**A genomic investigation of the putative contact zone
between divergent Brown Creeper (*Certhia americana*)
lineages: chromosomal patterns of genetic differentiation**

Journal:	<i>Genome</i>
Manuscript ID	gen-2015-0093.R1
Manuscript Type:	Article
Date Submitted by the Author:	14-Oct-2015
Complete List of Authors:	Manthey, Joseph; University of Kansas, Department of Ecology and Evolutionary Biology and Biodiversity Institute Robbins, Mark; University of Kansas, Biodiversity Institute Moyle, Robert; University of Kansas, Department of Ecology and Evolutionary Biology and Biodiversity Institute
Keyword:	contact zone, next generation sequencing, chromosomal patterns, sky islands

SCHOLARONE™
Manuscripts

A genomic investigation of the putative contact zone between divergent Brown Creeper (*Certhia americana*) lineages: chromosomal patterns of genetic differentiation

^{1,2}Manthey, Joseph D.; ¹Robbins, Mark B.; ^{1,2}Moyle, Robert G.

¹Biodiversity Institute, University of Kansas, Lawrence, KS 66045, USA

²Department of Ecology and Evolutionary Biology, University of Kansas, Lawrence, KS 66045, USA

Corresponding author: Manthey, Joseph D.

Dyche Hall, 1345 Jayhawk Blvd.

University of Kansas

Lawrence, KS 66045-7561

United States

Email: jdmanthey@gmail.com

Running head: Chromosomal differentiation in Brown Creeper contact zone

ABSTRACT

Sky islands, or montane forest separated by different lowland habitats, are highly fragmented regions that potentially limit gene flow between isolated populations. In the sky islands of the Madrean Archipelago (Arizona, USA), various taxa display different phylogeographic patterns, from unrestricted gene flow among sky islands to complex patterns with multiple distinct lineages. Using genomic-level approaches allows the investigation of differential patterns of gene flow, selection, and genetic differentiation among chromosomes and specific genomic regions between sky island populations. Here, we used thousands of SNPs to investigate the putative contact zone of divergent Brown Creeper (*Certhia americana*) lineages in the Madrean Archipelago sky islands. We found the two lineages to be completely allopatric (during the breeding season) with a lack of hybridization and gene flow between lineages and no genetic structure among sky islands within lineages. Additionally, the two lineages inhabit different climatic and ecosystem conditions and have many local primary song dialects in the southern Arizona mountain ranges. We identified a positive relationship between genetic differentiation and chromosome size, but the sex chromosome (Z) was not found to be an outlier. Differential patterns of genetic differentiation per chromosome may be explained by genetic drift—possibly in conjunction with non-random mating and non-random gene flow—due to variance in recombination rates among chromosomes.

Keywords: contact zone, chromosomal patterns, next-generation sequencing, sky islands

INTRODUCTION

Understanding the processes driving lineage divergence and limiting gene flow across species boundaries has been a longstanding focus of evolutionary biology (Mayr 1942). Highly fragmented landscapes escalate restrictions to gene flow among populations, potentially leading to increased genetic differentiation. The sky islands of the Madrean Archipelago represent one such example of a fragmented landscape for forest species; here, isolated patches of forest (i.e., the sky islands) are separated by desert, grassland, and scrub habitat. In poorly dispersing species, independent lineages have evolved on different sky islands, suggesting little dispersal between populations [e.g., jumping spiders (Masta 2000), bears (Atwood et al. 2011)]. In contrast, species with higher dispersal ability may show fewer restrictions on gene flow among sky islands [e.g., birds (Manthey & Moyle 2015)]. With new methods for obtaining reduced-representation genomic libraries [e.g., restriction-site associated DNA sequencing (RAD-seq); Miller et al. 2007] for many individuals, genetic structure and gene flow between sky islands can now be investigated at the genomic scale. Although researchers have taken a genomic approach to investigate gene flow and selection across species boundaries in many organisms, including—but not limited to—trees (Hersch-Green et al. 2014), mussels (Gardner & Wei 2015), snakes (Schield et al. 2015), fishes (Malek et al. 2012), birds (Lavretsky et al. 2015), and mammals (Janoušek et al. 2012), sky island taxa have received little attention at the genomic level.

Songbirds have high levels of interchromosomal synteny (Kawakami et al. 2014) and the genomic resources of a well-annotated genome (Estrildidae: *Taeniopygia guttata*; Warren et al. 2010). These patterns and tools thus allow the use of high-throughput sequence data of non-model songbirds, with the possibility to identify genomic regions biased toward a lack of gene flow or increased selective pressures (e.g., sex chromosomes). Biased levels of divergence, lack of gene flow, and greater selective pressures on sex chromosomes have been found in many avian contact zones using large genetic datasets, including flycatchers (*Ficedula*; Ellegren et al. 2012), chickadees (*Poecile*; Taylor et al. 2014), and sparrows (*Passer*; Elgvin et al. 2011), suggesting a common pattern in the Z chromosome's role in speciation. Additionally, this pattern has been found in ducks across populations and species (Lavrestsky et al. 2015).

In contrast, a recent study investigating genomic patterns of differentiation between two divergent lineages of Brown Creeper (*Certhia americana*) did not find biased differentiation of the Z chromosome between lineages, but rather a positive relationship between chromosome size and genetic differentiation (Manthey et al. 2015). The two divergent [~5% mitochondrial DNA (mtDNA) divergence; Manthey et al. 2011a] lineages come into contact in Arizona, although it is unclear exactly where the contact occurs. Marshall (1956) studied the plumage morphology of *C. americana* in Arizona (USA), where the defining differences between northern and southern forms were coloration of the rump (tawny vs. chestnut, respectively) and coloration of the underparts (white vs. sooty relative to throat color). He found the transition from northern to southern birds between the Santa Catalina Mountains and Huachuca Mountains (Fig. 1; his Fig. 11),

with intermediate colorations in the Santa Catalina, Rincon, Santa Rita, Chiricahua, and Huachuca ranges. This region includes the transition from temperate to subtropical coniferous forests (Wade et al. 2003), which may act as a mechanism for speciation or biased dispersal within a species. Indeed, the climatic extremes in this region (minimum precipitation of driest month and hottest temperature warmest month) were shown to correspond with genetic differentiation in the White-breasted Nuthatch (*Sitta carolinensis*; Manthey & Moyle 2015). The northern and southern forms of *C. americana* also differ in migratory behavior; the southern forms appear to be resident and the northern forms are partially migratory (Phillips et al. 1964). Wintering and migrant specimens of the northern form have been collected in southern Arizona, within the distribution of the southern form (Phillips et al. 1964).

