Hidden histories of gene flow in highland birds revealed with genomic markers

EUGENIA ZARZA,* BRANT C. FAIRCLOTH,† WHITNEY L.E. TSAI,* ROBERT W. BRYSON JR.,*‡
JOHN KLICKA‡ and JOHN E. MCCORMACK*

*Moore Laboratory of Zoology, Occidental College, Los Angeles, CA 90041, USA, †Department of Biological Sciences and Museum of Natural Science, Louisiana State University, Baton Rouge, LA 70803, USA, ‡Burke Museum of Natural History and Culture and Department of Biology, University of Washington, Seattle, WA 98195, USA

Abstract

Genomic studies are revealing that divergence and speciation are marked by gene flow, but it is not clear whether gene flow has played a prominent role during the generation of biodiversity in species-rich regions of the world where vicariance is assumed to be the principal mode by which new species form. We revisit a well-studied organismal system in the Mexican Highlands, *Aphelocoma* jays, to test for gene flow among Mexican sierras. Prior results from mitochondrial DNA (mtDNA) largely conformed to the standard model of allopatric divergence, although there was also evidence for more obscure histories of gene flow in a small sample of nuclear markers. We tested for these 'hidden histories' using genomic markers known as ultraconserved elements (UCEs) in concert with phylogenies, clustering algorithms and newer introgression tests specifically designed to detect ancient gene flow (e.g. ABBA/BABA tests). Results based on 4303 UCE loci and 2500 informative SNPs are consistent with varying degrees of gene flow among highland areas. In some cases, gene flow has been extensive and recent (although perhaps not ongoing today), whereas in other cases there is only a trace signature of ancient gene flow among species that diverged as long as 5 million years ago. These results show how a species complex thought to be a model for vicariance can reveal a more reticulate history when a broader portion of the genome is queried. As more organisms are studied with genomic data, we predict that speciation-with-bouts-of-gene-flow will turn out to be a common mode of speciation.

Keywords: ancient introgression, divergence-with-gene-flow, introgression, mito-nuclear discordance

Received 12 May 2016; revision received 10 August 2016; accepted 12 August 2016

Introduction

Genomic studies of species formation are increasingly revealing that divergence and speciation are marked by gene flow (Rheindt & Edwards 2011; Fontaine *et al.* 2015; Mallet *et al.* 2015; Payseur & Rieseberg 2016). This new appreciation for the role of gene flow in diversification departs from a more standard view of lineages splitting through clean, bifurcating histories that sunder through vicariance, never to cross paths again. It is no coincidence that our appreciation for the role of gene

flow during diversification has grown with our ability to collect more genomic data (Abbott *et al.* 2016). High-throughput sequencing now allows for broad genomic investigation, even at the level of whole genomes, making it possible to reconstruct species histories in situations where gene histories disagree (Jarvis *et al.* 2014), and specifically in cases of reticulate evolution resulting from gene flow (Yu *et al.* 2014). Whole-genome studies of recent speciation between sister species show that gene flow can be ongoing through the process of species formation (Ellegren *et al.* 2011; Osborne *et al.* 2016), with fixed species differences sometimes occurring in only a fraction of the genome affecting aspects of the phenotype (Poelstra *et al.* 2014).

Correspondence: John E. McCormack,
E-mail: mccormack@oxy.edu

While these recent studies have heightened our awareness of the role of gene flow in speciation, less clear is whether gene flow plays a prominent role in the generation of biodiversity in species-rich regions of the world where allopatry and vicariance are assumed to be the principal modes by which new species are formed (Haffer 1969). In the Mexican Highlands, for example, diversity and endemism have often been attributed solely to vicariance caused by dynamic geological activity, mountain formation and habitat fragmentation (Halffter 1987; Sullivan *et al.* 2000; Castoe *et al.* 2009; Bryson *et al.* 2011; McCormack *et al.* 2011). Alternative models, however, posit a more active role for recent and ancient gene flow in species evolution (Velo-Antón *et al.* 2013; Mastretta-Yanes *et al.* 2015). In considering alternative models, we draw a subtle, but important contrast between the classic divergence-with-gene-flow model of speciation (Nosil 2008; Pinho & Hey 2010), which implies continuous gene flow and focuses on the role of local adaptation in divergence, and what we call divergence-with-bouts-of-gene-flow, in which initial vicariance is crucial, gene flow is inconsistent through time (possibly involving nonsister taxa), and the result is a messy divergence process in which selection may or may not play a prominent role.

We revisit a well-studied system, jays in the genus *Aphelocoma*, to test for patterns of recent and ancient gene flow among lineages in the Mexican Highlands. We focus on four lineages in different mountain ranges (Fig. 1). One of these lineages is considered a separate species, the Transvolcanic Jay (*Aphelocoma ultramarina*), and is sister to the other three, which together comprise allopatric lineages of the Mexican Jay (*Aphelocoma wollweberi*). These lineages diverged between 1 and 5 million years ago (McCormack *et al.* 2011). Prior conclusions from studies relying mostly on mitochondrial DNA (mtDNA) conformed to the standard model of allopatric divergence and speciation, and showed little evidence for reticulate processes or gene flow among hundreds of individuals sampled throughout Mexico (McCormack *et al.* 2008b, 2011).

Several lines of evidence suggest, however, that despite the bifurcating and monophyletic patterns observed in mtDNA, there is a more obscure history of gene flow among the four lineages. First, despite no mtDNA haplotype sharing, results from 14 microsatellites (McCormack *et al.* 2008b) indicated a broad cline of nuclear gene flow between two of the *A. wollweberi* lineages in the Sierra Madre Oriental (East and Central; Fig. 1). Second, a prior phenotypic study of thousands of museum specimens suggested contact zones within *A. wollweberi* and between *A. wollweberi* and *A. ultramarina* (Pitelka 1951; Peterson 1991). Finally, although microsatellites assigned *A. wollweberi* and *A. ultramarina*
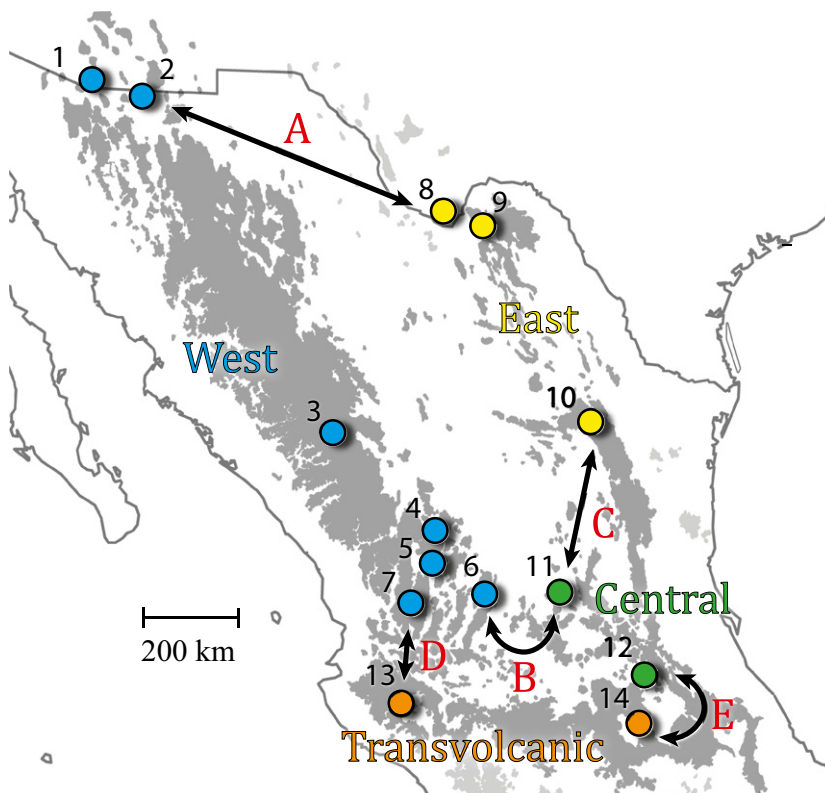

Fig. 1 Map of northern Mexico and the southwestern United States showing sampling locations for three lineages of Mexican Jays (*Aphelocoma wollweberi*) and Transvolcanic Jays (*Aphelocoma ultramarina*). Site numbers correspond to Table 1. Red letters denote gene flow hypotheses, with corresponding results from ABBA/BABA introgression tests found in Table 2.

perfectly to different genetic clusters, detailed inspection on a locus-by-locus basis revealed a distinctive macromutation in one microsatellite locus whose geographic distribution suggested ancient gene flow between the two species (McCormack & Venkatraman 2013).

