large number of migrants every generation and between which there is marked ecological differentiation (Jones *et al*. 2012; Nosil 2012; Nosil *et al*. 2012; Renaut *et al*. 2013; Harrison and Larson 2014; Malinsky

et al. 2015). These studies have shown that increased ecological differentiation, often when combined with geographic isolation, leads to ever increasing genome-wide differentiation and reduced heterogeneity in divergence through time. In contrast, this study focuses on lineages that diverged with minimal gene flow and that are ecologically and morphologically similar (Singhal and Moritz 2013); such divergences are an important, and in some ways understudied, component of Earth's biodiversity (Bickford *et al*. 2006). For these lineages, we find that the evolution of isolation across the genome is an iterative process.

Further, although heterogeneity in genomic divergence does not decrease as these lineage-pairs get older, we recover a marked decrease in the heterogeneity of introgression across the same time span. As such, our results underscore how time can lead to genomes "congealing." Given the current emphasis on how ecology drives lineage divergence (Schluter 2009; Nosil 2012), this work reminds us that history plays a pivotal role in species formation and maintenance, as well.

AUTHOR CONTRIBUTIONS

SS designed the project, SS and KB collected data, and SS analyzed data and wrote paper.

ACKNOWLEDGMENTS

We thank D.B. Wake for thoughtful conversations, during which he shared the idea that "history cleans up messes," which served as this article's inspiration. For advice, we gratefully acknowledge C. Moritz and M. Slatkin, for technical support, M. Chung and L. Smith, and for helpful comments on previous versions of this manuscript, S. Baird, I. Holmes, C. Moritz, M. Nidiffer, J. Peñalba, and two anonymous reviewers. Funding was provided by the Museum of Vertebrate Zoology Koford & Albert Preston Fund, NSF DDIG, and NSF Postdoctoral Fellowship in Biology. The Texas Advanced Computing Center (TACC) at The University of Texas at Austin provided grid resources that contributed to the research results reported within this article. This work used the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by NIH S10 Instrumentation Grants S10RR029668 and S10RR027303.

DATA ACCESSIBILITY

Data are available at the following locations:

- Scripts used in analysis, including a README for guidance: https://github.com/singhal/introgression_AWT
- Probes for exome capture at DataDryad: 10.5061/dryad.dg513
- Reference assemblies for targets at DataDryad: 10.5061/dryad.dg513
- Raw short-read data at NCBI SRA: PRJNA386488
- Summary of genotype calls and estimated allele frequencies at Data Dryad: 10.5061/dryad.dg513

LITERATURE CITED

Aljanabi, S., and I. Martinez. 1997. Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. Nucl. Acids Res. 25:4692–4693.

Baird, S. 1995. A simulation study of multilocus clines. Evolution 49:1038–1045.

Barton, N. 1983. Multilocus clines. Evolution 37:454–471.

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

Barton, N., and K. Gale. 1993. Genetic analysis of hybrid zones. Pp. 13–45 *in* R. G. Harrison, ed. Hybrid zones and the evolutionary process. Oxford Univ. Press, Oxford, U. K.

Bazykin, A. 1969. A hypothetical mechanism of speciation. Evolution 23:685–687.

Bi, K., D. Vanderpool, S. Singhal, T. Linderoth, C. Moritz, and J. M. Good. 2012. Transcriptome-based exon capture enables highly cost-effective comparative genomic data collection at moderate evolutionary scales. BMC Genomics 13:1.

Bickford, D., D. Lohman, N. Sodhi, P. Ng, R. Meier, K. Winker, K. Ingram, and I. Das. 2006. Cryptic species as a window on diversity and conservation. TREE 22:148–155.

Biron, I., S. Jackman, C. Nielsen, J. Qian, R. Varhol, G. Stazyk, R. Morin, Y. Zhao, M. Hirst, J. Schein, et al. 2009. De novo transcriptome assembly with ABySS. Bioinformatics 25:2872–2877.

