

# Rapid evolutionary divergence of a songbird population following recent colonization of an anthropogenic habitat

Guillermo Friis<sup>1</sup>, Jonathan Atwell<sup>2</sup>, Adam Fudickar<sup>2</sup>, Timothy Greives<sup>3</sup>, Pamela Yeh<sup>4</sup>, Trevor Price<sup>5</sup>, Ellen Ketterson<sup>2</sup>, and Borja Milá<sup>1</sup>

<sup>1</sup>Museo Nacional de Ciencias Naturales

<sup>2</sup>Indiana University

<sup>3</sup>North Dakota State University

<sup>4</sup>University of California Los Angeles

<sup>5</sup>University of Chicago Division of the Biological Sciences

October 6, 2021

## Abstract

Colonization of a novel environment by a few individuals can lead to rapid evolutionary change, yet evidence of the relative contributions of neutral and selective factors in promoting divergence during the early stages of colonization remain scarce. We explore the role of neutral and selective forces in the divergence of a unique urban population of the dark-eyed junco (*Junco hyemalis*), which became established on the campus of the University of California at San Diego (UCSD) in the early 1980s. Previous studies based on microsatellite loci documented significant genetic differentiation of the urban population as well as divergence in phenotypic traits relative to nearby montane populations, yet the geographic origin of the colonization and the factors involved remained uncertain. Our genome-wide SNP dataset confirmed the marked genetic differentiation of the UCSD population, and we identified the coastal subspecies *pinosus* from central California as its sister group instead of the neighboring mountain population. Demographic inference recovered a separation from *pinosus* as recent as 20 to 32 generations ago after a strong bottleneck, suggesting a role for drift in genetic differentiation. However, we also found significant associations between habitat variables and genome-wide variants linked to functional genes, some of which have been reported as potentially adaptive in birds inhabiting modified environments. These results suggest that the interplay between founder events and selection may result in rapid shifts in neutral and adaptive loci across the genome, and reveal the UCSD junco population as a case of contemporary evolutionary divergence in an anthropogenic environment.

## Introduction

Colonization of newly available habitats has led to some of the most conspicuous cases of biological diversification known in natural systems (e.g. Carson & Kaneshiro 1976; Lerner *et al.* 2011; Sato *et al.* 2001), raising questions about the relative roles of random founder effects and selection in population divergence (Carson 1975; Mayr 1963; Price *et al.* 2010; Templeton 1980). Population divergence following colonization can occur due to both local adaptation to novel selective pressures (Rundle & Nosil 2005; Schluter 2000), and to stochastic effects derived from rapid changes in effective population size (Charlesworth 2009; Lande 1980; Simpson 1953). Founder effects and directional selection may result in rapid shifts in allele frequencies and accelerated rates of trait evolution, as shown by instances of rapid evolution following colonization events at time scales of thousands to a few million years (e.g. Lerner *et al.* 2011; Millien 2006; Sato *et al.* 2001; Seehausen 2006; Wessel *et al.* 2013), to scales as short as a few decades (e.g. Chen *et al.* 2018; Jensen *et al.* 2017; Mathys & Lockwood 2011; Sendell-Price *et al.* 2020). In the context of global environmental change, contemporary colonization of modified habitats provides the opportunity to study evolution

in progress (Colautti & Lau 2015; Huey *et al.* 2000; Johnson & Munshi-South 2017; Perrier *et al.* 2020; Reznick & Ghalambor 2001; Salmón *et al.* 2021), and to understand the relative contributions of directional selection, phenotypic plasticity, gene flow and demography in driving rapid divergence in newly established populations (Campbell-Staton *et al.* 2020; Szulkin *et al.* 2020b).

A potential case of contemporary colonization is provided by a suburban population of the Oregon junco (*Junco [hyemalis] oregonus*), a small passerine that typically inhabits mixed-coniferous forests of western North America (Miller 1941; Nolan *et al.* 2002). In the early 1980s, a group of Oregon juncos became naturally established on the campus of the University of California at San Diego (UCSD) (McCaskie 1986; Rasner *et al.* 2004; Yeh 2004), on the coast of the Pacific Ocean. More recently, other Oregon junco populations have been identified in suburban areas of southern California, where they inhabit a mixed landscape of buildings and park areas. The early colonists from UCSD continued breeding and formed a resident population of about 70 pairs that has remained stable (Atwell *et al.* 2014; Yeh & Price 2004). Mark-recapture studies have shown that they are year-round residents at UCSD, but between September and May they are joined by small flocks of wintering Oregon juncos coming from other populations (Fudickar *et al.* 2017; Unitt 1984; Yeh 2004).

The Oregon junco is composed of various geographic forms that originated recently as part of the postglacial radiation of the dark-eyed junco (*Junco hyemalis*), following the northward recolonization of the North American continent as ice sheets receded after the Last Glacial Maximum (LGM), ca. 18,000 years ago (Friis *et al.* 2016; Mila *et al.* 2007). The Oregon junco has been traditionally divided into at least seven geographically structured forms showing considerable phenotypic and habitat variation (Dwight 1918; Friis *et al.* 2018; Miller 1941; Nolan *et al.* 2002). These forms include, from south to north: *townsendi* from the San Pedro Martir mountains in northern Baja California, Mexico; *pontilis* from the Sierra Juarez mountains, also in Baja California; *thurberi* from the mountains of southern to northern California; *pinosus* from the coastal region of central California; *montanus* from the interior regions of Oregon, Washington and British Columbia; *shufeldti* from coastal regions of Oregon and Washington; and *oreganus* from coastal British Columbia and southern Alaska (Miller 1941; Nolan *et al.* 2002) (Fig. 1).

Previous studies on the ecology and evolution of the UCSD junco population have generally assumed that it originated from the nearest breeding population of Oregon juncos, located in the Laguna Mountains, 70 km east of UCSD (Fig. 1), based on the appearance of birds sampled during the non-breeding season. This population belongs to the subspecies *thurberi* and inhabits elevations above 1,500 m, a montane habitat with a more extreme temperature regime than the milder Mediterranean climate of the UCSD campus (Unitt 1984; Yeh & Price 2004). Studies published to date have focused on the evolution of social signaling traits during colonization of a novel environment (Price *et al.* 2008; Reichard *et al.* 2020; Yeh 2004); the role of plasticity in population persistence during the early stages of colonization (Price *et al.* 2008; Yeh & Price 2004); patterns of morphological and genetic variation in comparison with other California populations, using both microsatellites (Rasner *et al.* 2004) and MHC loci (Whittaker *et al.* 2012) and hormonal changes underlying shifts in phenotypic and life-history traits during adjustments to urban environments (Atwell *et al.* 2012; Atwell *et al.* 2014; Fudickar *et al.* 2017). Based on phenotypes of over-wintering non-resident birds and proximity to potential sources, Yeh (2004) proposed that the population of UCSD juncos had likely originated from the *thurberi* population in the nearby mountains as a result of overwintering birds remaining to breed. Simulation-based analyses using microsatellite data reported in Rasner *et al.* (2004) were congruent with the hypothesis that the UCSD population experienced a founder event, with the  $N_e$  of the founding population ranging from 7 to 70 birds. Fudickar *et al.* (2017) used genome-wide SNP data to distinguish between year-round residents and wintering visitors at UCSD, and revealed marked differentiation of UCSD residents with respect to other Oregon junco populations. Divergence levels were in fact comparable to those found among dark-eyed junco lineages originated during the post-glacial radiation of the complex (Friis *et al.* 2016; Friis *et al.* 2018; Friis & Mila 2020), suggesting that high-throughput SNP data may provide a level of phylogenetic resolution not afforded by the more limited genetic sampling of previous studies.

Here we use genome-wide SNP data and extensive geographic sampling of Oregon junco forms across the

region to reconstruct the evolutionary and demographic history of the UCSD junco population. First, we use phylogenetic and co-ancestry analyses to identify the source junco population from which the UCSD population originated. We then apply demographic modelling to test whether the UCSD population is more likely to have resulted from a founder event ca. 30 generations ago, as previously assumed based on the date of appearance at UCSD, or instead diverged elsewhere for a longer period before colonizing the campus. We also use feather isotope ratios to determine whether birds that appear to be winter visitors at UCSD based on genetic data have indeed bred at a different latitude as opposed to being recent recruits into the resident breeding population. Finally, to infer potential selective pressures involved in rapid genetic divergence, we implement a genotype-environment association (GEA) analysis, and take advantage of a newly annotated version of the *Junco hyemalis* reference genome (Feng *et al.* 2020; Friis *et al.* 2018) to identify candidate loci under divergent selection.

## Materials and methods

### *Population sampling*

Oregon juncos were sampled across their breeding range using mist nets and seed-baited walk-in traps during the breeding season, between the months of April and June, and between the years 2004 and 2014. In addition to samples collected at UCSD in San Diego, California, we included in the study samples from all recognized Oregon junco subspecies, including *oreganus* from British Columbia, *montanus* and *shufeldti* from Oregon, *thurberi* and *pinosus* from California, and *pontilis* and *townsendi* from Baja California, Mexico (Fig. 1). At UCSD, birds were sampled both during the breeding season and the non-breeding season, and genetic and isotopic analyses were used to infer whether birds sampled in the non-breeding season were local residents or wintering visitors breeding elsewhere (see below).

Birds were marked with uniquely numbered aluminum bands to avoid resampling, and then released after processing at the site of capture. A blood sample was collected by venipuncture of the brachial vein and stored in Queen’s lysis buffer (Seutin 1991) or absolute ethanol at -80°C in the laboratory. All sampling activities were conducted in compliance with Animal Care and Use Program regulations at the University of California Los Angeles, University of California San Diego, North Dakota State University and Indiana University, and with State and Federal scientific collecting and bird banding permits in the USA, Canada and Mexico.

### *Genotyping-by-sequencing*

Genomic DNA was extracted from blood and tissue samples using a Qiagen DNeasy kit (Qiagen<sup>TM</sup>, Valencia, CA). We used genotyping-by-sequencing (GBS; Elshire *et al.* 2011) to obtain individual genotypes from a total of 174 juncos. Of these, 46 individuals were sampled at UCSD, 18 of them corresponding to confirmed resident breeders (referred to as ‘UCSD breeders’ hereafter), and 28 were sampled during the non-breeding season (referred to as ‘UCSD n.b.s.’ hereafter). In addition, we sequenced 128 Oregon juncos belonging to the following taxa (with sample sizes in parentheses): *townsendi* (16), *pontilis* (13), *pinosus* (14), *thurberi* (41; 23 of them from Laguna Mountains, and 18 from Tahoe), *montanus* (16), *shufeldti* (12) and *oreganus* (16) (Fig. 1, Table 1, Table S1 from Supporting Information). GBS libraries were prepared and sequenced at Cornell University’s Institute for Genomic Diversity, using restriction enzyme PstI for digestion. Sequencing of the 174 individually-barcoded libraries was carried out in five different lanes (along with another 301 junco samples intended for other analyses) of an Illumina HiSeq 2000, resulting in an average of 243.2 million good barcoded single-end reads 100 bp in length per lane.

