**Rapid postglacial diversification and long-term stasis within the
songbird genus *Junco*: phylogeographic and phylogenomic evidence**

Guillermo Friis¹, Pau Aleixandre¹, Ricardo Rodríguez-Estrella², Adolfo G. Navarro-
Sigüenza³, Borja Milá¹

¹*National Museum of Natural Sciences, Spanish National Research Council (CSIC),
Madrid 28006, Spain;*²*Centro de Investigaciones Biológicas del Noroeste (CIBNOR),
La Paz, Baja California Sur, Mexico;*³*Museo de Zoología, Facultad de Ciencias,
Universidad Nacional Autónoma de México (UNAM), Mexico City, Mexico*

Corresponding author: Guillermo Friis, National Museum of Natural Sciences - CSIC,
José Gutiérrez Abascal 2, Madrid 28006, Spain; Email: gfriis@mncn.csic.es; Tel: +34
914111328 x1266.

Running title: Speciation in the songbird genus *Junco*

Word count: 11,386

Abstract

Natural systems composed of closely-related taxa that vary in the degree of phenotypic divergence and geographic isolation, provide an opportunity to investigate the rate of phenotypic diversification and the relative roles of selection and drift in driving lineage formation. The genus *Junco* (Aves: Emberizidae) of North America includes parapatric northern forms that are markedly divergent in plumage pattern and color, in contrast to geographically isolated southern populations in remote areas that show moderate phenotypic divergence. Here, we quantify patterns of phenotypic divergence in morphology and plumage color, and use mitochondrial DNA genes, a nuclear intron, and genome-wide SNPs to reconstruct the demographic and evolutionary history of the genus to infer relative rates of evolutionary divergence among lineages. We found that geographically isolated populations have evolved independently for hundreds of thousands of years despite little differentiation in phenotype, in sharp contrast to phenotypically diverse northern forms, which have diversified within the last few thousand years as a result of the rapid postglacial recolonization of North America. SNP data resolved young northern lineages into reciprocally monophyletic lineages, indicating low rates of gene flow even among closely related parapatric forms, and suggesting a role for strong genetic drift or multifarious selection acting on multiple loci in driving lineage divergence. Juncos represent a compelling example of speciation-in-action, where the combined effects of historical and selective factors have produced one of the fastest cases of speciation known in vertebrates.

Key words: phylogeography, phylogenomics, rapid speciation, postglacial expansion, avian radiation, GBS

Introduction

One of the most remarkable patterns in nature is the difference in the degree and rate of phenotypic diversification among clades and geographic areas. At large macroevolutionary scales, morphological diversification shows marked variation across clades, and it has been suggested to be more strongly related to the number of species in the clade than to its age (Rabosky *et al.* 2012; Rabosky *et al.* 2013). The fact that species number is also decoupled from clade age, suggests the role of ecological limits to diversification and the interplay of a number of factors in determining patterns of species richness, including latitude, climate, ecology and biogeography (Currie *et al.* 2004; Phillimore *et al.* 2007; Rabosky 2009; Mahler *et al.* 2010; Yoder *et al.* 2010; Title & Burns 2015). In addition, different clades have shown different degrees of ‘evolvability’, or the capacity to modify their morphology in the face of strong selection, as exemplified in birds by the striking diversification of Hawaiian honeycreepers, ten times greater than that of the Hawaiian thrushes, despite having colonized the archipelago at similar times (Lovette *et al.* 2002). In general, however, uncertainty concerning past extinction events, branch lengths and ancestral trait values at large phylogenetic scales complicates efforts to distinguish between time-dependent and speciation-dependent processes (Ricklefs 2006). Clearly, understanding the factors responsible for patterns of species richness among clades and the evolvability of different lineages requires not only improved macroevolutionary approaches, but also more detailed studies at finer phylogenetic scales.

Taxonomic groups at inter- and intra-specific levels that cover a broad range of spatial scales and ecological conditions allow a closer study of the relative roles of selection and drift in driving evolutionary divergence, and the rate of phenotypic diversification and

speciation. Specifically, ideal scenarios for studying speciation mechanisms are provided by those natural systems that (i) are composed of recently evolved and closely related intra and inter-specific lineages or populations, minimizing the probability of extinction; (ii) are found at different spatial scales, from local to regional, and covering a range of ecological conditions; (iii) differ prominently in the degree of geographic isolation, from none to moderate gene flow; and (iv) show marked differentiation in phenotypic traits related to fitness.

One such system is provided by the songbird genus *Junco* of North and Central America, a striking case of geographic variation spanning the entire continuum from the local population to the species, and from complete isolation to freely interbreeding parapatric forms (Miller 1941; Nolan *et al.* 2002). In his extensive 1941 monograph, Alden Miller identified and described a total of 21 different “forms” or “types”, and grouped them based on phenotypic similarities, geographical distribution and potential evolutionary relationships into 10 separate species-level taxa. However, because several of those forms were found to interbreed freely in areas of parapatric contact, those 21 forms are currently grouped taxonomically into just four species: the divergent *Junco vulcani* in the highlands of Costa Rica, the island junco *Junco insularis* on Guadalupe Island (recognized recently based on work by Aleixandre *et al.* 2013), the yellow-eyed junco *Junco phaeonotus* in the highlands of Mexico and Guatemala, and the dark-eyed junco *Junco hyemalis*, which inhabits conifer and broadleaf forests across temperate and boreal North America (Sullivan 1999; Nolan *et al.* 2002). The yellow-eyed junco is divided into five moderately differentiated subspecific taxa: *J. p. palliatus* of northern Mexico, *J. p. phaeonotus* of central Mexico, *J. p. fulvescens* of Chiapas in southern Mexico, *J. p. alticola* of Guatemala, and *J. p.*

bairdi of the southern tip of the Baja California Peninsula. In contrast, the dark-eyed junco presents a striking diversity of plumage pattern and color, and is composed of at least 14 distinct and largely allopatric morphotypes classified into four major groups: the slate-colored junco [*hyemalis* group] in eastern and boreal North America, composed of three subspecific taxa, *J. h. hyemalis*, *J. h. carolinensis* and *J. h. cismontanus*); white-winged junco *J. h. aikenii* in the Black Hills of South Dakota; Oregon junco [*oreganus* group] across the West, composed of several distinct forms from Alaska to northern Baja California, including *oreganus*, *shufeldti*, *montanus*, *thurberi*, *pinosus*, *pontilis*, *townsendi* and the pink-sided junco *mearnsi* in the northern Rocky Mountains; and gray-headed junco [*caniceps* group] in the Rocky Mountains and southwestern USA, composed of *J. h. caniceps* and *J. h. dorsalis* (Fig. 1A, Table 1).

Here, we use an extensive taxonomic sampling of the genus that includes all major phenotypic forms, and we reconstruct the evolutionary and demographic history of the complex to assess differences in patterns and rates of phenotypic diversification among *Junco* lineages. We also explore the relative roles of historical factors (geography, shifts in distribution range, and time in allopatry) and selective factors (ecology and sexual selection), in explaining differences in phenotypic diversification among clades. In a previous study, Milá *et al.* (2007) used spatial patterns of mitochondrial DNA (mtDNA) control region diversity to show that dark-eyed junco forms presented genetic signatures of recent population expansion coincident with the aftermath of the Last Glacial Maximum (LGM) about 18,000 years ago. This was in contrast to lineages in Guatemala (*alticola*), and Guadalupe Island (*insularis*), which despite their lower divergence in plumage color, turned out to be an order of magnitude older (Milá *et al.* 2007; Aleixandre *et al.* 2013), suggesting that different rates of phenotypic diversification have operated in different clades and different geographic

areas. Here, in addition to mtDNA markers, we use a large set of genome-wide single-nucleotide polymorphism (SNP) loci to achieve the resolution needed to reconstruct the evolutionary history of the group, and to test the dark-eyed junco post-glacial diversification hypothesis from a genomic perspective. We also use morphometric and colorimetric datasets to quantify phenotypic differences among forms, assess correlations between phenotypic and genetic distances, and infer the contribution of natural and sexual selection in the radiation by assessing the relative divergence of morphological traits, and plumage coloration, respectively.

Materials and methods

Population sampling

Junco populations were sampled across the range of the genus using mist nets in order to obtain biological samples for DNA extraction. Each captured individual was aged, sexed, and marked with a numbered aluminum band. A blood sample was collected by venipuncture of the sub-brachial vein and stored in Queen's lysis buffer or absolute ethanol at -80°C in the laboratory. After processing, birds were released unharmed at the site of capture. All sampling activities were conducted in compliance with Animal Care and Use Program regulations at the University of California Los Angeles, and with state and federal scientific collecting permits in the USA and Mexico. Tissue samples from two individuals of *Junco vulcani* were kindly loaned by the Louisiana Museum of Natural History at Louisiana State University (catalog numbers: LSU16242, LSU16243).

Sequencing of mitochondrial and nuclear markers

We extracted genomic DNA from blood samples using a Qiagen DNeasy kit (QiagenTM, Valencia, CA) and amplified four regions of the mtDNA from 212 samples representing all major junco forms (Table 2): 343 base pairs (bp) of the hypervariable region I of the control region (CR); 640 bp of the cytochrome *c* oxidase I (COI); 945 bp of the NADH dehydrogenase subunit 2 (ND2) and 870 bp of the ATPase genes 8 & 6 (ATPase). We also amplified 489 bp from intron 5 of the nuclear fibrinogen beta chain (FGB) gene for a subset of 110 individuals (Table 2). PCR products were purified with an ethanol precipitation and sequenced in an ABI 3730X automated sequencer. The amplification of nuclear copies of the control region is unlikely since we amplified it together with a coding fragment that was unambiguously translated into its aminoacid sequence. Sequences were aligned using Sequencher 4.1.4 (GeneCodes) and the accuracy of variable sites was checked visually on the individual chromatograms. Sequence alignments for each marker were imported to DnaSP 5.10.01 (Librado & Rozas 2009) and haplotype data files were generated for each of them independently. Nuclear intron sequences were phased using 1000 MCMC iterations as implemented by the program PHASE in the DnaSP package.