Previous genetic studies (Manthey et al. 2011a,b; Manthey et al. 2015) of *C. americana* have included sparse sampling from Arizona, only including individuals from the Kaibab National Forest surrounding the Grand Canyon (northern lineage) and the Chiricahua Mountains (southern lineage; Fig. 1). The lack of genetic sampling across the putative contact zone, a transition in color pattern across multiple mountain ranges, and chromosomal variation in patterns of genetic differentiation in birds suggest that a genomic-level investigation of *C. americana* in Arizona is needed. Additionally, a genomic-level investigation in a Madrean Archipelago sky island taxon will build upon previous studies to infer comparative population genetic patterns and processes [e.g., birds (Manthey & Moyle 2015), mammals (Fitak et al. 2013, Atwood et al. 2011), invertebrates (Masta 2000, Smith & Farrell 2005), trees (Potter et al. 2013)].

Here, using large numbers of SNPs across the genome, we investigate a putative contact zone between two lineages of *C. americana* and address the following questions and hypotheses:

1. Where is the genetic break between lineages?

H₀: The genetic break matches morphology, and is clinal across several mountain ranges in Arizona.

H_A: The genetic break is narrow and lineages are allopatric.

2. Within the zone of plumage transition, do the two lineages show a positive relationship between genetic differentiation and chromosome size, as was shown in a broader phylogeographic study?

H₀: The positive relationship between chromosome size and genetic differentiation remains intact in the contact zone.

H_A: In the contact zone, this relationship breaks down.

MATERIALS AND METHODS

Sampling, sequencing, and SNP dataset creation

Fresh tissue samples of 45 *C. americana* were obtained across the putative contact zone in the sky islands of Arizona (Fig. 1, Table 1). The sky islands are islands of montane forest habitat separated by lowland scrub, grassland, and desert habitat. Four samples each of the northern and southern lineages were acquired from Utah, USA (northern) and

Jalisco, Mexico (southern) to use as pure parental samples (Table 1). A single *C. familiaris* sample was used as an outgroup (Table 1). We used a QIAGEN DNeasy blood and tissue kit to extract genomic DNA for each sample.

We performed a modified RAD-seq (Miller et al. 2007) protocol to obtain thousands of genetic loci spread across the genome. We used the restriction enzyme NdeI to digest DNA samples as the first step to create a reduced representation genomic library. Custom adapters with attached barcodes for multiplexing were ligated to digested DNA samples followed by pooling and purification of all samples using an Agencourt AMPure magnetic bead cleanup. To further reduce the genomic libraries, we size-selected fragments in the range of 500 to 600 bp using a Pippin Prep electrophoresis cassette (Sage Science). Libraries were cleaned again with a bead cleanup, followed by a brief polymerase chain reaction (PCR) of each library in duplicate. PCR conditions were as follows: an initial denaturation period of 98°C for 30 s, 14 cycles of 98°C for 10 s, 64°C for 30 s, and 72°C for 20 s, and a final extension of 72°C for 7 min. PCR reactions were cleaned a final time using a bead cleanup and were tested for DNA quality and quantity using an Agilent Tapestation and quantitative PCR at the University of Kansas Genome Sequencing Core Facility. All libraries were pooled and sequenced on a partial lane (~56%) of an Illumina HiSeq2500 100 bp single-end sequencing run.

We used the STACKS (Catchen et al. 2011) pipeline to assemble loci de novo from fastQ sequence files obtained from the Illumina sequencing run. Sequence reads of poor quality were removed using a quality threshold as an average phred score of ten in sliding windows of 15 bp using the *process_RADtags* python script included in STACKS.

Sequences lacking the restriction site were removed. The *ustacks*, *cstacks*, and *sstacks* modules of STACKS were used to assemble catalogs of loci for each individual as well as a catalog of overlapping loci among all individuals (mismatches allowed between individuals = 5). Lastly, the populations module of STACKS was used to create SNP datasets with the following restrictions: minimum stack depth of five, minimum minor allele frequency of 0.05, and observed heterozygosity of 0.5 or less to reduce inclusion of paralogous loci. With these conditions, two SNP datasets were created with different levels of coverage: 50%, and 75% coverage matrices. Here, coverage was determined for three groups: northern parentals, southern parentals, and individuals near the putative contact zone. For loci to be included, the coverage threshold needed to be met in each of these groups. For example, in the 75% coverage dataset, three northern parentals, three southern parentals, and 34 individuals in the contact zone needed to have data for a particular SNP in order for that SNP to be included.

SNP dataset analyses

We used the BLAST+ utility (Camacho et al. 2009) to match loci in our study to chromosomes in the Zebra Finch (*Taeniopygia guttata*). This is possible due to high levels of synteny in songbirds, but is limited in identifying chromosomal location due to high rates of intrachromosomal recombination in songbirds (Kawakami et al. 2014). Loci were determined a match to the Zebra Finch genome with 70% sequence identity and a maximum e-value of 0.01. We initially tested multiple e-values (0.01, 0.001, 0.0001), but

the number of loci matching each chromosome was strongly related between datasets ($R^2 > 0.99$); we therefore proceeded using the e-value less than 0.01 results.

For each chromosome, all matched loci were rerun in STACKS to estimate F_{ST} between northern and southern lineages (based on STRUCTURE results; See RESULTS and Fig. 1). Linear regression of chromosome size and F_{ST} was performed in R (RDC Team 2011), using chromosomes with a minimum of ten loci.