Our goal was to test for patterns of recent and ancient gene flow among the four lineages of *Aphelocoma* jays using a broad genomic subsample. We collected data from thousands of nuclear genomic regions using sequence capture of ultraconserved elements (UCEs). We were also able to assemble nearly complete mtDNA genomes from off-target reads. We compared the topology of mtDNA and nuclear trees to assess whether there was well-supported conflict between them, which often reflects gene flow (Carling & Brumfield 2008; Toews & Brelsford 2012; Peters *et al.* 2014) or gene flow coupled with selection (Ribeiro *et al.* 2011; Pavlova *et al.* 2013). After finding disagreement between the mtDNA and nuclear trees, we used single nucleotide polymorphisms (SNPs) to conduct population assignment tests, visualize discordant histories on a species tree, and test geographic hypothesis for gene flow with the ABBA/BABA method (Green *et al.* 2010; Durand *et al.* 2011).

Methods

Sampling and DNA extraction

We sampled 20 *Aphelocoma wollweberi* and four *Aphelocoma ultramarina*, using blood or tissue from 14 localities chosen to best represent the breadth of their distribution in the Mexican Highlands (Fig. 1 and Table 1). We complemented these samples with two samples from the closely related Western Scrub-Jay (*Aphelocoma californica*), which we used as outgroups for both the phylogenies and the ABBA/BABA tests. We extracted genomic DNA from tissue or blood using a Qiagen DNeasy Blood and Tissue extraction kit. We visualized extractions on an agarose gel to ensure fragments were >200 bp, and we quantified the resulting double-stranded DNA using a Qubit fluorometer (Life Technologies, Inc.).

Sequence capture and next-generation sequencing

To collect a representative subsample of the genome, we used sequence capture of ultraconserved elements (Bejerano *et al.* 2004). This method uses 180-bp synthetic baits to capture DNA data from thousands of orthologous loci identified from alignments of the chicken and *Anolis* genomes (Faircloth *et al.* 2012). Following capture, sequencing and assembly, variable base positions are mostly found in the regions flanking the UCEs. This

method has proven useful for solving phylogenetic questions at deep timescales (McCormack *et al.* 2013), as well as more recent questions at phylogeographic scales (Smith *et al.* 2014; McCormack *et al.* 2015; Manthey *et al.* 2016).

We followed the protocol for library preparation and UCE enrichment from Faircloth *et al.* (2012) which is available at ultraconserved.org. Briefly, we sheared 100 μL of DNA at a 20 ng/μL concentration to a size distribution with a peak between 400 and 600 bp using a Bioruptor ultrasonicator (Diagenode). We prepared genomic libraries for each sample for the Illumina platform with a KAPA LTP library preparation kit, attaching custom indexing tags (Faircloth & Glenn 2012) to DNA from each sample to allow pooling during the capture phase. We enriched three pools of eight samples each for 5060 UCE loci using a set of synthetic RNA probes (MYbaits_Tetrapods-UCE-5K kit; Mycroarray). After enrichment and recovery PCR, we verified that library size range was between 400 and 600 bp using a Bioanalyzer (Agilent Technologies). We quantified the enriched pools using qPCR and combined them in equimolar ratios prior to 250-bp paired-end sequencing on one run of an Illumina MiSeq.

Bioinformatic processing, assembly and alignment of UCEs

We followed the PHYLUCE (Faircloth 2015) pipeline to perform quality control on the demultiplexed reads, as well as to assemble UCE loci from the reads, and to align UCE loci for phylogenetic analysis. As part of this pipeline, we trimmed reads of adapter contamination and low-quality bases with Illumiprocessor (Faircloth 2013) and Trimmomatic (Bolger *et al.* 2014). We assembled reads into contigs using the PHYLUCE script assemblo_abyss, in concert with ABySS (Simpson *et al.* 2009), using kmer=51. We identified assembled contigs matching UCE loci by aligning each contig to the original UCE probe sequences with the PHYLUCE script match_contigs_to_probes, which uses LASTZ (Harris 2007). Finally, we aligned UCE loci using MAFFT v7.13 (Katoh & Standley 2013) by running the PHYLUCE script seqcap_align_2, using the default trimming algorithm.

Mitochondrial genome assembly

Mitochondrial DNA reads are abundant in off-target sequences resulting from sequence capture experiments (Picardi & Pesole 2012; do Amaral *et al.* 2015). We identified and assembled mtDNA genomes from off-target, trimmed Illumina reads. The complete mtDNA genome of the Rook (*Corvus frugilegus*) served as a reference (GenBank accession NC_002069) because this was the

Table 1 Specimens and genomic summary statistics

Map site	Museum	Catalogue No.	Species	Current subspecies taxonomy	Lineage	Locality	Latitude	Longitude	QC reads	UCEs	Avg. locus length	mtDNA reads	Avg. mtDNA coverage
1	FMNH	334259	*Aphelocoma wollweberi*	*arizonae*	West	Peña Blanca Lake, Arizona, USA	31.41	−111.13	2 720 552	4383	566	20 003	242
1	FMNH	334261	*A. wollweberi*	*arizonae*	West	Peña Blanca Lake, Arizona, USA	31.41	−111.13	3 275 672	4349	537	22 136	265
1	FMNH	334264	*A. wollweberi*	*arizonae*	West	Peña Blanca Lake, Arizona, USA	31.41	−111.13	2 653 668	4354	575	22 719	273
2	FMNH	343516	*A. wollweberi*	*arizonae*	West	Cananea, Sonora, Mex.	31.25	−110.06	2 758 592	4352	548	4426	55
3	CTR	02N9800	*A. wollweberi*	*wollweberi*	West	San Antonio, Durango, Mex.	25.02	−105.72	3 462 708	4326	604	281	6
3	CTR	02N9802	*A. wollweberi*	*wollweberi*	West	San Antonio, Durango, Mex.	25.02	−105.72	3 421 064	4364	624	307	6
4	MZFC	Conacyt340	*A. wollweberi*	*wollweberi*	West	Gachupines, Zacatecas, Mex.	23.16	−102.88	2 302 648	4350	571	13 004	153
4	MZFC	Conacyt360	*A. wollweberi*	*wollweberi*	West	Gachupines, Zacatecas, Mex.	23.16	−102.88	2 607 732	4415	621	22 339	267
5	FMNH	343575	*A. wollweberi*	*wollweberi*	West	Valparaiso, Zacatecas, Mex.	22.85	−103.63	3 803 632	4191	602	4625	60
7	FMNH	343520	*A. wollweberi*	*gracilis*	West	Villa Guerrero, Jalisco, Mex.	21.98	−103.66	2 681 656	4407	581	2822	37
6	FMNH	343570	*A. wollweberi*	*wollweberi*	West	El Jaguey, Jalisco, Mex.	22.2	−102.6	4 401 672	4183	526	6208	76
6	FMNH	393742	*A. wollweberi*	*wollweberi*	West	El Jaguey, Jalisco, Mex.	22.2	−102.6	2 691 228	4418	632	4685	59
11	FMNH	343552	*A. wollweberi*	*potosina*	Central	Bledos, San Luis Potosí, Mex.	21.86	−101.15	3 375 992	4246	506	4930	61
11	FMNH	393663	*A. wollweberi*	*potosina*	Central	Bledos, San Luis Potosí, Mex.	21.86	−101.15	3 019 600	4209	474	2987	37
12	FMNH	393655	*A. wollweberi*	*potosina*	Central	Jacala, San Luis Potosí, Mex.	21.05	−99	3 251 848	4189	506	7518	91
12	FMNH	393662	*A. wollweberi*	*potosina*	Central	Jacala, San Luis Potosí, Mex.	21.05	−99	3 947 868	4209	608	2704	36
10	FMNH	343534	*A. wollweberi*	*couchii*	East	El Diamante Pass, Coahuila, Mex.	25.36	−100.86	2 174 676	4344	504	2938	36
10	FMNH	393651	*A. wollweberi*	*couchii*	East	El Diamante Pass, Coahuila, Mex.	25.36	−100.86	3 151 220	4342	588	6003	76
9	CTR	04J0128	*A. wollweberi*	*couchii*	East	Vacas Canyon, Coahuila, Mex.	28.95	−102.61	2 113 732	4353	566	160	5

