Fall 12-15-2016

BIOGEOGRAPHY, INTERSPECIFIC INTROGRESSION, AND THE EVOLUTION OF HEMOGLOBIN GENES IN THE HIGH ANDES: THE EVOLUTIONARY HISTORY OF THE SOUTH AMERICAN SISKINS (Spinus)

Elizabeth Jane Beckman
University of New Mexico

Elizabeth J. Beckman
Candidate

Biology
Department

This dissertation is approved, and it is acceptable in quality and form for publication:

Approved by the Dissertation Committee:

Christopher C. Witt , Chairperson

Jeffrey C. Long

Michael J. Andersen

Zachary A. Cheviron

BIOGEOGRAPHY, INTERSPECIFIC INTROGRESSION AND THE EVOLUTION OF HEMOGLOBIN GENES IN THE HIGH ANDES: THE EVOLUTIONARY HISTORY OF THE SOUTH AMERICAN SISKINS (SPINUS)

by

ELIZABETH J. BECKMAN

B.S., Biology, Harvey Mudd College, 2006

DISSERTATION

Submitted in Partial Fulfillment of the
Requirements for the Degree of

Doctor of Philosophy
Biology

The University of New Mexico
Albuquerque, New Mexico

December, 2016

ACKNOWLEDGEMENTS

This work would not have been possible without the field efforts of many individuals from the Witt lab, the Museum of Southwestern Biology, and the Centro de Ornitología y Biodiversidad (CORBIDI) in Peru. I'd like to thank Phred Benham, Natalie Wright, Shane DuBay, Spencer Galen, C. Jonathan Schmitt, Andrew Johnson, Emil Bautista O., Thomas Valquí, Monica Flores and Walter Vargas C. for their excellent contributions to field expeditions. I gratefully acknowledge the generous donation of tissue loans from the following institutions: the Burke Museum; the Museum of Vertebrate Zoology at Berkeley, the Cornell University Museum of Vertebrates, the Biodiversity Institute and Natural History Museum, University of Kansas, the American Museum of Natural History; the Smithsonian Institution; the Field Museum of Natural History; the Academy of Natural Sciences, Philadelphia and the Louisiana State University Museum of Natural Science.

I would like to thank the members of my committee, Dr. Jeff Long, Dr. Zac Cheviron, and Dr. Mike Andersen for their support, advice and thoughtfulness through the years. I'd also like to thank Bryan McClean and Kayce Bell as well as the other students mentioned above for the many hours of productive discussion. My parents and my partner, Phred Benham, have been indefatigable supporters during my pursuit of this degree.

Lastly, I'd like to thank my advisor, Chris Witt, for his energy, high standards and patience. Chris's insight and enthusiasm for everything related to birds turned each setback into an opportunity, and revealed the true value of each discovery.

BIOGEOGRAPHY, INTERSPECIFIC INTROGRESSION, AND THE EVOLUTION OF HEMOGLOBIN GENES IN THE HIGH ANDES: THE EVOLUTIONARY HISTORY OF THE SOUTH AMERICAN SISKINS (*Spinus*)

by

Elizabeth J. Beckman

B.S., Biology, Harvey Mudd College, 2006

Ph.D., Biology, University of New Mexico, 2016

ABSTRACT

Landscape features, interspecific introgression, and adaptation work in concert to shape the evolutionary history of a clade. Understanding the independent and cumulative consequences of these evolutionary processes on diversification is critical to revealing the origins of extant biodiversity. Studying these processes within rapid radiations, a significant contributor to global biodiversity, can provide powerful insight into the process of diversification. To assess how diversification is shaped by these evolutionary forces, I examined the biogeographic history, patterns of interspecific introgression and adaptation to high elevation in a recent, rapid radiation of finches, the South American siskins (Fringillidae: *Spinus*). I found that this continental radiation colonized South America from North America and subsequently diversified at an exceptional rate in the high Andes. Further, my results show that sympatric siskin species within the high Andes form a monophyletic clade. I hypothesized that the close proximity of near relatives at high elevation could challenge species limits in *Spinus*. I investigated this hypothesis

using a genome-wide SNP dataset to construct phylogenetic trees and performed formal tests of introgression among high elevation species. I developed an approach for assessing introgression despite persistent phylogenetic uncertainty, and discovered evidence for multiple introgressive events among different high elevation *Spinus* species. Cold temperatures and decreased partial pressure of oxygen are chronic stressors on organisms living at high elevation. Finally, to understand the consequences of high elevation on adaptive divergence in *Spinus* , I sequenced all genes which encode the oxygen-transport protein hemoglobin across the *Spinus* clade and among several populations of a species with a wide elevational range. I identified multiple instances of non-synonymous mutations at the inter- and intra-specific level in both adult and embryonic hemoglobin proteins. These patterns of genetic variation within functionally significant loci across elevation suggest that hemoglobin genes have had a significant impact on adaptation and potentially diversification within the South American siskins.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter I:

Chapter II:

Chapter III:

LIST OF TABLES

Chapter I:

Chapter II:

Chapter III:

CHAPTER I: Phylogeny and biogeography of the New World siskins and goldfinches: Rapid, recent diversification in the Central Andes

ABSTRACT

Time-calibrated molecular phylogenies can help us to understand the origins of the diverse and unique Andean avifauna. Previous studies have shown that the tempo of diversification differed between the Andes and adjacent lowland regions of South America. Andean taxa were found to have speciated more recently and to have avoided the decelerated diversification that is typical of Neotropical lowland clades. The South American siskins, a Pleistocene finch radiation, may typify this Andean pattern. We investigated the phylogenetic biogeography of all the New World siskins and goldfinches in new detail. To understand the specific role of the Andes in siskin diversification, we asked: **(1)** Was diversification faster in Andean siskin lineages relative to non-Andean ones? **(2)** Did siskin lineages move into and out of the Andes at different rates? We found that siskin lineages in the Andes had higher diversification rates and higher outward dispersal rates than siskin lineages outside the Andes. We conclude that páramo expansion and contraction in response to Pleistocene climatic cycles caused accelerated diversification and outward dispersal in Andean siskins. The younger average age of bird species in the Andes compared to lowland South America may be attributable to bursts of recent diversification in siskins and several other vagile, open-habitat clades.

INTRODUCTION

The Andes of South America are a region of exceptional avian biodiversity and endemism (Myers et al., 2000; Orme et al., 2005). Dated molecular phylogenies from

birds and other taxa support the hypothesis that Andean biodiversity is primarily the result of allopatric speciation facilitated by orogenesis and climitalogical shifts since the mid-Neogene (Rull, 2008; Rull, 2011; Turchetto-Zolet et al., 2013). Several meta-analyses of Neotropical bird diversification patterns suggested that avian lineages experienced different tempos of diversification in the Andes and the adjacent lowlands (Fjeldså, 1994; Hawkins et al., 2006, 2011; Hoorn et al., 2010; Weir, 2006). Andean bird species tend to have speciated more recently than lowland species (Fjeldså, 1994; Hawkins et al., 2006, 2011), and lowland bird clades typically show a signature of significant deceleration in diversification rate while most Andean clades do not (Weir, 2006).

There are many potential explanations for the differences between high-elevation and low-elevation diversification patterns in South America (Rull, 2011). In birds, where the predominant mode of speciation is allopatric (Coyne and Price, 2000), the difference may ultimately reflect regional disparities in the opportunities for vicariance. High elevation taxa tend to have linear geographic ranges that may be prone to fragmentation (Graves, 1988). Alternatively, lowland species are more likely to be impacted by marine incursions due to the rise of sea levels resulting from climatic variation (Brumfield and Caparella, 1996; Nores, 2004). Climatic cycles during the Pleistocene comprise one of the principal vicariant causes of bird diversification in South America (Bush et al., 2007; Graves, 1985; Gutiérrez-Pinto et al., 2012; Weir, 2006). Palynological evidence from the last 2.8 million years demonstrates that these cycles caused dramatic shifts in habitat distributions across South America (Bush et al., 2007; Clapperton, 1993; Hooghiemstra and van der Hammen, 2004; Servant et al., 1993). Montane habitat moved up and down

slope (Hooghiemstra and van der Hammen, 2004; van der Hammen and Cleef, 1986) and spread widely into the lowlands (Bush and Oliveira, 2006; Bush et al., 2007). This likely resulted in episodes of expansion, contraction, and isolation for high elevation species. In the lowlands, glacial cycles influenced temperature, humidity and precipitation, but these fluctuations varied across the landscape (Bush et al., 2007). Shifting climates and the invasion of montane species in some regions likely caused range retractions for lowland species with unpredictable consequences for biotic interactions, vicariance, and extinction (Bush et al., 2007). The differences in the direct impacts of Pleistocene glacial cycles on high and low elevation taxa may have resulted in different regional tempos of diversification and inter-regional dispersal patterns.

Weir (2006) found a significant increase in average diversification rate of Andean birds during the last one million years, coincident with increased frequency and severity of glacial cycles (Hooghiemstra and van der Hammen, 2004). In contrast, the average diversification rate of lowland taxa remained static (Weir, 2006). Weir concluded that the impact of Pleistocene glacial cycles on avian diversification was greater in the Andes than in the tropical lowlands. The high average diversification rate of Andean taxa in this study was principally based on the inclusion of three rapid, recent radiations: *Cranioleuca* spinetails, *Muscisaxicola* ground-tyrants and *Spinus* siskins (formerly *Carduelis* or *Sporagra*). For continentally-distributed *Spinus* siskins, however, the phylogeny employed (van den Elzen et al., 2001) had sparse geographic sampling and poor resolution; further, the molecular dating method applied by Weir was not ideal for recent divergences (García-Moreno et al., 2004; Ho et al., 2005). Thus, increasing the geographic sampling and phylogenetic resolution for the South American siskin radiation

should provide insight into the tempo of avian Pleistocene diversification within and outside the Andes.

Biogeographic analysis of the South American siskin radiation should also provide perspective on inter-regional dispersal between the Andes and adjacent lowlands. In the core tanagers (Thraupini), a clade that diversified extensively in the Andes during the Miocene and Pleistocene, the Northern Andes had a significantly higher number of outward dispersal events than other regions of the Andes or tropical lowlands, including several key transitions from high to low elevations (Sedano and Burns, 2010). These results suggest that dispersal out of the Andes may contribute significantly to lowland biodiversity. However, in hummingbirds, another clade that proliferated in the Andes since the Miocene, there was little evidence that secondary colonization of the lowlands substantially impacted diversity (McGuire et al., 2014).

To further elucidate regional patterns of Pleistocene diversification and dispersal, we analyzed the evolutionary and biogeographic history of the South American siskins and their relatives, a monophyletic clade in the family Fringillidae that includes all of the New World goldfinches and siskins plus the Eurasian siskin (Arnaiz-Villena et al., 1998; van den Elzen et al., 2001; Zuccon et al., 2012). We follow the North American Checklist Committee (NACC) by treating all siskins in the Western hemisphere as *Spinus* (Chesser et. al., 2009). We refer to the North American goldfinches as *Astragalinus* as proposed by Remsen (2011), and Chesser and Rising (2013); previous phylogenetic results indicate this clade is sister to all other *Spinus* (Arnaiz-Villena et al., 2007; Arnaiz-Villena et al., 2007; Zuccon et al., 2012). A secondary goal of this study is to evaluate the validity of recently proposed generic changes (Chesser et al., 2009; Chesser and Rising, 2013;

Remsen, 2011), including the genus name *Sporagra* for the South American siskins (Remsen et al., 2014).

South American *Spinus* occupy humid and dry open habitats including fields, scrub and forest edges (Fjeldså and Krabbe, 1990; Ridgely and Tudor, 1980). Siskin species richness peaks at high elevations in the Peruvian Andes where three species (*S. magellanicus, S. atratus* and *S. crassirostris*) breed sympatrically and are joined by a fourth, *S. uropygialis*, during its non-breeding season (Fjeldså and Krabbe, 1990). Of the nine Andean siskins, four (*S. atratus, S. crassirostris, S. siemiradzkii, S. uropygialis*) occur only in the Central Andes and Austral Andes; one, *S. barbatus,* is restricted to the Chilean Andes and Patagonia, and two, *S. spinescens* and *S. xanthogastrus,* reside in the Northern Andes. *Spinus xanthogastrus* extends into the Central Andes and the Talamanca mountains of Costa Rica and Panama too. *Spinus magellanicus* and *S. olivaceus* occur in both the Northern and Central Andes; *S. magellanicus* is also found at low-mid elevations in Guyana, northern Argentina, and southern Brazil. Of the two non-Andean *Spinus* in South America, *S. cucullatus* lives at mid-elevations in the coastal mountains of Venezuela and the Guianan highlands, and *S. yarrellii* is restricted to low elevations in eastern Brazil.

In this study, we asked: **(1)** Was diversification faster in Andean siskin lineages relative to non-Andean ones? **(2)** Did siskin lineages move into and out of the Andes at different rates? To answer these questions, we estimated a time-calibrated phylogeny of the New World siskins and goldfinches. We used this phylogeny to calculate region-specific diversification rates and inter-regional dispersal rates between the Andes and adjacent lowlands.

MATERIAL AND METHODS

Taxonomic sampling

We obtained tissue samples for ten of eleven South American *Spinus* species from museum tissue collections (Fig. 1, Appendix). Species were represented by one to five individuals with the exception of the widespread, polytypic *Spinus magellanicus* for which we included 28 individuals representing eight geographic populations (Fig. 1). We also obtained tissues for 1-3 individuals of each of the five Central American, Caribbean, and North American *Spinus* species and each of the three species of North American goldfinches (*Astragalinus*). Sequences for *Spinus spinus*, the Eurasian siskin, were obtained from Genbank. We rooted the *Astragalinus*-*Spinus* clade using original sequences from *Coccothraustes vespertinus* and Genbank sequences of *Serinus canaria* that comprised a chimera of sequence fragments from three different individuals. At the time of sequencing, no tissue was available for *Spinus yarrellii* or *Spinus pinus perplexus*, an evolutionarily independent lineage (B. Milá, pers. comm.); however, a single cytochrome *b* sequence (*cytb)* was deposited for each taxa on Genbank (Supplemental Appendix 1). We excluded these taxa from our full phylogenetic analyses, and we subsequently utilized the published *cytb* sequences to determine their placement in our final phylogeny *a posteriori* (see Section 2.4).

DNA sequencing

We extracted DNA from muscle tissue samples with the Qiagen DNeasy Blood and Tissue kit (Qiagen, Valencia, CA) using the standard protocol. For 68 individuals

total, we amplified and sequenced three mitochondrial protein-coding genes (*cytb*, NADH dehydrogenase II, and NADH dehydrogenase III) as well as one Z-linked intron (intron 2 of muscle-specific kinase, MUSK) and one autosomal intron (intron 2 of myoglobin). DNA fragments were amplified in 15 μl reactions consisting of 15mM Tris/HCl, 50mM KCl, 2.5 mM $MgCl_2$, 0.025 U/μl AmpliTaq Gold, 200 μM of each dNTP and 0.5 μM each of locus-specific primers (Table 1) with 1 μl extracted genomic DNA. Polymerase chain reactions ran under the following conditions: an initial 8 minute step at 95°C, 35 cycles of 95°C for 30 seconds, a locus-specific annealing temperature (Table 1) for 30 seconds and 72°C for 1 minute, with a final extension of 10 minutes at 72°C. For difficult to amplify samples, we designed *Spinus* specific ND2 primers using Primer3 (Rozen and Skaletsky, 2000). We cleaned the PCR products with ExoSAP-IT (Affymetrix, Santa Clara, CA) to eliminate unused reagents and gel-purified them with Sephadex (Sigma-Aldrich, St. Louis, MO). Amplified DNA was sequenced with Big Dye Terminator 3.1 sequencing reaction mix (Applied Biosystems, Grand Island, NY) on ABI 3130 or ABI 377 Automated DNA Sequencers at the University of New Mexico Molecular Biology Facility (Albuquerque, NM).

We evaluated all sequences manually in Sequencher 4.10.1 (Gene Codes Corp., Ann Arbor, MI) for quality. Sequences for each locus were aligned first with Muscle 3.7 (Edgar, 2004) on the CIPRES portal (Miller et al., 2010) and reviewed in MacClade 4.08 (Maddison and Maddison, 2005). We considered double peaks in high quality nuclear loci heterozygotes and used PHASE (Stephens and Donnelly, 2003) to assign diploid haplotypes. Haplotypes with greater than 0.005 pairwise genetic distance across all five loci were included as unique sequences, denoted samples A and B of a single individual.

We tested for neutrality for each mitochondrial gene using Tajima's D (Tajima, 1996) and Fu's F_S (Fu, 1996) in Arlequin 3.5 (Excoffier et al., 2005); for these tests, we grouped individuals by species and by region within the widespread species, *S. magellanicus*. We also constructed maximum likelihood phylogenetic trees in MEGA 5 (Tamura et al., 2011) using second and third codon positions, respectively, in *cytb* to test the phylogenetic signal of non-synonymous and synonymous mutations.

Partitioning and Model Selection

Incorrect partitioning for genetic datasets may have major phylogenetic consequences (Lemmon & Moriarty, 2004; McGuire et al., 2007). Under-parameterization may result in "long branch attraction" (Huelsenbeck, 1995) while over-parameterization may result in increased variance and improper posterior distributions and priors (McGuire et al., 2007). We used the program PartitionFinder (Lanfear et al., 2012) to select the partition scheme and evolutionary models for our mitochondrial genes. We divided our mitochondrial genes by codon, gene and individual flanking tRNAs and performed a "greedy" search where branch length estimates were unlinked. The "greedy" search implements a heuristic search algorithm which allows the program to effectively explore the likelihood space without an exhaustive, computationally prohibitive search to select the most likely partition and model scheme (Lanfear et al., 2012). We performed two independent trials of the analysis. We assessed the best fit between the data and partitioning/modeling schemes through the Bayesian Inference Criteria (BIC) as recommended by Lanfear et al. (2012). We designated MUSK and MYO2 introns as independent partitions since they both occur in the nuclear genome and

are found on separate chromosomes. Evolutionary models for each nuclear locus were evaluated in MEGA 5 (Tamura et al., 2011) using BIC.

We used the selected partitioning and model schemes throughout all subsequent analysis. When the exact methodology was not available due to different phylogenetic program limitations, we favored over-parameterization to under-parameterization as in McGuire et al. (2007).

Phylogenetics

We constructed phylogenetic trees of each independent locus in MrBayes (Ronquist and Huelsenbeck, 2003) and RAxML 7.3.1 (Stamatakis, 2006). Mitochondrial genes were concatenated and treated as a single locus after preliminary analysis showed strong concordance among gene trees. We also conducted an independent Bayesian analysis of *cytb* including *S. yarrellii* and *S. pinus perplexus* (Supplemental Appendix 1). All MrBayes analyses were performed with four chains (1 cold, 3 incrementally heated) on four independent runs and returned a strict bifurcating consensus tree. Convergence and appropriate burnin values were assessed with AWTY (Nylander et al., 2008) and Tracer 1.5 (Rambaut & Drummond, 2009). MUSK and MYO2 Bayesian trees were each run for 30 million generations with a 10% burnin. The mitochondrial phylogeny was run for 40 million generations with a 25% burnin. We also concatenated MUSK, MYO2 and mtDNA, partitioned, and ran the analysis for 40 million generations with a 10% burnin.

We used RAxML 7.3.1 to build maximum likelihood trees for each independent locus (MUSK, MYO2, mtDNA) and the concatenated dataset as well. We used the rapid bootstrapping function (1000 replicates) with the model of evolution set to

GTRGAMMA as recommended for relatively small number of sequences (Stamatakis, 2006). The concatenated dataset was analyzed with joint branch optimization to minimize the impact of missing loci on branch lengths.

Divergence Date Estimation

We estimated divergence dates using the gene-specific rates of Lerner et al. (2011) which were based on geological time calibrations in the Hawaiian Honeycreepers (Fringillidae: Drepanidinae): *cytb* 0.014 subs/site/Ma, ND3 0.024 s/s/Ma & ND2 0.029 s/s/Ma. We calculated divergence dates with mitochondrial DNA alone using BEAST and all loci using the species tree method (*BEAST).

In the BEAST analysis (Drummond and Rambaut, 2007), we used unlinked, relaxed lognormal molecular clocks for each mitochondrial gene following Drummond et al. (2007), however we linked the substitution models and trees. The BEAST analysis included for 3 independent runs of 200 million generations each with 25% burnin; results were combined in LogCombiner 1.7.1 and convergence was evaluated in Tracer 1.5 (Rambaut & Drummond, 2009).

Species tree analysis in *BEAST (Heled and Drummond, 2010) uses coalescent analysis to evaluate the probability of independent gene histories in the evolution of species and populations (Edwards, 2009). This analysis is significantly improved by the inclusion of multiple individuals per species or population (Heled and Drummond, 2010); accordingly, we strove to include at least two individuals per species in the *BEAST dataset. *BEAST requires species or populations be designated *a priori*; we assigned each species as an operational taxonomic unit with the exception of *Spinus magellanicus*.

Spinus magellanicus has several recognized subspecies and a large range that includes diverse habitats and spans a major physical barrier, the Andes (Ridgely and Tudor, 1989). We used geographic populations of *S. magellanicus*, guided by subspecies boundaries, as independently evolving units in our *BEAST analysis. We used the methodology outlined above for BEAST for the mitochondrial genes, and we added MUSK and MYO2 to the analysis. The substitution, clock and tree models were unlinked for MUSK, MYO2 and mtDNA; nuclear clock rates were estimated under strict molecular clock models following the method of Drummond et al. (2007) and were calculated relative to the known mitochondrial rates with ploidy specified for each locus. All phylogenetic analyses were conducted on the CIPRES Science Gateway (Miller et al., 2010).

We adjusted the *BEAST dated species tree to have one tip representing each species in order to produce a 'complete time-tree' for diversification analyses. To do this, we removed redundant populations of *S. magellanicus* and estimated the placement of the missing taxa, *S. yarrellii* and *S. pinus perplexus*, based on partial data. The *BEAST phylogeny revealed that the Andean populations of *S. magellanicus* formed a single clade with low divergence and thus, may be represented as a single lineage; we retained only the population with the highest number of unique haplotypes (Lima Department, Peru). *Spinus yarrellii* and *S. pinus perplexus* (B. Milá, pers. comm.) represent independent lineages within the genus; however the only genetic sequence available for each was *cytb,* available through Genbank (Supplemental Appendix 1). We chose to place these two taxa in the complete time-tree after *BEAST analysis because they lacked data for the majority of informative characters, which could cause them to fit equally well in different parts the tree, lowering node support overall (Sanderson and Shaffer, 2002; Thomson and

Shaffer, 2010). In a BEAST analysis using methods summarized above, we found *S. yarrellii* to be sister to the eastern lowland clade of *S. magellanicus* (*S. magellanicus alleni*) and *S. pinus perplexus* sister to other *S. pinus*. To create complete time-trees, we took 5000 post-burnin *BEAST trees as well as the final *BEAST phylogeny, pruned the redundant Andean *S. magellanicus* populations, added *S. yarrellii* as sister to *S. m. alleni*, and added *S. p. perplexus* as sister to *S. pinus*. We used a custom R script to add *S. yarrellii* at the midpoint of the *S. m. alleni* branch and *S. p. perplexus* at the midpoint of the *S. pinus* branch in each tree (Supplemental methods). With this approach, the ages of both *S. yarrellii* and *S. p. perplexus* in the complete time-trees were well within their respective 95% highest posterior probability age estimates from the BEAST analysis that included both taxa.

Dispersal-vicariance reconstruction

We obtained geographic distributions of taxa from the literature (Davis, 1999; Dawson, 1997; Fjeldså & Krabbe, 1990; Hilty, 2003; Howard and Moore, 2003; Jetz et al., 2013; McGraw & Middleton, 2009; Restall et al., 2006; Ridgely and Tudor, 1989; Schulenberg et al., 2007; Watt and Willoughby, 2014) and divided the occupied area into nine biogeographic regions (Table 2). Region boundaries within South America represent a simplification of the areas of endemism published by Cracraft (1985) to reflect South American *Spinus* distribution limits more accurately. Note, due to the geographic population structure in *S. magellanicus*, we excluded Guyana as part of the range of this species since nothing is known about the genetic affiliation of this population to the

central Andean or *S. m. alleni* populations. We assumed that current geographic ranges are similar to previous range distribution for the same species.

We reconstructed the biogeographic history of *Spinus* using S-DIVA and Bayesian Binary MCMC (BBM) analyses implemented in the program RASP (Yu et al., 2012). S-DIVA uses an event-based three-dimensional cost matrix (Ronquist, 1996) to evaluate dispersal, extinction and vicariance events across a phylogeny. Given a set of trees generated through Bayesian analysis, S-DIVA incorporates phylogenetic uncertainty into reconstructions by evaluating support for each node in a final tree. BBM addresses phylogenetic and mapping uncertainty through Bayesian analysis to estimate ancestral areas across a set of given trees (Ronquist, 2004; Yu et al., 2012). In both analyses, we used 5000 complete time-trees modified from *BEAST (see above) and specified the complete *BEAST time-tree as the final tree. We restricted the maximum number of ancestral areas allowed (n) to two based on preliminary analyses comparing the S-DIVA values, analogous to the fit of the data to the area model, of n=2, n=3 and n=4. In the S-DIVA analysis, we restricted dispersal to adjacent regions only (Yu et al., 2012). We ran the BBM analysis for 50,000 generations with 10 chains at a temperature of 0.1. We specified fixed state frequencies (JC), equal among-site rate variation and a wide root distribution. Null distributions, distributions including areas outside the nine regions, were not allowed in any reconstruction. Note, for outgroups *Serinus* and *Coccothraustes* without complete sampling, we used the "common-equals-primitive" criterion as suggested by Ronquist (1996) to estimate the genus root distribution though we allowed a wider distribution if regions were equally common.

Diversification rate analyses

We compared our complete time-tree to the expectations of two diversification models, pure-birth and birth-death, using a lineage-through-time plot (laser: Rabosky, 2006a) in R (R Core Team 2013). In the pure-birth model, the increase in lineages over time is exponential and depends on a single birth rate, b, with no extinction (Yule, 1924). In the birth-death model, the rate of diversification is determined by the birth rate, b, minus the extinction rate, d (Kendall, 1948). We estimated birth and death rates for these models using our complete time-tree (Rabosky, 2006a), simulated 5000 trees for each model (geiger: Harmon et al., 2012) and plotted the average simulations on our lineage-through-time plot for a visual comparison.

To test for speciation rate variability in *Astragalinus* and *Spinus*, we used the gamma statistic (Pybus and Harvey, 2000) and a version of the ΔAIC_{RC} statistic (Rabosky, 2006b). A gamma statistic value of -1.645 or smaller for a completely sampled phylogeny indicates a significant deviation from the pure-birth model. The gamma statistic is robust to missing taxa though the number of missing lineages will impact the exact critical value. The power of the gamma statistic is greatest if there is a decrease in birth rate over time; it has little power to identify a rate shift if the birth rate increases over time or if extinction is moderate or high (Pybus and Harvey, 2000). We calculated the gamma statistic and a p-value for our result using the Monte Carlo Constant Rates test (laser: Rabosky, 2006a) with the *BEAST tree (2 missing taxa). In the MCCR test, a gamma statistic null distribution based on pure-birth simulations is generated and the p-value results from a one-tailed t-test. The ΔAIC_{RC} statistic is a likelihood method that compares the AIC value of the best constant rate and variable rates models. Like the

gamma statistic, the significance of the ΔAIC_{RC} statistic is evaluated by constructing a null distribution of ΔAIC_{RC} statistics from pure-birth simulation and conducting a one-tailed t-test. However, ΔAIC_{RC} statistic is capable of detecting both significant increases and decreases in birth rates in the presence of low or moderate extinction given complete taxon sampling. Using our complete time-tree, we applied the rate-constant diversifications models (pure-birth and birth-death). The rate variable diversification models that we considered were two density-dependent models: one exponential model ($r_t = r_0 N_t^{-x}$ where r is the rate of diversification, N_t is the number of lineages at time t and x dictates the rate of change of r_t), and one logistic model ($r_t = r_0 (1- N_t/K)$ where K represents a factor similar to carrying capacity). We also included a pure-birth model where the birth rate shifts once across time. As recommended by Rabosky (2006b), we only permitted rate shifts to occur at nodes; this method is conservative since permitting rate shifts at any time could bias our results towards rate variable models. We calculated ΔAIC_{RC} and an associated p-value (Rabosky, 2006a) for AICc, a more appropriate measure for model comparison with a small number of samples (Burnham and Anderson, 2002).

We also investigated rate variability in the *Astragalinus* and *Spinus* phylogeny using the Geographic State Speciation and Extinction model (GeoSSE, Goldberg et al., 2011) in R (diversitree: Fitzjohn 2013). The GeoSSE allowed us to estimate diversification parameters independently for two geographic regions including speciation (S) and extinction (X) rates for each area as well as dispersal (D) rates out of each area into the other. We designated two geographic regions: 1) the Andes and 2) Outside the Andes. A finer biogeographic scale would have been computationally prohibitive

(Goldberg et el., 2011). Importantly, unlike binary state speciation and extinction models (BiSSE), the GeoSSE allows a trait change (range shift) to be associated with a speciation event (FitzJohn, 2012) and measures this speciation rate as well ($S_{A/OA}$).

Using the complete time-tree, we compared the maximum likelihood of the full parameter GeoSSE model to a set of constrained models. For each constrained model, we held one parameter (speciation, extinction or dispersal) equivalent between the two geographic regions. For one model, we defined extinction equivalent and zero as in the pure-birth diversification model. We used AICc and Akaike weights to identify the most likely model for further analysis. Using the best-fit model, we estimated diversification parameters for *Astragalinus* and *Spinus* by using 1000 complete time-trees, defining geographic states for each taxa (Andes, Outside the Andes or in both regions) and using a Monte Carlo Markov Chain analysis (20000 steps/tree) to generate posterior distributions of parameter estimates for each tree. MCMC results were combined and analyzed as a time-series (coda; Plummer et al., 2012) after removing a 25% burnin. We summarized support for logical statements by evaluating the number of MCMC iterations for which the statement was true (Goldberg et al., 2011). The logical statements we tested were: **(1)** the speciation rate of the Andes (S_A) was greater than the speciation rate outside the Andes (S_{OA}), **(2)** the speciation rate of the Andes (S_A) was greater than the speciation rate associated with moving between the Andes and outside the Andes ($S_{A/OA}$), and **(3)** the dispersal rate out of the Andes (D_A) is greater than the dispersal rate into the Andes (D_{OA}).

RESULTS

Molecular data, models and partitions

We amplified and sequenced the three mitochondrial genes (*cytb*, ND2 and ND3; 2574 base pairs) for 65 of 66 tissue samples (Appendix) and all three loci (mtDNA plus nuclear introns MUSK and MYO2; 3989 base pairs) for 51 individuals (Appendix). We were unable to sequence nuclear loci from individuals of *S. olivaceus, S. siemiradzkii, S. spinescens* and *S. xanthogastrus* due to poor quality of samples.

We found mtDNA conformed with neutral expectations in all populations and species except in *cytb* and ND3 in central Peru *S. magellanicus* and ND2 in northern Peru *S. magellanicus* (Supplemental Table 1). Molecular convergence is unlikely to have affected phylogenetic signal because all well-resolved clades were also recovered in phylogenies estimated using only third codon positions of mtDNA.

The partition scheme with the best BIC score in the PartitonFinder analysis for the mitochondrial DNA included two partitions: (1) codon positions 1 and 2 for all three genes as well as ND3 tRNA flanking regions with evolutionary model HKY+I and (2) codon position 3 comprised with model GTR+I+G. Evolutionary models for the nuclear introns MUSK and MYO2 were HKY+I and Jukes-Cantor, respectively.

Phylogenetics

We summarize the major results and conflicts among datasets and methods on a clade-by-clade basis below. Unless otherwise stated, bootstrap support (bs) summarizes support from RAxML trees conducted for both mtDNA and all loci datasets; posterior probability (pp) summarizes Bayesian analyses with both mtDNA and all loci datasets,

the BEAST analysis and the *BEAST analysis (see Table 3 for support values from individual analyses). In summary, we list the lowest bs or pp recovered from a single analysis for an individual node.

Goldfinches

The goldfinches of North America (*Astragalinus lawrencei*, *A. tristis*, and *A. psaltria*) formed a monophyletic clade that was sister to all other *Spinus* with strong support (bs ≥ 88, pp ≥ 0.99, Fig. 2). *A. tristis* and *A. psaltria* formed a sister group to the exclusion of *A. lawrencei* (mtDNA: bs=89, pp≥0.97; all loci: bs=55, pp≥0.78).

Non-South American siskins

The siskins of North America and Central America (*Spinus dominicensis, S. atriceps, S. pinus, S. notatus*) and the Eurasian siskin, *S. spinus,* were more closely related to South American *Spinus* than to the goldfinches (bs≥95, pp≥0.99). The southern Central American *S. notatus* was sister to all South American *Spinus* (bs≥97, pp=1), and *S. dominicensis, S. atriceps, S. spinus* and *S. pinus* formed a monophyletic clade sister to *S. notatus* and the South American siskins with modest support (bs=67, pp≥0.76). mtDNA analyses, however, supported an alternate topology (Supplemental Fig. 1, see below). *Spinus atriceps* and *S. pinus* were sisters in all analyses (bs≥99, pp≥0.99). ***Spinus spinus:*** mtDNA BEAST analysis and all three-locus analyses recovered *S. spinus* as sister to *S. atriceps* and *S. pinus* (bs=72, pp≥0.95), although this node was poorly resolved in RAxML and Bayesian analyses with mitochondrial DNA only.

Spinus dominicensis: mtDNA and three-locus analyses disagreed on the placement of Hispaniola endemic species *S. dominicensis*. mtDNA analyses supported *S. dominicensis* as sister to the clade containing the South American siskins and *S. notatus* (bs=79, pp≥0.90, Supplemental Fig. 1). The three-locus analyses indicated that *S. dominicensis* was sister to *S. atriceps*, *S. pinus* and *S. spinus* (bs=67, pp≥0.76, Figs. 2 and Supplemental Fig. 4).

South American siskins

The South American *Spinus* formed a strongly supported monophyletic clade (bs=96, pp≥0.97) characterized by short internal nodes. Six taxa diverged near the base of the clade: *S. barbatus, S. cucullatus, S. olivaceus, S. spinescens, S. xanthogastrus* and a divergent lineage of *S. magellanicus* referred to hereafter as *S. m. alleni* (Fig. 2). A seventh lineage that diverges near the base of the South American *Spinus* clade we refer to hereafter as the Central Andean clade; it is comprised exclusively of species that occur in the Central Andes or Central and Austral Andes (*S. atratus, S. siemiradzkii, S. crassirostris, S. uropygialis,* and all Peruvian *S. magellanicus*). The Central Andean clade was recovered in all analyses (bs≥33, pp≥0.51), although support was strongest with the coalescent methods (BEAST, pp=0.99 and *BEAST, pp=0.93).

Spinus magellanicus: The widespread taxon *S. magellanicus* appeared in three separate parts of the South American *Spinus* clade in all analyses. First, samples from the lowlands of northeastern Argentina and eastern Bolivia that we tentatively refer to as *S. m. alleni* formed a monophyletic clade on a long branch. Second, within the Central Andean clade, some *S. magellanicus* were included in the *S. atratus* clade with strong

support (bs=93, pp=1, excluding *BEAST). These *S. magellanicus* individuals included all central Bolivian samples and a subset of individuals from high elevations in Lima and Cusco, Peru; some *S. magellanicus* and *S. atratus* in this clade shared identical mtDNA and nuDNA haplotypes (Supplemental Figs. 1, 2, and 3). Third, *S. magellanicus* samples from all Peruvian populations formed a shallow clade with *S. crassirostris* and *S. siemiradzkii* (bs≥53, pp≥0.99, excluding *BEAST) within the Central Andean clade. Some *S. magellanicus*, *S. crassirostris*, and *S. siemiradzkii* shared exact mtDNA and nuDNA haplotypes. In our *BEAST analysis, geographic populations rather than individuals comprised the terminal nodes; as a result, only the central Bolivian *S. magellanicus* population was recovered as sister to *S. atratus* in the *BEAST analysis (pp=0.26) despite "atrata" mtDNA haplotypes also occurring at low frequency in Lima and Cusco *S. magellanicus* populations (Fig. 2). In *BEAST, all Peruvian populations of *S. magellanicus* were grouped with *S. crassirostris* and *S. siemiradzkii* (pp=0.28; Fig. 2).

Spinus uropygialis: A sister relationship between *S. uropygialis* and the rest of the Central Andean clade (*S. atratus, S. siemiradzkii, S. crassirostris* and all Peruvian *S. magellanicus*) was well supported in *BEAST (pp=0.99) and moderately supported in mtDNA-only Bayesian analyses (pp=0.82). mtDNA-only and three-locus maximum likelihood, BEAST, three-locus Bayesian analyses were unable to resolve the placement of *S. uropygialis* with respect to other members of the Central Andean clade.

Divergence Dates

Divergence dates estimated with BEAST and *BEAST were similar though the young nodes were younger and the oldest nodes were older in *BEAST compared to

BEAST (Table 3). The 95% highest posterior density (HPD) ranges overlapped considerably between BEAST and *BEAST analyses at all shared nodes except the root. We present the three-locus *BEAST dates unless otherwise noted. The most recent common ancestor of *Astragalinus* and *Spinus* was estimated at ~2.88 million years ago (Ma) (95% HPD: 2.19-3.56 Ma, Table 3). The North American goldfinches (*A. tristis, A. psaltria,* and *A. lawrencei*), shared a common ancestor ~1.83 Ma (95% HPD: 1.25-2.45 Ma) and sister taxa *A. psaltria* and *A. tristis* diverged ~1.54 Ma (95% HPD: 0.98-2.15 Ma). Sisters *Spinus atriceps* and *S. pinus* diverged ~0.35 Ma (95% HPD: 0.02-0.64 Ma), and shared a common ancestor with *S. spinus*, the Eurasian siskin, at ~1.17 Ma (95% HPD: 0.70-1.70 Ma), and a common ancestor with *S. dominicensis* at ~1.52 Ma (95% HPD: 1.10-1.96 Ma) .

The most recent common ancestor of the South American *Spinus* and *S. notatus* was ~1.03 Ma (95% HPD: 0.56-1.45 Ma). The most recent common ancestor of *Spinus* in South America dated to ~0.55 Ma (95% HPD: 0.40-0.72 Ma). *BEAST estimated the ancestor of the Central Andean clade to be ~0.21 Ma (95% HPD: 0.10-0.32 Ma); however, the BEAST estimate for this node was older (~0.40 MA, 95% HPD: 0.26-0.55 Ma).

Dispersal-Vicariance Analysis

The ancestral range reconstruction using S-DIVA revealed a northern origin for *Astragalinus* and *Spinus* (Fig. 3, Table 4); all reconstructions of the most recent common ancestor included North America. The siskin ancestor dispersed southward into Central America and diversified there while goldfinches diversified within North America. The

S-DIVA reconstruction showed that (1) some siskin diversification occurred within Central America and (2) there were several independent dispersals between Central America and other regions. *A. psaltria* likely originated in North America and dispersed independently into Central America and South America. Similarly, *S. pinus,* which currently occurs widely in North America, most likely had a Central American origin. The ancestor of *S. spinus* probably colonized Europe and North Africa through dispersal from Central America. Although the S-DIVA result showed an *S. spinus* ancestor with a range including Central America and Europe, this reconstruction is most likely due to the failure of S-DIVA to account for the vast geographic distance between Central America and Europe which makes a continuous range between these two regions unlikely.

The ancestor of *S. notatus* and South American *Spinus* occupied Central America and either the Northern coastal mountains of South America (p=0.5) or the Northern Andes (p=0.3). The ancestor subsequently expanded across South America, southward through the Andes into Patagonia. Siskins colonized the Andes soon after reach South America, between ~0.47-0.55 Ma. The expansion of *Spinus* across South America includes several transitions into and out of the Andes and, beginning ~0.21 Ma, speciation within the Central and Austral Andes.

