



Detecting introgression despite phylogenetic uncertainty: The case of the South American siskins

Elizabeth J. Beckman^{1,2}  | Phred M. Benham¹ | Zachary A. Cheviron¹ | Christopher C. Witt²

¹Division of Biological Sciences, University of Montana, Missoula, Montana

²Department of Biology and Museum of Southwestern Biology, University of New Mexico, Albuquerque, New Mexico

Correspondence

Elizabeth J. Beckman, Division of Biological Sciences, University of Montana, Missoula, MT.

Email: libby.beckman@gmail.com

Funding information

Division of Environmental Biology, Grant/Award Number: 1146491; American Museum of Natural History; American Ornithological Society

Abstract

Genetic introgression among closely related species is a widespread phenomenon across the Tree of Life and could be an important source of adaptive variation during early stages of diversification. In particular, genomic studies have revealed that many rapidly radiating clades tend to have complex, reticulate evolutionary histories. Although rapid radiations appear to be susceptible to introgression, they present special challenges for its detection because formal tests require accurate phylogenies, and paradoxically, introgression itself may obscure evolutionary relationships. To address this methodological challenge, we assessed introgression in a recent, rapid avian radiation in the Andes, the South American siskins (*Spinus*). Using ~45,000 SNPs, we estimated the *Spinus* phylogeny using multiple analytical approaches and recovered four strongly conflicting topologies. We performed a series of complimentary introgression tests that included valid tests for each of the likely species trees. From the consistency of test results, we inferred multiple introgression events among Andean *Spinus* in a way that was robust to phylogenetic uncertainty in the species tree. Positive tests for introgression were corroborated by independent population structure and ancestral assignment analyses, as well as a striking geographic pattern of mitochondrial haplotype sharing among species. The methodological approach we describe could be applied using any genomewide data, including SNP data, for clades without fully resolvable species trees. Our discovery of multiple introgression events within the Andean radiation of *Spinus* siskins is consistent with an emerging paradigm, that introgression tends to accompany the early stages of diversification.

KEYWORDS

Andes, introgression, molecular evolution, phylogenomics, rapid radiation, *Spinus*

1 | INTRODUCTION

Introgression, the transfer of genetic material from one independent lineage into the gene pool of another (Mallet, 2005), commonly occurs between closely related species pairs across the Tree of Life (Payseur & Rieseberg, 2016). Introgression can impact functional traits (Lewontin & Birch, 1966; Miao, Wang, & Li, 2016) and shape

adaptive divergence by altering standing genetic variation (Grant & Grant, 2016; Parsons, Son, & Albertson, 2011; Pease, Haak, Hahn, & Moyle, 2016; Winger, 2017). Within rapid radiations where nascent species barriers may be challenged early in diversification, introgression may be of particular importance either as a creative or as homogenizing force (Seehausen, 2004). Genomic studies in several rapid radiations in plants, arthropods, and vertebrates have revealed

reticulate evolutionary histories with introgression occurring among multiple species within a clade (Alexander et al., 2016; Cui et al., 2013; Dasmahapatra et al., 2012; Fontaine et al., 2015; Gompert et al., 2014; Keller et al., 2013; Lamichhaney et al., 2015; Meyer, Matschiner, & Salzburger, 2016; Vargas, Ortiz, & Simpson, 2017). Within recent, speciose clades, alleles with major effects have introgressed with consequences on diet (Grant & Grant, 2016; Lamichhaney et al., 2015; Richards & Martin, 2017), behavioural reproductive isolation (Genner & Turner, 2012; Keller et al., 2013; Meier et al., 2017; Meyer et al., 2016), predation risk (Dasmahapatra et al., 2012; Wallbank et al., 2016), and geographic range (Fontaine et al., 2015; Wen, Yu, Hahn, & Nakhleh, 2016). All told, these results suggest that introgression may be a typical and important characteristic of rapid radiations (Mallet, Besansky, & Hahn, 2016; Seehausen, 2004). However, nuclear DNA (nDNA) introgression does not always occur when closely related lineages come into contact (Good, Vanderpool, Keeble, & Bi, 2015; Melo-Ferreira, Seixas, Cheng, Mills, & Alves, 2014), and introgression has been formally assessed in only a small fraction of the rapid radiations found across the Tree of Life. Defining the frequency and predictability of introgression in diverse lineages will be essential for understanding the early stages of diversification.

Studying introgression within rapid radiations is challenging. Formal introgression tests require accurate phylogenies (Patterson et al., 2012); paradoxically, introgression itself may obscure the phylogenetic relationships within recent radiations. Rapid radiations have two potential sources of gene heterogeneity: incomplete lineage sorting (ILS) (Edwards, 2009; Linkem, Minin, & Leaché, 2016; Maddison, 1997) and introgression (Pease & Hahn, 2015; Solís-Lemus, Yang, & Ané, 2016; Yu, Dong, Liu, & Nakhleh, 2014). The presence of incomplete lineage sorting, or retention of ancestral variation across divergence events, is more probable when speciation events occur in close succession (Maddison, 1997). Coalescent-based phylogenetic methods were developed to address this problem, but introgression violates the assumptions of widely used traditional and coalescent-based phylogenetic approaches (Edwards et al., 2016; Solís-Lemus et al., 2016; Yu, Degnan, & Nakhleh, 2012; Yu et al., 2014). When introgression is rampant, the true species tree may be represented in only a small fraction of loci across the genome (Fontaine et al., 2015); in other cases, a large proportion of the genome must be sequenced to estimate a statistically robust phylogeny (Cui et al., 2013; Lamichhaney et al., 2015). Phylogenetic uncertainty in rapid radiations may remain due to either source of gene-tree heterogeneity, even with genome-scale data (Linkem et al., 2016; Novikova et al., 2016; Suh, Smeds, & Ellegren, 2015). Often the extent and source of phylogenetic uncertainty are only revealed with exhaustive phylogenetic analyses (Crowl, Myers, & Cellinese, 2017; Shen, Hittinger, & Rokas, 2017; Smith, Moore, Brown, & Yang, 2015). Further, conventional methods for assessing phylogenetic support may over-represent the confidence for a specific reconstruction when large amounts of data are available (Jarvis et al., 2014; Smith et al., 2015; Suh, 2016). To elucidate evolutionary history despite the presence of ILS and introgression, researchers have

taken several different approaches, including comparing the fit of genomic data when migration is allowed or excluded in population models that incorporate colonization history derived from nongenetic information (Alexander et al., 2016); modifying existing introgression tests for more complex scenarios (Eaton, Hipp, González-Rodríguez, & Cavender-Bares, 2015; Morales, Jackson, Dewey, O'Meara, & Carstens, 2017; Pease & Hahn, 2015), and explicitly incorporating introgression into phylogenetic networks based on gene trees (Meyer et al., 2016; Solís-Lemus & Ané, 2016; Wen, Yu, & Nakhleh, 2016; Wen, Yu, Hahn et al., 2016; Yu et al., 2012, 2014). Despite this progress, a simple, general approach that can be applied to any unresolved phylogeny, including clades without genomic resources, is needed to establish the extent and significance of introgression in the early stages of diversification.

Here, we recommend a set of existing analyses that together allow conservative inference of introgression events even when the phylogeny remains unresolved. This approach could be broadly applicable to a wide range of organisms due to its simple data requirements: modest intraspecific sampling for the focal taxa, a nuclear genomewide SNP or sequence data set, and a mitochondrial or chloroplast DNA sequence data set. Appropriate SNP data sets can be produced by popular sequencing protocols (Leaché & Oaks, 2017; Ree & Hipp, 2015) and de novo assembly methods (Catchen et al., 2011; Eaton, 2014) that are affordable and accessible even for non-model taxa without genomic resources. With these type of data, we propose the following steps: (a) explore population structure using nDNA sequence divergence and population allele frequencies. In clades impacted by incomplete lineage sorting and/or introgression, different analytical methods may yield discordant phylogenetic hypotheses; (b) test for introgression by employing a suite of formal introgression tests based on site pattern and population allele frequencies; formal four-taxon introgression tests that are valid for each of the alternative phylogenetic hypotheses should be included; (c) identify introgression events by looking for consensus across tests; and (d) compare nDNA introgression patterns with the evolutionary history of a mitochondrial or chloroplast genome.

We applied this approach to test introgression in a rapid radiation of Neotropical finches, the South American siskins (*Spinus*). The South American clade of *Spinus*, with 11 endemic species on the continent, is recent (~0.55 million years old), diversified rapidly (Beckman & Witt, 2015) and has maximum diversity in the high Central Andes where three to four *Spinus* species co-occur over the course of the year (Fjeldså & Krabbe, 1990; Ridgely & Tudor, 1989). There is extensive mitochondrial haplotype sharing among sympatric Andean siskins; for example, all Andean individuals of the widespread *Spinus magellanicus* possessed mitochondrial DNA (mtDNA) haplotypes identical to either *S. atratus* or *S. crassirostris*, two high-elevation restricted taxa (Beckman & Witt, 2015). Reticulate evolutionary histories have been well-documented in a similarly rapid granivorous passerine bird radiation, the Darwin's finches, (Grant & Grant, 2016; Grant, Grant, & Petren, 2005; Lamichhaney et al., 2015; Lamichhaney et al., 2016), and may be characteristic of rapid Andean plant radiations (Pease et al., 2016; Vargas et al., 2017).

However, it is possible that mtDNA haplotype sharing in Andean *Spinus* could be the product of incomplete lineage sorting alone and that species barriers have been maintained in sympatry. Testing for introgression in *Spinus* is an appropriate challenge with which to demonstrate the recommendations we propose, and our study provides potential insight into the biogeography and genetics of reticulate evolutionary histories.

2 | METHODS

2.1 | Taxonomic sampling

We used museum collections and field expeditions to compile frozen tissue samples for 148 *Spinus* individuals including three species restricted to high elevation in the Andes, *S. atratus* ($N = 8$), *S. crassirostris* ($N = 16$) and *S. uropygialis* ($N = 20$); one widespread, polytypic species, *S. magellanicus*, sampled broadly throughout its western range (Ecuador [$N = 3$], Peru [$N = 89$], Bolivia [$N = 4$] and northern Argentina [$N = 4$]). These taxa are referred to as Andean *Spinus* hereafter. We included *S. cucullatus* ($N = 4$) sampled from Guyana as an outgroup. Previous work demonstrated this sampling should accurately represent mtDNA genetic diversity for these species (Beckman & Witt, 2015).

2.2 | DNA sequencing

2.2.1 | Genotyping-by-sequencing

We extracted DNA from muscle tissue samples with the Qiagen DNeasy Blood and Tissue kit (Qiagen, Valencia, CA, USA). We used a modified, double-digest, genotyping-by-sequencing (GBS) approach (Parchman et al., 2012) to sequence tens of thousands of loci from across the genome. We chose 40 individuals with high-quality extracts for GBS (Table 1) to represent the geographic, taxonomic and mtDNA diversity present in *Spinus* (Beckman & Witt, 2015). We included four individuals per species of *S. atratus*, *S. crassirostris* and *S. uropygialis*, and one *S. cucullatus*. We sequenced *S. magellanicus* from four regions: central Peru including the departments of Lima and Ancash, Peru ($N = 16$), the department of Cusco, Peru ($N = 5$), Cochabamba, Bolivia ($N = 3$), and northern Argentina ($N = 3$).

We followed Parchman et al. (2012) to generate sequencing libraries using a size-selection window of 500–600 bp. We pooled individual libraries in equimolar ratios to produce the pooled product which was sequenced on a single flow-cell lane of Illumina HiSeq 2500 at the W. M. Keck Sequencing Center at the University of Illinois, Urbana-Champaign. Sequencing resulted in >200 million single-end 100-bp reads with average Phred scores over 30.

We used the STACKS PIPELINE v1.35 (Catchen, Hohenlohe, Bassham, Amores, & Cresko, 2013; Catchen et al., 2011) to process the GBS raw reads into de novo loci. STACKS recovers homologous loci between different samples without the aid of a genome, and it is effective across a range of divergence levels. Its flexibility is due to three user-designated parameters; the raw read depth required to

assemble a locus (m), the maximum mismatches permitted within a locus in an individual (M) and the maximum mismatches allowed between individuals within a single locus (n). The appropriate values of these parameters vary based on the divergence expected between individuals, the depth of sequencing, and the sequence error rate (Mastretta-Yanes et al., 2015). Based on previous work (Catchen et al., 2013; Harvey et al., 2015; Mastretta-Yanes et al., 2015) and preliminary analysis of m with a single individual, we set $m = 4$ to maximize the number of high-confidence loci recovered. To test the robustness of our assembly to variation in assembly parameters, we processed the data through STACKS under five conditions: (B1) $n = 1$, $M = 2$, (B2) $n = 2$, $M = 2$, (B3) $n = 2$, $M = 3$, (B4) $n = 3$, $M = 3$, (B5) $n = 3$, $M = 4$. B1 had most restrictive assembly criteria (97% sequence similarity), B5 was most lenient (93% sequence similarity). We eliminated potentially confounded loci from highly repetitive regions during locus construction ($-\text{max_locus_stacks}=3$) and by filtering with RXSTACKS ($-\text{ln_lim}$, $-\text{conf_filter}$; Catchen et al., 2013).

We evaluated assemblies B1 through B5 by calculating the average number of unique alleles per locus for each. Under inappropriately strict mismatch parameters, a true homologous locus with distinct, divergent alleles may be artificially split, resulting in a lower value of average unique alleles per locus (Harvey et al., 2015). Comparable values of the average unique alleles per locus between assemblies indicate that the loci constructed in each are robust to parameter changes. To minimize the number of nonhomologous loci that were incorrectly combined, we chose the most restrictive assembly, assembly B3, from those with similar unique alleles per locus to use for all subsequent analyses.