### *Variant calling*

We evaluated GBS read quality using FASTQC (Andrews 2010) after sorting them by individual with AXE (Murray & Borevitz 2017) and performed the trimming and quality filtering treatment using Trim Galore (Krueger 2015), excluding all reads outside of a range of 40 to 90 bp long. Adapter removal stringency was set to 1 and the quality parameter ‘q’ to 20. GBS reads were then mapped to the *Junco hyemalis* genome (BioProject accession number PRJNA493001; for assembly details see Friis *et al.* 2018) using the

mem algorithm in the Burrows-Wheeler Aligner (BWA, Li & Durbin 2009). Read group assignment and generation of BAM files was carried out with Picard Tools version 2 (<http://broadinstitute.github.io/picard>). We used the Genome Analysis Toolkit (GATK, McKenna *et al.* 2010) version 3.6-0 to call the individual genotypes with the HaplotypeCaller tool. We then used the GenotypeGVCFs tool to gather all the per-sample GVCFs files generated in the previous step and produce a set of jointly-called SNPs and indels (GATK Best Practices, Auwera *et al.* 2013; DePristo *et al.* 2011) in the variant call format (*vcf*). We retained biallelic SNPs and applied GATK generic hard-filtering recommendations consisting on QualByDepth (QD) > 2.0; FisherStrand (FS) < 60.0; RMSMappingQuality (MQ) > 40; and MappingQualityRankSumTest (MQRankSum) > -12.5. Since GBS homologous reads have identical starts and ends, filters depending on position within reads (ReadPosRankSum and StrandOddsRatio) were not applied. Individuals with rates of missing data above 0.6 were excluded at this point. Using *vcftools* (Danecek *et al.* 2011), we then excluded those SNPs out of a range of coverage between 4 and 50 or with a genotyping phred quality score below 40. The resulting dataset (hereafter referred to as ‘Full Dataset’ in downstream analyses) consisted of 158 individuals (Table 1) and 1,448,676 SNPs with a per-individual average coverage of 4.25 and a missing data rate of 0.65.

### *Population structure and genetic diversity analyses*

To explore genome-wide population structure among Oregon junco forms, we applied principal components analysis (PCA) based on SNP data. SNP loci under linkage disequilibrium were filtered out from the ‘Full Dataset’ with *bcftools* (Danecek & McCarthy 2017) applying an  $R^2$  limit of 0.2 in windows of 10K bp. Using *vcftools*, positions with less than 50% of individuals genotyped for each taxon/population were removed from the data matrix, along with those presenting a minor allele frequency (MAF) below 0.02. We implemented a threshold for SNP loci showing highly significant deviations from Hardy-Weinberg equilibrium (HWE) with a p-value of  $10^{-4}$  to filter out false variants arising from the alignment of paralogous loci. To fulfil neutrality assumptions in population structure analyses, we used PCADAPT (Luu *et al.* 2017) to detect and exclude sites putatively under selection, applying a q-value threshold of 0.1, resulting in a final data matrix of 21,879 SNPs (hereafter referred to as the ‘Neutral Dataset’ in downstream analyses). The PCA was conducted using the function *snpgdsPCA* available in *SNPrelate* to obtain the eigenvectors to be plotted.

We also examined population divergence in Oregon juncos with the program STRUCTURE (Pritchard *et al.* 2000) and the ‘Neutral Dataset’. We converted the *vcf* file to STRUCTURE format using PGDspider (Lischer & Excoffier 2012) version 2.0.5.1. Bash scripts to perform the analyses were created with STRAUTO (Chhatre & Emerson 2016), and we ran the program five times per K value, with K ranging from 1 to 10 after running a preliminary analysis to infer the lambda value. The burn-in was set to 50,000 iterations and the analysis was run for an additional 100,000 iterations. Similarity scores among runs and graphics were computed with CLUMPAK (Kopelman *et al.* 2015).

We computed pairwise  $F_{ST}$  (Weir & Cockerham 1984) values from the ‘Neutral Dataset’ after excluding UCSD n.b.s. samples (SNP matrix size = 38,247) whose breeding range is uncertain, using the hierfstat R package (Goudet *et al.* 2015). We also used the program populations from the Stacks pipeline (Rochette *et al.* 2019) to compute the observed ( $H_O$ ) and expected heterozygosity ( $H_E$ ), and nucleotide diversity ( $\pi$ ) based on autosomal positions for each form.

### *Phylogenetic and co-ancestry analyses*

Phylogenetic analyses based on GBS data were conducted with the main objective of identifying the source lineage from which the UCSD population derived. For this test we excluded once again the UCSD n.b.s. individuals from the ‘Neutral Dataset’. A maximum likelihood phylogeny and a neighbor-net were produced using the programs IQ-TREE (Nguyen *et al.* 2015) and SplitsTree (Bryant & Moulton 2004), respectively. For the former, the ascertainment bias correction and the generalized time-reversible (GTR) model were implemented.

We further tested the origin of the UCSD junco population by exploring pairwise levels of co-ancestry among Oregon junco forms with the program RADpainter (Malinsky *et al.* 2018). Starting from the ‘Full Dataset’,

UCSD n.b.s. individuals were excluded. To detect subtler signs of differentiation, and because this program relies on linkage disequilibrium to achieve more refined assignment of nearest-neighbor relationships, we did not apply neutrality and linkage filters in this analysis. We did, however, increase the per-site missing data threshold to 0.75 as recommended by the authors, and applied the same MAF and HWE filters as in the PCA and STRUCTURE analyses, resulting in a matrix of 41,231 SNPs. We ran the analysis with 100K burn-in and 100K sampling iterations, and a thinning interval of 1,000 to produce a matrix of pairwise individual co-ancestry estimates, which were averaged over forms. For estimating 95% confidence intervals around pairwise co-ancestry values between UCSD and the different Oregon junco forms, we applied a non-parametric bootstrap procedure. Bootstrapping was carried out by splitting the SNP matrix in 1,000 blocks and randomly combining them for each one of the repetitions. A total of 100 analyses were run for confidence interval estimation.

*Breeding latitude of UCSD juncos captured during the non-breeding season* Hydrogen isotopic ratios in bird feathers correspond to those found in the localities where the feathers were grown, and since juncos molt their feathers at the end of the breeding season, isotopic ratios can be used to determine relative differences in breeding latitude (Chamberlain *et al.* 1996). To determine whether UCSD birds sampled in the non-breeding season (UCSD n.b.s.) were local breeders or winter visitors breeding elsewhere, we tested for correlation between deuterium isotope ratio ( $\delta D$ ) and posterior likelihoods of genetic population assignment. We obtained isotopic data from the 28 individuals sampled in UCSD over winter, to compare to two confirmed year-round residents used as controls. A single secondary feather was collected per bird and washed using a 2:1 chloroform: methanol solution. Approximately 0.50 mg of vane material was clipped from the proximal end of each feather and the hydrogen isotope composition was measured by continuous flow isotope ratio mass spectrometry using a Finnigan TC/EA interfaced to a Finnigan DeltaPlus XL mass spectrometer (Thermo Scientific, Bremen, Germany), as described in Wunder *et al.* (2012).  $\delta D$  values are reported in per mil notation (-172 2003). Benzoic acid, ( $\delta D = -61$ ) analyzed within analytical sequences with a precision of  $< \pm 4$  same 30 individuals (four failed to pass missing data filters, Table 1), and assignment likelihoods were computed with STRUCTURE under the same conditions, in this case only for  $K = 2$ . Correlation was tested by means of a linear regression between STRUCTURE scores and  $\delta D$  abundance.

#### *Demographic inference and molecular dating*

The marked genetic divergence of the UCSD juncos with respect to neighboring populations documented to date (Fudickar *et al.* 2017; Rasner *et al.* 2004), has been considered to be the result of the founder event in the 1980's and the novel selective forces associated with the newly colonized habitat. Alternatively, UCSD juncos may derive from a population that diverged elsewhere over a longer period before colonizing UCSD. The effect of a potential founder event just ca. 30 generations ago on the pattern of genetic variation observed in UCSD juncos needs to be confirmed, as the population could potentially be part of a larger population, so that no actual reduction in effective population size took place upon the colonization of the campus. To explore these competing hypotheses, we performed model comparison under the likelihood framework developed in fastSIMCOAL2 (Excoffier *et al.* 2013). Even though methods based on SFS for demographic inference using low density genomic data have been successfully applied to systems at comparable temporal scales (e.g. McCoy *et al.* 2014; Sendell-Price *et al.* 2020), GBS data have limitations when attempting to infer demographic events that took place just a few generations ago (Beichman *et al.* 2018). In an attempt to account for this, we implemented a general model that was as simple as possible, with a single cladogenetic event, and setting narrow bounds around splitting times for the different hypotheses to be tested. Fortunately, we had an approximate estimate of the current census size of the UCSD population, which was used to inform the models.

We established two temporal ranges (recent = 20-60 generations ago; and postglacial = 4,000 – 12,000 generations ago; generation time = 1.5 years) for two potential demographic events (population split from source lineage with or without a founder event) resulting in four possible models to compare: recent population split following a founder event (Fig. 2A); postglacial population split followed by a recent founder event (Fig. 2B); recent population split without a founder event (Fig. 2C); and postglacial population split

without a founder event (Fig. 2D). In addition, two different gene migration scenarios were implemented: a ‘strict isolation’ scenario with migration rates set to zero; and an ‘isolation with migration’ scenario where migration rates can vary freely. The current effective size of the UCSD population (Pop. Size 1 in Fig. 2) was estimated within a range of 35 to 280 individuals based on previous direct censuses (Atwell *et al.* 2014; Fudickar *et al.* 2017) in scenarios A and B, while it was allowed to vary in models C and D. The number of colonizers in scenarios A and B was set to vary between 2 and 280. Mutation rate was fixed at  $2.21 \times 10^{-9}$  in all the cases (Nam *et al.* 2010; Smeds *et al.* 2016), and a generation time of 1.5 years was assumed to calibrate parameters and estimate divergence times. The likelihood of each of the final eight models was estimated with fastSIMCOAL2 along with the time of population splitting, the current and ancestral effective population sizes, and the gene migration rates in those models where they were implemented.

As input data we used the folded site frequency spectra (SFS) generated from GBS data. We retained the samples corresponding to the two study groups from the ‘Full Dataset’ and filtered out sex-linked positions to avoid distortions related to loci with different effective sizes. We applied the HWE filter, resulting in a matrix of 828,699 SNP loci. Because singletons are important for estimating parameters and likelihoods, no MAF filters were applied. The SFS were generated with easySFS (<https://github.com/isaacovercast/easySFS>), maximizing the number of segregating sites as recommended by the authors (I. Overcast, pers. comm.). Parameter values with the highest likelihood were estimated under each of the models after 50 cycles of the algorithm, with 150,000 coalescent simulations per cycle. This procedure was replicated 100 times and the set of parameter values with the highest final likelihood was retained as the best point estimate. To compare models we applied the Akaike Information Criterion (AIC; Akaike 1998). For estimating 95% support limits for the parameters estimated under the best model, we followed the same bootstrap procedure than for RADpainter using blocks of 51 SNPs to produce 100 bootstrapped datasets.