Table 1 *Continued*

Map site	Museum	Catalogue No.	Species	Current subspecies taxonomy	Lineage	Locality	Latitude	Longitude	QC reads	UCEs	Avg. locus length	mtDNA reads	Avg. mtDNA coverage
8	CTR	05J0230	A. wollweberi	couchii	East	Chisos Basin, Texas, USA	29.24	−103.29	3 444 528	4349	601	341	7
14	FMNH	393674	A. ultramarina	ultramarina	Transvolcanic	Huitzilac, Morelos, Mex.	19.05	−99.26	2 584 848	4366	573	24159	298
14	FMNH	393675	A. ultramarina	ultramarina	Transvolcanic	Huitzilac, Morelos, Mex.	19.05	−99.26	1 669 548	4090	357	2625	28
13	FMNH	343585	A. ultramarina	colimae	Transvolcanic	Sayula, Jalisco, Mex.	19.78	−103.66	3 023 124	4178	469	11 231	129
13	FMNH	343586	A. ultramarina	colimae	Transvolcanic	Sayula, Jalisco, Mex.	19.78	−103.66	2 920 344	4316	555	5016	64
Outgroup	FMNH	334087	Aphelocoma californica	nevadae		Mt. Charleston, Nevada, USA	36.37	−115.63	—	—	—	—	—
Outgroup	FMNH	334000	A. californica	superciliosa		Gardnerville, Nevada, USA	38.83	−119.62	—	—	—	—	—

FMNH, Field Museum of Natural History; CTR, UCLA Center for Tropical Research Blood and Feather Collection; MZFC, Colección Ornitológica del Museo de Zoología 'Alfonso L. Herrera', UNAM, México.

Table 2 Tests of introgression hypotheses from Fig. 1

Hypothesis	P1	P2	P3	n ABBA	n BABA	D	Z	P
A	*02N9800* West	*343516* West	*05J0230* East	277	334	−0.09	−2.26	0.98
A	*02N9800* West	*334264* West	*04J0128* East	268	364	−0.13	−3.04	0.99
B	*334264* West	*393742* West	*393662* Central	484	287	0.26	7.04	<0.001
B	*02N9800* West	*343570* West	*343552* Central	418	289	0.18	4.79	<0.001
C	*334264* West	*393662* Central	*343534* East	598	204	0.49	15.34	<0.001
C	*334264* West	*343552* Central	*343534* East	514	203	0.43	12.41	<0.001
D	*02N9800* West	*343575* West	*343586* Trans	295	252	0.08	1.8	0.07
D	*334264* West	*conacyt360* West	*343586* Trans	310	273	0.06	1.46	0.14
E	*343534* East	*393662* Central	*393674* Trans	252	159	0.23	4.54	<0.001
E	*343534* East	*393655* Central	*393674* Trans	216	176	0.1	1.96	0.12
NULL	*02N9800* West	*393742* West	*conacyt360* West	418	405	0.02	0.44	0.66
NULL	*02N9800* West	*343570* West	*conacyt360* West	375	419	−0.06	−1.51	0.13

Outgroup (O4) for all tests was a Western Scrub-Jay (*Aphelocoma californica*); hypotheses correspond to geneflow scenarios on the map in Fig. 1; nABBA, number of ABBA patterns; nBABA, number of BABA patterns.

closest related corvid mtDNA genome available. We used MITOBIM v1.7 (Hahn *et al.* 2013), a Perl wrapper for MIRA 4.0.2 (Chevreux *et al.* 1999) that takes a baiting and iterative mapping approach for assembly. We annotated the assembled mtDNA regions de novo with the MITOchondrial genome annotation Server (MITOS) (Bernt *et al.* 2013). To facilitate alignment, we selected and aligned only the 13 protein-coding mtDNA regions with MAFFT and built a partitioned, concatenated matrix.

Phylogenies from concatenated data

We created a concatenated data matrix of UCE loci including outgroups. We allowed some missing data, dropping a UCE locus only if it was absent from more than 25% of the samples (i.e. more than six individuals). We analysed this unpartitioned, concatenated data matrix with RAXML v8.0.19 (Stamatakis 2014) using the GTRGAMMA model of evolution by searching for the best-scoring maximum-likelihood (ML) tree, performing 1000 bootstrap searches and reconciling the best ML tree with the bootstrap replicates. We followed the same ML tree search and bootstrapping strategy to analyse a concatenated matrix composed of protein-coding mtDNA loci partitioned according to locus.

Calling and analysing SNPs

In addition to tree-based approaches, we also wanted to employ methods capable of identifying individuals and populations with evidence for introgression, which is facilitated by SNP data. We called SNPs using the assembled UCE contigs of the ingroup individual with the highest number of recovered UCEs as a reference. We indexed and mapped the trimmed Illumina reads to the reference assembly with the BWA-MEM algorithm,

which is suitable for reads ranging from 70 bp to 1 Mbp (Li 2013). After sorting reads with SAMTOOLS (Li *et al.* 2009), we removed PCR duplicates from the resulting BAM files with Picard (http://broadinstitute.github.io/picard/) and used GATK 3.2 (McKenna *et al.* 2010) to realign the mapped reads around indels, to call variants, to quality-filter, to remove indels and to select only biallelic variants. We further manipulated the resulting SNP calls with VCFTOOLS (Danecek *et al.* 2011) to eliminate all positions with missing data, to select one SNP every 900 bp (when loci were longer than 900 bp) and to convert variants to number format. We generated a species tree from this SNP matrix in SNAPP 1.1.10 (Bryant *et al.* 2012), implemented in BEAST 2.2.1 (Bouckaert *et al.* 2014). We ran two instances of SNAPP for two million generations using default priors, saving the output every 1000 steps. We combined tree and parameter files from both runs with LOGCOMBINER 2.1.3. We displayed the full set of likely species trees with DENSITREE v2.2.1 (Bouckaert 2010), which is expected to show fuzziness in parts of the tree that conflict due to gene flow or other causes.

We also analysed the SNP data in STRUCTURE v2.3.4 (Pritchard *et al.* 2000) to assign individuals to a predetermined number of genetic groups using their multilocus genetic profiles. We ran simulations under the admixture model to detect signals of gene flow among lineages. We assumed an admixture model and correlated allele frequencies. We ran the data assuming different numbers of genetic clusters ($K = 2$ to $K = 8$) for one million generations and five replicates. These preliminary analyses suggested that $K < 6$ was much more likely than $K > 6$ scenarios. This led us to increase our computing effort, focusing on $K = 2$ to $K = 6$, increasing the number of generations to 10 million with 10 replicates for each value of K.