To control the amount of missing data, we used the parameters r and p in the STACKS populations program to filter the SNP data set; r defines the proportion of individuals required per population for a locus to be considered, whereas p refers the number of populations required for a locus (Catchen et al., 2013). We treated each species as a taxonomic unit in the populations program except for *S. magellanicus*. Informed by previous work (Beckman & Witt, 2015) and preliminary analyses, we analysed *S. magellanicus* as three populations defined by latitude: 8°–14°S, hereafter “Peru” ($N = 21$); 15°–21°S, hereafter “Bolivia” ($N = 3$), and 22°–28°S, hereafter “Argentina” ($N = 3$). These designations resulted in seven *Spinus* lineages.

2.2.2 | Mitochondrial genes

For all 148 *Spinus* samples, excepting 53 *S. magellanicus*, we amplified and sequenced the mitochondrial genes cytochrome b (*cytb*), NADH dehydrogenase II (ND2), and NADH dehydrogenase III (ND3) as in Beckman and Witt (2015). We found ND3 was sufficient to identify major mtDNA clades and sequenced the gene for the remaining 53 *S. magellanicus*. Sequencing was conducted at the University of New Mexico Molecular Core Facility (Albuquerque, NM). We edited chromatograms by eye in SEQUENCHER 4.10.1 (GeneCodes Corp., Ann Arbor, MI, USA), aligned loci with MUSCLE v3.7 (Edgar 2004) on the CIPRES PORTAL (Miller, Pfeiffer, & Schwartz, 2010) and confirmed the alignments by eye in MACCLADE v4.08 (Maddison & Maddison 2005).

TABLE 1 Genotyping-by-sequencing sampling. Elevation (Elev.) in metres. Reads with >30 mean Phred score are reported as Raw Reads; loci are from assembly B3. SetC and SetE columns indicate whether a sample was included in respective subsets

Individual	Species	Elev.	Locality	Latitude	Longitude	Raw reads	Loci	SetC	SetE
MSB:Bird:34123	<i>Spinus atratus</i>	4,300	PERU: Apurimac: Abancay: Chacoche	-14.063	-73.009	1,847,044	150,964	X	—
MSB:Bird:33129	<i>S. atratus</i>	3,835	PERU: Cusco: Ollantaytambo: Choquechaca	-13.188	-72.231	1,913,373	157,428	X	X
MSB:Bird:33088	<i>S. atratus</i>	4,400	PERU: Cusco: Ollantaytambo: Choquechaca	-13.188	-72.231	1,786,290	151,718	X	X
FMNH:AVES:391992	<i>S. atratus</i>	4,140	PERU: Lima: Huarochiri: Japani	-11.683	-76.533	1,469,527	120,267	—	—
MSB:Bird:33077	<i>S. crassirostris</i>	4,200	PERU: Cusco: Ollantaytambo: Choquechaca	-13.188	-72.231	2,384,107	211,144	X	X
MSB:Bird:33091	<i>S. crassirostris</i>	4,200	PERU: Cusco: Ollantaytambo: Choquechaca	-13.188	-72.231	2,302,031	201,734	X	X
MSB:Bird:31528	<i>S. crassirostris</i>	3,973	PERU: Lima: Huarochiri: Carampoma	-11.628	-76.434	2,516,502	214,128	X	X
MSB:Bird:31505	<i>S. crassirostris</i>	3,981	PERU: Lima: Huarochiri: Carampoma	-11.628	-76.434	2,678,762	216,611	X	X
USNM:Birds:B12867	<i>S. cucullatus</i>	230	GUYANA			2,319,237	191,068	X	X
UWBM:ORN:70271	<i>S. magellanicus</i>	209	ARGENTINA: Misiones: Posadas	-26.955	-55.088	1,687,108	129,920	X	X
UWBM:ORN:70272	<i>S. magellanicus</i>	209	ARGENTINA: Misiones: Posadas	-26.955	-55.088	1,447,694	123,440	X	X
UWBM:ORN:70716	<i>S. magellanicus</i>	340	ARGENTINA: Tucumán: San Miguel de Tucumán	-27.022	-65.645	1,714,660	135,705	X	X
FMNH:AVES:334723	<i>S. magellanicus</i>	2,470	BOLIVIA: Cochabamba: Cochabamba-Oruro Rd.	-17.504	-66.329	1,951,949	166,978	X	X
FMNH:AVES:396033	<i>S. magellanicus</i>	2,470	BOLIVIA: Cochabamba: Cochabamba-Oruro Rd.	-17.504	-66.329	1,745,484	143,165	X	—
FMNH:AVES:334722	<i>S. magellanicus</i>	2,470	BOLIVIA: Cochabamba: Cochabamba-Oruro Rd.	-17.504	-66.329	1,906,384	162,422	X	X
MSB:Bird:34881	<i>S. magellanicus</i>	2,972	PERU: Ancash: Santa: Macate	-8.755	-78.048	1,796,953	150,017	—	—
MSB:Bird:34999	<i>S. magellanicus</i>	3,714	PERU: Ancash: Caraz: Pueblo Libre: SW Caraz	-9.101	-77.866	2,207,409	178,262	X	—
MSB:Bird:35015	<i>S. magellanicus</i>	3,714	PERU: Ancash: Caraz: Pueblo Libre: SW Caraz	-9.101	-77.866	1,997,706	160,239	—	—
MSB:Bird:34249	<i>S. magellanicus</i>	<100	PERU: Ancash: Huarmey	-10.068	-78.136	2,371,488	189,171	X	X
MSB:Bird:34232	<i>S. magellanicus</i>	<100	PERU: Ancash: Huarmey	-10.068	-78.136	2,279,538	179,967	X	—
MSB:Bird:34250	<i>S. magellanicus</i>	<100	PERU: Ancash: Huarmey	-10.068	-78.136	2,109,526	171,765	X	—
MSB:Bird:34251	<i>S. magellanicus</i>	<100	PERU: Ancash: Huarmey	-10.068	-78.136	1,619,697	132,965	—	—
MSB:Bird:34882	<i>S. magellanicus</i>	2,972	PERU: Ancash: Santa: Macate	-8.755	-78.048	1,243,386	97,967	—	—
MSB:Bird:33140	<i>S. magellanicus</i>	3,018	PERU: Cusco: Ollantaytambo: Choquechaca	-13.188	-72.231	1,637,057	135,922	—	—
MSB:Bird:33112	<i>S. magellanicus</i>	3,835	PERU: Cusco: Ollantaytambo: Choquechaca	-13.188	-72.231	2,207,694	149,100	—	—
MSB:Bird:27201	<i>S. magellanicus</i>	3,380	PERU: Cusco: Urubamba: 7.9 km NW Urubamba	-13.249	-72.169	2,365,420	185,163	X	X
MSB:Bird:27175	<i>S. magellanicus</i>	3,500	PERU: Cusco: Urubamba: 7.9 km NW Urubamba	-13.249	-72.169	1,925,996	163,831	X	—
MSB:Bird:27176	<i>S. magellanicus</i>	3,500	PERU: Cusco: Urubamba: 7.9 km NW Urubamba	-13.249	-72.169	1,963,590	154,841	X	X
MSB:Bird:32938	<i>S. magellanicus</i>	935	PERU: Lima: 2.3 km E Nieve Nieve	-12.030	-76.651	2,042,449	172,064	X	—
MSB:Bird:32907	<i>S. magellanicus</i>	935	PERU: Lima: 2.3 km E Nieve Nieve	-12.030	-76.651	2,439,440	185,163	X	—
MSB:Bird:28301	<i>S. magellanicus</i>	3750	PERU: Lima: San Pedro de Casta: Carhuayumac	-11.762	-76.549	1,671,815	137,642	—	—

(Continues)

TABLE 1 (Continued)

Individual	Species	Elev.	Locality	Latitude	Longitude	Raw reads	Loci	SetC	SetE
MSB:Bird:28373	<i>S. magellanicus</i>	3,800	PERU: Lima: San Pedro de Casta: Chinchaycocha	−11.761	−76.568	2,159,135	179,865	X	X
MSB:Bird:33408	<i>S. magellanicus</i>	3,905	PERU: Lima: Huarochiri: San Pedro de Casta	−11.768	−76.535	2,011,512	168,659	—	—
MSB:Bird:33488	<i>S. magellanicus</i>	4,140	PERU: Lima: Huarochiri: San Pedro de Casta	−11.770	−76.532	2,202,776	180,957	X	X
MSB:Bird:36423	<i>S. magellanicus</i>	121	PERU: Lima: Lima, Comas	−11.944	−77.063	1,973,024	160,481	—	—
MSB:Bird:36390	<i>S. magellanicus</i>	214	PERU: Lima: Lima: Pachacamac: Lurin Valley	−12.153	−76.836	2,681,732	218,246	X	X
MSB:Bird:33453	<i>S. uropygialis</i>	4,131	PERU: Lima: Huarochiri: San Pedro de Casta	−11.769	−76.533	2,345,234	206,819	X	X
MSB:Bird:33454	<i>S. uropygialis</i>	4,131	PERU: Lima: Huarochiri: San Pedro de Casta	−11.769	−76.533	1,863,579	160,701	X	—
MSB:Bird:33470	<i>S. uropygialis</i>	4,131	PERU: Lima: Huarochiri: San Pedro de Casta	−11.769	−76.533	2,174,660	195,620	X	X
MSB:Bird:33471	<i>S. uropygialis</i>	4,131	PERU: Lima: Huarochiri: San Pedro de Casta:	−11.769	−76.533	1,420,740	119,999	—	—

Abbreviations for museums are MSB: Museum of Southwestern Biology; FMNH: Field Museum of Natural History; USNM: Smithsonian Institute; UWBM: Burke Museum.

2.3 | Population structure

We assessed the population genetic structure of Andean *Spinus* with the GBS data by performing a principal component analysis (PCA) with the R-package ADEGENET v2.0.1 (Jombart & Ahmed, 2011) in R v3.2.4 (R Core Team, 2016). The input included all variant sites from the best assembly with $r = 0.5$. As the initial PCA separated only two lineages from the all other taxa, we ran a second PCA omitting the two divergent taxa to assess structure within the remaining populations.

We used the model-based likelihood clustering algorithm implemented in ADMIXTURE v1.3 (Alexander, Novembre, & Lange, 2009) and the GBS data to assign Andean *Spinus* individuals to populations and admixture fractions to individuals without a priori population assignment. The input included the first SNP from every variable locus in assembly B3 with $r = 0.5$ and $p = 6$. To identify the appropriate number of populations (k) for the data, we calculated the cross-validation error (CV error) for k equal to 1–10 and identified k with the lowest values (Alexander et al., 2009).

2.4 | Phylogenetics

We estimated a maximum-likelihood (ML) phylogenetic tree from the concatenated GBS data using the hybrid, MPI/PThreads version of RAXML v8.1.17 (Pfeiffer & Stamatakis, 2010; Stamatakis, 2014). The computation efficiency of RAXML permitted the inclusion of tens of thousands of loci for all 40 individuals sequenced; the inclusion of more sites can reveal strong support for nodes that are weakly supported by data subsets and other phylogenetic methods (Jarvis et al., 2014). The rapid computational time of RAXML also allowed a thorough exploration of the effect of missing data on the phylogeny, a

critical component to the analysis of SNP data sets with large amounts of missing data (Davey et al., 2011). To assess the consequences of missing data on topology and node support, we constructed a series of SNP data sets using assembly B3 in which we varied the missing data parameters r from 0.5 to 0.75, and p from 4 to 7. We concatenated all variant sites within a data set and used the Felsenstein invariant-site correction with molecular evolution model ASC_GTRGAMMA in RAXML (Leaché, Banbury, Felsenstein, de Oca, & Stamatakis, 2015) with 500 bootstrap replicates. The number of invariant sites present in phylogenetic construction will impact the branch lengths and therefore the likelihood of each tree (Leaché et al., 2015; Lewis, 2001). The Felsenstein invariant-site correction allays this problem and reduces computational time compared to an analysis with complete, concatenated loci. For the final phylogeny, hereafter RAXML-B3 tree, we selected the r and p combination which minimized missing data and maximized phylogenetic support for internal nodes. We estimated the RAXML-B3 phylogeny using concatenated, complete loci (89 bp/locus), GTRGAMMA and 500 bootstrap replicates.

We generated a phylogenetic tree using the multispecies coalescent with the program SNAPP (Bryant, Bouckaert, Felsenstein, Rosenberg, & Roychoudhury, 2012) and the GBS DATA. In this approach, unlinked, biallelic sites have independent gene histories that are constrained by the ultimate species tree (Bryant et al., 2012; Edwards, 2009). Gene heterogeneity is attributed solely to differences in the time of coalescence and incomplete lineage sorting (Bryant et al., 2012). For this analysis, we maximized the number of sites covered across individuals by selecting 20 high coverage individuals including *S. cucullatus* and 2–6 individuals of the remaining populations (Set E, Table 1). We extracted the first variant site from each locus that was present in >66% of individuals in each population ($r = 0.66$,

$p = 7$). To reduce computational time, we binned SNPs into three equal sized data sets and ran each in a separate SNAPP run. We grouped *S. magellanicus* from Bolivia and Argentina, hereafter *S. magellanicus* Bol. Arg., into a single operational taxonomic unit analysis based on the RAXML analyses. We estimated the mutation rates U and V within BEAUTI 2 (Bouckaert et al., 2014) from the data and used the default prior specifications. Each data set was run for three million generations with a 10% burnin. We used TRACER v1.6 (Rambaut, Suchard, Xie, & Drummond, 2014) to evaluate convergence. Once we established concordance among independent runs, we combined the trees and evaluated support for individual nodes in DENSITREE v2.2.4 (Bouckaert & Heled, 2014).