#### *Genotype-environment association analysis*

We used genotype-environment association (GEA) analysis to identify selective pressures potentially driving rapid adaptive divergence of UCSD juncos. Here, we used redundancy analysis (RDA), a canonical ordination method that combines regression and principal components analysis enabling an estimate of the variance of a set of response variables explained by a number of constraining or explanatory variables (Borcard *et al.* 2011; Legendre & Legendre 1998; Van Den Wollenberg 1977). In RDA applied to GEA studies, explanatory variables are usually environmental parameters, while response variables correspond to the individual genotypes at each variant position. We chose RDA because of its ability to accommodate multiple explanatory variables simultaneously. We used geo-referenced environmental data, extracted for the coordinates at which junco breeding individuals were sampled to capture the distinct environmental conditions at UCSD with respect to the native range of Oregon juncos. We started by selecting a subset of the 19 variables available in the WorldClim database (Hijmans *et al.* 2005), chosen *a priori* according to their relevance to junco ecology (Miller 1941; Nolan *et al.* 2002). They measured mean temperature and precipitation over the year (BIO1 and BIO12); mean temperature and precipitation over the warmest quarter (BIO10 and BIO18), which corresponds generally to the junco breeding season; isothermality, which refers to how the range of day-to-night temperature differs from the range of summer-to-winter (BIO3); and seasonality of temperature and precipitation (BIO4 and BIO15). We also included three vegetation variables from the Moderate Resolution Imaging Spectroradiometer (MODIS) satellites as available in <https://modis.gsfc.nasa.gov>: percent tree cover (TREE), Normalized Difference Vegetation Index (NDVI, a measure of canopy greenness), and NDVI’s annual standard deviation (std\_NDVI). In addition, we included the high-quality elevation data provided by the NASA Shuttle Radar Topographic Mission (SRTM), downloadable from <http://www2.jpl.nasa.gov/srtm>, for a total of eleven variables. To reduce the number of explanatory variables while retaining those maximizing differences between habitats at UCSD and elsewhere, we computed the sum of the pairwise Mahalanobis distances (Mahalanobis, 1936) between UCSD juncos and each of the other forms (as listed in Table 1) for each variable separately. Environmental variables whose sums across the pairwise computed distances were higher than average were retained, consisting of BIO1, BIO10, BIO4, NDVI and SRTM. Based on the variance inflation factor (VIF), BIO1 was eliminated from the set to reduce correlation among explanatory variables. The SNP dataset implemented as the matrix of response variables was obtained by excluding

UCSD n.b.s. individuals from the ‘Full Dataset’, as well as positions with rates of missing data over 0.5 computed over each form. The HWE filter was also applied for a final dataset of 132,343 SNPs. Because RDA does not allow missing data, we replaced non-genotyped positions with the major frequency allele. This approach may hinder genotype-environment signals, resulting in a higher false-negative rate, but allows us to screen a larger fraction of the genome.

*Identification of candidate genes* To ask if signals of genotype-environment association recovered in the RDA were driven by selection rather than by demography related effects, we searched for functional genes associated with the loci showing the highest contribution to the correlation patterns. Candidate regions under selection were identified as those containing SNPs with RDA loadings that departed six times or more the standard deviation from mean loading (Kamvar *et al.* 2017, tutorial by Forester in [popgen.nescent.org/2018-03-27\\_RDA\\_GEA.html](http://popgen.nescent.org/2018-03-27_RDA_GEA.html)). We extracted 2,000-bp regions flanking these SNPs (1,000 bp from the SNP of interest in each direction of the sequence, a conservative distance in terms of linkage disequilibrium for passerines, Backström *et al.* 2006), and overlapping annotated genes were identified by similarity using BLAST2 (Tatusova & Madden 1999) and the annotation of the B10K consortium for the *Junco* genome (Feng *et al.* 2020). To further explore how adaptive variability is structured among Oregon junco forms and test for divergence specifically in UCSD juncos, we computed pairwise  $F_{ST}$  values (Weir & Cockerham 1984) based on identified candidate gene loci among all forms, as well as per-locus between UCSD and its more closely related group (see Results).

All R analyses were carried out in RStudio v 1.1.463 (R.Studio.Team 2015) with R v3.5.3 (R.Core.Team 2019). See Appendix I in Supporting Information for R scripts.

## Results

*Neutral genetic structure among Oregon junco forms* A plot of PC1 against PC2 (11.5% and 8.7% of explained variance, respectively) revealed considerable differentiation among Oregon junco forms. Most UCSD birds formed a highly differentiated cluster from the rest of the groups, with levels of divergence comparable to those of the southern isolated forms in Baja California, *pontilis* and *townsendi*. The form *pinosus* also formed a separate cluster, although distance with respect to the remaining forms was less pronounced. Northern forms *oreganus*, *shufeldti* and *montanus* showed very limited levels of structure and grouped into a single genetic cluster, which also included some of the UCSD n.b.s. individuals (Fig. 3A, upper plot). The plot of PC1 against PC3 (4.0% of explained variance) also revealed *thurberi* birds from Laguna Mountains as a differentiated population (Fig. 3A, lower plot).

The STRUCTURE analysis was consistent with PCA results and provided additional information on genetic differentiation among localities (Fig. 3B). At  $K = 2$ , most UCSD birds separated from the rest, and three breeders showed a limited degree of admixture with non-UCSD birds. Most UCSD n.b.s. birds belonged to the resident population, but five of them were clearly identified as winter visitors from further north. The plot for  $K = 3$  and 4 also revealed *townsendi* and *pontilis* as highly divergent groups with respect to other junco populations and with respect to each other, although they share a considerable amount of variance. The form *pinosus* formed a clearly differentiated cluster at  $K = 5$ . Birds from Laguna Mountains also appeared as a separate population at  $K = 6$ , again congruently with the PCA and ruling out this population as the source of UCSD n.b.s. birds, which instead cluster with birds at more northern latitudes (either *thurberi* -Tahoe, *montanus*, *shufeldti* or *oreganus*; Fig. 3B).

$F_{ST}$  values among forms based on neutral loci were generally congruent with the PCA and STRUCTURE results, with forms from Baja California and UCSD showing the highest degree of differentiation. The remaining forms showed lower levels of pairwise differentiation, yet *thurberi* from Laguna Mountains and *pinosus* had values that were moderately higher than those from northern populations (Table 2). Both heterozygosity and nucleotide diversity indices revealed lower levels of genetic variability at UCSD and *townsendi* compared to other populations, especially those corresponding to *thurberi* and more boreal forms (Table 1).

*Breeding latitude of UCSD juncos sampled in the non-breeding season* The linear regression between STRUC-

STRUCTURE scores for  $K = 2$  and  $\delta D$  values recovered a highly significant signal ( $P = 2.52 \times 10^{-8}$ ), where the two control UCSD breeders and 17 other genetically close individuals showed high isotopic ratios (Fig. S1, Supporting Information). In turn, the five UCSD n.b.s. individuals that were genetically different from UCSD breeders showed evidence of having grown their feathers elsewhere, and thus represent wintering individuals that do not breed at UCSD.

*Phylogenetic and co-ancestry analyses* Phylogenetic reconstructions based on maximum likelihood and neighbor-joining algorithms were consistent with population structure analyses in showing marked differentiation of Baja California and UCSD, as well as for *pinosus* and southern *thurberi*. In the unrooted maximum likelihood tree, *pinosus* and UCSD individuals formed two reciprocally monophyletic sister clades with strong support (Fig. 4). In turn, relatively low structure among northern Oregon juncos was detected in the analysis, in contrast to forms in Baja California, which clustered together in a differentiated clade with close affinity to southern *thurberi* individuals from Laguna Mountains (as suggested by  $K = 4$  and  $5$  in the STRUCTURE analysis, Fig. 3). In the neighbor-net, UCSD and *pinosus* also appear as sister groups, yet the degree of differentiation with respect other forms is considerably less pronounced (Fig. S2). *Forms townsendi* and *pontilis* appear as sister clades, highly divergent from the remaining junco forms. Southern *thurberi* individuals from Laguna Mountains cluster together as well in this analysis, while little differentiation is observed among northern forms.

The RADpainter genetic ancestry analysis consistently recovered high average values of intra-population co-ancestry in those forms showing marked differentiation and geographic isolation, namely *townsendi*, *pontilis*, *pinosus* and the UCSD population. UCSD individuals showed highest co-ancestry values with respect to *pinosus* individuals (Fig. 5A), a pattern that was significant according to a bootstrap analysis (Fig. 5B).

#### *Demographic inference*

Eight different models were compared to identify the most likely scenario under which UCSD juncos diverged from its closest relative, the *pinosus* form, using fastSIMCOAL2. AIC revealed that the most likely scenario was the one consistent with a recent population split due to a founder event (Fig. 2A) under a model of isolation with migration, and clearly rejected alternative models (Table 3). Parameter estimates revealed a very recent colonization time of the urban environment ( $TDIV = 44$  years, 95% CI = [30 - 47]) by a very small number of effective founders, from two to three individuals ( $N_e = 2$  individuals; 95% CI = [2 - 3]). The estimate for the current effective population size of UCSD was  $N_e = 252$  individuals, 95% CI = [215 - 274]. The best estimate for UCSD population growth rate since colonization was estimated at  $r = 0.159$ , 95% CI = [0.150 - 0.244]. The estimate for *pinosus* effective population size was high, at  $N_e = 5.86 \times 10^6$  individuals, 95% CI = [ $5.86 \times 10^6$  -  $5.87 \times 10^6$ ]. The analysis also recovered a narrow, limited rate of gene migration from *pinosus* to UCSD ( $MIG_{pin-UCSD} = 6.02 \times 10^{-2}$ , 95% CI = [ $5.87 \times 10^{-2}$  -  $8.03 \times 10^{-2}$ ]). The migration rate range from UCSD to *pinosus* varied by several orders of magnitude and was close to zero, revealing little impact of the parameter in the SFS ( $MIG_{UCSD-pin} = 2.01 \times 10^{-6}$ , 95% CI = [ $9.38 \times 10^{-10}$  -  $2.57 \times 10^{-5}$ ]) (Table 3).

Despite the marked difference in AIC scores supporting a recent population split due to a founder event, it should be noted that none of the next best-fitting models included a recent population split, and that the model including a recent population split due to a founder event in strict isolation was not among those at the top (Table 3). This may be due to the effect of gene flow during the establishment and genetic differentiation of the UCSD population, resulting in highly different site frequency spectra when comparing scenarios of recent colonization with and without migration. However, it could also reflect a lack of resolution of the analysis in discriminating the historical factors shaping the current patterns of genetic variability in the UCSD population of juncos, so that these results should be interpreted with caution.

*Genotype association analysis* In the RDA, four explanatory environmental variables accounted for 7.65% ( $R^2 = 0.0765$ ) of the variability in the SNP dataset, and the model was highly significant ( $P = 0.001$ ). Both mean temperature of the warmest quarter and temperature seasonality showed large contributions to the RDA2 axis, while variability in NDVI and elevation was more evenly captured by both RDA1 and RDA2 axes

(Fig. 6). RDA association scores revealed a strong positive correlation of mean temperature of the warmest quarter with genetic differences of UCSD residents, and negative correlations with the remaining explanatory variables. Correlation patterns were less conspicuous in other junco forms, mostly differentiated along RDA1. Forms *townsendi* and *pontilis* showed negative correlations with vegetation greenness, with the former also loading heavily on elevation. The form *pinosus* presented general associations of the same sign as UCSD, yet much less pronounced. The remaining, more boreal forms presented moderate and positive correlations with vegetation greenness and temperature seasonality, while loading negatively on mean temperature of the warmest quarter, an inverse pattern with respect to UCSD juncos (Fig. 6).