Introgression tests

To detect potential past hybridization, we used the ABBA/BABA test (Green *et al.* 2010; Durand *et al.* 2011). In a phylogeny with topology (((P1,P2))P3)O4), where an allele B arises via mutation from allele A in the common ancestor of P1, P2 and P3 (where O4 is the outgroup), it is expected – in the absence of other evolutionary forces – that stochastic lineage sorting will produce equal numbers of (((A,B))B)A) and (((B,A))B)A) patterns in the descendent lineages. An excess of ABBA patterns is expected when there has been gene flow between nonsister lineages P2 and P3, provided that P1 and P3 are not also exchanging genes. Such asymmetries in the frequency of shared, derived variants are detectable using Patterson's *D*-statistic (Green *et al.* 2010; Durand *et al.* 2011). The ABBA/BABA test has been applied to SNP data derived from whole genomes (Green *et al.* 2010), RADseq protocols (Eaton & Ree 2013; Rheindt *et al.* 2014; Streicher *et al.* 2014) and exon capture data (Heliconius Genome Consortium 2012), but thus far not to SNPs mined from UCEs.

We performed 12 comparisons to test different hypotheses for gene flow (letters in Fig. 1). For each comparison, we assumed a tree topology (((P1,P2))P3)O4), where P represents ingroup taxa and O the outgroup, in this case a Western Scrub-Jay (FMNH 334087). The ABBA/BABA method tests for evidence of gene flow between P3 and either P1 or P2 (traditionally P2 is designated as the lineage hypothesized to be exchanging genes with P3). We used the mtDNA phylogeny as a guide to identify P1, P2 and P3, making an assumption that this topology reflects the original history of divergence per McCormack *et al.* (2011), with later gene flow resulting from secondary contact. Thus, we made sure P1 and P2 were more closely related to one another than either was to P3. We chose P2 to be the lineage hypothesized to have gene flow with P3, which, if true, would result in more ABBA patterns relative to BABA patterns as indicated by a positive Patterson's *D*-statistic (Green *et al.* 2010; Durand *et al.* 2011). We tested each of the six hypotheses depicted in Fig. 1 using two different sets of individuals to ensure that results were not driven by outlier data at the individual level. We tested two null hypotheses where P1, P2 and P3 were all drawn from the same lineage, expecting no differences in ABBA vs. BABA patterns.

The *D*-statistic was originally applied to continuous genomewide data (Green *et al.* 2010). To compute the standard error, Green *et al.* (2010) took a block jackknife approach where the resampling block size was larger than the extent of linkage disequilibrium. Our data set comprises UCEs located on average 188 150 ± 12 485 bp apart in the chicken genome (Faircloth *et al.* 2012).

Given this distance, it is possible that jackknifing among UCE loci will overcome the extent of linkage disequilibrium. Because we mapped against a reduced reference genome that we created by concatenating UCE loci, we used a block size of 500 bp (a number near the average size of UCE loci, see Results) to resample among loci. Additionally, to test the effect of block size on the outcome of each test, we resampled with 200 and 1000 bp block sizes.

To isolate SNPs for use in the ABBA/BABA tests, we mapped the trimmed Illumina reads of each sample to the outgroup with BWA. After removing PCR duplicates from the mapped BAM files with Picard, we followed the GATK pipeline as above and added a base recalibration step. Given that there is not a SNP database available for these species, we followed the base recalibration procedure suggested by the GATK best practices for nonmodel organisms. This consists of performing an initial round of SNP calling on the original uncalibrated data, selecting the SNPs with the highest confidence (a minimum emission and call quality of 40) and using these as the database of known SNPs. We executed four rounds of base recalibration on the original data. We used the recalibrated, mapped read files as input to the program ANGSD v0.70 (Korneliussen *et al.* 2014), which makes use of genotype likelihood estimates rather than SNP calls, counts ABBA/BABA variants from the mapped data, and estimates the *D*-statistic. For each individual, we only considered bases with a minimum map score and quality of 30. We executed the jackknife procedures with the jackknife.R script included in ANGSD to obtain standard deviations and Z-scores for the *D*-statistic. We then converted Z-scores to their equivalent *P*-values.

Results

UCE and mtDNA trees from concatenated data

On average, we obtained around 1.5 million reads per sample (range: 834 774 to 2 200 836 reads) after initial quality control and adapter trimming (Table 1). We assembled an average of 336 000 contigs per sample with an N50 between 586 and 764 bp. We recovered an average of 4303 UCE contigs per sample (Table 1). A total of 4188 contigs were present in at least 20 ingroup and outgroup individuals, forming an aligned matrix of 2 205 383 bp.

We assembled the mtDNA genomes of 20 of the individuals from between 2625 and 44 213 reads at an average read depth of 152×. For the four samples that were derived from blood (02N9800 and 02N9802 from site 3, and 05J0230 and 04J0128 from sites 8 and 9), we obtained a low number of mtDNA reads per sample

(range: 160–341), resulting in low read depth and many ambiguous base calls. Despite the low number of mtDNA reads, coverage of nuclear UCE loci enriched from these blood samples was high and on a par with tissue-derived samples. Thus, we removed these four individuals from the mtDNA analysis, but retained them in the UCE analysis. The final, aligned mtDNA matrix contained 11 352 bp comprising 13 concatenated, protein-coding regions.

The ML phylogeny of the mtDNA genomes (Fig. 2B) showed the same topology as that described in previous studies using one or two mtDNA regions (McCormack *et al.* 2008b, 2011). There was a deep split between Transvolcanic Jays and Mexican Jays. Within Transvolcanic Jays, the subspecies in the east (*Aphelocoma ultramarina ultramarina*) and west (*A. ultramarina colimae*) were reciprocally monophyletic. Within Mexican Jays, the East, Central and West groups were each monophyletic, with the East group sister to the Central + West groups. Within the West group, samples from the north (*A. w. arizonae*) and south (*A. w. wollweberi*) in the Sierra Madre Occidental showed reciprocal monophyly. As in previous studies, two divergent mtDNA clades were present even within *A. w. arizonae* (McCormack *et al.* 2008a).

The ML phylogeny from concatenated UCE data (Fig. 2A) differed from the mtDNA phylogeny by showing less evidence for monophyly of the previously described clades. Where both phylogenies agreed was in the deep split between Mexican and Transvolcanic Jays, the reciprocal monophyly of the two groups within Transvolcanic Jays, and the monophyly of the East lineage of Mexican Jays. However, the UCE tree did not place the East group of Mexican Jays sister to all other Mexican Jay groups, but rather recovered the East group as sister only to the Central group, which itself was not monophyletic. The West group was also not monophyletic, contrasting with the mtDNA tree. Rather, some West individuals appeared more closely related to the Central group. The central Sierra Madre Occidental population within the West lineage (site 3 in Fig. 1) also showed conflicting phylogenetic affinities, with the UCE tree supporting a sister relationship with northern populations and the mtDNA tree supporting a sister relationship with southern populations. The long terminal tips in the UCE phylogeny are likely an artefact of singletons, possibly caused by sequencing error, that are interpreted by the phylogenetic program as derived variants for terminal taxa.