Lastly, we used a quartet inference approach with SVDquartets (Chifman & Kubatko, 2014, 2015), implemented in PAUP* v4.0a147 (Swofford, 2002). This coalescent method calculates the best tree topology by identifying a valid split across all loci for each quartet. The quartet trees are assembled with a variant of Quartet FM (Reaz, Bayzid, & Rahman, 2014) to create the final phylogeny. Again, we maximized the number of sites shared among individuals in SVDquartets by including 29 individuals (Set C, Table 1) with filtering parameters $r = 0.66$, $p = 6$. Note, we used $p = 6$ because unlike SNAPP, SVDquartets will permit a site to be completely missing on one population. Changing the population parameter p from 7 to 6 allowed us to include more individuals and sites in the SVDquartet analysis than the SNAPP analysis. We searched all possible quartets for all analyses. We estimated “lineage trees,” where each individual is treated as a terminal branch, and “species trees”, where members of each population are defined a priori.

For individuals sequenced for mitochondrial genes *cytb*, ND3, and ND2, we ran PARTITIONFINDER 2 (Lanfear, Frandsen, Wright, Senfeld, & Calcott, 2016; Stamatakis, 2014) using the greedy search algorithm (Lanfear, Calcott, Ho, & Guindon, 2012), linked branch lengths and the Akaike's information criterion (AIC) to select an appropriate partition and evolutionary model scheme for phylogenetic analysis. Using the evolutionary model GTRGAMMA, and four partitions (*cytb* codon 1; ND3 and ND2 codon 1; codon 2; and codon 3), we estimated a maximum-likelihood phylogeny from the mitochondrial data using RAXML v7.2.7 (Stamatakis, 2014) on the CIPRES SCIENCE GATEWAY V. 3.1 (Miller et al., 2010) with 500 bootstrap replicates.

2.5 | nDNA introgression tests

To test for introgression among Andean *Spinus* with the GBS data, we used three tree-based methods: the ABBA/BABA test on fixed sites (Green et al., 2010), the ABBA/BABA test on polymorphic sites (Durand, Patterson, Reich, & Slatkin, 2011; Patterson et al., 2012), and the four-population test (Reich, Thangaraj, Patterson, Price, & Singh, 2009). These tests varied in their assumptions and power (Alexander et al., 2009; Martin, Cutler et al., 2015; Patterson et al., 2012; Rheindt, Fujita, Wilton, & Edwards, 2014). We defined 31 well-supported four-taxon trees (referred to as Trees 1–31 hereafter) from our RAXML-B3 phylogeny (Table 2); a subset of these trees was

also valid for topologies produced by SNAPP and SVDquartets (Table 2). To understand how the signal of introgression might vary across the landscape (Eaton et al., 2015), we used the PCA results to delineate four geographic populations with distinct genetic variation within *S. magellanicus*: Central Peru (CP, from Lima and Ancash), Southern Peru (SP, from Cusco), Bolivia and Argentina. We defined positive evidence for introgression as (a) a raw z-score $>|2.0|$, the equivalent to an uncorrected p -value of <0.0455 for a two-tailed test and (b) passing the false discovery rate (FDR) threshold of 0.05 (Benjamini & Hochberg, 1995) to mitigate the consequences of multiple tests for each method. This FDR approach was robust to particular dependencies among tests, and controlled the rate of type I errors while minimizing the rate of type II errors that may arise from more conservative multiple test corrections (Benjamini & Yekutieli, 2001).

We applied the ABBA/BABA test (Green et al., 2010), a site pattern frequency test, to each of the 31 trees using biallelic sites in which sister taxa were fixed for different alleles, for example, A and B. Given a strictly bifurcating population history among four taxa, for example ((Taxon1, Taxon2) Taxon3) Taxon4, we expect the allelic patterns “ABBA,” where Taxon2 and Taxon3 share allele B, and “BABA”, where Taxon1 and Taxon3 share allele B, with equal probability across the genome. Thus, we predict a 1:1 ratio of “ABBA” and “BABA” patterns at unlinked loci, and the summary statistic D should equal zero (Green et al., 2010; Patterson et al., 2012; Rheindt & Edwards, 2011). For Trees 1–31, we selected loci that were fixed within each taxon, but variable among the four taxa, and extracted the first variable site from each locus. We counted “ABBA” and “BABA” sites to calculate D and tested the null hypothesis that $D = 0$ through 1,000 bootstrap replicates. The latter steps were conducted in R v3.2.4 (R Core Team, 2016) with code modified from EVOBIR v1.1 (Blackmon, 2016; Streicher et al., 2014). The sign of the D statistic, given $D \neq 0$, indicated which of the two possible taxa (Taxon1 or Taxon2) had experienced introgression with Taxon3.

We also conducted the ABBA/BABA test on polymorphic loci for every tree possible from the 31 tree data set; the loci that contribute to the polymorphic and fixed ABBA/BABA methods are mutually exclusive sets. Again, we expected $D = 0$ for a strictly bifurcating population history; however, we calculated D from population allele frequencies of independent polymorphic loci (Durand et al., 2011; Patterson et al., 2012). Thus, we only conducted this analysis on trees that included three or more individuals per population. Further, we generated two data sets: “Set A” included all individuals; “Set C” included only high coverage individuals (see Table 1). Because we required all loci to be present for every individual in the set ($r = 1$), “Set A” data sets had more accurate population allele frequency estimates, but fewer total loci than in “Set C”. We calculated D and an associated z-score for the null hypothesis that $D = 0$ through 1,000 bootstrap replicates with code modified from EVOBIR.

Finally, we used the four-population test to assess introgression patterns (Reich et al., 2009). Given an unrooted tree, for example (Taxon1, Taxon2), (Taxon3, Taxon4), the allele frequency differences between sisters Taxon1 and Taxon2 should be uncorrelated with the

TABLE 2 ABBA/BABA and four-population test results. Introgression test results based on trees 1–31; each tree is arranged so Taxon1 and Taxon2 are sister, potentially introgressing taxa are Taxon2 and Taxon3, and Taxon4 is closest to the root. Z-score and number of GBS unlinked, biallelic sites per test presented. Significant Z-scores are in bold; † indicates a Z-score that is >2.0, but did not pass false discovery threshold. Trees 1–31 were derived from RAXML results. Tree Topology indicates whether a tree is valid for the alternative phylogenies: SNAPP-S1 (S1), SNAPP-S2 (S2) and SVDquartet (SVD) trees. Set A includes all individuals; Set C includes a subset (Table 1). *S. magellanicus* abbreviations refer to central Peru (CP), southern Peru (SP), Peru (P), Bolivia (B), Argentina (A), or Bolivia and Argentina (BA)

Tree	Tree topology			Taxon1	Taxon2	Taxon3	Taxon4	ABBA/BABA fix		Four-population test				ABBA/BABA population			
	S1	S2	SVD					SET A		SET C		SET A		SET C		SET A	
								Z	Sites	Z	Sites	Z	Sites	Z	Sites	Z	Sites
1	–	–	–	<i>S. uropy.</i>	<i>S. atrat.</i>	<i>S. crass.</i>	<i>S. cucul.</i>	6.45	389	–	–	–	–	–	–	–	–
2	X	X	X	<i>S. mag.</i> SP	<i>S. atrat.</i>	<i>S. crass.</i>	<i>S. cucul.</i>	0.29	190	–	–	–	–	–	–	–	–
3	X	–	–	<i>S. uropy.</i>	<i>S. mag.</i> B	<i>S. atrat.</i>	<i>S. crass.</i>	0.54	259	–0.33	18,522	0.30	15,160	0.03	2,030	0.41	1,654
4	–	–	–	<i>S. uropy.</i>	<i>S. mag.</i> B	<i>S. atrat.</i>	<i>S. cucul.</i>	2.02*	432	–	–	–	–	–	–	–	–
5	–	–	–	<i>S. uropy.</i>	<i>S. mag.</i> BA	<i>S. atrat.</i>	<i>S. cucul.</i>	0.38	426	–	–	–	–	–	–	–	–
6	–	–	–	<i>S. uropy.</i>	<i>S. mag.</i> P	<i>S. crass.</i>	<i>S. cucul.</i>	4.82	141	–	–	–	–	–	–	–	–
7	–	X	X	<i>S. mag.</i> BA	<i>S. mag.</i> P	<i>S. crass.</i>	<i>S. cucul.</i>	7.15	118	–	–	–	–	–	–	–	–
8	X	X	X	<i>S. atrat.</i>	<i>S. mag.</i> P	<i>S. crass.</i>	<i>S. cucul.</i>	2.24	56	–	–	–	–	–	–	–	–
9	X	X	X	<i>S. atrat.</i>	<i>S. mag.</i> CP	<i>S. crass.</i>	<i>S. cucul.</i>	2.36	69	–	–	–	–	–	–	–	–
10	–	–	–	<i>S. uropy.</i>	<i>S. mag.</i> CP	<i>S. crass.</i>	<i>S. cucul.</i>	5.36	169	–	–	–	–	–	–	–	–
11	–	X	X	<i>S. mag.</i> BA	<i>S. mag.</i> CP	<i>S. crass.</i>	<i>S. cucul.</i>	7.10	138	–	–	–	–	–	–	–	–
12	X	–	–	<i>S. mag.</i> BA	<i>S. uropy.</i>	<i>S. atrat.</i>	<i>S. crass.</i>	0.29	281	1.00	12,708	0.82	10,967	1.49	1,209	1.00	1,034
13	X	–	–	<i>S. mag.</i> A	<i>S. uropy.</i>	<i>S. atrat.</i>	<i>S. crass.</i>	1.54	294	–1.17	13,163	–1.59	11,220	1.17	1,042	1.31	852
14	–	–	–	<i>S. mag.</i> B	<i>S. uropy.</i>	<i>S. crass.</i>	<i>S. cucul.</i>	0.99	369	–	–	–	–	–	–	–	–
15	–	–	–	<i>S. mag.</i> BA	<i>S. uropy.</i>	<i>S. crass.</i>	<i>S. cucul.</i>	0.30	400	–	–	–	–	–	–	–	–
16	–	–	–	<i>S. mag.</i> B	<i>S. uropy.</i>	<i>S. mag.</i> P	<i>S. cucul.</i>	0.20	110	–	–	–	–	–	–	–	–
17	X	–	–	<i>S. mag.</i> B	<i>S. uropy.</i>	<i>S. mag.</i> P	<i>S. crass.</i>	0.35	69	2.09	17,649	2.19	8,341	2.16	1,711	2.20	653
18	X	–	–	<i>S. mag.</i> A	<i>S. uropy.</i>	<i>S. mag.</i> P	<i>S. crass.</i>	0.70	74	–2.51	13,220	–2.31	6,993	2.79	932	2.13	419
19	–	–	–	<i>S. mag.</i> BA	<i>S. uropy.</i>	<i>S. mag.</i> P	<i>S. cucul.</i>	0.69	136	–	–	–	–	–	–	–	–
20	–	–	–	<i>S. mag.</i> B	<i>S. uropy.</i>	<i>S. mag.</i> SP	<i>S. cucul.</i>	1.21	400	–	–	–	–	–	–	–	–
21	X	–	–	<i>S. mag.</i> B	<i>S. uropy.</i>	<i>S. mag.</i> SP	<i>S. crass.</i>	0.99	233	2.29	19,827	2.46	13,869	1.74	1,961	2.19	1,346
22	–	–	–	<i>S. mag.</i> BA	<i>S. uropy.</i>	<i>S. mag.</i> SP	<i>S. cucul.</i>	1.61	399	–	–	–	–	–	–	–	–
23	X	–	–	<i>S. mag.</i> BA	<i>S. uropy.</i>	<i>S. mag.</i> SP	<i>S. crass.</i>	0.96	249	3.08	13,537	3.13	10,351	2.77	1,138	3.46	853
24	X	–	–	<i>S. mag.</i> A	<i>S. uropy.</i>	<i>S. mag.</i> SP	<i>S. crass.</i>	2.49	240	–3.70	14,035	–4.13	10,389	2.74	1,027	4.11	720
25	X	–	–	<i>S. mag.</i> BA	<i>S. uropy.</i>	<i>S. mag.</i> CP	<i>S. crass.</i>	0.30	95	–1.78	12,660	–1.46	6,950	2.34	1,184	1.54	556
26	X	–	–	<i>S. mag.</i> B	<i>S. uropy.</i>	<i>S. mag.</i> CP	<i>S. crass.</i>	0.41	88	2.02	19,200	2.83	9,758	2.06	1,950	3.03	838
27	–	–	–	<i>S. mag.</i> B	<i>S. uropy.</i>	<i>S. mag.</i> CP	<i>S. cucul.</i>	0.42	133	–	–	–	–	–	–	–	–
28	–	–	–	<i>S. mag.</i> BA	<i>S. uropy.</i>	<i>S. mag.</i> CP	<i>S. cucul.</i>	0.30	164	–	–	–	–	–	–	–	–
29	X	–	–	<i>S. mag.</i> BA	<i>S. uropy.</i>	<i>S. mag.</i> CP	<i>S. crass.</i>	0.54	120	1.88	13,491	1.68	7,914	2.19	1,313	1.41	682
30	X	–	–	<i>S. mag.</i> A	<i>S. uropy.</i>	<i>S. mag.</i> CP	<i>S. crass.</i>	1.97	97	–2.39	14,157	–2.12	7,998	2.63	1,086	1.59	517
31	X	X	X	<i>S. mag.</i> CP	<i>S. mag.</i> SP	<i>S. atrat.</i>	<i>S. crass.</i>	0.00	76	–4.09	16,701	–2.60	8,114	2.96	2,335	1.03	1,073

allele frequency differences between sisters Taxon3 and Taxon4 (Reich et al., 2009, 2010). Summed across the genome, a violation of this prediction indicates the data do not support a strictly tree-like

relationship among the four taxa, and introgression has occurred. Again, we required all loci to be present for each individual ($r = 1$) in a tree and generated “Set A” and “Set C”. We calculated the test

statistic f_4 and a z-score for each tree in TREEMIX v1.12 (Pickrell & Pritchard, 2012). In the small likelihood of linkage disequilibrium impacting our findings (assemblies in STACKS are unordered with respect to genomic region), we assigned block size to 20 to account for linkage disequilibrium; when we compared block sizes 1 50, 20, 100 and 200 in five trees, the significance of the results did not change for each respective tree.