The ancestral range reconstructions based on the Bayesian Binary MCMC (BBM) and S-DIVA were similar (Table 4), except that the BBM analysis commonly found narrower ranges to be more likely than broad ranges. For example, BBM reconstructed North America as the sole ancestral range for goldfinches and siskins with a probability of 0.68, with lesser support for Europe or Central America as the ancestral areas. Expansive range reconstructions including Europe/North Africa and North America or

Central America were not present in the BBM analysis. BBM results show the ancestor of *Spinus spinus*, *S. atriceps* and *S. pinus* resided in Central America (pp = 0.56), suggesting that the current European distribution of *S. spinus* resulted from long-distance dispersal.

Diversity Analyses

Standard Diversification models

The accumulation of species in *Astragalinus* and *Spinus* was near constant through time (Fig. 4), implying a steady rate of net diversification (Nee, 2006). Using the MCCR test, we determined that the critical value of the gamma statistic was -1.808; a gamma statistic more negative than -1.808 would indicate a deceleration in speciation rate. However, we estimated gamma at 0.983 (p-value= 0.86) for goldfinches and siskins*;* thus, there was no evidence for a deceleration in speciation rate (Pybus and Harvey, 2000).

The standard diversification model that best fit the *Astragalinus* and *Spinus* phylogeny was the pure-birth model (AICc: -46.03, Akaike weight: 0.34), though the constant birth-death diversification model had a similar fit (AICc: -45.68, Akaike weight: 0.29). The best rate variable model was the density dependent exponential model (AICc: -44.41, Akaike weight: 0.15). The difference in AICc score between the best rate-variable and best rate-constant model,

ΔAIC_{RC}, was low (-1.618**,** p-value: 0.397), well within expectations for a phylogeny generated under a pure-birth model.

In sum, the results of the gamma statistic and the ΔAIC_{RC} indicate the pure-birth model of diversification is an appropriate model for describing lineage accumulation in *Astragalinus* and *Spinus* (Table 5). The overall birth rate for this clade, estimated using the pure-birth model, was 1.04/Ma.

Geographic State Speciation and Extinction model

To investigate diversification rate variability on a finer scale, we assessed speciation, extinction and dispersal in and outside of the Andes using the Geographic State Speciation and Extinction model (GeoSSE). We used maximum likelihood to test the full, seven-parameter model against biologically-relevant constrained models (Table 6). The best-fit model of the five tested was the no-extinction model ($X_A \sim X_{OA} \sim 0$) (AICc: 53.46, Akaike weight: 0.81); hence, we used that model to estimate speciation and dispersal rates using Bayesian analysis (see methods). Notably, the worst fit model was the six-parameter model constraining speciation to be equivalent between the two regions (AICc: 66.31, Akaike weight: 0.001).

Using the no-extinction GeoSSE model and Bayesian analysis (Table 7), we found that speciation within the Andes (mean: 3.36/Ma ± 0.0085 SE) was significantly higher than speciation outside the Andes (mean: 0.68/Ma ± 0.0014 SE); this result was supported by a posterior probability of 0.99 (Fig. 5). We also found that dispersal out of the Andes (mean: 1.61/Ma ± 0.0059 SE) was significantly greater than dispersal into the Andes (mean: 0.21/Ma ± 0.0003 SE); this conclusion was supported by a posterior probability of 0.96. The results were concordant with the maximum likelihood analysis of the no-extinction GeoSSE model (Table 6), though dispersal rate out of the Andes in the

Bayesian analysis was almost three times that predicted from the likelihood analysis. Both methods showed that speciation rate within the Andes and dispersal rate out of the Andes were significantly greater than speciation and dispersal rates outside the Andes.

Some discrepancies between these analyses are expected since our Bayesian analysis incorporated the phylogenetic uncertainty of 1000 trees and the maximum likelihood analysis used the summary tree only. Nonetheless, the disparity between estimates for the speciation rate associated with inter-regional transitions ($S_{A/OA}$) was large (Bayesian mean: 2.891/Ma ± 0.0116 SE, maximum likelihood: 5.30×10^{-8}/Ma). There was substantial standard error and a long right tail to the posterior distribution of our Bayesian estimate of this parameter, $S_{A/OA}$, which suggests the speciation rate associated with inter-regional transitions was difficult to estimate (Table 7, Fig. 5). The error and posterior distributions of other GeoSSE parameters were far smaller, suggesting that the disparity between maximum likelihood and Bayesian estimates for $S_{A/OA}$ should not undermine confidence in the regional speciation rates (S_A, S_{OA}) or the regional dispersal rates (D_A, D_{OA}).

DISCUSSION

We produced a complete, dated phylogeny of all the New World goldfinch and siskin species, including the Eurasian siskin, using three mitochondrial genes, one sex-linked intron and one autosomal intron. This phylogeny has implications for the role of the Andes in bird diversification, inter-regional variation in diversification, Pleistocene biogeography and generic-level taxonomy.

Taxonomy

The North American goldfinches *Astragalinus tristis, A. lawrencei* and *A. psaltria* formed a monophyletic clade sister to all New World siskins. This result confirms previous studies (Arnaiz-Villena et al., 2007; Arnaiz-Villena et al., 2008; Zuccon et al., 2012) and supports recent proposals to the South American Checklist Committee (SACC; see Remsen, 2011) and NACC (Chesser and Rising, 2013) to assign these taxa to a separate genus, *Astragalinus* (Cabanis, 1851).

We found that the New World siskins and the Eurasian siskin formed a well supported, monophyletic clade. The clade including the South American *Spinus* species and the Central American *S. notatus* was strongly supported as monophyletic, consistent with previous studies (Arnaiz-Villena et al., 1998; Arnaiz-Villena et al., 2007; Nguembock et al., 2009; Zuccon et al., 2012). The three locus analyses recovered the remaining non-South American siskins (*Spinus dominicensis, S. atriceps, S. pinus* and *S. spinus*) as monophyletic and sister to *S. notatus* plus all South American *Spinus*. However, support for this relationship was moderate (bs= 67, pp=0.94 Bayesian, pp=0.76 *BEAST) and is at odds with the mtDNA-only analysis where *S. dominicensis* was sister to *S. notatus* and South American *Spinus*. The structure of the MUSK-only phylogeny also supports the sister relationship between *S. dominicensis* and the *S. notatus*/South American *Spinus* clade (Supplemental Fig. 2). The relationship of *S. dominicensis* in the three locus-analyses was influenced by strong phylogenetic signal from MYO2 (Supplemental Fig. 3). In the MYO2 phylogeny, *S. dominicensis* was nested inside a clade with *S. pinus, S. atriceps, S. spinus* and *A. tristis;* this clade was sister to a clade containing *S. notatus, A. psaltria, A. lawrencei, S. cucullatus,* and some *S. magellanicus*

individuals. All remaining South American *Spinus* formed a monophyletic clade sister to the two former clades. We suspect that the MYO2 phylogeny does not reflect the true species tree and resulted from incomplete lineage sorting. This is a viable hypothesis considering that the autosomal MYO2 has a larger effective population size than mtDNA or the Z-chromosome MUSK gene. Further, the MYO2 tree incorrectly places *A. lawrencei* and *A. psaltria* nested within the South American siskins (Supplemental Fig. 3, Supplemental Fig. 4); recent introgression of the myoglobin gene can be rejected as the origin of this pattern since lineages that are sister only in the MYO2 tree are physically isolated. Our results demonstrate that reciprocal monophyly between a northern and southern siskin clade would be difficult to establish because of the uncertainty about the placement of the Caribbean species as well as the Eurasian siskin, whose placement was poorly resolved using mtDNA alone (Supplemental Fig. 2). Hence, we favor lumping all New World siskins and the Eurasian siskin into the single genus, *Spinus* (Koch, 1816), to be referred to as the New World siskins. A single genus would reflect the unambiguous monophyly of the New World siskins and their overall morphological and behavioral similarity (Dawson, 1997; Ridgely and Tudor, 1989).

Our phylogenetic analysis revealed two divergent lineages of *S. magellanicus*, one located in the Peruvian and central Bolivian Andes and the other, referred to tentatively as *S. m. alleni*, from the lowlands of eastern Bolivia and northern Argentina. We are uncertain of the subspecies identification for these lowland birds; however, there are at least three subspecies in this region including *S. m. alleni, S. m. icterius* and *S. m. bolivianus.* We refrain from proposing species status for *S. m. alleni* based on a single phylogenetically informative locus, mtDNA, but genomic, morphological and vocal

characters should be assessed to establish whether species recognition is warranted for lowland populations of *S. magellanicus* east of the Andes. In particular, additional sampling is needed across the large geographic range of *S. magellanicus* in the eastern lowlands, the Northern Andes and the Tepuis (Ridgely and Tudor, 1989).

Our BEAST and *BEAST analyses revealed a Central Andean siskin clade with extensive haplotype sharing among sympatric high-elevation species. The Central Andean clade, comprising five species (*S. magellanicus* from Peru and central Bolivia, *S. atratus, S. crassirostris, S. siemiradzkii* and *S. uropygialis*), is present in all phylogenetic reconstructions, but has greatest support in the BEAST and *BEAST trees (posterior probability equaled 0.99 and 0.93, respectively). The difference in support for the Central Andean clade may be caused by the increased power of coalescent methods to resolve the relationships of recently diverged lineages by modeling the stochastic loss of shared ancestral polymorphisms over recent timescales (Bowie et al., 2006; Edwards, 2009). Incomplete lineage sorting, where shared ancestral polymorphism is retained in evolutionarily independent lineages (Edwards, 2009), may be the source of the extensive haplotype sharing that we observed between Central Andean *S. magellanicus* and other Central Andean species. Alternatively, mitochondrial introgression from high elevation sympatric congeners *S. crassirostris* and/or *S. atratus* into *S. magellanicus* would cause a similar pattern of haplotype sharing. To test whether incomplete lineage sorting or ancient introgression is responsible for the observed pattern of haplotype sharing pattern will require many informative nuclear loci among Central Andean taxa (Peters et al., 2007; Rheindt et al., 2014).

Diversification and dispersal rates

The diversification rate of the goldfinches and siskins in the New World was high throughout the Pleistocene at ~1.04 species/Ma. For perspective, Jetz et al. (2012) reported that the lower bound for the top 25% quartile for bird diversification rates was 0.243/Ma. Examples of exceptionally diverse clades in birds, such as the 9-primaried oscines (including Cardinalidae, Emberizidae, Icteridae, Parulidae and Thraupidae) and hummingbirds (Trochilidae) on average have lower diversification rates, 0.13-0.38 species/Ma (Barker et al., 2013) and ~0.23 species/Ma (McGuire et al., 2014), respectively, though rates are heterogeneous within both clades. Within Trochilidae, for example, the diversification rate of the Bee clade, ~0.87 species/Ma, was closer to the goldfinch and siskin diversification rate. However, the diversification rate of the white-eyes, *Zosterops*, one of the avian "great speciators," far exceeds this at 1.95-2.63 species/Ma (Moyle et al., 2009).

We found a significantly higher diversification rate in the Andes, ~3.4 species/Ma, than outside the Andes, ~0.7 species/Ma. The Andean siskin diversification rate is similar to the speciation rates of the greatest animal speciators (Mendelson and Shaw, 2005; Moyle et al., 2009). It is also close to the high diversification rates of plants that diversified extensively in the Andes during the Pleistocene such as *Lupinus* (Hughes and Eastwood, 2006) and several Andean páramo lineages (Madriñán et al., 2013). In this respect, our result corroborates the conclusions of Weir (2006), whose meta-analysis suggested that average diversification rates of Andean taxa were higher than those of non-Andean taxa.

Weir (2006) also predicted that if glacial cycles facilitated diversification, diversification rates should increase with increased severity of glacial cycles. We did not find increased glacial cycle severity and frequency in the last one million years coincided with a diversification rate shift in *Spinus*. However, the analysis we conducted to test for a rate shift over time has low power in small clades (Rabosky, 2006b), so this result is not a definitive rejection of glacial cycles as a cause of siskin speciation. Siskin colonization of the Andes most likely occurred near the beginning of the late Pleistocene, and siskin diversification rates peaked in the Andes during the late Pleistocene; these findings are consistent with siskin diversification having been facilitated by the joint effects of fluctuating climate and topographic heterogeneity.

We found an increased rate of dispersal from the Andes to the non-Andes, supporting the idea that the Andes may have served as a species pump to the lowlands (Fjeldså and Rahbek, 2006). The result is consistent with previous findings that Andean radiations contributed significantly to lowland diversity (Batalha-Filho et al., 2013; García-Moreno et al., 1999; Sedano and Burns, 2010; Voelker, 1999); however, the directional dispersal needs to be systematically tested in other groups. If the direction of dispersal was predominantly from high to low elevation for Pleistocene clades, one would expect to find more young species at low elevations than observed (Hawkins et al., 2011). In the Miocene diversification of *Thamnophilus* antshrikes (a primarily lowland radiation), transitions from low to high elevations or to broad elevational ranges were more frequent than transitions from high to low elevations (Brumfield and Edwards, 2007). Similarly, in the core tanagers (a primarily Andean radiation), dispersals from high elevations to low elevations made up a minority of dispersal events though several

of the high-low shifts had significant impacts on clade diversity (Sedano and Burns, 2010). In the hummingbirds, Andean and lowland radiations were largely independent (McGuire et al. 2014) and Pleistocene radiations in the Andes did not cause lowland colonization (Benham et al. 2014). Given these previous findings, it is not yet clear whether the pattern of Andean dispersal observed among the siskins will prove to be a general pattern or an exception.

The role of transitions into and out of the Andes in generating diversity requires further research (Brumfield and Edwards, 2007). In this study, the GeoSSE model did a poor job of estimating the rate of speciation associated with transitions between Andean and non-Andean regions. Studies that examine local adaptation to elevation among populations (DuBay and Witt, 2014; Projecto-Garcia et al., 2014) and explicitly model elevational transitions (Brumfield and Edwards, 2011; Campagna et al., 2011) may be more successful at elucidating the role of shifts in elevation in causing speciation.

Biogeography & Regional Diversification Patterns

The ancestor of the goldfinches and siskins arrived in North America sometime after ~4.1 Ma. During the early Pleistocene, the goldfinches occupied North America and the siskin ancestor dispersed southward. During the mid-Pleistocene, the goldfinches diversified in North America and the siskins diversified in Central America. Siskins colonized South America in the late Pleistocene and dispersed widely across the continent, including the entire latitudinal extent of the Andes, early in the radiation. These results suggest that diversification of goldfinches and siskins occurred in the context of Pleistocene glacial cycles and that the South American siskins radiated

primarily in the late Pleistocene, coincident with severe and frequent glaciations (Hooghiemstra and van der Hammen, 2004).

There are several proposed mechanisms for how Pleistocene glaciations may have impacted the physical isolation and diversification of Andean avifauna (Fjeldså, 1994; Graves, 1985; Haffer, 1969). Of these, the páramo expansion hypothesis (van der Hammen and Cleef, 1986) has the greatest explanatory power for the South American *Spinus* clade because it is an open-habitat taxon that is generally associated with páramo or puna grasslands at high elevations (Fjeldså and Krabbe, 1990). This hypothesis has been invoked to explain the exceptional endemism (van der Hammen and Cleef, 1986) and high diversification rates found in páramo plants (Madriñán et al., 2013; Särkinen et al., 2012).

Páramo expansion and contraction may have been a primary cause of the high diversification rate of Andean siskins as well as other páramo bird species such as the ground-tyrants (*Muscisaxicola*) (Chesser 2000, Weir 2006). In support of this, palynological studies from different regions of the Andes indicate elevational shifts in vegetation zones during Pleistocene glacial cycles (Hooghiemstra and van der Hammen, 2004; van der Hammen and Cleef, 1986). During dry, cool glacial periods, the Andean montane forest belt narrowed and moved to lower elevations. Simultaneously, the páramo/puna grassland habitat, located above the upper limit of montane forest, extended to lower elevations and expanded to several times its previous area. This expansion of páramo provided continuous grassland habitat between high elevation regions, "páramo islands", that were isolated from one another during interglacial periods. Repeated glacial cycles may have provided páramo species with the opportunity for dispersal,

colonization, and secondary contact during dry, cool periods of glaciation and the chance to evolve in isolation on "páramo islands" during interglacials and wet, cool glacial periods that marked the peak of glacial extent (Madriñan et al., 2013; van der Hammen and Cleef, 1986). The South American siskins and the ground-tyrants both show strong local patterns of diversification, suggesting that physical isolation on "paramo islands" could be important in facilitating speciation across moderate geographic distances (e.g. within the Central Andes) as well as at larger scales (e.g. between regions of the Andes).

There is evidence for grassland expansion outside of the Andes during Pleistocene glacial cycles as well, however the extent is poorly understood (Bush and Oliveira, 2006; Bush et al., 2007; Clapperton, 1993; Rull, 2011; Servant et al., 1993), in part due to the dearth of appropriate palynological sites (Bush et al., 2007). While páramo expansion in the Andes provided migration corridors between "páramo islands", Bush et al. (2007, p. 65) describes the heterogeneity of Amazonia during glacial cycles as "a series of patches and connectivity as a shuffling within the patchwork rather than migration in corridors." The habitat shifts at low elevations may have facilitated diversification; Campagna et al. (2012) cited the expansion of dry, open habitats at low elevations during Pleistocene cycles as a source of diversity in the rapid radiation of southern capuchino seedeaters (*Sporophila*). However, our result that siskin diversification rates are significantly higher in the Andes than the non-Andes suggests that fluctuations in the area occupied by non-forest habitats had a greater impact on diversity in the Andes than outside the Andes. This finding could indicate that grassland expansions at low elevations had a localized, rather than regional, impact (Bush et al. 2007); alternatively, it may be a result of differences in

biotic interactions and niche availability between low and high elevations (Brumfield and Edwards, 2007; Bush et al., 2007; Rull, 2011).

The role of intrinsic characteristics

The high diversification rate of *Spinus* may also be a consequence of certain intrinsic dispersal behaviors in siskins and other finches (Price, 2011). The impact of dispersal ability on diversification rate can be two-fold: a high dispersal propensity may permit individuals to colonize new areas, it may also maintain connectivity across a large geographic range, preventing isolation and local adaptation (Burney and Brumfield, 2009). Evidence for a strong relationship between dispersal ability and diversification rate is mixed; high speciation rates can be correlated with low dispersal propensity (Claramunt et al., 2012; Burney and Brumfield, 2009), moderate dispersal abilities (Diamond et al., 1976) or high dispersal abilities (Owens et al., 1999; Phillimore et al., 2006). The dispersal propensities of siskins, including irruptive migration, may predispose these finches to successfully colonize new regions. Price (2011) and Lerner et al. (2011) hypothesize that irruptive behavior, specifically the nomadic and unpredictable migration of finch flocks in response to food shortage, may be one of the shared traits that have led to the successful radiation of finches such as the Hawaiian Honeycreepers and crossbills as well as other modern seed-eating birds such as Darwin's finches and the capuchino seedeaters (Campagna et al., 2012; Sick 1985).. Along similar lines, Chesser (2000) suggests that species with ancestry in the southern Andes that colonized the central Andes may have done so through the loss of migratory behavior in a population. Moyle et al. (2009) argued that group dispersal in the social white-eyes (*Zosterops*) of the

Old World tropics has contributed to their high speciation rate. Through irruptive migration, loss of migration or group dispersal, flocks of birds traveling together may establish new resident populations distant from their original population center. A difference in dispersal propensity might explain the difference between páramo lineages with high diversification rates (siskins, ground-tyrants) and those with moderate ones like cinclodes (*Cinclodes*; Chesser, 2004), miners, (*Geositta*; Cheviron et al., 2005) or chat-tyrants (*Ochthoeca*; García-Moreno et al., 1998) .

CONCLUSIONS

Our phylogenetic analysis of the New world goldfinches and siskins revealed that goldfinches and siskins form monophyletic, sister clades; we recommend goldfinches should be grouped in the genus *Astragalinus* while the New World siskins (including the single Old World species, Eurasian siskin) should be grouped together in the genus *Spinus*. We found evidence of mtDNA divergence between *S. magellanicus* of the Central Andes and the eastern lowlands, indicating that *S. magellanicus* is comprised of at least two species. We discovered a Central Andean clade of five taxa, three of which breed sympatrically at high elevations; these close evolutionary relationships are rare among sympatric bird species (Weir and Price, 2011). Further study is warranted to understand the formation and maintenance of species that are phenotypically distinct but genetically identical, at least at a few loci.

Regional diversification and dispersal rates varied within the goldfinch and siskin clade; the Andean siskin lineages had a higher diversification rate and higher outward dispersal rate than non-Andean lineages. The rapid diversification of the Andean siskins

likely reflects the interplay between dispersal ability and dispersal opportunity, mediated by páramo expansion and contraction in response to Pleistocene climate cycles. Given this, we predict that other páramo-associated, vagile Andean groups will be found to have undergone fast diversification during the Pleistocene. The general difference in age between high and low elevation bird species in South America (Hawkins et al., 2007, 2011; Weir, 2006) is at least partly attributable to rapid Pleistocene diversification of siskins and other Andean open-habitat birds.

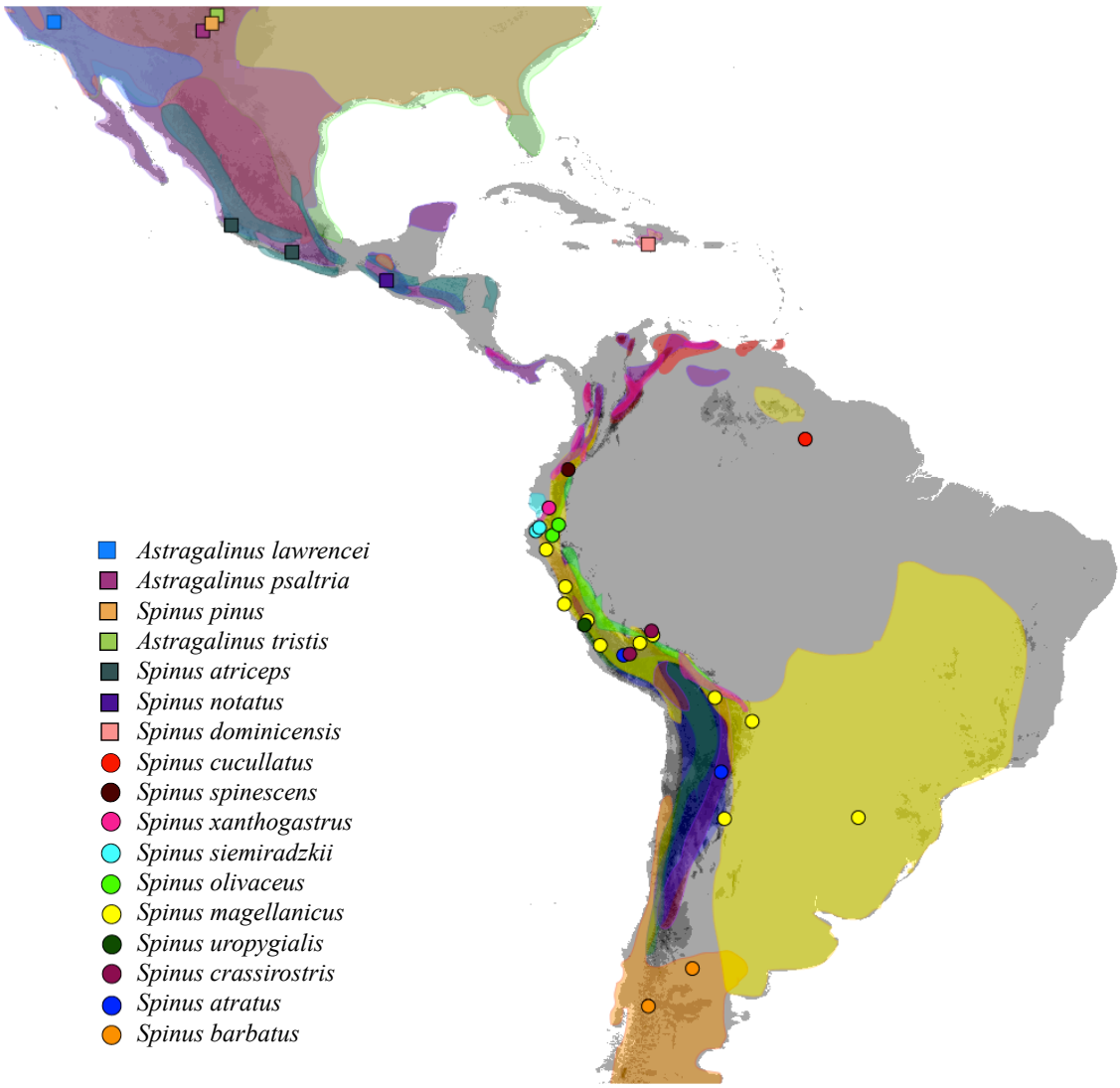

Fig. 1. Sampling localities for *Astragalinus* and *Spinus* tissue samples. Each point represents 1-6 specimens sampled.

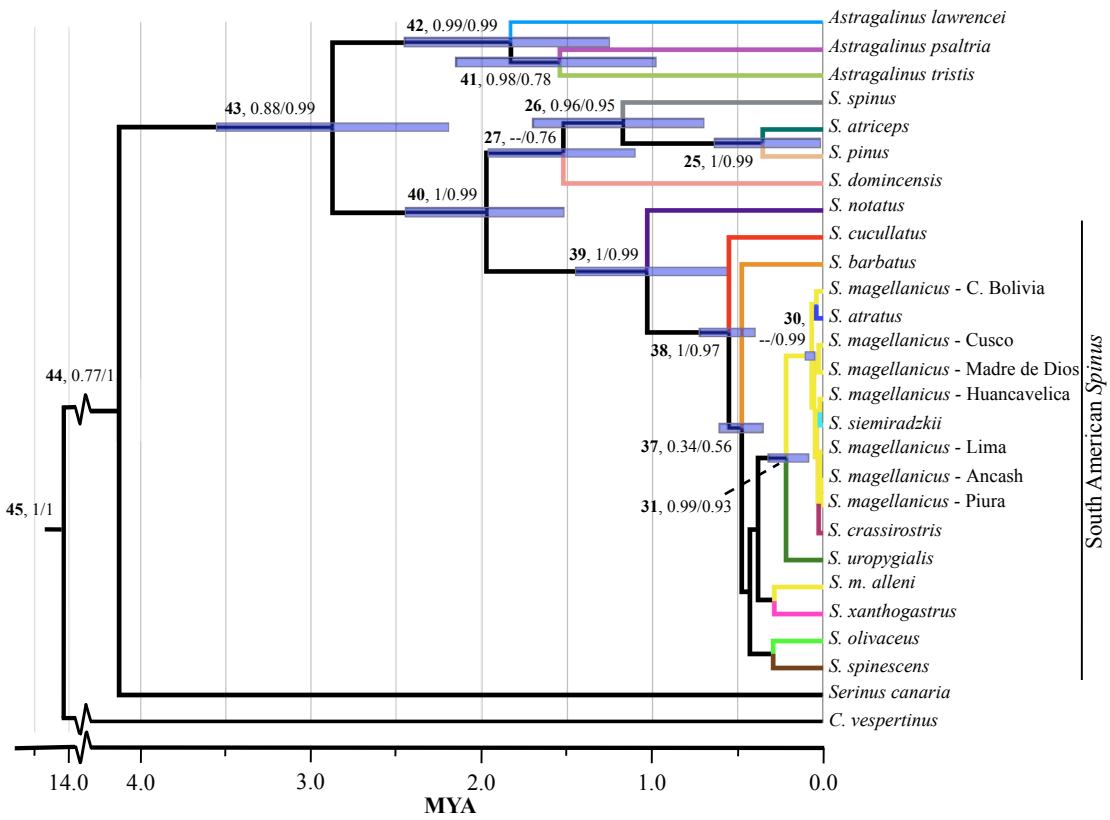

Fig. 2. Species tree for *Astragalinus* and *Spinus*. Tree generated with mitochondrial protein-coding genes (*cytb*, ND2 & ND3) and nuclear introns (MUSK & MYO2) in *BEAST. Branch colors indicate species. Node bars indicate 95% HPD for age (in millions of years). Labels at nodes include node numbers (bold) and node support for BEAST and *BEAST, before and after slash, respectively.

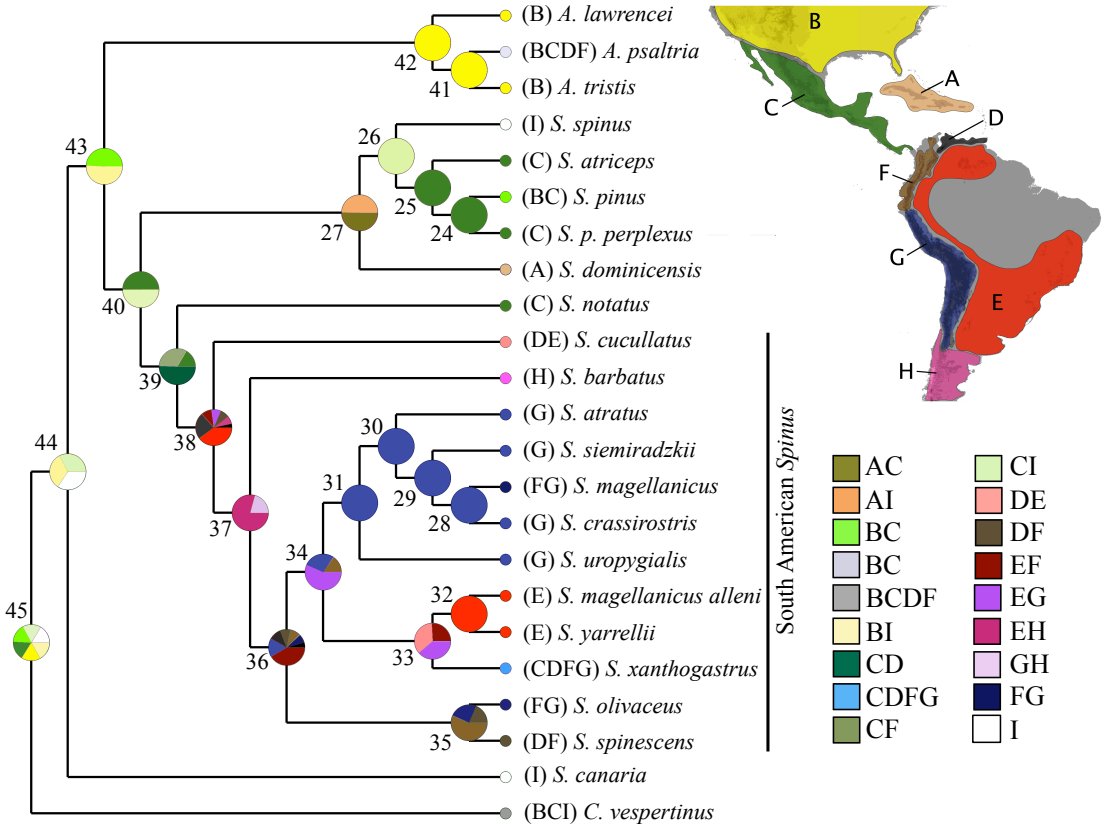

Fig. 3. Dispersal-Vicariance Results for *Astragalinus* and *Spinus*. Linearized S-DIVA results based on the species tree. Biogeographic regions A-I are color coded according to embedded key. Pie charts at the nodes indicate support for particular ancestral range reconstructions. Outgroups are *Coccothraustes vespertinus* and *Serinus canaria.*

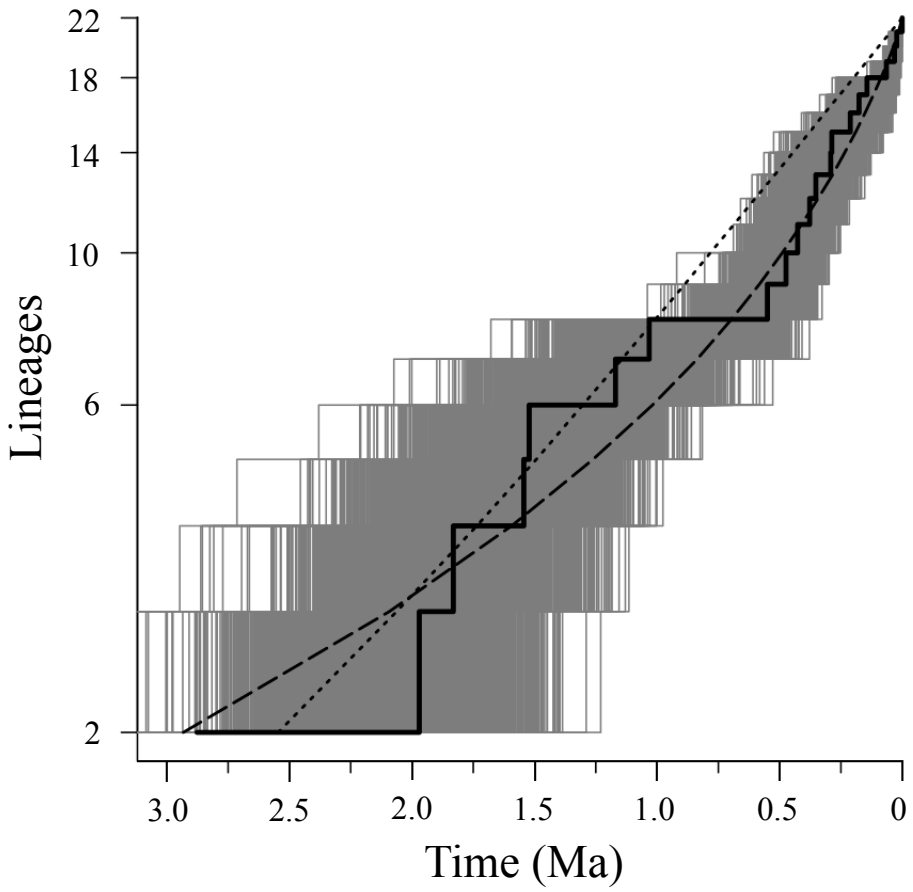

Fig. 4. Lineage through time plot for *Astragalinus* and *Spinus*. Based on the complete dated species tree (solid black), 5000 complete species trees from *BEAST posterior distribution (gray), average pure-birth simulation (dotted black) and average birth-death simulation (dashed black).

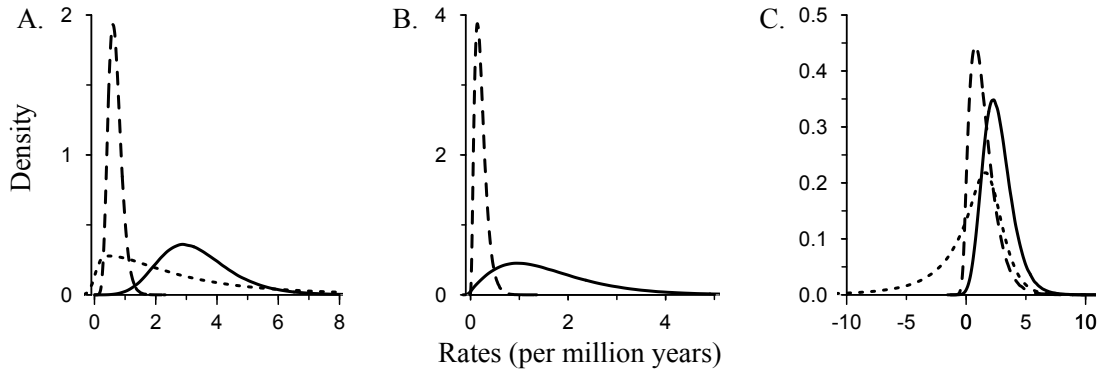

Fig 5. Parameter estimation from the no-extinction GeoSSE diversification model for *Astragalinus* and *Spinus.* A) Speciation rates within the Andes (solid), outside the Andes (dashed) and speciation associated with movement between regions (dotted). B) Dispersal rates out of the Andes (solid) and into the Andes (dashed). C) Differences between rates of different regions: the difference between the speciation rate within the Andes and outside of it (solid), the difference between disperal rate out of the Andes and into them (dashed), and the difference between the speciation rate within the Andes and the speciation rate associated with movement between regions (dotted).

Table 1. Primers and annealing temperatures (°C).

Locus	Length	°C	Primer	Sequence (5'-3')	Citation
cytb	1020	54	L14841	GCTTCCATCCAACATCTCAGCATGATGAAA	Kocher et al., 1989
			H4a	AAGTGGTAAGTCTTCAGTCTTTGGTTTACAAGACC	Harshman 1996
ND2	1021	50	L5219	CCCATACCCCGAAAATGATG	Sorenson et al., 1999
			H6313	CTCTTATTTAAGGCTTTGAAGGC	Sorenson et al., 1999
	369	50	ND2s2F	CCCTCCAATCACCCTCCTAT	This study
			ND2s2R	AGTGTGGCTGCAGGGATTAT	This study
	395	50	ND2s3F	TCACAATCAGCCTGCTCCTA	This study
			ND2s3R	GGGTGATTGGAGGGAGTTTT	This study
ND3	351	50	L10647	TTYGAAGCMGCMGCMTGATSCTG	Mindel et al., 1999
			H11151	GATTTGTTGAGCCGAAATCAA	Chesser 1999
MUSK	453	50	MUSK-E3F	CTTCCATGCACTACAATGGGAAA	Clark and Witt, 2006
			MUSK-E4R	CTCTGAACATTGTGGATCCTCAA	Clark and Witt, 2006
	641	50	AnaMUSK-F	AAATAATAGAAGGCTTAAAGG	DuBay and Witt, 2012
			AnaMUSK-R	CTCTGGACATTGTGTATCCTT	DuBay and Witt, 2012
MYO2	747	59	E2F1	GAAGATCTGAAGAAACATGGAGCTA	DuBay and Witt, 2012
			E3R1	CAATGACCTTGATAATGACTTCAGA	DuBay and Witt, 2012

Table 2. Biogeographic regions assigned to taxa for S-DIVA and BBM analysis. The Andes in this analysis are divided into three regions: the Northern Andes (F), which extend from the Colombian Andes to northern Peru (~6° S); the Central Andes and Austral Andes (G), which reach from 6° S to 25° S, near Catamarca, Argentina; and the Chilean Andes (H) which extend from 25° S to the tip of South America. The area that includes the Chilean Andes also includes Patagonia.

Area:	**Region code:**
Caribbean	A
North America	B
Central America	C
Northern Coastal Mountains, South America	D
Dry eastern lowlands, South America	E
Northern Andes	F
Central Andes and Austral Andes	G
Chilean Andes and Patagonia	H
Europe and North Africa (Old World)	I
Taxon:	
Coccothraustes vespertinus	BCI
Serinus canaria	I
Astragalinus lawrencei	B
Astragalinus psaltria	BCDF
Astragalinus tristis	B
Spinus atratus	G
Spinus atriceps	C
Spinus barbatus	H
Spinus crassirostris	G
Spinus cucullatus	DE
Spinus dominicensis	A
Spinus magellanicus	FG
Spinus magellanicus alleni	E
Spinus notatus	C
Spinus olivaceus	FG
Spinus pinus	BC
Spinus siemiradzkii	G
Spinus spinescens	DF
Spinus spinus	I
Spinus uropygialis	G
Spinus xanthogastrus	CDFG

Table 3. Divergence dates from BEAST mtDNA only and *BEAST total evidence analyses. Node names correspond to Fig. 2; support for each node is indicated as bootstrap / posterior probability from MrBayes / posterior probability from BEAST (mtDNA) or *BEAST (all loci). Ages are in millions of years; age ranges correspond to 95% highest posterior density (HPD). Node names correspond to Fig. 2. Alternate topologies for the placement of *Spinus dominicensis* indicated as 27A (mtDNA only trees, Supplemental Fig. 1) and 27B (total evidence trees, Figs. 2 and Supplemental Fig. 4). NA indicates support not estimated while -- indicates a node is not present in a particular analysis.