SNPs, STRUCTURE and SNAPP results

We used FMNH 393742, which produced 4418 UCE loci, as our reference for mapping and SNP calling. We mapped between 22 and 31% of the trimmed reads to
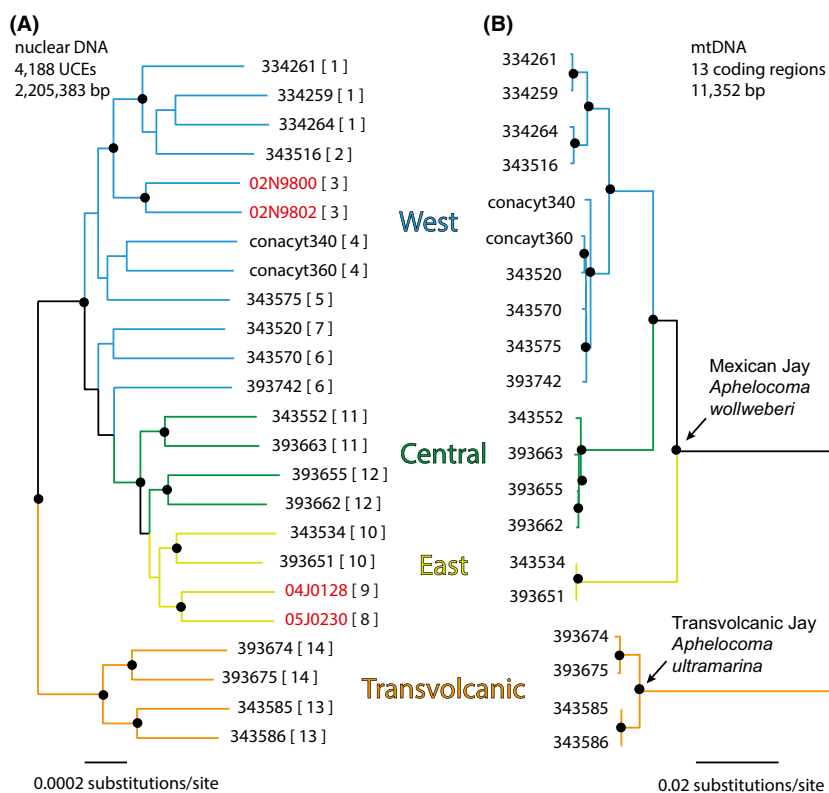

Fig. 2 Maximum-likelihood trees for (A) UCE data and (B) mtDNA data showing the evolutionary relationships of individuals of *Aphelocoma wollweberi* and *Aphelocoma ultramarina*. Tip labels are museum numbers followed by site numbers from Fig. 1. Nodes with bootstrap support >75% are marked with a black dot. Four individuals (noted in red font) were dropped from the mtDNA analysis due to low coverage.

the reference for SNP calling. The rest of the reads were off-target sequences. On average, we detected and removed 5.7% of reads per sample as PCR duplicates. After SNP calling and variant filtration with GATK, we obtained a total of 8870 biallelic SNPs, with an average of 2.2 SNPs per UCE locus. After eliminating all positions with missing data and selecting one SNP every 900 bp, 2501 SNPs remained, each with an average read depth of 25×.

The SNAPP species tree obtained from 2501 SNPs was similar to the UCE tree generated from concatenated data, but the cloudogram revealed more about the conflict within the genetic data for those individuals that differed in their placement between the mtDNA and UCE trees (Fig. 3A). For example, certain West individuals appear to have conflicting genetic signal across loci and affinities to both the West and Central groups in the SNAPP tree (black arrow in Fig. 3A). The distribution of posterior trees was approximately evenly split between these two evolutionary scenarios. In contrast, there was no observable conflict in the SNAPP tree for two other cases where the mtDNA and UCE trees differed. A sister relationship between the East and Central groups was strongly supported by the SNAPP tree. Likewise, individuals from site 3 within the West group were unanimously supported as being part of the northern lineage.

Results from STRUCTURE were largely consistent with the SNAPP species tree, but provided a more nuanced understanding of the genetic affinities of individuals (Fig. 3B). STRUCTURE runs of $K = 4$ have the highest likelihood and identified genetic clusters corresponding to the Transvolcanic Jay and three groups within Mexican Jays. These three groups did not conform perfectly to the three geographic groups West, East and Central. Instead, STRUCTURE detected two genetic groups within the West (north and south) and did not split the East and Central groups. West individuals (site 6) that appeared to share ancestry with the Central group in the UCE and SNAPP trees showed mixed assignment in the STRUCTURE plots. Within the West group, individuals from site 3 showed signs of mixed assignment between north and south lineages.

Introgression tests

Results of the ABBA/BABA tests (Table 2) also suggest introgression between lineages based on 630 SNPs with patterns appropriate for testing. The null test where all individuals were drawn from the same lineage produced results with negligible deviations from the expected 50/50 ratio of ABBA/BABA patterns and the lowest Z-scores of any tests. The closest test to a positive control (Hypothesis C) was between the East and
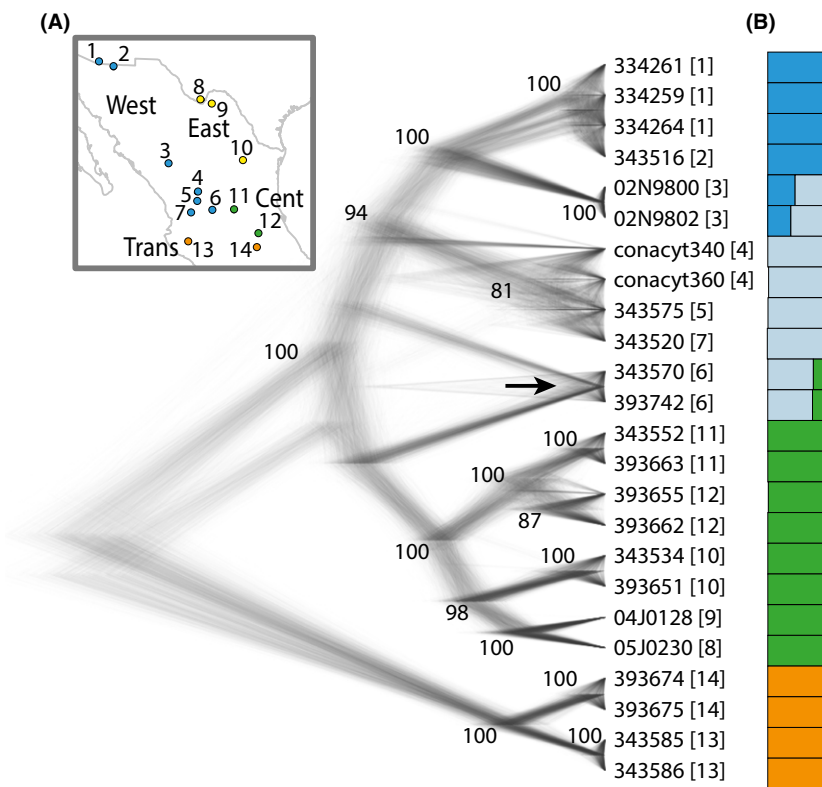

Fig. 3 (A) SNAPP tree with map inset. Bootstrap support is shown for nodes >75%. Tip labels are museum numbers followed by site numbers relating to map inset. The black arrows denote two individuals at site 6 with evolutionary affinities to both West and Central lineages. (B) STRUCTURE plot generated using 2501 informative SNPs gathered from UCE loci. Colours represent assignment to four genetic clusters: West (north), dark blue; West (south), light blue; Central + East, green; Transvolcanic, orange.

Central lineages, where several lines of evidence point to introgression. This comparison produced the highest Z-scores. We observed the next highest Z-scores between the West and Central lineages (Hypothesis B), where the STRUCTURE plots and SNAPP trees suggested introgression. In the two tests for introgression between the most divergent lineages, Mexican and Transvolcanic Jays (Hypotheses D and E), we detected slightly elevated ABBA patterns, but they were only significant in one of four tests. Finally, in the case of Hypothesis A, where gene flow has never been hypothesized, there was an excess of BABA patterns (P-value > 0.95), which not only suggests little gene flow between the tested lineages, but that there was relatively more gene flow between the nonfocal lineages.