3 | RESULTS

3.1 | Genotyping-by-sequencing metrics

Our GBS data set contained 82.9 million raw reads for the 40 *Spinus* samples with a median of 1.99 million reads per individual (Table 1). The average number of unique alleles per locus was lowest under the most restrictive assembly parameters (B1) and increased from B1 to B3. The average number of unique alleles per locus was comparable among the more liberal assemblies B3, B4 and B5 (Supporting Information Figure S1) which suggests these assemblies consist of true, homologous loci. We selected assembly B3 for all subsequent analyses as it had the most restrictive STACKS mismatch parameters while still accurately representing the overall diversity of the data. In assembly B3, we recovered a median ~163,800 high-quality loci per individual. With missing data parameters $r = 0.5$ and p set to 4, 5, 6, and 7, respectively, we generated data sets that included between 27,460 and 59,148 variable loci across all individuals. These data sets contained 12–23% missing data; as the number of populations required per locus (p) increased, the number of variable loci recovered and the proportion of missing decreased (Supporting Information Table S1).

3.2 | nDNA population structure

In our GBS principal component analysis, individuals from each *Spinus* taxon tended to be more similar to each other than other *Spinus* taxa (Figure 1). In the full analysis, PC1 and PC2 represented 10.12% and 7.05% of the genetic variation, respectively; these axes separated *S. uropygialis* and *S. crassirostris* from each other and all other taxa (Figure 1a). With *S. crassirostris* and *S. uropygialis* excluded, PC1 and PC2, representing 8.56% and 4.75%, respectively, of the genetic variation (Figure 1b), effectively separated *S. atratus*, *S. magellanicus* Argentina, and *S. magellanicus* Central Peru. *Spinus magellanicus* Bolivia was placed approximately equidistant from *S. magellanicus* Argentina and *S. atratus* suggesting *S. magellanicus* Bolivia individuals may share history with both taxa, although this may also be an artefact of isolation by distance (Novembre & Stephens, 2008). *S. magellanicus* Southern Peru individuals were more similar to *S. atratus* on the PC2 axis than to *S. magellanicus* Central Peru.

We found the ADMIXTURE model best fit the GBS data when the number of populations (k) was set to 1, 2 or 4 (Figure 2a). The ancestral assignment analysis with $k = 2$ (Figure 2b), reconstructed one population with *S. uropygialis* and *S. crassirostris*, and a second

containing all others; two individuals from *S. magellanicus* Bol. Arg. shared ancestry with both populations. The clustering analysis of $k = 4$ (Figure 2), however, closely mirrored a priori *Spinus* taxonomic units. The four populations corresponded to *S. magellanicus* Bol. Arg., *S. crassirostris*, *S. uropygialis* and *S. magellanicus* from Peru lumped with *S. atratus*. The clustering analysis was unable to discriminate between Peruvian *S. magellanicus* and *S. atratus* regardless of k value. In the reconstruction with $k = 4$, several individuals were admixed; this included two *S. atratus* who shared ancestry with *S. magellanicus* Bol. Arg. Six individuals that were identified a priori as *S. magellanicus* Peru shared ancestry with *S. crassirostris*.

3.3 | nDNA phylogenetics

Our GBS maximum-likelihood phylogenetic analyses with assembly B3 resulted in a novel phylogenetic hypothesis for *Spinus*. Below we summarize the evolutionary relationships resolved in the RAXML-B3 phylogeny, reporting bootstrap support (bs) for the concatenated all-loci maximum likelihood. The RAXML-B3 phylogeny was constructed with $r = 0.5$, $p = 6$ for a total of 45,246 variable loci and 17.6% missing data across the alignment. In additional analyses, we found that the topology of the RAXML phylogeny was robust across different GBS assemblies (B1–B5, Supporting Information Table S1).

In the RAXML-B3 phylogeny (Figure 3), the following taxa formed reciprocally monophyletic clades (bs 100 for each clade): *S. crassirostris*, *S. atratus*, *S. uropygialis*, *S. magellanicus* Bol. Arg. and *S. magellanicus* Peru. Using *S. cucullatus* as the root, we found that *S. crassirostris* diverged first (bs 96). Within the remaining taxa, *S. magellanicus* from Argentina and Cochabamba, Bolivia, formed a well-supported monophyletic clade (bs 100); *S. magellanicus* Bol. Arg. was sister to *S. uropygialis* (bs 100). *S. magellanicus* Peru formed a monophyletic clade (bs 63) which was sister to *S. atratus* (bs 81). Within Peruvian *S. magellanicus*, individuals from Southern Peru formed a monophyletic clade with strong support (bs 100); individuals from Central Peru did not form a monophyletic clade due to the placement of one sample as sister to the Southern Peru clade.

The results of the RAXML-B3 phylogeny were well supported across missing data regimes with the exception of the monophyly of *S. magellanicus* Peru, and its sister relationship with *S. atratus* (Supporting Information Table S1). The RAXML-B3 topology for *S. magellanicus* Peru and *S. atratus* was supported in data sets with $p = 6$. With $p = 6$, a locus that was completely absent in one population could be included in the data set. An alternative topology where *S. atratus* was sister to the clade containing *S. uropygialis* and *S. magellanicus* Bol. Arg. was recovered in data sets with $p = 4, 5$ and 7; in a subset of those trees, *S. magellanicus* from Southern Peru was sister to the *S. uropygialis*/*S. magellanicus* Bol. Arg./*S. atratus* clade (Supporting Information Figure S2, Table S1). Support for the conflicting topologies was low across the differing nodes (bs ranged from $29 \leq$ to ≤ 85).

The topology of *S. magellanicus* Peru and *S. atratus* was impacted by missing data and the number of variable sites. In particular, we found missing data were unevenly distributed among taxa in data sets

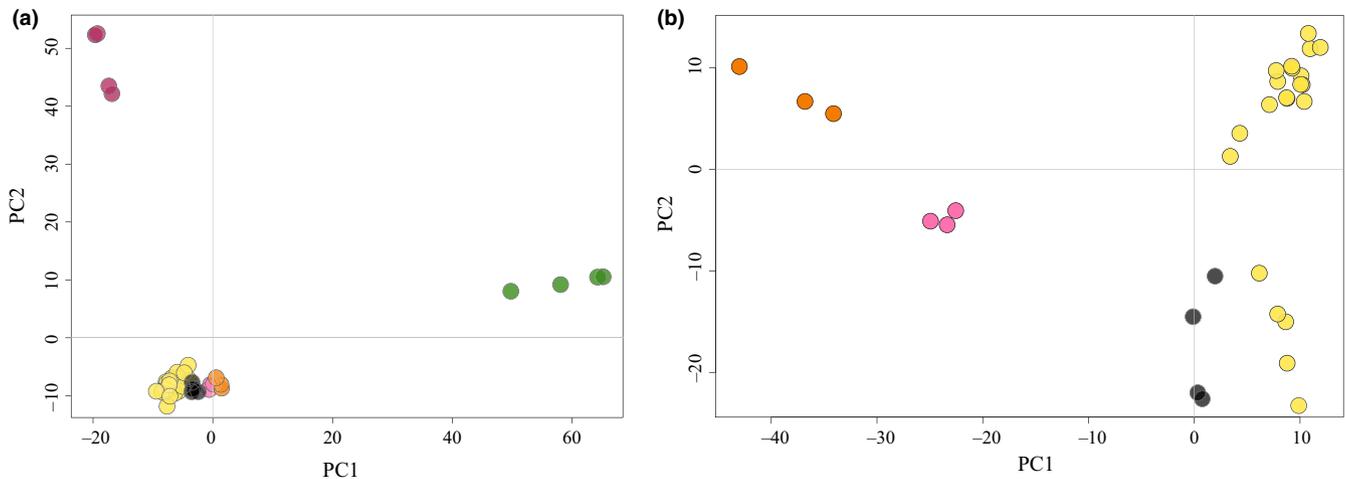


FIGURE 1 Principal component analysis of *Spinus* nDNA. Analyses were produced with 45,246 unlinked, variant sites from assembly B3. Individual birds are coloured by taxon or clade as follows: *S. crassirostris*: maroon, *S. atratus*: black, *S. magellanicus* Peru: yellow, *S. magellanicus* Cochabamba, Bolivia: pink, *S. magellanicus* Argentina: orange, *S. uropygialis*: green. (a) Includes all individuals. (b) Excludes *S. crassirostris* and *S. uropygialis*. *S. magellanicus* Peru is coloured as Southern Peru (light green) or Central Peru (yellow) [Colour figure can be viewed at wileyonlinelibrary.com]

with $p = 4$ and 5. Permitting multiple populations to lack data for certain loci could have introduced bias into the phylogenetic analysis, although the effect of structured missing data on maximum-likelihood estimation has not been systematically investigated (Leaché et al., 2015). Restricting the analysis to loci that were present in all populations ($p = 7$) eliminated over one-third of the sites variable within Andean *Spinus* with $p = 6$. Fewer variable sites within the ingroup could impact topology and diminish support for internal nodes. We selected $p = 6$ as a reasonable compromise between maximizing the number of loci and minimizing bias due to structured missing data.

The sister relationship of *S. magellanicus* Peru and *S. atratus* was recovered in additional unrooted phylogenetic analyses as well as the ADMIXTURE results; this suggests that the methodology, lack of informative sites or missing data affected the power of RAXML to recover this relationship. A conservative interpretation of the RAXML phylogenetic uncertainty is that *S. magellanicus* Central Peru, *S. magellanicus* Southern Peru and *S. atratus* form a four-taxon polytomy with the ancestor of the *S. magellanicus* Bol. Arg. and *S. uropygialis* clade. Based on the above reasoning, we report modest support for monophyly in *S. magellanicus* Peru.

We found that the rooting of the *Spinus* tree in RAXML was sensitive to the proportion of invariant sites in the model. In a SNP data set without the Felsenstein invariant-site correction, we recovered an unrooted tree in which *S. cucullatus* was sister to the ancestor of *S. uropygialis* and *S. magellanicus* Bol. Arg (see “RAXML-no invariant-site correction” in Figure 4a). This root placement was in strong disagreement with the RAXML-B3 phylogeny and the Felsenstein invariant-site corrected analyses (Figure 4a).

Our SNAPP analyses (Supporting Information Figure S3) supported most of the evolutionary relationships recovered in RAXML, but the SNAPP results differed in the placement of the root (Figure 4). In all SNAPP trees, the first taxon to diverge after *S. cucullatus* was

S. uropygialis (pp 100), not *S. crassirostris* as in RAXML (bs 97). Apart from the placement of the root, the SNAPP tree with the strongest posterior probability (pp), referred hereafter as SNAPP-S1 tree, was equivalent to the RAXML-B3 unrooted topology excepting *S. cucullatus*: *S. magellanicus* Peru was sister to *S. atratus* (pp 100), *S. magellanicus* Bol. Arg. was sister to *S. uropygialis*, and *S. crassirostris* was excluded from both of these clades. SNAPP also revealed strong, conflicting signal within the GBS data (Supporting Information Figure S3). The RAXML-B3 unrooted topology was recovered in 58.7% of trees sampled; however, the second most frequently sampled unrooted topology (SNAPP-S2 tree), 29.3% of trees, placed *S. atratus* and *S. magellanicus* Peru as sister, with *S. magellanicus* Bol. Arg. sister to them; *S. uropygialis* and *S. crassirostris* were sisters. The third SNAPP unrooted topology (11.9% of trees) had *S. magellanicus* Peru and *S. atratus* sister, and *S. magellanicus* Bol. Arg. and *S. crassirostris* sister, with *S. uropygialis* equally related to both clades. We hypothesize that conflicting signal within the data prevented convergence (ESS > 100) for demographic parameters (e.g., theta, upon which the coalescent model depends). Consistent with this interpretation, ESS values did not increase with increasing sampling generations.

The topology recovered by SVDquartets (Supporting Information Figure S4) was similar to the SNAPP-S2 tree. The support for the nodes in conflict with the RAXML-B3 and SNAPP-S1 topology was low in “lineage” trees (bs: 63 for *S. crassirostris*/*S. uropygialis* clade, 77 for *S. magellanicus* Bol. Arg. sister to *S. magellanicus* Peru/*S. atratus*) and high in “species” trees (bs: 91, 98). In this analysis, *S. cucullatus* was placed sister to the ancestor of *S. crassirostris* and *S. uropygialis* (Figure 4).