Genotyping of genome-wide SNP loci

We used genotyping-by-sequencing (Elshire *et al.* 2011) to obtain individual genotypes from 95 juncos belonging to the following taxa (with sample sizes in parentheses): *hyemalis* (7), *aikeni* (7), *thurberi* (8), *montanus* (7), *mearnsi* (7), *caniceps* (8), *dorsalis* (8), *palliatu*s (7), *phaeonotus* (8), *fulvescens* (3), *alticola* (7), *insularis* (7), *bairdi* (7), and *vulcani* (2) (Table 2). GBS libraries were prepared and analyzed at Cornell

University's Institute for Genomic Diversity (IGD), using restriction enzyme PstI for digestion (Baldassarre *et al.* 2014; Harvey & Brumfield 2015). Sequencing of the 95 individually-barcoded libraries was carried out in one lane of an Illumina HiSeq 2000, resulting in 191.5 million good single-end reads about 100 bp in length, and 28.4 million individual tags (unique sequences from one or more barcoded reads, Glaubitz *et al.* 2014) corresponding to 499,971 tag loci (individual tags with the same starting position and strand), out of which 261,481 were polymorphic. Sequencing of two of the 95 samples failed (1 *thurberi* and 1 *montanus*). We used the GBS analysis pipeline version 3.0.160 (Glaubitz *et al.* 2014), an extension to the Java program TASSEL (Bradbury *et al.* 2007), to call the SNPs from the sequenced GBS library, using as a reference the *Zonotrichia albicollis* genome [GenBank accession: ARWJ000000000] (Romanov *et al.* 2011). To account for genotyping errors, the TASSEL3-GBS pipeline excludes individual tags represented by less than five reads before genotyping. During variant discovery, loci for which less than 10% of the taxa had one or more tags were not used, and SNPs with a minor allele frequency (MAF) below 0.01 were excluded. Average sequencing error rate used to decide between heterozygous and homozygous calls was set by default to 0.1. We set a minimum sample call rate for each SNP of 0.8 and a minimum SNP call rate for each sample of 0.1. The resulting data matrix included 258,933 SNPs for 93 samples.

Phylogenetic analysis and molecular dating

We constructed a phylogeny using the concatenated mtDNA markers and the FGB intron 5, for a total of 110 individuals. The sequences were analyzed with Partition Finder (Lanfear *et al.* 2012) which found statistical support for partitioning by gene but not by codon position. We then used jModelTest (Posada 2008) for determining the

model of molecular evolution of each marker following the Bayesian Information Criterion (BIC), which were Hasegawa-Kishino-Yano with gamma distributed rate of among-site variation (HKY + G) for COI, ND2 and CR; Tamura-Nei (TrN) + G for ATPase; and HKY for the FGB intron. We simultaneously estimated tree topology and node divergence times through Bayesian inference using the program BEAST 1.8.1 (Drummond *et al.* 2012). A preliminary run using a lognormal relaxed clock model for all markers yielded moderate to low deviation from clock-like behavior for ATPase, CR and especially COI (Coefficient of variation = 0.185 as checked in TRACER v.1.6 (Rambaut *et al.* 2014), while ND2 and FGB intron 5 showed high substitution rate variability across the tree. Following Lerner *et al.* (2011), we fixed an evolutionary rate of 0.016 substitutions per site per million years for the COI gene under a strict clock model, while estimating the rates under a strict clock model for ATPase and CR, and a lognormal relaxed clock model for ND2 and FGB. A coalescent model was chosen over a birth-death model based on the path sampling (PS) and stepping stone (SS) sampling methods which have been shown to outperform the harmonic mean and Akaike's information criterion (AICM) to compute marginal likelihoods (Baele *et al.* 2012). Since we do not know the precise demographic history of the genus, we implemented the Bayesian Skyline coalescent model as available in BEAST, which can fit a wide range of demographic scenarios while taking into account phylogenetic uncertainty (Drummond *et al.* 2005). We set the number of groups to six based on the number of main lineages in the haplotype network (see Results) and used a piecewise-linear Skyline model. We ran the analysis for 400M iterations, sampling the chains every 8,000 steps. Convergence and effective sample sizes (ESS) were checked with TRACER v.1.6. We also constructed haplotype networks for all four mtDNA markers (COI, ATPase, ND2 and CR) using the median-joining algorithm (Bandelt *et al.* 1999) in the

program Network 4.6.1.2 (fluxus-engineering.com). As an alternative method for dating the postglacial expansion, we also estimated the time to the most recent common ancestor (TMRCA) of the young yellow-eyed and dark-eyed forms using the COI evolutionary rate in IMA2 (Hey 2010), and applying a pure isolation model that excluded migration. Based on preliminary runs, we set priors for maximum time of population splitting and maximum population size of 20 and 60, respectively. Generation time was set to 2 years since juncos do not usually breed successfully their first year. We run 40 chains of 30 million steps each, sampling every 100 steps after a 100,000-step burn-in, with heating in a geometric increment model with a term 1 of 0.96 and a term 2 of 0.9.

We calculated Nei's (Nei 1987) unbiased haplotypic diversity (h) and nucleotide diversity (π) indices for all lineages using the four mitochondrial markers in Arlequin 3.5.1.2 (Excoffier *et al.* 2005). To test for past sudden changes in population size we used Fu's test of neutrality, which detects departures from neutrality in scenarios with biased frequencies of rare alleles or young mutations in non-recombinant sequences (Fu 1997). We used Arlequin to obtain Fu's F_s values, and interpreted large and significant negative values as an excess of recent mutations caused by a recent population expansion. We also used Arlequin to compute the average number of pairwise differences corrected for intragroup variation for the ND2 mitochondrial marker.

Genome-wide structure from SNP data

To explore genome-wide population structure among recently diverged junco taxa and assess its correspondence with phenotypic groups, we ran a principal components analysis (PCA) using SNP data. We constructed a data matrix excluding older lineages

(*vulcani*, *bairdi*, *insularis*, *alticola* and *fulvescens*) as well as samples with a proportion of genotyped variable sites below 0.6, or with more than 20% heterozygous positions. Variant sites missing in more than 85% of the samples or presenting a minor allele frequency (MAF) below 0.03 were also removed. We filtered out SNPs putatively under selection using BayeScan (Foll & Gaggiotti 2008). Original *vcf* files were converted to BayeScan format using PGDSpider version 2.0.5.1 (Lischer & Excoffier 2012). We used default options in BayeScan and set the thinning interval size to 100 to ensure convergence. Under these conditions, the neutral model has a prior likelihood 10 times higher than the model with selection at any given locus. We defined junco populations according to phenotype (see Table 1), and for each SNP we obtained the posterior probability for the selection model and the F_{ST} coefficient averaged over populations. For outlier detection, we implemented a false discovery rate of 0.1. From a starting dataset of 101,743 loci, 597 significant outliers were excluded. Finally, to filter out the SNPs under linkage disequilibrium (LD) we used the function `snpGdsLDpruning` from the {SNPrelate} package (Zheng *et al.* 2012) in R version 3.2.2 (R Core Team 2015). We applied the correlation coefficient method with a threshold of 0.2 (`method="corr", ld.threshold=0.2`), resulting in a final data matrix of 24,832 SNPs. We then used the function `snpGdsPCA` also available in {SNPrelate} to perform the PCA and obtain the eigenvectors to be plotted.

Phylogenomic analyses with SNP data

To reconstruct phylogenetic relationships among junco forms using SNP data, we constructed two different data matrices, one including all taxa except the divergent volcano junco, and the other including the northern forms only (excluding older lineages *bairdi*, *insularis*, *alticola* and *fulvescens*). We applied the same quality filtering

controls than in the construction of the data matrix for the PCA (see above), except that the value for the threshold for missing genotyped sites was raised to 0.8 for the full-genus matrix. To fulfill neutrality assumptions for phylogenetic analyses, we filtered out SNPs putatively under selection from each matrix using the software BayeScan with the same settings used for the PCA. Once outliers were excluded, we created different matrices corresponding to different slices of the distribution of per-locus global- F_{ST} values among forms (Keller *et al.* 2013; Puebla *et al.* 2014). We tested the phylogenetic resolution of neutral SNPs using three different F_{ST} intervals: (i) 100% of the SNP dataset, (ii) top 25% F_{ST} SNPs, and (iii) top 10% F_{ST} SNPs (Table 3).

We constructed maximum likelihood phylogenies with RAxML 8.1.11 (Stamatakis 2014) in the CIPRES Science Gateway (Miller *et al.* 2010). We implemented a General Time Reversible (GTR) + gamma model of sequence evolution and applied the Lewis ascertainment bias correction (Stamatakis 2014) to the full genus tree which improved branch support. We used the rapid bootstrap algorithm (Stamatakis *et al.* 2008) with 100 replicates to compute a support value for the topology. We repeated the northern radiation phylogenetic reconstruction for the northern types including three samples of *fulvescens*, from Chiapas, so we could root the tree and discuss the evolutionary history of the group. Finally, to assess the significance of the phylogenetic signal in our datasets, we replicated the phylogenetic analysis on a matrix of randomized individuals assigned stochastically to an equal number of taxa, and using the top 25% F_{ST} SNPs to generate a maximum likelihood tree (Table 3).

To compare genetic and phenotypic distances, we computed a matrix of pairwise Nei's distances from the entire SNP dataset after filtering for selection using the package {adegenet} (Jombart 2008) in R version 3.2.2.