Discussion

Our current understanding of species limits, taxonomy and the speciation process is based largely on estimates of genetic patterns and processes inferred from one or a few markers. This is changing. New studies that sequence more individuals, use more markers and include markers from a broader portion of the genome are revealing complex portraits of speciation that have previously been overlooked (Cui *et al.* 2013; Mallet *et al.* 2015; Li *et al.* 2016; Wen *et al.* 2016). *Aphelocoma* jays have been a case study for speciation (Pitelka 1951; Peterson 1992; McCormack *et al.* 2008b), niche evolution (Rice *et al.* 2003; McCormack *et al.* 2010) and the role of isolation and Earth history events like glacial cycles and mountain formation in the generation of biodiversity (McCormack *et al.* 2011). Most conclusions from these studies were based on mtDNA data, where monophyletic gene trees and neatly bifurcating phylogenies lent themselves to an interpretation of species formation focused exclusively on vicariance. Where nuclear data have supplemented mtDNA, hints of gene flow have emerged that were not apparent in mtDNA alone (McCormack *et al.* 2008b; McCormack & Venkatraman 2013). In this study, by sampling broadly from the genome, a more complex story of speciation in *Aphelocoma* jays emerges, one where recent (although perhaps not current) gene flow among mountain ranges has played an important role, and one where the footprint of more ancient gene flow is present, although scarcely detectable.

The biggest difference between the story recorded in the mtDNA genome and that recorded in the nuclear genome is among the relationships of the three major genetic lineages of the Mexican Jay, *A. wollweberi*, which we here call East, West and Central. The mtDNA phylogeny we built from all protein-coding genes agrees with prior work (McCormack *et al.* 2008b). This

phylogeny shows East sister to, and quite divergent from, West+Central (Fig. 2B). The UCE data, meanwhile, strongly suggest introgression between the East and Central lineages along the Sierra Madre Oriental. East and Central are sister lineages in the concatenated tree (Fig. 2) and the species tree (Fig. 3A). Introgression tests between East and Central (Hypothesis C) produced the highest Z-scores of any of the tested gene-flow scenarios (Table 2). In a sense, this is the closest scenario we have to a positive control for the introgression tests, as prior work using microsatellites (McCormack *et al.* 2008b) and phenotype (Pitelka 1951) also both indicate gene flow between East and Central.

Given the strong evidence for nuclear introgression between East and Central, it is striking that there is no evidence for mtDNA mixing. As noted before, East and Central are monophyletic in the mtDNA phylogeny and are not even sister lineages. In a prior study, there was no evidence for a single mtDNA haplotype shared among 48 Central and 270 East individuals sampled (McCormack *et al.* 2008b). What does this lack of mtDNA mixing say about whether gene flow between East and Central is ongoing today? For one, it suggests there is little long-distance female dispersal. Otherwise, one would expect to find the occasional mismatched haplotype of a female that dispersed from one geographic area to the other. Rather, our results suggest that if introgression occurs today, it must happen across a very narrow cline, similar to the two hybridizing Western Scrub-Jay lineages that meet in the mountains of western Nevada (Gowen *et al.* 2014). In this scenario, mtDNA introgression is restricted by the effects of Haldane's rule, in which hybrid females are more likely to be sterile and therefore mtDNA is less likely to cross lineage boundaries (Carling & Brumfield 2008; Gowen *et al.* 2014). Another possibility is that gene flow might have stopped sufficiently long ago that introgressed mtDNA haplotypes were lost from populations due to genetic drift, while introgressed nuclear alleles persisted due to their larger effective population size.

While gene flow between East and Central was expected based on prior research, our observation of relatively high admixture between West and Central was unexpected. Neither the prior microsatellite analysis nor phenotypic work suggested a high degree of mixing between the West and Central lineages (McCormack *et al.* 2008b). However, the species tree, assignment tests and introgression tests all agree that the two individuals from site 6 share ancestry between West and Central. These same individuals were included in the microsatellite study of McCormack *et al.* (2008b), where they grouped with other West individuals. The mtDNA haplotypes of individuals from site 6 were either shared with other West individuals or were most

closely related to other West haplotypes. What these results show is that genomic data can help detect gene flow where even prior nuclear data did not. In this case, we increased the amount of nuclear data from 13 microsatellite loci to thousands of UCE loci and UCE-derived SNPs. Retrospectively, we discovered that a connection between West and Central had been hypothesized over 100 years ago based on careful phenotypic study (Nelson 1899). There is a certain degree of irony in the idea that painstaking phenotypic studies carried out long before the discovery that DNA is the hereditary material might, in fact, provide a wealth of hypotheses regarding the subtle influence of introgression on the appearance of populations. That phenotype is, of course, likely to be under the control of numerous nuclear genetic markers whose introgression occurs in a similar manner to the markers we queried.

Turning from more recent divergence among lineages within the Mexican Jay to the oldest split between Mexican Jays and Transvolcanic Jays, our results show a weak, but consistent signal of introgression. According to the phylogenies we inferred from both nuclear and mtDNA, there is no evidence for gene flow. The two species are reciprocally monophyletic and separated by long branches. Yet, the two introgression tests between the West lineage of Mexican Jays and a nearby Transvolcanic Jay population (Hypothesis D) both showed elevated ABBA patterns (although not statistically significant). Between the Central lineage of Mexican Jays and a nearby Transvolcanic population (Hypothesis E), ABBA patterns were also elevated, and one of the two tests was significant. The significant test amounted to an excess of 100 ABBA SNPs. For comparison, in the tests between East and Central, where gene flow was apparent from multiple lines of evidence, ABBA SNPs were elevated by almost three times as much.

If the above can be taken as suggestive evidence for gene flow between Mexican Jays and Transvolcanic Jays, that gene flow appears to be ancient based on reciprocal monophyly in the phylogenies and clear differences in genetic clustering. Ancient gene flow was the same conclusion drawn by a study that examined a macromutation in a single microsatellite locus (McCormack & Venkatraman 2013). Mexican Jays usually have very short alleles at this microsatellite locus, whereas Transvolcanic Jays were nearly fixed for alleles almost four times as long. The only place where Mexican Jays had long alleles was near the range border with Transvolcanic Jays, and there was a decreasing clinal pattern of the long allele moving away from the range border. However, even Mexican Jays with two long 'Transvolcanic' copies of the allele did not differ from other Mexican Jays in appearance or in their mtDNA type. This scenario mirrors, in some respects, the situation with modern humans and Neanderthals, which show no trace of introgression in their phenotypes or mtDNA, only in a small percentage of their nuclear genomes (Green *et al.* 2010).

Ultraconserved elements have been used previously to infer evolutionary history at both deep and shallow timescales (Faircloth *et al.* 2012; Smith *et al.* 2014; Manthey *et al.* 2016). However, the basic idea of enriching reads from around highly conserved anchor regions often leads to the question of whether enough variability will be sequenced from the flanking DNA, especially when these data are used to investigate questions at recent timescales. This problem can be compounded when analyses are limited to those that use individual gene trees because some UCE loci lack informative variation. This can produce nonbifurcating gene trees and add noise that can mislead certain types of analysis (O'Neill *et al.* 2013; Xi *et al.* 2015; Manthey *et al.* 2016; Meiklejohn *et al.* 2016).