3.4 | nDNA introgression tests

The formal introgression tests were based on 31 four-taxon trees, all of which were compatible with our RAXML-B3 tree. Sixteen of the 31

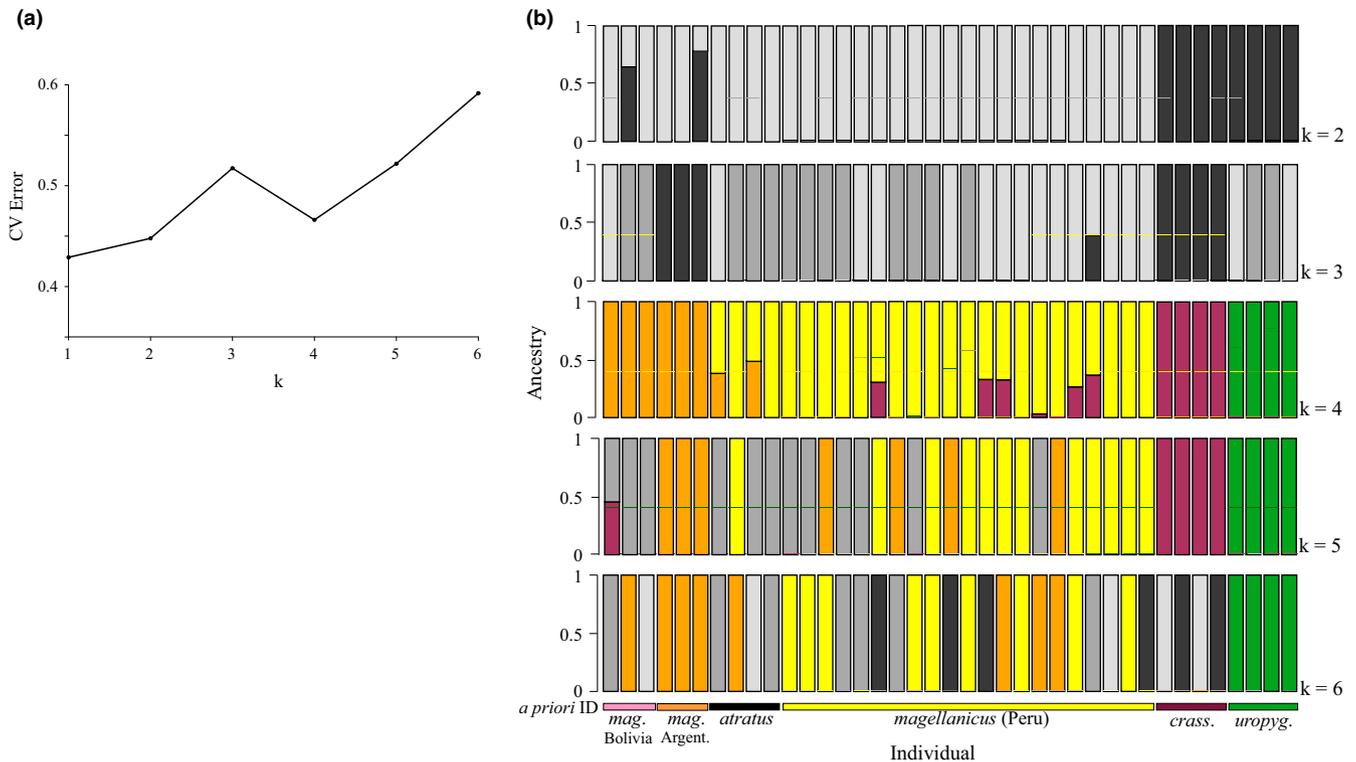


FIGURE 2 ADMIXTURE results for *Spinus* nDNA. Analysis conducted with 45,246 unlinked, variant sites assembly B3; number of populations (k) varied from 2 to 6. (a) CV error for $k = 2$ through 6; a low CV error indicates data are well described by corresponding k . (b) Individual birds are colour-coded by their a priori identification and, in the case of *S. magellanicus*, population of origin. Reconstructed populations that did not roughly correspond to a priori IDs were given grey-scale colouring [Colour figure can be viewed at wileyonlinelibrary.com]

trees were also consistent with the SNAPP-S1 tree; six were valid for the SNAPP-S2 tree and the SVDquartets tree (Table 2). With the 31 trees, we were able to conduct multiple tests for introgression between specific pairs of taxa; the number of SNPs differed among tests based on coverage per individual and variation within loci (Supporting Information Table S1).

We found strong evidence for introgression between *S. crassirostris* and Peruvian *S. magellanicus* (Table 2); this was true when *S. magellanicus* Peru (Trees 6–8) or *S. magellanicus* Central Peru (Trees 9–11) was used as the terminal taxon. Z-scores from the ABBA/BABA fixed tests for *S. crassirostris* and *S. magellanicus* from Peru ranged from 2.24 to 7.15; trees 6, 7 10 and 11 passed the FDR threshold of 0.05. The evidence for introgression was robust for three tree topologies and marginal for the fourth. The number of sites impacted the power of individual tests and may explain the low z-scores for some tests.

We also found strong evidence for introgression between *S. magellanicus* Peru and *S. uropygialis*; relevant Trees 16–30 were valid for the RAXML and SNAPP-S1 phylogenies (Table 2). The signal of introgression was almost uniformly positive (z-score > 2.0) in introgression tests based on polymorphic sites (though not all passed the FDR threshold), but mostly absent in the ABBA/BABA test based on fixed sites. Z-scores were high in tests for introgression between *S. uropygialis* and *S. magellanicus* Southern Peru and lower with all Peruvian *S. magellanicus* or *S. magellanicus* Central Peru, including some z-scores that did not pass the FDR threshold. This suggests

that introgression patterns between *S. uropygialis* and Peruvian *S. magellanicus* were spatially structured, with higher incidence in Southern Peru.

Likewise, there was modest evidence, valid for all tree topologies, that *S. magellanicus* from Southern Peru has introgressed with *S. atratus*. Introgression was well supported in 3 of 4 polymorphic site tests (Tree 31, Table 2), but not the fixed-difference ABBA/BABA test; this difference is likely a consequence of the few fixed sites available for the latter.

Evidence for introgression between *S. magellanicus* Bolivia and *S. atratus* was marginal. Tree 4, valid for the RAXML tree only, had a z-score of 2.02, but did not pass the FDR threshold. There was no support in Test 3, which was valid for RAXML and SNAPP-S1 trees.

There was no definitive evidence for introgression among the following: *S. crassirostris* and *S. magellanicus* Bolivia or Bol. Arg. (Trees 12–15); *S. atratus* and *S. uropygialis* (Trees 12 and 13); or *S. crassirostris* and *S. uropygialis* (Trees 14 and 15). Results for the former two taxon pairs were valid for the SNAPP-S1 tree; however, no tests apply to SNAPP-S2 or SVDquartets phylogenies. Further, the test results from Trees 16–30 firmly reject introgression between *S. magellanicus* Bol. Arg. and *S. magellanicus* Peru.

3.5 | Mitochondrial DNA

We sequenced the genes ND3, ND2 and *cytb* for 95 *Spinus* samples and trimmed each to 416, 777 and 753 bp, respectively. The

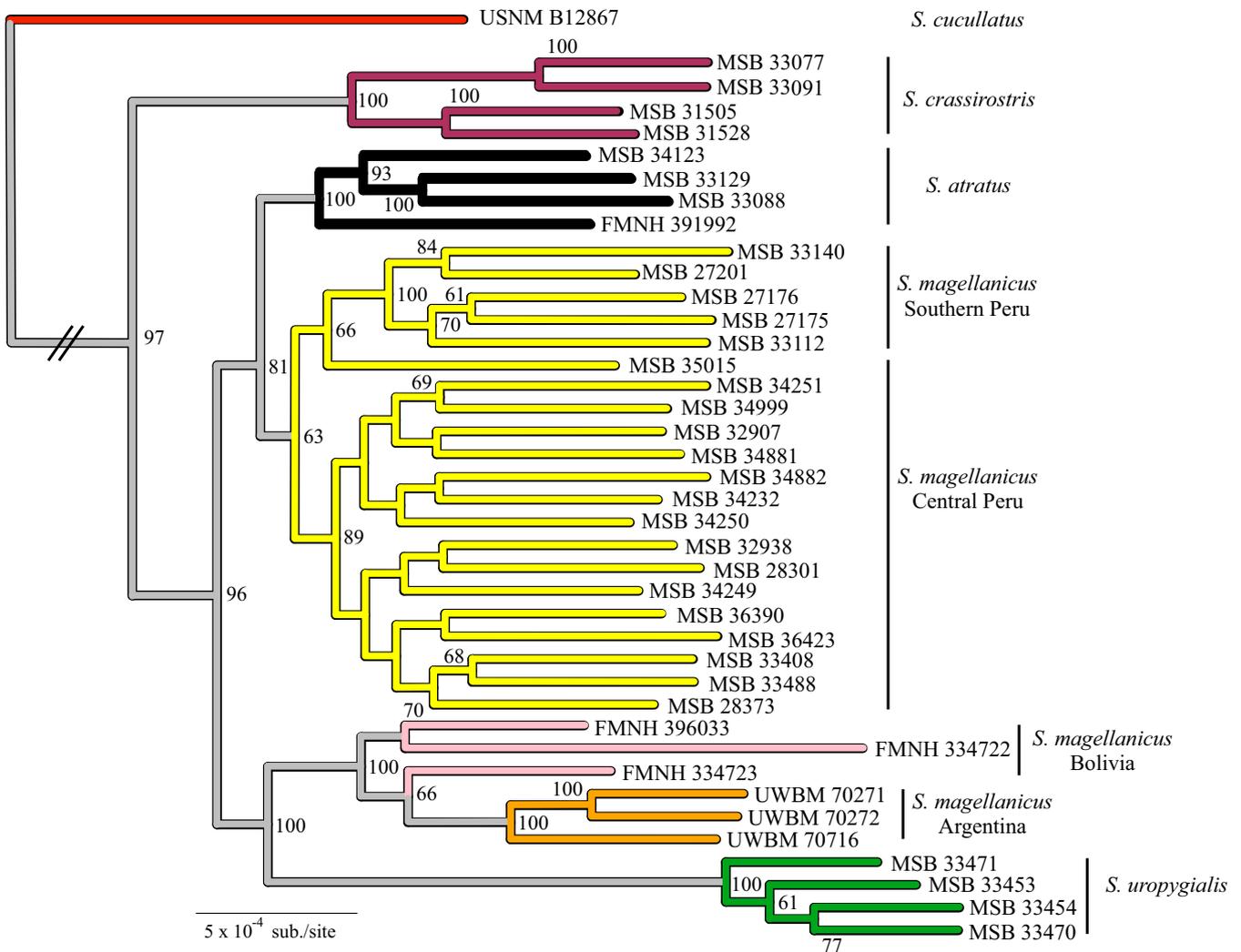


FIGURE 3 Maximum-likelihood phylogeny of *Spinus* nDNA. Tree generated in RAxML with 45,246 concatenated loci from assembly B3. Branches coloured by taxonomic units: *S. cucullatus*: red, *S. crassirostris*: maroon, *S. atratus*: black, *S. magellanicus* Peru: yellow, *S. magellanicus* Bolivia: pink, *S. magellanicus* Argentina: orange, *S. uropygialis*: green. Node support represents 500 bootstrap replicates [Colour figure can be viewed at wileyonlinelibrary.com]

concatenated mtDNA phylogeny revealed five well-supported monophyletic clades corresponding to *Spinus crassirostris* (bs 84), *S. uropygialis* (bs 60), *S. atratus* (bs 97), *S. cucullatus* (bs 100) and *S. magellanicus* from Argentina (bs 100, Figure 5a); individuals from each of these populations were found only in their respective clade. However, *S. magellanicus* from Peru and Bolivia were found in the clades corresponding to *S. crassirostris* and *S. atratus* (Figure 5a). The above results were recapitulated with ND3 alone; thus, we used ND3 to characterize 100 samples of *S. magellanicus* into *S. crassirostris*, *S. atratus* or *S. magellanicus* Argentina clades. In *S. magellanicus* individuals, the *S. crassirostris* haplotype group predominated on the western slope of the Andes in Ecuador and Peru at all elevations (Figure 5b). The *S. atratus* haplotype group increased in frequency at high elevations near Cusco, Peru, and was fixed in Cochabamba, Bolivia, at 2,470 m. In the foothills and plains southeast of the central Andes, we found only *S. magellanicus* Argentina haplotypes (Figure 5b).

4 | DISCUSSION

We documented introgression in a rapid radiation, the South American siskins, despite persistent uncertainty within the true species tree. Although we applied a suite of phylogenetic methods, some key internal nodes of the *Spinus* phylogeny remained unresolved. While considering a set of alternative phylogenetic trees, we tested for nDNA introgression among *Spinus* taxa using phylogeny-based tests for both fixed and polymorphic loci. We corroborated these results with independent population structure analyses for the nuclear and mitochondrial genomes. The preponderance of the results was consistent with multiple introgression events occurring among *Spinus* species, which likely contributed to the conflicting phylogenetic signals that we detected.