### Candidate genes

From a total of 132,343 SNPs, 115 showed absolute loading scores on the RDA1 and RDA2 distributions deviating from the mean by more than six times the standard deviation. Out of these 115 loci, the 2000-bp flanking region of 23 of them (20%) mapped against a total of 15 annotated genes. Potentially relevant reported functions of the identified genes in vertebrates included the following: heart and muscle development (DPF3; Lange *et al.* 2008); heavy metal sensitiveness and detoxification (ABCB6; Rakvác *et al.* 2019); high-pitched hearing (KCNQ4; Liu *et al.* 2011); laying performance and egg quality (IOV7; Huang *et al.* 2019; Wu *et al.* 2018); DNA crosslink repair and tolerance to UV light and camptothecin-induced DNA damage (FANCM; Rosado *et al.* 2009); megakaryocytopoiesis, platelet formation (TPOR; Ocampo Daza & Larhammar 2018); trait regulation of energy balance through effects on feeding behavior and adaptive thermogenesis (KSR2; Guo *et al.* 2017); and modulation of fasting blood glucose (G6PC2; Matschinsky 2005; O'Brien 2013), possibly in response to stress (Boortz *et al.* 2016). Pairwise  $F_{ST}$  values based on SNP loci linked to functional candidate genes revealed relative patterns of differentiation similar to those based on neutral markers, but yielded considerably higher absolute values. UCSD showed the highest pairwise  $F_{ST}$  scores with isolated forms *townsendi* (0.54) and *pontilis* (0.28), an average of 0.32 with northern forms, and the lowest value with *pinosus* (0.21) (Table 2). Per-locus  $F_{ST}$  values between UCSD and the sister group *pinosus* showed moderate to high levels of divergence (0.26 to 0.51) in genes TPOR, ABCB6, FANCM, KCNQ4 and IOV7, with the rest of the genes ranging from 0 to 0.2  $F_{ST}$  values (Table S2, Supporting Information).

## Discussion

### Marked genetic differentiation of the UCSD juncos

Our results confirmed the striking differentiation of the urban population of Oregon juncos inhabiting the UCSD campus. Even when time of divergence was allowed to vary between 20 and 60 generations, parameter estimates, and bootstrap procedures recovered a time of divergence highly consistent with the time at which the juncos were first observed to breed at UCSD in the 1980s, with a narrow 95% CI of 20 to 32 generations, lending support to the recent founder event hypothesis. Recovered levels of divergence for the UCSD residents that were comparable to those of long-established, geographically isolated forms like *townsendi* and *pontilis* from Baja California sky islands, and in sharp contrast with the low genetic structure found among more northern forms. Both the PCA and the STRUCTURE analyses based on selectively neutral or near-neutral SNPs, revealed that the UCSD juncos represent a genetically differentiated population, although some degree of recent admixture with northern forms was detected. Juncos have become established in other suburban and natural areas of southern California in recent years (including the campus at UCLA in 2007, pers. obs.), and the degree to which the UCSD population is genetically isolated from these will require more extensive sampling in the region. At first sight, the pattern of divergence of UCSD juncos is comparable to that of other dark-eyed junco forms, known to have originated several thousand years ago (Friis *et al.* 2016). However, demographic inference analyses yielded results that are consistent with a very recent origin of the urban population. The inferred number of founders was as low as two to three individuals, just below the seven to 70 individuals previously estimated using microsatellite markers (Rasner *et al.* 2004). The analysis also favored a scenario of isolation with migration over a strict isolation model.

The migration rate from *pinosus* to UCSD was estimated at 6.02% with fastSIMCOAL2. This estimate is

consistent with results from Yeh and Price (2004), who reported an immigration rate of 7% (95% CI: 4%-10%) based on observed ratios of banded to unbanded birds over a four-year period, although immigrants from other populations besides *pinosus* may be included in that estimate (see below). In any case, our results suggest that some degree of gene flow occurred during the differentiation of the urban population of juncos. In contrast, the inferred migration rate from UCSD to *pinosus* was negligible.

Our analyses also recovered a strong correlation between  $\delta D$  in feathers and population assignment likelihoods in UCSD individuals sampled during the non-breeding season, confirming the presence of both local residents and genetically distinct individuals from northern latitudes that do not represent recent immigrants but rather wintering visitors. These visitors were also genetically distinct from the juncos from Laguna Mountains and more similar to northern Oregon juncos.

#### *Random effects and selection promote rapid evolutionary change*

Rapid divergence of phenotypic traits is often attributed to strong selective pressures imposed by novel biotic and abiotic environments (e.g. Chen *et al.* 2018; Jensen *et al.* 2017; Mathys & Lockwood 2011; Sendell-Price *et al.* 2020), or to a combination of selective and neutral factors (Grant *et al.* 2001; Kliber & Eckert 2005; Kolbe *et al.* 2012). The juncos from UCSD represent an extreme case of rapid evolutionary divergence, apparently caused by the colonization of an urban environment. Such degree of differentiation taking place over comparably short periods is only met by a few other documented examples in nature (e.g. Chen *et al.* 2018; Hendry & Kinnison 1999; Koskinen *et al.* 2002; Lescak *et al.* 2015; Sendell-Price *et al.* 2020). The small number of colonizers estimated suggests the role of severe founder effects and genetic drift in the process of differentiation of the UCSD juncos (Kolbe *et al.* 2012), a scenario also congruent with the lower values of genetic diversity at UCSD based on neutral markers.

However, we also found evidence of local adaptation in promoting evolutionary divergence between UCSD juncos and closely related Oregon junco forms. A redundancy analysis based on the most dissimilar set of our geo-referenced ecological variables between the UCSD and natural habitats, revealed a distinctive pattern of associations for the urban population, positively correlating with mean temperature during the breeding season and negatively with vegetation greenness, temperature seasonality and elevation. Although these ecological parameters were particularly distinct for the UCSD campus with respect to native ranges, they may not encompass key aspects of the urban environment (e.g., unsuitable habitat due to human structures, contamination, human-processed foods, mechanical noise or distressing stimuli) with which they may, in turn, present spatial correlation (Szulkin *et al.* 2020a). Much of the recovered correlation signal is likely to also be due to population history-related effects (Forester *et al.* 2016; Frichot *et al.* 2015; Wang & Bradburd 2014), yet controlling for neutral genetic structure in RDA approaches may result in a high loss of power to detect signals of selection without necessarily reducing the rate of false positives (Forester *et al.* 2018). RDA outliers linked to functional genes suggest, however, a role for selective pressures in shaping allele frequencies in potentially adaptive loci in the newly colonized habitat. Among the functions associated to detected candidate genes, at least two (ABCB6 and KCNQ4) seem particularly relevant in anthropogenic environments. The ABCB6 gene shows significant sequence identity to HMT-1 (heavy metal tolerance factor 1) proteins, whose evolutionarily conserved role is to confer tolerance to heavy metals through the intracellular sequestration of metal complexes (Rakvács *et al.* 2019). Several studies have reported negative effects of metals (such as cadmium, copper, lead, mercury, nickel and selenium) on fitness traits of birds occurring in anthropogenic habitats (Burger & Gochfeld 1988; Eeva & Lehikoinen 1996; Geens *et al.* 2010; Takekawa *et al.* 2002). In turn, KCNQ4 encodes a protein that forms a voltage-gated potassium channel for the regulation of electrical signaling, expressed in the sensory receptors of the auditory system of all vertebrates (Kharkovets *et al.* 2000). KCNQ4 has also been found to be involved in high-frequency hearing in echolocating bats (Eriksson & Wikteliu 2011). A significantly higher minimum frequency in the vocal signaling of the UCSD juncos with respect to natural habitats have been previously reported (Cardoso & Atwell 2010, 2011; Reichard *et al.* 2020; Slabbekoorn *et al.* 2007), arguably as a strategy to reduce masking by anthropogenic low-frequency noise. Importantly, both genes showed high levels of divergence with respect to *pinosus*, as evidenced by  $F_{ST}$  values. Other genes showing signs of differentiation between these two groups

were involved in thrombopoiesis, DNA damage repair, or egg laying performance and quality, which could also be involved in adaptive processes driving differentiation between *pinosus* and UCSD juncos. These results suggest a specific role for anthropogenic selective pressures in shaping adaptive variability among UCSD and forest Oregon juncos, which in combination with extreme founding conditions, may have resulted in an exceptionally fast process of differentiation at both functional and neutral loci in the resident juncos at UCSD.

### *Geographic origin of the UCSD juncos*

Phylogenetic and co-ancestry analyses strongly supported the form *pinosus* as the sister group of the UCSD juncos, and not the population at the nearby Laguna Mountains as reported in previous studies. Phylogenetic, co-ancestry or population structure analyses failed to detect any significant admixture or clustering of birds breeding or wintering at UCSD with the *thurberi* individuals from Laguna Mountains. One UCSD breeder did present an intermediate assignment probability to the Laguna Mountains population, while one of the likely residents detected among the UCSD wintering birds showed some admixture with *pinosus*, which is consistent with the migration rates estimated in fastSIMCOAL2 analyses. These two birds also occupy intermediate positions between the *pinosus* -northern Oregon juncos and the UCSD clusters in the PCA, suggesting these two individuals are likely the result of recent introgression.

The population genomics analyses presented here thus support a recent colonization of urban San Diego by a few individuals from the *pinosus* population up the coast, followed by limited gene flow from other Oregon junco forms over the last ca. 40 years. Although gene flow between divergent populations could have acted as a homogenizing force, eroding population differentiation (Slatkin 1987; Wright 1931), there is theoretical and empirical evidence showing that under conditions of rapid population growth following a founder event, divergence of allele frequencies established during colonization may be resistant to decay by gene exchange (Boileau *et al.* 1992; Lomba *et al.* 2020). Moreover, moderate admixture can also contribute to colonization success by minimizing deleterious effects of reduced intra-population diversity or by introducing favorable alleles and enhancing evolutionary potential (Keller & Taylor 2010). That introgression between taxa contributes to adaptation and sorting of genes has become increasingly recognized as a major contributor to population divergence and speciation (Abbott *et al.* 2013; Lamichhaney *et al.* 2015; Wegener *et al.* 2019). Based on our genomic analyses and on the migration rates reported here and in Yeh and Price (2004), as well as in the phenotypic studies in Yeh (2004), it seems that several over-wintering birds on the UCSD campus came from populations other than *pinosus*, and some of these may have remained to breed on campus. It thus appears likely that there has been introgression into the UCSD population from forms other than *pinosus*, and its relevance remains to be determined.

The form *pinosus*, from which the UCSD population originated, is a non-migratory form found along the coastal ranges of central California and occurs in a range of Mediterranean habitats, displaying a tolerance for low zonal conditions not found in *thurberi* (Miller 1941). Individuals from the *pinosus* population would generally be expected to produce higher fitness progeny at UCSD than *thurberi* individuals. The amount of white on the tail feathers of *pinosus* reduced relative to *thurberi* (Miller 1941), and more similar to UCSD individuals. Indeed, Price *et al.* (2008) inferred strong selection on juvenile juncos carrying excess white in their tail, which they attributed to both direct selection on white, and correlated responses to greater dispersal propensity, both of which would be most prominent in offspring from *thurberi* immigrants, and lead to lower rates of introgression. Nevertheless, some alleles may successfully introgress from other subspecies, and the large pool from which immigrants appear to be drawn from increases chances of adaptive alleles such as the ones we have inferred from entering into the UCSD population.

### **Conclusions and implications for previous work**

Analyses based on genome-wide markers confirmed the marked genetic divergence of the Oregon juncos at UCSD, a population that became established in the early 1980s. Three further, fundamental aspects of the evolutionary history of the UCSD population have also been established here based on genomic data: (i) the UCSD junco population is most closely related to the coastal form *pinosus*, a group showing ecological

and phenotypic similarities; (ii) environmental factors played a role in driving rapid divergence at functional loci, some of which may be involved in adaptations to an anthropogenic habitat; and (iii) demographic inference supports a scenario of contemporary population divergence as a result of a founder event by a very small number of colonizers, consistent with a major role for genetic drift in the process of differentiation. Together these results suggest that a combination of drift and directional selection can result in rapid shifts in both neutral and adaptive loci across the genome, and reveal the Oregon juncos from UCSD as a rare contemporary case of population divergence following the colonization of an anthropogenic environment.