Phenotypic data and analysis

Morphological and colorimetric data were collected from 639 museum specimens representing all main junco forms, deposited at various natural history museums (Acknowledgements, Table S1 and Appendix I). A wing ruler was used to measure unflattened wing length to the nearest 0.5 mm, and dial calipers of 0.1-mm precision were used to measure tail length, tarsus length, bill culmen, exposed bill culmen, and bill width and depth, following Milá *et al.* (2008). All measurements were taken by a single observer (BM). We examined overall morphological differentiation among junco forms using male data in a discriminant function analysis (DFA) after transforming all variables using natural logarithms. Analyses were conducted in SPSS 17.0 (SPSS Inc. 2008).

We obtained reflectance spectra from the same museum skins used to measure morphological traits. We used a JAZ-EL200 spectrophotometer with a deuterium-tungsten light source via a bifurcate optical fiber probe (Ocean OpticsTM). The reflectance captor probe was mounted in a black rubber holder which excluded all external light and maintained the probe fixed at a distance of 3 mm from the feather surface at a 90° angle (as in Schmitz-Ornes 2006; Chui & Doucet 2009). The spectrum of each measurement ranged from 300 to 700 nm and consisted of three replicate measurements of three different readings per replicate, taken on each of six plumage patches: crown, nape, back, breast, flank and belly. Replicates were averaged before analysis. All reflectance data is expressed as the percentage of reflectance from a white standard (WS-1, Ocean OpticsTM). The white standard was measured after each specimen and the spectrophotometer was recalibrated regularly. All measurements were taken by a single observer (GF).

We obtained colorimetric variables by applying the avian visual model by Stoddard and Prum (2008), based on Goldsmith's (1990) tetrahedral color space for spectral data. We used the R-package {pavo} (Maia *et al.* 2013) to calculate the relative quantum catch for each cone using the function *vismodel*, and obtained the spherical coordinates describing the hue (Θ and ϕ) and the achieved chroma (r) using the function *tcs*. We included the normalized brilliance as a fourth variable, computed as described in Stoddard and Prum (2008). Variables were used in a DFA once again excluding females. Analyses were carried out with SPSS 17.0 and R version 3.2.2.

Phenotypic versus genetic distance

To compare genetic and phenotypic distances we used pairwise Nei's distances based on SNPs (see "*Phylogenomic analysis with SNP data*") along with Euclidean distances among group centroids from the DFAs of phenotypic data computed in SPSS 17.0. All distance values were rescaled to 1 and graphed in two heat-map color matrices using the function *image* from the R-package {graphics}. In these heat-map plots, the upper hemi-matrix depicts pairwise distances for phenotypic traits (morphology or color), and the lower hemi-matrix shows pairwise genetic distances based on SNPs. We also tested the correlation between genetic and Euclidean distances from morphometric and colorimetric traits by means of a Mantel test in the R-package {vegan} (Oksanen *et al.* 2005).

Results

Phylogeography and divergence times

Analysis of mtDNA and intron sequences revealed the existence of six well-supported phylogenetic clades corresponding, respectively, to junco populations in Costa Rica (*vulcani*), the tip of Baja California (*bairdi*), Guatemala (*alticola*), Guadalupe Island (*insularis*), Chiapas in southern Mexico (*fulvescens*), and a large clade composed of all remaining taxa from mainland Mexico, USA and Canada (Fig. 1B, Table 4). The form *bairdi* appears as the most basal lineage in the yellow-eyed/dark-eyed complex, with the clade including the rest of taxa divided into two major subclades, one formed by *alticola* and *insularis*, and the other by *fulvescens* and a single shallow clade containing all northern taxa. Molecular dating of the split between *fulvescens* and the northern clade yielded a mean value of 79,600 years before present [95% Highest Posterior Density interval (HPD): 41,300-121,800] while the TMRCA for the northern clade itself resulted in a mean value of 58,300 years (95% HPD: 31,300-88,600) when estimated with BEAST 1.8.1, and 16,132 years before present (95% HPD: 6,874-36,335) when using IMa2.

Haplotype networks for each of the mtDNA markers show topologies that are largely consistent with that of the Bayesian phylogeny (Figs. 2 and S1), although careful inspection reveals several interesting differences among genes in patterns of genetic structure. Thus, the ND2 gene provides clear separation of the older lineages (*vulcani*, *bairdi*, *insularis* and *alticola*) yet mainland yellow-eyed and dark-eyed junco forms are clustered into a single haplogroup (Fig. 2A). In contrast, ATPase genes show relatively less differentiation between some isolated lineages like *insularis* and *alticola*, but instead separate mainland yellow-eyed juncos (*phaeonotus*) from all dark-eyed junco forms to the north (Fig. 2B). The COI gene shows a similar pattern to that of ND2, with more marked divergence of isolated clades (Fig. S1A), and the CR, the only non-coding

region used, is the least informative marker at the lineage level, likely due to homoplasy resulting from its high mutation rate (Fig. S1B). A consistent pattern across markers is the presence of a “star-like” pattern of haplotypes in the north, with a high-frequency haplotype surrounded by several closely-related, low-frequency haplotypes, a pattern associated with a recent population expansion (Hewitt 1996).

Genetic diversity and historical demography

Sequencing of the four mitochondrial regions for 212 individuals produced 24 haplotypes for COI (640 bp), 28 haplotypes for ATPase (870 bp), 34 haplotypes for ND2 (985 bp), and 19 haplotypes for the non-coding CR (349 bp); resulting in 85 different haplotypes for the concatenated sequence (2844 bp). Indices of haplotype and nucleotide diversity were generally high (above 0.8 and 0.001 for h and π , respectively) except for three particularly small and isolated populations: *aikeni* ($h = 0.607$; $\pi = 0.0002$), *bairdi* ($h = 0.473$; $\pi = 0.0004$) and *fulvescens* ($h = 0.679$; $\pi = 0.0002$) (Table 5). Three of the taxa (*hyemalis*, *oreganus* and *phaeonotus*) showed evidence of population expansions as determined by significantly negative F_s values (Table 5).

Genome-wide structure from principal components analysis

A PCA of the entire SNP dataset revealed a pattern of clustering clearly consistent with the phenotypically divergent northern forms of junco (Figs. 3 and S2). The analysis recovered a clustering pattern remarkably congruent with both the geographic distribution and phenotypic characteristics of junco taxa. PCA revealed a clear separation between the yellow and dark-eyed juncos, but also between the northernmost forms *aikeni* and *hyemalis* with respect the rest of the groups. Structure was evident

even between the subspecific groups of the Oregon junco (*montanus* and *thurberi*) and those from the yellow-eyed junco (*palliatu*s and *phaeonotus*), showing relatively less separation among groups from the middle of the geographic distribution.

Phylogenomic reconstruction

A maximum likelihood phylogeny based on 32,674 selectively neutral SNP loci corresponding to the top 25% of the global- F_{ST} distribution was consistent with the mtDNA phylogeny in showing a marked contrast between isolated southern lineages and recently diverged forms in mainland North America (Fig. 4A). Forms *bairdi* and *insularis* showed the longest branches, whereas Guatemalan *alticola*, which appeared as sister to *insularis* in the mtDNA phylogeny, shows a shorter branch within the mainland cluster (Fig. 4A). Interestingly, in this phylogeny *fulvescens* from Chiapas appears as sister to *alticola*, whereas in the mtDNA phylogeny it clusters with the clade composed of mainland yellow-eyed juncos and dark-eyed juncos. The remaining taxa show a pattern of reciprocal monophyly, although node support values were low. To increase resolution among recently diverged northern forms, we analyzed these taxa separately (excluding *bairdi*, *insularis*, *alticola* and *fulvescens*). The resulting ML unrooted tree revealed a pattern of strong reciprocal monophyly among junco taxa with bootstrap values ranging from 87 to 100%, except the nodes for *dorsalis* and *mearnsi* which received values of 63% and 62%, respectively (Fig. 4B). Support for nodes among the main clades was generally strong, except for that between *hyemalis* and *aikeni*, which was 30%. The dataset that included only the top 10% of SNPs from the global F_{ST} distribution showed slightly reduced phylogenetic resolution, with one *caniceps* individuals grouping with the *mearnsi* group (Fig. 5C). The dataset that included all SNPs showed less resolution, separating yellow-eyed juncos from dark-eyed juncos and

resolving *dorsalis* and *hyemalis* forms while showing reduced resolution for the remaining recently diverged clades (Fig. 5A).

Rooting the SNP tree with *fulvescens* from Chiapas revealed a striking pattern of diversification consistent with a northward sequence of cladogenetic events (Fig. 6). Thus, relative to *fulvescens*, *phaeonotus* (central Mexico) appears as the most basal taxon, followed by *palliatu*s (northern Mexico), *dorsalis* (Arizona and New Mexico), *caniceps* (Rocky Mts.), and finally the most recently diversified taxa in the north (*oreganus*, *hyemalis* and *aikeni*) with gradually decreasing node support.

The phylogeny based on a randomized dataset (global Fst for top 25% = 0.036-0.15, Table 3) showed little structure and failed to produce monophyletic clusters among randomly assigned individuals (Fig. S3), indicating that the phylogenetic patterns recovered with the SNP datasets are robust.

Phenotypic differentiation and correlations with genetic distance

The plot of the two first discriminant functions from a DFA of morphometric variables revealed considerable differentiation among southern lineages, in contrast to northern forms, which generally showed extensive overlap (Fig. 7A). In contrast, colorimetric variables did show a clear separation between some northern taxa, yet widespread overlap among most of the older lineages (Fig. 7B), with the exception of *insularis* (GUJU), which clusters next to distantly-related *mearnsi* (PSJU), illustrating a case of evolutionary convergence in plumage coloration within the junco system as reported previously (Aleixandre *et al.* 2013). Heat-map plots illustrate the markedly different correlation between genetic distances and morphometric and colorimetric distances, respectively (Fig. 7C, D). The pattern is particularly striking for northern forms (upper

left quadrant of the matrices), which show little genetic divergence (yellow cells below the diagonal) and morphological divergence (predominantly yellow cells above the diagonal in Fig. 7C), and marked plumage color divergence (predominantly orange and red cells above the diagonal in Fig. 7D). A Mantel test revealed a significant correlation between matrices of genetic and morphological distances ($r = 0.595$, $P < 0.001$), but not between genetic and colorimetric distances ($r = -0.169$, $P = 0.837$).