All analyses were performed for both SNP datasets. To identify any potential admixture or hybridization, we used the program STRUCTURE (Pritchard et al. 2000) with an a priori number of genetic clusters equaling two ($K = 2$). Initially, we inferred lambda by estimating the $K = 1$ likelihood and allowing lambda to converge. We used this inferred value of lambda in five subsequent replicates with $K = 2$, the same F_{ST} assumed in the two groups, and using the admixture model. The burn-in was set as a period of 50,000 steps followed by 100,000 MCMC iterations. All replicates converged on the same general parameter estimates, and the mean of all replicates was used for reporting assignment of individuals to genetic clusters. Within each of the lineages, STRUCTURE was run for multiple values of K ; no additional genetic clustering was identified.

In addition to STRUCTURE analyses, we examined population genetic structure using Discriminant Analysis of Principal Components (DAPC; Jombart et al. 2010), implemented in the R package adegenet (Jombart & Ahmed 2011). DAPC transforms the SNP data using principal components analysis and then uses discriminant analysis to identify genetic clusters. Because two lineages were again identified, DAPC was run

with the northern and southern lineages separately; here, the most likely number of genetic clusters within each lineage was one (Supp. Info.).

Song and environmental characteristics

During June 2013, song recordings were collected from multiple localities in the contact zone (Table 2), although not enough sampling to perform an extensive quantitative analysis. Here, our goal was to find potentially identifying characteristics between the song of northern and southern lineages. We used the terminology of syllables used by Baptista and Krebs (2000), who studied *C. americana* song dialects in California, USA. The song repertoire of *C. americana* generally consists of five types of syllables (as described by Baptista and Krebs 2000; Fig. 2): (1) srih—rapidly frequency modulated note, (2) note complex—group of notes forming an obvious cohesive unit, (3) sigmoid—note shaped like a sideways letter “S” on the spectrogram, (4) tüt—note that rapidly drops in frequency and then remains sustained, and (5) whistle—a frequency sustained note. We used the Raven Pro v1.4 (Cornell Lab of Ornithology) software to audibly and visually inspect songs. Measurements were taken of the following song attributes: total length, frequency range, number of syllables, and types of syllables. All songs are deposited in the Macaulay Library (Table 2), with the specific song sequences analyzed in Supp. Info.

Because another species (*Sitta carolinensis*) was shown to have genetic structure based on temperature and precipitation extremes in the Madrean Archipelago sky

islands of Arizona, USA (Manthey & Moyle 2015), we obtained environmental data (minimum precipitation of driest month and maximum temperature of warmest month) from worldclim.org (Hijmans et al. 2005) to test this possibility in *C. americana*. We obtained all specimen records of *C. americana* around the putative contact zone (Arizona, USA, Sonora, Mexico, and Chihuahua, Mexico) from VertNet, an online repository of museum-vouchered specimen data. Specimens were restricted to breeding months (late May-early August) to prevent inclusion of migrants. To assess if northern and southern localities (based on genetic results in Fig. 1) significantly differed in environmental characteristics, we performed a multivariate analysis of variance (MANOVA) in R (R Development Core Team 2012).

RESULTS

Number of reads and RAD-tags per individual was highly variable (Table 1). The 50% and 75% coverage matrices had 15,531 and 7,995 loci, with ~2.8 SNPs per locus on average (Table 3). STRUCTURE analyses, for both the 50% and 75% coverage matrices, identified a strong and well-defined split between northern and southern populations (Fig. 1B). All individuals sorted to their respective lineages with greater than 90% probability assignment (Fig. 1B), indicating a lack of hybridization between lineages. However, if hybridization rarely occurs, our limited sampling may not be sufficient to identify hybrids. Within each lineage, the most likely number of genetic clusters was one, suggesting a lack of fine-scale spatial structure. Similar to STRUCTURE analyses,

DAPC identified the same individuals sorting to two genetic clusters and a lack of genetic structure within each of the lineages (Supp. Info.).

The parental populations (Utah and Mexico) had 2296 and 766 fixed differences for the 50% and 75% coverage matrices, respectively. Note that these numbers may be inflated due to small sample sizes. Most SNPs fixed between parental populations were similarly fixed in sampled localities of Arizona (Fig. 1D). In the contact zone specifically, most SNPs were private alleles (SNPs found in only one lineage), with more than 1.5% of SNPs fixed between lineages (Fig. 1C), even between populations as close as 70 km.

The number of loci per chromosome was highly related to chromosome size ($R^2 > 0.98$), suggesting our data were spread evenly across the genome (based on BLAST+ results). When F_{ST} was computed per chromosome across the putative contact zone between northern and southern lineages, genetic differentiation was positively related to chromosome size (Fig. 3A). Additionally, the pattern was similar between contact zone samples and parental samples (Fig. 3B), although the scale of F_{ST} was different between groups, likely due to different sample sizes.

Northern and southern localities near the contact zone differed significantly in environmental characteristics (Fig. 4), likely a reflection of the latitudinal difference between lineage distributions. The geographic break between lineages is at the split between temperate and subtropical coniferous forests (Wade et al. 2003), however, and may represent a real difference in climate regimes inhabited by each lineage. In this region, the northern lineage inhabits forests dominated by Ponderosa Pine (*Pinus ponderosa*); in contrast, the southern lineage populations are in the northernmost

Madrean forest localities, where *Pinus* species are interspersed with Madrean oak species (Genus: *Quercus*). Further south (northern Mexico), the sky island habitat of *C. americana* becomes dominated by mixed oak communities.

No differences between lineages were apparent in song length, frequency range, or song structure (Table 2). All individuals' songs began with the Srih note, ranged between 1.12 and 1.53 seconds in duration, and had similar frequency ranges (3.0 – 8.2 kilohertz). Differences in song mostly existed among sampling localities; all but one locality (Santa Rita Mountains) with multiple individuals had the same beginning and ending syllables (Table 2). In the Santa Rita Mountains, the first four syllables were always the same, with the only difference being whether the song ended with a whistle. These results suggest that each locality has developed a specific local dialect, independent of genetic lineage, although more song recordings are needed to test this.