One way to alleviate this perceived 'gene-tree problem' is to expand the phylogenetic toolkit to include methods that use individual gene trees as well as those that use concatenated data sets. Using gene-tree methods alongside concatenation has, in fact, always been a part of the UCE analytical pipeline, despite recent misrepresentations (Gatesy & Springer 2014). Another way to alleviate the 'gene-tree problem' is to gather the informative SNPs from UCE loci and use these variant sites in appropriate analyses (McCormack *et al.* 2015; Manthey *et al.* 2016). In this study, we were able to gather 2501 informative SNPs from our UCE loci for a species-tree analysis and an average of 630 SNPs with appropriate patterns for ABBA/BABA introgression tests. While far below the number typically culled from whole genomes for such analyses, this number is on a par with the number of SNPs generated by other reduced-representation methods. For example, Manthey *et al.* (2016) used RADseq to generate 2960 SNPs for species-tree analysis, and Rheindt *et al.* (2014) used RADseq to generate 954 SNPs for species-tree analysis and 889 SNPs for ABBA/BABA tests.

This study shows how a species complex previously thought to be a model for isolation and vicariance can reveal a more reticulate history when a broader portion of the genome is queried. What remains to be seen is whether 'hidden histories' will reveal themselves in other vertebrate cases as well, especially species with even lower dispersal, like reptiles and amphibians, which show strong phylogenetic structure across the Mexican Highlands in mtDNA (Bryson *et al.* 2011, 2012; Rovito *et al.* 2013). Our results reinforce the growing view that mtDNA gene trees, while useful in many ways, do not accurately represent the rich histories of

species and populations, which often involve gene flow between sister and nonsister lineages. As more speciation scenarios are studied with genomic data, we expect that the only thing we will discover that is especially unique about the complex evolutionary story involving, for example, humans and Neanderthals is that it was the first to be studied carefully with genomes. Speciation-with-bouts-of-gene-flow might turn out to be the most common mode of speciation.

Acknowledgements

References

Abbott RJ, Barton NH, Good JM (2016) Genomics of hybridization and its evolutionary consequences. *Molecular Ecology*, **25**, 2325–2332.

do Amaral FR, Neves LG, Resende Jr MF *et al.* (2015) Ultraconserved elements sequencing as a low-cost source of complete mitochondrial genomes and microsatellite markers in non-model amniotes. *PLoS ONE*, **10**, e0138446.

Bejerano G, Pheasant M, Makunin I *et al.* (2004) Ultraconserved elements in the human genome. *Science*, **304**, 1321–1325.

Bernt M, Donath A, Jühling F *et al.* (2013) MITOS: improved de novo metazoan mitochondrial genome annotation. *Molecular Phylogenetics and Evolution*, **69**, 313–319.

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.

Bouckaert RR (2010) DensiTree: making sense of sets of phylogenetic trees. *Bioinformatics*, **26**, 1372–1373.

Bouckaert R, Heled J, Kühnert D *et al.* (2014) BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, **10**, e1003537.

Bryant D, Bouckaert R, Felsenstein J, Rosenberg NA, RoyChoudhury A (2012) Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. *Molecular Biology and Evolution*, **29**, 1917–1932.

Bryson RW, Murphy RW, Lathrop A, Lazcano-Villareal D (2011) Evolutionary drivers of phylogeographical diversity in the highlands of Mexico: a case study of the *Crotalus triseriatus* species group of montane rattlesnakes. *Journal of Biogeography*, **38**, 697–710.

Bryson RW, García-Vázquez UO, Riddle BR (2012) Diversification in the Mexican horned lizard *Phrynosoma orbiculare* across a dynamic landscape. *Molecular Phylogenetics and Evolution*, **62**, 87–96.

Carling MD, Brumfield RT (2008) Haldane's rule in an avian system: using cline theory and divergence population genetics to test for differential introgression of mitochondrial, autosomal, and sex-linked loci across the *Passerina* bunting hybrid zone. *Evolution*, **62**, 2600–2615.

Castoe TA, Daza JM, Smith EN *et al.* (2009) Comparative phylogeography of pitvipers suggests a consensus of ancient Middle American highland biogeography. *Journal of Biogeography*, **36**, 88–103.

Chevreux B, Wetter T, Suhai S (1999) Genome sequence assembly using trace signals and additional sequence information. *Proceedings of the German Conference on Bioinformatics (GCB)*, **99**, 45–56.

Cui R, Schumer M, Kruesi K, Walter R, Andolfatto P, Rosenthal GG (2013) Phylogenomics reveals extensive reticulate evolution in Xiphophorus fishes. *Evolution*, **67**, 2166–2179.

Danecek P, Auton A, Abecasis G *et al.* (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156–2158.

Durand EY, Patterson N, Reich D, Slatkin M (2011) Testing for ancient admixture between closely related populations. *Molecular Biology and Evolution*, **28**, 2239–2252.

Eaton DA, Ree RH (2013) Inferring phylogeny and introgression using RADseq data: an example from flowering plants (Pedicularis: Orobanchaceae). *Systematic Biology*, **62**, 689–706.

Ellegren H, Smeds L, Burri R *et al.* (2011) The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature*, **491**, 756–760.

Faircloth BC (2013) Illumiprocessor: a Trimmomatic wrapper for parallel adapter and quality trimming. Available from https://github.com/faircloth-lab/illumiprocessor/.

Faircloth BC (2015) PHYLUCE is a software package for the analysis of conserved genomic loci. *Bioinformatics*, **32**, 786–788.

Faircloth BC, Glenn TC (2012) Not all sequence tags are created equal: designing and validating sequence identification tags robust to indels. *PLoS ONE*, **7**, e42543.

Faircloth BC, McCormack JE, Crawford NG, Harvey MG, Brumfield RT, Glenn TC (2012) Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Systematic Biology*, **61**, 717–726.

Fontaine MC, Pease JB, Steele A *et al.* (2015) Extensive introgression in a malaria vector species complex revealed by phylogenomics. *Science*, **347**, 1258524.

Gatesy J, Springer MS (2014) Phylogenetic analysis at deep timescales: unreliable gene trees, bypassed hidden support, and the coalescence/concatalescence conundrum. *Molecular Phylogenetics and Evolution*, **80**, 231–266.

Gowen FC, Maley JM, Cicero C *et al.* (2014) Speciation in Western Scrub-Jays, Haldane's rule, and genetic clines in secondary contact. *BMC Evolutionary Biology*, **14**, 135.

Green RE, Krause J, Briggs AW *et al.* (2010) A draft sequence of the Neandertal genome. *Science*, **328**, 710–722.

Haffer J (1969) Speciation in Amazonian forest birds. *Science*, **165**, 131–137.

Hahn C, Bachmann L, Chevreux B (2013) Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. *Nucleic Acids Research*, **41**, e129.

Halffter G (1987) Biogeography of the montane entomofauna of Mexico and Central America. *Annual Review of Entomology*, **32**, 95–114.

Harris RS (2007) *Improved Pairwise Alignment of Genomic DNA*. PhD Dissertation, State University, Pennsylvania, University Park.

Heliconius Genome Consortium (2012) Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature*, **487**, 94–98.

Jarvis ED, Mirarab S, Aberer AJ *et al.* (2014) Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science*, **346**, 1320–1331.

Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, **30**, 772–780.

Korneliussen TS, Albrechtsen A, Nielsen R (2014) ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics*, **15**, 356.

Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997*.

Li H, Handsaker B, Wysoker A *et al.* (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.

Li G, Davis B, Eizirik E, Murphy W (2016) Phylogenomic evidence for ancient hybridization in the genomes of living cats (Felidae). *Genome Research*, **26**, 1–11.

Mallet J, Besansky N, Hahn M (2015) How reticulated are species? *BioEssays*, **38**, 140–149.

Manthey JD, Campillo LC, Burns KJ, Moyle RG (2016) Comparison of target-capture and restriction-site associated DNA sequencing for phylogenomics: a test in cardinalid tanagers (Aves, Genus: *Piranga*). *Systematic Biology*, **65**, 640–650.