Based on the new evidence regarding the origin of the UCSD juncos presented here, some conclusions from previous studies on this system regarding rates of evolutionary change and phenotypic plasticity may have to be reassessed. Specifically, differences in morphological fitness traits (Rasner et al. 2004), sexual signaling traits (Yeh 2004; Price et al. 2008), and shifts in a range of behavioral and life-history traits (Yeh and Price 2004; Price et al. 2008) were thought to be due to rapid evolutionary change, as were associated differences in hormones and behavior (Atwell et al. 2012; Atwell et al. 2014; Fudickar et al. 2017). Studies comparing individuals from UCSD to those from the nearby mountain *thurberi* population attributed differences to a recent transition from a mountain climate to a more benign Mediterranean climate, and from a wild to an anthropogenic urban habitat. These comparative studies remain relevant to understanding phenotypic differences between junco populations adapted to different environments, although assumptions regarding rapid evolution may not be accurate given the new evidence on phylogeny and co-ancestry.

For example, UCSD juncos’ greater phenotypic similarity in tail white to their closest relative *pinosus*, rather than to the mountain *thurberi* individuals they were previously compared to, suggests that prior explanations may have overestimated the role of rapid evolutionary change in plumage variation. Further, prior research comparing UCSD and the montane *thurberi* phenotypes could not distinguish between two potential explanations, adaptation to an anthropogenic environment or to a mild Mediterranean climate. The fact that *pinosus* shares some aspects of a mild Mediterranean climate with UCSD suggests that selection reported here is associated more strongly with the anthropogenic environment at UCSD rather than milder climate. Looking forward, our new findings provide the opportunity to develop new comparative studies between UCSD and *pinosus* individuals that will contribute to effectively addressing the role of urban habitats in driving rapid evolutionary divergence.

**Acknowledgements**

Pau Aleixandre, Elena Berg, Setefilla Buenavista, John McCormack, Adán Oliveras de Ita, Vicente Rodríguez, Kristen Ruegg, and Alvar Veiga provided help with field sampling. Funding was provided by grants CGL-2011-25866 and CGL-2015-66381 from Spain’s Ministerio de Ciencia e Innovación to BM, and USA National Science Foundation IOS-1257527 to TJG.

**Data Accessibility** *Junco hyemalis* reference genome deposited in NCBI (Accession number: QZWM00000000). SNP ‘Full Dataset’, isotopic and georeferenced datasets are deposited in Dryad, as well as the model files for the demographic analysis with fastSIMCOAL2 and the R script for the redundancy analysis. Link for reviewers: <https://datadryad.org/stash/share/RAlwLVkew2K4HudDCqjp455t2yWE6JmRFNvVu55yWY0>

**Table 1.** Oregon junco forms, number of genotyped individuals and population genetic diversity indices per locality. Indices include observed ( $H_O$ ) and expected ( $H_E$ ) heterozygosity. Figures in the last column correspond to the number of samples included in the analyses after filtering those with a rate of missing data above 0.6. State abbreviations are the following: British Columbia (BC) in Canada; Oregon (OR) and California (CA) in the USA; and Baja California Norte (BCN) in Mexico. For additional locality information see Table S1 in the Supplementary Information. The abbreviation n.b.s denotes UCSD individuals sampled during the non-breeding season.

| Form            | State | Localities                    | $H_O$ | $H_E$ | $\pi$ | N (analyzed) |
|-----------------|-------|-------------------------------|-------|-------|-------|--------------|
| <i>oreganus</i> | BC    | Banks, Porcher, and Susan Is. | 0.177 | 0.181 | 0.188 | 16 (15)      |

| Form                     | State | Localities            | H <sub>O</sub> | H <sub>E</sub> | π     | N (analyzed) |
|--------------------------|-------|-----------------------|----------------|----------------|-------|--------------|
| <i>shufeldti</i>         | OR    | Willamette N. F.      | 0.180          | 0.179          | 0.188 | 12 (12)      |
| <i>montanus</i>          | OR    | Wallowa N. F.         | 0.176          | 0.179          | 0.187 | 16 (14)      |
| <i>pinosus</i>           | CA    | Santa Cruz Mts.       | 0.180          | 0.177          | 0.185 | 14 (12)      |
| <i>thurberi</i> (Tahoe)  | CA    | Tahoe                 | 0.179          | 0.182          | 0.188 | 18 (18)      |
| <i>thurberi</i> (Laguna) | CA    | Laguna Mts.           | 0.178          | 0.183          | 0.188 | 23 (21)      |
| UCSD n.b.s.              | CA    | San Diego             | -              | -              |       | 28 (24)      |
| UCSD breeders            | CA    | San Diego             | 0.166          | 0.164          | 0.171 | 18 (14)      |
| <i>pontilis</i>          | BCN   | Juarez Mts.           | 0.173          | 0.171          | 0.179 | 13 (12)      |
| <i>townsendi</i>         | BCN   | San Pedro Martir Mts. | 0.157          | 0.157          | 0.163 | 16 (16)      |
| <b>Total</b>             |       |                       |                |                |       | 174 (158)    |

**Table 2.** Pairwise  $F_{ST}$  values based on independent, selectively neutral SNP loci (below diagonal), and on RDA outlier SNP loci linked to functional genes (above diagonal) for all Oregon junco forms.

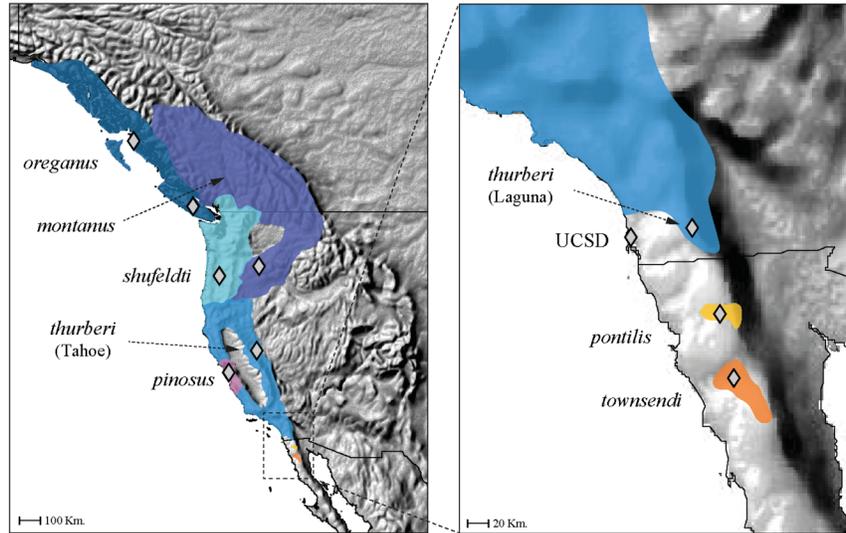
|                          | <i>oreganus</i> | <i>shufeldti</i> | <i>montanus</i> | <i>pinosus</i> | <i>thurberi</i> (Tahoe) | <i>thurberi</i> (Laguna) | UCSD  | <i>pontilis</i> |
|--------------------------|-----------------|------------------|-----------------|----------------|-------------------------|--------------------------|-------|-----------------|
| <i>oreganus</i>          |                 | 0.004            | 0.041           | 0.012          | 0.025                   | 0.000                    | 0.291 | 0.191           |
| <i>shufeldti</i>         | 0.014           |                  | 0.000           | 0.009          | 0.000                   | 0.000                    | 0.343 | 0.182           |
| <i>montanus</i>          | 0.013           | 0.014            |                 | 0.011          | 0.000                   | 0.008                    | 0.340 | 0.160           |
| <i>pinosus</i>           | 0.019           | 0.019            | 0.018           |                | 0.028                   | 0.000                    | 0.206 | 0.115           |
| <i>thurberi</i> (Tahoe)  | 0.012           | 0.012            | 0.011           | 0.016          |                         | 0.000                    | 0.342 | 0.145           |
| <i>thurberi</i> (Laguna) | 0.014           | 0.014            | 0.013           | 0.018          | 0.011                   |                          | 0.287 | 0.172           |
| UCSD                     | 0.026           | 0.026            | 0.026           | 0.027          | 0.023                   | 0.024                    |       | 0.275           |
| <i>pontilis</i>          | 0.022           | 0.023            | 0.022           | 0.027          | 0.020                   | 0.020                    | 0.033 |                 |
| <i>townsendi</i>         | 0.032           | 0.032            | 0.031           | 0.035          | 0.029                   | 0.030                    | 0.041 | 0.022           |

**Table 3.** AIC scores of all models and parameter estimates for the best model (Model A – IM) computed in fastSIMCOAL2 based on site frequency spectrum data. Models tested correspond to those in Fig. 2. IM and SI refer to ‘isolation with migration’ and ‘strict isolation’, respectively.  $\Delta$  Likelihood refers to the difference between the maximum observed and maximum estimated likelihood for our data under each model. Time of divergence is expressed in years, assuming a generation time of 1.5 years.

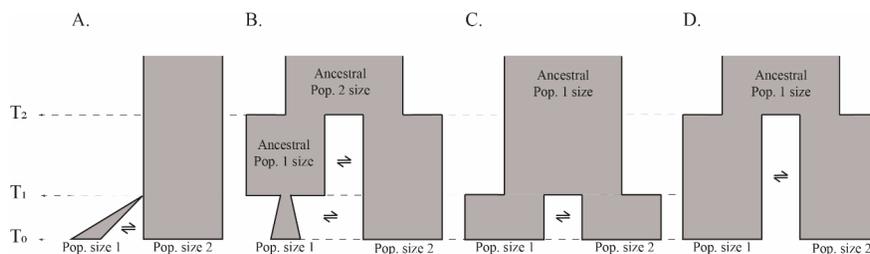
| Model                         | $\Delta$ Λικελιθοδ      | AIC  | $\Delta$ AIC                                       | $\Delta$ AIC |
|-------------------------------|-------------------------|--|--|--------------|
| Model A - IM                  | 67,732.987              | 939,268.006  | 0.000  | 0.000        |
| Model B - SI                  | 67,945.960              | 940,250.783  | 982.777  | 982.777      |
| Model B - IM                  | 67,956.734              | 940,308.399  | 1,040.393  | 1,040.393    |
| Model A - SI                  | 68,033.065              | 940,645.916  | 1,377.910  | 1,377.910    |
| Model D - SI                  | 68,106.265              | 940,983.015  | 1,715.009  | 1,715.009    |
| Model D - IM                  | 68,107.213              | 940,991.381  | 1,723.374  | 1,723.374    |
| Model C - IM                  | 68,260.534              | 941,697.450  | 2,429.444  | 2,429.444    |
| Model C - IM                  | 68,312.247              | 941,931.597  | 2,663.591  | 2,663.591    |
| <b>Model A Parameters</b>     | <b>Estimate</b>         | <b>CI 95%</b>                                      | <b>CI 95%</b>                                      |              |
| N <sub>e</sub> UCSD           | 252                     | 215 - 274  | 215 - 274  |              |
| N <sub>e</sub> <i>pinosus</i> | 5,859,774               | 5,857,240 - 5,865,925                              | 5,857,240 - 5,865,925                              |              |
| N <sub>e</sub> colonizers     | 2                       | 2 - 3  | 2 - 3  |              |
| Time of divergence            | 44                      | 30 - 47  | 30 - 47  |              |
| MIG <i>pinosus</i> - UCSD     | 6.02 x 10 <sup>-2</sup> | 5.87 x 10 <sup>-2</sup> - 8.03 x 10 <sup>-2</sup>  | 5.87 x 10 <sup>-2</sup> - 8.03 x 10 <sup>-2</sup>  |              |
| MIG UCSD - <i>pinosus</i>     | 2.01 x 10 <sup>-6</sup> | 9.38 x 10 <sup>-10</sup> - 2.57 x 10 <sup>-5</sup> | 9.38 x 10 <sup>-10</sup> - 2.57 x 10 <sup>-5</sup> |              |

| Model            | $\Delta$ Likelihood | AIC           | $\Delta$ AIC* | $\Delta$ AIC* |
|------------------|---------------------|---------------|---------------|---------------|
| UCSD growth rate | 0.159               | 0.150 – 0.244 | 0.150 – 0.244 |               |