Node	mtDNA support	BEAST mean age	BEAST age range	All loci support	*BEAST mean age	*BEAST age range
25	99 / 1 / 1	0.610	0.264 - 1.015	100 / 1 / 0.99	0.352	0.015 - 0.635
26	-- / 0.88 / 0.96	1.399	0.829 - 2.009	72 / 0.99 / 0.95	1.170	0.697 - 1.699
27A	79 / 0.98 / 0.90	1.650	1.206 - 2.457	-- / -- / --	---	---
27B	-- / -- / --	---	---	67 / 0.94 / 0.76	1.522	1.100 - 1.962
30	-- / -- / --	---	---	-- / -- / 0.99	0.064	0.037-0.097
31	35 / 0.86 / 0.99	0.399	0.262 - 0.549	51 / 0.76 / 0.93	0.212	0.096 - 0.317
37	43 / -- / 0.34	0.560	0.406 - 0.721	51 / -- / 0.56	0.474	0.353 - 0.607
38	96 / 1 / 1	0.622	0.450 - 0.820	98 / 1 / 0.97	0.550	0.399 - 0.721
39	97 / 1 / 1	1.128	0.792 - 1.481	98 / 1 / 0.99	1.032	0.558 - 1.449
40	95 / 1 / 1	1.927	1.456 - 2.457	0.97 / 0.99 /0.99	1.970	1.520 - 2.448
41	89 / 0.97 /0.98	1.262	0.715 - 1.848	55 / 0.89 / 0.78	1.544	0.983 - 2.149
42	88 / 1 / 0.99	1.718	1.064 - 2.409	96 / 1 / 0.99	1.831	1.251 - 2.452
43	87 / 1 / 0.88	2.768	2.076 - 3.543	94 / 1 / 0.99	2.877	2.194 - 3.555
44	na / na / 0.77	3.457	2.396 - 4.488	na / na / 1	4.127	3.115 - 5.197
45	na / na / 1	4.688	2.883 - 6.983	na / na / 1	14.438	10.616 - 18.615

Table 4. S-DIVA and Bayesian Binary Markov dispersal and vicariance anaylsis results for the complete species tree. Node names correspond to Fig. 3. PP indicates posterior probability of node in the phylogeny. Letter codes correspond to biogeographic regions: A- Caribbean; B- North America; C- Central America; D- Northern coastal mountains, South America; E- Eastern lowlands, South America; F- Northern Andes; G- Central Andes and Austral Andes; H- Chilean Andes and Patagonia; and I- Europe and North Africa. Support for specific reconstructions are indicated by the integer to the right of the letter code. Reconstructions with <5% support omitted.

Node	P	S-DIVA	BBM
24	1.00	C 100	C 85, BC 15
25	1.00	C 100	C 97
26	0.95	CI 100	C 56, I 33
27	0.76	AI 50, AC 50	A 42, C 40, I 9
28	0.58	G 100	G 83, FG 17
29	0.90	G 100	G 96
30	1.00	G 100	G 99
31	0.93	G 100	G 97
32	1.00	E 100	E 96
33	0.20	DE 39, EG 35, EF 26	G 38, FG 16, F 13, EG 5
34	0.11	EG 57, G 28, FG 15	G 65, FG 17, F 10
35	0.34	F 57, FG 25, DF 18	G 63, FG 28, DF 6
36	0.25	EF 42, G 16, D 11, DF 9, F 9, FG 8	G 43, F 28, FG 20
37	0.56	EH 79, GH 21	H 45, G 21, F 10, D 8
38	0.97	E 40, D 24, EF 9, EG 8, DF 8, EH 7	D 42, E 24, C 10, H 8
39	1.00	CD 51, CF 33, C 16	C 81, D 5
40	1.00	CI 50, C 50	C 68, B 9, A 8, I 5
41	0.78	B 100	B 90
42	1.00	B 100	B 97
43	1.00	BI 51, BC 49	B 68, I 11, C 10
44	1.00	I 34, BI 33, CI 33	I 59, BI 17, B 9
45	1.00	I 17, BI 17, CI 17, BC 17, B 16, C 16	BI 54, CI 30, BC 8

Table 5. Standard diversification models fit to *BEAST *Astragalinus* and *Spinus* phylogeny. LH indicates the negative log likelihood of the model. Rates are per million years. T_s indicates time, in million years ago, of rate shift in Pure-birth, 2-rate model. K and x correspond to parameters that modulate speciation rate in density dependent models.

Model	LH	AICc	Akaike weights	Birth rate	Death rate	2nd Birth rate	T_s	K	x
Pure-birth	24.12	-46.03	0.340	1.04	0	---	---	---	---
Birth-death	25.17	-45.68	0.286	1.77	1.36	---	---	---	---
Pure-birth, 2-rate	25.71	-44.00	0.123	1.00	0	1.56	0.064	---	---
Density Dependent exponential	24.54	-44.41	0.151	0.51	0	---	---	---	-0.33
Density Dependent logistic	24.12	-43.57	0.100	1.04	0	---	---	$3.6x10^5$	---

Table 6. Geographic state speciation & extinction model (GeoSSE) maximum likelihood model fits for *Astragalinus* and *Spinus*. Geographic regions defined as Andes (A) and outside the Andes (OA). Parameters include measures of model fit (log likelihood, LH; AICc; Akaike weights), speciation rate (S), extinction rate (X) and dispersal rates (D) in per million years. The full, seven-parameter model was constrained to four other models including setting extinction in both regions equal to one another (X_A~X_{OA}) and to zero (X_A~X_{OA}~0), setting dispersal in both regions equal to one another (D_A~D_{OA}), and setting speciation in both regions equal to one another (S_A~S_{OA}).

Model	LH	AICc	Akaike Weight	S_A	S_{OA}	$S_{A/OA}$	X_A	X_{OA}	D_A	D_{OA}
Full	-27.83	67.41	0.066	3.047	0.591	3.01×10^{-9}	2.447	4.43×10^{-7}	0.929	0.366
X_A~X_{OA}~0	-26.91	62.45	0.794	2.719	0.667	5.30×10^{-8}	0	0	0.580	0.138
X_A~X_{OA}	-29.91	70.19	0.017	2.508	1.009	1.75×10^{-6}	2.407	2.407	3.16	7.00×10^{-9}
D_A~D_{OA}	-28.03	66.31	0.115	3.492	0.584	1.05×10^{-7}	3.461	1.00×10^{-8}	0.66	0.66
S_A~S_{OA}	-30.60	71.62	0.008	1.069	1.069	2.65×10^{-5}	6.45×10^{-6}	0.228	0.55	0.14

Table 7. Parameters and cross correlations for no-extinction GeoSSE model for *Astragalinus* and *Spinus*. Geographic regions defined as the Andes (A) and outside the Andes (OA). Parameter values generated via MCMC searches across 1000 *BEAST species trees; means and time-series standard errors presented. Speciation (S) and dispersal (D) rates given per million years. $S_{A/OA}$ indicates the speciation rate associated with movement between regions.

		Correlations				
	Mean ± SE	**S_A**	**S_{OA}**	**$S_{A/OA}$**	**D_A**	**D_{OA}**
S_A	3.358 ± 0.0085	1	---	---	---	---
S_{OA}	0.679 ± 0.0014	- 0.0118	1	---	---	---
$S_{A/OA}$	2.891 ± 0.0116	0.1872	- 0.0501	1	---	---
D_A	1.605 ± 0.0059	0.0835	- 0.0651	0.1276	1	---
D_{OA}	0.212 ± 0.0003	- 0.0205	0.0688	0.0201	- 0.0205	1

CHAPTER II: The challenge of testing introgression when introgression obscures phylogeny: the case of the South American siskins

ABSTRACT:

Recent genomic studies from across the tree of life have confirmed that introgression of genes from one species to another is a regular occurrence, at least during the early stages of diversification. If a rapid radiation were accompanied by rampant introgression, its phylogeny could become impossible to resolve. This creates a paradox, because tests for introgression require that the phylogeny is known without error. Here we attempted to estimate the phylogeny and test for introgression in a recent, rapid bird radiation, the South American siskins (*Spinus*). We used genotype-by-sequencing to assemble and align ~60,000 loci. We applied maximum likelihood, quartets, and species-tree phylogenetic methods on data matrices that varied in assembly parameters and included different proportions of missing data. Four conflicting topologies emerged, each with strong support in some analyses. We used each of these topologies to design formal tests for introgression using three phylogeny-based methods. All three methods and all four topologies revealed evidence of introgression between *Spinus* populations. This result implicates introgression as the cause of phylogenetic discordance, regardless of which topology represents the true species history. Patterns of apparent mitochondrial genome capture among siskin species were consistent with the introgression events revealed by the nuclear loci and suggest that genetic adaptation played a major role in the movement of genes among species.

INTRODUCTION:

The ultimate factors that contribute to rapid diversification remain obscure despite intense study. Different authors have proposed a role for ecological opportunity (Dobzhansky, 1948; Mahler et al., 2010; Schluter, 2000), connectivity among habitats (Fjeldså, Bowie, and Rahbek, 2012; Gavrilets and Losos, 2009; Schluter, 2000) and intrinsic characteristics that act as key innovations (Lerner et al., 2011; Price, 2011; Simpson, 1953). Recent genomic studies, however, have shown that introgression among close relatives may play a significant role in generating ecological variation and facilitating diversification in rapid radiations (Keller et al. 2012; Lamichhaney et al., 2015; Stankowski and Streisfeld, 2015; The Heliconius Genome Consortium et al., 2012). On a fundamental level, the direct consequence of introgression is reticulate gene histories in the genomes of the introgressed taxa; introgressed genes reflect their own history, not that of the species tree. Reticulate gene histories from introgression pose a major challenge to phylogenetic estimation (Wen et al., 2016; Yu et al., 2011). Paradoxically, accurate phylogenies are required for many formal introgression tests (Patterson et al., 2012). Identifying best practices for phylogenetic analysis when multiple lineages have introgressed within a clade is essential to understanding the significance of introgression to global biodiversity.

Introgression, the movement of alleles from one lineage into the gene pool of another, can be a powerful evolutionary force (Mallet, 2005; Seehausen, 2004). Introgression challenges existing species limits and may result in species collapse (Seehausen, Alphen, and Witte, 1997; Vonlanthen et al., 2012), the elimination of introgressed individuals through negative selection (Barton, 1983; Fishman and Willis,

2001; Rieseberg, Whitton, and Gardner, 1999), and the reinforcement of existing species barriers (Sætre and Sæther, 2010; Servedio and Noor, 2003; Taylor et al., 2006). Alternatively, when novel genetic material is successfully transferred into a naive species (Huerta-Sánchez et al., 2014; Mallet, 2005; Reich et al., 2010), introgression may act as a creative force by producing new gene combinations and original phenotypes (Abbott et al., 2013; Burke and Arnold, 2001; Dobzhansky, 1937; Wright, 1931). Seehausen (2004) argued that introgression between close relatives, in particular, should elevate the heritable diversity of ecologically relevant traits and facilitate a complete exploration of the ecological niche space available.

Recent genomic studies have revealed that introgression occurred multiple times within several adaptive radiations (Keller et al., 2012; Lamichhaney et al., 2015; Stankowski and Streisfeld, 2015; The Heliconius Genome Consortium et al., 2012), sometimes to such an extent that the true species history is represented in a minority of the genome (Fontaine et al., 2015). The ecological consequences of these bouts of introgression have included the interspecific transfer of an allele with a major effect on a known selective target, bill shape, in the Galapagos finches (Grant, Grant, and Petren, 2005; Grant and Grant, 2016; Lamichhaney et al., 2015) as well as the generation of novel phenotypes and new species in African rift lake cichlids (Genner and Turner, 2012; Keller et al., 2012; Parsons, Son, and Albertson, 2011). Adaptive introgression, the transmission of beneficial alleles from one species to another, has also been identified in rapid continental radiations of plants and invertebrates (Fontaine et al., 2015; Pease et al., 2016; Stankowski and Streisfeld, 2015; The Heliconius Genome Consortium et al., 2012). In the most challenging of these investigations, authors relied on whole genomic

sequencing (Lamicchaney et al., 2015; The Heliconius Genome Consortium et al., 2012; Fontaine et al., 2015). Similarly, many of the network-based phylogenetic innovations designed to address introgression rely on long homologous regions of the genome (Wen, Yu, and Nakhleh, 2016; Yu, Degnan, and Nakhleh, 2012). This level of sequencing is unlikely to be available to many researchers, despite advancing next-generation sequencing methods. The need to expand the survey for introgression across biodiversity, however, is essential to understand the importance of introgression in ecology and diversification across the globe. A more reticulate tree of life challenges assumptions in the evolution of phenotypic traits (Rabosky et al., 2013), the origins of new species (Grant and Grant, 2009; Schluter, 2000; Seehausen, 2004) and phylogenetic approaches for shallow and deep divergences (Mallet, Besansky, and Hahn, 2015; Wen, Yu, and Nakhleh, 2016; Yu, Degnan, and Nakhleh, 2012).

With the objective to uncover and address the phylogenetic challenges that accompany introgression with sequencing short of whole genomes, we evaluated the evidence of introgression in a rapid, continental radiation, the South American siskins (Fringillidae: *Spinus*). Despite deep taxonomic divergence, this finch radiation bears a striking resemblance to the Galapagos finches (family Thraupidae). Both the Galapagos finches (~0.9 million years old) and South American *Spinus* clades (~0.55 million years old) are young, have high diversification rates and demonstrate extensive sympatry among closely related species. Species are morphologically differentiated, including bill shape and size. *Spinus* diversified across the South American continent, with exceptionally high rates of diversification within the high Andes. Due to the similarities between the Galapagos finches and *Spinus* radiations, introgression is possible among

Spinus and the clade is a good candidate for investigating phylogenetic signal in the presence of introgression. Further, evidence of introgression in *Spinus* will add perspective on the introgression in rapid bird radiations and the ecological conditions in which it occurs.

Introgression may be an important factor in the diversification of *Spinus* since in the high Andes of Ecuador, Peru and Bolivia, there is sympatry among three species (*Spinus atratus, S. crassirostris* and *S. magellanicus*) during the breeding season. A fourth species, *S. uropygialis*, joins the community during the non-breeding season (Fjeldså and Krabbe, 1990; Ridgely and Tudor, 1989). The sympatric species differ in plumage, body size and habitat choice. *S. crassirostris* also has a significantly deeper, more robust bill than its congeners (Ridgely and Tudor, 1989, Beckman, unpubl. data). Lastly, these species differ in their elevational ranges: *S. atratus* and *S. crassirostris* are found above ~3500 m, *S, uroypygialis* has a lower elevational limit of ~2000 m (Ridgely and Tudor, 1989), and *S. magellanicus* is found from sea level up to 5000 m (Fjeldså and Krabbe, 1990). In a previous multilocus phylogenetic study, we found that these four species shared a most recent common ancestor ~0.2 million years ago and are each other's closest relatives (Beckman and Witt, 2015).

The sympatry of recently diverged taxa in the high central Andes is likely to challenge reproductive isolation among *Spinus*. Indeed, we discovered extensive mitochondrial DNA (mtDNA) haplotype sharing among three *Spinus* species whose distributions overlap, with modest levels of mtDNA structure explained by species limits, geography and elevation (Beckman and Witt, 2015). Incomplete lineage sorting (ILS), caused by recent divergence times, short internal branches, and large population sizes

could be the source of the mtDNA haplotype sharing pattern among central Andean siskins (Edwards, 2009). Alternatively, the pattern may be the result of introgression of the mitochondrial genome among *Spinus* species. The morphometric and plumage differences among *Spinus* species (Beckman, unpubl. data) suggest that if introgression is present, or has been in the past, it has been rare. Frequent hybridization is often identifiable by morphology in birds (McDonald et al., 2001; Rheindt et al., 2014).

Mitochondrial introgression has been inferred in many animals, including most families of birds (Rheindt and Edwards, 2011), mammals (Good, Vanderpool, Keeble, and Bi, 2015), lizards (Leaché, 2009) and fish (Dowling and DeMarais, 1993). In many cases, natural selection may drive introgression at this locus, however this has been rarely investigated (but see Toews et al., 2013). mtDNA is a potential target for natural selection with respect to elevation due to its role in cellular respiration (Ballard and Whitlock, 2004; Blier, Breton, Desrosiers, and Lemieux, 2006; Boratyński et al., 2011; Mishmar et al., 2003). Previous studies showed that natural selection on mtDNA at high elevation may relate to thermal tolerance (Gering, Opazo, and Storz, 2009) and the reduction of reactive oxygen species (ROS) under hypoxic conditions (Graham R. 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* (Cheviron and Brumfield, 2009).

The introgression of mtDNA from high elevation restricted taxa into generalist species may permit the recipient species to expand their upper elevational range and persist at higher elevations. Range expansions accompanied by adaptive introgression have been documented in several systems (Fontaine et al., 2015; Lewontin and Birch,

1966) including the introgression of an allele that might be responsible for increased survival in high altitude hypoxia (Hackinger et al., 2016; Huerta-Sánchez et al., 2014), but see Lou et al. (2015) .

In this study, we assessed the evidence for mitochondrial and nuclear introgression among *Spinus* in the central Andes. Our null hypothesis (H_o) was that no interspecific introgression has occurred among *Spinus* in either genome. To test this hypothesis with mtDNA, we asked whether the genetic distances between interspecific haplotypes, and the geographic structure of the locus across the Andes were consistent with ILS as the sole source of mtDNA haplotype sharing. Our ILS expectations were contingent on a phylogenetic hypothesis constructed from high resolution nuclear DNA (nDNA) to serve as the "true" *Spinus* species tree. Our alternative hypothesis for mtDNA (H_a1) was that haplotype sharing reflects rare interspecific introgression events among Andean *Spinus.* To test our null hypothesis with nDNA, we assessed population structure in Andean *Spinus* with upwards of ~60,000 loci generated from Genotyping-by-Sequencing (Elshire et al., 2011). We used the phylogenetic hypotheses to conduct formal tests for introgression for all possible species pairs in the nuclear genome using both fixed and polymorphic biallelic sites; results that deviated from neutral expectations offered support for our alternative hypothesis (H_a2), that interspecific introgression has occurred in nDNA in Andean *Spinus*. Lastly, we considered if support for our alternative hypotheses was consistent between the mtDNA and nDNA. Support for introgression in only one of the two genomes does not undercut the evidence for introgression; however concurrent results between the mtDNA and nDNA suggest that the conclusion is robust for both.

METHODS:

Taxonomic Sampling:

We used museum collections and field expeditions to compile frozen tissue samples for 148 individuals from the genus *Spinus* including: three high elevation restricted species, *S. atratus* (N=8), *S. crassirostris* (N=16), and *S. uropygialis* (N=20); one widespread, polytypic species, *S. magellanicus*, sampled broadly across its distribution (Fig. 1) including Ecuador (N=3), Peru (N=89), Bolivia (N=4) and northern Argentina (N=4). These taxa as referred to as Andean *Spinus* hereafter. We also included one known outgroup from northern South America (Beckman and Witt, 2015), *S. cucullatus* (N=4). Previous work (Beckman and Witt, 2015) suggested this sampling should accurately represent mtDNA genetic diversity for these species.

DNA Sequencing:

Mitochondrial genes:

We extracted DNA from muscle tissue samples with the Qiagen DNeasy Blood and Tissue kit (Qiagen, Valencia, CA) using the standard protocol. For each sample, excepting 53 *S. magellanicus,* we amplified and sequenced the mitochondrial genes cytochrome b (*cytb*), NADH dehydrogenase II (ND2) and NADH dehydrogenase III (ND3) as described in Beckman and Witt (2015). The remaining *S. magellanicus* were sequenced for ND3 to extend geographic sampling; ND3 was sufficient to identify major mtDNA haplotypes (see Results). Sequencing was conducted at the University of New Mexico Molecular Core Facility (Albuquerque, NM). Mitochondrial loci were evaluated by eye in Sequencer 4.10.1 (GeneCodes Corp., Ann Arbor, MI) for quality. Sequences

for each locus were aligned first with Muscle 3.7 (Edgar, 2004) on the CIPRES portal (Miller et al., 2010) and alignments were confirmed by eye in MacClade 4.08 (Maddison and Maddison, 2005).

Genotyping-by-sequencing:

We used genotyping-by-sequencing (GBS) (Elshire et al., 2011) to generate sequences for thousands of loci. We chose 40 individuals (Table 1) for GBS based on the results of our mitochondrial sequencing; these included four samples per species of *S. atratus, S. crassirostris* and *S. uropygialis*, one *S. cucullatus*, and 27 samples of *S. magellanicus* from four regions: central Peru including the departments of Lima and Ancash (16), Cusco, Peru (5), Cochabamba, Bolivia (3), and northern Argentina (3).

To generate libraries, we followed the protocol outlined in Parchman et al. (2012). Briefly, this method employed two restriction enzymes (*Eco*RI and *Mse*1) to digest whole genomic DNA. Individual samples were barcoded with unique 8-10 base pair (bp) by ligation, then amplified with PCR (see Parchman et al., 2012 for details). After amplification, we pooled samples and size-selected 500-600 bp DNA fragments using gel extraction and a Qiagen gel purification kit (Qiagen, Valencia, CA). Product from gel extractions were pooled and submitted for sequencing 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 processed the raw reads by demultiplexing them and removing barcodes with the *process_radtags* function in the program Stacks (Catchen et al., 2011), resulting in 89

bp reads. We utilized the Stacks, v1.35 (Catchen et al., 2013; Catchen et al., 2011) pipeline to construct *de novo* assemblies from our GBS data. Stacks is a computationally robust, efficient, open source software designed to recover homologous loci from closely related individuals using short read sequence data without the aid of a genome (Catchen et al., 2013). Stacks works effectively across different levels of divergence due to its flexibility which depends on three user-designated parameters. These parameters specify the raw read depth required to assemble a locus (*m*), the maximum distance in nucleotides permitted within a locus in an individual (*M*) and the maximum distance 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 an initial exploration of *m* = 2, 3, 4 and 5 with a single individual, we set *m*=4 to maximize the number of high-confidence loci recovered. We took steps to eliminate potentially confounded loci from highly repetitive regions of the genome during locus construction (--max_locus_stacks = 3, see Catchen et al., 2013) and by including the rxstacks filters (--lnl_lim, --conf_filter) in our pipeline (see Stacks online manual). 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% similarity), B5 was most lenient (93% similarity).

We evaluated assemblies B1-B5 by calculating the average number of unique alleles per locus for each dataset. 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). Alternatively, comparable values of the average unique alleles per locus across different similarity thresholds should indicate the assembly is robust to the changes in Stacks parameters. We also constructed maximum likelihood phylogenetic trees including all individuals using RAxML v 8.1.17 (Stamatakis, 2014) to test the consistency of topology and node support among the five datasets (see Phylogenetics for details). Of the acceptable datasets, we chose the most restrictive (highest percent similarity) dataset to use for all subsequent analyses in order to reduce the number of over-merged loci, that is, distinct loci that have been incorrectly combined. This assembly will be referred to as **B3** hereafter.

The Stacks parameters *r* and *p* in the populations program were used in downstream analyses to manipulate the amount of missing data permitted. The proportion of individuals required per population for a locus to be included in a dataset was defined by *r*. The number of populations required for a locus to be included was defined by *p* (Catchen et al., 2013). We treated each species as a taxonomic unit in the populations program except for *S. magellanicus*. Based on previous work and preliminary analyses, we analyzed as *S. magellanicus* as three populations based on latitude: 8^{o} to 14^{o} S, hereafter Peru; 15^{o}-21^{o} S, hereafter Bolivia, and 22^{o} to 28^{o} S, hereafter Argentina.

Population Structure:

Network analysis:

To visualize the mitochondrial genetic variation among individuals, we concatenated the genes *cytb*, ND2 and ND3 for all individuals with complete data,

identified haplotypes and constructed a haplotype network using an infinite sites model in R v3.2.4 (R Core Team, 2016) with the packages adegenet v2.0.1 (Jombart and Ahmed, 2011) and pegas v0.6 (Paradis, 2010). We performed the same analysis with the ND3 gene only that included all 100 *S. magellanicus* individuals.

To test our null hypothesis that introgression has not occurred among Andean *Spinus*, we had two predictions. First, previous analyses (Beckman and Witt, 2016) showed that individuals of the widespread species *S. magellanicus* possessed three divergent mtDNA haplotype groups. If this mtDNA diversity is due to ILS, it follows that the ancestral population of this lineage was large. Current populations of this ancestral population could be also be large, maintained by gene flow among populations, which should result in similar allele frequencies for the three divergent haplotypes across geography. Alternatively, current populations could be smaller, with or without gene flow among populations. In this scenario, the allele frequency of mtDNA haplotypes should drift randomly within isolated populations, and allele frequencies should vary among populations, but not predictably with geography. Finding the geographic mtDNA variation in *S. magellanicus* similar to these two scenarios would permit us to accept our null hypothesis, H_o1. A third scenario, in which mtDNA variation in *S. magellanicus* is structured across geography is consistent with ILS if populations are small and isolated with biased gene flow among certain populations. However, geographically structured mtDNA is also concordant with a rare introgression event in which an allele was introduced into *S. magellanicus* via introgression and diffused, via neutral or non-neutral processes, by gene flow across the landscape.

Our second prediction was that, when two species possess the same mtDNA haplotype group, the mean *interspecific* mtDNA genetic distance should be equal to or greater than the mean *intraspecific* genetic distance of a species. Note, we did not assess intraspecific genetic distance in *S. magellanicus* since it is polytypic. The expected interspecific genetic distance should be greater between species pairs that are more divergent in the nDNA phylogeny. This expectation is founded on principle that isolated lineages should acquire unique silent variation over time; so even if two lineages inherit one ancestral allele, over time those alleles should diverge.

Phylogenetics:

We constructed a maximum likelihood (ML) phylogenetic tree from the GBS data using the hybrid, MPI/PThreads version of RAxML v8.1.17 (Pfeiffer and Stamatakis, 2010; Stamatakis, 2014). The computational efficiency of the ML approach, based on an molecular evolutionary model (GTR + Γ for the hybrid version), allowed us to include of tens of thousands of loci. One possible weakness of the method is that it assumes a single phylogenetic history for all loci, an assumption that is almost certainly false and could lead to a misrepresentation of the true species (Edwards, 2009). Our ML phylogenetic analysis had two stages: first, to detail the consequences of different Stacks assemblies (see above) and missing data on the topology and resolution, and second, to run a complete analysis on the best-performing dataset. First, we constructed a series of individual-level SNP datasets in which we varied *r* from 0.5 to 0.75*, p* from 4 to the maximum possible of 7*,* and the Stacks assembly B1 through B5. We concatenated all variant sites within a dataset, and used the Felsenstein invariant site correction with

molecular evolution model ASC_GTRGAMMA in RAxML to account for the invariant sites present in the original sequence data (Leaché et al., 2015). 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). In the most extreme case, the difference of branch lengths due to invariant sites may impact which tree topology is best supported (Leaché, personal comm.). The Felsenstein invariant site correction also reduced computational time compared to an analysis with complete, concatenated loci. We ran 500 bootstrap replicates for each ML tree. For the final phylogeny, we chose the best Stacks assembly (B3). We then selected the dataset which had the least amount of missing data while still retaining enough phylogenetic information to resolve the evolutionary relationships within Andean *Spinus*. We constructed the final phylogeny using concatenated, complete loci (89 bp per locus), GTRGAMMA and 500 bootstrap replicates.

We also generated a phylogenetic tree using the multi-species coalescent approach implemented in the program SNAPP (Bryant et al., 2012) using the GBS data. In this approach, based on a combination of diffusion and coalescent theories, unlinked, bi-allelic sites have independent gene histories that are constrained by the ultimate species tree (Edwards 2009, Bryant et al., 2012). In it, heterogeneity among the historical signals from different genes is attributable solely to differences in the time of coalescence and incomplete lineage sorting (Bryant et al., 2012). We selected 20 high coverage individuals including the single *S. cucullatus* and 2 - 6 individuals of the remaining, previously defined populations (Set E in Table 1). In the final SNP dataset, we selected 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 these SNPs into 3 equal sized datasets and ran each in a separate SNAPP run. For the multi-species coalescent, we grouped *S. magellanicus* from Bolivia and Argentina, hereafter *S. magellanicus* Bol. Arg., into a single operational taxonomic unit analysis based on RAxML phylogenetic tree results. We also grouped *S. magellanicus* from Cusco, Lima and Ancash, hereafter *S. magellanicus* Peru. We estimated the mutation rates U and V within BEAUti 2 from the data and used the default prior specifications. Each dataset was run for 3 million generations with a 10% burnin. We used Tracer, v 1.6 (Rambaut, Suchard, Xie, and Drummond, 2014) to evaluate convergence. Once we established concordance among independent runs, we combined trees from each run and evaluated support for individual nodes in Densitree v. 2.2.4 (Bouckaert and Heled, 2014).

Lastly, we used a quartet inference approach based in SVDquartets (Chifman and Kubatko, 2014, 2015), implemented in PAUP* v 4.0a147 (Swofford, 2002). This method, based on coalescent theory, calculates the best tree topology by identifying a valid split across all loci for each quartet. The best quartet trees are then assembled with QFM (Reaz, Bayzid, and Rahman, 2014) to create the final phylogeny. We included 29 individuals (Set C in Table 1), and constructed both "lineage trees" and "species trees". In "lineage trees", each individual was treated as a terminal branch. In "species trees", we defined members of each population *a priori* as in SNAPP; for each quartet, one individual was chosen to represent each species. We extracted the first variant site from each locus that was present in >66% of individuals in six of seven populations (r = 0.66, p=6) for these analyses. SVDquartets should permit missing data (Chifman and Kubatko,

2015); however, we used similar parameters to those used in SNAPP to permit methodological comparisons.

Principal component analysis:

We assessed the population genetic structure of all Andean *Spinus* individuals with the GBS data by performing a principal component analysis (PCA) in R v3.2.4 (R Core Team, 2016) with the R-package adegenet v2.0.1 (Jombart and Amhed, 2011). The input included all variant sites from assembly B3 with $r = 0.5$. Since 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.

Clustering analysis:

We used the model-based likelihood clustering algorithm implemented in the program Admixture v1.3 (Alexander, Novembre, and Lange, 2009) to assign individuals to populations and admixture fractions to individuals without *a priori* labels. The input included the first SNP from every variable locus in assembly B3 with $r = 0.5$ and $p = 6$. Outgroup *S. cucullatus* was excluded from this analysis. To identify the appropriate number of populations (k) for the data, we calculated the cross-validation error for k equal to one through 10 and identified k with the lowest values as suggested by Alexander, Novembre and Lange (2009).

Introgression tests:

We conducted three tree-based methods to explicitly test whether introgression has occurred among Andean *Spinus*. For these methods, we defined 31 tests (Test 1-31) based on well-supported four taxon trees derived from our "Best" RAxML phylogenetic analyses (Table 2); a subset of these tests were valid for the alternative topologies recovered in SNAPP and SVDquartets (see Results). We also included geographic populations instead of complete clades in a subset of tests (*e.g.* instead of using all members of *S. magellanicus* Bol. Arg., we sometimes used only *S. magellanicus* Bolivia). Including geographic populations instead of complete clades gave us insight into how the signal of introgression might vary across the landscape (Eaton et al., 2015). In certain analyses, we divided *S. magellanicus* Peru into *S. magellanicus* Central Peru (CP), containing samples from the departments of Lima and Ancash, and *S. magellanicus* Southern Peru (SP), containing birds from the department of Cusco. Similarly, *S. magellanicus* Bol. Arg. treated as *S. magellanicus* Bolivia and *S. magellanicus* Argentina in some tests. For all methods, we defined positive evidence for introgression (rejecting H_o2) as an absolute value of the z-score > 2.0; this is equivalent to a p-value for a two-tailed test of 0.0455.

We performed the ABBA/BABA method (Green et al., 2010) for fixed sites on every test. For this method, we considered only biallelic sites in which sister taxa were fixed for different alleles, *e.g.* A and B. The essence of the ABBA/BABA method is that, given a strictly bifurcating population history among four taxa, *e.g.* (((Taxon1, Taxon2) Taxon3) Taxon 4), we expect two allelic patterns with equal probability in unlinked loci across the genome. The ABBA pattern indicates Taxon2 and Taxon3 share one allele (B)

while Taxon1 and Taxon4 share another (A). The BABA pattern indicates Taxon1 and Taxon3 share one allele (B) to the exclusion of Taxon2 and Taxon4 (A). Across the genome, we predict a 1:1 ratio of ABBA and BABA patterns at unlinked loci, and the corresponding summary statistic D should equal zero (Green et al., 2010; Patterson et al., 2012; Rheindt and Edwards, 2011). For every test, we extracted the first variable site from loci that were fixed within every taxon ($r = 0$, $p = 4$). We then counted ABBA and BABA sites in order to calculate D, and tested the null hypothesis that D = 0 through 1000 bootstrap replicates. The latter steps were conducted in R v3.2.4 (R Core Team, 2016) with code modified from the package evobiR (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 (Patterson et al., 2012).

We also conducted the ABBA/BABA method on polymorphic loci for every test possible. Again, the prediction was that D equals zero for a strictly bifurcating population history; however, we calculated D from population allele frequencies of independent polymorphic loci (Durand, Patterson, Reich, and Slatkin, 2011; Patterson et al., 2012). Thus, we only ran conducted this analysis on tests that included three or more individuals per population. The loci that contribute to the polymorphic and fixed ABBA/BABA methods are mutually exclusive from one another. Further, we generated two datasets for each test in order to investigate the impact of the number of individuals per population vs. the total number of SNPs on the test outcome. In "Set A", we included all individuals. In "Set C", we included only high coverage individuals (see Table I). Because we required all loci to be present for every individual in the set ($r = 1$), "Set A" datasets had more accurate population allele frequency estimates due to the higher number of individuals

included, but fewer total loci than in "Set C". We calculated D and an associated z-score for the null hypothesis that D=0 through 1000 bootstrap replicates with code modified from the evobiR package (Blackmon, 2016) in R v3.2.4 (R Core Team, 2016).

Finally, we also used the four population test to assess introgression patterns (Reich et al., 2009). The four population test is based on the concept that, given an unrooted tree (Taxon1, Taxon2),(Taxon3, Taxon4), the allele frequency differences between sisters Taxon1 and Taxon2 should be uncorrelated with the allele frequency differences between sisters Taxon3 and Taxon 4 (Reich et al., 2009; Reich et al., 2010). Summed across the genome, a violation of this prediction indicates the genetic data does not support a strictly tree-like relationship among the four taxa, and introgression has occurred. Again, for a test, we required all loci to be present for each individual ($r = 1$), and generated datasets "Set A" (all individuals) and "Set C" (high coverage individuals). We calculated the test statistic f_4 for the expected tree and a z-score indicating if the genetic data supports the phylogenetic hypothesis in Treemix v1.12 (Pickrell and Pritchard, 2012). To account for linkage disequilibrium, we assigned block size to 20 in Treemix. However, we considered linkage disequilibrium unlikely to impact our results since Stacks assemblies are completely unordered with respect to genomic region; any association among loci order and physical location within the genome would be due to chance and unlikely. When we varied block size from 1 to 200 in 5 tests for comparison, it made no noticeable difference in results.

RESULTS:

mtDNA haplotype network:

We sequenced the genes ND3, ND2, and *cytb* for 95 *Spinus* samples and trimmed each to complete sequences of 416 bp, 777 bp and 753 bp, respectively. The concatenated mtDNA haplotype network revealed five central haplotype groups corresponding to *Spinus crassirostris, S. uropygialis, S. atratus, S. cucullatus* and *S. magellanicus* from Argentina (Fig. 1A); individuals from each taxon were only found in their respective haplotype groups. *S. magellanicus* from Peru and Bolivia were found in the haplotype groups of *S. crassirostris* and *S. atratus* (Fig. 1A). *S. magellanicus* Peru included some variation within haplotype groups, but individuals also shared identical mtDNA haplotypes with *S. crassirostris* and *S. atratus. S. magellanicus* Bolivia differed in three nucleotides from the central *S. atratus* haplotype; however, this genetic distance was on par with the intraspecific variation in *S. atratus* (Fig. 1A). The haplotype networks of each individual gene reflected the concatenated haplotype structure, as expected for genes on the non-recombining mitochondrial genome; thus, we used ND3 to characterize 100 *S. magellanicus* samples into *S. crassirostris, S. atratus* or *S. magellanicus* Argentina haplotype groups. In *S. magellanicus,* the *S. crassirostris* haplotype group predominated on the western slope of the Andes in Ecuador and Peru at all elevations (Fig. 1B). The *S. atratus* haplotype group increased in frequency southwards, at high elevations near Cusco, Peru, and fixed in Cochabamba, Bolivia at 2470 m. In the low foothills and plains southeast of the central Andes, we found only *S. magellanicus* Argentina haplotypes (Fig. 1B).

Genotyping-by-sequencing metrics:

Using the genotyping-by-sequencing approach, we sequenced 82.9 million raw reads for the 40 *Spinus* samples with a median of 1.99 million raw reads per individual (range: 1.24-2.68 million raw reads, Table 1). With minimum read depth (*m*) equal to 4, we recovered a median of ~163,800 total, high quality loci per individual in Stacks assembly B3 (Table 1), post rxstacks processing (Catchen et al., 2013). With r=0.5 and p set to 4, 5, 6 and 7 (see methods), we generated datasets that included between 59,148 and 27,460 variable loci across all individuals with assembly B3. These datasets contained 12.1% to 23% missing data; as the number of populations required (*p*) increased, the number of variable loci recovered decreased and the proportion of missing data decreased (Table S1). Results from other assemblies (B1, B2, B4 and B5) were similar (Table S1).

We found that the average number of unique alleles per locus was lowest under the most restrictive Stacks assembly parameters (97% similarity, assembly B1) and increased from B1 to B3 (95% similarity). Differences among the more liberal assemblies B3, B4 (94%) and B5 (93%) were negligible (Fig. S1). These results suggest that assemblies B1 and B2 are over-splitting loci and artificially reducing diversity within a locus. Assemblies B3, B4 and B5 are effectively equivalent; in them, the diversity within a locus is appropriately described by the mismatch parameters. Maximum likelihood phylogenetic tree topology of all individuals was invariant with respect to assembly; tree topology was more sensitive to the impact of structured missing data (Table S1, see below). Based on these results, we selected Stacks 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. Stacks assembly B3 minimized both over-merging and over-splitting in loci assembled *de novo*.

GBS Phylogenetics:

Our GBS maximum likelihood phylogenetic analyses across assemblies B1-B5 and missing data regimes resulted in a novel phylogenetic hypothesis for *Spinus* based on nuclear DNA. We summarize below the evolutionary relationships resolved in the "Best" phylogeny, reporting bootstrap support (bs) for the concatenated complete loci maximum likelihood. The "Best" 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. The "Best" dataset optimized the phylogenetic support for internal nodes while minimizing missing data (Table S1).

In the "Best" phylogeny (Fig. 2), 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 *S. crassirostris* diverged first (bs 96). Within the remaining taxa, *S. magellanicus* from Argentina and Cochabamba, Bolivia formed a well supported the monophyletic clade (bs 100) *S. magellanicus* Bol. Arg., which was sister to *S. uropygialis* (bs 96). *S. magellanicus* Peru formed a monophyletic clade (bs 63) which was sister to *S. atratus* (bs 81). *S. magellanicus* from Southern Peru formed a monophyletic clade with strong support (bs 100).

The results of the "Best" phylogeny were well supported across assemblies B1-B5 and missing data regimes (Table S1) with the exception of the monophyly of *S. magellanicus* Peru, and its sister relationship with *S. atratus.* The "Best" topology for *S.*

magellanicus Peru and *S. atratus* was supported in datasets with $p = 6$. With p = 6, a loci that was completely absent in one population could be included in the dataset. An alternative topology where *S. atratus* was sister to the clade containing *S. uropygialis* and *S. magellanicus* Bol. Arg. was recovered in datasets with $p = 4$, 5 and 7; in a subset of trees, *S. magellanicus* from Southern Peru was sister to the *S. uro./S. mag.* Bol. Arg./*S. atr.* clade (Fig. 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 was disproportionately distributed among taxa in datasets with p = 4 or p = 5. Permitting multiple populations to lack data for a single locus may have introduced bias into the phylogenetic analysis, though the effect of structured missing on maximum likelihood estimation needs to be systematically investigated (Leaché et al., 2015). Requiring all populations to be present for a locus (p = 7) may have had consequences as well since loci that were not recovered in our single outgroup individual were eliminated. Datasets with p = 7 possessed two thirds of the loci recovered in p = 6 analyses. The one third of loci that were eliminated from p = 6 to p = 7 were, by definition, loci with variation within the Andean *Spinus*; fewer variable sites within the in-group could impact topology and support for internal nodes. We suggest p = 6 is a reasonable balance between maximizing the number of loci and minimizing the unknown impact of structured missing data.