We report definitive evidence for interspecific introgression between each of the following pairs of taxa: (a) *S. magellanicus* Peru

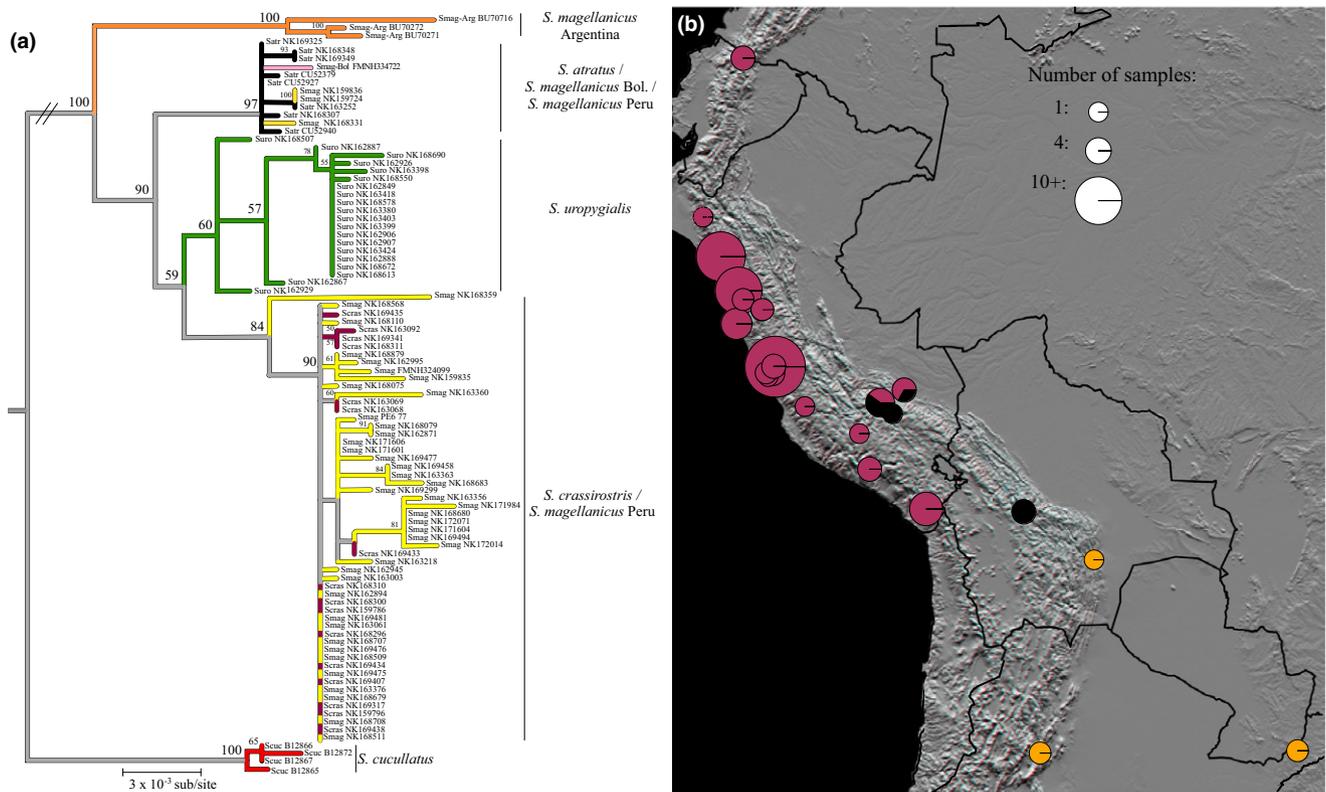


FIGURE 5 mtDNA variation within *Spinus*. (a) Maximum-likelihood phylogeny based on ND3 (416 bp), ND2 (777 bp) and *cytb* (753 bp); colours correspond to *S. crassirostris* (maroon, 16 samples), *S. uropygialis* (green, 20), *S. atratus* (black, 8), *S. cucullatus* (red, 4) and *S. magellanicus* from Peru (yellow, 43), Cochabamba, Bolivia (pink, 1), eastern Bolivia/northern Argentina (orange, 3). (b) Map of *S. magellanicus* mtDNA for 100 specimens sampled in central and western South America; clades shown in A are indicated by colour: *crassirostris* clade (maroon), *atratus* clade (black) or Argentina clade (orange) [Colour figure can be viewed at wileyonlinelibrary.com]

challenge of finding phylogenetically informative sites between recently diverged taxa; this problem may be exacerbated in phylogenetic network approaches where the initial inputs are informative, independent gene trees or their derivatives (Solís-Lemus & Ané, 2016; Wen, Yu, Hahn et al., 2016; Wen, Yu, & Nakhleh, 2016). Improvements in computational efficiency allow the network software SNAQ (Solís-Lemus & Ané, 2016) to incorporate as many as 1,000 gene trees; however, this is well below the number of loci we used here, and individual gene trees, even of longer loci, will tend to be poorly resolved in shallow divergences due to the paucity of variable sites per locus. Systematic study is needed to determine whether network analyses such as SNAQ can potentially utilize large numbers of poorly resolved gene trees. Second, we found that multiple species pairs within Andean *Spinus* have introgressed; this problem of multiple reticulation events has been reported in other rapid radiations as well (Fontaine et al., 2015; Lamichhane et al., 2015). Developing network analyses that retain the power to infer reticulation in complex scenarios involving many species is a significant computational challenge (Solís-Lemus & Ané, 2016). The ability to estimate the direction and magnitude of historic gene flow in clades that include many taxa, including parapatric and sympatric species pairs, would contribute to our understanding of how and when introgression occurs (Schumer et al., 2017; Schumer et al., 2018), as well

as how it may contribute to adaptive radiations (Kagawa & Takimoto, 2018; Lamichhane et al., 2017).

We found that sampling *S. magellanicus* across a broad geographic area was important for providing insight about how introgression has occurred among Andean *Spinus* populations. For example, we found that *S. magellanicus* from Bolivia introgressed with *S. atratus* while genetically similar *S. magellanicus* from Argentina did not. ADMIXTURE results (Figure 2, $k = 4$) revealed two *S. atratus* individuals with a high proportion of *S. magellanicus* Bolivia–Argentina ancestry, suggesting that these individuals may have resulted from recent hybridization or backcrossing. Thus, introgression with *S. atratus* may be recent and localized to Bolivia rather than an ancient reticulation event. In contrast, the introgression between *S. crassirostris* and *S. magellanicus* from Peru does not appear to be spatially structured; this could indicate that introgression has occurred across the entire extent of Peru where these lineages are sympatric or that introgression between these two lineages was ancestral. Future work comparing the genomic patterns in these two cases could reveal whether introgression in *Spinus* reflects recent interactions or ancient ones.

Analyses of mtDNA variation revealed geographic structure associated with elevation; 9 of 10 *S. magellanicus* that possessed *S. atratus*-like mtDNA were captured at elevations between 2,500 and

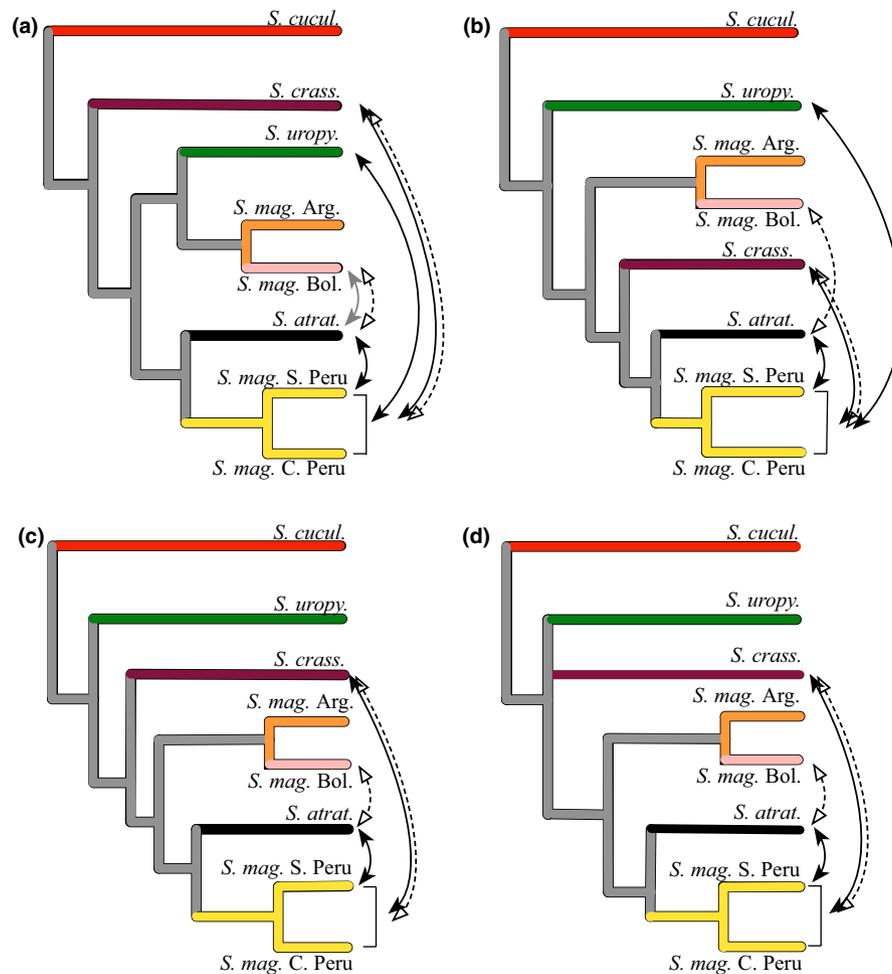


FIGURE 6 Summary of introgression results for alternative *Spinus* topologies. (a) RAxML-B3 cladogram. (b) SNAPP-S1 tree. (c) SNAPP-S2 cladogram (S2). (d) SVDquartet lineage tree; nodes with bootstrap support <70 collapsed. Bracket at tips indicates analyses with *S. magellanicus* Peru. Solid black arrows indicate significant results in formal introgression tests valid for each tree topology (see Table 2); solid grey arrows indicate significant results prior to false discovery rate correction; and dashed arrows indicate introgression detected using ADMIXTURE (see Figure 2). Formal tests (solid arrows) are only shown if the tests were valid for the respective tree topology [Colour figure can be viewed at wileyonlinelibrary.com]

3,800 m. *S. atratus* itself is restricted to elevations >3,500 m while *S. magellanicus* is found from 0 to 4,500 m (Ridgely & Tudor, 1989). The clade consisting of *S. magellanicus* from Bolivia and Argentina (including birds collected as low as 340 m) was well supported by nDNA; yet, Bolivian *S. magellanicus* from 2,550 m were fixed for the *S. atratus* mtDNA haplotype. mtDNA is a potential target for natural selection with respect to elevation due to its role in cellular respiration (Ballard & Whitlock, 2004; Blier, Breton, Desrosiers, & Lemieux, 2006; Boratyński et al., 2011; Mishmar et al., 2003), thermal tolerance (Gering, Opazo, & Storz, 2009) and the reduction of reactive oxygen species under hypoxic conditions (Scott et al., 2011). Local adaptation of mtDNA to elevation is suspected to have caused fine-scale mtDNA structure along the west slope of the Central Andes in the widespread sparrow *Zonotrichia capensis* (Chevion & Brumfield, 2009). The distribution of the *S. atratus* mtDNA haplotype in *S. magellanicus* is consistent with elevationally driven adaptive introgression. In light of the circumstantial evidence cited above and the

theoretical results of Bonnet, Leblois, Rousset, and Crochet (2017), we tentatively conclude adaptive introgression is likely explanation for mitonuclear discordance in *Spinus*.

5 | CONCLUSIONS

Significant challenges remain for rigorously identifying introgression in natural populations. Here, we employed a practical approach for evaluating phylogenetic conflict and overcoming it to achieve robust inference of introgression among closely related species; our recommendations are broadly applicable to nonmodel organisms due to the simple data requirements and the capacity to deal with phylogenetic uncertainty. Our discovery of multiple introgression events within the Andean radiation of *Spinus* siskins is consistent with an emerging paradigm that introgression tends to accompany the early stages of diversification (Dasmahapatra et al., 2012; Fontaine et al.,

2015; Grant & Grant, 2016; Martin, Davey, & Jiggins, 2015; Pease et al., 2016). Conservative inference of genetic introgression is essential because of numerous recent reports that introgressed genes can have profound consequences for trait variation (Lamichhaney et al., 2016; Richards & Martin, 2017), and diversification (Barrera-Guzmán, Aleixo, Shawkey, & Weir, 2018; Kagawa & Takimoto, 2018; Lamichhaney et al., 2017).

ACKNOWLEDGEMENTS

For tissue samples, we thank the University of Washington Burke Museum (Sharon Birks), the Smithsonian Institute (Christopher Milensky, Michael Braun); the Academy of Natural Sciences, Philadelphia (Nate Rice); the Field Museum of Natural History (David Willard, John Bates); the Cornell University Museum of Vertebrates (Kimberly Bostwick) and the Museum of Southwestern Biology, University of New Mexico, as well as the many prescient collectors. We gratefully acknowledge funding from NSF (DEB 1146491), the Frank M. Chapman Memorial Fund at the American Museum of Natural History, and the American Ornithological Society. Last, we thank Bryan Carstens, Michael Gruenstaedl, and two anonymous reviewers for their valuable comments.

DATA ACCESSIBILITY

Mitochondrial DNA sequences are available on GenBank as accessions KT221117-KT221365 and MH360743-MH360977. The raw GBS sequence data are accessible at the Nucleotide Short Read Archive, SRA Accession No.: SRP146082.

Sample localities are available in Table 1 and the Supporting Information Appendix S1; hyperlinks to Arctos database record are provided when possible.

AUTHOR CONTRIBUTIONS

E.J.B. and C.C.W. designed the study. E.J.B., P.M.B. and C.C.W. collected the samples. E.J.B. and P.M.B. sequenced the samples. E.J.B. analysed the data. P.M.B., Z.A.C. and C.C.W. advised on the analyses. E.J.B. and C.C.W. drafted the manuscript. P.M.B. and Z.A.C. improved the manuscript.