DISCUSSION

High throughput sequencing to investigate genetic structure and gene flow

The use of high-throughput sequencing has been critical in identifying chromosomal levels of differentiation between species at contact zones (e.g., Ellegren et al. 2012, Taylor et al. 2014), usually identifying higher relative effects (e.g., divergence, lack of introgression, selection) of sex chromosomes compared to autosomes. Studies with thousands of SNPs have been used to provide genomic scale resolution in contact zone

studies compared to traditional studies using mtDNA or microsatellites alone. Large panels of SNPs provide an increasingly robust method to diagnose hybridization or lack thereof. Here, using more than 40,000 SNPs we identified a lack of hybridization and a sharp geographic break between two genetically distinct lineages (Fig. 1). Interestingly, this pattern is in contrast to Marshall's (1956) evidence for a more gradual cline across several mountain ranges based on morphology. This may be caused by multiple factors; we may have not sampled the genetic variation that contributes to morphological variation (e.g., *Sporophila* seedeaters, Campagna et al. 2015). Alternatively, morphological differences could be due to an environment by genotype interaction; here, a cline in environmental conditions (i.e., latitudinal cline) may contribute to a cline in morphological characters.

The observed genetic pattern is not unique to *C. americana*, however. The sky islands between the Rocky Mountains and Sierra Madre Occidental span the transition between temperate and subtropical coniferous forests (Wade et al. 2003). Here, at the southern range limits of the Rocky Mountains, higher elevations are dominated by Ponderosa Pine. This transitions to the northern range limits of the Sierra Madre Occidental, where higher elevations are dominated by mixed pine (Genus *Pinus*) and Madrean oak species (Genus *Quercus*). As such, many montane songbird species (> 50) have their northern or southern range limits in this region. The specific geographic break identified here in *C. americana* is evident in three other bird species, including the southern range limit of the Mountain Chickadee (*Poecile gambeli*) and the northern range limit of two species: Mexican Chickadee (*Poecile sclateri*), and the montane range of the

Eastern Bluebird (*Sialis sialis*). Between *C. americana* lineages, divergence likely occurred during the Pleistocene (Manthey et al. 2011b); however, during the Pleistocene glacial cycles there is little evidence for increased distributional overlap among lineages during glacial maxima based on ecological niche modeling (Manthey et al. 2014), suggesting mechanisms other than climate cycles have restricted gene flow among lineages.

Chromosomal patterns and divergence hypotheses

In the contact zone, we found a positive relationship between genetic differentiation and chromosome size (Fig. 3A). The Z chromosome had the highest F_{ST} between lineages (arrows in Fig. 3A) and one of the narrower cline widths (Fig. 3C), although it did not appear to be an outlier based on relationships with chromosome size. Many studies of hybridizing bird species have found elevated genetic differentiation on the Z chromosome relative to autosomes (e.g., Carling et al. 2010; Storchová et al. 2010) when using intron sequence data and small numbers of loci. Indeed, in this study system (*C. americana*), elevated Z chromosome differentiation has been shown in analyses of introns alone (Manthey & Spellman 2014). This discrepancy suggests that using only intron sequences biases inference of chromosome wide patterns of differentiation. However, a genome-level analysis of *Ficedula* flycatchers still estimated the Z chromosome to be an order of magnitude more differentiated on average than autosomes (Ellegren et al. 2012), suggesting small-scale genetic sequencing may be biased in only some cases.

An alternative hypothesis to explain this pattern may be that only recent speciation events exhibit elevated sex chromosome differentiation. Relative to *Ficedula* flycatchers (Ellegren et al. 2012) and *Passerina* buntings (Carling & Brumfield 2008), the two *Certhia* lineages are more divergent based on mitochondrial and nuclear sequence data (Manthey et al. 2011a,b,2015). Under a time-since-speciation scenario, biased genomic introgression (i.e., less on sex chromosomes) may have ceased much earlier in *Certhia* due to lack of interbreeding between lineages.

A final hypothesis is that sex chromosome evolution is similar to autosomal evolution when genetic mechanisms involved with postzygotic isolation between taxa do not occur [e.g., large-Z effect or Haldane's rule (Coyne & Orr 1989)]. In this scenario, a lack of hybridization prevents increased differentiation on the Z to manifest. Here, prezygotic isolation would have needed to occur throughout the divergence process. Behavioral differences between populations could cause divergence on chromosomes via genetic drift and non-random mating. This possibility may also be the isolation mechanism between *Poecile gambeli* and *P. sclateri*, sister species (Harris et al. 2013) that do not hybridize (McCarthy 2006), occupy the same habitat, and have the identical genetic break shown here in *C. americana*.

Some behavioral characteristics potentially minimizing interbreeding between lineages include non-random gene flow, non-random mating, and natal philopatry. Non-random gene flow may lead to population differentiation with or without the presence of fitness variation (Edelaar & Bolnick 2012). Variation in dispersal ability, individual behavior, and habitat preference genes alone can lead to biased gene flow

and subsequent genetic differentiation through time (Edelaar & Bolnick 2012). Short natal dispersal distance (Davis 1978) and natal philopatry (Cramp & Perrins 1993) in creepers could increase the effects of non-random gene flow. Because northern and southern forms of *C. americana* at least overlap seasonally (Phillips et al. 1964) and there is still a distinct genetic break between lineages (Fig. 1) suggests that behavioral characteristics may minimize gene flow among lineages.