Mastretta-Yanes A, Moreno-Letelier A, Piñero D, Jorgensen TH, Emerson BC (2015) Biodiversity in the Mexican highlands and the interaction of geology, geography and climate within the Trans-Mexican Volcanic Belt. *Journal of Biogeography*, **42**, 1586–1600.

McCormack JE, Venkatraman MX (2013) A distinctive genetic footprint of ancient hybridization. *The Auk*, **130**, 469–475.

McCormack JE, Bowen BS, Smith TB (2008a) Integrating paleoecology and genetics of bird populations in two sky island archipelagos. *BMC Biology*, **6**, 28.

McCormack JE, Peterson AT, Bonaccorso E, Smith TB (2008b) Speciation in the highlands of Mexico: genetic and phenotypic divergence in the Mexican jay (*Aphelocoma ultramarina*). *Molecular Ecology*, **17**, 2505–2521.

McCormack JE, Zellmer AJ, Knowles LL (2010) Does niche divergence accompany allopatric divergence in *Aphelocoma* jays as predicted under ecological speciation?: insights from tests with niche models. *Evolution*, **64**, 1231–1244.

McCormack JE, Heled J, Delaney KS, Peterson AT, Knowles LL (2011) Calibrating divergence times on species trees versus gene trees: implications for speciation history of *Aphelocoma* jays. *Evolution*, **65**, 184–202.

McCormack JE, Harvey MG, Faircloth BC, Crawford NG, Glenn TC, Brumfield RT (2013) A phylogeny of birds based on over 1,500 loci collected by target enrichment and high-throughput sequencing. *PLoS ONE*, **8**, e54848.

McCormack J, Tsai WLE, Faircloth BC (2015) Sequence capture of ultraconserved elements from bird museum specimens. *Molecular Ecology Resources*, **16**, 1189–1203.

McKenna A, Hanna M, Banks E *et al.* (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, **20**, 1297–1303.

Meiklejohn KA, Faircloth BC, Glenn TC, Kimball RT, Braun EL (2016) Analysis of a rapid evolutionary radiation using ultra-conserved elements: evidence for a bias in some multispecies coalescent methods. *Systematic Biology*, **65**, 612–627.

Nelson EW (1899) Descriptions of new birds from Mexico. *The Auk*, **16**, 25–31.

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

O'Neill EM, Schwartz R, Bullock CT *et al.* (2013) Parallel tagged amplicon sequencing reveals major lineages and phylogenetic structure in the North American tiger salamander (*Ambystoma tigrinum*) species complex. *Molecular Ecology*, **22**, 111–129.

Osborne OG, Chapman MA, Nevado B, Filatov DA (2016) Maintenance of species boundaries despite ongoing gene flow in ragworts. *Genome Biology and Evolution*, **8**, 1038–1047.

Pavlova A, Amos JN, Joseph L *et al.* (2013) Perched at the mito-nuclear crossroads: divergent mitochondrial lineages correlate with environment in the face of ongoing nuclear gene flow in an australian bird. *Evolution*, **67**, 3412–3428.

Payseur BA, Rieseberg LH (2016) A genomic perspective on hybridization and speciation. *Molecular Ecology*, **25**, 2337–2360.

Peters JL, Winker K, Millam KC *et al.* (2014) Mito-nuclear discord in six congeneric lineages of Holarctic ducks (genus *Anas*). *Molecular Ecology*, **23**, 2961–2974.

Peterson AT (1991) Geographic variation in the ontogeny of beak coloration of gray-breasted jays (*Aphelocoma ultramarina*). *The Condor*, **93**, 448–452.

Peterson AT (1992) Phylogeny and rates of molecular evolution in the *Aphelocoma* jays (Corvidae). *The Auk*, **109**, 133–147.

Picardi E, Pesole G (2012) Mitochondrial genomes gleaned from human whole-exome sequencing. *Nature Methods*, **9**, 523–524.

Pinho C, Hey J (2010) Divergence with gene flow: models and data. *Annual Review of Ecology, Evolution and Systematics*, **41**, 215–230.

Pitelka FA (1951) *Speciation and Ecologic Distribution in American Jays of the Genus Aphelocoma*. University of California Press, Berkeley and Los Angeles, California.

Poelstra JW, Vijay N, Bossu CM *et al.* (2014) The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. *Science*, **344**, 1410–1414.

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

Rheindt FE, Edwards SV (2011) Genetic introgression: an integral but neglected component of speciation in birds. *The Auk*, **128**, 620–632.

Rheindt FE, Fujita MK, Wilton PR, Edwards SV (2014) Introgression and phenotypic assimilation in *Zimmerius* flycatchers (Tyrannidae): population genetic and phylogenetic inferences from genome-wide SNPs. *Systematic Biology*, **63**, 134–152.

Ribeiro ÂM, Lloyd P, Bowie RC (2011) A tight balance between natural selection and gene flow in a southern african arid-zone endemic bird. *Evolution*, **65**, 3499–3514.

Rice NH, Martinez-Meyer E, Peterson AT (2003) Ecological niche differentiation in the *Aphelocoma* jays: a phylogenetic perspective. *Biological Journal of the Linnean Society*, **80**, 369–383.

Rovito SM, Parra-Olea G, Hanken J, Bonett RM, Wake DB (2013) Adaptive radiation in miniature: the minute salamanders of the Mexican highlands (Amphibia: Plethodontidae: *Thorius*). *Biological Journal of the Linnean Society*, **109**, 622–643.

Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJ, Birol I (2009) ABySS: a parallel assembler for short read sequence data. *Genome Research*, **19**, 1117–1123.

Smith BT, Harvey MG, Faircloth BC, Glenn TC, Brumfield RT (2014) Target capture and massively parallel sequencing of ultraconserved elements for comparative studies at shallow evolutionary time scales. *Systematic Biology*, **63**, 83–95.

Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, **30**, 1312–1313.

Streicher JW, Devitt TJ, Goldberg CS, Malone JH, Blackmon H, Fujita MK (2014) Diversification and asymmetrical gene flow across time and space: lineage sorting and hybridization in polytypic barking frogs. *Molecular Ecology*, **23**, 3273–3291.

Sullivan J, Arellano E, Rogers DS (2000) Comparative phylogeography of Mesoamerican highland rodents: concerted versus independent response to past climatic fluctuations. *The American Naturalist*, **155**, 755–768.

Toews DP, Brelsford A (2012) The biogeography of mitochondrial and nuclear discordance in animals. *Molecular Ecology*, **21**, 3907–3930.

Velo-Antón G, Parra J, Parra-Olea G, Zamudio K (2013) Tracking climate change in a dispersal-limited species: reduced spatial and genetic connectivity in a montane salamander. *Molecular Ecology*, **22**, 3261–3278.

Wen D, Yu Y, Hahn MW, Nakhleh L (2016) Reticulate evolutionary history and extensive introgression in mosquito species revealed by phylogenetic network analysis. *Molecular Ecology*, **25**, 2361–2372.

Xi Z, Liu L, Davis CC (2015) Genes with minimal phylogenetic information are problematic for coalescent analyses when gene tree estimation is biased. *Molecular Phylogenetics and Evolution*, **92**, 63–71.

Yu Y, Dong J, Liu KJ, Nakhleh L (2014) Maximum likelihood inference of reticulate evolutionary histories. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, 16448–16453.

Data accessibility

Raw sequence data and UCE fasta files are available from the NCBI SRA: SRA289270 (BioProject PRJNA317781). PHYLUCE code used throughout the manuscript is available under an open-source licence from https://github.com/faircloth-lab/phyluce. DNA alignments, inputs for STRUCTURE and SNAPP, and files and scripts associated with SNP calling are available from the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.s8v80.

All authors designed the research; E.Z., W.L.E.T. and B.C.F. performed molecular work; E.Z. analysed the data; all authors wrote the manuscript.