Burri, R., A. Nater, T. Kawakami, C. F. Mugal, P. I. Olason, L. Smeds, A. Suh, L. Dutoit, S. Bureš, L. Z. Garamszegi, et al. 2015. Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of ficedula flycatchers. Genome Res. 25:1656–1665.

Charlesworth, B., M. Morgan, and D. Charlesworth. 1993. The effect of deleterious mutations on neutral molecular variation. Genetics 134:1289–1303.

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

Currat, M., M. Ruedi, R. Petit, and L. Excoffier. 2008. The hidden side of invasions: massive introgression by local genes. Evolution 62:1908–1920.

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

Dobzhansky, T. 1934. Studies on hybrid sterility. i. spermatogenesis in pure and hybrid *Drosophila pseudoobscura*. Zeitschrift fur Zellforschung und mikroskopische Anatomie 21:169–221.

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

Endler, J. 1977. Geographic variation, speciation and clines. Princeton Univ. Press, Princeton, NJ.

Falcon, S., and R. Gentleman. 2007. Using gostats to test gene lists for go term association. Bioinformatics 23:257–258.

Feder, A. F., D. A. Petrov, and A. O. Bergland. 2012. Ldx: estimation of linkage disequilibrium from high-throughput pooled resequencing data. PLoS One 7:e48588.

Feder, J. L., P. Nosil, A. C. Wacholder, S. P. Egan, S. H. Berlocher, and S. M. Flaxman. 2014. Genome-wide congealing and rapid transitions across the speciation continuum during speciation with gene flow. J. Heredity 105:810–820.

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

Fisher, R. 1950. Gene frequencies in a cline determined by selection and diffusion. Biometrics 6:353–361.

Fitzpatrick, B. M., J. R. Johnson, D. K. Kump, J. J. Smith, S. R. Voss, and H. B. Shaffer. 2010. Rapid spread of invasive genes into a threatened native species. Proc. Natl. Acad. Sci. 107:3606–3610.

Gompert, Z., and C. Buerkle. 2012. bgc: software for Bayesian estimation of genomic clines. Mol. Ecol. Resources 12:1168–1176.

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

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

Harrison, R. G., and E. L. Larson. 2014. Hybridization, introgression, and the nature of species boundaries. J. Heredity 105:795–809.

———. 2016. Heterogeneous genome divergence, differential introgression, and the origin and structure of hybrid zones. Mol Ecol. 105:795–809.

Hodges, E., M. Rooks, Z. Xuan, A. Bhattacharjee, D. B. Gordon, L. Birzuela, W. R. McCombie, and G. Hannon. 2009. Hybrid selection of discrete genomic intervals on custom-designed microarrays for massively parallel sequencing. Nat. Protoc. 4:960–974.

Huang, X., and A. Madan. 1999. CAP3: a DNA sequence assembly program. Genome Res. 9:868–877.

Janousek, V., L. Wang, K. Luzynski, P. Dufkova, M. Vyskocilova, M. Nachman, P. Munclinger, M. Macholan, J. Pialek, and P. Tucker, 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.

Jones, F., M. Grabherr, Y. Chan, P. Russell, E. Mauceli, J. Johnson, R. Swofford, M. Pirun, M. Zody, S. White, et al. 2012. The genomic basis of adaptive evolution in threespine sticklebacks. Nature 484:55–61.

Kent, W. 2002. BLAT–the BLAST-like alignment tool. Genome Res. 12:656–664.

Key, K. 1968. The concept of stasipatric speciation. Syst. Biol. 17:14–22.

Kirkpatrick, M., and V. Ravigne. 2002. Speciation by natural and sexual selection: models and experiments. Am. Nat. 159:S22–S35.

Kruuk, L., S. Baird, K. Gale, and N. Barton. 1999. A comparison of multilocus clines maintained by environmental selection or by selection against hybrids. Genetics 153:1959–1971.