The conservative interpretation of these results is that *S. magellanicus* Central Peru, S. magellanicus Southern Peru, and *S. atratus* form a soft polytomy with the

ancestor of *S. magellanicus* Bol. Arg. and *S. uropygialis*. *S. magellanicus* from Peru resembled a large population which has not fully differentiated from its congeners due to its size and insufficient time though geographic differentiation has developed in Southern Peru. This latter interpretation is consistent with morphological, morphometric and behavioral data (Beckman and Witt, 2015 for discussion) which unite *S. magellanicus* from Peru. Considering potential phylogenetic biases (see above), we report the support for monophyly in *S. magellanicus* Peru , recovered in the "Best" phylogeny, was modest. The "Best" sister relationship between *S. magellanicus* Peru and *S. atratus* was recovered in all other population structure analyses, suggesting that the methodology, lack of informative sites or missing data affected the power of RAxML to recover this relationship.

We found that the rooting of the *Spinus* tree in RAxML was highly sensitive to the proportion of invariant sites (Fig. 3). In an SNP dataset without the Felsenstein invariant site correction, we recovered an unrooted tree in which *S. cucullatus* was placed sister to ancestor of *S. uropygialis* and *S. magellanicus* Bol. Arg. This root placement was in strong disagreement with the complete, concatenated locus analysis which placed *S. cucullatus* sister to *S. crassirostris* in the unrooted tree (bs 97) and Felsenstein invariant site corrected analyses.

Our SNAPP analyses supported most of the evolutionary relationships recovered in the RAxML analyses. The SNAPP results differed in the placement of the root compared to the RAxML phylogenies (Fig. 3). In all SNAPP trees, the first taxon to diverge after *S. cucullatus* was *S. uropygialis* (pp 100); in RAxML "Best" tree, the first to diverge after *S. cucullatus* was *S. crassirostris* with 97 bootstrap support (bs). Despite

this rooting issue, the SNAPP tree with the strongest posterior probability (pp), referred hereafter as SNAPP tree S1, was equivalent to the "Best" unrooted topology (*S. cucullatus* excluded): *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 clades. The SNAPP results also revealed strong, conflicting signal within the GBS data. In addition to the "Best" topology recovered in 58.7% of trees sampled, we also found two other topologies among central Andean *Spinus* (Fig. S3). The second most frequently sampled topology (SNAPP tree S2), 29.3% of trees, placed *S. atratus* and *S. magellanicus* Peru sister, with *S. magellanicus* Bol. Arg. sister to them; *S. uropygialis* and *S. crassirostris* were sisters. The third 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 the conflicting signal within the data prevented convergence (ESS > 100) for important population demographic parameters like theta upon which the coalescent model depends; ESS values did not increase with increasing generations.

The topology recovered by SVDquartets (Fig. S4) was similar to the SNAPP tree S2. The support for the nodes in conflict with the "Best" and S1 topology was low in "lineage" trees (bs: 63 for *S. cras./S. uro* clade, 77 for *S. mag.* Bol. Arg. sister to *S. mag.* Peru/*S. atr.*) while "species" trees had high support for them (bs: 91, 98). In this analysis, *S. cucullatus* was placed sister to the ancestor of *S. crassirostris* and *S. uropygialis* (Fig. 3).

GBS Principal Component Analysis:

As in our phylogenetic analysis, individuals in our PCA from one *Spinus* taxon were more similar to each other than other *Spinus* taxa (Fig. 4). Most of the genetic variation of each taxon was best described on orthologous axes. In the full analysis, PC1 and PC2 represented 10.12% and 7.05% of the genetic variation in the data respectively; these axes separated *S. uropygialis* and *S. crassirostris* from each other and all other taxa (Fig. 4A). The axes PC1 and PC2, representing 8.56% and 4.75%, respectively, of the variation in the nested analysis (Fig. 4B, no *S. crassirostris* or *S. uropygialis*) effectively separated *S. atratus*, *S. magellanicus* Argentina, and *S. magellanicus* Central Peru along orthologous axes as well. However, *S. magellanicus* Bolivia was placed approximately equidistant from *S. magellanicus* Argentina and *S. atratus* suggesting *S. magellanicus* Bolivia individuals may contain components of both taxa, though this may also be an artifact of isolation by distance (Novembre and Stephens, 2008). *S. magellanicus* Southern Peru were placed in the lower right quadrant of the nested PCA; these individuals share more similar values with *S. atratus* on the PC2 axis than with *S. magellanicus* Central Peru.

GBS Clustering Analysis:

Using CV error to evaluate the support for different numbers of populations to be reconstructed in Admixture (k), we found the best support, the lowest CV error, for analyses with k = 1, 2 or 4 (Fig. 5A). The clustering analysis with k = 2 (Fig. 5B), reconstructed two populations, one containing *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 (Fig. 5), however, closely mirrored *a priori Spinus* taxonomic units. The four populations corresponded to *S. magellanicus* Bol. Arg., *S. crassirostris*, *S. uropygialis* and, lastly, *S. magellanicus* from Peru lumped with *S. atratus.* Thus, the clustering analysis was unable to discriminate between Peruvian *S. magellanicus* and *S. atratus.* In the reconstruction with k=4, there were also several individuals that are admixed; this included two *a priori S. atratus* who shared ancestry with the *S. magellanicus* Bol. Arg. cluster. Another six individuals *a priori* designated as *S. magellanicus* from Peru shared ancestry with *S. crassirostris.* Other reconstructions (k = 3, 5, 6 and more) fit the data poorly (Fig. 5A) and were difficult to interpret in light of expected taxonomic designations.

Introgression tests:

We derived 31 separate tests based on well supported four-taxon trees from our "Best" RAxML tree. These tests were robust to the alternate topologies recovered in RAxML with the following exception; we treated *S. magellanicus* Peru as a single population in 8 tests (see Phylogenetics for justification). Sixteen of the 31 tests applied to the SNAPP tree S1 (Table 2); six were valid for the SNAPP tree S2 and the SVDquartet tree. With the 31 tests, we were able to conduct multiple tests for introgression between specific pairs of taxa (see Test 1 and 2 in Table 2). The number of SNPs included in each test depended on the variable loci that overlapped with complete coverage among the four taxa, and the number of ABBA or BABA-like patterns in fixed and polymorphic loci. We report the result of all RAxML tests and specify the validity of the tests with respect to alternate topologies.

First, we found strong evidence that *S. crassirostris* had introgressed with Peruvian *S. magellanicus* (Table 2); this was true when *S. magellanicus* Peru was considered (Tests 6-8) as well as when only *S. magellanicus* Central Peru was included (Tests 9-11). The evidence for introgression was valid across all four Spinus tree topologies (Fig. 6) though only a subset of the tests applied for alternative topologies. Z-scores based on bootstrap replicates of the ABBA/BABA fixed tests for *S. crassirostris* and *S. magellanicus* from Peru ranged from 2.24 to 7.15. The variation among z-scores may be explained by the reduced number of sites available in Tests 8 and 9, or by a history of introgression between *S. crassirostris* and *S. atratus*. Tests 1 and 2 give mixed support for introgression between *S. crassirostris* and *S. atratus*; however, the significant result in Test 1 was computed with over twice as many variable loci as Test 2, suggesting the Test 2 result may be due to lack of power. Introgression between *S. crassirostris* and *S. atratus* may have inadvertently reduced the strength of the introgression pattern in Tests 8 and 9.

Second, we found modest evidence, valid for all tree topologies, for introgression between *S. atratus* and *S. magellanicus* from Southern Peru. Introgression was well supported in 3 of 4 population approaches (Test 31, Table 2), but not the fixed ABBA/BABA test; this difference may be a consequence of the few fixed sites available for the latter. Note, that this test was contingent upon the monophyly of *S. magellanicus* Peru. The reciprocal monophyly of *S. magellanicus* Southern Peru and *S. magellanicus* Central Peru was unambiguous in Set C due to the removal of two low coverage individuals.

Third, evidence for introgression between *S. magellanicus* Bolivia and *S. atratus* was mixed. Support for introgression was significant (z-score=2.02) in the ABBA/BABA fixed Test 4; however, there was no support in Test 3, the only valid test for SNAPP tree S1. With respect to the ML tree, Test 3 was identical to Test 4 excepting the outgroup (*S. cucullatus* in Test 4, *S. crassirostris* in Test 3). The disparity among tests could again be the result of introgression between *S. crassirostris* and *S. atratus* (see above). Introgression among *S. crassirostris* and *S. atratus* would serve to artificially increase the support for the expected tree in Test 3, despite introgression between *S. magellanicus* from Bolivia and *S. atratus.* When the entire *S. magellanicus* Bol. Arg. clade or only *S. magellanicus* from Argentina was included in the analyses (Tests 5, 12, 13), evidence for introgression with *S. atratus* was negligible suggesting that introgression was geographically structured between these two taxa. There were no valid tests for SNAPP tree S2 or the SVDquartet tree for this taxon pair.

There was no evidence for introgression among *S. crassirostris* and *S. magellanicus* Bolivia or Bol. Arg. (Tests 12-15), *S. atratus* and *S. uropygialis* (Tests 12 and 13), or *S. crassirostris* and *S. uropygialis* (Tests 14 and 15). Results for the former two taxon pairs were also valid for the SNAPP tree S1; no tests apply to SNAPP tree S2 or the SVDquartet tree. The independence of these lineages lends additional support to the introgression patterns reported above since complex relationships among these specific taxa did not need to be accounted for in the interpretation of the results.

Lastly, we found evidence for introgression between *S. magellanicus* Peru and *S. uropygialis* (Tests 16-30); these results are valid for the RAxML tree and SNAPP tree S1 (Table 2). The signal of introgression was almost uniformly positive (z-score > 2.0) in

population-based introgression tests (four population and ABBA/BABA population test), but mostly absent in the ABBA/BABA test based on fixed sites. Z-scores were highest in tests for introgression between *S. uropygialis* and *S. magellanicus* Southern Peru. Z-scores in tests that included all Peruvian *S. magellanicus* or *S. magellanicus* Central Peru were lower, including some that were insignificant. These findings suggest weak support for introgression between *S. uropygialis* and Peruvian *S. magellanicus*. The results of Tests 16-30 also reject introgression between *S. magellanicus* Bol. Arg. and *S. magellanicus* Peru.

DISCUSSION:

We documented introgression in a rapid radiations, the South American siskins, despite lingering uncertainty about the true species tree. We assessed population structure in the mitochondrial and nuclear genomes through clustering algorithms, principal component analyses, network building and phylogenetic approaches. Resolving the internal nodes of *Spinus* was challenging despite tens of thousands of loci; these difficulties highlighted sources of conflict in building a robust phylogenetic hypothesis in rapid radiations with complex histories. We formally tested for introgression among *Spinus* populations using phylogenetic-based tests for fixed and polymorphic loci. We conclude that the confluence of evidence supports multiple introgression events among *Spinus* populations. Interspecific introgression of at least one locus with adaptive potential across an elevational gradient, the mitochondrial genome, has occurred within this radiation.

Population structure:

With the GBS approach, we were able to successfully resolve individual *Spinus* populations corresponding to known species and geographic populations of the widespread *S. magellanicus*. However, the differences among phylogenetic methods demonstrated that, despite the upwards of 60,000 short variable loci, the phylogenetic hypothesis of *Spinus* was difficult to reconstruct. The challenge of reconstructing the *Spinus* phylogeny was likely due to the presence of both incomplete lineage sorting and introgression within the radiation. Incomplete lineage sorting (ILS), the retention of ancestral polymorphism across divergence events, is likely in rapid, recent radiations due to the short internal branches and recent divergence times (Edwards, 2009). ILS may result in gene heterogeneity and, under certain conditions, can mislead concatenated maximum likelihood analyses (Kubatko and Degnan, 2007). Multi-species coalescent models, like SNAPP and SVDquartets, incorporate ILS; however, introgression is not included in any of these models despite its contribution to gene heterogeneity. Introgression can mislead maximum likelihood analyses (Fontaine et al., 2015). In multi-species coalescent analyses, it can reduce the frequency with which the true species tree is recovered, and overestimate demographic parameters which impact branch lengths (Leaché et al., 2014). The divergence among introgressing species as well as the amount of introgression (Fontaine et al., 2015; Leaché et al., 2014) determine the magnitude of these consequences. Network analyses for phylogenomic data are a potential solution for resolving phylogenies when introgression has occurred. Current research has focused on comparing large genomic regions rather than SNP data (Gerard, Gibbs, and Kubatko,

2011; Liu et al., 2014; Wen et al., 2016; Wen, Yu, and Nakhleh, 2016; Yu, Degnan, and Nakhleh, 2012; Yu et al., 2011).

We evaluated the two major phylogenetic disagreements among methods in light of their methodology, and in the context of other population structure results. First, in the topology recovered in the SNAPP tree S1 and the "Best" RAxML tree, *S. crassirostris* was sister to the *S. magellanicus* Peru/*S. atratus* clade in the unrooted tree. In the alternate topology, recovered in the SNAPP tree S2 and SVDquartets, *S. magellanicus* Bol. Arg. was sister to the *S. magellanicus* Peru/*S. atratus* clade. The presence of the two topologies in all SNAPP runs, combined constituting ~90% of all trees sampled, implied that there were two distinct phylogenetic signals within the data. ILS was an insufficient explanation for the observed heterogeneity; otherwise, SNAPP would have converged on a single tree with appropriate demographic parameters. Instead, the program did a poor job of estimating demographic parameters.

We suggest the best interpretation of these alternate tree topologies is that the unrooted tree in which *S. crassirostris* was sister to the *S. magellanicus* Peru/*S. atratus* clade represents the true species history, and the alternate tree reflects a history of introgression between *S. magellanicus* Bol. Arg. and *S. atratus*. This interpretation was supported by formal introgression Test 4 (applicable only to the RAxML tree), and the Admixture results (k=4), in which *S. atratus* individuals were recovered as admixed with *S. magellanicus* Bol. Arg. The PCA, in which *S. magellanicus* Bolivia birds were equidistant from *S. magellanicus* Argentina and *S. atratus* also supports this interpretation, though introgression is not the sole acceptable explanation for this pattern. Lastly, the fact that the maximum likelihood phylogeny and the most frequently sampled

SNAPP tree have the same unrooted topology, despite methodological differences (Edwards, 2009; Edwards et al., 2016; Gatesy and Springer, 2014; Springer and Gatesy, 2016) suggests that this topology is the correct one. The reason for the alternate topology within SVDquartets, an approach which proposes valid splits among quartets as an extension of coalescent theory (Chifman and Kubatko, 2015), is unclear; the impact of introgression on this method has not been rigorously addressed.

The second phylogenetic conflict among our methods was the placement of the root, *S. cucullatus* (see Fig. 3). We hypothesize that, due to the close relatedness and history of introgression among Andean *Spinus* species, the placement of *S. cucullatus* was unusually sensitive to long-branch attraction and evolutionary-model misspecification. The consequences of introgression on branch length estimation in the different phylogenetic methods may have led to the support for alternate topologies for the relatively distant outgroup, *S. cucullatus.*

This hypothesis was supported by a RAxML reconstruction of SNP-only data with no Felsenstein invariant site correction. In this analysis, branch lengths were changed by an order of magnitude and the position of *S. cucullatus* was altered (Fig. 3); Ford et al. (2015) observed a similar issue in a radiation of African cichlids in which multiple bouts of introgression had occurred. Considering the failure of SNAPP to estimate demographic parameters with high confidence, and the fourfold more variable loci included in the RAxML analysis, we suspect that the latter was more likely to have provided the correct root (Jarvis et al., 2014); however, we consider the true root to remain unknown.

Despite these uncertainties, the overall concordance of the "Best" maximum likelihood tree, the SNAPP tree S1, Admixture and the PCA, gave us a manageable set of well-supported trees with which to construct the formal introgression tests.

Evidence of introgression:

Nuclear introgression:

Evidence for introgression of nuclear genes among multiple *Spinus* species was supported by population structure results and multiple, formal introgression tests. Corroboration of disparate methods is important because the assumptions and power vary across these tests (Alexander, Novembre, and Lange, 2009; Martin, Davey, and Jiggins, 2015; Patterson et al., 2012; Rheindt et al., 2014). Introgression has occurred between: (1) *S. magellanicus* Peru and *S. crassirostris*; (2) *S. magellanicus* Southern Peru and *S. atratus*; (3) *S. magellanicus* Bolivia and *S. atratus*; and (4) *S. magellanicus* Peru and *S. uropygialis*. Each of these species pairs either had significant results for multiple of our 31 formal introgression tests, significant results among different formal testing methods (*e.g.* ABBA/BABA fixed vs. ABBA/BABA population tests), or at least one significant formal test and evidence of introgression in Admixture. We also report that evidence for introgression in pairs 1, 2 and 4 was supported in formal tests that applied to SNAPP tree S1. Formal tests that applied to SNAPP tree S2 and the SVDquartets tree supported introgression in pairs 1 and 2. We consider this strong evidence that introgression has occurred multiple times among Andean *Spinus*.

The results for mtDNA and nDNA introgression are concordant in all but one species pair (*S. magellanicus* Peru and *S. uropygialis*). This bolsters the support for the

mitochondrial evidence for introgression and suggests that in this case, the mitochondrial genome offered insight into the evolutionary forces impacting Andean *Spinus*.

Mitochondrial introgression:

We predicted that (1) geographic structure within the widespread *S. magellanicus*, and (2) low genetic distance between interspecific haplotypes within the same haplotype group would be consistent with our alternative hypothesis, H_A1, that mitochondrial genome introgression had occurred. We found both these predictions were supported among specific species pairs. The north/south cline in mtDNA in *S. magellanicus* was concordant with the introduction of the *S. atratus* mtDNA in Southern Peru and Bolivia, and subsequent spread northwards and an increase in allele frequency in the South. However, ILS with limited gene flow among populations is also a valid explanation for this pattern (see methods). We also found low genetic distance between interspecific haplotypes in the same haplotype group in the following pairs: *S. magellanicus* Peru and *S. crassirostris*, *S. magellanicus* Peru and *S. atratus*, and *S. magellanicus* Bolivia and *S. atratus*. Individuals of the former two species pairs shared identical mtDNA haplotypes. The evidence for introgression based on genetic distance, however, is strongest between *S. magellanicus* Bolivia and *S. atratus* since the most probable nDNA phylogeny (see above) shows greater divergence between *S. magellanicus* Bolivia and *S. atratus* than within the other putatively introgressing species pairs. The mtDNA interspecific genetic distance between *S. magellanicus* Bolivia and *S. atratus* is comparable to the intraspecific genetic distance between *S. atratus* individuals, suggesting that the Bolivian birds have not acquired unique variation despite a relatively longer time since the most recent

common ancestor of both populations. Taken together, our results are consistent with introgression between *S. magellanicus* Peru and *S. crassirostris*, *S. magellanicus* Peru and *S. atratus*, and *S. magellanicus* Bolivia and *S. atratus*, however the null hypothesis of no introgression cannot be completely rejected (see methods). The best evidence for mtDNA introgression is between *S. magellanicus* Bolivia and *S. atratus.*

The mitochondrial genetic variation with the widespread species, *S. magellanicus*, likely reflects a history of introgression with two high elevation restricted species, *S. atratus* and *S. crassirostris*. Interestingly, we found that all but one *S. magellanicus* possessing *S. atratus* mtDNA was captured above 2500 m. Further, despite strong support for the *S. mag.* Bol. Arg. clade, which included samples from ≤340 m in Argentina, *S. magellanicus* from 2550 m in Bolivia were fixed for the *S. atratus* mtDNA. This pattern suggests that elevation may be an important factor in the retention of the introgressed *S. atratus* allele; the mitochondrial genome is a known target for natural selection across elevation in other systems (Boratyński et al., 2011; Gering, Opazo, and Storz, 2009; Mishmar et al., 2003), including birds (Scott et al., 2009). Discordant histories between mitochondrial and nuclear genomes are abundant, and often attributed to introgression. The way in which introgressed mitochondrial genomes rise in frequency, sometimes to fixation (genome capture), may be the result of neutral processes and the consequences of an expected small effective population size for the mitochondrial genome (Ballard and Whitlock, 2004). However, natural selection with respect to elevation may be an important factor in some systems; our preliminary examination of the distribution of *S. atratus* mtDNA in *S. magellanicus* suggests this may be the case in *Spinus*. Future work should explicitly address this possibility in *Spinus* and other systems

by formally testing the association of introgressed mtDNA haplotypes with elevation, identifying non-synonymous mutations and assessing whether introgressed mtDNA have fitness consequences.

CONCLUSIONS:

Our discovery of multiple introgression events within the Andean radiation of *Spinus* siskins is consistent with an emerging paradigm in biodiversity genomics, that introgression commonly occurs during early stages of diversification. A crucial consequence of this phenomenon, as illustrated in our study, is that estimating the phylogenetic history of these radiations can be exceptionally difficult or impossible. Comparative analyses that depend on the phylogeny can be difficult as a result. The phylogenetic challenge arising from introgression persists with genomic-scale data sets, causing a major challenge for efforts to resolve the tree of life. Our results resemble the findings of Lamichhaney et al., (2015) which found extensive introgression among multiple species pairs in the Darwin's finches (Thraupidae). Together, the results of these studies suggest that introgression may be an important evolutionary force in rapid avian radiations. Introgression may also be a central source of phylogenetic uncertainty in bird clades, young and old. Rapid radiations have occurred frequently in the avian tree of life, including at the base of Neoaves (Jarvis et al., 2014; McCormack, 2013; Prum et al., 2015) the base of Passeriformes (Suh et al., 2015), the tanagers (Thraupidae) (Burns et al., 2014), the Hawai'ian honeycreepers (Lerner et al., 2011), and the *Zosterops* white-eyes (Moyle et al., 2009). All of these rapid radiations have been difficult to resolve. Our

analyses suggest that introgression can cause conflicting phylogenetic signal in these cases, possibly causing the species tree to become difficult or impossible to discover.

Our results highlight the importance of developing phylogenetic methods capable of explicitly modeling gene heterogeneity from ILS as well as introgression. This is an area of active research (Wen et al., 2016; Yu et al., 2011), but we have provided an example for how to meet this challenge: by employing multiple phylogenetic methods, conducting additional population structure analyses, and thoroughly investigating the consequences of missing data and assembly parameters on our results. Through this effort, we identified four possible phylogenetic hypotheses for *Spinus,* that differ mostly but not entirely on the appropriate placement of the root in Andean *Spinus*. Recognizing areas of phylogenetic uncertainty by using multiple methods of phylogenetic inference proved essential to conducting valid downstream tests of introgression.

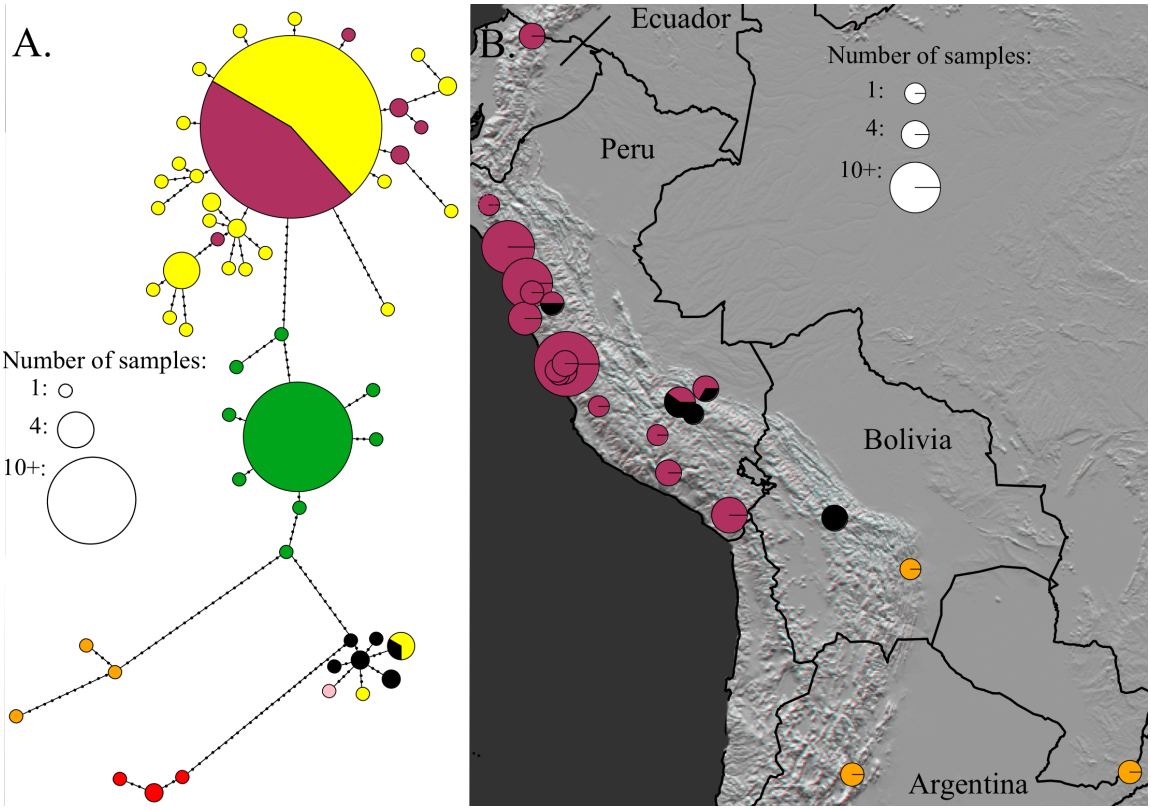

Fig. 1. Mitochodrial genetic variation within *Spinus*. A. MtDNA haplotype network constructed from concatenated ND3 (416 bp), ND2 (777 bp) and *cytb* (753 bp) sequences from *S. crassirostris* (maroon, 16 samples), *S. uropygialis* (green, 20), *S. atratus* (black, 8), *S. cucullatus* (red, 4) and *S. magellanicus* from Peru (yellow, 43), Cochabamba, Bolivia (pink, 1), eastern Bolivia/northern Argentina (orange, 3). **B.** *S. magellanicus* ND3 haplotype occurrence based on 100 samples overlaid on topographic map of western South America; haplotypes were classified as *S. crassirostris* type (maroon), *S. atratus* type (black) or northern Argentina type (orange). For both panels, size of circle is proportional to sampling; see respective scales.

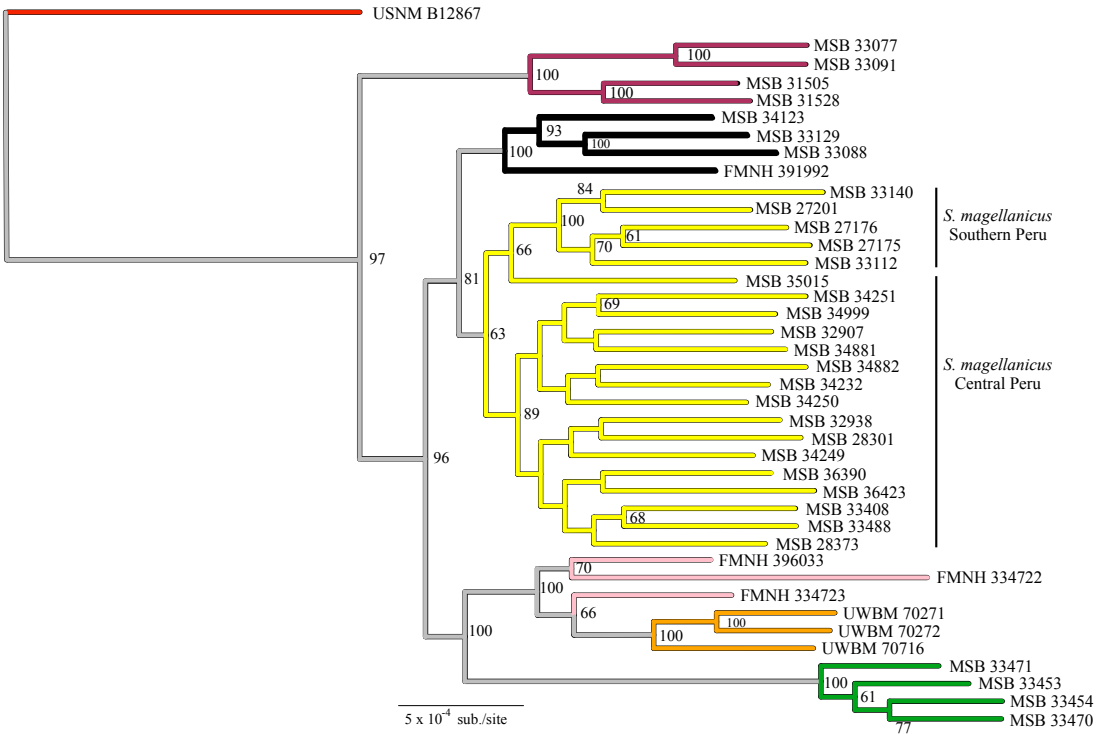

Fig. 2. Maximum likelihood phylogeny of *Spinus* nuclear DNA. Tree generated in RAxML with 45,246 full, concatenated loci from GBS assembly B3. Branches colored by taxonomic units: *S. cucullatus*: red, *S. crassirostris*: maroon, *S. atratus*: black, *S. magellanicus* Peru: yellow, *S. magellanicus* Bolivia: pink, *S. magellanicus* Argentina: orange, *S. uropygialis*: green. Node support represents 500 bootstrap replicates.

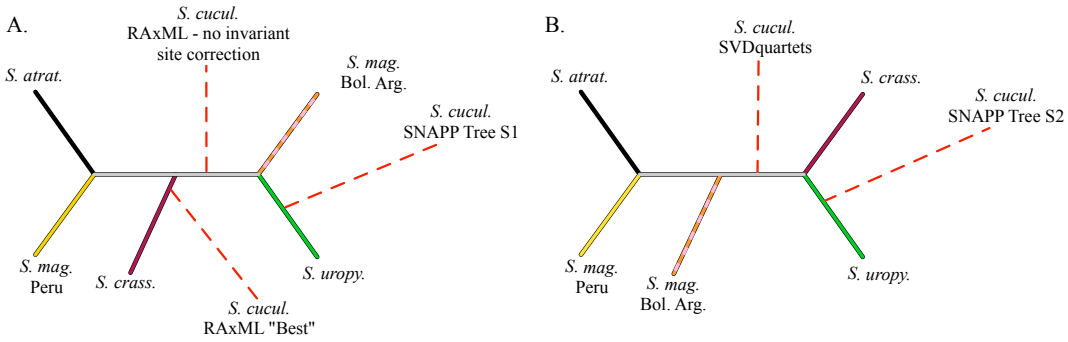

Fig. 3. Alternate root placement and phylogeny topologies of *Spinus*. Summary of unrooted Andean *Spinus* tree topologies. The dashed red line indicates alternate placements of *S. cucullatus*, the outgroup for the Andean *Spinus*, based on phylogenetic analysis. **A.** Topology of the RAxML "Best" analysis and the most frequently recovered SNAPP tree (S1). A RAxML SNP-only analyses without the Felsenstein invariant site correction for reference in discussion. **B.** Topology of the second most frequently recovered SNAPP tree (S2), and the SVDquartets lineage and species trees (SVD). In both, branches are colored by taxon: *S. crassirostris*: maroon, *S. atratus*: black, *S. magellanicus* Peru: yellow, *S. magellanicus* Bolivia/Argentina: orange, *S. uropygialis*: green.

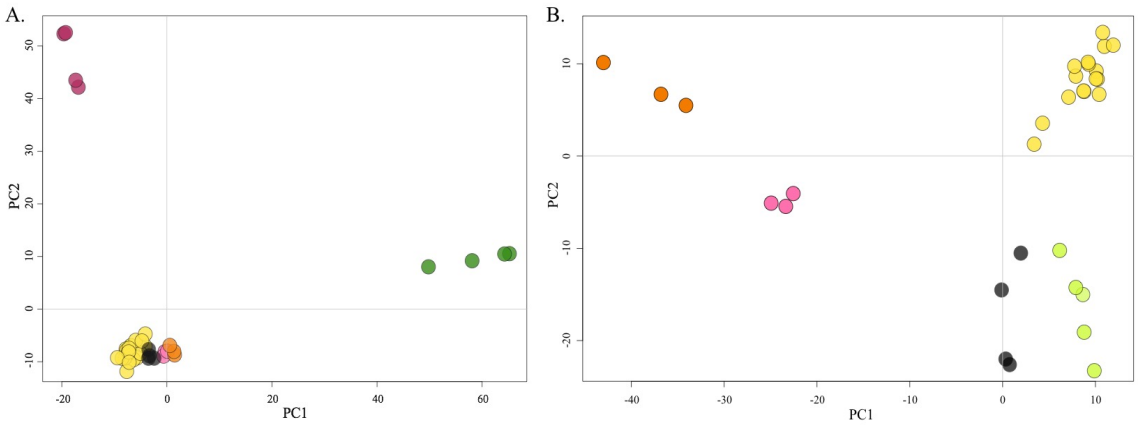

Fig. 4. Principal component analysis of *Spinus* nuclear DNA. All analyses produced with 45,246 unlinked, variant sites from genotyping-by-sequencing, assembly B3. Individuals colored by taxonomic units: *S. cucullatus*: red, *S. crassirostris*: maroon, *S. atratus*: black, *S. magellanicus* Peru: yellow, *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 are colored as Southern Peru (light green) or Central Peru (yellow).

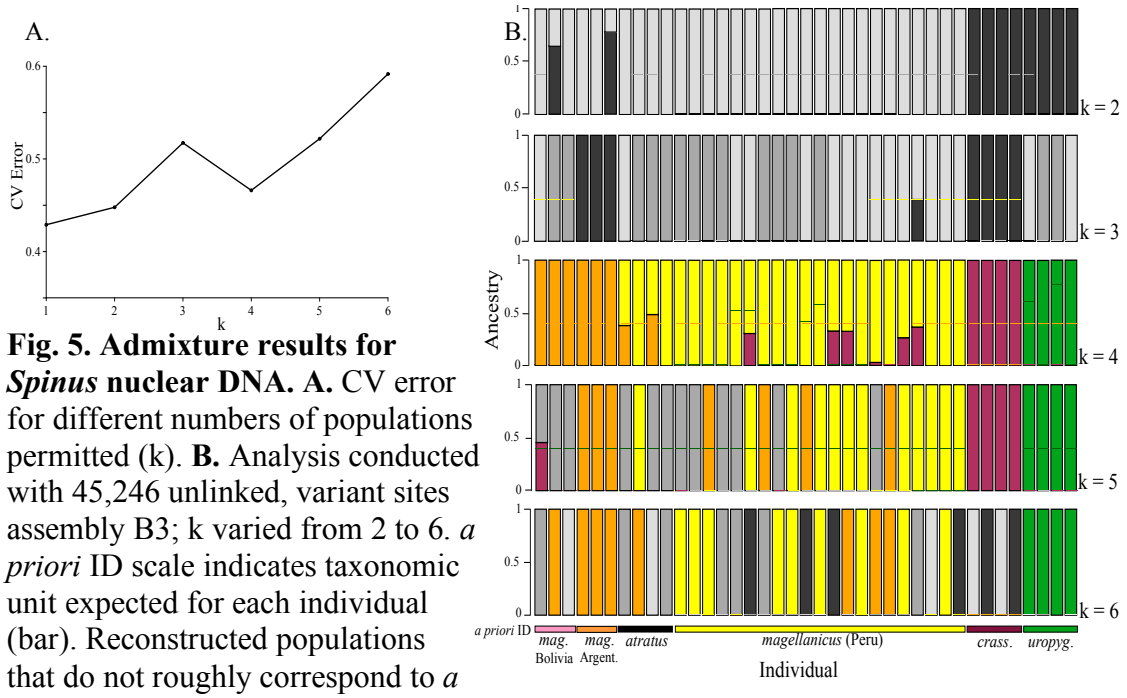

Fig. 5. Admixture results for *Spinus* nuclear DNA. A. CV error for different numbers of populations permitted (k). **B.** Analysis conducted with 45,246 unlinked, variant sites assembly B3; k varied from 2 to 6. *a priori* ID scale indicates taxonomic unit expected for each individual (bar). Reconstructed populations that do not roughly correspond to *a priori* IDs were given gray-scale

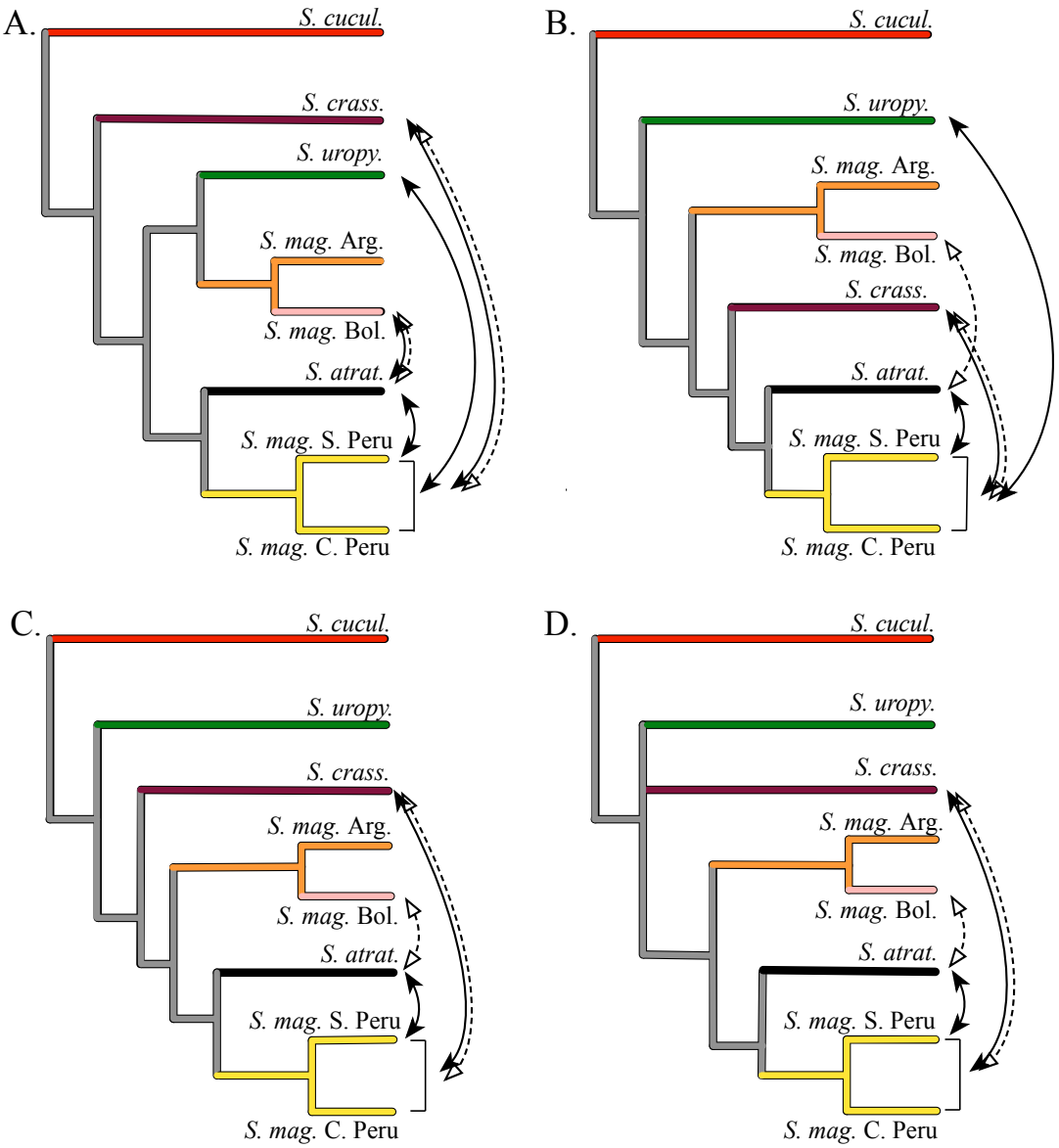

Fig. 6. Summary of introgression results for alternative *Spinus* tree toplogies. Solid arrows indicate significant results in formal introgression tests valid for each tree topology; dashed arrows indicate introgression in Admixture. **A.** "Best" RAxML cladogram. **B.** Most frequently recovered SNAPP cladogram (S1). **C.** Second most frequently recovered SNAPP cladogram (S2). **D.** SVDquartet lineage tree; nodes with bootstrap support <70 collapsed. Bracket at tips indicates analyses with *S. magellanicus* Peru.