Discussion

Young and old junco clades: rapid diversification vs. long-term stasis

Our results reveal the existence of closely related evolutionary lineages within the genus *Junco* that have diversified at strikingly different rates. Lineages that have been geographically isolated in oceanic or sky islands for over two hundred thousand years have diverged markedly from each other in neutral genetic markers, but relatively little in plumage color and pattern. In sharp contrast, the postglacial colonization of North America by one of the yellow-eyed lineages from southern Mexico, has caused a burst of diversification that has resulted in no less than six phenotypically and genetically distinct dark-eyed junco taxa. These young lineages differ most prominently in plumage coloration and patterning, both between each other and with respect to ancestral yellow-eyed taxa to the south, yet show relatively limited divergence in morphological traits. Morphological divergence in older lineages is marked but variable, with *bairdi* showing average characters compared to the significantly enlarged *alticola*, which inhabits the highlands of Guatemala and eastern Chiapas at nearly 4,000 m elevation, and the small size and elongated bill of *insularis* on Guadalupe Island, a species that has likely been

subjected to strong divergent selection to feed on a limited set of insular food types (Aleixandre *et al.* 2013). Overall, it is apparent that within juncos, species richness is clearly decoupled from clade age, and phenotypic diversification appears to show a better correlation with the rate of cladogenesis than with lineage age, patterns which have been shown to emerge at macroevolutionary scales (Rabosky & Adams 2012; Rabosky *et al.* 2012; Rabosky *et al.* 2013).

Understanding the micro-evolutionary mechanisms that give rise to such different rates of diversification within a single bird genus will require careful testing of the relative roles of natural and sexual selection, and even drift, in driving divergence, yet some prominent differences between northern and southern lineages seem relevant. One is that old southern lineages have remained isolated in areas with relatively unchanging habitat in climatically stable regions over many thousands of years, in contrast to young northern lineages, which have diversified across a large continental land mass spanning a latitudinal gradient with marked environmental variation. In both cases, the patchy distribution of suitable junco habitat in many parts of the junco range, where populations are isolated in high-elevation sky islands, may have promoted periods of complete isolation, where both selection and drift can act quickly to drive phenotypic divergence. In any case, our results are consistent with studies showing that the evolvability of clades may be related to their capacity to diversify across climatic gradients, as shown for example in the tanagers (Thraupidae), where species richness is correlated with climatic niche evolution (Title & Burns 2015), or in African lake cichlids, where a combination of ecological factors and sexual selection explain the probability of adaptive radiation in African lakes (Wagner *et al.* 2012).

Rapid postglacial speciation along a latitudinal environmental gradient

Our mtDNA and genome-wide datasets confirm a striking case of rapid phenotypic diversification in dark-eyed juncos from a yellow-eyed junco ancestor. Phylogeographic data based on several mtDNA coding regions reveal a clear population expansion of dark-eyed junco morphs across North America as evidenced by star-like haplotype phylogenies and significant tests of demographic expansion, confirming previous results obtained with smaller datasets (Milá *et al.* 2007). Numerous studies on avian taxa have documented similar postglacial expansions using phylogeographic and genetic diversity data (Seutin *et al.* 1995; Milá *et al.* 2000; 2006; Hansson *et al.* 2008; Malpica & Ornelas 2014; Alvarez *et al.* 2015), although none have shown levels of phenotypic diversification comparable to those in juncos. Our coalescence analyses provide estimates of the time since the junco expansion ranging from 41,300 to 121,800 years ago when measured at the *fulvescens*-northern juncos node, and from 6,800 to 88,600 years ago when computing the TMRCA for the northern taxa. However, given the date of Last Glacial Maximum 18,000 years ago, the IMA2 mean value estimate of 16,132 years ago is clearly more consistent with available geo-climatic data than the 58,300 years estimated with BEAST, which corresponds to a period when suitable junco habitat was lacking in most of the North American continent.

In addition to mtDNA sequence data, our analysis of genome-wide SNP loci provides unprecedented insight into the rapid junco radiation. As in previous studies using genome-wide SNPs to reconstruct the evolutionary history of recently radiated complexes, we obtained a better resolution of phylogenetic relationships among phenotypically differentiated taxa when using loci in the top 25% of a global- F_{ST} distribution than when including all available loci. Recent studies by Puebla *et al.* (2014) and Keller *et al.* (2013) obtained the best phylogenetic signal among recently diverged fish taxa when using only the top 10% and 25% of the global- F_{ST} distribution,

respectively. The lack of complete congruence between high- F_{ST} loci and the entire dataset is likely due to the stochastic nature of the coalescence process among loci, with the probability of finding loci with histories that are incongruent with the species tree increasing when large numbers of loci are used. Using high- F_{ST} loci among **genetically and phenotypically differentiated groups** (following the exclusion of outlier loci potentially under selection), is thus a way of reducing the influence of loci with incongruent histories due to incomplete lineage sorting, drift, divergent selection or balancing selection.

The resolution achieved in our **PCA and** phylogenomic analyses clarifies the evolutionary history of the juncos in three major ways. First, it confirms the young age of dark-eyed junco lineages relative to southern lineages, showing short-branched but reciprocally monophyletic clades for the different forms. This demonstrates that continent-wide sharing of mtDNA haplotypes in North America represents shared ancestral polymorphism due to recent divergence, and thus rules out the alternative explanation of introgressive hybridization and mtDNA capture in the early stages of the northward expansion, as has been shown to have occurred in other species with similar distributional histories (Milá *et al.* 2011). In fact, our SNP data reveal a **possible** case of past mtDNA introgressive hybridization in *fulvescens* of Chiapas, in southern Mexico. Individuals of *fulvescens* carry mtDNA haplotypes closely related to North American forms, yet genome-wide SNP data shows they are actually more closely related to *alticola* of Guatemala, which is more consistent with geography and phenotype. A likely scenario causing this example of mitochondrial discordance is the introduction of mtDNA haplotypes into the *fulvescens* population through incidental hybridization with a few *phaeonotus* females,

543 followed by the subsequent fixation of those haplotypes in drift-prone small highland
544 populations.

545 Second, the dark-eyed junco phylogeny based on neutral genome-wide SNPs
546 recovers with remarkable precision the south-to-north sequence of cladogenetic events,
547 from yellow-eyed forms in Mexico (*phaeonotus* and *palliatus*), through the first dark-
548 eyed forms in the southern USA (*dorsalis* and *caniceps*), to the most recently evolved
549 forms in the north (*hyemalis* and *oreganus*), once again confirming the role of a rapid
550 postglacial expansion in driving the fast radiation. Consistent with this scenario are the
551 node support values on the tree, which decrease gradually with increasing latitude,
552 showing the lowest support for the most recently evolved northern forms.

553 And third, the clear pattern of **genetic structure and even** reciprocal monophyly
554 among young dark-eyed junco clades indicates that they represent relatively
555 homogeneous genetic entities with limited gene flow among them. This degree of
556 differentiation among recently diverged forms is particularly striking given that the
557 SNPs sampled correspond to only about 0.4% of the junco's ~1.2 Gb genome, and that
558 outlier loci potentially under selection were filtered out for this analysis. Unlike other
559 avian models in which divergence appears to be limited to a reduced number of regions
560 or "genomic islands of divergence" (Ellegren *et al.* 2012; Poelstra *et al.* 2014) the clear
561 separation of taxa using a small subset of neutral genome-wide markers suggests that
562 divergence has taken place at the level of the entire genome. This may indicate a role
563 for local adaptation of junco forms to environmental conditions along the latitudinal
564 gradient, which is thought to result from relatively weak multifarious selection acting on
565 many loci across the genome as opposed to strong selection on a few loci (Dambroski &
566 Feder 2007; Michel *et al.* 2010; Parchman *et al.* 2013; Egea-Serrano *et al.* 2015). When
567 selection acts on many loci, the combined effect of divergent selection can drive a

global reduction in effective gene flow, thus favoring the establishment of new mutations and leading to “genome hitchhiking” (Feder & Nosil 2010; Feder *et al.* 2013), a process with the potential to drive genome-wide divergence even in the early stages of population divergence (Michel *et al.* 2010; Roesti *et al.* 2012). Thus, a relatively small number of markers may be sufficient to detect genome-wide divergence caused by multifarious selection driven by local adaptation. Alternatively, the low-level, genome-wide divergence found among junco forms could have been caused by drift if the original populations suffered strong reductions in effective population size during geographic isolation. The large number of differences encountered across the genome within a few thousand years, and the relatively slow **substitution** rate under drift compared to selection (even low-level multifarious selection), suggests drift may be a less parsimonious explanation for the observed patterns of variation, but higher-density genome scans will be necessary to further explore the genomic landscape of this unique radiation in order to assess the relative importance of drift, multifarious selection, and strong selection acting on specific loci across the junco genome.

The junco radiation stands out as one of the fastest known in vertebrates. Other famous avian radiations like those of *Zosterops* white-eyes (Moyle *et al.* 2009), *Setophaga* warblers (Lovette *et al.* 2010), or Hawaiian drosophilid honeycreepers (Lerner *et al.* 2011), gave rise to vast number of species, but over relatively longer periods spanning several million years. One case that is remarkably comparable in timescale and number of resulting taxa is that of southern capuchinos, a clade of South American *Sporophila* seedeaters, thought to have originated within the last 40,000 years and whose eight members differ conspicuously in melanic plumage coloration

(Campagna *et al.* 2015). Unlike juncos, however, most morphs are largely sympatric across ecologically similar habitat, and signal for genetic structure using similar genome-wide markers to the ones used here, is very weak (Campagna *et al.* 2015). This lack of phylogenetic signal in random genome-wide SNPs points to a major role for a few loci of strong effect in determining southern capuchino phenotypes, with the rest of the genome showing no differentiation due to either shared ancestral polymorphism or widespread gene flow in unselected genomic regions. Future research into the evolutionary mechanisms involved and the respective genomic signatures in both systems, will help us understand the relative roles of gene flow, selection and historical context in driving fast divergence and speciation.