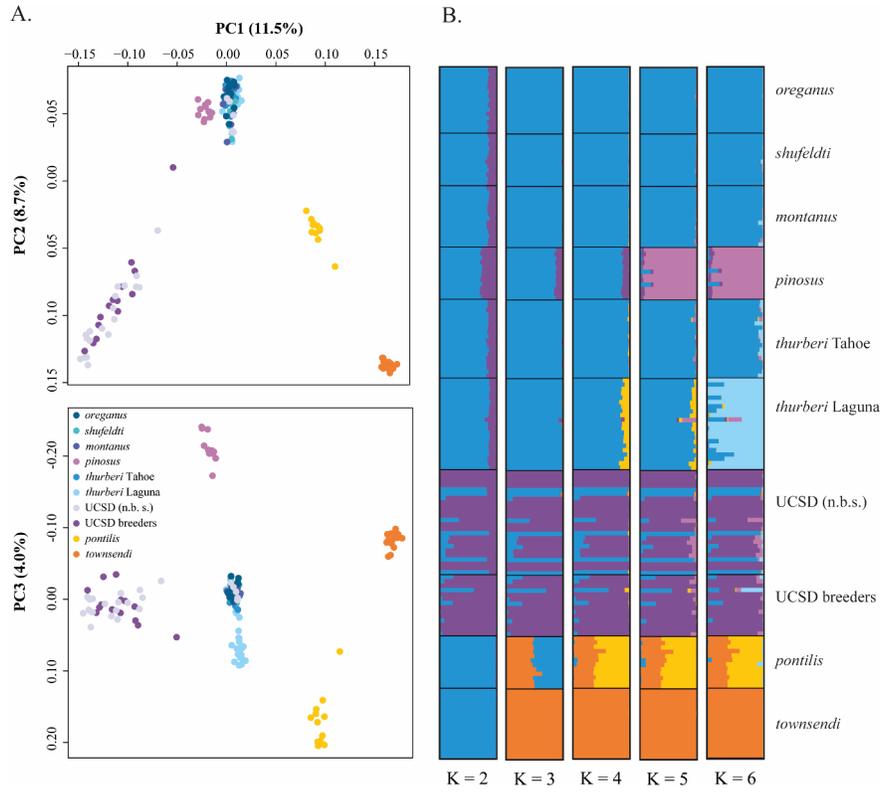
**Figure 1. Geographic distribution of the Oregon junco forms.** Breeding ranges (colored areas) and sampling sites (diamonds) of all the Oregon junco forms and samples included in this study.



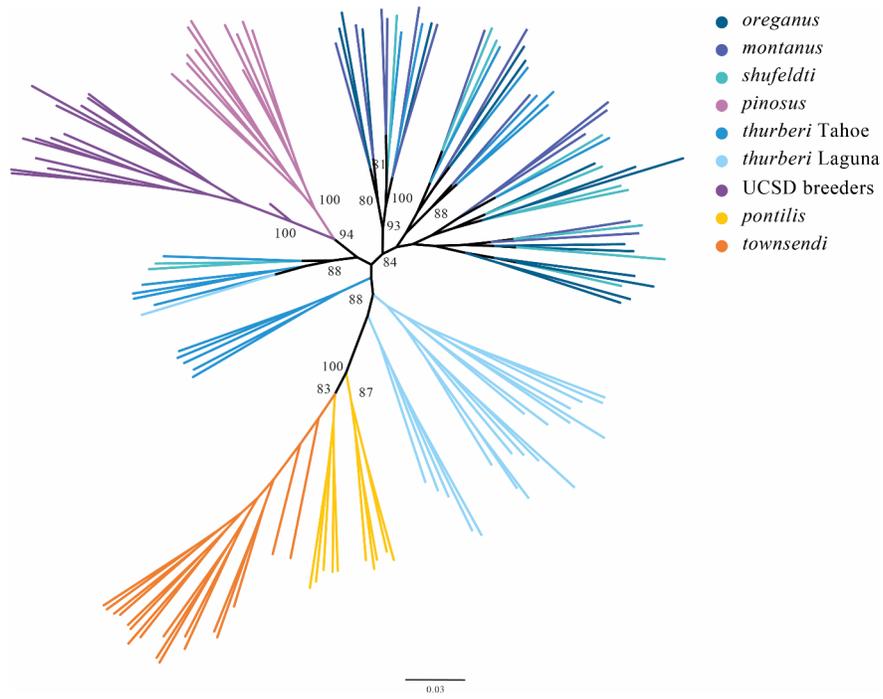
**Figure 2. Demographic inference in UCSD juncos.** Summary of the four demographic models and parameters implemented in fastSIMCOAL2 for likelihood comparison using genome-wide SNP data. Branch width represents population size. Values of T correspond to the present time ( $T_0$ ); a recent split/founder event about [20 - 100] generations ago ( $T_1$ ); and a postglacial split about [5 - 20]  $\times 10^3$  generations ago ( $T_2$ ).



**Figure 3. Neutral genetic structure of the Oregon junco forms.** (A) Plot of axes PC1 against PC2 (above) and PC3 (below) from a PCA based on 21,879 selectively neutral genome-wide SNPs. Percentage of explained variation is shown in parentheses. Colors match those in Fig. 1. (B) STRUCTURE analysis based on the same SNP set for K values 2 to 6. On each vertical plot, individuals are represented by horizontal bars, and colors correspond to the posterior assignment probabilities of belonging to each genetic cluster. The abbreviation n.b.s denotes UCSD individuals sampled during the non-breeding season.

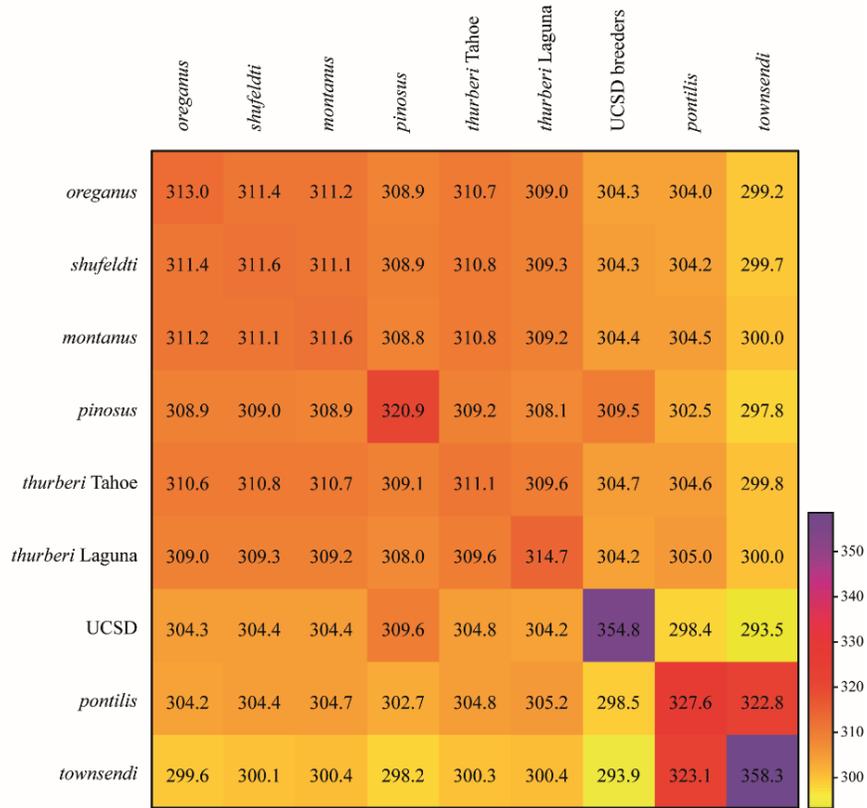


**Figure 4. Phylogenetic relationships among Oregon junco forms.** Unrooted maximum likelihood phylogeny based on 14,175 selectively neutral SNP loci. Bootstrap values equal to or over 80% are shown. Branch colors correspond to those in Fig. 1.

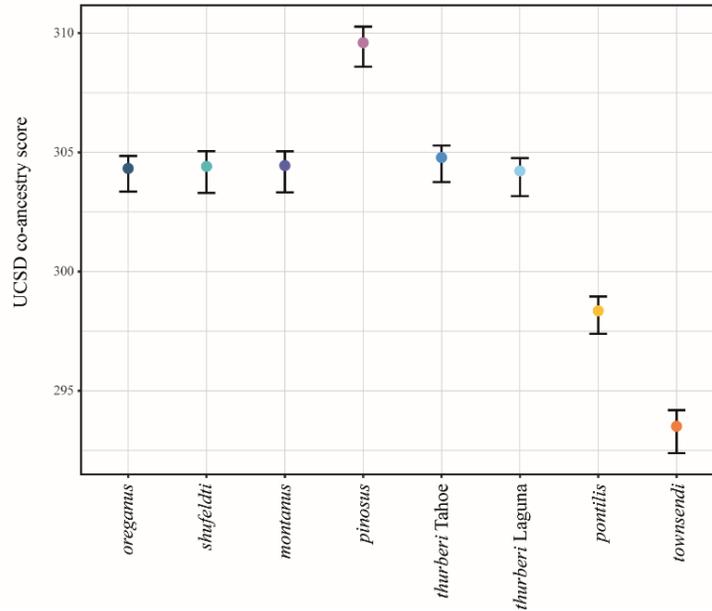


**Figure 5. Co-ancestry among Oregon junco populations.** (A) Individual-based pairwise co-ancestry scores based on 41,231 SNPs, computed with RADpainter and averaged over Oregon junco forms. (B) Co-ancestry scores for UCSD individuals and each of the other Oregon junco populations with 95% confidence intervals obtained by bootstrap.

A.