Multiple lines of evidence suggest non-random mating, with or without other sources of non-random gene flow, may be causing the patterns observed in *C. americana*. First, local dialect formation appears to be common, as has been shown in California, USA (Baptista & Krebs 2006) and now in Arizona, USA (Table 2). Song variation has been suggested as a cue for finding locally-adapted mates and thus promoting co-divergence of song and genetics (Slabbekoorn & Smith 2002), but has also been shown to differ widely among populations without speciation events [e.g., White-crowned Sparrows (*Zonotrichia leucophrys*), Marler & Tamura 1962]. Additionally, each lineage is associated with significantly different climates (Fig. 4)—and associated ecosystems—a potential context for non-random dispersal and gene flow.

Lastly, the genomic patterns observed here (Figs. 1 & 3)—namely chromosome-differentiation relationships—are similar between local (i.e., in the contact zone) and highly disjunct (e.g., eastern USA and southern Mexico; Manthey et al. 2015) populations. The same patterns between geographically and ecologically disparate populations suggest selection is not the driving force of differentiation, as it would be unlikely for such widespread populations to be undergoing the same selective pressures

(Manthey et al. 2015). Rather, genetic drift—in conjunction with non-random mating and possibly non-random gene flow—is causing the differential patterns of genetic differentiation observed between lineages (Fig. 3). Genetic drift alone could cause divergence patterns correlated to chromosome size because of differential average recombination; recombination rates scale with chromosome size due to meiotic crossover requirements (Lynch 2007).

CONCLUSIONS

We sequenced thousands of SNPs to identify the geographic limits of divergent *C. americana* lineages, determine the amount of hybridization—or lack thereof—between lineages, and identify differential patterns of genetic differentiation between chromosomes. We found that the two lineages are completely allopatric during the breeding season, with the geographic break in southern Arizona, USA. Based on our sampling, there does not appear to be gene flow between the two lineages. Genetic differentiation in the contact zone is higher on larger chromosomes. The two lineages differ in climate and ecosystems and exhibit local dialects throughout southern Arizona. Genetic drift, possibly with some combination of non-random mating and non-random gene flow, is likely causing the differential chromosomal patterns of genetic differentiation because of variance in recombination rates among chromosomes.

ACKNOWLEDGEMENTS

We would like to thank Mark Geiger for assistance with specimen collection. We would like to thank the state and federal agencies and wildlife officers for their cooperation and help obtaining permits that contributed to this research (U.S. Department of Fish and Wildlife and the Arizona Game and Fish Department). We would like to thank John Klicka and Sharon Birks at the University of Washington Burke Museum for tissue loans. We thank John Kelly and Patrick Monnahan for lab work advice and materials. This work was funded through an American Museum of Natural History Frank M. Chapman Memorial Fund Grant, an American Ornithologists' Union Student Research Grant, an NSF Doctoral Dissertation Improvement Grant (DEB-1406989), NSF DEB 1241181, and through the NSF IGERT C-CHANGE Program #0801522 at the University of Kansas. We thank the use of facilities at the COBRE Genome Sequencing Core Laboratory at the University of Kansas, funded through NIH award number P20GM103638.

DATA ACCESSIBILITY

All raw data is deposited in the NCBI Sequence Read Archive accessioned under BioProject #: . All song files are uploaded to Macaulay Library (see Table 2).

AUTHOR CONTRIBUTIONS

JDM and MBR collected specimens. MBR recorded song data. JDM collected genomic data and analyzed all data. JDM and RGM designed the research and all authors worked on completion of the manuscript.

REFERENCES

- Atwood, TC., Young, J.K., Beckmann, J.P., Breck, S.W., Fike, J., Rhodes, O.E., Bristow, K.D. 2011. Modeling connectivity of black bears in a desert sky island archipelago. *Biological Conservation*, 144: 2851-2862.
- Baptista, L.F. and Krebs, R. 2000. Vocalizations and relationships of Brown Creepers *Certhia americana*: a taxonomic mystery. *Ibis*, 142: 457-465.
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. 2009. BLAST+: architecture and applications. *BMC Bioinformatics*, 10: 421.

- Campagna, L., Gronau, I., Silveira, L.F., Siepel, A., Lovette, I.J. 2015. Distinguishing noise from signal in patterns of genomic divergence in a highly polymorphic avian radiation. *Molecular Ecology*, 24: 4238-4251.
- Carling, M.D. and Brumfield, R.T. 2008. Integrating phylogenetic and population genetic analyses of multiple loci to test species divergence hypotheses in *Passerina* buntings. *Genetics*, 178: 363-377.
- Carling, M.D., Lovette, I.J., and Brumfield, R.T. 2010. Historical divergence and gene flow: coalescent analyses of mitochondrial, autosomal and sex-linked loci in *Passerina* buntings. *Evolution*, 64: 1762-1772.
- Catchen, J.M., Hohenlohe, P.A., Bassham, S., Amores, A., and Cresko, W.A. 2013. Stacks: an analysis tool for population genomics. *Molecular Ecology*, 22: 3124-3140.
- Cramp, S. and Perrins, C.M. 1993. The Birds of the Western Palearctic: Flycatchers to Shrikes, vol. 7. Oxford University Press, Oxford, UK.
- Coyne, J.A. and Orr, H.A. 1989. Two rules of speciation. In *Speciation and its consequences* (eds. Otte, D. & Endler, J.), Sinauer, Sunderland, MA, USA.
- Davis, C.M. 1978. A nesting study of the Brown Creeper. *Living Bird*, 17: 237-263.
- Edelaar, P. and Bolnick, D.I. 2012. Non-random gene flow: an underappreciated force in evolution and ecology. *Trends in Ecology and Evolution*, 27: 659-665.
- Elgvin, T.O., Hermansen, J.S., Fijarczyk, A., Bonnet, T., Borge, T., Saether, S.A. et al. 2011. Hybrid speciation in sparrows II: a role for sex chromosomes? *Molecular Ecology*, 20: 3823-3837.