Table 1. Genotyping-By-Sequencing Sampling. Elevation (Elev.) in meters. Raw reads indicates total reads with >30 mean Phred score. Loci indicates number of loci in assembly B3 with 4x depth, after rxstacks processing. SetC and SetE columns show whether the sample was included in subsets. Abbreviations for museums are MSB: Museum of Southwestern Biology; FMNH: Field Museum of Natural History; USNM: Smithsonian Institution; UWBM: Burke Museum.

Individual	Species	Elev.	Locality	Latitude	Longitude	Raw reads	Loci	SetC	SetE
MSB:Bird:34123	*Spinus atratus*	4300	PERU: Apurimac: Abancay: Chacoche	-14.063	-73.009	1 847 044	150 964	X	---
MSB:BIRD:33129	*S. atratus*	3835	PERU: Cusco: Ollantaytambo: Choquechaca	-13.188	-72.231	1 913 373	157 428	X	X
MSB:BIRD:33088	*S. atratus*	4400	PERU: Cusco: Ollantaytambo: Choquechaca	-13.188	-72.231	1 786 290	151 718	X	X
FMNH:BIRD:391992	*S. atratus*	4140	PERU: Lima: Huarochiri: Japani	-11.683	-76.533	1 469 527	120 267	---	---
MSB:BIRD:33077	*S. crassirostris*	4200	PERU: Cusco: Ollantaytambo: Choquechaca	-13.188	-72.231	2 384 107	211 144	X	X
MSB:BIRD:33091	*S. crassirostris*	4200	PERU: Cusco: Ollantaytambo: Choquechaca	-13.188	-72.231	2 302 031	201 734	X	X
MSB:BIRD:31528	*S. crassirostris*	3973	PERU: Lima: Huarochiri: Carampoma	-11.628	-76.434	2 516 502	214 128	X	X
MSB:BIRD:31505	*S. crassirostris*	3981	PERU: Lima: Huarochiri: Carampoma	-11.628	-76.434	2 678 762	216 611	X	X
USNM:BIRD:B12867	*S. cucullatus*	230	GUYANA: Wiwitau Mtn., E. Rupununi Savannah	2.867	-59.267	2 319 237	191 068	X	X
UWBM:BIRD:70271	*S. magellanicus*	209	ARGENTINA: Misiones: Posadas	-26.955	-55.088	1 687 108	129 920	X	X
UWBM:BIRD:70272	*S. magellanicus*	209	ARGENTINA: Misiones: Posadas	-26.955	-55.088	1 447 694	123 440	X	X
UWBM:BIRD:70716	*S. magellanicus*	340	ARGENTINA: Tucumán: San Miguel de Tucumán	-27.022	-65.645	1 714 660	135 705	X	X
FMNH:BIRD:334723	*S. magellanicus*	2470	BOLIVIA: Cochabamba: Cochabamba-Oruro Rd.	-17.504	-66.329	1 951 949	166 978	X	X
FMNH:BIRD:396033	*S. magellanicus*	2470	BOLIVIA: Cochabamba: Cochabamba-Oruro Rd.	-17.504	-66.329	1 745 484	143 165	X	---
FMNH:BIRD:334722	*S. magellanicus*	2470	BOLIVIA: Cochabamba: Cochabamba-Oruro Rd.	-17.504	-66.329	1 906 384	162 422	X	X
MSB:Bird:34881	*S. magellanicus*	2972	PERU: Ancash: Santa: Macate	-8.755	-78.048	1 796 953	150 017	---	---
MSB:Bird:34999	*S. magellanicus*	3714	PERU: Ancash: Caraz: Pueblo Libre: SW Caraz	-9.101	-77.866	2 207 409	178 262	X	---
MSB:Bird:35015	*S. magellanicus*	3714	PERU: Ancash: Caraz: Pueblo Libre: SW Caraz	-9.101	-77.866	1 997 706	160 239	---	---
MSB:Bird:34249	*S. magellanicus*	<100	PERU: Ancash: Huarmey	-10.068	-78.136	2 371 488	189 171	X	X
MSB:Bird:34232	*S. magellanicus*	<100	PERU: Ancash: Huarmey	-10.068	-78.136	2 279 538	179 967	X	---
MSB:Bird:34250	*S. magellanicus*	<100	PERU: Ancash: Huarmey	-10.068	-78.136	2 109 526	171 765	X	---
MSB:Bird:34251	*S. magellanicus*	<100	PERU: Ancash: Huarmey	-10.068	-78.136	1 619 697	132 965	---	---
MSB:Bird:34882	*S. magellanicus*	2972	PERU: Ancash: Santa: Macate	-8.755	-78.048	1 243 386	97 967	---	---
MSB:Bird:33140	*S. magellanicus*	3018	PERU: Cusco: Ollantaytambo: Choquechaca	-13.188	-72.231	1 637 057	135 922	---	---
MSB:Bird:33112	*S. magellanicus*	3835	PERU: Cusco: Ollantaytambo: Choquechaca	-13.188	-72.231	2 207 694	149 100	---	---
MSB:Bird:27201	*S. magellanicus*	3380	PERU: Cusco: Urubamba: 7.9km NW Urubamba	-13.249	-72.169	2 365 420	185 163	X	X
MSB:Bird:27175	*S. magellanicus*	3500	PERU: Cusco: Urubamba: 7.9km NW Urubamba	-13.249	-72.169	1 925 996	163 831	X	---
MSB:Bird:27176	*S. magellanicus*	3500	PERU: Cusco: Urubamba: 7.9km NW Urubamba	-13.249	-72.169	1 963 590	154 841	X	X
MSB:Bird:32938	*S. magellanicus*	935	PERU: Lima: 2.3km E Nieve Nieve	-12.030	-76.651	2 042 449	172 064	X	---
MSB:Bird:32907	*S. magellanicus*	935	PERU: Lima: 2.3km E Nieve Nieve	-12.030	-76.651	2 439 440	185 163	X	---
MSB:Bird:28301	*S. magellanicus*	3750	PERU: Lima: San Pedro de Casta: Carhuayumac	-11.762	-76.549	1 671 815	137 642	---	---
MSB:Bird:28373	*S. magellanicus*	3800	PERU: Lima: San Pedro de Casta: Chinchaycocha	-11.761	-76.568	2 159 135	179 865	X	X
MSB:Bird:33408	*S. magellanicus*	3905	PERU: Lima: Huarochiri: San Pedro de Casta	-11.768	-76.535	2 011 512	168 659	---	---
MSB:Bird:33488	*S. magellanicus*	4140	PERU: Lima: Huarochiri: San Pedro de Casta	-11.770	-76.532	2 202 776	180 957	X	X
MSB:Bird:36423	*S. magellanicus*	121	PERU: Lima: Lima, Comas	-11.944	-77.063	1 973 024	160 481	---	---
MSB:Bird:36390	*S. magellanicus*	214	PERU: Lima: Lima: Pachacamac: Lurin Valley	-12.153	-76.836	2 681 732	218 246	X	X
MSB:BIRD:33453	*S. uropygialis*	4131	PERU: Lima: Huarochiri: San Pedro de Casta	-11.769	-76.533	2 345 234	206 819	X	X
MSB:BIRD:33454	*S. uropygialis*	4131	PERU: Lima: Huarochiri: San Pedro de Casta	-11.769	-76.533	1 863 579	160 701	X	---
MSB:BIRD:33470	*S. uropygialis*	4131	PERU: Lima: Huarochiri: San Pedro de Casta	-11.769	-76.533	2 174 660	195 620	X	X
MSB:Bird:33471	*S. uropygialis*	4131	PERU: Lima: Huarochiri: San Pedro de Casta:	-11.769	-76.533	1 420 740	119 999	---	---

Table 2. ABBA/BABA and Four Population Test Results. ABBA/BABA and four population tests in *Spinus* based on fixed and polymorphic sites; tests constructed from four-taxon trees derived from RAxML phylogenies. All tests valid for RAxML trees where *S. magellanicus* Peru is monophyletic. Tests valid for alternate phylogenies are indicated by an X in Tree Topology columns including: the most frequently recovered tree in SNAPP (S1), the 2nd most recovered tree in SNAPP (S2) and the SVDquartets tree (SVD). Z-score and number of unlinked, biallelic sites from GBS data contributing to each test presented. Tests (T) are arranged so Taxon1 and Taxon2 are sister, and introgressing taxa are Taxon2 and Taxon3. Tests with significant evidence for introgression are in bold. All individuals are included in Set A; Set C includes a subset (see Table 1). Abbreviations of *S. magellanicus* populations refer to central Peru (CP), southern Peru (SP), Peru (P), Bolivia (B), Argentina (A) or Bolivia and Argentina (BA).

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

CHAPTER III:

Evolution in adult and embryonic hemoglobin genes in the Andean siskins (*Spinus*).

ABSTRACT:

Local adaptation may cause isolation by genetic mechanisms or spatial separation, and therefore provides a critical step towards speciation. Theory suggests taxa distributed across an elevational gradients may be prone to speciate via local adaptation since populations are spatially structured and subject to persistent selection due to temperature and oxygen availability. Previous research has identified hemoglobin (Hb), which binds and carries O2 from the lungs to respiring tissues in the body, as a selective target with respect to elevation. To investigate the influence of local adaptation across elevation gradients on diversification, we assessed non-synonymous variation in all hemoglobin loci in a recent radiation of finches, the South American siskins (genus: *Spinus*). We hypothesized that adaptation to high-altitude hypoxia in Andean *Spinus* has promoted diversification within and among species. We found that most amino acid variation across the *Spinus* clade in the hemoglobin loci transcribed in adults was structured by elevational range, and we identified three amino acid positions (α^{60pi}, β^{21A} and β^{135H}) that were significantly differentiated between high and low populations in Peruvian *S. magellanicus.* . Our result supports a role for elevation in promoting functional divergence among *Spinus* across the Andes, even across relatively short distances.

INTRODUCTION:

Local adaptation may cause isolation by genetic mechanisms or spatial separation, and therefore provides a critical step towards speciation (Sobel et al., 2010). Local adaptation is defined as the development of specialized phenotypes in local populations

that enhance fitness in one habitat but not another (Savolainen et al., 2013). In order for local adaptation to occur, the strength of selection must be strong relative to the level of gene flow (Kawecki & Ebert, 2004; Garant et al., 2007; Räsänen & Hendry, 2008; Cheviron & Brumfield, 2009; Savolainen et al., 2013) and gene flow must be restricted somewhat among populations (Kawecki & Ebert, 2004). Under certain circumstances, local adaptation may provide the initial disruption gene flow between populations (Sobel et al., 2010), providing realized habitat isolation, and may serve as a precursor to ecological speciation (Schluter, 2001; Jones et al., 2012; Wright et al., 2013). Theoretical work suggests the potential for speciation facilitated by local adaptation is greatest among systems where natural selection is moderately strong and persistent over time, loci are of large effect and populations are spatially structured (Doebeli & Dieckmann, 2003; Kawecki & Ebert, 2004). Given these predictions, species that are distributed across impactful environmental gradients, such as mountain slopes, should be especially prone to speciation, and clades that occur in heterogeneous environments should have higher rates of diversification as a result.

Populations distributed across elevational gradients often vary clinally, likely because they are subject to strong, omnipresent selection due to temperature and O_2 availability (Storz et al., 2010a). O_2 is the critical final electron acceptor in cellular respiration; profound physiological adjustments are required to insure that its uptake at the blood-gas barrier and its delivery to respiring tissues will be maintained despite ambient hypoxia (Storz et al., 2010a). Reduced partial pressure of oxygen (PO_2) poses a challenge to metabolically active species like birds, for which flight requires a 15-20 fold increase in metabolic rate (Ward et al., 2002). Previous work has identified several genes

that are targets of selection with respect to elevation. Mitochondrial DNA (mtDNA) may be under selection at high altitudes due to its role in thermogenesis (Mishmar et al., 2003; Gering et al., 2009) or to mitigate the deleterious effects of reactive oxygen species (ROS) produced in hypoxic conditions (Scott et al., 2011). Selection on hemoglobin (Hb) subunits α^A, α^D and β^A to alter Hb-O_2 binding affinity and increase O_2 distribution efficiency has been well characterized in a number of high elevation taxa (Jessen et al., 1991; McCracken et al., 2009; Storz et al., 2010a; Storz et al., 2010b; Projecto-Garcia et al., 2013; Wilson et al. , 2013). The strong selective pressures associated with cold temperature and hypoxia should favor local adaptation between high and low populations and may result in divergence along this axis. However the accumulation of specialized phenotypes may be eroded through gene flow (Garant et al., 2007), and the elevational movements of individuals that occur in seasonal environments (Hua et al., 2016). We extend existing work by assessing functional variation in all hemoglobin loci, including the loci which contribute nearly exclusively to embryonic hemoglobin isoforms (α^{pi}, β^H, β^{rho} and $\beta^{epsilon}$; Opazo et al., 2015)), in a radiation that has a complex and rapid history of diversification in the Andes.

To investigate the influence of local adaptation across elevation gradients on diversification, we assessed inter- and intra-specific patterns of hemoglobin variation in a recent radiation of finches, the South American siskins (genus: *Spinus*). The South American *Spinus* shared a most recent common ancestor approximately one million years ago; the diversification rate for this clade is in the top quartile for birds, and exceptionally high in the Andes (Beckman & Witt, 2015). The *Spinus* clade includes species with a diverse set of elevational ranges and range amplitudes, including high elevation

specialists (≥3000 m), species that do not occur at high elevation (<3000 m), and one elevation generalist, a polytypic species that is distributed from sea level to ~4300 m (*Spinus magellanicus*, Ridgely & Tudor, 1989; Fjeldså & Krabbe, 1990). From sea level to the top of its range over 4000+ m, *S. magellanicus* experiences a 40% reduction in PO_2 (West 1996). The high rate of diversification across the heterogeneous, Andean landscape in *Spinus* suggests that local adaptation to environmental variables such as PO_2 and temperature may be pivotal in facilitating speciation.

In this study, we assessed the hypothesis that adaptation to high-altitude hypoxia in Andean *Spinus* has promoted diversification within and among species. We predicted that, in the absence of any effect of elevation on divergence, hemoglobin functional variation should reflect known relationships among *Spinus* taxa (Beckman, Chapter 2). Conversely, hemoglobin variation segregating based on elevational range would be evidence that elevation has impacted divergence among *Spinus* taxa through selection on standing variation (as in Colosimo, 2005), *de novo* mutations or introgressed alleles (see Beckman, Chapter 2). Second, if local adaptation to elevation has occurred in *Spinus magellanicus*, we predicted that non-synonymous mutations in hemoglobin should vary significantly with elevation and be differentiated between high and low populations. We tested these predictions by identifying the non-synonymous mutations in all avian hemoglobin loci across the *Spinus* clade, and in two parallel elevational transects of *S. magellanicus* on the west slope of the Peruvian Andes. We included all seven hemoglobin loci in our analysis since all may contribute to an individual's oxygen transport in passerines. Three of these loci are found in avian adult hemoglobin isoforms (α^A, α^D, and β^A) and four are found exclusively in embryonic hemoglobin isoforms (α^{pi},

β^{rho}, $\beta^{epsilon}$, and β^{H}) (Opazo et al., 2015). We also sequenced 500 genomically distributed exonic regions that should represent neutral population structure among *S. magellanicus.*

METHODS:

Sampling:

We assembled frozen tissue samples for 88 individuals from the genus *Spinus* from museum collections and our own field expeditions (see Appendix). This included one representative of each of the following lowland or mid-elevation taxa (found only below 3000 m): *S. notatus, S. cucullatus, S. olivaceus, S. siemiradzkii,* and *S. barbatus*. We included one to two individuals of the following highland taxa found above 3000 m: *S. atratus*, *S. crassirostris* and *S. uropygialis*. For the widespread, polytypic *S. magellanicus*, we sampled one individual each from the following localities: northern Argentina at 209 m; eastern Bolivia at 430 m; and central Andean Bolivia at 2470 m. Lastly, we collected *S. magellanicus* along two linear transects in the Peruvian departments of Ancash (N=39) and Lima (N=37) (Fig. 1). For each transect, we sampled three to ten individuals at each of six elevations from 100 m to 4100 m; exact elevations differed between transects.

Sequencing:

We used a target capture approach (reviewed in Jones & Good, 2015) to sequence all seven hemoglobin loci including coding, intronic and flanking regions for 88 *Spinus*. To do this, we constructed alignments for each hemoglobin locus using whole genome data from *Serinus canaria, Zonotrichia leucophrys* and *Geospiza fortis*, and mRNA from

across the nine-primaried oscines. Alignments were built with MUSCLE v3.7 (Edgar, 2004) and evaluated by eye. We designed 120 base pair probes to tile across each hemoglobin alignment for target enrichment. Using the *Serinus canaria* genome, we also designed one to two probes for each of 500 exons from independent loci, spread across the avian genome. These 500 additional loci should were intended to represent the distribution of population divergence levels found across the coding parts of the *Spinus* genome.

We extracted DNA from muscle tissue samples with the Qiagen DNeasy Blood and Tissue kit (Qiagen, Valencia, CA) using the manufacturer's protocol. We prepared libraries for each individual with the NEBNext Ultra DNA library prep kit (New England Biolabs, Ipswich, MA), and NEBNext Multiplex Oligos for Illumina (New England Biolabs, Ipswich, MA). We aimed for an insert of 400 to 500 base pairs and used the custom target enrichment kit, MYbaits (Mycroarray, Ann Arbor, MI). Insert length and library quality were assessed with a Bioanalyzer (Agilent) before and after hybridization. We pooled and sequenced all 88 individuals in a single Illumina HiSeq 2500 lane (160 bp, paired-end). Sequencing was performed at the Keck Center at the University of Illinois, Champaign-Urbana. Raw reads were demultiplexed and filtered for PHRED score > 30 at the Keck Center.

To build reference alignments for each complete hemoglobin locus, we combined sequence data for six high coverage individuals and aligned the data to the mRNA sequences for hemoglobin loci α^A, α^D and β^A using aTRAM v1.04 (Allen, Huang, Cronk, & Johnson, 2015) with the assembler Velvet v1.2.10 (Zerbino & Birney, 2008). The program aTRAM conducts a series of BLAST searches to rapidly extract sequence

fragments that align to a user-specified reference. This reference is updated prior to each BLAST search to incorporate the newly extracted fragments; thus the reference gains in length and accuracy with each iteration. If data support multiple, alternative references, as in the case of paralogous genes like hemoglobin, aTRAM will build and report each unique reference. To correctly identify the independent, paralogous hemoglobin loci, we interleaved successive aTRAM analyses with hand curation of the best aTRAM results in Geneious v6 (http://www.geneious.com, Kearse et al., 2012). During hand curation, we considered (1) how suggested loci compared to non-machine generated material in Genbank, (2) coverage, including the presence of paired-end reads vs single reads, and (3) the ability to recover regions across all *Spinus*. We favored putative loci that were supported by paired-end reads that spanned intronic and exonic regions. By alternating aTRAM runs with hand-curation, we extracted complete loci including the divergent introns and conserved exons for all seven hemoglobin loci. In the case of the 500 exonic regions, we used the 120 base pair probes as our initial reference in aTRAM and selected the longest, well-supported aTRAM sequence as our final reference.

We constructed alignments for all loci by conducting a sensitive search in Bowtie2 v2.2.6 (Langmead & Salzberg, 2012) based on the refined reference sequences for each individual. Hemoglobin assemblies were run through custom, locus-specific python scripts to remove paralogous reads. After processing, we called the variable nucleotide sites across all individuals using SAMtools v1.3.1 (Li et al., 2009) and bcftools v1.3.1 (Li et al., 2009). Variable sites that did not pass coverage and quality thresholds were eliminated from future analysis. We conducted an additional evaluation by-to confirm variable sites in low coverage individuals for all hemoglobin loci.

Analysis:

We characterized all variable sites in hemoglobin exons as silent or non-synonymous mutations. To visualize the amino acid sequence variation, we identified haplotypes for non-synonymous nucleotide changes in each locus and constructed a haplotype network using an infinite sites model in R v3.2.4 (R Core Team, 2016) with the packages adegenet v2.0.1 (Jombart and Ahmed, 2011) and pegas v0.6 (Paradis, 2010).

A subset of these non-synonymous nucleotide sites were variable within at least one of the *S. magellanicus* elevational transects, with the minor allele frequency greater than 0.05. For the remaining analyses, we considered the Ancash and Lima elevational transects independently. For each transect, we binned individuals into six elevational populations: 100 m, 700 m, 1500 m, 2300 m, 3000 m, and 3800 m for Ancash; <400 m, 900 m, 1600 m, 2500 m, 3100 m and 3700 m for Lima. We calculated the frequencies of each allele in each population and examined the relationship between allele frequency and elevation using linear regression methods in R v3.2.4 (R Core Team, 2016).

To test the significance of differentiation in non-synonymous mutations between high and low *S. magellanicus* populations, we conducted an Fst outlier analysis for each transect in R v3.2.4 (R Core Team, 2016) with the package "pegas" v0.6 (Paradis, 2010). For this analysis, we calculated Fst for all variable sites in the 500-exon dataset present in both the high (> 3700 m) and low (< 1000 m) populations. We compared the differentiation of each variable hemoglobin site to the genome-wide Fst distribution for SNPs in coding DNA.

RESULTS:

Sequencing:

Our target-capture sequencing approach resulted in 411 million high quality reads (Phred score >30), with an average of 2.3 million reads per individual (S.D.=0.81 million reads). From this data, we assembled seven hemoglobin coding loci using aTRAM. Six out of the seven loci had complete introns and exons (Table 1), whereas one β-globin sequence was incomplete (see below). Seven hemoglobin loci were present in all 88 individuals; average coverage per exon was 7.2X, though coverage varied by locus. From the sequence data of six, pooled, high-coverage individuals, we built reference sequences for 472 of the 500 genomically distributed exons included in our target enrichment. We successfully aligned an average of 83.4 individuals to 452 exonic regions.

Hemoglobin locus structure:

Using manually curated sequences available on Genbank, we unambiguously identified five of the seven hemoglobin loci that we recovered: α^{A}, α^{D}, α^{pi}, β^{A}, and β^{H}. The remaining two loci, β^{rho} and $\beta^{epsilon}$, shared identical reference sequences across exon 1, intron 1 and exon 2. However, in intron 2, the sequences diverged significantly. Based on coverage and the presence of paired-end reads overlapping exon 2 and intron 2, we have confidence that we recovered two unique loci with identical 5' end sequences. We reconstructed exon 3 for one of these loci, hereafter $\beta^{rho\text{-}like\ 1}$. For the second locus, $\beta^{rho\text{-}like\ 2}$, we truncated exon 3 after 44 nucleotides since we were unable to distinguish, due to modest sequencing coverage, between an exon 3 identical to that of $\beta^{rho\text{-}like\ 1}$ and a long

repeat region. The remarkable reference sequence similarity at the 5' end of $\beta^{\text{rho-like 1}}$ and $\beta^{\text{rho-like 2}}$ is consistent with gene conversion having occurred recently between β^{rho} and β^{epsilon}. Both loci were recovered across the *Spinus* clade, suggesting that gene conversion had occurred prior to the divergence of South American *Spinus.*

Hemoglobin variation across Spinus:

We identified amino acid variation across the *Spinus* clade in all hemoglobin loci, and characterized it into three patterns. First, we found that the amino acid variants were associated with elevation for seven out of 10 polymorphic sites in loci that contribute to the major hemoglobin isoforms in adult birds (α^{A}, α^{D}, and β^{A}). *Spinus* populations that occurred exclusively below 3000 m shared an ancestral amino acid state, inferred on the basis of identity to the amino acid state observed in an outgroup, *S. notatus* (Figs. 2, 3). *Spinus* species and *S. magellanicu*s populations that extended above 3000 m were fixed (see α^{34A}) or polymorphic (see β^{21A}) for an alternative amino acid. These high elevation taxa included *S. uropygialis, S. crassirostris, S. atratus* and Andean *S. magellanicus* from Peru and Bolivia. The low/high dichotomy was also present in β^{56H} and β^{135H}. This hemoglobin pattern was not congruent with known evolutionary history of *S. magellanicus* from Bolivia and Argentina. Previous genomic analyses revealed these populations to be sister (Beckman, Chapter 2). However, *S. magellanicus* from Andean Bolivia shared the high elevation allele in five of the seven amino acid positions in α^{A}, α^{D}, and β^{A} where an alternative was present only in high elevation lineages. At corresponding positions, the birds from eastern Bolivia and Argentina possessed the ancestral amino acid state that is typical of lowland populations and species.

We found some variable sites at which all *Spinus* taxa except Andean *S. magellanicus* shared an ancestral amino acid state. This was true for the majority of variable sites in $\beta^{\text{rho-like 1}}$ and $\beta^{\text{rho-like 2}}$, which were rare, minor alleles found only in Andean *S. magellanicus*; it was also true for some minor alleles that occur at moderate frequency, such as α^{60pi} (see Fig. 2).

Third, many amino acid variants were distributed unpredictably across *Spinus* taxa with respect to phylogenetic history (see Beckman, chapter 2) and elevational range; examples include β^{20A}, which is present in high-elevation *S. atratus, S. uropygialis* and Peruvian *S. magellanicus*, as well as the lower elevation *S. olivaceus* (a mid-montane species that does not reach as high as 3000 m). Imperfect correspondence between amino acid variation and elevation characterized the embryonic hemoglobin loci α^{pi} and β^{H}. The locus β^{H}, in particular, had a high number of variable and polymorphic amino acid sites; 13 total sites compared to an average four non-synonymous sites per locus for other hemoglobin chains. The majority of the sites in β^{H} were variable across the *Spinus* clade, and were generally unpredictable based on phylogeny or elevation (Fig. 3).

Intraspecific variation across elevation:

Among *S. magellanicus* from Peru, we discovered fifteen non-synonymous hemoglobin polymorphisms that had minor allele frequencies greater than 0.05. Of these fifteen sites, four had a significant relationship with elevation on at least one transect (Table 2; Fig. 4). Three of these sites occurred in Ancash (α^{60pi}, β^{74H}, and β^{137H}) and one site (β^{135H}) in Lima. No site was significant across both transects. All significant sites were in the embryonic hemoglobin loci α^{pi} and β^{H}.

Amino acids changes at α^{60pi}, were from a serine at low elevation to phenylalanine at high, and in β^{74H}, from aspartic acid (low) to asparagine (high). Both cases resulted in alterations of the polarity and charge of the amino acid functional group. At β^{135H}, the amino acid isoleucine increased in frequency with elevation compared to the alternative valine. Alanine at β^{137H} increased in frequency with elevation as well, compared to valine. Both these amino acid changes did not alter the nonpolar functional type of the amino acid side chains. However, Valβ^{135H} to isoleucine changes the hydrocarbon side chain from three to four carbons while the change from valine to alanine at β^{137H} reduces the number of carbons from three to one in the side chain.

The association between allelic variation and elevation in *S. magellanicus* at β^{20A} and β^{21A}, was not significant despite the suggestive low/high pattern observed across *Spinus* taxa (Fig. 5). However, the steep shift in allele frequency in Ancash from 3000 m to 3800 m for β^{21A}, implies that a linear model may not be the ideal way to describe the elevational shift from non-polar alanine to polar, uncharged threonine.

To test the background level of differentiation between high and low populations, we calculated Fst for each SNP found in the 452 genomically distributed exons (see methods) between low (<1000 m) and high (>3700 m) populations on each transect. Ninety-five percent of loci exhibited Fst values less than 0.2299 (Ancash) and 0.1655 (Lima). We used this distribution to compare Fst values for each variable site in loci that contribute to adult hemoglobin (α^{A}, α^{D}, and β^{A}) as well as non-synonymous sites that varied significantly with elevation (Table 3). We found SNPs with Fst values that exceeded the 95 percentile in hemoglobin α^{pi}, β^{A}, and β^{H}. Five of the eight significant SNPs were non-synonymous mutations; these included α^{60pi}, β^{21A}, β^{27H},β^{45H} and β^{135H}

(Fig. 6). β^{135H} was the only site to be significantly differentiated in both Ancash and Lima; however, α^{60pi} was nearly significant on both transects, with an Fst in the 95 percentile for Ancash and in the 90th percentile in Lima (Fig 6).

DISCUSSION:

We set out to test whether adaptation to high-altitude hypoxia in Andean *Spinus* has promoted diversification within and among species. We did this by sequencing the functionally relevant hemoglobin loci that encode the embryonic and adult hemoglobin isoforms, the tetrameric protein essential for oxygen-transport throughout the body. We characterized the amino acid variation patterns across multiple *Spinus* species and several populations of the widespread *Spinus magellanicus* from different elevations. We found that most amino acid variation across the *Spinus* clade at major adult hemoglobin loci transcribed in adults was structured by elevational range, and we identified three amino acid positions (α^{60pi}, β^{21A} and β^{135H}) that were significantly differentiated between high and low populations in Peruvian *S. magellanicus.*

Interspecific patterns of amino acid variation in hemoglobin:

We predicted that hemoglobin amino acid variation would segregate based on elevational range limits rather than the known evolutionary relationships in *Spinus* based on neutral genome-wide SNPs. The phylogeny of Spinus is not fully resolved due to rapid bursts of speciation with rampant introgression (see discussion Beckman, Chapter 2). However, certain nodes with the *Spinus* tree have been resolved with high confidence. This allowed us to compare hemoglobin variation to neutral divergence

history. Specifically, based on a phylogenetic analysis of ~45,000 SNPs from across the genome, we reported *S. magellanicus* from Cochabamba, Bolivia was more closely related to *S. magellanicus* from northern Argentina than to *S. magellanicus* from Peru (Beckman, Chapter 2). Thus, if hemoglobin variation were to reflect divergence history, *S. magellanicus* from Bolivia and Argentina in this study should share amino acid identity. However, in five out of seven amino acid substitutions in which an alternative allele was defined in high *Spinus* in loci encoding adult hemoglobin, we found that lowland Bolivian and Argentina birds shared the ancestral allele while Andean Bolivian birds shared a derived allele with high elevation *Spinus* species.

There are a number of possible explanations for this pattern in *S. magellanicus* including selection on identical mutations that arose *de novo* and selection on standing variation (Colosimo, 2005; F. C. Jones et al., 2012). However, we argue that the most likely explanation is adaptive introgression between Andean Bolivian *S. magellanicus* and the high Andean specialist, *S. atratus*. We previously reported modest evidence for introgression between these lineages in formal introgression tests (Beckman, Chapter 2). There is precedence for the adaptive introgression of high-altitude alleles among other species. Natarajan et al. (2015) concluded that a high-altitude adapted β^A allele was introduced by interspecific introgression between waterfowl species (*Anas georgica* and *Anas flavirostris*). Further, an allele responsible for increased survival in high altitude hypoxia in humans may be the result of adaptive introgression from the archaic Denisovan lineage (Huerta-Sánchez et al., 2014; Hackinger et al., 2016, but see Lou et al., 2015).

Overall, we find that non-synonymous hemoglobin variation reflected elevational range limits more than neutral divergence history in the loci α^{A}, α^{D}, and β^{A}. This conclusion did not hold for polymorphisms in loci which contribute exclusively to embryonic hemoglobin with the exception of β^{135H} (see below).

Evidence for natural selection with respect to elevation:

We report strong evidence for population differentiation in *Spinus magellanicus* with respect to elevation in three hemoglobin amino acid positions: α^{60pi}, β^{21A} and β^{135H}. The only site we recovered which contributes to adult hemoglobin isoforms was β^{21A}; this amino acid polymorphism consists of a change from the ancestral, nonpolar alanine to polar, uncharged threonine. Support for population structure at this SNP was modest. It was a significant Fst outlier in Ancash, where it showed a precipitous decline in allele frequency from 3000 m to 3700 m on that transect. The *Spinus* clade-wide results are concordant with the hypothesis that the threonine at position 21 confers a functional advantage at high elevations. Threonine is fixed in *S. atratus* and *S. crassirostris,* two species that generally do not occur lower than ~3400 m, and the site is polymorphic in the high-altitude migrant species, *S. uropygialis* (Fjeldså & Krabbe, 1990). The population structure result at β^{21A} in *S. magellanicus* is further supported by preliminary functional analyses of Hb-O_2 binding affinity. Natarajan et al. (2016) report that the amino acid change from alanine to threonine reduces the P_{50} of the major hemoglobin isoform, indicating an increased affinity for O_2. The high Hb-O_2 affinity facilitates O_2 uptake at the blood/gas barrier in the lungs when the ambient partial pressure of O_2 is low

(Storz et al., 2010). This amino acid position, to our knowledge, has not previously been reported as a target for selection in avian hemoglobin (Natarajan et al., 2016).

We found evidence for natural selection with respect to elevation in the embryonic hemoglobin sites α^{60pi} and β^{135H}. Support for this conclusion was bolstered by multiple lines of evidence. Both loci varied significantly with elevation on one transect (α^{60pi}: Ancash; β^{135H}: Lima). α^{60pi} was a significant Fst outlier in Ancash and in the upper 90th percentile in Lima while β^{135H} was a significant outlier for both transects. Support for differentiation at the same site on independent transects, 350 km apart from one another, suggests that these sites are real targets for natural selection with respect to the partial pressure of oxygen. However, evidence for population structure at a SNP in the absence of functional data should be evaluated carefully (Cheviron et al., 2014). The taxonomic structure of the β^{135H} polymorphism, where the alternative amino acid isoleucine was fixed in all other high elevation *Spinus*, and low elevation *Spinus* were fixed for valine adds corroboration that β^{135H} has a significant role in high altitude adaptation. α^{60pi} is only polymorphic in Peruvian *S. magellanicus*.

Our finding of molecular high-altitude adaptaion in hemoglobin loci that exclusively embryonic, α^{pi} and β^{H}, is novel in birds. Most research into intraspecific variation in hemoglobin has focused on the loci which are involved in oxygen transport in adults (McCracken et al., 2009; Wilson et al., 2013; Cheviron et al., 2014; Galen et al., 2015). Our findings suggest that the embryonic stage may also be an important target for high altitude adaptation in birds. Amino acid substitutions in the hemoglobin subunits and their functional consequences may be crucial components to understanding fitness costs at the embryonic stage. Further, little is known about the relative abundances of

alternative embryonic transcripts across development. The lineage specific abundances of embryonic hemoglobin transcripts are only characterized for the chicken (*Gallus gallus*) and Zebra finch (*Taeniopygia guttata*) (Alev et al., 2009). In adult birds, the relative abundances of α^A and α^D transcripts helps determine the average hemoglobin-oxygen affinity. Modifications in the transcription of the higher affinity locus, α^D, can result in a higher or lower lineage-specific hemoglobin affinity, but this mechanism does not appear to be an important component of adaptation to high altitudes (Zhang et al., 2014; Opazo et al., 2015). Hemoglobin functional studies and expression assays for each embryonic locus could provide additional insight into how early developmental stages might contribute to high altitude adaptation and diversification along elevational gradients.

Gene conversion among β^{rho} and $\beta^{epsilon}$:

Our target capture sequencing approach revealed an example of gene conversion among embryonic β^{rho} and $\beta^{epsilon}$. Gene conversion between these loci was previously documented in the zebra finch, golden-collared manakin, downy woodpecker, medium ground finch and chicken among others (Reitman et al., 1993; Hoffmann et al., 2010; Opazo et al., 2015). Little is known about the mechanism by which gene conversion occurs between these loci, the frequency of gene conversion across avian diversity or its consequences. Our target capture sequence method is a promising approach for identifying additional examples of hemoglobin gene conversion across birds, and could be used in addressing these questions.

CONCLUSIONS:

We found that several non-synonymous mutations in the subunits of the oxygen transport protein hemoglobin were associated with elevation in a widespread South American siskin, *Spinus magellanicus*. They were also associated predictably with elevational distribution limits among species across the *Spinus* clade. Our result supports a role for elevation in promoting functional divergence among *Spinus* across the Andes, even across relatively short distances. The Ancash transect was ~86 km from low to high populations. These findings are concordant with other avian studies that have found intraspecific adaptation to elevation via hemoglobin (Wilson et al., 2013; Galen et al., 2015). Among species that are continuously distributed and highly vagile, like finches (Ridgely & Tudor, 1989; Price, 2011), selection upon hemoglobin loci may ultimately reduce gene flow, providing an initial isolating barrier between high and low populations with important evolutionary consequences (Sobel et al., 2010).

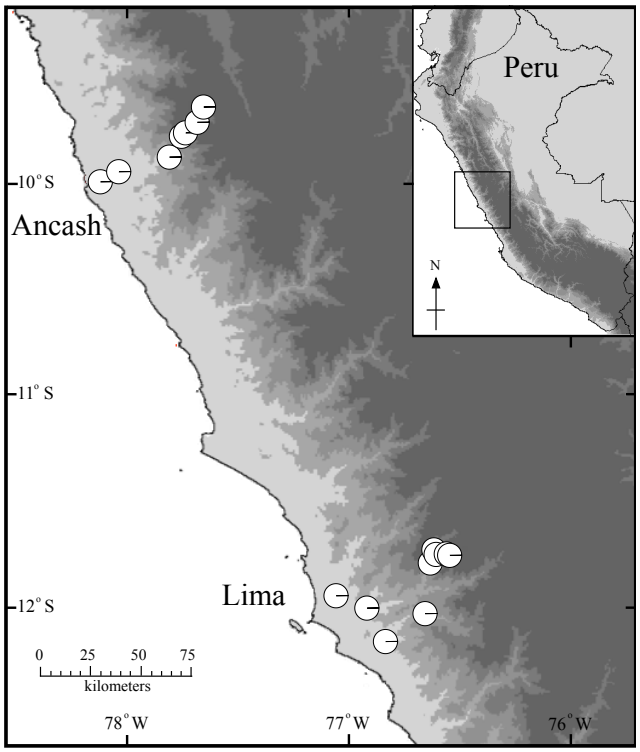

Fig. 1. *S. magellanicus* transect sampling localities. Circles represent three to eight individuals. Inset, upper right, shows area of location of transects (black box) within Peru.