Langmead, B., C. Trapnell, M. Pop, and S. Salzberg. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10:25.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map (SAM) format and SAMtools. Bioinformatics 25:2078–2079.

Liu, B., J. Yuan, S. Yiu, Z. Li, Y. Xie, Y. Chen, Y. Shi, H. Zhang, Y. Li, T. Lam, and R. Luo. 2012. COPE: an accurate k-mer based pair-end reads connection tool to facilitate genome assembly. Bioinformatics 28:2870–2874.

Lohse, M., A. Bolger, A. Nagel, A. Fernie, J. Lunn, M. Stitt, and B. Usadel. 2012. RobiNA: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics. Nucleic Acids Research 40:622–627.

Magoč, T., and S. Salzberg. 2011. FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics 27:2957–2963.

Malinsky, M., R. J. Challis, A. M. Tyers, S. Schiffels, Y. Terai, B. P. Ngatunga, E. A. Miska, R. Durbin, M. J. Genner, and G. F. Turner. 2015. Genomic islands of speciation separate cichlid ecomorphs in an east african crater lake. Science 350:1493–1498.

Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. Available at URL http://journal.embnet.org/index.php/embnetjournal/article/view/200.

Martinsen, G., T. Whitham, R. Turek, and P. Keim. 2001. Hybrid populations selectively filter gene introgression between species. Evolution 55:1325–1335.

Meyer, M., and M. Kircher. 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harb. Protoc. 6:pdb.prot5448.

Moritz, C., C. Hoskin, J. MacKenzie, B. Phillips, M. Tonione, N. Silva, J. VanDerWal, S. Williams, and C. Graham. 2009. Identification and dynamics of a cryptic suture zone in a tropical rainforest. Proc. Roy. Soc. B 276:1235–1244.

Muller, H. 1942. Isolation mechanisms, evolution and temperature. Biol. Symposium 6:71–125.

Nadeau, N. J., A. Whibley, R. T. Jones, J. W. Davey, K. K. Dasmahapatra, S. W. Baxter, M. A. Quail, M. Joron, M. L. Blaxter, J. Mallet, et al. 2012. Genomic islands of divergence in hybridizing heliconius butterflies identified by large-scale targeted sequencing. Phil. Trans. R. Soc. B 367:343–353.

Nei, M., and W.-H. Li. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc. Natl. Acad. Sci. 76:5269–5273.

Nosil, P. 2012. Ecological Speciation. Oxford Univ. Press, Oxford, U.K.

Nosil, P., T. Parchman, J. Feder, and Z. Gompert. 2012. Do highly divergent loci reside in genomic regions affecting reproductive isolation? a test using next-generation sequence data in Timema stick insects. BMC Evo. Bio. 12:164.

Parchman, T., Z. Gompert, M. Braun, R. Brumfield, D. McDonald, J. Uy, G. Zhang, E. Jarvis, B. Schlinger, and C. Buerkle. 2013. The genomic consequences of adaptive divergence and reproductive isolation between species of manakins. Mol. Ecol. 22:3304–17.

Parla, J., I. Iossifov, I. Grabill, M. Spector, M. Kramer, and W. McCombie. 2011. A comparative analysis of exome capture. Genome Biol. 12:R97.

Phillips, B., S. Baird, and C. Moritz. 2004. When vicars meet: a narrow contact zone between morphologically cryptic phylogeographic lineages of the rainforest skink, *Carlia rubrigularis*. Evolution 58:1536–1548.

Pool, J., and R. Nielsen. 2009. Inference of historical changes in migration rate from the lengths of migrant tracts. Genetics 181:711–719.

Porter, A., R. Wenger, H. Geiger, A. Scholl, and A. Shapiro. 1997. The *Pontia daplidice-edusa* hybrid zone in northwestern Italy. Evolution 52:1561–1573.