- Ellegren, H., Smeds, L., Burri, R., Olason, P.I., Backström, N., Kawakami, T. et al. 2012. The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature*, 491: 756-760.
- Fitak, R.R., Koprowski, J.L., Culver, M. 2013. Severe reduction in genetic variation in a montane isolate: the endangered Mount Graham red squirrel (*Tamiasciurus hudsonicus grahamensis*). *Conservation Genetics*, 14: 1233-1241.
- Gardner, J.P.A. and Wei, K.J. 2014. The genetic architecture of hybridisation between two lineages of greenshell mussels. *Heredity*, 114: 344-355.
- Harris, R.B., Carling, M.D., and Lovette, I.J. 2014. The influence of sampling design on species tree inference: a new relationship for the new world chickadees (Aves: *Poecile*). *Evolution*, 68: 501-513.
- Hersch-Green, E.I., Allan, G.J., and Whitham, T.G. 2014. Genetic analysis of admixture and patterns of introgression in foundation cottonwood trees (Salicaceae) in southwestern Colorado, USA. *Tree Genetics and Genomes*, 10: 527-539.
- Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., and Jarvis, A. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25: 1965-1975.
- Janoušek, V., Wang, L., Luzynski, L., Dufková, P., Vyskočilová, 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*. *Molecular Ecology*, 21: 3032-3047.

- Jombart, T., Ahmed, I. 2011. Adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*, 27: 3070-3071.
- Jombart, T., Devillard S., Balloux, F. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, 11: 94.
- Kawakami, T., Smeds, L., Backström, N., Husby, A., Qvarnström, A., Mugal, C.F., et al. 2014. A high-density linkage map enables a second-generation collared flycatcher genome assembly and reveals the patterns of avian recombination rate variation and chromosomal evolution. *Molecular Ecology*, 23: 4035–4058.
- Lavretsky, P., DaCosta, J.M., Hernández-Baños, B.E., Engilis, A., Sorenson, M.D., Peters, J.L. 2015. Speciation genomics and a role for the Z chromosome in the early stages of divergence between Mexican ducks and mallards. *Molecular Ecology*.
- Lynch, M. 2007. *The Origins of Genome Architecture*. Sinauer Associates, Sunderland, Massachusetts.
- Malek, T.B., Boughman, J.W., Dworkin, I., and Peichel, C.L. 2012. Admixture mapping of male nuptial colour and body shape in a recently formed hybrid population of threespine stickleback. *Molecular Ecology*, 21: 5265-5279.
- Manthey, J.D., Klicka, J., and Spellman, G.M. 2011a. Cryptic diversity in a widespread North American songbird: phylogeography of the Brown Creeper (*Certhia americana*). *Molecular Phylogenetics and Evolution*, 58: 502-512.

- Manthey, J.D., Klicka, J., Spellman, G.M. 2011b. Isolation-driven divergence: speciation in a widespread North American songbird (Aves: Certhiidae). *Molecular Ecology*, 20: 4371-4384.
- Manthey, J.D., Klicka, J., and Spellman, G.M. 2015. Chromosomal patterns of diversity and differentiation in creepers: a next-gen phylogeographic investigation of *Certhia americana*. *Heredity*, 115: 165-172.
- Manthey, J.D. and Moyle, R.G. 2015. Isolation by environment in White-breasted Nuthatches (*Sitta carolinensis*) of the Madrean Archipelago sky islands: a landscape genomics approach. *Molecular Ecology*, 24: 3628-3638.
- Manthey, J.D. and Spellman G.M. 2014. Increased differentiation and reduced gene flow in sex chromosomes relative to autosomes between lineages of the brown creeper *Certhia americana*. *Journal of Avian Biology*, 45: 149-156.
- Manthey, J.D., Klicka, J., Spellman, G.M. 2014. Effects of climate change on the evolution of Brown Creeper (*Certhia americana*) lineages. *The Auk*, 131: 559-570.
- Marler, P. and Tamura, M. 1962. Song "dialects" in three populations of White-crowned Sparrows. *Condor*, 64: 368-377.
- Marshall, J.T. 1956. Summer birds of the Rincon Mountains, Saguaro National Monument, Arizona. *Condor*, 58: 81-97.
- Masta, S.E. 2000. Phylogeography of the jumping spider *Habronattus pugillis* (Araneae: Salticidae): recent vicariance of sky island populations? *Evolution*, 54: 1699-1711.
- Mayr, E. 1942. Systematics and the origin of species from the viewpoint of a zoologist. Harvard University Press, Cambridge, MA, USA.

- McCarthy, E.M. 2006. Handbook of avian hybrids of the world. Oxford University Press, New York, NY, USA.
- Miller, M.R., Dunham, J.P., Amores, A., Cresko, W.A., and Johnson, E.A. 2007. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Research*, 17: 240-248.
- Phillips, A.J., Marshall, J., Monson, G. 1964. The Birds of Arizona. University of Arizona Press, Tucson, AZ.
- Potter, K.M., Hipkins, V.D., Mahalovich, M.F., Mean, R.E. 2013. Mitochondrial DNA haplotype distribution patterns in *Pinus ponderosa* (Pinaceae): range-wide evolutionary history and implications for conservation. *American Journal of Botany*, 100: 1562-1579.
- Pritchard, J.K., Stephens, M., and Donnelly, P. 2000. Inference of population structure using multilocus genotype data. *Genetics*, 155: 945–959.
- R Development Core Team. 2012. R: A Language for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
- Schild, D.R., Card, D.C., Adams, R.H., Jezkova, T., Reyes-Velcasco, J., Proctor, F.N., et al. 2015. Incipient speciation with biased gene flow between two lineages of the Western Diamondback Rattlesnake (*Crotalus atrox*). *Molecular Phylogenetics and Evolution*, 83: 213-223.
- Slabbekoorn, H. and Smith, T.B. 2002. Bird song, ecology, and speciation. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, 357: 493-503.