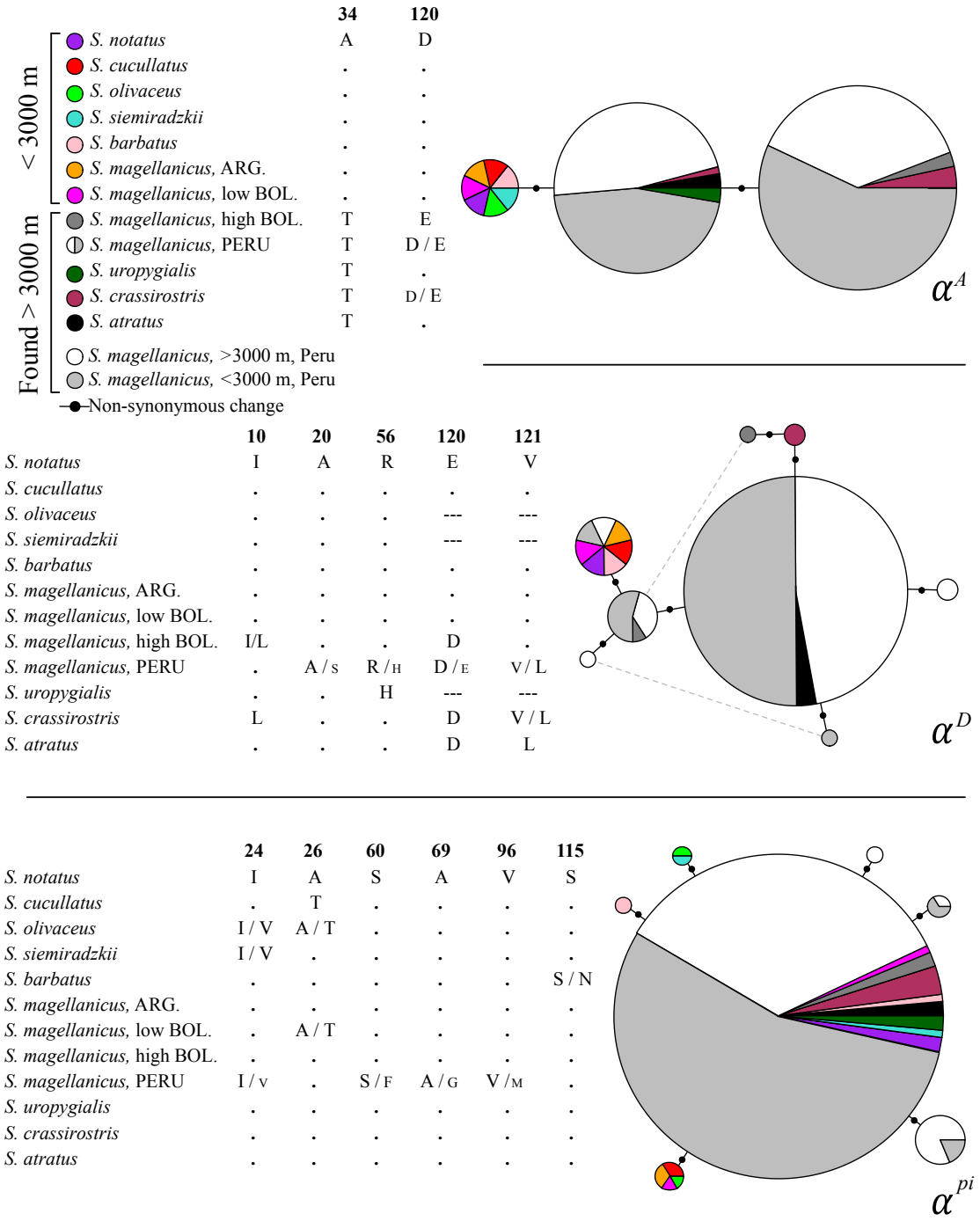

Fig. 2. Non-synonymous changes in α–hemoglobin loci across *Spinus*. For each locus, amino acid variation by population is indicated in a table (left) and haplotype network (right). The table depicts variation by amino acid position relative to outgroup *S. notatus*; font size indicates relative frequency if two alleles are present in a population. Color key for haplotype network located upper right; Circle size demonstrates relative haplotype frequency among sampled individuals.

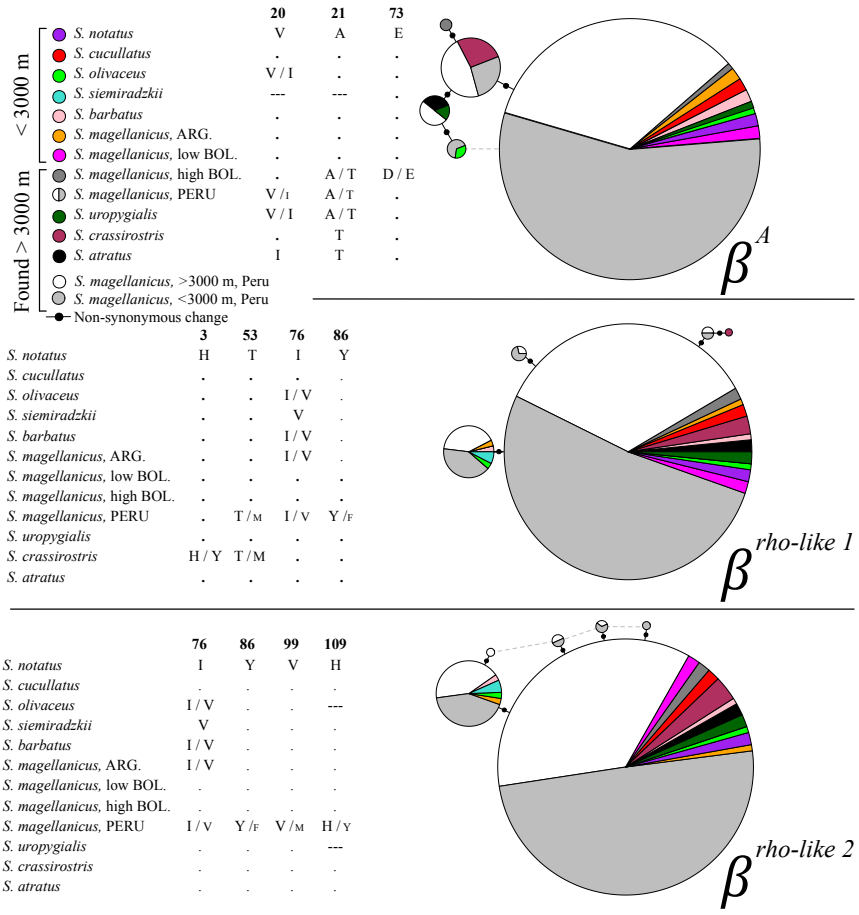

Fig. 3. Non-synonymous changes in β–hemoglobin across *Spinus*. For each locus, amino acid variation indicated in a table (left) and haplotype network (right). The table depicts variation by amino acid position relative to outgroup *S. notatus*; font size indicates relative frequency in population. Color key for haplotype network located upper right; circle size demonstrates relative haplotype frequency.

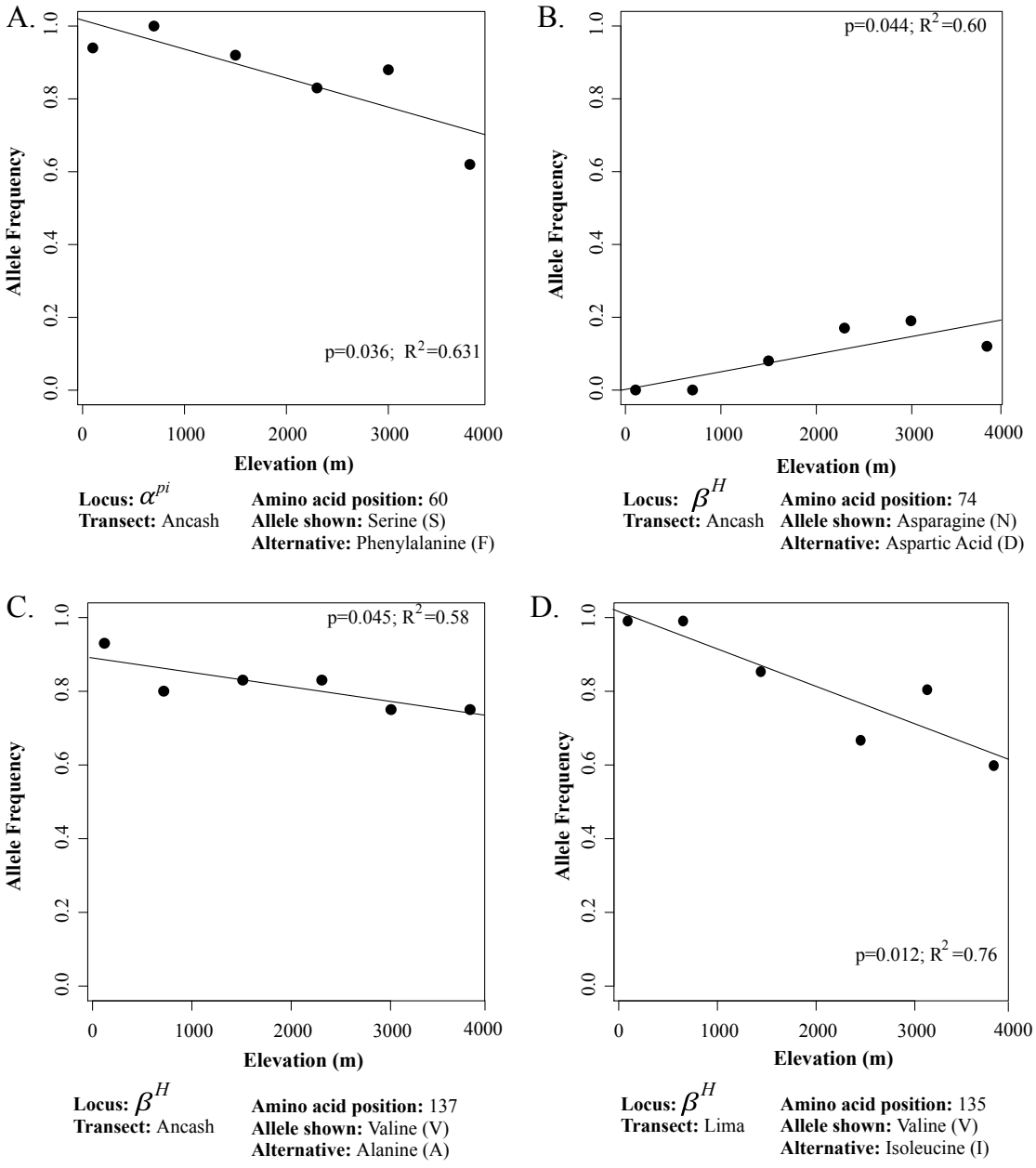

Fig. 4. Linear regressions between non-synonymous allele frequency and elevation in *Spinus magellanicus*. Strength of relationship indicated by P-value and R^2 shown on each graph. **A.** Hemoglobin α^{60pi}, Ancash transect. **B.** Hemoglobin β^{74H}, Ancash transect. **C.** Hemoglobin β^{137H}, Ancash transect. **D.** Hemoglobin β^{135H}, Lima transect.

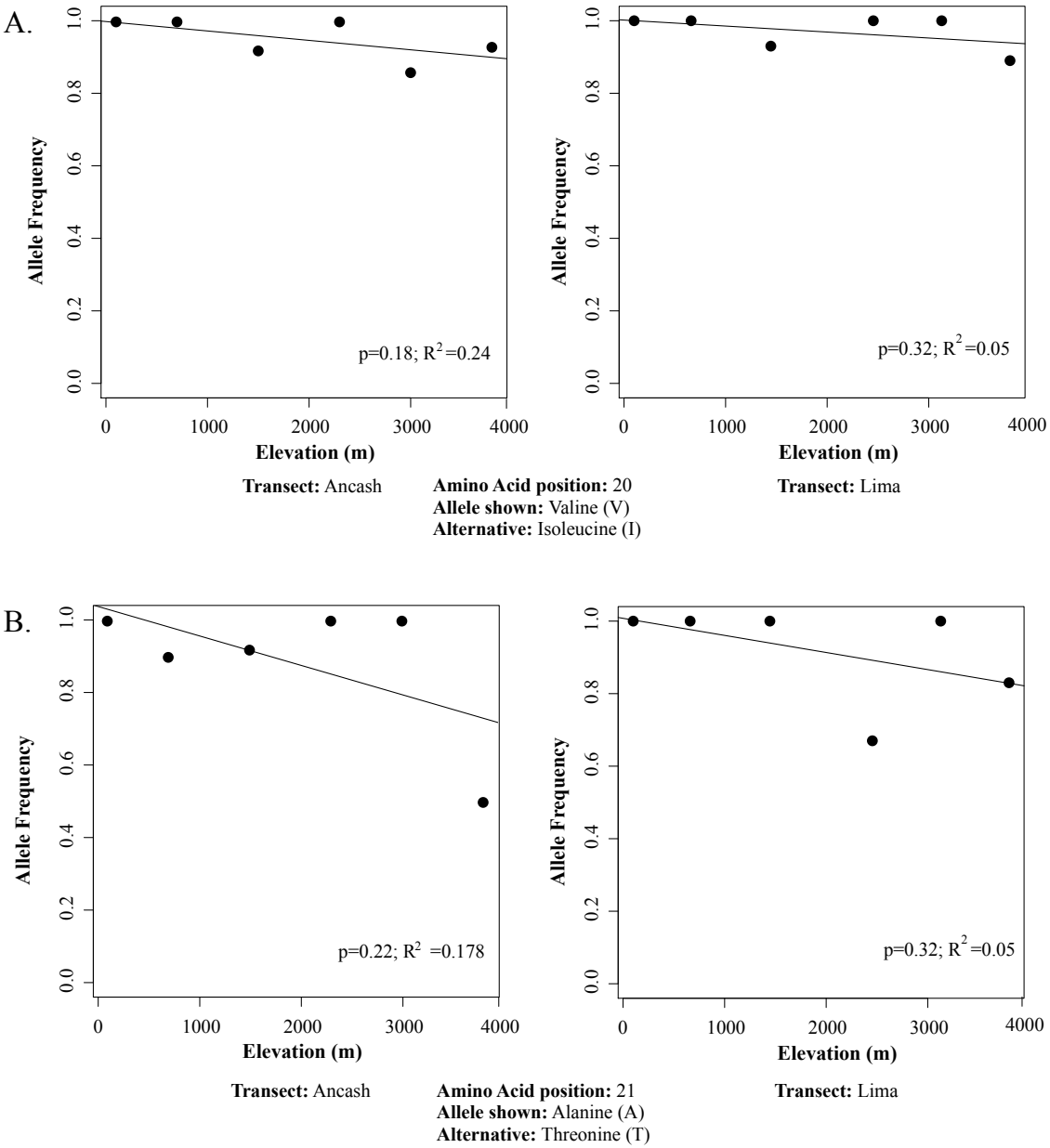

Fig. 5. Linear regressions between non-synonymous allele frequency and elevation in *Spinus magellanicus*. Strength of relationship indicated by P-value and R^2 shown on each graph. **A.** Hemoglobin β^{20A}. **B.** Hemoglobin β^{21A}.

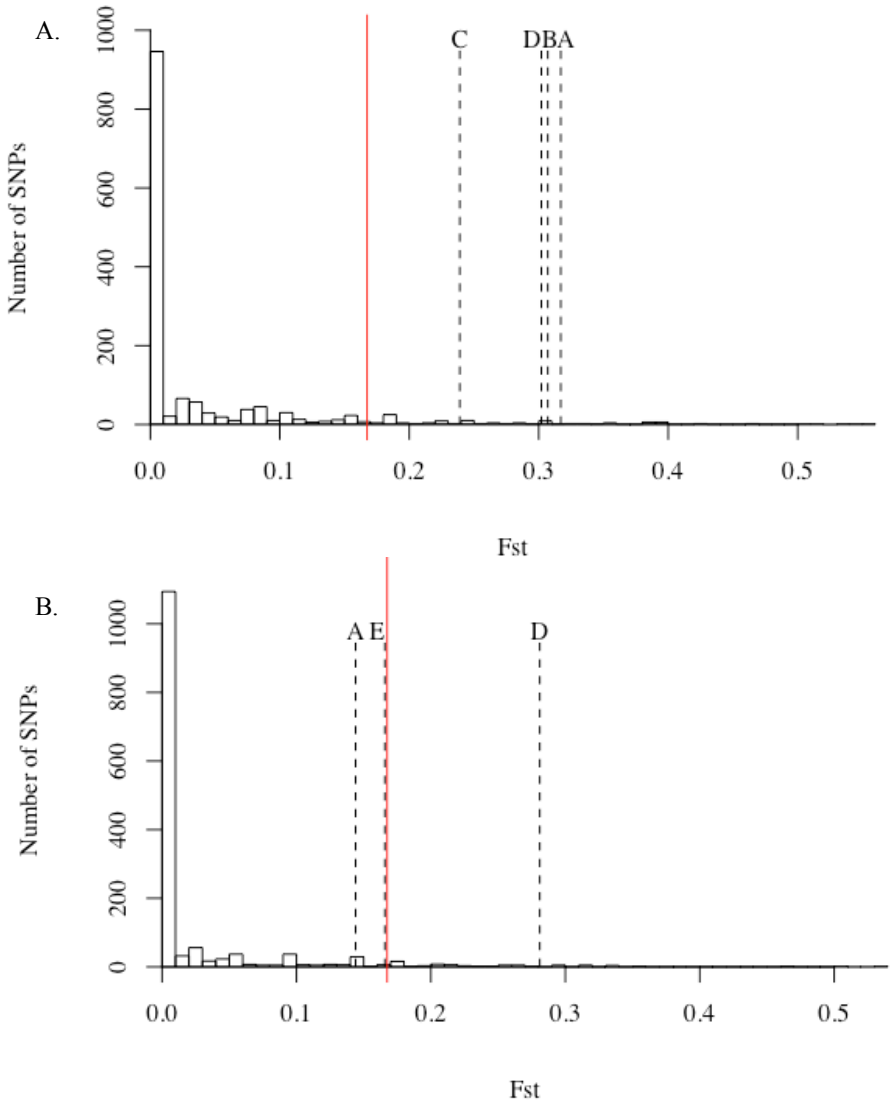

Fig 6. Distribution of Fst values between *S. magellanicus* populations from <1000 m and >3700 m in Ancash (A) and Lima (B). Fst values greater than 95% of all variable sites are right of the red line. Dashed lines indicate transect specific Fst for α^{60pi} (A), β^{21A}(B), β^{27H} (C), β^{135H} (D), and E: β^{45H} (E).

Table 1. Summary of characteristics in hemoglobin reference sequences. References were assembled from pooled *Spinus* sequence data in aTRAM. Columns indicate the number of base pairs included in the total reference, exonic regions, intronic regions and flanking regions for each globin locus.

Locus	Total	Exons	Introns	Flanking
α^{A}	827	429	256	142
α^{D}	807	426	366	15
α^{pi}	2968	426	1500	1042
β^{A}	2196	444	929	823
β^{H}	1907	444	803	660
$\beta^{\text{rho-like 1}}$	2067	444	702	921
$\beta^{\text{rho-like 2}}$	997	360*	512	125

Table 2. Linear regression results of non-synonymous hemoglobin allele frequency changes by elevation in Peruvian *S. magellanicus.* Amino acid column indicates position of mutation in amino acid sequence. Regressions with p value < 0.05 are in bold.

		Ancash		Lima	
Locus	**Amino Acid**	**R^2**	**p value**	**R^2**	**p value**
α^A	120	-0.25	0.925	-0.20	0.705
α^D	121	-0.24	0.853	-0.09	0.482
α^{pi}	60	0.63	**0.036**	0.05	0.320
β^A	20	0.24	0.180	0.06	0.319
β^A	21	0.18	0.220	0.05	0.323
$\beta^{rho\text{-}like\ 1}$	76	-0.18	0.641	0.19	0.215
$\beta^{rho\text{-}like\ 2}$	76	-0.24	0.850	-0.09	0.493
β^H	14	-0.20	0.711	-0.17	0.630
β^H	30	-0.23	0.818	0.25	0.175
β^H	45	-0.17	0.634	-0.25	0.921
β^H	56	-0.14	0.570	-0.18	0.640
β^H	74	0.60	**0.044**	-0.24	0.856
β^H	76	0.25	0.179	-0.21	0.722
β^H	135	0.12	0.258	0.77	**0.012**
β^H	137	0.58	**0.045**	0.24	0.186

Table 3. Fst outliers in hemoglobin between *S. magellanicus* populations below 1000 m and above 3700 m. Amino acid is abbreviated to AA. Bold Fst values were greater than 95% of all Fst values per transect, calculated from all high quality SNPs from 452 genomically distributed exonic regions.

Locus	Position	Mutation type	Fst Ancash	Fst Lima
α^{pi}	AA 60	Non-synonymous	**0.317**	0.144
α^{pi}	Flanking	---	-0.043	**0.248**
β^{A}	AA 21	Non-synonymous	**0.307**	-0.036
β^{A}	Intron 2	---	**0.235**	-0.089
β^{A}	AA 124	Silent	**0.306**	-0.066
β^{H}	AA 27	Non-synonymous	**0.239**	-0.074
β^{H}	AA 45	Non-synonymous	-0.07	**0.1657**
β^{H}	AA 135	Non-synonymous	**0.302**	**0.281**

Appendix A. Sample table for Chapter 1. The museum abbreviations are MSB: Museum of Southwestern Biology; MVZ: Berkeley Museum of Vertebrate Zoology; CUMV: Cornell University Museum of Vertebrates; KU: Biodiversity Institute and Natural History Museum, University of Kansas AMNH: American Museum of Natural History; USNM: Smithsonian Institution; UWBM: Burke Museum;. FMNH: Field Museum of Natural History; ANSP: Academy of Natural Sciences, Philadelphia.

Voucher/source	Species	Locality	ND2	ND3	cytb	Myo2	MUSK
MSB:Bird:29180	C. vespertinus	USA; NM; USFS Rd. 439	Ready	Ready	Ready	Ready	Ready
NRM:20026502	Serinus canaria	Captivity	–	JN715576	–	JN715301	–
unknown	Serinus canaria	Captivity	EU327666	–	–	–	–
unknown	Serinus canaria	Unknown	–	–	AY914135	–	–
MVZ:Bird:176979	Astragalinus lawrencei	USA; CA; Weldon	Ready	–	Ready	Ready	Ready
MVZ:Bird:176985	Astragalinus lawrencei	USA; CA; Weldon	Ready	Ready	Ready	Ready	Ready
MSB:Bird:29194	Astragalinus psaltria	USA; NM; Albuquerque	Ready	Ready	Ready	Ready	Ready
MSB:Bird:29161	Astragalinus tristis	USA; NM; Taos County	Ready	Ready	Ready	Ready	Ready
MSB:Bird:34123	Spinus atratus	PER; Apurimac; Lag. Anantay	Ready	Ready	Ready	Ready	Ready
CUMV:52927	Spinus atratus	ARG; Jujuy; Quebralena	Ready	Ready	Ready	Ready	Ready
CUMV:52940	Spinus atratus	ARG; Jujuy; Quebralena	Ready	Ready	Ready	Ready	–
AMNH:DOT4593	Spinus atriceps	MEX; Chiapas; Papales	Ready	Ready	Ready	Ready	Ready
AMNH:DOT7151	Spinus atriceps	MEX	Ready	Ready	Ready	Ready	Ready
AMNH:DOT10430	Spinus barbatus	ARG; Neuquen; Centenario	Ready	Ready	Ready	Ready	Ready
AMNH:DOT10438	Spinus barbatus	ARG; Neuquen; Centenario	Ready	Ready	Ready	Ready	Ready
KU:11739	Spinus barbatus	ARG; Rio Negro; El Bolson	Ready	Ready	Ready	Ready	Ready
KU:11772	Spinus barbatus	ARG; Rio Negro; El Bolson	Ready	Ready	Ready	Ready	Ready
KU:11773	Spinus barbatus	ARG; Rio Negro; El Bolson	Ready	Ready	Ready	Ready	Ready
MSB:Bird:27138	Spinus crassirostris	PER; Cusco; N Urubamba	Ready	Ready	Ready	Ready	Ready
MSB:Bird:33092	Spinus crassirostris	PER; Cusco; Choquechaca	Ready	Ready	Ready	Ready	Ready
MSB:Bird:34207	Spinus crassirostris	PER; Apurimac; Lag. Anantay	Ready	Ready	Ready	Ready	Ready
USNM:B12867	Spinus cucullatus	GUY; Wiwitau Mountain, E. Rupununi Sav.	Ready	Ready	Ready	Ready	Ready
USNM:B12868	Spinus cucullatus	GUY; Wiwitau Mountain, E. Rupununi Sav.	Ready	Ready	Ready	Ready	Ready
AMNH:DOT3049	Spinus dominicensis	DOM; Pedernales; El Aceitillar	Ready	Ready	Ready	Ready	Ready
AMNH:DOT6927	Spinus dominicensis	DOM; Pedernales; El Aceitillar	Ready	Ready	Ready	–	Ready
AMNH:DOT6930	Spinus dominicensis	DOM; Pedernales; El Aceitillar	Ready	Ready	Ready	Ready	Ready
MSB:Bird:27175	Spinus magellanicus	PER; Cusco; NW Urubamba	Ready	Ready	Ready	Ready	Ready
MSB:Bird:27176	Spinus magellanicus	PER; Cusco; NW Urubamba	Ready	Ready	Ready	Ready	Ready
MSB:Bird:28282	Spinus magellanicus	PER; Lima; Carhuayumac	Ready	Ready	Ready	Ready	Ready
MSB:Bird:28342	Spinus magellanicus	PER; Lima; Carhuayumac	Ready	Ready	Ready	Ready	Ready
MSB:Bird:31497	Spinus magellanicus	PER; Lima; Carampoma	Ready	Ready	Ready	Ready	Ready
MSB:Bird:28373	Spinus magellanicus	PER; Lima; Chinchaycocha	Ready	Ready	Ready	Ready	Ready
MSB:Bird:28399	Spinus magellanicus	PER; Lima; Chinchaycocha	Ready	Ready	Ready	Ready	Ready
MSB:Bird:33112	Spinus magellanicus	PER; Cusco; Choquechaca	Ready	Ready	Ready	Ready	Ready

Appendix A cont.

Voucher/source	Species	Locality	ND2	ND3	cytb	Myo2	MUSK
MSB:Bird:33140	Spinus magellanicus	PER; Cusco; Choquechaca	Ready	Ready	Ready	Ready	Ready
MSB:Bird:33369	Spinus magellanicus	PER; Lima; Potago	Ready	Ready	Ready	Ready	Ready
MSB:Bird:33464	Spinus magellanicus	PER; Lima; Potago	Ready	Ready	Ready	Ready	Ready
MSB:Bird:34073	Spinus magellanicus	PER; Piura; Abra Porculla	Ready	Ready	Ready	Ready	Ready
MSB:Bird:34232	Spinus magellanicus	PER; Ancash; Huarmey	Ready	Ready	Ready	Ready	Ready
MSB:Bird:34251	Spinus magellanicus	PER; Ancash; Huarmey	Ready	Ready	Ready	Ready	Ready
MSB:Bird:34268	Spinus magellanicus	PER; Ancash; Huarmey	Ready	Ready	Ready	Ready	Ready
MSB:Bird:34828	Spinus magellanicus	PER; Ancash; Santa	Ready	Ready	Ready	–	–
MSB:Bird:34867	Spinus magellanicus	PER; Ancash; Santa	Ready	Ready	Ready	–	–
MSB:Bird:34875	Spinus magellanicus	PER; Ancash	Ready	Ready	Ready	–	–
MSB:Bird:31065	Spinus magellanicus	PER; Huancavelica; Castrovirreyna	Ready	Ready	Ready	Ready	Ready
UWBM:54421	Spinus magellanicus	ARG; Tucumán; San Miguel de Tucumán	Ready	Ready	Ready	Ready	Ready
UWBM:70271	Spinus magellanicus	ARG; Misiones; Posadas	Ready	Ready	Ready	Ready	Ready
UWBM:70272	Spinus magellanicus	ARG; Misiones; Posadas	Ready	Ready	Ready	Ready	Ready
UWBM:70716	Spinus magellanicus	ARG; Tucumán; San Miguel de Tucumán	Ready	Ready	Ready	Ready	Ready
UWBM:77454	Spinus magellanicus	BOL; Santa Cruz; Gutiérrez	Ready	Ready	Ready	Ready	Ready
FMNH:Bird:324099	Spinus magellanicus	PER; Madre de Dios; Hacienda Amazonia	Ready	Ready	Ready	Ready	Ready
FMNH:Bird:334722	Spinus magellanicus	BOL; Cochabamba; Cochabamba-Oruro Rd.	Ready	Ready	Ready	Ready	–
FMNH:Bird:334723	Spinus magellanicus	BOL; Cochabamba; Cochabamba-Oruro Rd.	Ready	Ready	Ready	Ready	Ready
FMNH:Bird:396033	Spinus magellanicus	BOL	Ready	Ready	Ready	Ready	–
UWBM:82846	Spinus notatus	MEX; Guerrero; Chilpancingo	Ready	Ready	Ready	Ready	Ready
FMNH:Bird:393899	Spinus notatus	MEX; Jalisco; Sierra de Manantlan	Ready	Ready	Ready	Ready	Ready
ANSP:18770	Spinus olivaceus	ECU; Zamora-Chinchipe; N Zumba	Ready	Ready	Ready	–	–
ANSP:19746	Spinus olivaceus	ECU; Zamora-Chinchipe; Paquisha	Ready	Ready	Ready	–	–
MSB:Bird:29160	Spinus pinus	USA; NM; Santa Fe	Ready	Ready	Ready	Ready	Ready
unknown	Spinus pinus perplexus	GTM; Quetzaltenago	–	–	DQ246804	–	–
ANSP:18991	Spinus siemiradzkii	ECU; Loja; N Zapotillo	Ready	Ready	Ready	–	–
ANSP:18992	Spinus siemiradzkii	ECU; Loja; N Zapotillo	Ready	Ready	Ready	–	–
ANSP:19701	Spinus siemiradzkii	ECU; Loja; W Pozul	Ready	Ready	Ready	–	–
ANSP:18789	Spinus spinescens	ECU; Carchi; CA. SE Impueran	Ready	Ready	Ready	–	–
ANSP:18807	Spinus spinescens	ECU; Carchi; CA. SE Impueran	Ready	Ready	Ready	–	Ready
NRM:986184	Spinus spinus	Sweden	JN715426	JN715517	–	JN715244	–
AHNU:A0001	Spinus spinus	Unknown	HQ915866	HQ915866	HQ915866	–	–
MSB:Bird:28312	Spinus uropygialis	PER; Lima; Carhuayumac	Ready	Ready	Ready	Ready	Ready
MSB:Bird:28326	Spinus uropygialis	PER; Lima; Carhuayumac	Ready	Ready	Ready	Ready	Ready
MSB:Bird:28329	Spinus uropygialis	PER; Lima; Carhuayumac	Ready	Ready	Ready	Ready	Ready
MSB:Bird:33471	Spinus uropygialis	PER; Lima; Potago	Ready	Ready	Ready	Ready	Ready

Appendix A cont.

Voucher/source	Species	Locality	ND2	ND3	cytb	Myo2	MUSK
ANSP:18543	Spinus xanthogastrus	ECU; Azuay; S Zhucay	Ready	Ready	Ready	–	–
Unknown	Spinus yarrellii	Captivity: BRA; Recife	–	–	U83200.1	–	–

Appendix B: Table of individuals sequenced for mitochondrial genes *cytb*, ND2, and ND3 in Chapter 2. Elevation (Elev.) is reported in meters. The museum abbreviations are MSB: Museum of Southwestern Biology; FMNH: Field Museum of Natural History; USNM: Smithsonian Institution; UWBM: Burke Museum; CUMV: Cornell University Museum of Vertebrates.

Individual	Species	Elev.	Locality	Latitude	Longitude	*cytb*	ND2	ND3
CUMV:Bird:52379	*Spinus atratus*	3650	Argentina: Jujuy: Quebralena	-23.27500	-65.76300	*Pending*	*Pending*	*Pending*
CUMV:Bird:52927	*S. atratus*	3650	Argentina: Jujuy: Quebralena	-23.27500	-65.76300	KT221306.1	KT221176.1	KT221240.1
CUMV:Bird:52940	*S. atratus*	3650	Argentina: Jujuy: Quebralena	-23.27500	-65.76300	KT221307.1	KT221177.1	KT221241.1
MSB:Bird:34099	*S. atratus*	4300	Peru: Apurimac: Abancay: Chacoche	-14.06005	-73.00160	*Pending*	*Pending*	*Pending*
MSB:Bird:34123	*S. atratus*	4300	Peru: Apurimac: Abancay: Chacoche	-14.06337	-73.00925	KT221305.1	KT221175.1	KT221239.1
MSB:Bird:33088	*S. atratus*	4400	Peru: Cusco: Ollantaytambo: Choquechaca	-13.18808	-72.23137	*Pending*	*Pending*	*Pending*
MSB:Bird:33129	*S. atratus*	3835	Peru: Cusco: Ollantaytambo: Choquechaca	-13.18808	-72.23137	*Pending*	*Pending*	*Pending*
MSB:Bird:28400	*S. atratus*	3800	Peru: Lima: San Pedro de Casta	-11.76077	-76.56767	*Pending*	*Pending*	*Pending*
MSB:Bird:34091	*S. crassirostris*	4300	Peru: Apurimac: Abancay: Chacoche	-14.06705	-73.02272	*Pending*	*Pending*	*Pending*
MSB:Bird:34115	*S. crassirostris*	4300	Peru: Apurimac: Abancay: Chacoche	-14.07003	-73.01275	*Pending*	*Pending*	*Pending*
MSB:Bird:34181	*S. crassirostris*	4300	Peru: Apurimac: Abancay: Chacoche	-14.05850	-73.00120	*Pending*	*Pending*	*Pending*
MSB:Bird:34207	*S. crassirostris*	4300	Peru: Apurimac: Abancay: Chacoche	-14.06635	-73.02360	KT221317.1	KT221187.1	KT221251.1
MSB:Bird:34208	*S. crassirostris*	4300	Peru: Apurimac: Abancay: Chacoche	-14.06635	-73.02360	*Pending*	*Pending*	*Pending*
MSB:Bird:34209	*S. crassirostris*	4300	Peru: Apurimac: Abancay: Chacoche	-14.06635	-73.02360	*Pending*	*Pending*	*Pending*
MSB:Bird:34212	*S. crassirostris*	4300	Peru: Apurimac: Abancay: Chacoche	-14.06635	-73.01895	*Pending*	*Pending*	*Pending*
MSB:Bird:27128	*S. crassirostris*	4330	Peru: Cusco: Urubamba: Urubamba	-13.19933	-72.16000	*Pending*	*Pending*	*Pending*
MSB:Bird:27138	*S. crassirostris*	4300	Peru: Cusco: Urubamba: Urubamba	-13.19933	-72.16000	KT221315.1	KT221185.1	KT221249.1
MSB:Bird:33077	*S. crassirostris*	4200	Peru: Cusco: Ollantaytambo: Choquechaca	-13.18808	-72.23137	*Pending*	*Pending*	*Pending*
MSB:Bird:33081	*S. crassirostris*	4206	Peru: Cusco: Ollantaytambo: Choquechaca	-13.18808	-72.23137	*Pending*	*Pending*	*Pending*
MSB:Bird:33091	*S. crassirostris*	4200	Peru: Cusco: Ollantaytambo: Choquechaca	-13.18808	-72.23137	*Pending*	*Pending*	*Pending*
MSB:Bird:33092	*S. crassirostris*	4200	Peru: Cusco: Ollantaytambo: Choquechaca	-13.18808	-72.23137	KT221316.1	KT221186.1	KT221250.1
MSB:Bird:31504	*S. crassirostris*	3981	Peru: Lima: Huarochiri: Carampoma	-11.62778	-76.43412	*Pending*	*Pending*	*Pending*
MSB:Bird:31505	*S. crassirostris*	3981	Peru: Lima: Huarochiri: Carampoma	-11.62778	-76.43412	*Pending*	*Pending*	*Pending*
MSB:Bird:31528	*S. crassirostris*	3973	Peru: Lima: Huarochiri: Carampoma	-11.62765	-76.43425	*Pending*	*Pending*	*Pending*
USNM:Bird:B12865	*S. cucullatus*	~150	Guyana: Wiwitau Mtn., E. Rupununi Sava.	02.86667	-59.26667	*Pending*	*Pending*	*Pending*
USNM:Bird:B12866	*S. cucullatus*	~150	Guyana: Wiwitau Mtn., E. Rupununi Sava.	02.86667	-59.26667	*Pending*	*Pending*	*Pending*
USNM:Bird:B12867	*S. cucullatus*	~150	Guyana: Wiwitau Mtn., E. Rupununi Sava.	02.86667	-59.26667	KT221318.1	KT221188.1	KT221252.1
USNM:Bird:B12872	*S. cucullatus*	~150	Guyana: Wiwitau Mtn., E. Rupununi Sava.	02.86667	-59.26667	*Pending*	*Pending*	*Pending*
UWBM:Bird:54421	*S. magellanicus*	300	Argentina: Tucumán: San Miguel Tucumán	-27.04800	-65.66500	---	---	KT221276.1
UWBM:Bird:70271	*S. magellanicus*	209	Argentina: Provincia de Misiones: Posadas	-26.95520	-55.08770	KT221343.1	KT221213.1	KT221277.1
UWBM:Bird:70272	*S. magellanicus*	209	Argentina: Provincia de Misiones: Posadas	-26.95520	-55.08770	KT221344.1	KT221214.1	KT221278.1
UWBM:Bird:70716	*S. magellanicus*	300	Argentina: Tucumán: San Miguel Tucumán	-27.02200	-65.64500	KT221345.1	KT221215.1	KT221279.1
UWBM:Bird:77454	*S. magellanicus*	945	Bolivia: Santa Cruz: Gutiérrez	-19.40000	-63.45000	---	---	KT221280.1
FMNH:Bird:334722	*S. magellanicus*	2470	Bolivia: Cochabamba: Cochabamba-Oruro	-17.50394	-66.32923	KT221348.1	KT221218.1	KT221282.1

Appendix B cont.

Individual	Species	Elev.	Locality	Latitude	Longitude	*cytb*	ND2	ND3
FMNH:Bird:334723	*S. magellanicus*	2470	Bolivia: Cochabamba: Cochabamba-Oruro	-17.50394	-66.32923	---	---	KT221283.1
FMNH:Bird:396033	*S. magellanicus*	2470	Bolivia: Cochabamba: Cochabamba-Oruro	-17.50394	-66.32923	---	---	KT221284.1
ANSP:Bird:18817	*S. magellanicus*	3035	Ecuador: Carchi: SE Impueran	00.45000	-77.86667	---	---	*Pending*
ANSP:Bird:18864	*S. magellanicus*	3035	Ecuador: Carchi: SE Impueran	00.45000	-77.86667	---	---	*Pending*
ANSP:Bird:18865	*S. magellanicus*	3035	Ecuador: Carchi: SE Impueran	00.45000	-77.86667	---	---	*Pending*
MSB:Bird:34232	*S. magellanicus*	<100	Peru: Ancash: Huarmey	-10.06812	-78.13603	KT221335.1	KT221205.1	KT221269.1
MSB:Bird:34249	*S. magellanicus*	<100	Peru: Ancash: Huarmey	-10.06812	-78.13603	*Pending*	*Pending*	*Pending*
MSB:Bird:34250	*S. magellanicus*	<100	Peru: Ancash: Huarmey	-10.06812	-78.13603	*Pending*	*Pending*	*Pending*
MSB:Bird:34251	*S. magellanicus*	<100	Peru: Ancash: Huarmey	-10.06812	-78.13603	KT221336.1	KT221206.1	KT221270.1
MSB:Bird:34255	*S. magellanicus*	<100	Peru: Ancash: Huarmey	-10.06812	-78.13603	*Pending*	*Pending*	*Pending*
MSB:Bird:34268	*S. magellanicus*	<100	Peru: Ancash: Huarmey	-10.06812	-78.13603	KT221337.1	KT221207.1	KT221271.1
MSB:Bird:34828	*S. magellanicus*	2972	Peru: Ancash: Santa: Macate	-08.75478	-78.04845	---	---	KT221272.1
MSB:Bird:34867	*S. magellanicus*	2972	Peru: Ancash: Santa: Macate	-08.75478	-78.04845	---	---	KT221273.1
MSB:Bird:34875	*S. magellanicus*	2972	Peru: Ancash: Santa: Macate	-08.75478	-78.04845	KT221340.1	KT221210.1	KT221274.1
MSB:Bird:34876	*S. magellanicus*	2972	Peru: Ancash: Santa: Macate	-08.75478	-78.04845	---	---	*Pending*
MSB:Bird:34877	*S. magellanicus*	2972	Peru: Ancash: Santa: Macate	-08.75478	-78.04845	---	---	*Pending*
MSB:Bird:34878	*S. magellanicus*	2972	Peru: Ancash: Santa: Macate	-08.75478	-78.04845	*Pending*	*Pending*	*Pending*
MSB:Bird:34879	*S. magellanicus*	2972	Peru: Ancash: Santa: Macate	-08.75478	-78.04845	---	---	*Pending*
MSB:Bird:34880	*S. magellanicus*	2972	Peru: Ancash: Santa: Macate	-08.75478	-78.04845	*Pending*	*Pending*	*Pending*
MSB:Bird:34881	*S. magellanicus*	2972	Peru: Ancash: Santa: Macate	-08.75478	-78.04845	---	---	*Pending*
MSB:Bird:34882	*S. magellanicus*	2972	Peru: Ancash: Santa: Macate	-08.75478	-78.04845	---	---	*Pending*
MSB:Bird:34999	*S. magellanicus*	3714	Peru: Ancash: Caraz: Pueblo Libre: Caraz	-09.10097	-77.86556	---	---	*Pending*
MSB:Bird:35015	*S. magellanicus*	3714	Peru: Ancash: Caraz: Pueblo Libre: Caraz	-09.10097	-77.86556	---	---	*Pending*
MSB:Bird:35694	*S. magellanicus*	4073	Peru: Ancash			---	---	*Pending*
MSB:Bird:35695	*S. magellanicus*	4073	Peru: Ancash			---	---	*Pending*
MSB:Bird:36033	*S. magellanicus*	3775	Peru: Ancash: Santa: Macate	-08.74247	-78.04070	---	---	*Pending*
MSB:Bird:36034	*S. magellanicus*	3775	Peru: Ancash: Santa: Macate	-08.74247	-78.04070	---	---	*Pending*
MSB:Bird:36094	*S. magellanicus*	3350	Peru: Ancash: Huaylas: Santo Toribio	-08.83621	-77.92580	---	---	*Pending*
MSB:Bird:33653	*S. magellanicus*	3600	Peru: Apurimac: Aymaraes: Caraybamba	-14.40760	-73.08772	*Pending*	*Pending*	*Pending*
MSB:Bird:35537	*S. magellanicus*	3200	Peru: Arequipa: Condesuyos: Chiquibamba	-15.81433	-72.66712	---	---	*Pending*
MSB:Bird:35561	*S. magellanicus*	3200	Peru: Arequipa: Condesuyos: Chiquibamba	-15.81433	-72.66712	---	---	*Pending*
MSB:Bird:35572	*S. magellanicus*	3200	Peru: Arequipa: Condesuyos: Chiquibamba	-15.81433	-72.66712	---	---	*Pending*
MSB:Bird:35238	*S. magellanicus*	2500	Peru: Cajamarca: Contumaza: Contumaza	-07.39803	-78.77827	---	---	*Pending*
MSB:Bird:35258	*S. magellanicus*	2500	Peru: Cajamarca: Contumaza: Contumaza	-07.39803	-78.77827	*Pending*	*Pending*	*Pending*
MSB:Bird:35282	*S. magellanicus*	2500	Peru: Cajamarca: Contumaza: Contumaza	-07.39803	-78.77827	---	---	*Pending*
MSB:Bird:35288	*S. magellanicus*	2500	Peru: Cajamarca: Contumaza: Contumaza	-07.39803	-78.77827	*Pending*	*Pending*	*Pending*
MSB:Bird:35310	*S. magellanicus*	2550	Peru: Cajamarca: Contumaza: Contumaza	-07.40383	-78.77978	---	---	*Pending*

Appendix B cont.