ORCID

Elizabeth J. Beckman  <http://orcid.org/0000-0002-8303-2475>

REFERENCES

- Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*, 1655–1664. <https://doi.org/10.1101/gr.094052.109>
- Alexander, A. M., Su, Y.-C., Oliveros, C. H., Olson, K. V., Travers, S. L., & Brown, R. M. (2016). Genomic data reveals potential for hybridization, introgression, and incomplete lineage sorting to confound phylogenetic relationships in an adaptive radiation of narrow-mouth frogs. *Evolution*, *85*, 10–15. <https://doi.org/10.1111/evo.13133>
- Arcila, D., Ortí, G., Vari, R., Armbruster, J., Stiasny, M., Kyung, K., ... Betancur-R, R. (2017). Genome-wide interrogation advances resolution of recalcitrant groups in the tree of life. *Nature Ecology & Evolution*, *1*, 0020. <https://doi.org/10.1038/s41559-016-0020>
- Ballard, J. W. O., & Whitlock, M. C. (2004). The incomplete natural history of mitochondria. *Molecular Ecology*, *13*, 729–744. <https://doi.org/10.1046/j.1365-294X.2003.02063.x>
- Barrera-Guzmán, A. O., Aleixo, A., Shawkey, M. D., & Weir, J. T. (2018). Hybrid speciation leads to novel male secondary sexual ornamentation of an Amazonian bird. *Proceedings of the National Academy of Sciences of the United States of America*, *115*, E218–E225. <https://doi.org/10.1073/pnas.1717319115>
- Beckman, E. J., & Witt, C. C. (2015). Phylogeny and biogeography of the New World siskins and goldfinches: Rapid, recent diversification in the Central Andes. *Molecular Phylogenetics and Evolution*, *87*, 28–45. <https://doi.org/10.1016/j.ympev.2015.03.005>
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society B*, *57*, 289–300.
- Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics*, *29*, 1165–1188.
- Blackmon, H. (2016). R package: evobIR. v 1.1.
- Blier, P. U., Breton, S., Desrosiers, V., & Lemieux, H. (2006). Functional Conservatism in mitochondrial evolution: Insight from hybridization of Arctic and Brook Charrs. *Journal of Experimental Zoology. Part B, Molecular and Developmental Evolution*, *306*, 425–432. [https://doi.org/10.1002/\(ISSN\)1552-5015](https://doi.org/10.1002/(ISSN)1552-5015)
- Bonnet, T., Leblois, R., Rousset, F., & Crochet, P. A. (2017). A reassessment of explanations for discordant introgressions of mitochondrial and nuclear genomes. *Evolution*, *000*, 1–45. <https://doi.org/10.1111/evo.13296>
- Boratyński, Z., Alves, P., Berto, S., Koskela, E., Mappes, T., & Melo-Ferreira, J. (2011). Introgression of mitochondrial DNA among Myodes voles: Consequences for energetics? *BMC Evolutionary Biology*, *11*, 355. <https://doi.org/10.1186/1471-2148-11-355>
- Bouckaert, R., & Heled, J. (2014). DensiTree 2: Seeing trees through the forest. *bioRxiv*, 12401. <https://doi.org/10.1101/012401>
- Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C., Xie, D., ... Drummond, A. J. (2014). BEAST 2: A software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, *10*, e1003537. <https://doi.org/10.1371/journal.pcbi.1003537>
- Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N. A., & 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. <https://doi.org/10.1093/molbev/mss086>
- 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. <https://doi.org/10.1111/mec.13314>
- Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., Postlethwait, J. H., & De Koning, D.-J. (2011). Stacks: Building and genotyping loci de novo from short-read sequences. *G3 (Bethesda, Md.)*, *1*, 171–182. <https://doi.org/10.1534/g3.111.000240>
- Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: An analysis tool set for population genomics. *Molecular Ecology*, *22*, 3124–3140. <https://doi.org/10.1111/mec.12354>
- Chevron, Z. A., & Brumfield, R. T. (2009). Migration-selection balance and local adaptation of mitochondrial haplotypes in Rufous-Collared Sparrows (*Zonotrichia Capensis*) along an elevational gradient. *Evolution*, *63*, 1593–1605. <https://doi.org/10.1111/j.1558-5646.2009.00644.x>
- Chifman, J., & Kubatko, L. (2014). Quartet inference from SNP data under the coalescent model. *Bioinformatics*, *30*, 3317–3324. <https://doi.org/10.1093/bioinformatics/btu530>

- Chifman, J., & Kubatko, L. (2015). Identifiability of the unrooted species tree topology under the coalescent model with time-reversible substitution processes, site-specific rate variation, and invariable sites. *Journal of Theoretical Biology*, *374*, 35–47. <https://doi.org/10.1016/j.jtbi.2015.03.006>
- Crowl, A. A., Myers, C., & Cellinese, N. (2017). Embracing discordance: Phylogenomic analyses provide evidence for allopolyploidy leading to cryptic diversity in a Mediterranean Campanula (Campanulaceae) clade. *Evolution*, *71*, 913–922. <https://doi.org/10.1111/evo.13203>
- Cui, R., Schumer, M., Kruesi, K., Walter, R., Andolfatto, P., & Rosenthal, G. G. (2013). Phylogenomics reveals extensive reticulate evolution in Xiphophorus fishes. *Evolution*, *67*, 2166–2179. <https://doi.org/10.1111/evo.12099>
- Dasmahapatra, K. K., Walters, J. R., Briscoe, A. D., Davey, J. W., Whibley, A., Nadeau, N. J., ... Jiggins, C. D. (2012). Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature*, *487*, 94–98. <https://doi.org/10.1038/nature11041>
- Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M., & Blaxter, M. L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, *12*, 499–510. <https://doi.org/10.1038/nrg3012>
- Durand, E. Y., Patterson, N., Reich, D., & Slatkin, M. (2011). Testing for ancient admixture between closely related populations. *Molecular Biology and Evolution*, *28*, 2239–2252. <https://doi.org/10.1093/molbev/msr048>
- Eaton, D. (2014). PyRAD: Assembly of *de novo* RADseq loci for phylogenetic analyses. *Bioinformatics*, *30*, 1844–1849. <https://doi.org/10.1093/bioinformatics/btu121>
- Eaton, D. A. R., Hipp, A. L., González-Rodríguez, A., & Cavender-Bares, J. (2015). Historical introgression among the American live oaks and the comparative nature of tests for introgression. *Evolution*, *69*, 2587–2601. <https://doi.org/10.1111/evo.12758>
- Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*, 1792–1797. <https://doi.org/10.1093/nar/gkh340>
- Edwards, S. V. (2009). Is a new and general theory of molecular systematics emerging? *Evolution*, *63*, 1–19. <https://doi.org/10.1111/j.1558-5646.2008.00549.x>
- Edwards, S. V., Xi, Z., Janke, A., Faircloth, B. C., McCormack, J. E., Glenn, T. C., ... Davis, C. C. (2016). Implementing and testing the multi-species coalescent model: A valuable paradigm for phylogenomics. *Molecular Phylogenetics and Evolution*, *94*, 447–462. <https://doi.org/10.1016/j.ympev.2015.10.027>
- Fjeldså, J., & Krabbe, N. K. (1990). *Birds of the high Andes*. Copenhagen, Denmark: Apollo Books.
- Fontaine, M. C., Pease, J. B., Steele, A., Waterhouse, R. M., Neafsey, D. E., Sharakhov, I. V., ... Besansky, N. J. (2015). Extensive introgression in a malaria vector species complex revealed by phylogenomics. *Science*, *347*, 1258524. <https://doi.org/10.1126/science.1258524>
- Genner, M. J., & Turner, G. F. (2012). Ancient hybridization and phenotypic novelty within Lake Malawi's Cichlid fish radiation. *Molecular Biology and Evolution*, *29*, 195–206. <https://doi.org/10.1093/molbev/msr183>
- Gerard, D., Gibbs, H. L., & Kubatko, L. (2011). Estimating hybridization in the presence of coalescence using phylogenetic intraspecific sampling. *BMC Evolutionary Biology*, *11*, 291. <https://doi.org/10.1186/1471-2148-11-291>
- Gering, E. J., Opazo, J. C., & Storz, J. F. (2009). Molecular evolution of cytochrome b in high- and low-altitude deer mice (genus *Peromyscus*). *Heredity*, *102*, 226–235. <https://doi.org/10.1038/hdy.2008.124>
- Gompert, Z., Lucas, L. K., Buerkle, C. A., Forister, M. L., Fordyce, J. A., & Nice, C. C. (2014). Admixture and the organization of genetic diversity in a butterfly species complex revealed through common and rare genetic variants. *Molecular Ecology*, *23*, 4555–4573. <https://doi.org/10.1111/mec.12811>
- Good, J. M., Vanderpool, D., Keeble, S., & Bi, K. (2015). Negligible nuclear introgression despite complete mitochondrial capture between two species of chipmunks. *Evolution*, *69*, 1961–1972. <https://doi.org/10.1111/evo.12712>
- Grant, P. R., & Grant, B. R. (2016). Introgressive hybridization and natural selection in Darwin's finches. *Biological Journal of the Linnean Society*, *117*, 812–822. <https://doi.org/10.1111/bij.12702>
- Grant, P. R., Grant, B. R., & Petren, K. (2005). Hybridization in the recent past. *The American Naturalist*, *166*, 56–67.
- Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., ... Pääbo, S. (2010). A draft sequence of the neandertal genome. *Science*, *328*, 710–722. <https://doi.org/10.1126/science.1188021>
- Harvey, M. G., Judy, C. D., Seeholzer, G. F., Maley, J. M., Graves, G. R., & Brumfield, R. T. (2015). Similarity thresholds used in short read assembly reduce the comparability of population histories across species. *PeerJ*, *3*, e864v1. <https://doi.org/10.7287/peerj.preprints.864v1>
- Jarvis, E. D., Mirarab, S., Aberer, A. J., Li, B., Houde, P., Li, C., ... Zhang, G. (2014). Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science*, *346*, 1320–1331. <https://doi.org/10.1126/science.1253451>
- Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*, *27*, 3070–3071. <https://doi.org/10.1093/bioinformatics/btr521>
- Kagawa, K., & Takimoto, G. (2018). Hybridization can promote adaptive radiation by means of transgressive segregation. *Ecology Letters*, *21*, 264–274. <https://doi.org/10.1111/ele.12891>
- Keller, I., Wagner, C. E., Greuter, L., Mwaiko, S., Selz, O. M., Sivasundar, A., ... Seehausen, O. (2013). Population genomic signatures of divergent adaptation, gene flow and hybrid speciation in the rapid radiation of Lake Victoria cichlid fishes. *Molecular Ecology*, *22*, 2848–2863. <https://doi.org/10.1111/mec.12083>
- Kubatko, L. S., Degnan, J. H., & Collins, T. (2007). Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Systematic Biology*, *56*, 17–24. <https://doi.org/10.1080/10635150601146041>
- Lamichaney, S., Berglund, J., Almén, M. S., Maqbool, K., Grabherr, M., Martinez-Barrio, A., ... Andersson, L. (2015). Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature*, *518*, 371–375. <https://doi.org/10.1038/nature14181>
- Lamichaney, S., Han, F., Berglund, J., Wang, C., Almén, M. S., Webster, M. T., ... Andersson, L. (2016). A beak size locus in Darwin's finches facilitated character displacement during a drought. *Science*, *352*, 470–474. <https://doi.org/10.1126/science.aad8786>
- Lamichaney, S., Han, F., Webster, M. T., Andersson, L., Grant, B. R., & Grant, P. R. (2017). Rapid hybrid speciation in Darwin's finches. *Science*, *359*, 224–228. <https://doi.org/10.1126/science.aao4593>
- Lanfear, R., Calcott, B., Ho, S. Y. W., & Guindon, S. (2012). PartitionFinder: Combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution*, *29*, 1695–1701. <https://doi.org/10.1093/molbev/mss020>
- Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T., & Calcott, B. (2016). PartitionFinder 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution*, *34*, 772–773.
- Leaché, A. D., Banbury, B. L., Felsenstein, J., de Oca, A., & Stamatakis, A. (2015). Short tree, long tree, right tree, wrong tree: New acquisition bias corrections for inferring SNP phylogenies. *Systematic Biology*, *64*, 1032–1047. <https://doi.org/10.1093/sysbio/syv053>
- Leaché, A. D., & Oaks, J. R. (2017). The utility of single nucleotide polymorphism (SNP) data in phylogenetics. *Annual Review of Ecology, Evolution, and Systematics*, *48*, 69–84. <https://doi.org/10.1146/annurev-ecolsys-110316-022645>
- Lewis, P. O. (2001). A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic Biology*, *50*, 913–925. <https://doi.org/10.1080/106351501753462876>