Remington, C. 1968. Suture zones of hybrid interaction between recently joined biotas. Pp. 321–428 *in* T. Dobzhansky, M. Hecht, and W. Steere, eds. Evolutionary biology. Plenum, New York.

Renaut, S., C. Grassa, S. Yeaman, B. Moyers, Z. Lai, N. Kane, J. Bowers, J. Burke, and L. Rieseberg. 2013. Genomic islands of divergence are not affected by geography of speciation in sunflowers. Nat. Comm. 4:1827.

Renaut, S., G. L. Owens, and L. H. Rieseberg. 2014. Shared selective pressure and local genomic landscape lead to repeatable patterns of genomic divergence in sunflowers. Mol. Ecol. 23:311–324.

Roux, C., C. Fraisse, J. Romiguier, Y. Anciaux, N. Galtier, and N. Bierne. 2016. Shedding light on the grey zone of speciation along a continuum of genomic divergence. PLoS Biol. 14:e2000234.

Schlötterer, C., R. Tobler, R. Kofler, and V. Nolte. 2014. Sequencing pools of individuals [mdash] mining genome-wide polymorphism data without big funding. Nat. Rev. Genet. 15:749–763.

Schluter, D. 2001. Ecology and the evolution of the species. TREE 16:372–380.

———. 2009. Evidence for ecological speciation and its alternative. Science 323:737–741.

Singhal, S. 2013. De novo transcriptomic analyses for non-model organisms: an evaluation of methods across a multi-species data set. Mol. Ecol. Res. 13:403–416.

Singhal, S., E. M. Leffler, K. Sannareddy, I. Turner, O. Venn, D. M. Hooper, A. I. Strand, Q. Li, B. Raney, C. N. Balakrishnan, et al. 2015. Stable recombination hotspots in birds. Science 350:928–932.

Singhal, S., and C. Moritz. 2012. Strong selection maintains a narrow hybrid zone between morphologically cryptic lineages in a rainforest lizard. Evolution 66:1474–1489.

———. 2013. Reproductive isolation between phylogeographic lineages scales with divergence. Proc. Royal Soc. Lond. B Biol. Sci. 280: 20132246.

Slatkin, M. 1973. Gene flow and selection in a cline. Genetics 75:733–756.

———. 1975. Gene flow and selection in a two-locus system. Genetics 81:787–802.

Stern, D. L., and V. Orgogozo. 2009. Is genetic evolution predictable? Science 323:746–751.

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

Trifonov, V., N. Vorobieva, and W. Rens. 2009. FISH with and without COT1 DNA. Pp. 99–109 *in* T. Liehr, ed. Fluorescence in situ hybridization FISH. Springer Berlin, Berlin, Germany.

Turner, J. R. 1967. Why does the genotype not congeal? Evolution 21:645–656.

Van Doren, B. M., L. Campagna, B. Helm, J. C. Illera, I. J. Lovette, and M. Liedvogel. 2017. Correlated patterns of genetic diversity and differentiation across an avian family. Mol. Ecol. Available at http://onlinelibrary.wiley.com/doi/10.1111/mec.14083/full.

Vijay, N., M. Weissensteiner, R. Burri, T. Kawakami, H. Ellegren, and J. B. Wolf. 2017. Genome-wide signatures of genetic variation within and between populations-a comparative perspective. bioRxiv P. 104604. Avaliable at http://onlinelibrary.wiley.com/doi/10.1111/mec.14195/full.

Weir, B. S., and C. C. Cockerham. 1984. Estimating f-statistics for the analysis of population structure. Evolution 38:1358–1370.

Williams, S., J. VanDerWal, J. Isaac, L. Shoo, C. Storlie, S. Fox, E. Bolitho, C. Moritz, C. Hoskin, and Y. Williams. 2010. Distributions, life history characteristics, ecological specialization and phylogeny of the rainforest vertebrates in the Australian Wet Tropics bioregion. Ecology 91: 2493.

Wu, C. 2001. The genic view of the process of speciation. J. Evol. Bio. 14:851–865.