A potential role for sexual selection

The marked differences in plumage pattern and color among recently diversified dark-eyed junco morphs suggest the role of sexual selection, an important mechanism in divergence and speciation (Lande 1981; West-Eberhard 1983; Wilson *et al.* 2000; Coyne & Orr 2004; Ritchie 2007). Evidence for the direct role of sexual selection in modifying junco phenotypes is limited, but studies at the population level have shown that sexual selection can modify the amount of white on tail feathers in slate-colored and Oregon juncos, a trait known to be involved in mate choice (Hill *et al.* 1999; Yeh 2004; McGlothlin *et al.* 2005; Price *et al.* 2008). At the level of the entire complex, a testable prediction of the action of sexual selection in evolutionary divergence is the appearance of sexual dimorphism in secondary sexual characters (Darwin 1872). Based on the examination of thousands of junco specimens, Miller (1941) assessed the amount of sexual dimorphism in head color among the different junco taxa, and he concluded that dimorphism appears to increase with latitude: it is absent in *vulcani*, *bairdi*,

alticola, and *fulvescens*, very low in *phaeonotus*, *palliatu*s, *dorsalis*, and *insularis*, slight in *caniceps*, and clearly present in the various *oreganus* and *hyemalis* taxa. Although it remains to be properly quantified, this pattern would be congruent with both latitude and the phylogenetic pattern we have reported, suggesting that sexual selection may have played a role in the fast diversification of northern forms. Indeed, an interaction between sexual selection and local adaptation has been shown to promote speciation in recent models, showing that sexual ornaments can function as indicators of local adaptation and increase disruptive selection driven by ecological factors (van Doorn *et al.* 2009).

Taxonomic implications

Diagnosability through plumage coloration has been and remains the single most important criterion in avian species designation (Watson 2005; Sangster *et al.* 2011) . This is the case even though phenotypic divergence has often been shown to be a poor predictor of evolutionary history (Olsson *et al.* 2005; Navarro-Sigüenza *et al.* 2008; Gill 2014). Juncos are a prime example, and our results reveal major incongruences between the group's evolutionary history and current taxonomy. Traditional taxonomic classifications based on eye color (yellow vs. dark) are clearly paraphyletic, and counterintuitively, the more uniform-looking yellow-eyed juncos in the south turned out to be the most evolutionarily divergent lineages, with the most distinct forms among dark-eyed juncos of North America being the youngest. Even before these findings, junco taxonomy has never been simple, and the lack of consensus on how junco diversity must be classified has led to a convoluted taxonomic history (Ketterson & Atwell 2016; Milá *et al.* 2016).

Based on a combination of genetic and phenotypic data, the island junco (*Junco insularis*) from Guadalupe Island was proposed as a separate species in a previous study (Aleixandre *et al.* 2013), and it has been recently adopted by the American Ornithologists' Union (Chesser *et al.* 2014) and the IOC World Bird List (Gill & Donsker 2016). Based on similar criteria, Baird's junco (*J. bairdi*), the Guatemala junco (*J. alticola*), the Chiapas junco (*J. fulvescens*), and several dark-eyed junco forms, may be also recognized as separate species in due time. Specifically, recently diverged dark-eyed junco forms breeding in the USA and Canada, currently grouped into a single species (*Junco hyemalis*), will have to undergo a thorough reexamination of geographic variation in phenotypic traits in light of the genomic data presented here, and we believe this will likely result in the recognition of at least four species-level taxa, namely the gray-headed junco (*J. caniceps*), the Oregon junco (*J. oreganus*), the white-winged junco (*J. aikenii*) and the slate-colored junco (*J. hyemalis*). This taxonomic arrangement would in fact be quite similar to the one proposed by Alden Miller (1941), whose keen observations and tireless efforts may well be vindicated now, over seventy years later, by the genomic data he lacked. However, given the complex patterns of phenotypic variation at different geographic scales and the varying degrees of intergradation among some forms within the group, issuing conclusive and detailed taxonomic recommendations will require in-depth descriptions, analyses and discussions that are beyond the scope of the present paper, and thus will be presented elsewhere.

Conclusion

Our analyses confirm a striking case of rapid postglacial speciation in the genus *Junco* and reveal marked differences in the rate and degree of phenotypic diversification among clades. The rapid diversification of distinctly colored dark-eyed junco forms across North America took place since the Last Glacial Maximum just 18,000 years ago. In contrast, isolated lineages in the south show relative phenotypic stasis, with plumage color showing little change over hundreds of thousands of years. The pattern of genome-wide divergence among northern forms despite their recent divergence, suggests that in addition to color differences, which are known to be controlled by a small number of pigmentation loci, junco forms have diverged in many other traits likely related to adaptation to local environmental conditions. In-depth genome scans will be necessary to identify genomic regions under selection and reveal the genomic architecture of rapid diversification. Our within-genus study reveals a compelling example of speciation-in-action in juncos, where the combined effect of historical and selective factors has produced one of the fastest cases of speciation known in vertebrates.

Acknowledgements

We thank Jonathan Atwell, Elena Berg, Steve Burns, Jatziri Calderón, Adrián Gutiérrez, Alfonsina Hernández, Fritz Hertel, Ellen Ketterson, John McCormack, Adán Oliveras de Ita, César Ríos, Sahid Robles, Vicente Rodríguez, Whitney Tsai, Rich Van Buskirk and Alvar Veiga for invaluable help in the field, and Boris Delahaie for his kind assistance with colorimetric analysis. We are grateful to the following museum curators and collection managers for allowing us access to their junco specimens: Philip Unitt at

the San Diego Natural History Museum (SDMNH), Kimball Garrett at Los Angeles Museum of Natural History (LAMNH), Carla Cicero at the Museum of Vertebrate Zoology (MVZ), John McCormack and James Maley at The Moore Laboratory of Zoology at Occidental College (MLZ), Chris Milensky at The National Museum of Natural History (NMNH), and Paul Sweet at The American Museum of Natural History (AMNH). Thanks to Robb Brumfield at the Louisiana Museum of Natural History at Louisiana State University for loaning us two volcano junco tissue samples. A special research permit to sample on Guadalupe Island was issued by the Secretaría de Medio Ambiente y Recursos Naturales of Mexico (Permit No. SGPA/DGVVS/01217/11). Funding was provided in part by a grant from the Spanish Ministry of Science and Innovation (CGL-2011-25866) to BM, and a CSIC-CONACyT bilateral Spain-Mexico project (2008MX0009) to BM and AGNS.

References

- Aleixandre P, Hernández Montoya J, Milá B (2013) Speciation on oceanic islands: Rapid adaptive divergence vs. cryptic speciation in a Guadalupe Island songbird (Aves: *Junco*). *PLoS ONE*, **8**, e63242.
- Alvarez S, Salter JF, McCormack JE, Milá B (2015) Speciation in mountain refugia: phylogeography and demographic history of the pine siskin and black-capped siskin complex. *Journal of Avian Biology*, **47**, 335-345.
- Baele G, Lemey P, Bedford T, Rambaut A, Suchard MA, Alekseyenko AV (2012) Improving the accuracy of demographic and molecular clock model comparison while

714 accommodating phylogenetic uncertainty. *Molecular Biology and Evolution*, **29**, 2157-
715 2167.

716 Baldassarre DT, White TA, Karubian J, Webster MS (2014) Genomic and
717 morphological analysis of a semipermeable avian hybrid zone suggests asymmetrical
718 introgression of a sexual signal. *Evolution*, **68**, 2644-2657.

719 Bandelt HJ, Forster P, Rohl A (1999) Median-joining networks for inferring
720 intraspecific phylogenies. *Molecular Biology and Evolution*, **16**, 37-48.

721 Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007)
722 TASSEL: software for association mapping of complex traits in diverse samples.
723 *Bioinformatics*, **23**, 2633-2635.

724 Campagna L, Gronau I, Silveira LF, Siepel A, Lovette IJ (2015) Distinguishing noise
725 from signal in patterns of genomic divergence in a highly polymorphic avian radiation.
726 *Molecular Ecology*, **24**, 4238-4251.

727 Chesser RT, Banks RC, Cicero C, et al (2014) Fifty-fifth supplement to the American
728 Ornithologists' Union check-list of North American birds. *The Auk*, **131**, CSi-CSxv.

729 Chui CKS, Doucet SM (2009) A test of ecological and sexual selection hypotheses for
730 geographical variation in coloration and morphology of golden-crowned kinglets
731 (*Regulus satrapa*). *Journal of Biogeography*, **36**, 1945-1957.

732 Coyne JA, Orr HA (2004) *Speciation* Sinauer Associates, Inc, Sunderland,
733 Massachusetts.

734 Currie DJ, Mittelbach GG, Cornell HV, et al (2004) Predictions and tests of climate-
735 based hypotheses of broad-scale variation in taxonomic richness. *Ecology Letters*, **7**,
736 1121-1134.

737 Dambroski HR, Feder JL (2007) Host plant and latitude-related diapause variation in
738 *Rhagoletis pomonella*: a test for multifaceted life history adaptation on different stages
739 of diapause development. *Journal of Evolutionary Biology*, **20**, 2101-2112.

740 Darwin C (1872) *The descent of man, and selection in relation to sex* D. Appleton.

741 Drummond AJ, Rambaut A, Shapiro B, Pybus OG (2005) Bayesian coalescent
742 inference of past population dynamics from molecular sequences. *Molecular Biology*
743 *and Evolution*, **22**, 1185-1192.