Individual	Species	Elev.	Locality	Latitude	Longitude	*cytb*	ND2	ND3
MSB:Bird:35311	*S. magellanicus*	2550	Peru: Cajamarca: Contumaza: Contumaza	-07.40383	-78.77978	---	---	*Pending*
MSB:Bird:35316	*S. magellanicus*	2550	Peru: Cajamarca: Contumaza: Contumaza	-07.40383	-78.77978	---	---	*Pending*
MSB:Bird:35317	*S. magellanicus*	2550	Peru: Cajamarca: Contumaza: Contumaza	-07.40383	-78.77978	---	---	*Pending*
MSB:Bird:35326	*S. magellanicus*	2550	Peru: Cajamarca: Contumaza: Contumaza:	-07.40383	-78.77978	---	---	*Pending*
MSB:Bird:35339	*S. magellanicus*	2550	Peru: Cajamarca: Contumaza: Contumaza	-07.40383	-78.77978	---	---	*Pending*
MSB:Bird:35340	*S. magellanicus*	2550	Peru: Cajamarca: Contumaza: Contumaza	-07.40383	-78.77978	---	---	*Pending*
MSB:Bird:35345	*S. magellanicus*	2550	Peru: Cajamarca: Contumaza: Contumaza	-07.40383	-78.77978	*Pending*	*Pending*	*Pending*
MSB:Bird:35348	*S. magellanicus*	2550	Peru: Cajamarca: Contumaza: Contumaza	-07.40383	-78.77978	---	---	*Pending*
MSB:Bird:35400	*S. magellanicus*	2550	Peru: Cajamarca: Contumaza: Contumaza	-07.40383	-78.77603	---	---	*Pending*
MSB:Bird:27068	*S. magellanicus*	3120	Peru: Cusco: Quispicanchis: Huacarpay	-13.62533	-71.71800	*Pending*	*Pending*	*Pending*
MSB:Bird:27175	*S. magellanicus*	3500	Peru: Cusco: Urubamba: Urubamba	-13.24933	-72.16917	KT221323.1	KT221193.1	KT221257.1
MSB:Bird:27176	*S. magellanicus*	3500	Peru: Cusco: Urubamba: Urubamba	-13.24933	-72.16917	KT221324.1	KT221194.1	KT221258.1
MSB:Bird:27201	*S. magellanicus*	3380	Peru: Cusco: Urubamba: Urubamba	-13.24933	-72.16917	---	---	*Pending*
MSB:Bird:33112	*S. magellanicus*	3835	Peru: Cusco: Ollantaytambo: Choquechaca	-13.18808	-72.23137	KT221330.1	KT221200.1	KT221264.1
MSB:Bird:33140	*S. magellanicus*	3018	Peru: Cusco: Ollantaytambo: Choquechaca	-13.18808	-72.23137	KT221331.1	KT221201.1	KT221265.1
MSB:Bird:31065	*S. magellanicus*	3630	Peru: Huancavelica: 6km S Castrovirreyna	-13.33633	-75.32483	KT221341.1	KT221211.1	KT221275.1
MSB:Bird:33292	*S. magellanicus*	3932	Peru: Lima: San Pedro de Casta: Potago	-11.76418	-76.54300	*Pending*	*Pending*	*Pending*
MSB:Bird:33349	*S. magellanicus*	3905	Peru: Lima: San Pedro de Casta: Potago	-11.76770	-76.53462	*Pending*	*Pending*	*Pending*
MSB:Bird:33369	*S. magellanicus*	3905	Peru: Lima: San Pedro de Casta: Potago	-11.76770	-76.53462	---	---	KT221266.1
MSB:Bird:33383	*S. magellanicus*	3905	Peru: Lima: San Pedro de Casta: Potago	-11.76770	-76.53462	---	---	*Pending*
MSB:Bird:33460	*S. magellanicus*	4131	Peru: Lima: San Pedro de Casta: Potago	-11.76938	-76.53278	*Pending*	*Pending*	*Pending*
MSB:Bird:33489	*S. magellanicus*	4140	Peru: Lima: San Pedro de Casta: Potago	-11.76958	-76.53193	*Pending*	*Pending*	*Pending*
MSB:Bird:28282	*S. magellanicus*	3750	Peru: Lima: San Pedro de Casta: Carhuay.	-11.76188	-76.54887	KT221325.1	KT221195.1	KT221259.1
MSB:Bird:28301	*S. magellanicus*	3750	Peru: Lima: San Pedro de Casta: Carhuay.	-11.76188	-76.54887	*Pending*	*Pending*	*Pending*
MSB:Bird:28342	*S. magellanicus*	3750	Peru: Lima: San Pedro de Casta: Carhuay.	-11.76188	-76.54887	KT221326.1	KT221196.1	KT221260.1
MSB:Bird:31431	*S. magellanicus*	372	Peru: Lima: : San Juan de Lurigancho	-12.00148	-76.92082	*Pending*	*Pending*	*Pending*
MSB:Bird:31439	*S. magellanicus*	372	Peru: Lima: San Juan de Lurigancho	-12.00148	-76.92082	*Pending*	*Pending*	*Pending*
MSB:Bird:31497	*S. magellanicus*	3945	Peru: Lima: Huarochiri: Carampoma	-11.62683	-76.43457	KT221327.1	KT221197.1	KT221261.1
MSB:Bird:28373	*S. magellanicus*	3800	Peru: Lima: San Pedro de Casta: Chinchay.	-11.76077	-76.56767	KT221328.1	KT221198.1	KT221262.1
MSB:Bird:28466	*S. magellanicus*	2550	Peru: Lima: San Pedro de Casta: Pariangan.	-11.74283	-76.60617	*Pending*	*Pending*	*Pending*
MSB:Bird:28470	*S. magellanicus*	2550	Peru: Lima: San Pedro de Casta: Pariangan.	-11.74283	-76.60617	*Pending*	*Pending*	*Pending*
MSB:Bird:28473	*S. magellanicus*	2550	Peru: Lima: San Pedro de Casta: Pariangan.	-11.74283	-76.60617	*Pending*	*Pending*	*Pending*
MSB:Bird:31704	*S. magellanicus*	3750	Peru: Lima: San Pedro de Casta: Carhuay.	-11.76188	-76.54887	*Pending*	*Pending*	*Pending*
MSB:Bird:32903	*S. magellanicus*	935	Peru: Lima: 2.3km E Nieve Nieve	-12.02950	-76.65078	*Pending*	*Pending*	*Pending*
MSB:Bird:32907	*S. magellanicus*	935	Peru: Lima: 2.3km E Nieve Nieve	-12.02950	-76.65078	*Pending*	*Pending*	*Pending*
MSB:Bird:32938	*S. magellanicus*	935	Peru: Lima: 2.3km E Nieve Nieve	-12.02950	-76.65078	*Pending*	*Pending*	*Pending*
MSB:Bird:33290	*S. magellanicus*	3980	Peru: Lima: San Pedro de Casta: Potago	-11.76360	-76.54250	*Pending*	*Pending*	*Pending*

Appendix B cont.

Individual	Species	Elev.	Locality	Latitude	Longitude	*cytb*	ND2	ND3
MSB:Bird:33290	*S. magellanicus*	3980	Peru: Lima: San Pedro de Casta: Potago	-11.76360	-76.54250	*Pending*	*Pending*	*Pending*
MSB:Bird:33408	*S. magellanicus*	3905	Peru: Lima: San Pedro de Casta: Potago	-11.76770	-76.53462	---	---	*Pending*
MSB:Bird:33461	*S. magellanicus*	4140	Peru: Lima: San Pedro de Casta: Potago	-11.76958	-76.53193	*Pending*	*Pending*	*Pending*
MSB:Bird:33464	*S. magellanicus*	4082	Peru: Lima: San Pedro de Casta: Potago	-11.76847	-76.53325	KT221333.1	KT221203.1	KT221267.1
MSB:Bird:33488	*S. magellanicus*	4140	Peru: Lima: San Pedro de Casta: Potago	-11.76958	-76.53193	*Pending*	*Pending*	*Pending*
MSB:Bird:35944	*S. magellanicus*	3243	Peru: Lima: Huarochiri: Carhuayumac	-11.76188	-76.54887	---	---	*Pending*
MSB:Bird:35960	*S. magellanicus*	3177	Peru: Lima: Huarochiri: San Pedro de Casta	-11.75720	-76.59718	---	---	*Pending*
MSB:Bird:36390	*S. magellanicus*	214	Peru: Lima: Pachacamac: Lurin Valley	-12.15305	-76.83622	---	---	*Pending*
MSB:Bird:36423	*S. magellanicus*	121	Peru: Lima: Lima, Comas	-11.94425	-77.06345	---	---	*Pending*
MSB:Bird:36577	*S. magellanicus*	4153	Peru: Lima: San Pedro de Casta: Potago	-11.75720	-76.59718	---	---	*Pending*
FMNH:324099	*S. magellanicus*	500	Peru: Madre de Dios: Hacienda Amazonia	-12.68442	-71.24497	*KT221347.1*	*KT221217.1*	*KT221281.1*
FMNH: 398593	*S. magellanicus*	243	Peru: Madre de Dios: Quebrada Aguas Cali.	-12.66690	-71.27060	---	---	*Pending*
FMNH: 398594	*S. magellanicus*	243	Peru: Madre de Dios: Quebrada Aguas Cali.	-12.66690	-71.27060	---	---	*Pending*
MSB:Bird:34073	*S. magellanicus*	2100	Peru: Piura: Huancabamba: Abra Porculla	-05.84000	-79.50533	*KT221334.1*	*KT221204.1*	*KT221268.1*
MSB:Bird:35095	*S. magellanicus*	3728	Peru: Tacna: Candarave: San Pedro	-17.22700	-70.26208	---	---	*Pending*
MSB:Bird:35100	*S. magellanicus*	3728	Peru: Tacna: Candarave: San Pedro	-17.22700	-70.26208	---	---	*Pending*
MSB:Bird:35439	*S. magellanicus*	2200	Peru: Tacna: Candarave: Curibaya	-17.39005	-70.34580	---	---	*Pending*
MSB:Bird:35440	*S. magellanicus*	2200	Peru: Tacna: Candarave: Curibaya	-17.39005	-70.34580	---	---	*Pending*
MSB:Bird:35490	*S. magellanicus*	2975	Peru: Tacna: Candarave: Quilahuani	-17.32088	-70.24773	---	---	*Pending*
MSB:Bird:35500	*S. magellanicus*	2975	Peru: Tacna: Candarave: Quilahuani	-17.32088	-70.24773	---	---	*Pending*
MSB:Bird:35506	*S. magellanicus*	2975	Peru: Tacna: Candarave: Quilahuani	-17.32088	-70.24773	---	---	*Pending*
MSB:Bird:33471	*S. uropygialis*	4131	Peru: Lima: San Pedro de Casta: Potago	-11.76938	-76.53278	*KT221364.1*	*KT221234.1*	*KT221298.1*
MSB:Bird:28329	*S. uropygialis*	3750	Peru: Lima: San Pedro de Casta: Carhuay.	-11.76188	-76.54887	*KT221363.1*	*KT221233.1*	*KT221297.1*
MSB:Bird:31708	*S. uropygialis*	3808	Peru: Lima: San Pedro de Casta: Carhuay.	-11.76373	-76.54882	*Pending*	*Pending*	*Pending*
MSB:Bird:28263	*S. uropygialis*	3750	Peru: Lima: San Pedro de Casta: Carhuay.	-11.76188	-76.54887	*Pending*	*Pending*	*Pending*
MSB:Bird:28278	*S. uropygialis*	3750	Peru: Lima: San Pedro de Casta: Carhuay.	-11.76188	-76.54887	*Pending*	*Pending*	*Pending*
MSB:Bird:28295	*S. uropygialis*	3750	Peru: Lima: San Pedro de Casta: Carhuay.	-11.76188	-76.54887	*Pending*	*Pending*	*Pending*
MSB:Bird:28296	*S. uropygialis*	3750	Peru: Lima: San Pedro de Casta: Carhuay.	-11.76188	-76.54887	*Pending*	*Pending*	*Pending*
MSB:Bird:28311	*S. uropygialis*	3750	Peru: Lima: San Pedro de Casta: Carhuay.	-11.76188	-76.54887	*Pending*	*Pending*	*Pending*
MSB:Bird:28312	*S. uropygialis*	3750	Peru: Lima: San Pedro de Casta: Carhuay.	-11.76188	-76.54887	*KT221361.1*	*KT221231.1*	*KT221295.1*
MSB:Bird:28326	*S. uropygialis*	3750	Peru: Lima: San Pedro de Casta: Carhuay.	-11.76188	-76.54887	*KT221362.1*	*KT221232.1*	*KT221296.1*

Appendix B cont.

Individual	Species	Elev.	Locality	Latitude	Longitude	*cytb*	ND2	ND3
MSB:Bird:31726	S. uropygialis	3808	Peru: Lima: San Pedro de Casta: Carhuay.	-11.76373	-76.54882	*Pending*	Pending	Pending
MSB:Bird:31727	S. uropygialis	3808	Peru: Lima: San Pedro de Casta: Carhuay.	-11.76373	-76.54882	*Pending*	Pending	Pending
MSB:Bird:31731	S. uropygialis	3777	Peru: Lima: San Pedro de Casta: Carhuay.	-11.76300	-76.54885	*Pending*	Pending	Pending
MSB:Bird:31746	S. uropygialis	3808	Peru: Lima: San Pedro de Casta: Carhuay.	-11.76373	-76.54882	*Pending*	Pending	Pending
MSB:Bird:31752	S. uropygialis	3808	Peru: Lima: San Pedro de Casta: Carhuay.	-11.76373	-76.54882	*Pending*	Pending	Pending
MSB:Bird:33288	S. uropygialis	3980	Peru: Lima: San Pedro de Casta: Potago	-11.76360	-76.54250	*Pending*	Pending	Pending
MSB:Bird:33331	S. uropygialis	4056	Peru: Lima: San Pedro de Casta: Potago	-11.76770	-76.53462	*Pending*	Pending	Pending
MSB:Bird:33359	S. uropygialis	3905	Peru: Lima: San Pedro de Casta: Potago	-11.76770	-76.53462	*Pending*	Pending	Pending
MSB:Bird:33394	S. uropygialis	3940	Peru: Lima: San Pedro de Casta: Potago	-11.76770	-76.53462	*Pending*	Pending	Pending
MSB:Bird:33453	S. uropygialis	4131	Peru: Lima: San Pedro de Casta: Potago	-11.76938	-76.53278	*Pending*	Pending	Pending

Appendix C. Individuals sampled for target-capture sequencing in Chapter 3. Elevation (Elev.) is reported in meters. The museum abbreviations are MSB: Museum of Southwestern Biology; KU: Biodiversity Institute and Natural History Museum, University of Kansas; USNM: Smithsonian Institution; UWBM: Burke Museum; FMNH: Field Museum of Natural History; LSUMNS: Louisiana State University Museum of Natural Science.

Museum	Tissue/NK	Catalog	Taxon	Locality	Elev.	Latitude	Longitude
MSB	169325	34099	*Spinus atratus*	PERU: Apurimac: Abancay: Chacoche: Laguna Anantay	4300	-14.0600	-73.0016
KU	11773	98352	*S. barbatus*	ARGENTINA: Rio Negro: NW El Bolson, Cerro Perito Moreno	1100	-41.7900	-71.5633
MSB	168311	33092	*S. crassirostris*	PERU: Cusco: Ollantaytambo: Choquechaca	4200	-13.1880	-72.2313
MSB	172338	35612	*S. crassirostris*	PERU: Tacna: Candarave: San Pedro	3728	-17.2270	-70.2620
USNM	B12865		*S. cucullatus*	GUYANA: Wiwitau Mountain, East Rupununi Savannah	150	2.8666	-59.2666
UWBM	70271	70271	*S. magellanicus*	ARGENTINA: Provincia de Misiones: Posadas	209	-26.9552	-55.0877
FMNH	334722		*S. magellanicus*	BOLIVIA: Cochabamba: Cochabamba-Oruro Rd, km 29	2470	-17.5039	-66.3292
LSUMNS	B-18761		*S. magellanicus*	BOLIVIA: Santa Cruz: Cordillera: 130 km E Charagua.	430	-19.7928	-61.9678
MSB	169458	34232	*S. magellanicus*	PERU: Ancash: Huarmey: Huarmey	39	-10.0681	-78.1360
MSB	169475	34249	*S. magellanicus*	PERU: Ancash: Huarmey: Huarmey	39	-10.0681	-78.1360
MSB	169476	34250	*S. magellanicus*	PERU: Ancash: Huarmey: Huarmey	39	-10.0681	-78.1360
MSB	169477	34251	*S. magellanicus*	PERU: Ancash: Huarmey: Huarmey	39	-10.0681	-78.1360
MSB	169481	34255	*S. magellanicus*	PERU: Ancash: Huarmey: Huarmey	39	-10.0681	-78.1360
MSB	169494	34268	*S. magellanicus*	PERU: Ancash: Huarmey: Huarmey	39	-10.0681	-78.1360
MSB	220665	43359	*S. magellanicus*	PERU: Ancash: Huarmey: Huarmey	116	-10.0237	-78.0500
MSB	220664	43358	*S. magellanicus*	PERU: Ancash: Huarmey: Huarmey	116	-10.0237	-78.0500
MSB	220633	43327	*S. magellanicus*	PERU: Ancash: Huarmey: Malbas: Yanaparin	662	-9.9610	-77.8147
MSB	220634	43328	*S. magellanicus*	PERU: Ancash: Huarmey: Malbas: Yanaparin	662	-9.9610	-77.8147
MSB	220620	43314	*S. magellanicus*	PERU: Ancash: Huarmey: Malbas: Yanaparin	672	-9.9607	-77.8142
MSB	220621	43315	*S. magellanicus*	PERU: Ancash: Huarmey: Malbas: Yanaparin	672	-9.9607	-77.8142
MSB	220616	43310	*S. magellanicus*	PERU: Ancash: Huarmey: Malbas: Yanaparin	676	-9.9625	-77.8139
MSB	220631	43325	*S. magellanicus*	PERU: Ancash: Huarmey: Malbas: Yanaparin	723	-9.8586	-77.8073
MSB	220655	43349	*S. magellanicus*	PERU: Ancash: Aija: Coris: San Damian	1430	-9.8899	-77.7744
MSB	220561	43255	*S. magellanicus*	PERU: Ancash: Aija: Coris: Irman	1566	-9.8750	-77.7656
MSB	220643	43337	*S. magellanicus*	PERU: Ancash: Aija: Coris: San Damian	1578	-9.8676	-77.7566
MSB	220654	43348	*S. magellanicus*	PERU: Ancash: Aija: Coris: Irman	1726	-9.8593	-77.7457
MSB	220653	43347	*S. magellanicus*	PERU: Ancash: Aija: Coris: Irman	1726	-9.8593	-77.7457
MSB	220596	43290	*S. magellanicus*	PERU: Ancash: Aija: Coris: Irman	1804	-9.8521	-77.7426
MSB	220593	43287	*S. magellanicus*	PERU: Ancash: Aija: Coris: Irman	2120	-9.8386	-77.7200
MSB	220608	43302	*S. magellanicus*	PERU: Ancash: Aija: Coris: Irman	2121	-9.8521	-77.7374
MSB	220595	43289	*S. magellanicus*	PERU: Ancash: Aija: Coris: Irman	2300	-9.8281	-77.7013

Appendix C cont.

Museum	Tissue/NK	Catalog	Taxon	Locality	Elev.	Latitude	Longitude
MSB	220538	43232	*S. magellanicus*	PERU: Ancash: Aija: Huacllan: Huacllan	2930	-9.8073	-77.6837
MSB	220537	43231	*S. magellanicus*	PERU: Ancash: Aija: Huacllan: Huacllan	2940	-9.8071	-77.6838
MSB	220541	43235	*S. magellanicus*	PERU: Ancash: Aija: Huacllan: Huacllan	3023	-9.8054	-77.6826
MSB	220542	43236	*S. magellanicus*	PERU: Ancash: Aija: Huacllan: Huacllan	3023	-9.8054	-77.6826
MSB	220544	43238	*S. magellanicus*	PERU: Ancash: Aija: Huacllan: Huacllan	3023	-9.8054	-77.6826
MSB	220545	43239	*S. magellanicus*	PERU: Ancash: Aija: Huacllan: Huacllan	3023	-9.8054	-77.6826
MSB	220539	43233	*S. magellanicus*	PERU: Ancash: Aija: Huacllan: Huacllan	3052	-9.8059	-77.6834
MSB	220540	43234	*S. magellanicus*	PERU: Ancash: Aija: Huacllan: Huacllan	3052	-9.8055	-77.6834
MSB	220516	43210	*S. magellanicus*	PERU: Ancash: Aija: La Merced: Ullucuran	3831	-9.7370	-77.6525
MSB	220518	43212	*S. magellanicus*	PERU: Ancash: Aija: La Merced: Ullucuran	3831	-9.7370	-77.6525
MSB	220519	43213	*S. magellanicus*	PERU: Ancash: Aija: La Merced: Ullucuran	3831	-9.7370	-77.6525
MSB	220520	43214	*S. magellanicus*	PERU: Ancash: Aija: La Merced: Ullucuran	3831	-9.7370	-77.6525
MSB	220521	43215	*S. magellanicus*	PERU: Ancash: Aija: La Merced: Ullucuran	3831	-9.7370	-77.6525
MSB	220522	43216	*S. magellanicus*	PERU: Ancash: Aija: La Merced: Ullucuran	3831	-9.7370	-77.6525
MSB	220529	43223	*S. magellanicus*	PERU: Ancash: Aija: La Merced: Ullucuran	3832	-9.7209	-77.6727
MSB	220530	43224	*S. magellanicus*	PERU: Ancash: Aija: La Merced: Ullucuran	3832	-9.7209	-77.6727
MSB	175254	36423	*S. magellanicus*	PERU: Lima: Lima: Comas	121	-11.9442	-77.0634
MSB	175221	36390	*S. magellanicus*	PERU: Lima: Lima: Pachacamac: Lurin Valley	214	-12.1530	-76.8362
MSB	162993	31429	*S. magellanicus*	PERU: Lima: San Juan de Lurigancho: Huachipa	372	-12.0014	-76.9208
MSB	162995	31431	*S. magellanicus*	PERU: Lima: San Juan de Lurigancho: Huachipa	372	-12.0014	-76.9208
MSB	163003	31439	*S. magellanicus*	PERU: Lima: San Juan de Lurigancho: Huachipa	372	-12.0014	-76.9208
MSB	168075	32903	*S. magellanicus*	PERU: Lima: Huarochiri: Antioquia: Nieve Nieve	935	-12.0295	-76.6507
MSB	168079	32907	*S. magellanicus*	PERU: Lima: Huarochiri: Antioquia: Nieve Nieve	935	-12.0295	-76.6507
MSB	168109	32937	*S. magellanicus*	PERU: Lima: Huarochiri: Antioquia: Nieve Nieve	935	-12.0295	-76.6507
MSB	168110	32938	*S. magellanicus*	PERU: Lima: Huarochiri: Antioquia: Nieve Nieve	935	-12.0295	-76.6507
MSB	221240		*S. magellanicus*	PERU: Lima: Huarochiri: San Pedro de Casta: Vequil	1600	-11.7989	-76.6239
MSB	221241		*S. magellanicus*	PERU: Lima: Huarochiri: San Pedro de Casta: Vequil	1600	-11.7989	-76.6239
MSB	221214		*S. magellanicus*	PERU: Lima: Huarochiri: San Pedro de Casta: Vequil	1650	-11.7989	-76.6239
MSB	221213		*S. magellanicus*	PERU: Lima: Huarochiri: San Pedro de Casta: Vequil	1657	-11.8017	-76.6236
MSB	221238		*S. magellanicus*	PERU: Lima: Huarochiri: San Pedro de Casta: Vequil	1657	-11.7988	-76.6239
MSB	221239		*S. magellanicus*	PERU: Lima: Huarochiri: San Pedro de Casta: Vequil	1657	-11.7988	-76.6239
MSB	221207		*S. magellanicus*	PERU: Lima: Huarochiri: San Pedro de Casta: Vequil	1666	-11.8169	-76.6230
MSB	163356	28466	*S. magellanicus*	PERU: Lima: Huarochiri: San Pedro de Casta: Pariangancha	2550	-11.7428	-76.6061

Appendix C cont.

Museum	Tissue/NK	Catalog	Taxon	Locality	Elev.	Latitude	Longitude
MSB	163360	28470	*S. magellanicus*	PERU: Lima: Huarochiri: San Pedro de Casta: Pariangancha	2550	-11.7428	-76.6061
MSB	163363	28473	*S. magellanicus*	PERU: Lima: Huarochiri: San Pedro de Casta: Pariangancha	2550	-11.7428	-76.6061
MSB	221228		*S. magellanicus*	PERU: Lima: Huarochiri: San Pedro de Casta: San Pedro de Casta	3166	-11.7669	-76.5967
MSB	172798	35960	*S. magellanicus*	PERU: Lima: Huarochiri: San Pedro de Casta: San Pedro de Casta	3177	-11.7572	-76.5971
MSB	221225		*S. magellanicus*	PERU: Lima: Huarochiri: San Pedro de Casta: San Pedro de Casta	3226	-11.7632	-76.5961
MSB	221231		*S. magellanicus*	PERU: Lima: Huarochiri: San Pedro de Casta: San Pedro de Casta	3230	-11.7635	-76.5960
MSB	221232		*S. magellanicus*	PERU: Lima: Huarochiri: San Pedro de Casta: San Pedro de Casta	3230	-11.7635	-76.5960
MSB	221233		*S. magellanicus*	PERU: Lima: Huarochiri: San Pedro de Casta: San Pedro de Casta	3230	-11.7635	-76.5960
MSB	221234		*S. magellanicus*	PERU: Lima: Huarochiri: San Pedro de Casta: San Pedro de Casta	3244	-11.7657	-76.5953
MSB	221235		*S. magellanicus*	PERU: Lima: Huarochiri: San Pedro de Casta: San Pedro de Casta	3244	-11.7657	-76.5953
MSB	162894	28301	*S. magellanicus*	PERU: Lima: Huarochiri: San Pedro de Casta: Carhuayumac	3750	-11.7618	-76.5488
MSB	163376	31704	*S. magellanicus*	PERU: Lima: Huarochiri: San Pedro de Casta: Carhuayumac	3750	-11.7618	-76.5488
MSB	163413	31741	*S. magellanicus*	PERU: Lima: Huarochiri: San Pedro de Casta: Carhuayumac	3840	-11.7635	-76.5488
MSB	168588	33369	*S. magellanicus*	PERU: Lima: Huarochiri: San Pedro de Casta: Potago	3905	-11.7677	-76.5346
MSB	168627	33408	*S. magellanicus*	PERU: Lima: Huarochiri: San Pedro de Casta: Potago	3905	-11.7677	-76.5346
MSB	163061	31497	*S. magellanicus*	PERU: Lima: Huarochiri: Carampoma	3945	-11.6268	-76.4345
MSB	168683	33464	*S. magellanicus*	PERU: Lima: Huarochiri: San Pedro de Casta: Potago	4082	-11.7684	-76.5332
MSB	168680	33461	*S. magellanicus*	PERU: Lima: Huarochiri: San Pedro de Casta: Potago	4140	-11.7695	-76.5319
MSB	168707	33488	*S. magellanicus*	PERU: Lima: Huarochiri: San Pedro de Casta: Potago	4140	-11.7695	-76.5319
MSB	175529	36577	*S. magellanicus*	PERU: Lima: Huarochiri: San Pedro de Casta	4153	-11.7572	-76.5971
UWBM	82845		*S. notatus*	MEXICO: Guerrero: Chilpancingo	2200	17.5700	-99.6400
MSB	176738	41706	*S. olivaceus*	PERU: Amazonas: Santa Maria de Nieve: Imaza	1503	-5.2130	-78.6435
LSUMNS	B-66764		*S. siemiradzkii*	PERU: Tumbes: Tumbes	25	-3.5650	-80.4445
MSB	168578	33359	*S. uropygialis*	PERU: Lima: Huarochiri: San Pedro de Casta: Potago	3905	-11.7677	-76.5346

REFERENCES CITED:

Abbott, R., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J. E., Bierne, N., … Zinner, D. (2013). Hybridization and speciation. *Journal of Evolutionary Biology*, *26*, 229–246. http://doi.org/10.1111/j.1420-9101.2012.02599.x

Alev, C., Shinmyozu, K., McIntyre, B. A. S., & Sheng, G. (2009). Genomic organization of zebra finch alpha and beta globin genes and their expression in primitive and definitive blood in comparison with globins in chicken. *Development Genes and Evolution*, *219*, 353–360.

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast Model-Based Estimation of Ancestry in Unrelated Individuals. *Genome Research*, *19*, 1655–1664. http://doi.org/10.1101/gr.094052.109.vidual

Allen, J. M., Huang, D. I., Cronk, Q. C., & Johnson, K. P. (2015). aTRAM - automated target restricted assembly method: a fast method for assembling loci across divergent taxa from next-generation sequencing data. *BMC Bioinformatics*, *16*, 98. http://doi.org/10.1186/s12859-015-0515-2

Arnaiz-Villena, A., Álvarez-Tejado, M., Ruíz-del-Valle, V., García-de-la-Torre, C., Varela, P., Recio, M.J., Ferre, S., Martínez-Laso, J., (1998). Phylogeny and rapid Northern and Southern Hemisphere speciation of goldfinches during the Miocene and Pliocene Epochs. *Cell. Mol. Life Sci. 54,* 1031-1041.

Arnaiz-Villena, A., Ruiz-del-Valle, V., Moscoso, J., Serrano-Vela, J.I., Zamora, J., (2007). mtDNA phylogeny of North American *Carduelis pinus* group. *Ardeola 54*, 1-14.

Arnaiz-Villena, A., Ruiz-del-Valle, V., Moscoso, J., Serrano-Vela, I., Zamora, J., (2008). Mitochondrial DNA phylogeneic defintion of a group of 'arid-zone' Carduelini finches. *Open Ornithol. J. 1*, 1-7.

Ballard, J. W. O., & Whitlock, M. C. (2004). The incomplete natural history of mitochondria. *Molecular Ecology*, *13*, 729–744. http://doi.org/10.1046/j.1365-294X.2003.02063.x

Barton, N. H. (1983). Multilocus Clines. *Evolution*, *37*, 454. http://doi.org/10.2307/2408260

Batalha-Filho, H., Irestedt, M., Fjeldså, J., Ericson, P.G.P., Silveira, L.F., Miyaki, C.Y., (2013). Molecular systematics and evolution of the *Synallaxis ruficapilla* complex (Aves: Furnariidae) in the Atlantic Forest. *Mol. Phylogenet. Evol. 67*, 86-94.

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. http://doi.org/10.1016/j.ympev.2015.03.005

Benham, P. M., A. M. Cuervo, J. A. McGuire, & C. C. Witt. (2014). Biogeography of the Andean metaltail hummingbirds: contrasting evolutionary histories of treeline and habitat-generalist clades. *Journal of Biogeography. 42,* 763-777. http://doi.org/10.1111/jbi.12452

Blackmon, H. (2016). R package: evobiR.

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.
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. http://doi.org/10.1186/1471-2148-11-355
Bouckaert, R., & Heled, J. (2014). DensiTree 2: Seeing Trees Through the Forest. *bioRxiv*, 012401. http://doi.org/10.1101/012401
Bowie, R.C.K., Fjeldså, J., Hackett, S.J., Bates, J.M., Crowe, T.M., (2006). Coalescent models reveal the relative roles of ancestral polymorphism, vicariance and dispersal in shaping phylogeographic structure of an African montane forest robin. *Mol. Phylogenet. Evol. 38*, 171-188.
Brumfield, R.T. and Capparella, A. P. (1996). Historical diversification of birds in northwestern South America: a molecular perspective on the role of vicariant events. *Evolution. 50*, 1607-1624.
Brumfield, R.T., Edwards, S.V., (2007). Evolution into and out of the Andes: A Bayesian analysis of historical diversification in *Thamnophilus* antshrikes. *Evolution 61*, 346-367.
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. http://doi.org/10.1093/molbev/mss086
Burke, J. M., & Arnold, M. L. (2001). Genetics and the fitness of hybrids. *Annu. Rev. Genet.*, *35*, 31–52.
Burney, C.W., Brumfield, R.T., (2009). Ecology predicts levels of genetic differentiation in Neotropical birds. *Am. Nat. 174*, 358-368.
Burnham, K.P., Anderson, D.R., (2002). Model selection and multimodel inference: A practical infromation-theoretical approach. Springer-Verlag, New York.
Burns, K. J., Shultz, A. J., Title, P. O., Mason, N. A., Barker, F. K., Klicka, J., … Lovette, I. J. (2014). Phylogenetics and diversification of tanagers (Passeriformes: Thraupidae), the largest radiation of Neotropical songbirds. *Molecular Phylogenetics and Evolution*, *75*, 41–77. http://doi.org/10.1016/j.ympev.2014.02.006
Bush, M.B., Oliveira, P.E., (2006). The rise and fall of the refugial hypothesis of Amazonian speciation: a paleo-ecological perspective. Biota Neotropica. 6 (n1) DOI: 10.1590/S1676-06032006000100002
Bush, M.B, Gosling, W.D., Colinvaux, P.A., (2007). Climate change in the lowlands of the Amazon Basin. in: Bush, M., Flenley, J., Gosling, W. (Eds.), Tropical rainforest responses to climatic change. eds. Springer Praxis Books. pp. 55-76.
Cabanis., (1851). *Astragalinus*, type *F. tristis* Lin. Mus. Hein. I, 159.
Campagna, L., Geale, K., Handford, P., Lijtmaer, D.A., Tubaro, P.L., Lougheed, S.C., (2011). A molecular phylogeny of the Sierra-Finches (*Phrygilus,* Passeriformes): Extreme polyphyly in a group of Andean specialists. *Mol. Phylogenet. Evol. 61*, 521-533.
Campagna, L., Benites, P., Lougheed, S.C., Lijtmaer, D.A., Di Giacomo, A.S., Eaton, M.D., Tubaro, P.L., (2012). Rapid phenotypic evolution during incipient speciation in a continental avian radiation. *Proc. R. Soc. B. 279*, 1847-1856.

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. http://doi.org/10.1111/mec.12354

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. *Genes|Genomes|Genetics*, *1*, 171–182. http://doi.org/10.1534/g3.111.000240

Chesser, R.T., (1999). Molecular systematics of the rhinocryptid genus *Pteroptochos*. *Condor 101*, 439–446.

Chesser, R.T., (2000). Evolution in the high Andes: The phylogenetics of *Muscisaxicola* ground-tyrants. *Mol. Phylogenet. Evol. 15*, 369-380.

Chesser, R.T., (2004). Systematics, evolution, and biogeography of the South American ovenbird genus *Cinclodes*. *Auk 121*, 757-766.

Chesser, R.t., Banks, R.C., Barker, F.K., Cicero, C., Dunn, J.L., Kratter, A.W., Lovette, I.J., Rasmussen, P.C., Remsen Jr., J.V., Rising, J.D., Stotz, D.F., Winker, K., (2009). Fiftieth supplement to the American Ornithologists' Union Check-List of North American birds. *Auk 126*, 705-714.

Chesser, R.T & Rising J., (2013). Transer *Spinus notatus, S. xanthogastrus*, and *S. cucullatus* to *Sporagra*. in North and Middle America Classification Committee Proposal set 2014-A.

Cheviron, Z.A., Capparella, A.P., Vuilleumier, F., (2005). Molecular phylogenetic relationships among the *Geositta* miners (Furnariidae) and biogeographic implications for Avian speciation in Fuego-Patagonia. *Auk 122*,158-174.

Cheviron, 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. http://doi.org/10.1111/j.1558-5646.2009.00644.x

Cheviron, Z. A., Natarajan, C., Projecto-Garcia, J., Eddy, D. K., Jones, J., Carling, M. D., … Storz, J. F. (2014a). Integrating Evolutionary and Functional Tests of Adaptive Hypotheses: A Case Study of Altitudinal Differentiation in Hemoglobin Function in an Andean Sparrow, Zonotrichia capensis. *Molecular Biology and Evolution*, *31*, 2948–2962. http://doi.org/10.1093/molbev/msu234

Chifman, J., & Kubatko, L. (2014). Quartet Inference from SNP Data Under the Coalescent Model. *Bioinformatics*, *30*, 3317–3324. http://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. http://doi.org/10.1016/j.jtbi.2015.03.006

Clapperton, C., (1993). Nature of environmental changes in South America at the Last Glacial Maximum. *Palaeogeogr., Palaeoclimatol., Palaeoecol. 101*, 189-208.

Claramunt, S., Derryberry, E.P., Remsen Jr, J.V., Brumfield, R.T., (2012). High dispersal ability inhibits speciation in a continental radiation of passerine birds. *Proc. R. Soc. B 279*, 1567-1574.

Clark, W.S., Witt, C.C., (2006). First known specimen of a hybrid Buteo: Swainson's Hawk (*Buteo Swainsoni*) x Rough-legged Hawk (*B. lagopus*) from Louisiana. *Wilson J. Ornithol. 118*, 42-52.