Yang, Z. 2007. Paml 4: phylogenetic analysis by maximum likelihood. Mol. Bio. Evol. 24:1586–1591.

Associate Editor: S. Baird
Handling Editor: M. Servedio

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Figure S1: Basic sampling scheme used in this study.
Figure S2: Results from simulations exploring how increasing sequencing effort influences our ability to infer allele frequencies from pooled populations accurately.
Figure S3: Results from simulations exploring role of sampling drift in inferring cline widths from pooled populations.
Figure S4: Results from simulations exploring role of sampling drift in inferring cline centers from pooled populations.
Figure S5: Summary of bioinformatics and inference pipeline used in this study.
Figure S6: Results from simulations estimating the false-negative rate in inferring loci with a "sweep" pattern from pooled populations.
Figure S7: Results from simulations estimating the false-positive rate in inferring loci with a "sweep" pattern from pooled populations.
Figure S8: Density histograms comparing distributions of summary statistics (d_{xy}, *dN/dS*, *FST*) for all transcripts sequenced for all seven focal lineages (in red) and for the subset of transcripts targeted on exome capture arrays (in blue).
Figure S9: Specificity, or proportion of cleaned reads mapping onto target exons, summarized across all libraries for a given capture.
Figure S10: Correlation in coverage across the same loci from different libraries from the same capture experiment.
Figure S11: For orthologs shared across multiple exome capture arrays, correlation in coverage between different capture experiments.
Figure S12: Density plots of locus-wide coverage by capture experiment, with frequencies shown on the left *y*-axis.
Figure S13: Correlation between coverage at a given locus and its sequence divergence (d_{xy}) between lineages in the contact zone.
Figure S14: Relationship between coverage at a given locus and GC-content at that locus.
Figure S15: Correlation between allele frequencies estimated from individual genotyping (Singhal and Moritz 2013) and allele frequencies estimated from sequencing anonymously pooled populations of the same individuals.
Figure S16: Variance in allele frequency estimates across nearly -fixed and fixed single nucleotide polymorphisms (SNPs) between the mtDNA sequences of the two lineages meeting in each contact zone.
Figure S17: Type of clines inferred at filtered SNPs by each contact zone.
Figure S18: Distributions for A. cline width and B. cline center across an average of 3.2K clines from loci captured across all four lineage-pairs.
Figure S19: A test for concordance in cline widths across contact zones.
Table S1: Summary of geographic locations and sample sizes of populations included in this work.
Table S2: Summary of exome capture array designs and resulting assemblies.
Table S3: Summary of data collected, coverage, and specificity across sequenced populations.
Table S4: Summary of single nucleotide polymorphisms (SNPs) discovered in the captured targets.
Table S5: Pearson correlations for gene-by-gene comparisons for three different metrics that characterize patterns of locus evolution between lineages.
Table S6: Comparison of cline widths estimated from individual genotypes across many fewer markers ($N = 2–11$), as previously published in (Singhal and Moritz 2013), and cline widths estimated from pooled data across many more markers ($N = 1.5K – 13.4K$).
Table S7: Pearson correlations between summary statistics for genomic divergence and cline width across the four sampled transects.
Table S8: The number of Gene Ontology (GO) terms that had significantly narrower cline widths than the background cline width, across different α-values.

Table S9: The number of significant Gene Ontology (GO) terms that are shared between contacts.
Table S10: Pearson correlations across average cline widths for Gene Ontology (GO) terms across all six contact comparisons.
Table S11: Spearman correlations between locus-specific measures of relative differentiation (F_{ST}) and diversity (π).
Table S12: Spearman correlations between locus-specific measures of GC* and average cline width at that locus.
Table S13: Linear-model fitting results used to determine if the significant correlation in cline widths between contacts (Fig. 4) is merely because cline widths are correlated to F_{ST} (Table S7), which is also correlated between contacts (Table S5).