744 Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian phylogenetics with
745 BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, **29**, 1969-1973.

746 Egea-Serrano A, Hangartner S, Laurila A, Räsänen K (2015) Multifarious selection
747 through environmental change: acidity and predator-mediated adaptive divergence in
748 the moor frog (*Rana arvalis*). *Proceedings of the Royal Society of London B: Biological*
749 *Sciences*, **281**, 20133266.

750 Ellegren H, Smeds L, Burri R, et al (2012) The genomic landscape of species
751 divergence in *Ficedula* flycatchers. *Nature*, **491**, 756-760.

752 Elshire RJ, Glaubitz JC, Sun Q, et al (2011) A robust, simple Genotyping-by-
753 Sequencing (GBS) approach for high diversity species. *Plos One*, **6**, e19379.

754 Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): An integrated
 755 software package for population genetics data analysis. *Evolutionary Bioinformatics*, **1**,
 756 47-50.

757 Feder JL, Flaxman SM, Egan SP, Comeault AA, Nosil P (2013) Geographic mode of
 758 speciation and genomic divergence. *Annual Review of Ecology, Evolution, and*
 759 *Systematics*, **44**, 73-97.

760 Feder JL, Nosil P (2010) The efficacy of divergence hitchhiking in generating genomic
 761 islands during ecological speciation. *Evolution*, **64**, 1729-1747.

762 Foll M, Gaggiotti O (2008) A Genome-Scan method to identify selected loci
 763 appropriate for both dominant and codominant markers: A Bayesian perspective.
 764 *Genetics*, **180**, 977-993.

765 Fu YX (1997) Statistical tests of neutrality of mutations against population growth,
 766 hitchhiking and background selection. *Genetics*, **147**, 915-925.

767 Gill FB (2014) Species taxonomy of birds: Which null hypothesis? *The Auk*, **131**, 150-
 768 161.

769 Gill, F & D Donsker, Eds. (2016) IOC World Bird List (v 6.2). *doi* :
 770 *10.14344/IOC.ML.6.2*.

771 Glaubitz JC, Casstevens TM, Lu F, Harriman J, Elshire RJ, Sun Q, Buckler ES (2014)
 772 TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. *Plos One*,
 773 **9**, e90346.

774 Goldsmith TH (1990) Optimization, constraint, and history in the evolution of eyes.
775 *Quarterly Review of Biology*, **65**, 281-322.

776 Hansson B, Hasselquist D, Tarka M, Zehindjiev P, Bensch S (2008) Postglacial
777 colonisation patterns and the role of isolation and expansion in driving diversification in
778 a passerine bird. *PloS One*, **3**, e2794.

779 Harvey MG, Brumfield RT (2015) Genomic variation in a widespread Neotropical bird
780 (*Xenops minutus*) reveals divergence, population expansion, and gene flow. *Molecular*
781 *Phylogenetics and Evolution*, **83**, 305-316.

782 Hewitt GM (1996) Some genetic consequences of ice ages, and their role in divergence
783 and speciation. *Biological Journal of the Linnean Society*, **58**, 247-276.

784 Hey J (2010) Isolation with migration models for more than two populations. *Molecular*
785 *Biology and Evolution*, **27**, 905-920.

786 Hill JA, Enstrom DA, Ketterson ED, Nolan V, Jr, Ziegenfus C (1999) Mate choice
787 based on static versus dynamic secondary sexual traits in the dark-eyed junco.
788 *Behavioral Ecology*, **10**, 91-96.

789 Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic
790 markers. *Bioinformatics*, **24**, 1403-1405.

791 Keller I, Wagner CE, Greuter L, et al (2013) Population genomic signatures of
792 divergent adaptation, gene flow and hybrid speciation in the rapid radiation of Lake
793 Victoria cichlid fishes. *Molecular Ecology*, **22**, 2848-2863.

794 Ketterson ED, Atwell JW (2016) The junco: a common bird and a classic subject for
 795 descriptive and experimental studies in evolutionary and integrative biology. In:
 796 *Snowbird: Integrative Biology and Evolutionary Diversity in the Junco* (eds. Ketterson
 797 ED, Atwell JW) Chicago University Press, Chicago.

798 Lande R (1981) Models of Speciation by Sexual Selection on Polygenic Traits.
 799 *Proceedings of the National Academy of Sciences of the United States of America-*
 800 *Biological Sciences*, **78**, 3721-3725.

801 Lanfear R, Calcott B, Ho SYW, Guindon S (2012) PartitionFinder: Combined selection
 802 of partitioning schemes and substitution models for phylogenetic analyses. *Molecular*
 803 *Biology and Evolution*, **29**, 1695-1701.

804 Lerner HRL, Meyer M, James HF, Hofreiter M, Fleischer RC (2011) Multilocus
 805 resolution of phylogeny and timescale in the extant adaptive radiation of hawaiian
 806 honeycreepers. *Current Biology*, **21**, 1838-1844.

807 Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA
 808 polymorphism data. *Bioinformatics*, **25**, 1451-1452.

809 Lischer HEL, Excoffier L (2012) PGDSpider: an automated data conversion tool for
 810 connecting population genetics and genomics programs. *Bioinformatics*, **28**, 298-299.

811 Lovette IJ, Bermingham E, Ricklefs RE (2002) Clade-specific morphological
 812 diversification and adaptive radiation in Hawaiian songbirds. *Proceedings of the Royal*
 813 *Society B*, **269**, 37-42.

814 Lovette IJ, Pérez-Emán JL, Sullivan JP, et al (2010) A comprehensive multilocus
 815 phylogeny for the wood-warblers and a revised classification of the Parulidae (Aves).
 816 *Molecular Phylogenetics and Evolution*, **57**, 753-770.

817 Mahler DL, Revell LJ, Glor RE, Losos JB (2010) Ecological opportunity and the rate of
 818 morphological evolution in the diversification of greater antillean anoles. *Evolution*, **64**,
 819 2731-2745.

820 Maia R, Eliason CM, Bitton P, Doucet SM, Shawkey MD (2013) pavo: an R package
 821 for the analysis, visualization and organization of spectral data. *Methods in Ecology and*
 822 *Evolution*, **4**, 906-913.

823 Malpica A, Ornelas JF (2014) Postglacial northward expansion and genetic
 824 differentiation between migratory and sedentary populations of the broad-tailed
 825 hummingbird (*Selasphorus platycercus*). *Molecular Ecology*, **23**, 435-452.

826 McGlothlin JW, Parker PG, Nolan Jr. V, Ketterson ED (2005) Correlational selection
 827 leads to genetic integration of body size and an attractive plumage trait in dark-eyed
 828 juncos. *Evolution*, **59**, 658-671.

829 Michel AP, Sim S, Powell THQ, Taylor MS, Nosil P, Feder JL (2010) Widespread
 830 genomic divergence during sympatric speciation. *Proceedings of the National Academy*
 831 *of Sciences*, **107**, 9724-9729.

832 Milá B, Aleixandre P, Alvarez-Nordstrom S, McCormack JE (2016) More than meets
 833 the eye: lineage diversity and evolutionary history of dark-eyed and yellow-eyed juncos.
 834 In: *Snowbird: Integrative Biology and Evolutionary Diversity in the Junco* (eds.
 835 Ketterson ED, Atwell JW) Chicago University Press, Chicago.

836 Milá B, Girman D, Kimura M, Smith T (2000) Genetic evidence for the effect of a
837 postglacial population expansion on the phylogeography of a North American songbird.
838 *Proceedings of the Royal Society B*, **267**, 1033-1040.

839 Milá B, Toews DPL, Smith TB, Wayne RK (2011) A cryptic contact zone between
840 divergent mitochondrial DNA lineages in southwestern North America supports past
841 introgressive hybridization in the yellow-rumped warbler complex (Aves: *Dendroica*
842 *coronata*). *Biological Journal of the Linnean Society*, **103**, 696-706.

843 Milá B, Wayne RK, Smith TB (2008) Ecomorphology of migratory and sedentary
844 populations of the yellow-rumped warbler (*Dendroica Coronata*). *Condor*, **110**, 335-
845 344.

846 Milá B, McCormack JE, Castaneda G, Wayne RK, Smith TB (2007) Recent postglacial
847 range expansion drives the rapid diversification of a songbird lineage in the genus
848 *Junco*. *Proceedings of the Royal Society B*, **274**, 2653-2660.

849 Milá B, Smith TB, Wayne RK (2006) Postglacial population expansion drives the
850 evolution of long-distance migration in a songbird. *Evolution*, **60**, 2403-2409.

851 Miller M, Pfeiffer W, Schwartz T (2010) Creating the CIPRES Science Gateway for
852 inference of large phylogenetic trees. Gateway Computing Environments Workshop, 1-
853 8.

854 Miller AH (1941) Speciation in the avian genus *Junco*. *University of California*
855 *Publications in Zoology*, **44**, .173-434.

856 Moyle RG, Filardi CE, Smith CE, Diamond J (2009) Explosive Pleistocene
857 diversification and hemispheric expansion of a "great speciator". *Proceedings of the*
858 *National Academy of Sciences of the United States of America*, **106**, 1863-1868.

859 Navarro-Sigüenza AG, Peterson AT, Nyari A, García-Deras GM, García-Moreno J
860 (2008) Phylogeography of the *Buarremon* brush-finch complex (Aves, Emberizidae) in
861 Mesoamerica. *Molecular Phylogenetics and Evolution*, **47**, 21-35.

862 Nei M (1987) *Molecular evolutionary genetics*. Columbia university press.

863 Nolan V,Jr, Ketterson ED, Cristol DA, et al (2002) Dark-eyed Junco (*Junco hyemalis*).
864 In: *The Birds of North America* (eds. Poole A, Gill F) The Birds of North America, Inc,
865 Philadelphia, Pennsylvania.