- Smith, C.I., Farrell, B.D. 2005. Phylogeography of the longhorn cactus beetle *Moneilema appressum* LeConte (Coleoptera: Cerambycidae): was the differentiation of the Madrean sky islands driven by Pleistocene climate changes? *Molecular Ecology*, 14: 3049-3065.
- Storchová, R., Reif, J., and Nachman, M.W. 2010. Female heterogamety and speciation: reduced introgression of the Z chromosome between two species of nightingales. *Evolution*, 64: 456-471.
- Taylor, S.A., Curry, R.L., White, T.A., Ferretti, V., and Lovette, I. 2014. Spatiotemporally consistent genomic signatures of reproductive isolation in a moving hybrid zone. *Evolution*, 68: 3066-3081.
- Wade, T.G., Riitters, K.H., Wickham, J.D., and Jones, K.B. 2003. Distribution and causes of global forest fragmentation. *Conservation Ecology*, 7.
- Warren, W.C., Clayton, D.F., Ellegren, H., Arnold, A.P., Hillier, L.W., Künstner, A., et al. 2010. The genome of a songbird. *Nature*, 464: 757-762.

Table 1. List of samples and their associated RAD-seq coverage. For each individual, the locality (as in Fig. 1), collection number [either from University of Washington Burke Museum (UWBM) or University of Kansas Natural History Museum (all others)], percent completeness in the 50% and 75% coverage matrices (C50 and C75, respectively), number of sequencing reads (# Reads), number of RAD-tags, and coverage of sequencing reads for the 50% coverage matrix dataset (RAD Cov.; mean and standard deviation).

Locality	Coll. Number	C50	C75	# Reads	RAD-tags	RAD Cov.
Chiricahua	MBR8595	89.3	95.7	2,193,670	48,677	48.5 (51.1)
Chiricahua	MBR8596	86.3	93.2	1,874,272	47,317	40.7 (41.4)
Chiricahua	JDM057	91.8	96.5	3,102,638	50,786	62.8 (71.8)
Chiricahua	MBR8601	88.2	95.9	1,614,960	34,682	40.6 (44.5)
Chiricahua	KU31226	87.6	94.2	2,468,865	52,090	14.7 (11.5)
Huachuca	MBR8586	89.6	95.7	2,475,852	52,235	52.0 (53.1)
Huachuca	MBR8584	89.2	95.7	2,110,794	40,465	48.9 (51.7)
Huachuca	MBR8585	86.5	92.6	3,582,197	45,483	73.1 (84.6)
Huachuca	KU31227	91.8	97.3	2,344,278	38,860	56.4 (62.4)
Jalisco	UWBM110712	88.1	93.5	3,912,641	65,526	73.9 (84.5)
Jalisco	UWBM110615	92.7	97.7	1,656,810	42,467	39.0 (36.5)
Jalisco	UWBM110621	87.7	93.0	4,114,168	55,404	77.1 (86.2)
Jalisco	UWBM117091	91.3	97.5	2,883,402	39,628	69.5 (78.2)
Santa Ritas	JDM063	88.3	95.2	1,206,553	34,904	30.6 (29.7)
Santa Ritas	JDM064	89.4	95.6	1,831,829	42,147	41.4 (41.7)
Santa Ritas	MBR8609	89.6	95.9	2,997,954	48,142	64.7 (68.6)

Santa Ritas	JDM065	91.5	97.1	1,738,408	42,413	42.1 (40.3)
Kaibab North Rim	JDM002	86.2	93.2	2,037,171	50,198	42.8 (43.4)
Kaibab North Rim	JDM009	91.3	97.4	2,179,302	46,187	50.9 (52.3)
Kaibab North Rim	JDM010	91.1	96.1	3,039,178	52,035	67.6 (64.6)
Kaibab South Rim	JDM061	91.7	97.4	2,467,519	39,978	59.7 (64.5)
Kaibab South Rim	JDM062	90.7	95.7	1,456,471	40,073	35.3 (32.0)
Mogollon Rim	JDM012	90.5	96.1	3,599,547	48,270	79.8 (86.4)
Mogollon Rim	JDM013	92.7	97.9	2,947,298	40,714	71.2 (77.1)
Mogollon Rim	JDM014	89.1	95.6	2,897,455	56,549	55.4 (58.5)
Mogollon Rim	JDM015	79.9	90.5	863,742	26,774	26.4 (24.7)
Mogollon Rim	JDM016	45.6	54.2	155,422	14,710	7.6 (4.3)
Mogollon Rim	JDM017	70.1	82.1	414,123	22,121	14.7 (11.5)
Mogollon Rim	JDM018	63.1	74.9	291,260	20,417	10.6 (7.4)
Mogollon Rim	JDM043	69.1	79.7	879,162	44,515	19.5 (16.3)
Mogollon Rim	JDM045	51.6	57.9	212,617	18,040	7.0 (3.9)
Pinal	MBR8580	90.7	96.6	2,677,802	43,901	59.7 (64.5)
Pinal	MBR8579	91.7	96.9	3,098,164	42,654	72.2 (79.4)
Pinal	MBR8604	89.2	95.3	2,505,339	59,776	50.7 (50.2)
Pinaleño	JDM033	79.6	89.9	686,733	33,435	17.6 (14.6)
Pinaleño	JDM034	91.6	97.0	1,366,277	46,048	32.5 (28.8)
Pinaleño	JDM035	54.1	63.9	650,917	27,576	19.6 (17.4)
Pinaleño	JDM036	89.1	95.8	978,434	31,435	26.7 (27.0)
Pinaleño	JDM037	56.6	67.4	250,007	18,556	10.1 (6.8)
Prescott	MBR8578	89.1	94.7	3,920,836	58,548	78.3 (87.1)