- Lewontin, R. C., & Birch, L. C. (1966). Hybridization as a source of variation for adaptation to new environments. *Evolution*, *20*, 315. <https://doi.org/10.2307/2406633>
- Linkem, C. W., Minin, V. N., & Leaché, A. D. (2016). Detecting the anomaly zone in species trees and evidence for a misleading signal in higher-level skink phylogeny (Squamata: Scincidae). *Systematic Biology*, *65*, 465–477. <https://doi.org/10.1093/sysbio/syw001>
- Liu, K. J., Dai, J., Truong, K., Song, Y., Kohn, M. H., & Nakhleh, L. (2014). An HMM-based comparative genomic framework for detecting introgression in eukaryotes. *PLoS Computational Biology*, *10*, e1003649. <https://doi.org/10.1371/journal.pcbi.1003649>
- Maddison, W. P. (1997). Gene trees in species trees. *Systematic Biology*, *46*, 523. <https://doi.org/10.2307/2413694>
- Maddison, D. R., & Maddison, W. P. (2005). *MacClade 4: Analysis of phylogeny and character evolution. Version 4.08a*. Sunderland, MA: Sinauer Associates.
- Mallet, J. (2005). Hybridization as an invasion of the genome. *Trends in Ecology and Evolution*, *20*, 229–237. <https://doi.org/10.1016/j.tree.2005.02.010>
- Mallet, J., Besansky, N., & Hahn, M. W. (2016). How reticulated are species? *BioEssays*, *38*, 140–149. <https://doi.org/10.1002/bies.201500149>
- Martin, C. H., Cutler, J. S., Friel, J. P., Dening Touokong, C., Coop, G., & Wainwright, P. C. (2015). Complex histories of repeated gene flow in Cameroon crater lake cichlids cast doubt on one of the clearest examples of sympatric speciation. *Evolution*, *69*, 1406–1422. <https://doi.org/10.1111/evo.12674>
- Martin, S. H., Davey, J. W., & Jiggins, C. D. (2015). Evaluating the use of ABBA-BABA statistics to locate introgressed loci. *Molecular Biology and Evolution*, *32*, 244–257. <https://doi.org/10.1093/molbev/msu269>
- Mastretta-Yanes, A., Arrigo, N., Alvarez, N., Jorgensen, T. H., Piñero, D., & Emerson, B. C. (2015). Restriction site-associated DNA sequencing, genotyping error estimation and de novo assembly optimization for population genetic inference. *Molecular Ecology Resources*, *15*, 28–41. <https://doi.org/10.1111/1755-0998.12291>
- McVay, J. D., Hipp, A. L., & Manos, P. S. (2017). A genetic legacy of introgression confounds phylogeny and biogeography in oaks. *Proceedings of the Royal Society B*, *284*, 20170300. <https://doi.org/10.1098/rspb.2017.0300>
- Meier, J. I., Marques, D. A., Mwaiko, S., Wagner, C. E., Excoffier, L., & Seehausen, O. (2017). Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nature Communications*, *8*, 14363. <https://doi.org/10.1038/ncomms14363>
- Melo-Ferreira, J., Seixas, F. A., Cheng, E., Mills, L. S., & Alves, P. C. (2014). The hidden history of the snowshoe hare, *Lepus americanus*: Extensive mitochondrial DNA introgression inferred from multilocus genetic variation. *Molecular Ecology*, *23*, 4617–4630. <https://doi.org/10.1111/mec.12886>
- Meyer, B. S., Matschiner, M., & Salzburger, W. (2016). Disentangling incomplete lineage sorting and introgression to refine species-tree estimates for Lake Tanganyika Cichlid fishes. *Systematic Biology*, *66*, 531–550. <https://doi.org/10.1093/sysbio/syw069>
- Miao, B., Wang, Z., & Li, Y. (2016). Genomic analysis reveals hypoxia adaptation in the Tibetan Mastiff by introgression of the grey wolf from the Tibetan Plateau. *Molecular Biology and Evolution*, *34*, 734–743. <https://doi.org/10.1093/molbev/msw274>
- Miller, M. A., Pfeiffer, W., & Schwartz, T. (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In Proceedings of the Gateway Computing Environments Workshop (GCE), 14 Nov. 2010, New Orleans, LA. 1–8.
- Mishmar, D., Ruiz-Pesini, E., Golik, P., Macaulay, V., Clark, A. G., Hosseini, S., ... Wallace, D. C. (2003). Natural selection shaped regional mtDNA variation in humans. *Proceedings of the National Academy of Sciences of the United States of America*, *100*, 171–176. <https://doi.org/10.1073/pnas.0136972100>
- Morales, A., Jackson, N., Dewey, T., O'Meara, B., & Carstens, B. (2017). Speciation with gene flow in North American *Myotis* bats. *Systematic Biology*, *66*, 440–452. <https://doi.org/10.1093/sysbio/syw100>
- Novembre, J., & Stephens, M. (2008). Interpreting principal component analyses of spatial population genetic variation. *Nature Genetics*, *40*, 646–649. <https://doi.org/10.1038/ng.139>
- Novikova, P. Y., Hohmann, N., Nizhynska, V., Tsuchimatsu, T., Ali, J., Muir, G., ... Nordborg, M. (2016). Sequencing of the genus *Arabidopsis* identifies a complex history of nonbifurcating speciation and abundant trans-specific polymorphism. *Nature Genetics*, *48*, 1077–1082. <https://doi.org/10.1038/ng.3617>
- Parchman, T. L., Gompert, Z., Mudge, J., Schilkey, F. D., Benkman, C. W., & Buerkle, C. A. (2012). Genome-wide association genetics of an adaptive trait in lodgepole pine. *Molecular Ecology*, *21*, 2991–3005. <https://doi.org/10.1111/j.1365-294X.2012.05513.x>
- Parsons, K. J., Son, Y. H., & Albertson, R. C. (2011). Hybridization promotes evolvability in African Cichlids: Connections between transgressive segregation and phenotypic integration. *Evolutionary Biology*, *38*, 306–315. <https://doi.org/10.1007/s11692-011-9126-7>
- Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., ... Reich, D. (2012). Ancient admixture in human history. *Genetics*, *192*, 1065–1093. <https://doi.org/10.1534/genetics.112.145037>
- Payseur, B. A., & Rieseberg, L. H. (2016). A genomic perspective on hybridization and speciation. *Molecular Ecology*, *25*, 2337–2360. <https://doi.org/10.1111/mec.13557>
- Pease, J. B., Haak, D. C., Hahn, M. W., & Moyle, L. C. (2016). Phylogenomics reveals three sources of adaptive variation during a rapid radiation. *PLOS Biology*, *14*, e1002379. <https://doi.org/10.1371/journal.pbio.1002379>
- Pease, J. B., & Hahn, M. W. (2015). Detection and polarization of introgression in a five-taxon phylogeny. *Systematic Biology*, *64*, 651–662. <https://doi.org/10.1093/sysbio/syv023>
- Pfeiffer, W., & Stamatakis, A. (2010). Hybrid mpi/pthreads parallelization of the raxml phylogenetics code. In *Parallel & Distributed Processing, Workshops and Phd Forum (IPDPSW)*, 2010 IEEE International Symposium On, 208.
- Pickrell, J. K., & Pritchard, J. K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics*, *8*, e1002967. <https://doi.org/10.1371/journal.pgen.1002967>
- R Core Team (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.r-project.org/>
- Rambaut, A., Suchard, M., Xie, D., & Drummond, A. J. (2014). *Tracer v1.6*. Retrieved from <http://beast.bio.ed.ac.uk/Tracer>
- Reaz, R., Bayzid, M. S., & Rahman, M. S. (2014). Accurate phylogenetic tree reconstruction from quartets: A heuristic approach. *PLoS One*, *9*, e0104008. <https://doi.org/10.1371/journal.pone.0104008>
- Ree, R. H., & Hipp, A. L. (2015). Inferring phylogenetic history from restriction site associated DNA (RADseq). In E. Hörandl & M. S. Appelhans (Eds.), *Regnum vegetabile* (pp. 181–204). Next-Generation Sequencing in Plant Systematics. Königstein, Germany: International Association for Plant Taxonomy, IAPT.
- Reich, D., Green, R. E., Kircher, M., Krause, J., Patterson, N., Durand, E. Y., ... Pääbo, S. (2010). Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature*, *468*, 1053–1060. <https://doi.org/10.1038/nature09710>
- Reich, D., Thangaraj, K., Patterson, N., Price, A. L., & Singh, L. (2009). Reconstructing Indian population history. *Nature*, *461*, 489–494. <https://doi.org/10.1038/nature08365>
- Rheindt, F. E., & Edwards, S. V. (2011). Genetic introgression: An Integral but neglected component of speciation in birds. *The Auk*, *128*, 620–632. <https://doi.org/10.1525/auk.2011.128.4.620>
- Rheindt, F. E., Fujita, M. K., Wilton, P. R., & Edwards, S. V. (2014). Introgression and phenotypic assimilation in zimmerius flycatchers (Tyrannidae): Population genetic and phylogenetic inferences from genome-

- wide SNPs. *Systematic Biology*, 63, 134–152. <https://doi.org/10.1093/sysbio/syt070>
- Richards, E. J., & Martin, C. H. (2017). Adaptive introgression from distant Caribbean islands contributed to the diversification of a microendemic adaptive radiation of trophic specialist pupfishes. *PLoS Genetics*, 13, e1006919. <https://doi.org/10.1371/journal.pgen.1006919>
- Ridgely, R. S., & Tudor, G. (1989). *The birds of South America* (Vol. I). Austin, TX: University of Texas Press.
- Schumer, M., Powell, D. L., Delclós, P. J., Squire, M., Cui, R., Andolfatto, P., & Rosenthal, G. G. (2017). Assortative mating and persistent reproductive isolation in hybrids. *Proceedings of the National Academy of Sciences of the United States of America*, 114, 10936–10941. <https://doi.org/10.1073/pnas.1711238114>
- Schumer, M., Xu, C., Powell, D. L., Durvasula, A., Skov, L., Holland, C., ... Przeworski, M. (2018). Natural selection interacts with recombination to shape the evolution of hybrid genomes. *Science*, 360, 656–660.
- Scott, G. R., Schulte, P. M., Egginton, S., Scott, A. L. M., Richards, J. G., & Milsom, W. K. (2011). Molecular evolution of cytochrome c oxidase underlies high-altitude adaptation in the bar-headed goose. *Molecular Biology and Evolution*, 28, 351–363. <https://doi.org/10.1093/molbev/msq205>
- Seehausen, O. (2004). Hybridization and adaptive radiation. *Trends in Ecology and Evolution*, 19, 198–207. <https://doi.org/10.1016/j.tree.2004.01.003>
- Shen, X., Hittinger, C. T., & Rokas, A. (2017). Contentious relationships in phylogenomic studies can be driven by a handful of genes. *Nature Ecology & Evolution*, 1, 126. <https://doi.org/10.1038/s41559-017-0126>
- Smith, S. A., Moore, M. J., Brown, J. W., & Yang, Y. (2015). Analysis of phylogenomic datasets reveals conflict, concordance, and gene duplications with examples from animals and plants. *BMC Evolutionary Biology*, 15, 150. <https://doi.org/10.1186/s12862-015-0423-0>
- Solís-Lemus, C., & Ané, C. (2016). Inferring phylogenetic networks with maximum pseudolikelihood under incomplete lineage sorting. *PLoS Genetics*, 12, e1005896. <https://doi.org/10.1371/journal.pgen.1005896>
- Solís-Lemus, C., Yang, M., & Ané, C. (2016). Inconsistency of species tree methods under gene flow. *Systematic Biology*, 65, 843–851. <https://doi.org/10.1093/sysbio/syw030>
- Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30, 1312–1313.
- Streicher, J. W., Devitt, T. J., Goldberg, C. S., Malone, J. H., Blackmon, H., & Fujita, M. K. (2014). Diversification and asymmetrical gene flow across time and space: Lineage sorting and hybridization in polytypic barking frogs. *Molecular Ecology*, 23, 3273–3291. <https://doi.org/10.1111/mec.12814>
- Suh, A. (2016). The phylogenomic forest of bird trees contains a hard polytomy at the root of Neoaues. *Zoologica Scripta*, 45, 50–62. <https://doi.org/10.1111/zsc.12213>
- Suh, A., Smeds, L., & Ellegren, H. (2015). The dynamics of incomplete lineage sorting across the ancient adaptive radiation of neoavian birds. *PLoS Biology*, 13, e1002224. <https://doi.org/10.1371/journal.pbio.1002224>
- Swofford, D. (2002). PAUP*. *Phylogenetic analysis using parsimony (* and Other Methods)*. Version 4.0a147. Sunderland, MA: Sinauer Associates.
- Than, C., Ruths, D., & Nakhleh, L. (2008). PhyloNet: A software package for analyzing and reconstructing reticulate evolutionary relationships. *BMC Bioinformatics*, 9, 322. <https://doi.org/10.1186/1471-2105-9-322>
- Thom, G., Amaral, F. R., Hickerson, M. J., Aleixo, A., Araujo-Silva, L. E., Ribas, C. C., ... Miyaki, C. Y. (2018). Phenotypic and genetic structure support gene flow generating gene tree discordances in an Amazonian floodplain endemic species. *Systematic Biology*, 67, 700–718. <https://doi.org/10.1093/sysbio/syy004>
- Vargas, O. M., Ortiz, E. M., & Simpson, B. B. (2017). Conflicting phylogenomic signals reveal a pattern of reticulate evolution in a recent high-Andean diversification (Asteraceae: Astereae: *Diplostegium*). *New Phytologist*, 214, 1736–1750. <https://doi.org/10.1111/NPH.14530>
- Wallbank, R. W. R., Baxter, S. W., Pardo-Díaz, C., Hanly, J. J., Martin, S. H., Mallet, J., ... Jiggins, C. D. (2016). Evolutionary novelty in a butterfly wing pattern through enhancer shuffling. *PLoS Biology*, 14, e1002353. <https://doi.org/10.1371/journal.pbio.1002353>
- Wen, D., Yu, Y., Hahn, M. W., & Nakhleh, L. (2016). Reticulate evolutionary history and extensive introgression in mosquito species revealed by phylogenetic network analysis. *Molecular Ecology*, 25, 2361–2372. <https://doi.org/10.1111/mec.13544>
- Wen, D., Yu, Y., & Nakhleh, L. (2016). Bayesian inference of reticulate phylogenies under the multispecies network coalescent. *PLoS Genetics*, 12, e1006006. <https://doi.org/10.1371/journal.pgen.1006006>
- Winger, B. M. (2017). Consequences of divergence and introgression for speciation in Andean cloud forest birds. *Evolution*, 71, 1815–1831. <https://doi.org/10.1111/evo.13251>
- Yu, Y., Degnan, J. H., & Nakhleh, L. (2012). The probability of a gene tree topology within a phylogenetic network with applications to hybridization detection. *PLoS Genetics*, 8, e1002660. <https://doi.org/10.1371/journal.pgen.1002660>
- Yu, Y., Dong, J., Liu, K. J., & 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. <https://doi.org/10.1073/pnas.1407950111>
- Yu, Y., Than, C., Degnan, J. H., & Nakhleh, L. (2011). Coalescent histories on phylogenetic networks and detection of hybridization despite incomplete lineage sorting. *Systematic Biology*, 60, 138–149. <https://doi.org/10.1093/sysbio/syq084>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Beckman EJ, Benham PM, Cheviron ZA, Witt CC. Detecting introgression despite phylogenetic uncertainty: The case of the South American siskins. *Mol Ecol*. 2018;27:4350–4367. <https://doi.org/10.1111/mec.14795>