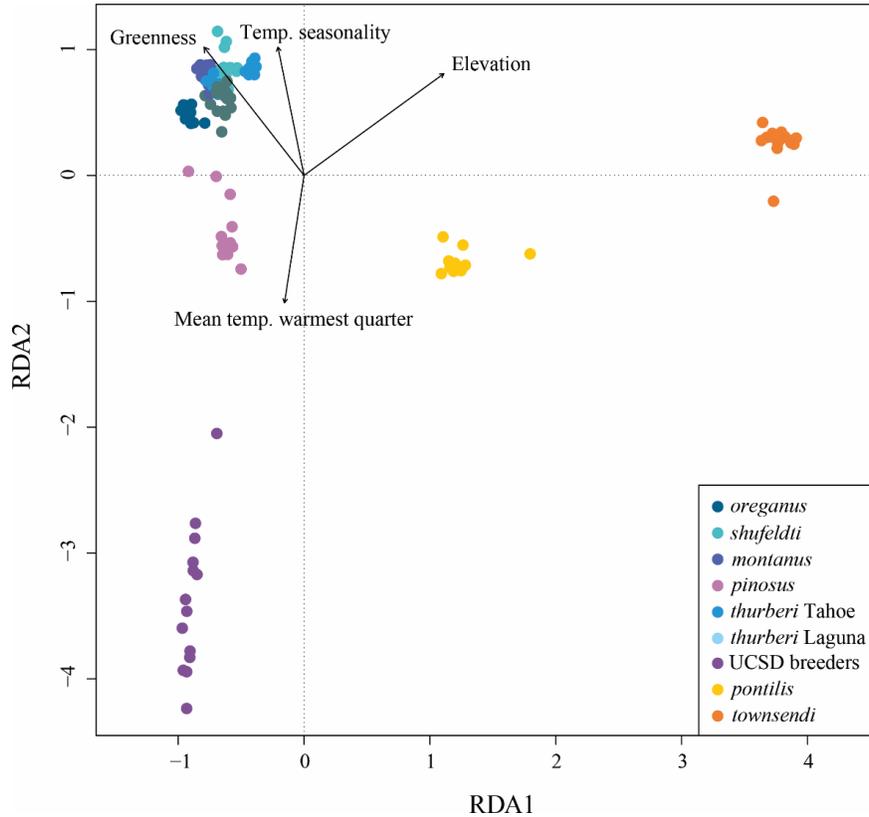


B.



**Figure 6. Signals of selection in UCSD juncos.** Redundancy analysis (RDA) between 132,343 SNP response variables and four explanatory variables corresponding to environmental factors (mean temperature of the warmest quarter, temperature seasonality, greenness, and elevation) for Oregon junco forms. Points

represent the projection of individual genotypes on the first two RDA axes. Marker colors correspond to those on the range map in Fig. 1. The explanatory variables are shown within the space defined by RDA1 and RDA2 by labeled vectors, and their contribution to each axis is represented by the length of their orthogonal projections over the scale bars along the x and y axes. Vector arrows indicate the direction of the gradient of variation for the corresponding environmental parameter. The loading score of each sample point on each explanatory variable can be obtained by an orthogonal projection on the corresponding plotted vector.



## Literature cited

Abbott R, Albach D, Ansell S, *et al.* (2013) Hybridization and speciation. *Journal of Evolutionary Biology* **26** , 229-246.

Akaike H (1998) Information theory and an extension of the maximum likelihood principle. In: *Selected papers of hirotugu akaike* , pp. 199-213. Springer.

Andrews S (2010) FastQC: a quality control tool for high throughput sequence data.

Atwell JW, Cardoso GC, Whittaker DJ, *et al.* (2012) Boldness behavior and stress physiology in a novel urban environment suggest rapid correlated evolutionary adaptation. *Behavioral Ecology* **23** , 960-969.

Atwell JW, Cardoso GC, Whittaker DJ, Price TD, Ketterson ED (2014) Hormonal, behavioral, and life-history traits exhibit correlated shifts in relation to population establishment in a novel environment. *The American Naturalist* **184** , E147-E160.

Auwerwa GA, Carneiro MO, Hartl C, *et al.* (2013) From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. *Current protocols in bioinformatics* , 11.10. 11-11.10. 33.

- Backström N, Qvarnström A, Gustafsson L, Ellegren H (2006) Levels of linkage disequilibrium in a wild bird population. *Biology Letters* **2** , 435-438.
- Beichman AC, Huerta-Sanchez E, Lohmueller KE (2018) Using genomic data to infer historic population dynamics of nonmodel organisms. *Annual Review of Ecology, Evolution, and Systematics* .
- Boileau MG, Herbert PND, Schwartz SS (1992) Non-equilibrium gene frequency divergence: persistent founder events in natural populations. *Journal of Evolutionary Biology* **5** , 25-39.
- Boortz KA, Syring KE, Dai C, *et al.* (2016) G6PC2 modulates fasting blood glucose in male mice in response to stress. *Endocrinology* **157** , 3002-3008.
- Borcard D, Gillet F, Legendre P (2011) Chapter 6. Canonical Ordination. In: *Numerical Ecology with R* , pp. 153-226. Springer.
- Bryant D, Moulton V (2004) Neighbor-Net: An agglomerative method for the construction of phylogenetic networks. *Molecular Biology and Evolution* **21** , 255-265.
- Burger J, Gochfeld M (1988) Effects of lead on growth in young herring gulls (*Larus argentatus*). *Journal of toxicology and environmental health* **25** , 227-236.
- Campbell-Staton SC, Winchell KM, Rochette NC, *et al.* (2020) Parallel selection on thermal physiology facilitates repeated adaptation of city lizards to urban heat islands. *Nature Ecology & Evolution* **4** , 652-658.
- Cardoso GC, Atwell JW (2010) Directional cultural change by modification and replacement of memes. *Evolution: International Journal of Organic Evolution* **65** , 295-300.
- Cardoso GC, Atwell JW (2011) On the relation between loudness and the increased song frequency of urban birds. *Animal Behaviour* **82** , 831-836.
- Carson HL (1975) The genetics of speciation at the diploid level. *The American Naturalist* **109** , 83-92.
- Carson HL, Kaneshiro KY (1976) *Drosophila* of Hawaii: systematics and ecological genetics. *Annual Review of Ecology and Systematics* **7** , 311-345.
- Chamberlain C, Blum J, Holmes RT, *et al.* (1996) The use of isotope tracers for identifying populations of migratory birds. *Oecologia* **109** , 132-141.
- Charlesworth B (2009) Effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics* **10** , 195-205.
- Chen Y, Shenkar N, Ni P, *et al.* (2018) Rapid microevolution during recent range expansion to harsh environments. *BMC Evolutionary Biology* **18** , 187.
- Colautti RI, Lau JA (2015) Contemporary evolution during invasion: evidence for differentiation, natural selection, and local adaptation. *Molecular Ecology* **24** , 1999-2017.
- Danecek P, Auton A, Abecasis G, *et al.* (2011) The variant call format and VCFtools. *Bioinformatics* **27** , 2156-2158.
- Danecek P, McCarthy SA (2017) BCFtools/csq: haplotype-aware variant consequences. *Bioinformatics* **33** , 2037-2039.
- DePristo MA, Banks E, Poplin R, *et al.* (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics* **43** , 491-498.
- Dwight J (1918) The geographical distribution of color and of other variable characters in the genus *Junco* : a new aspect of specific and subspecific values. *Bull. Am Mus. Nat. Hist.* **38** , 269-309.
- Eeva T, Lehikoinen E (1996) Growth and mortality of nestling great tits (*Parus major*) and pied flycatchers (*Ficedula hypoleuca*) in a heavy metal pollution gradient. *Oecologia* **108** , 631-639.

- Eriksson H, Wiktelius S (2011) Impact of chlorpyrifos used for desert locust control on non-target organisms in the vicinity of mangrove, an ecologically sensitive area. *International Journal of Pest Management* **57** , 23-34.
- Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll M (2013) Robust demographic inference from genomic and SNP data. *Plos Genetics* **9** , e1003905.
- Feng S, Stiller J, Deng Y, *et al.* (2020) Dense sampling of bird diversity increases power of comparative genomics. *Nature* **587** , 252-257.
- Forester BR, Jones MR, Joost S, Landguth EL, Lasky JR (2016) Detecting spatial genetic signatures of local adaptation in heterogeneous landscapes. *Molecular Ecology* **25** , 104-120.
- Forester BR, Lasky JR, Wagner HH, Urban DL (2018) Comparing methods for detecting multilocus adaptation with multivariate genotype-environment associations. *Molecular Ecology* .
- Frichot E, Schoville SD, de Villemereuil P, Gaggiotti OE, Francois O (2015) Detecting adaptive evolution based on association with ecological gradients: orientation matters! *Heredity* **115** , 22-28.
- Friis G, Aleixandre P, Rodriguez-Estrella R, Navarro-Siguenza AG, Mila B (2016) Rapid postglacial diversification and long-term stasis within the songbird genus *Junco* : phylogeographic and phylogenomic evidence. *Molecular Ecology* **25** , 6175-6195.
- Friis G, Fandos G, Zellmer AJ, *et al.* (2018) Genome-wide signals of drift and local adaptation during rapid lineage divergence in a songbird. *Molecular Ecology* **27** , 5137-5153.
- Friis G, Mila B (2020) Change in sexual signalling traits outruns morphological divergence across an ecological gradient in the post-glacial radiation of the songbird genus *Junco*. *Journal of Evolutionary Biology* **33** , 1276-1293.
- Fudickar AM, Greives TJ, Abolins-Abols M, *et al.* (2017) Mechanisms associated with an advance in the timing of seasonal reproduction in an urban songbird. *Frontiers in Ecology and Evolution* **5** , 85.
- Geens A, Dauwe T, Bervoets L, Blust R, Eens M (2010) Haematological status of wintering great tits (*Parus major*) along a metal pollution gradient. *Science of the Total Environment* **408** , 1174-1179.
- Goudet J, Jombart T, Goudet MJ (2015) Package ‘hierfstat’.
- Grant PR, Grant BR, Petren K (2001) A population founded by a single pair of individuals: establishment, expansion, and evolution. In: *Microevolution rate, pattern, process* , pp. 359-382. Springer.
- Guo L, Costanzo-Garvey DL, Smith DR, *et al.* (2017) Kinase Suppressor of Ras 2 (KSR2) expression in the brain regulates energy balance and glucose homeostasis. *Molecular metabolism* **6** , 194-205.
- Hendry AP, Kinnison MT (1999) Perspective: the pace of modern life: measuring rates of contemporary microevolution. *Evolution* **53** , 1637-1653.
- Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* **25** , 1965-1978.
- Huang T, Ma J, Gong Y, Feng Y (2019) Polymorphisms in the ovoinhibitor gene (OIH) and their association with egg quality of Xinhua E-strain chickens. *British Poultry Science* **60** , 88-93.
- Huey RB, Gilchrist GW, Carlson ML, Berrigan D, Serra L (2000) Rapid evolution of a geographic cline in size in an introduced fly. *Science* **287** , 308-309.
- Jensen AJ, Hansen LP, Johnsen BO, Karlsson S (2017) Rapid evolution of genetic and phenotypic divergence in Atlantic salmon following the colonisation of two new branches of a watercourse. *Genetics Selection Evolution* **49** , 22.
- Johnson MT, Munshi-South J (2017) Evolution of life in urban environments. *Science* **358** .