Prescott	JDM051	94.0	98.7	3,718,143	40,543	90.8 (102.8)
Prescott	JDM055	90.4	97.0	1,346,722	34,398	33.8 (35.4)
Prescott	MBR8602	93.1	98.0	1,902,484	53,109	42.6 (38.8)
Prescott	MBR8603	89.0	94.8	4,669,850	53,519	96.8 (104.4)
Santa Catalina	JDM027	41.1	47.4	122,107	14,648	5.6 (2.5)
Santa Catalina	JDM028	91.5	96.8	2,755,533	45,462	64.6 (65.0)
Santa Catalina	JDM029	70.0	82.7	524,465	22,115	19.7 (16.7)
Santa Catalina	JDM030	93.2	97.8	1,735,413	39,986	42.1 (42.7)
Santa Catalina	JDM032	53.5	64.6	349,605	26,160	10.1 (6.6)
Utah	UWBM114838	68.5	81.5	460,367	36,076	9.8 (6.4)
Utah	UWBM111063	92.6	96.8	2,916,596	42,087	67.5 (72.0)
Utah	UWBM114811	86.8	94.7	644,668	34,290	16.4 (12.6)
Utah	UWBM111062	79.9	92.2	687,820	28,984	18.9 (17.3)
Outgroup - England	KU6761	79.1	84.3	2,085,339	49,129	37.4 (41.3)

Table 2. Song repertoires of *C. americana* near the contact zone. Based on limited sample sizes, songs did not noticeably differ in length or frequency range between lineages and varied in the number of syllables from three to five. All syllable types are as in Figure 2. In the Santa Rita Mountains, birds appear to have a unique syllable, where the end of a sigmoid is rapidly frequency modulated (similar to a srih syllable). Recording number refers to Macaulay Library reference number.

Locality	#	1	2	3	4	5
Prescott	202873	Srih	Long Sigmoid	Note Complex	-	-
Prescott	202882	Srih	Long Sigmoid	Note Complex	-	-
Prescott	202884	Srih	Sigmoid	Whistle	Note Complex	-
Prescott	202888	Srih	Whistle	Whistle	Note Complex	-
Santa Catalina	203222	Srih	Sigmoid	Sigmoid	Whistle	-
Santa Catalina	203223	Srih	Sigmoid	Sigmoid	Whistle	-
Santa Rita	203232	Srih	Srih-Sigmoid	Sigmoid	Sigmoid	Whistle
Santa Rita	203233	Srih	Srih-Sigmoid	Sigmoid	Sigmoid	Whistle
Santa Rita	203234	Srih	Srih-Sigmoid	Sigmoid	Sigmoid	-
Huachuca	203253	Srih	Sigmoid	Tüt	Tüt	Tüt
Pinaleño	203292	Srih	Whistle	Long Sigmoid	Whistle	-
Pinaleño	203298	Srih	Whistle	Long Sigmoid	Whistle	-
Mog. Rim (East)	203628	Srih	Sigmoid	Whistle	Note Complex	-

Table 3. Statistics for the 50% and 75% coverage matrices (CM). Number of loci, single nucleotide polymorphisms (SNPs), SNPs per locus (PL; mean and standard deviation), SNPs per locus in the contact zone (PL CZ), and percent missing data in each dataset (Miss.).

Dataset	Loci	SNPs	SNPs PL	SNPs PL CZ	Miss.
50% CM	15,531	44,400	2.86 (1.59)	2.65 (1.50)	17.11%
75% CM	7,995	22,700	2.84 (1.53)	2.62 (1.44)	10.08%

Draft

Figure 1. (A) Sampling map in Arizona, USA. Gray areas correspond to montane forest. Inset shows locations of parental populations (solid circles) used in this study. (B) STRUCTURE results for the 75% coverage matrix (75% CM) dataset. Each bar represents the probability of population assignment to northern (gray) or southern (white) lineages. All individuals sorted with population assignment values greater than 0.9. (C) Proportion of shared (white), private (gray), and fixed (black) polymorphisms in Arizona between the two lineages [based on genetic structure in part (B)]. (D) Proportion of SNPs fixed between parental populations at each locality in Arizona. Black and white indicate proportion of fixed SNPs for the northern and southern alleles, respectively. Gray indicates the population is polymorphic with both northern and southern alleles.

Figure 2. Definition of syllables used in song analysis. Song vocabulary follows Baptista and Krebs 2000. Frequency of each note is not on the same scale (i.e., the Sigmoid is not at an innately higher frequency than a Srih).

Figure 3. (A) Relationship of genetic differentiation (F_{ST}) between lineages in the contact zone and chromosome size for the 50% and 75% coverage SNP matrices (CM). (B) Comparison of chromosomal patterns of genetic differentiation between lineages in the contact zone and in parental populations. Arrows indicate the Z chromosome.

Figure 4. Scatter plot of environmental variables (minimum precipitation of driest month and maximum temperature of warmest month) for both lineages and results of a multivariate analysis of variance (MANOVA) between lineages of these two environmental variables.

Draft

Frequency

Note
Complex

Sigmoid

Srih

Tüt

Whistle

Time

A F_{ST} Contact Zone0.1
0.2
0.3
0.4

1.0

1.5

2.0

log Chromosome Size

Genome

Page 36 of 51

● 50% CM

 $R^2 = 0.728$ $p < 0.001$

○ 75% CM

 $R^2 = 0.606$ $p < 0.001$ **B** F_{ST} Parentals0.1
0.2
0.3
0.4

0.05

0.15

0.25

0.35

 F_{ST} Contact Zone

● 50% CM

 $R^2 = 0.891$ $p < 0.001$

○ 75% CM

 $R^2 = 0.703$ $p < 0.001$

MANOVA

Pillai trace = 0.645

Approx. $f = 91.76$

$p < 0.001$

Supplementary Information:

DAPC Results. Bayesian Information Criterion plot indicating the most likely number of genetic clusters for the 50% and 75% coverage datasets, on the left and right respectively. The lowest value (in both cases 2) indicates the assumed best number of genetic clusters. Below these, the assignment of the individuals from each locality to genetic cluster 1 or 2.

Screen captures of songs. All full recordings are available via Macaulay Library (#s in Table 2). The screen captures indicate the specific song sequence analyzed for each individual.

Prescott 63

Prescott 81

Prescott 83

Prescott 91

Santa Catalina 15

Santa Catalina 17

Santa Rita 42

Santa Rita 45

Santa Rita 47

Huachuca 89

Pinalenos 149

Pinalenos 160

Mogollon Rim east 50