866 Oksanen J, Kindt R, O'Hara B (2005) Vegan: R functions for vegetation ecologists.
867 URL: <http://cc.oulu.fi/jarioksa/softhelp/vegan.html>, .

868 Olsson U, Alström P, Ericson PG, Sundberg P (2005) Non-monophyletic taxa and
869 cryptic species—evidence from a molecular phylogeny of leaf-warblers (*Phylloscopus*,
870 Aves). *Molecular Phylogenetics and Evolution*, **36**, 261-276.

871 Parchman T, Gompert Z, Braun M, et al (2013) The genomic consequences of adaptive
872 divergence and reproductive isolation between species of manakins. *Molecular Ecology*,
873 **22**, 3304-3317.

874 Phillimore AB, Orme CDL, Davies RG, et al (2007) Biogeographical basis of recent
875 phenotypic divergence among birds: A global study of subspecies richness. *Evolution*,
876 **61**, 942-957.

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

879 Posada D (2008) jModelTest: Phylogenetic model averaging. *Molecular Biology and*
880 *Evolution*, **25**, 1253-1256.

881 Price TD, Yeh PJ, Harr B (2008) Phenotypic plasticity and the evolution of a socially
882 selected trait following colonization of a novel environment. *The American Naturalist*
883 **172**, S49-S62.

884 Puebla O, Bermingham E, McMillan WO (2014) Genomic atolls of differentiation in
885 coral reef fishes (*Hypoplectrus spp.*, Serranidae). *Molecular Ecology*, **23**, 5291-5303.

886 R Core Team (2015) R: A language and environment for statistical computing. R
887 Foundation for Statistical Computing, Vienna, Austria. URL [https://www.R-](https://www.R-project.org/)
888 [project.org/](https://www.R-project.org/).

889 Rabosky DL, Adams DC (2012) Rates of morphological evolution are correlated with
890 species richness in salamanders. *Evolution*, **66**, 1807-1818.

891 Rabosky DL (2009) Ecological limits and diversification rate: alternative paradigms to
892 explain the variation in species richness among clades and regions. *Ecology Letters*, **12**,
893 735-743.

894 Rabosky DL, Slater GJ, Alfaro ME (2012) Clade age and species richness are
895 decoupled across the eukaryotic tree of life. *PLoS Biol*, **10**, e1001381.

896 Rabosky D, Santini F, Eastman J, et al (2013) Rates of speciation and morphological
 897 evolution are correlated across the largest vertebrate radiation. *Nature Communications*,
 898 **4**, 1958.

899 Rambaut A, Suchard M, Xie D, Drummond A (2014) Tracer v1. 6. *Computer program*
 900 *and documentation distributed by the author, website* <http://beast.bio.ed.ac.uk/Tracer>.

901 Ricklefs RE (2006) Global variation in the diversification rate of passerine birds.
 902 *Ecology*, **87**, 2468-2478.

903 Ritchie MG (2007) Sexual selection and speciation. *Annual Review of Ecology*,
 904 *Evolution, and Systematics*, **38**, 79-102.

905 Roesti M, Hendry AP, Salzburger W, Berner D (2012) Genome divergence during
 906 evolutionary diversification as revealed in replicate lake-stream stickleback population
 907 pairs. *Molecular Ecology*, **21**, 2852-2862.

908 Romanov MN, Dodgson JB, Gonser RA, Tuttle EM (2011) Comparative BAC-based
 909 mapping in the white- throated sparrow, a novel behavioral genomics model, using
 910 interspecies overgo hybridization. *BMC Research Notes*, **4**, 211.

911 Sangster G, Collison J, Crochet P, et al (2011) Taxonomic recommendations for British
 912 birds: seventh report. *Ibis*, **153**, 883-892.

913 Schmitz-Ornes A (2006) Using colour spectral data in studies of geographic variation
 914 and taxonomy of birds: examples with two hummingbird genera, *Anthracothorax* and
 915 *Eulampis*. *Journal of Ornithology*, **147**, 495-503.

916 Seutin G, Ratcliffe LM, Boag PT (1995) Mitochondrial DNA homogeneity in the
 917 phenotypically diverse redpoll finch complex (Aves: Carduelinae: *Carduelis flammea-*
 918 *hornemanni*). *Evolution*, **49**, 962-973.

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

921 Stamatakis A, Hoover P, Rougemont J (2008) A rapid bootstrap algorithm for the
 922 RAxML web servers. *Systematic Biology*, **57**, 758-771.

923 Stoddard MC, Prum RO (2008) Evolution of avian plumage color in a tetrahedral color
 924 space: A phylogenetic analysis of new world buntings. *American Naturalist*, **171**, 755-
 925 776.

926 Sullivan KA (1999) Yellow-eyed junco (*Junco phaeonotus*). In: *The Birds of North*
 927 *America* (eds. Poole A, Gill F) The Birds of North America Inc, Philadelphia, PA.

928 Title PO, Burns KJ (2015) Rates of climatic niche evolution are correlated with species
 929 richness in a large and ecologically diverse radiation of songbirds. *Ecology Letters*, **18**,
 930 433-440.

931 van Doorn GS, Edelaar P, Weissing FJ (2009) On the origin of species by natural and
 932 sexual selection. *Science (New York, N.Y.)*, **326**, 1704-1707.

933 Wagner C, Harmon L, Seehausen O (2012) Ecological opportunity and sexual selection
 934 together predict adaptive radiation. *Nature*, **487**, 366-369.

935 Watson DM (2005) Diagnosable versus distinct: evaluating species limits in birds.
 936 *Bioscience*, **55**, 60-68.

937 West-Eberhard MJ (1983) Sexual selection, social competition and speciation.
 938 *Quarterly Review of Biology*, **58**, 155-183.

939 Wilson AB, Noack-Kunnmann K, Meyer A (2000) Incipient speciation in sympatric
 940 Nicaraguan crater lake cichlid fishes: sexual selection versus ecological diversification.
 941 *Proceedings of the Royal Society of London B: Biological Sciences*, **267**, 2133-2141.

942 Yeh PJ (2004) Rapid evolution of a sexually selected trait following population
 943 establishment in a novel habitat. *Evolution*, **58**, 166-174.

944 Yoder JB, Clancey E, Des Roches S, et al (2010) Ecological opportunity and the origin
 945 of adaptive radiations. *Journal of Evolutionary Biology*, **23**, 1581-1596.

946 Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS (2012) A high-
 947 performance computing toolset for relatedness and principal component analysis of
 948 SNP data. *Bioinformatics (Oxford, England)*, **28**, 3326-3328.

949

950 **Data Accessibility:**

951 Sequences deposited in GenBank (accessions KX460825 - KX461036; KX461037 -
 952 KX461248; KX461249 - KX461460; KX461461 - KX461570; KX461571 -
 953 KX461782); Datasets deposited in Dryad doi:10.5061/dryad.30374

954

955 **Author Contributions:**

956 GF and BM designed the study; GF, PA, RRE, AGNS and BM conducted field work
 957 and collected phenotypic data; GF, PA generated molecular data; GF, PA and BM
 958 analyzed the data; GF and BM wrote the paper with input from all authors.

Table 1. Taxonomy of junco forms based on Miller (1941). Those forms included in this study are shown in bold. The four taxa reported under “Species” are those currently recognized by the American Ornithologists’ Union (2016).

Country	Species	Forms	Common name	Code	
USA and Canada	<i>J. hyemalis</i>	<i>hyemalis</i> <i>cismontanus</i> <i>carolinensis</i>	Slate-colored junco	SCJU	“Dark-eyed junco” (DEJU)
		<i>aikeni</i>	White-winged junco	WWJU	
		<i>caniceps</i>	Gray-headed junco	GHJU	
		<i>dorsalis</i>	Red-backed junco	RBJU	
		<i>mearnsi</i>	Pink-sided junco	PSJU	
		<i>oreganus</i> <i>shufeldti</i> <i>montanus</i> <i>pinosus</i> <i>thurberi</i>	Oregon junco	ORJU	
Mexico	<i>J. hyemalis</i> <i>insularis</i> <i>J. phaeonotus</i>	<i>pontilis</i> <i>townsendi</i>			
		<i>insularis</i>	Guadalupe junco	GUJU	
		<i>phaeonotus</i> <i>bairdi</i> <i>fulvescens</i>	Yellow-eyed junco	YEJU	“Yellow-eyed junco” (YEJU)
				BAJU	
				CHJU	
Guatemala	<i>alticola</i>		GTJU		
Costa Rica	<i>J. vulcani</i>	<i>vulcani</i>	Volcano junco	VOJU	

Table 2. Number of genotyped individuals per junco form and molecular marker. State abbreviations are the following: Virginia (VA), Wyoming (WY), Alaska (AK), British Columbia (BC), California (CA), Idaho (ID), Utah (UT), and Arizona (AZ) in the USA; Coahuila (COAH), Mexico City (CM), Durango (DGO), Michoacán (MICH), Nuevo León (NL), Oaxaca (OAX), and Chiapas (CHIS) in Mexico; and Huehuetenango (HUE) in Guatemala.

<i>Group</i>	<i>State</i>	<i>mtDNA</i>	<i>FGB-I5</i>	<i>SNPs</i>
<i>hyemalis</i>	VA	6	16	7
<i>aikeni</i>	WY	8	6	7
<i>oreganus</i>	AK, BC, CA, ID	36	18	15
<i>mearnsi</i>	WY	8	3	7
<i>caniceps</i>	UT	9	4	8
<i>dorsalis</i>	AZ	10	4	8
<i>phaeonotus</i>	AZ, COAH, CM, DGO, MICH, NL, OAX	76	19	15
<i>fulvescens</i>	CHIS	8	5	3
<i>insularis</i>	Guadalupe Is.	23	12	7
<i>bairdi</i>	BCS	11	11	8
<i>alticola</i>	HUE	15	10	8
<i>vulcani</i>	Costa Rica	2	2	2
Total		212	110	95

Table 3. Data matrices used in phylogenomic analysis. Shown for each matrix are number of samples, number of SNP loci and F_{ST} ranges for datasets used in phylogenomic reconstructions.