- Kamvar ZN, Lopez-Urbe MM, Coughlan S, *et al.* (2017) Developing educational resources for population genetics in R: an open and collaborative approach. *Molecular Ecology Resources* **17** , 120-128.
- Keller S, Taylor D (2010) Genomic admixture increases fitness during a biological invasion. *Journal of Evolutionary Biology* **23** , 1720-1731.
- Kharkovets T, Hardelin J-P, Safieddine S, *et al.* (2000) KCNQ4, a K<sup>+</sup> channel mutated in a form of dominant deafness, is expressed in the inner ear and the central auditory pathway. *Proceedings of the National Academy of Sciences* **97** , 4333-4338.
- Kliber A, Eckert CG (2005) Interaction between founder effect and selection during biological invasion in an aquatic plant. *Evolution* **59** , 1900-1913.
- Kolbe JJ, Leal M, Schoener TW, Spiller DA, Losos JB (2012) Founder effects persist despite adaptive differentiation: a field experiment with lizards. *Science* **335** , 1086-1089.
- Koskinen MT, Haugen TO, Primmer CR (2002) Contemporary fisherian life-history evolution in small salmonid populations. *Nature* **419** , 826-830.
- Krueger F (2015) Trim Galore!: A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files.
- Lamichhaney S, Berglund J, Almen MS, *et al.* (2015) Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature* **518** , 371-375.
- Lande R (1980) Genetic variation and phenotypic evolution during allopatric speciation. *The American Naturalist* **116** , 463-479.
- Lange M, Kaynak B, Forster UB, *et al.* (2008) Regulation of muscle development by DPF3, a novel histone acetylation and methylation reader of the BAF chromatin remodeling complex. *Genes & development* **22** , 2370-2384.
- Legendre P, Legendre L (1998) Numerical ecology: second English edition. *Developments in environmental modelling* **20** .
- Lerner HRL, Meyer M, James HF, Hofreiter M, Fleischer RC (2011) Multilocus resolution of phylogeny and timescale in the extant adaptive radiation of hawaiian honeycreepers. *Current Biology* **21** , 1838-1844.
- Lescak EA, Bassham SL, Catchen J, *et al.* (2015) Evolution of stickleback in 50 years on earthquake-uplifted islands. *Proceedings of the National Academy of Sciences* **112** , E7204-E7212.
- Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25** , 1754-1760.
- Liu Z, Li S, Wang W, *et al.* (2011) Parallel evolution of KCNQ4 in echolocating bats. *PLOS One* **6** , e26618.
- Lombal AJ, O'dwyer JE, Friesen V, Woehler EJ, BurrIDGE CP (2020) Identifying mechanisms of genetic differentiation among populations in vagile species: historical factors dominate genetic differentiation in seabirds. *Biological Reviews* **95** , 625-651.
- Luu K, Bazin E, Blum MG (2017) pcadapt: an R package to perform genome scans for selection based on principal component analysis. *Molecular Ecology Resources* **17** , 67-77.
- Malinsky M, Trucchi E, Lawson DJ, Falush D (2018) RADpainter and fineRADstructure: population inference from RADseq data. *Molecular Biology and Evolution* **35** , 1284-1290.
- Mathys BA, Lockwood JL (2011) Contemporary morphological diversification of passerine birds introduced to the Hawaiian archipelago. *Proceedings of the Royal Society B: Biological Sciences* **278** , 2392-2400.
- Matschinsky FM (2005) Glucokinase, glucose homeostasis, and diabetes mellitus. *Current diabetes reports* **5** , 171-176.

- Mayr E (1963) *Animal species and evolution* Belknap Press, Cambridge, MA.
- McCaskie G (1986) Wood warblers to finches. *American Birds***40** , 1257.
- McCoy RC, Garud NR, Kelley JL, Boggs CL, Petrov DA (2014) Genomic inference accurately predicts the timing and severity of a recent bottleneck in a nonmodel insect population. *Molecular Ecology***23** , 136-150.
- McKenna A, Hanna M, Banks E, *et al.* (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* **20** , 1297-1303.
- Mila B, McCormack JE, Castaneda G, Wayne RK, Smith TB (2007) Recent postglacial range expansion drives the rapid diversification of a songbird lineage in the genus *Junco*. *Proceedings of the Royal Society B-Biological Sciences* **274** , 2653-2660.
- Miller A (1941) Speciation in the avian genus *Junco* .*University of California Publications in Zoology* **44** , 173-434.
- Millien V (2006) Morphological evolution is accelerated among island mammals. *PLOS Biology* **4** .
- Murray KD, Borevitz JO (2017) Axe: rapid, competitive sequence read demultiplexing using a trie. *bioRxiv* , 160606.
- Nam K, Mugal C, Nabholz B, *et al.* (2010) Molecular evolution of genes in avian genomes. *Genome biology* **11** , R68.
- Nguyen L-T, Schmidt HA, Von Haeseler A, Minh BQ (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* **32** , 268-274.
- Nolan VJ, Ketterson ED, Cristol DA, *et al.* (2002) Dark-eyed Junco (*Junco hyemalis* ). In: *The Birds of North America* (eds. Poole A, Gill F). The Birds of North America, Inc., Philadelphia, Pennsylvania.
- O'Brien RM (2013) Moving on from GWAS: functional studies on the G6PC2 gene implicated in the regulation of fasting blood glucose. *Current diabetes reports* **13** , 768-777.
- Ocampo Daza D, Larhammar D (2018) Evolution of the receptors for growth hormone, prolactin, erythropoietin and thrombopoietin in relation to the vertebrate tetraploidizations. *General and Comparative Endocrinology* **257** , 143-160.
- Perrier C, Caizergues A, Charmantier A (2020) Adaptation genomics in urban environments. Oxford University Press.
- Price T, Phillimore A, Awodey M, Hudson R (2010) Ecological and geographical influences on the allopatric phase of island speciation. *In search of the causes of evolution: From field observations to mechanisms* , 251-281.
- Price TD, Yeh PJ, Harr B (2008) Phenotypic plasticity and the evolution of a socially selected trait following colonization of a novel environment. *American Naturalist* **172** , S49-S62.
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* **155** , 945-959.
- R\_Core\_Team (2019) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- R\_Studio\_Team (2015) RStudio: Integrated Development for R. R Studio, Inc., Boston, MA.
- Rakvac Z, Kucsma N, Gera M, *et al.* (2019) The human ABCB6 protein is the functional homologue of HMT-1 proteins mediating cadmium detoxification. *Cellular and Molecular Life Sciences* **76** , 4131-4144.
- Rasner CA, Yeh P, Eggert LS, *et al.* (2004) Genetic and morphological evolution following a founder event in the dark-eyed junco, *Junco hyemalis thurberi* . *Molecular Ecology***13** , 671-681.

- Reichard DG, Atwell JW, Pandit MM, *et al.* (2020) Urban birdsongs: higher minimum song frequency of an urban colonist persists in a common garden experiment. *Animal Behaviour* **170** , 33-41.
- Reznick DN, Ghalambor CK (2001) The population ecology of contemporary adaptations: what empirical studies reveal about the conditions that promote adaptive evolution. In: *Microevolution rate, pattern, process* , pp. 183-198. Springer.
- Rochette NC, Rivera-Colon AG, Catchen JM (2019) Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. *Molecular Ecology* **28** , 4737-4754.
- Rosado IV, Niedzwiedz W, Alpi AF, Patel KJ (2009) The Walker B motif in avian FANCM is required to limit sister chromatid exchanges but is dispensable for DNA crosslink repair. *Nucleic Acids Research* **37** , 4360-4370.
- Rundle HD, Nosil P (2005) Ecological speciation. *Ecology Letters* **8** , 336-352.
- Salmon P, Jacobs A, Ahren D, *et al.* (2021) Continent-wide genomic signatures of adaptation to urbanisation in a songbird across Europe. *Nature Communications* **12** , 1-14.
- Sato A, Tichy H, O'HUigin C, *et al.* (2001) On the Origin of Darwin's Finches. *Molecular Biology and Evolution* **18** , 299-311.
- Schluter D (2000) *The ecology of adaptive radiation* Oxford University Press, Oxford.
- Seehausen O (2006) African cichlid fish: a model system in adaptive radiation research. *Proceedings of the Royal Society B* **273** , 1987-1998.
- Sendell-Price AT, Ruegg KC, Clegg SM (2020) Rapid morphological divergence following a human-mediated introduction: the role of drift and directional selection. *Heredity* **124** , 535-549.
- Seutin G (1991) Preservation of avian blood and tissue samples for DNA analyses. *Canadian Journal of Zoology* **69** , 82-90.
- Simpson G (1953) The major features of evolution. *and G. Simpson, Tempo and Mode in* , 143.
- Slabbekoorn H, Yeh P, Hunt K (2007) Sound transmission and song divergence: a comparison of urban and forest acoustics. *The Condor* **109** , 67-78.
- Slatkin M (1987) Gene flow and the geographic structure of animal populations. *Science* **236** , 787-792.
- Smeds L, Qvarnstrom A, Ellegren H (2016) Direct estimate of the rate of germline mutation in a bird. *Genome Research* **26** , 1211-1218.
- Szulkin M, Garroway CJ, Corsini M, *et al.* (2020a) How to quantify urbanisation when testing for urban evolution? *Urban evolutionary biology* **13** .
- Szulkin M, Munshi-South J, Charmantier A (2020b) *Urban evolutionary biology* Oxford University Press, USA.
- Takekawa JY, Wainwright-De La Cruz S, Hothem R, Yee J (2002) Relating body condition to inorganic contaminant concentrations of diving ducks wintering in coastal California. *Archives of Environmental Contamination and Toxicology* **42** , 60-70.
- Tatusova TA, Madden TL (1999) BLAST 2 Sequences, a new tool for comparing protein and nucleotide sequences. *FEMS microbiology letters* **174** , 247-250.
- Templeton AR (1980) The theory of speciation via the founder principle. *Genetics* **94** , 1011-1038.
- Unitt P (1984) *The birds of San Diego county* San Diego Society of.
- Van Den Wollenberg AL (1977) Redundancy analysis an alternative for canonical correlation analysis. *Psychometrika* **42** , 207-219.

- Wang IJ, Bradburd GS (2014) Isolation by environment. *Molecular Ecology* **23** , 5649-5662.
- Wassenaar L, Hobson K (2003) Comparative equilibration and online technique for determination of non-exchangeable hydrogen of keratins for use in animal migration studies. *Isotopes in environmental and health studies* **39** , 211-217.
- Wegener JE, Pita-Aquino JN, Atutubo J, Moreno A, Kolbe JJ (2019) Hybridization and rapid differentiation after secondary contact between the native green anole (*Anolis carolinensis*) and the introduced green anole (*Anolis porcatu*s). *Ecology and Evolution* **9** , 4138-4148.
- Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution*. **38** , 1358-1370.
- Wessel A, Hoch H, Asche M, *et al.* (2013) Founder effects initiated rapid species radiation in Hawaiian cave planthoppers. *Proceedings of the National Academy of Sciences* **110** , 9391-9396.
- Whittaker DJ, Dapper AL, Peterson MP, Atwell JW, Ketterson ED (2012) Maintenance of MHC Class IIB diversity in a recently established songbird population. *Journal of Avian Biology* **43** , 109-118.
- Wright S (1931) Evolution in Mendelian populations. *Genetics* **16** , 97.
- Wu Y, Liang H, Zhang H, *et al.* (2018) The Differential Expression and SNP Analysis of the Ovoinhibitor Gene in the Ovaries of Laying Duck Breeds (*Anas Platyrhynchos* ). *Brazilian Journal of Poultry Science* **20** , 281-286.
- Wunder MB, Jehl Jr JR, Stricker CA (2012) The early bird gets the shrimp: confronting assumptions of isotopic equilibrium and homogeneity in a wild bird population. *Journal of Animal Ecology* **81** , 1223-1232.
- Yeh PJ (2004) Rapid evolution of a sexually selected trait following population establishment in a novel habitat. *Evolution* **58** , 166-174.
- Yeh PJ, Price TD (2004) Adaptive phenotypic plasticity and the successful colonization of a novel environment. *The American Naturalist* **164** , 531-542.