Colosimo, P. F. (2005). Widespread Parallel Evolution in Sticklebacks by Repeated Fixation of Ectodysplasin Alleles. *Science*, *307*, 1928–1933. http://doi.org/10.1126/science.1107239

Coyne, J.A., Price, T.D., (2000). Little evidence for sympatric speciation in island birds. *Evolution 54*, 2166-2171.

Cracraft, J., (1985). Historical biogeography and patterns of differentiation within the South American avifauna: areas of endemism. in: Buckley, P. A. et al. (Eds.). Neotropical ornithology. American Ornithologists Union, Lawrence , Kansas , pp. 49–84.

Davis, J.N., (1999). Lawrence's Goldfinch (*Spinus lawrencei*), The Birds of North America Online (A. Poole, Ed.). Ithaca: Cornell Lab of Ornithology; Retrieved from the Birds of North America Online: http://bna.birds.cornell.edu/bna/species/480.

Dawson, W.R., (1997). Pine Siskin (*Spinus pinus*), The Birds of North America Online (A. Poole, Ed.). Ithaca: Cornell Lab of Ornithology; Retrieved from the Birds of North America Online: http://bna.birds.cornell.edu/bna/species/280.

Diamond, J.M., Gilpin, M.E., Mayr, E., (1976). Species-distance relation for birds of the Solomon Archipelago, and the paradox of the great speciators. *Proc. Natl. Acad. Sci.. 73*, 2160-2164.

Dobzhansky, T. (1937). *Genetics and the origin of species*. New York, NY: Columbia University Press.

Dobzhansky, T. (1948). Genetics of natural populations. XVI. Altitudinal and seasonal changes produced by natural selection in certain populations of Drosophila psuedoobscura and Drosophila persimilis. *Genetics*, *33*, 158–176.

Doebeli, M., & Dieckmann, U. (2003). Speciation along environmental gradients. *Nature*, *421*, 259–264. http://doi.org/10.1038/nature01274

Dowling, T. E., & DeMarais, B. D. (1993). Evolutionary significance of introgressive hybridization in cyprinid fishes. *Nature*, *362*, 444–446. http://doi.org/10.1038/362444a0

Drummond, A.J., Rambaut, A., (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol. 7*, 214. doi:10.1186/1471-2148-7-214.

DuBay, S.G., Witt, C.C., (2012). An improved phylogeny of the Andean tit-tyrants (Aves, Tyrannidae): More characters trump sophisticated analyses. *Mol. Phylogenet. Evol. 64*, 285-296.

DuBay, S.G., Witt, C.C., (2014). Differential high-altitude adaption and restricted gene flow across a mid-elevation hybrid zone in Andean tit-tyrant flycatchers. *Mol. Ecol. 23*, 3551-3565. http:///doi.org/10.1111/mec.12836

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. http://doi.org/10.1093/molbev/msr048

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.

http://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.
Edwards, S. V. (2009). Is a new and general theory of molecular systematics emerging? *Evolution*, *63*, 1–19. http://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 multispecies coalescent model: A valuable paradigm for phylogenomics. *Molecular Phylogenetics and Evolution*, *94*, 447–462. http://doi.org/10.1016/j.ympev.2015.10.027
Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE*, *6*, 1–10. http://doi.org/10.1371/journal.pone.0019379
Excoffier, L., Laval, G., and Schneider, S. 2005. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online. 1,* 47-50.
FitzJohn, R.G., Goldberg, E.E., Magnuson-Ford, K., 2013. diversitree: comparative phylogenetic analyses of diversification. R package version 0.9-6. Available at: http://www.zoology.ubc.ca/prog/diversitree.
Fishman, L., & Willis, J. H. (2001). Evidence for Dobzhansky-Muller Incompatibilites Contributing To the Sterility of Hybrids Between Mimulus Guttatus and M . Nasutus. *Evolution*, *55*, 1932–1942.
Fjeldså, J., Bowie, R. C. K., & Rahbek, C. (2012). The Role of Mountain Ranges in the Diversification of Birds. *Annual Review of Ecology, Evolution, and Systematics*, *43*, 249-265 http://doi.org/10.1146/annurev-ecolsys-102710-145113
Fjeldså, J., & Krabbe, N. K. (1990). *Birds of the High Andes*. Copenhagen: Apollo Books.
Fjeldså, J., (1994). Geographic patterns for relict and young species of birds in Africa and South America and implications for conservation priorities. *Biodivers. Conserv. 3*, 207-226.
Fjeldså, J., Rahbek, C., (2006). Diversification of tanagers, a species rich bird group, from lowlands to montane regions of South America. *Integr. Comp. Biol. 46*, 72-81.
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*, 1258524–1258524. http://doi.org/10.1126/science.1258524
Ford, A. G. P., Dasmahapatra, K. K., Rüber, L., Gharbi, K., Cezard, T., & Day, J. J. (2015). High levels of interspecific gene flow in an endemic cichlid fish adaptive radiation from an extreme lake environment. *Molecular Ecology*, *24*, 3421–3440. http://doi.org/10.1111/mec.13247
Fu, Y. X., (1997). Statistical tests of neutrality of mutations against population growth, hitchhiking and backgrounds selection. *Genet. 147:* 915-925.
Galen, S. C., Natarajan, C., Moriyama, H., Weber, R. E., Fago, A., Benham, P. M., … Witt, C. C. (2015). Contribution of a mutational hot spot to hemoglobin adaptation in high-altitude Andean house wrens. *Proceedings of the National*

Academy of Sciences, *112*, 13958–13963 http://doi.org/10.1073/pnas.1507300112
Garant, D., Forde, S. E., & Hendry, A. P. (2007). The multifarious effects of dispersal and gene flow on contemporary adaptation. *Functional Ecology*, *21*, 434–443. http://doi.org/10.1111/j.1365-2435.2006.01228.x
García-Moreno, J., Arctander, P., Fjeldså, J., (1998). Pre-Pleistocene differentiation among chat-tyrants. *Condor 100*, 629-640.
García-Moreno, J., Arctander, P., Fjeldså, J., (1999). A case of rapid diversification in the Neotropics: Phylogenetic relationships among Cranioleuca *Spinetails* (Aves, Furnairiidae). *Mol. Phylogenet. Evol. 12*, 273-281.
García-Moreno, J., Navarro-Sigüenza, A.G., Peterson, A.T., Sánchez-González, L.A., (2004). Genetic variation coincides with geographic structure in the common bush-tanager (*Chlorospingus ophthalmicus*) complex from Mexico. *Mol. Phylogenet. Evol. 33,* 186-196.
Gatesy, J., & Springer, M. S. (2014). Phylogenetic analysis at deep timescales: Unreliable gene trees, bypassed hidden support, and the coalescence/concatalescence conundrum. *Molecular Phylogenetics and Evolution*, *80*, 231–266. http://doi.org/10.1016/j.ympev.2014.08.013
Gavrilets, S., & Losos, J. B. (2009). Adaptive Radiation : Contrasting theory with data. *Science*, *323*, 732–737.
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. http://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. http://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. http://doi.org/10.1038/hdy.2008.124
Goldberg, E.E., Lancaster, L.T., Ree, R.H., 2011. Phylogenetic inference of reciprocal effects between geographic range evolution and diversification. *Syst. Biol. 60*, 451-465.
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. http://doi.org/10.1111/evo.12712
Grant, P. R., & Grant, B. R. (2009). The secondary contact phase of allopatric speciation in Darwin's finches. *Proceedings of the National Academy of Sciences of the United States of America*, *106*, 20141–20148. http://doi.org/10.1073/pnas.0911761106
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. http://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. http://doi.org/10.1086/430331
Graves, G.R., (1985). Elevational correlates of speciation and intraspecific geographic variation in plumage in Andean forest birds. *Auk 102*, 556-576.
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. http://doi.org/10.1126/science.1188021

Gutiérrez-Pinto, N., Cuervo, A.M., Miranda, J., Pérez-Emán, J.L., Brumfield, R.T., Cadena, C.D., (2012). Non-monophyly and deep genetic differentiation across low-elevation barriers in a neotropical montane bird (*Basileuterus tristiatus*; Aves: Parulidae). *Mol. Phylogenet. Evol. 64*, 156-165.

Hackinger, S., Kraaijenbrink, T., Xue, Y., Mezzavilla, M., Asan, van Driem, G., … Ayub, Q. (2016). Wide distribution and altitude correlation of an archaic high-altitude-adaptive EPAS1 haplotype in the Himalayas. *Human Genetics*, *135*, 393–402. http://doi.org/10.1007/s00439-016-1641-2

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

Harmon, L.J., Weir, J.T., Brock, C.D., Glor, R.E., Challenger, W., (2008). GEIGER: investigating evolutionary radiations. *Bioinformatics 24*,129-131.

Harshman, J., (1996). Phylogeny, evolutionary rates and ducks. Thesis, University of Chicago.

Harvey, M. G., Judy, C. D., Seeholzer, G. F., Maley, J. M., Graves, G. R., & Brumfield, R. T. (n.d.). Similarity thresholds used in short read assembly reduce the comparability of population histories across species. *PeerJ*. http://doi.org/10.7287/peerj.preprints.864v1

Hawkins, B.A., Diniz-Filho, J.A.F., Jaramillo, C.A., Soeller, S.A., (2006). Post-Eocene climate change, niche conservatism and the latitudinal diversity gradient of New World birds. J. *Biogeography 33*, 770-780.

Hawkins, B.A., McCain, C.M., Davies, T.J., Buckley, L.B., Anacker, B.L., Cornell, H.V., Damschen, E.I., Grytnes, J., Harisson, S., Holt, R.D., Kraft, N.J.B. Stephens, P.R., (2012). Different evolutionary histories underlie congruent species richness gradients of birds and mammals. J. *Biogeography 39*, 825-841.

Heled, J., Drummond, A.J., (2010). Bayesian Inference of Species Trees from Multilocus Data. *Mol. Biol. Evol. 27*, 570-580.

Hilty, S.L., (2003). Birds of Venezuela, 2nd ed. Princeton University Press, Princeton, New Jersey.

Ho, S.Y.W., Phillips, M.J., Cooper, A. Drummond, A.J., (2005). Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. *Mol. Biol. Evol. 22*, 1561-1568.

Hoffmann, F. G., Storz, J. F., Gorr, T. A., & Opazo, J. C. (2010). Lineage-specific patterns of functional diversification in the ??-and ??-globin gene families of tetrapod vertebrates. *Molecular Biology and Evolution*, *27*, 1126–1138. http://doi.org/10.1093/molbev/msp325

Hooghiemstra, H., van der Hammen, T., (2004). Quaternary Ice-Age Dynamics in the Colombian Andes: Developing an Understanding of Our Legacy. *Phil. Trans. R. Soc. B Biol. Sci. 359*, 173-181.

Hoorn, C., Wesselingh, F.P., ter Steege, H., Bermudez, M.A., Mora, A., Sevink, J., Sanmartín, I., Sanchez-Meseguer, A., Anderson, C.L., Figueiredo, J.P., Jaramillo, C., Riff, D., Negri, F.R., Hooghiemstra, H., Lundberg, J., Stadler, T., Särkinen, T., Antonelli, A., (2010). Amazonia through time: Andean uplift, climate change, landscape evolution and biodiversity. *Science 330*, 927 - 31.

Howard, R., Moore, L. (2003). The Howard & Moore Complete Checklist of the Birds of the World. Princeton University Press. Princeton, New Jersey.
Hua, X., Janzen, D., McCain, C., Moritz, C., Patton, J., Schneider, C., … Loeschcke, V. (2016). The impact of seasonality on niche breadth, distribution range and species richness: a theoretical exploration of Janzen's hypothesis. *Proceedings. Biological Sciences / The Royal Society*, *283*, 233–249. http://doi.org/10.1098/rspb.2016.0349
Huelsenbeck, J.P., (1995). Performance of phylogenetic methods in simulation. *Syst. Biol. 44*, 17-48.
Huerta-Sánchez, E., Jin, X., Asan, Bianba, Z., Peter, B. M., Vinckenbosch, N., … Nielsen, R. (2014). Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature*, *512*, 194–7. http://doi.org/10.1038/nature13408
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. http://doi.org/10.1126/science.1253451
Jessen, T.-H., Weber, R. E., Fermi, G., Tame, J., & Braunitzer, G. (1991). Adaptation of bird hemoglobins to high altitudes: Demonstration of molecular mechanism by protein engineering. *Proceedings of the National Academy of Sciences of the United States of America*, *88*, 6519–6522. http://doi.org/10.1073/pnas.88.15.6519
Jetz, W., Thomas, G.H., Joy, J.B., Hartmann, K., Mooers, A.O., (2012). The global diversity of birds in space and time. *Nature 491*, 444-448.
Jetz, W., Guralnick, R., Steele, A., Vaidya, G., Schank, C., Malczyk, J., Meyer, C., Erb, P., Auer, T., (2013). Map of Life. Available at www.mappinglife.org.
Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*, *27*, 3070–3071. http://doi.org/10.1093/bioinformatics/btr521
Jones, F. C., Grabherr, M. G., Chan, Y. F., Russell, P., Mauceli, E., Johnson, J., … Kingsley, D. M. (2012). The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, *484*, 55–61. http://doi.org/10.1038/nature10944
Jones, M. R., & Good, J. M. (2015). Targeted capture in evolutionary and ecological genomics. *Molecular Ecology*, *25*, 185-202 . http://doi.org/10.1111/mec.13304
Kawecki, T. J., & Ebert, D. (2004). Conceptual issues in local adaptation. *Ecology Letters*, *7*, 1225–1241. http://doi.org/10.1111/j.1461-0248.2004.00684.x
Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., … Drummond, A. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, *28*, 1647–1649. http://doi.org/10.1093/bioinformatics/bts199
Keller, I., Wagner, C. E. Greuter, L., Mwaiko, S., Selz, O.M. Sivasundar, A., Wittwer, S. and Seehausen, O. (2012). 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. http://doi.org/10.1111/mec.12083.
Kendall, D.G., 1948. On the generalized birth-and-death proces. *Ann. Math. Stat. 19*, 1-15.
Koch., 1816. *Spinus*: type *F. spinus* Linneaus. Bayr Zool. p 233.
Kocher, T.D., Thomas, W.K., Meyer, A., Edwards, S.V., Paabo, S., Villablanca, F.X., AC Wilson, A.C., 1989. Dynamics of mitochondrial DNA evolution in animals:

Amplification and sequencing with conserved primers. *Proc. Natl. Acad. Sci. 86,* 6196-6200.

Kubatko, L. S., & Degnan, J. H. (2007). Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Systematic Biology*, *56*, 17–24. http://doi.org/10.1080/10635150601146041

Kumar, S., Skjæveland, Å., Orr, R.J.S., Enger, P., Ruden, T., Mevik, B., Burki, F., Botnen, A., Shalchian-Tabrizi, K., 2009. AIR: A batch-oriented web program package for construction of supermatrices ready for phylogenomic analyses. *BMC Bioinformatics 10*, 357.

Lamichhaney, 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. http://doi.org/10.1038/nature14181

Lanfear, R., Calcott, B., Ho, S.Y.W., Guindon, S., 2012. PartitionFinder: Combined Selection of Partitioning Schemes and Substitution Models for Phylogenetic Analyses. *Mol. Biol. Evol. 29*, 1695-1701.

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods*, *9*, 357–359. http://doi.org/10.1038/nmeth.1923

Leaché, A. D. (2009). Species tree discordance traces to phylogeographic clade boundaries in north American Fence lizards (Sceloporus). *Systematic Biology*, *58*, 547–559. http://doi.org/10.1093/sysbio/syp057

Leaché, A. D., Banbury, B. L., Felsenstein, J., de Oca, A. nieto-M., & Stamatakis, A. (2015). Short Tree, Long Tree, Right Tree, Wrong Tree: New Acquisition Bias Corrections for Inferring SNP Phylogenies. *Systematic Biology*, *64*, 1032–1047. http://doi.org/10.1093/sysbio/syv053

Leaché, A. D., Harris, R. B., Rannala, B., & Yang, Z. (2014). The influence of gene flow on species tree estimation: A simulation study. *Systematic Biology*, *63*,17–30. http://doi.org/10.1093/sysbio/syt049

Lemmon, A.R., Moriarty, E.C., (2004). The importance of proper model assumption in Bayesian phylogenetics. *Syst. Biol. 53*, 265-277.

Lerner, H.R., Meyer, M., James, H.F., Hofreiter, M., Fleischer, R.C., (2011). Multilocus resolution of phylogeny and timescale in the extant adaptive radiation of Hawaiian Honeycreepers. *Curr. Biol. 21,* 1838-1844.

Lewis, P. O. (2001). A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic Biology*, *50*, 913–925. http://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. http://doi.org/10.2307/2406633

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*, 2078–2079. http://doi.org/10.1093/bioinformatics/btp352

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*, http://doi.org/10.1371/journal.pcbi.1003649

Lou, H., Lu, Y., Lu, D., Fu, R., Wang, X., Feng, Q., … Xu, S. (2015). A 3.4-kb Copy-Number Deletion near EPAS1 Is Significantly Enriched in High-Altitude Tibetans but Absent from the Denisovan Sequence. *American Journal of Human Genetics*, *97*, 54–66. http://doi.org/10.1016/j.ajhg.2015.05.005

Maddison, D.R., Maddison, W.P., (2005). MacClade 4: Analysis of phylogeny and character evolution. Version 4.08. Sinauer Associates, Sunderland, Massachusetts.

Madriñán, S., Cortés, A.J., Richardson, J.E., (2013). Páramo is the world's fastest evolving and coolest biodiversity hotspot. *Front. Genet. 4* (article 192), 1- 7.

Mahler, D. L., Revell, L. J., Glor, R. E., & Losos, J. B. (2010). Ecological opportunity and the rate of morphological evolution in the diversification of greater Antillean anoles. *Evolution*, *64*, 2731–2745. http://doi.org/10.1111/j.1558-5646.2010.01026.x

Mallet, J. (2005). Hybridization as an invasion of the genome. *Trends in Ecology and Evolution*, *20*, 229–237. http://doi.org/10.1016/j.tree.2005.02.010

Mallet, J., Besansky, N., & Hahn, M. W. (2015). How reticulated are species? *BioEssays*. *38*, http://doi.org/10.1002/bies.201500149

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. http://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. http://doi.org/10.1111/1755-0998.12291

McCormack, J. (2013). A phylogeny of birds based on over 1,500 loci collected by target enrichment and high-throughput sequencing. *PLoS One*, 1–38.

McCracken, K. G., Barger, C. P., Bulgarella, M., Johnson, K. P., Sonsthagen, S. A., Trucco, J., … Sorenson, M. D. (2009). Parallel evolution in the major haemoglobin genes of eight species of Andean waterfowl. *Molecular Ecology*, *18*, 3992–4005. http://doi.org/10.1111/j.1365-294X.2009.04352.x

McDonald, D. B., Clay, R. P., Brumfield, R. T., & Braun, M. J. (2001). Sexual selection on plumage and behavior in an avian hybrid zone: experimental tests of male-male interactions. *Evolution; International Journal of Organic Evolution*, *55*, 1443–51. http://doi.org/10.1111/j.0014-3820.2001.tb00664.x

McGraw, K.J., Middleton, A.L., (2009). American Goldfinch (*Spinus tristis*), The Birds of North America Online (A. Poole, Ed.). Ithaca: Cornell Lab of Ornithology; Retrieved from the Birds of North America Online: http://bna.birds.cornell.edu/bna/species/080.

McGuire, J.A., Witt, C.C., Altshuler, D.L., Remsen Jr., J.V., (2007). Phylogenetic systematics and biogeography of hummingbirds: Bayesian and maximum likelihood analyses of partitioned data and selection of an appropriate partitioning strategy. *Syst. Biol. 56*, 837-856.

McGuire, J.A., Witt, C.C., Remsen Jr., J.V., Corl, A., Rabosky, D.L. Altshuler, D.L., Dudley, R., (2014). Molecular phylogenetics and the diversification of hummingbirds. *Curr. Biol. 24*, 910-916.

Mendelson, T.C., Shaw, K.L., (2005). Rapid speciation in an arthropod. *Nature. 433*, 375-396.

Miller, M.A., Pfeiffer, W., Schwartz, T., (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Proc. Gateway Computing Environment Workshop (GCE), 14 Nov 2010, New Orleans, LA, 1-8.

Mindell, D.P., Sorenson, M.D., Dimcheff, D.E., (1998). An extra nucleotide is not translated in mitochondrial ND3 of some birds and turtles. *Mol. Biol. Evol. 15*,1568–1571.

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*, *100*, 171–176. http://doi.org/10.1073/pnas.0136972100

Moyle, R. G., Filardi, C. E., Smith, C. E., & Diamond, J. (2009). Explosive Pleistocene diversification and hemispheric expansion of a "great speciator". *Proceedings of the National Academy of Sciences of the United States of America*, *106*, 1863–1868. http://doi.org/10.1073/pnas.0809861105

Myers, N., Mittermeier, R.A., Mittermeier, C.G., da Fonseca, G.A.B., Kent, J., (2000). Biodiversity hotspots for conservation priorities. *Nature 403*, 853-858.

Natarajan, C., Hoffmann, F. G., Weber, R. E., Fago, A., Witt, C. C., & Storz, J. F. (2016). Predictable convergence in hemoglobin function has unpredictable molecular underpinnings. *Science*, *in press*.

Natarajan, C., Projecto-Garcia, J., Moriyama, H., Weber, R. E., Muñoz-Fuentes, V., Green, A. J., … Storz, J. F. (2015). Convergent Evolution of Hemoglobin Function in High-Altitude Andean Waterfowl Involves Limited Parallelism at the Molecular Sequence Level. *PLoS Genetics*, *11*, 1–25. http://doi.org/10.1371/journal.pgen.1005681

Nee, S., (2006). Birth-Death Models in Macroevolution. *Ann. Rev. Ecol. Evol. Syst. 37*, 1-17.

Nguembock, B., Fjeldså, J., Pasquet, E., (2008). Molecular phylogeny of Carduelinae (Aves, Passeriformes, Fringillidae) proves polyphyletic origin of the genera *Serinus* and *Carduelis* and suggests redefined generic limits. *Mol. Phylogenet. Evol. 51*, 169-181.

Nores, M. (2004). The implications of Tertiary and Quaternary sea level rise events for avian distribution patterns in the lowlands of northern South America. *Global Ecol. Biogeogr. 13*, 149-161.

Novembre, J. and Stephens, M. (2008). Interpreting principal component analyses of spatial population genetic structure. *Nature genetics. 40*, 646-649. http://doi.org/10.1038/ng.139

Nylander, J.A.A., Wilgenbusch, J.C., Warren, D.L., Swofford, D.L., (2008a). AWTY (are we there yet?): a system for graphical exploration of MCMC convergence in Bayesian phylogenetics. *Bioinformatics 24*, 581-583.

Orme, C.D.L., Davies, R.G., Burgess, M., Eigenbrod, F., Pickup, N., Olson, V.A., Webster, A.J., Ding, T., Rasmussen, P.C., Ridgely, R.S., Stattersfield, A.J., Bennett, P.M., Blackburn, T.M., Gaston, K.J., Owens, I.P.F., (2005). Global hotspots of species richeness are not congruent with endemism or threat. *Nature 436*, 1016-1019.

Opazo, J. C., Hoffmann, F. G., Natarajan, C., Witt, C. C., Berenbrink, M., & Storz, J. F. (2015). Gene turnover in the avian globin gene families and evolutionary changes in hemoglobin isoform expression. *Molecular Biology and Evolution*, *32*, 871–887. http://doi.org/10.1093/molbev/msu341

Owens, I.P.F., Bennett, P.M., Harvey, P.H., (1999). Species richness among birds: Body size, life history, sexual selection or ecology? *Proc. R. Soc. B. 266*, 933-939.

Paradis, E. (2010). Pegas: An R package for population genetics with an integrated-modular approach. *Bioinformatics*, *26*, 419–420. http://doi.org/10.1093/bioinformatics/btp696

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. http://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. http://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. http://doi.org/10.1534/genetics.112.145037

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. http://doi.org/10.1371/journal.pbio.1002379

Peters, J.L., Zhuravlev, Y., Fefelov, I., Logie, A., Omland, K.E., (2007). Nuclear loci and coalescent methods support ancent hybridization as cause of mitochondrial paraphyly between Gadwall and Falcated Duck (*Anas* spp.) *Evolution 61*, 1992-2006.

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.

Phillimore, A.B., Freckleton, R.P., Orme, C.D.L., Owens, I.P.F., (2006). Ecology predicts large-scale patterns of phylogenetic diversification in birds. *Am. Nat. 168*, 220-229.

Pickrell, J. K., & Pritchard, J. K. (2012). Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. *PLoS Genetics*, *8*(11). http://doi.org/10.1371/journal.pgen.1002967

Plummer, M., Best, N., Cowles, K., Vines, K., Sarkar, D., Almond, R., (2012). coda: Output analysis and diagnostics for MCMC. R package version 0.16-1. R package version 2.3. http://CRAN.R-project.org/package=coda.

Price, T.D., 2011. Adaptive radiations: There's something about finches. *Curr. Biol. 21*, R953-R955.

Projecto-Garcia, J., Natarajan, C., Moriyama, H., Weber, R.E., Fago, A., Cheviron, A.Z., Dudley, R., McGuire, J.A., Witt, C.C., Storz, J.F., (2013). Repeated elevational transitions in hemoglobin function during the evolution of Andean hummingbirds. *Proc. Natl. Acad. Sci. 110*, 20669-20674.

Prum, R. O., Berv, J. S., Dornburg, A., Field, D. J., Townsend, J. P., Lemmon, E. M., &

Lemmon, A. R. (2015). A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature*, *526*, 569–573. http://doi.org/10.1038/nature15697

Pybus OG & Harvey PH., (2000). Testing Macro-Evolutionary Models Using Incomplete Molecular Phylogenies. *Proc. R. Soc. Lond. B. 267*, 2267-2272.

R Core Team. (2016). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Rabosky, D. L., Santini, F., Eastman, J., Smith, S. A., Sidlauskas, B., Chang, J., & Alfaro, M. E. (2013). Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation. *Nature Communications*, *4*, 1958. http://doi.org/10.1038/ncomms2958

Rabosky, D.L., (2006a). Laser: A maximum likelihood toolkit for detecting temporal shifts in diversification rates from molecular phylogenies. *Evol. Bioinform. Online. 2*, 257-250.

Rabosky, D.L., (2006b). Likelihood methods for detecting temporal shifts in diversification rates. *Evolution 60*, 1152-1164.

Rambaut, A., Suchard, M., Xie, D., & Drummond, A. J. (2014). Tracer v1.6.

Rambaut, A., Drummond, A.J. (2009). Tracer. Version 1.5. Available at http://tree.bio.ed.ac.uk/software/tracer.

Räsänen, K., & Hendry, A. P. (2008). Disentangling interactions between adaptive divergence and gene flow when ecology drives diversification. *Ecology Letters*, *11*, 624–636. http://doi.org/10.1111/j.1461-0248.2008.01176.x

Reaz, R., Bayzid, M. S., & Rahman, M. S. (2014). Accurate phylogenetic tree reconstruction from quartets: A heuristic approach. *PLoS ONE*, *9*, http://doi.org/10.1371/journal.pone.0104008

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. http://doi.org/10.1038/nature09710

Reich, D., Thangaraj, K., Patterson, N., Price, A. L., & Singh, L. (2009). Reconstructing Indian population history. *Nature*, *461*, 489–94. http://doi.org/10.1038/nature08365

Reitman, M., Grasso, J. a, Blumenthal, R., & Lewit, P. (1993). Primary sequence, evolution, and repetitive elements of the Gallus gallus (chicken) beta-globin cluster. *Genomics*, *18*, 616–26. http://doi.org/10.1016/S0888-7543(05)80364-7

Remsen, J.V., Jr., (2011). Proposal 488 to the South American Classification Committee: Resurrect *Sporagra* for South American goldfinches and siskins.

Remsen, J. V., Jr., Cadena, C.D., Jaramillo, A., Nores, M., Pacheco, J.F., Pérez-Emán, J., Robbins, M.B., Stiles, F.G., Stotz, D.F., Zimmer, K.J., Version 12 July 2014. A classification of the bird species of South America. American Ornithologists' Union. http://www.museum.lsu.edu/~Remsen/SACCBaseline.html.

Restall, R., Rodner, C., Lentino, M., (2006). Birds of Norther South America: An Identification Guide, Volume I. Yale University Press.

Rheindt, F.E., Fujita, M.K., Wilton, P.R., Edwards, S.V., (2014). Introgression and phenotypic assimilation in *Zimmerius* flycatches (Tyrannidae): Population genetic and phylogenetic inferences from genome-wide SNPs. *Syst. Biol. 63*, 134-152.

Rheindt, F. E., & Edwards, S. V. (2011). Genetic Introgression: An Integral but neglected

component of speciation in birds. *The Auk*, *128*, 620–632. http://doi.org/10.1525/auk.2011.128.4.620
Ridgely, R. S., & Tudor, G. (1989). *The Birds of South America. Vol. I.* Austin, TX.: University of Texas Press.
Rieseberg, L. H., Whitton, J., & Gardner, K. (1999). Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. *Genetics*, *152*, 713–727.
Ronquist, F. Huelsenbeck, J.P., (2003). MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinform. 19*,1572-1574.
Ronquist, F., (1996). Dispersal-vicariance analysis: a new approach to the quantification of historical biogeography. *Syst. Biol. 46*, 195-203.
Ronquist, F., (2004). Bayesian inference of character evolution. *TRENDS Ecol. Evol. 19*, 475-481.
Rozen, S., Skaletsky, H.J., (2000). Primer3 on the WWW for general users and for biologist programers. in: Krawetz, S., Misener, S. (Eds.), Bioinformatics Methods and Protocols: Methods in Molecular Biology. Humana Press, Totowa, NJ, pp. 365-386.
Rull, V., (2008). Speciation timing and neotropical biodiversity: the Tertiary-Quaternary debate in the light of molecular phylogenetic evidence. *Mol. Ecol. 17*, 2722-2729.
Rull, V., (2011). Neotropical biodiversity: timing and potential drivers. *TRENDS Ecol. Evol. 26*, 508 - 513.
Sætre, G. P., & Sæther, S. A. (2010). Ecology and genetics of speciation in Ficedula flycatchers. *Molecular Ecology*, *19*,1091–1106. http://doi.org/10.1111/j.1365-294X.2010.04568.x
Sanderson, M.J. and Shaffer, H. B. (2002). Troubleshooting molecular phylogenetic analyses. *Ann. Rev. Ecol. Syst. 33*, 49-72.
Särkinen, T., Pennington, R.T., Lavin, M., Simon, M.F., Hughes, C.E., (2012). Evolutionary islands in the Andes: persistence and isolation explain high endemism in Andean dry tropical forests. *J. Biogeography 39*, 884-900.
Savolainen, O., Lascoux, M., & Merilä, J. (2013). Ecological genomics of local adaptation. *Nature Reviews. Genetics*, *14*, 807–20. http://doi.org/10.1038/nrg3522
Schulenberg, T.S., Stotz, D.F., Lane, D.F., O'Neill, J.P., Parker III, T.A., (2007). Birds of Peru. Princeton University Press. Princeton, New Jersey.
Schluter, D. (2001). Ecology and the origin of species. *Trends in Ecology and Evolution*. *7,* 372-380 http://doi.org/10.1016/S0169-5347(01)02198-X
Schluter, D. (2000). *The Ecology of Adaptive Radiation*. (R. M. May & P. H. Harvey, Eds.) (Oxford Ser). Oxford: Oxford University Press.
Scott, G. R., Meir, J. U., Hawkes, L. A., Frappell, P. B., & Milsom, W. K. (2011). Point: high altitude is for the birds! *J Appl Physiol*, *111*, 1514–1515. http://doi.org/10.1152/japplphysiol.00821.2011
Scott, G. R., Egginton, S., Richards, J. G., & Milsom, W. K. (2009). Evolution of muscle phenotype for extreme high altitude flight in the bar-headed goose. *Proc Biol Sci*, *276*, 3645–3653. http://doi.org/10.1098/rspb.2009.0947\rrspb.2009.0947
Seehausen, O. (2004). Hybridization and adaptive radiation. *Trends in Ecology and Evolution*, *19*, 198–207. http://doi.org/10.1016/j.tree.2004.01.003
Seehausen, O., Alphen, J., & Witte, F. (1997). Cichild fish diversity threatened by

eutrophication that curbs sexual selection. *Science*, *277*, 1808–1810.
Sedano, R.E., Burns, K.J., 2010. Are the Northern Andes a species pump for Neotropical birds? Phylogenetics and biogeography of a clade of Neotropical tanagers (Aves: Thraupini*). J. Biogeography 37*, 325-343.
Servant, M., Maley, J., Turcq, B., Absy, M., Brenac, P., Fournier, M., Ledru, M., 1993. Tropical forest changes during the Late Quaternary in African and South American lowlands. *Glob. Planet. Change 7*, 25-40.
Servedio, M. R., & Noor, M. A. (2003). The role of reinforcement in speciation: theory and data. *Annual Review of Ecology, Evolution, and Systematics*, *34*, 339–364.
Sick, H. (1985). Birds in Brazil: a Natural History. Trans. W. Belton. Princeton University Press, Princeton, NJ.
Simpson, G. G. (1953). *The major features of evolution*. New York, NY: Columbia University Press.
Sobel, J. M., Chen, G. F., Watt, L. R., & Schemske, D. W. (2010). The biology of speciation. *Evolution*, *64*, 295–315. http://doi.org/10.1111/j.1558-5646.2009.00877.x
Sorenson, M.D., Ast, J.C., Dimcheff, D.E., Yuri, T., Mindell, D.P., (1999). Primers for a PCR-based approach to mitochondrial genome sequencing in birds and other vertebrates. *Mol. Phylogenet. Evol. 12*,105-114.
Springer, M. S., & Gatesy, J. (2016). The gene tree delusion. *Molecular Phylogenetics and Evolution*, *94*, 1–33. http://doi.org/10.1016/j.ympev.2015.07.018
Stamatakis, A. (2014). RAxML Version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 1–2.
Stamatakis, A., 2006. RAxML-VI-HPC: Maximum Likelihood-based Phylogenetic Analyses with Thousands of Taxa and Mixed Models. *Bioinformatics 22*, 2688-2690.
Stankowski, S., & Streisfeld, M. A. (2015). Introgressive hybridization facilitates adaptive divergence in a recent radiation of monkeyflowers. *Proceedings of the Royal Society of London B*, *282*, 1666. http://doi.org/10.1098/rspb.2015.1666
Stephens, M., Donnelly, P., (2003). A comparison of Bayesian methods for haplotype reconstruction from population genotype data. *Am. J. Hum. Genet., 73*,1162-1169.
Storz, J. F., Runck, A. M., Moriyama, H., Weber, R. E., & Fago, A. (2010). Genetic differences in hemoglobin function between highland and lowland deer mice. *The Journal of Experimental Biology*, *213*, 2565–2574. http://doi.org/10.1242/jeb.042598
Storz, J. F., Scott, G. R., & Cheviron, Z. A. (2010). Phenotypic plasticity and genetic adaptation to high-altitude hypoxia in vertebrates. *The Journal of Experimental Biology*, *213*, 4125–36. http://doi.org/10.1242/jeb.048181
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. http://doi.org/10.1111/mec.12814
Suh, A., Smeds, L., & Ellegren, H. (2015). The dynamics of incomplete lineage sorting across the ancient adaptive radiation of neoavian birds. *PLoS Biol. 13*: e1002224. http://dx.doi.org/10.1371/journal.pbio.1002224
Swofford, D. (2002). PAUP*. Phylogenetic Analysis Using Parsimony (* and Other

Methods). Version 4.0a147. Sunderland, Massachusetts: Sinauer Associates.
Tajima, F. (1996). The amount of DNA polymorphism maintained in a finite population when the neutral mutation rate varies among sites. *Genet. 143:* 1457-1465.
Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., (2011). MEGA5: Molecular Evolutionary Genetics Analysis using Likelihood, Distance, and Parsimony methods. *Mol. Biol. Evol. 28*, 2731-2739.
Taylor, E. B., Boughman, J. W., Groenenboom, M., Sniatynski, M., Schluter, D., & Gow, J. L. (2006). Speciation in reverse: Morphological and genetic evidence of the collapse of a three-spined stickleback (Gasterosteus aculeatus) species pair. *Molecular Ecology*, *15*, 343–355. http://doi.org/10.1111/j.1365-294X.2005.02794.x
The Heliconius Genome Consortium, Dasmahapatra, K. K., Walters, J. R., Briscoe, A. D., Davey, J. W., Whibley, A., … Jiggins, C. D. (2012). Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature*, *487*, 94–98. http://doi.org/10.1038/nature11041
Thomson, R.C., Shaffer, H.B., (2010). Sparce supermatrices for phylogenetic inference: Taxonomy, alignment, rogue taxa and the phylogeny of living turtles. *Syst. Biol. 59*, 42-58.
Turchetto-Zolet, A.C., Pinheiro, F., Salgueiro, F., Palma-Silva, C., (2013). Phylogeographical patterns shed light on evolutionary process in South America. *Mol. Ecol. 22,* 1193-1213.
van den Elzen, R., Guillén, J., Ruiz-del-Valle, V., Allende, L.M., Lowy, E., Zamora, J., Arnaiz-Villena, A., (2001). Both morphological and molecular characters support speciation of South American siskins by sexual selection. *Cell. Mol. Life Sci. 58*, 2117-2128.
van der Hammen, T., Cleef, A.M., (1986). Development of the high Andean páramo flora and vegetation. in: Vuilleumier, F., Monasterio, M. (Eds.), High Altitude tropical biogeography. Oxford Univ. Press. Oxford. pp. 153-201.
Voelker, G., (1999). Dispersal, vicariance and clocks: Historical biogeography and speciation in a cosmopolitan passerine genus (Anthus: Motacillidae). *Evolution 53*, 1536-1552.
Vonlanthen, P., Bittner, D., Hudson, A. G., Young, K. A., Müller, R., Lundsgaard-Hansen, B., … Seehausen, O. (2012). Eutrophication causes speciation reversal in whitefish adaptive radiations. *Nature*, *482*, 357–362. http://doi.org/10.1038/nature10824
Ward, S., Bishop, C. M., Woakes, a J., & Butler, P. J. (2002). Heart rate and the rate of oxygen consumption of flying and walking barnacle geese (Branta leucopsis) and bar-headed geese (Anser indicus). *The Journal of Experimental Biology*, *205*, 3347–3356.
Watt, D.J., Willoughby, E.J., (1999). Lesser Goldfinch (Spinus psaltria), The Birds of North America Online (A. Poole, Ed.). Ithaca: Cornell Lab of Ornithology; Retrieved from the Birds of North America Online: http://bna.birds.cornell.edu/bna/species/392.
Weir, J.T., (2006). Divergent timing and patterns of species accumulation in lowland and highland neotropical birds. *Evolution 60*, 842-855.

Weir, J.T., Price, M., (2011). Andean uplift promotes lowland speciation through vicariance and dispersal in *Dendrocincla* woodcreepers. *Mol. Ecol. 20*, 4550-4563.

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. http://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. http://doi.org/10.1371/journal.pgen.1006006

Wilson, R. E., Peters, J. L., & McCracken, K. G. (2013). Genetic and phenotypic divergence between low- and high-altitude populations of two recently diverged cinnamon teal subspecies. *Evolution*, *67*, 170–184. http://doi.org/10.1111/j.1558-5646.2012.01740.x

Wright, K. M., Lloyd, D., Lowry, D. B., Macnair, M. R., & Willis, J. H. (2013). Indirect Evolution of Hybrid Lethality Due to Linkage with Selected Locus in Mimulus guttatus. *PLoS Biology*, *11*, http://doi.org/10.1371/journal.pbio.1001497

Wright, S. (1931). Evolution in Mendelian populations. *Genetics*, *16*, 97–159.

Yu, Y., Harris, A.J., He, X.J., (2012). RASP (Reconstruct Ancestral State in Phylogenies) 2.1b. Available at http://mnh.scu.edu.cn/soft/blog/RASP.