Dataset	Number of samples	F_{ST} percentile	Number of SNPs	F_{ST} range
Full genus (except <i>vulcani</i>)	79	100%	130,700	0.085-0.371
		Top 25%	32,674	0.180-0.371
		Top 10%	13,070	0.186-0.371
Northern forms	64	100%	101,145	0.053-0.205
		Top 25%	25,286	0.063-0.205
		Top 10%	10,114	0.067-0.205
Northern forms and <i>fulvescens</i>	67	100%	105,572	0.061-0.258
		Top 25%	26,393	0.086-0.258
		Top 10%	10,557	0.090-0.258
Randomized dataset	64	Top 25%	25,387	0.036-0.150

Table 4. Genetic distances among junco forms. Nei's distances based on SNP data from genotyping-by-sequencing (above diagonal) and the average number of pairwise changes among forms corrected for intragroup variation for the ND2 gene (below diagonal, with intra-population divergence values along diagonal, in bold **italics**).

	<i>hyemalis</i>	<i>aikeni</i>	<i>mearnsi</i>	<i>oreganus</i>	<i>thurberi</i>	<i>montanus</i>	<i>caniceps</i>	<i>dorsalis</i>	<i>pallidus</i>	<i>phaeonotus</i>	<i>fulvescens</i>	<i>alticola</i>	<i>insularis</i>	<i>bairdi</i>	<i>vulcani</i>
<i>hyemalis</i>	0.933	0.079	0.075	-	0.063	0.071	0.079	0.064	0.089	0.112	0.160	0.129	0.182	0.250	-
<i>aikeni</i>	0.033	0.000	0.074	-	0.069	0.073	0.079	0.074	0.084	0.102	0.168	0.130	0.180	0.252	-
<i>mearnsi</i>	0.033	0.000	2.000	-	0.064	0.070	0.077	0.066	0.083	0.104	0.158	0.124	0.179	0.247	-
<i>oreganus</i>	0.037	0.004	0.004	0.514	-	-	-	-	-	-	-	-	-	-	-
<i>thurberi</i>	-	-	-	-	-	0.061	0.068	0.054	0.075	0.100	0.147	0.116	0.170	0.235	-
<i>montanus</i>	0.006	0.000	0.000	-0.011	-	0.500	0.073	0.064	0.077	0.098	0.155	0.120	0.172	0.242	-
<i>caniceps</i>	0.033	0.000	-0.028	0.004	-	0.000	0.222	0.068	0.081	0.100	0.160	0.125	0.179	0.248	-
<i>dorsalis</i>	0.033	0.000	0.000	0.004	-	0.000	0.000	0.200	0.070	0.097	0.132	0.108	0.177	0.238	-
<i>pallidus</i>	0.322	0.289	0.239	0.293	-	0.289	0.289	0.289	2.108	0.084	0.152	0.113	0.181	0.252	-
<i>phaeonotus</i>	0.029	0.004	-0.009	0.008	-	0.000	0.004	0.004	0.293	0.383	0.175	0.130	0.200	0.274	-
<i>fulvescens</i>	1.033	1.000	1.000	1.004	-	0.833	1.000	1.000	1.289	1.004	0.000	0.120	0.268	0.326	-
<i>alticola</i>	13.290	13.257	13.257	13.261	-	13.090	13.257	13.257	13.432	13.261	12.257	0.686	0.229	0.293	-
<i>insularis</i>	15.458	15.439	15.439	15.443	-	15.272	15.439	15.439	15.618	15.394	14.526	4.592	1.644	0.271	-
<i>bairdi</i>	15.033	15.000	15.000	14.917	-	14.833	15.000	15.000	15.175	15.004	14.000	7.057	9.395	0.182	-
<i>vulcani</i>	29.033	29.000	29.000	29.004	-	28.833	29.000	29.000	29.175	28.955	28.000	24.924	25.308	28.000	0.000

Table 5. Genetic diversity and demographic history of junco forms. Indices are based on concatenated mitochondrial markers (CR, COI, ND2 and ATPase) and include haplotype diversity (h), nucleotide diversity (π) and Fu's test of population expansion (F_s). Asterisks correspond to $P \leq 0,02$.

986

<i>Taxa</i>	<i>n</i>	<i>No. haps.</i>	<i>H</i>	π	F_s
<i>hyemalis</i>	6	6	1.000 +/- 0.096	0.0012 +/- 0.0008	-3.03*
<i>aikeni</i>	8	3	0.607 +/- 0.164	0.0002 +/- 0.0002	-0.48
<i>oreganus</i>	36	17	0.800 +/- 0.068	0.0008 +/- 0.0005	-11.41*
<i>mearnsi</i>	8	6	0.893 +/- 0.111	0.0010 +/- 0.0007	-1.82
<i>caniceps</i>	9	5	0.806 +/- 0.120	0.0010 +/- 0.0007	-0.08
<i>dorsalis</i>	10	6	0.889 +/- 0.075	0.0009 +/- 0.0006	-1.34
<i>phaeonotus</i>	76	27	0.934 +/- 0.013	0.0010 +/- 0.0006	-18.19*
<i>fulvescens</i>	8	3	0.679 +/- 0.122	0.0005 +/- 0.0004	0.83
<i>insularis</i>	23	9	0.889 +/- 0.032	0.0016 +/- 0.0009	0.04
<i>bairdi</i>	11	3	0.473 +/- 0.162	0.0004 +/- 0.0003	0.90
<i>alticola</i>	15	7	0.895 +/- 0.043	0.0012 +/- 0.0007	-0.36
<i>vulcani</i>	2	2	1.000 +/- 0.500	0.0004 +/- 0.0005	0.00

987

988

989 **Figure 1.** Geographic distribution of phenotypic variation and phylogenetic
990 relationships among *Junco* forms. A. Distribution map of the different junco forms.
991 Colored areas correspond to the breeding ranges of the major forms (see Table 1 for a
992 detailed nomenclature). Dots represent isolated localities with hybrid/intermediate
993 individuals. B. Ultrametric Bayesian phylogeny based on four mitochondrial regions
994 (CR, ND2, COI and ATPase) and the FGB-I5 intron. Posterior probability values are
995 shown at each node. Black asterisks correspond to additional nodes with support values
996 above 0.9. The node separating *fulvescens* and the clade of recently radiated northern
997 forms (blue asterisk) was timed at 79,600 years before present (95% HPD: 41,300-
998 121,800) using BEAST, while the TMRCA of the northern form clade along (red
999 asterisk) was estimated at 58,300 years ago (95%HPD: 31,300-88,600) in BEAST, and
1000 16,132 years ago (95%HPD: 6,874-36,335) in IMA2.

A.

B.

Figure 2. MtDNA structure and diversity across the junco range. Breeding range (gray areas), sampling sites (black dots), mtDNA haplotype frequency per locality (pie charts) and median-joining haplotype networks for ND2 (A) and ATPase (B) genes. On the maps, pie chart size is proportional to the number of samples, and colors correspond to the different haplotypes in the population depicted in the networks on the right, which also correspond to colors used in Fig. 1. For the correspondence between haplotype distribution and the geographic range of different junco forms see Fig. 1. In the networks, haplotypes are represented by circles sized in proportion to their frequency in the population. Each branch represents a single nucleotide change, with additional mutations indicated by bars along branches. Numbers in squares indicate nucleotide changes along longer branches.

A.

1018 B.

1019

Figure 3. Genetic structure of northern junco forms based on a principal components analysis of selectively neutral genome-wide SNPs. Marker colors correspond to those on the range map on Fig. 1A, except for ORJU and YEJU, which are divided into two additional taxa shown by different shades.

Figure 4. Phylogenetic relationships among junco forms based on genome-wide SNP markers. A. Unrooted maximum likelihood phylogenies of (A) all junco taxa based on 32,674 selectively neutral SNP loci, and (B) recently diverged northern junco forms only (excluding highly divergent taxa) based on 25,286 selectively neutral SNP loci. See Methods section for SNP filtering criteria. Branch colors correspond to those on Fig. 1A.

Figure 5. Variation in phylogenetic signal among SNP datasets from across the global-
 F_{ST} distribution. From an initial dataset of 101,145 neutral SNPs, shown are unrooted
maximum likelihood phylogenies of northern junco forms (right panels) based on SNP
matrices of different ranges within the global- F_{ST} distribution (left panels),
corresponding to 100% (A), top 25% (B), and top10% SNPs from the global- F_{ST}
distribution (C). Blue bars correspond to the SNPs used, gray bars correspond to the
SNPs excluded (if any), and black lines correspond to the accumulated percentage of
SNPs over F_{ST} classes.

A.

B.

C.

Figure 6. Phylogenetic footprint of a recent northward expansion using SNP data.

Maximum likelihood phylogeny of northern junco forms based on a subset of neutral SNP loci restricted to the 25% with highest global F_{ST} (26,393 loci, $F_{ST} = 0.09$ to 0.26).

The Chiapas junco (*fulvescens*, CHJU) from southern Mexico is used to root the tree.

Figure 7. Phenotypic differences among junco forms and their correlation with genetic distance. The first two discriminant functions in a discriminant function analysis (DFA) based on morphological variables (A) and plumage color variables (B) show patterns of phenotypic differentiation among junco morphs. Heat-map plots showing the correlation between pairwise SNP-based genetic distances among forms (below diagonal) and Euclidean distances among DFA centroids (above diagonal) for morphometric (C) and colorimetric (D) variables. Morphometric and colorimetric divergence values between forms show generally opposite patterns among old and recently diverged junco lineages.