Received: 20 September 2021 | Revised: 14 January 2022 | Accepted: 1 March 2022

DOI: 10.1111/mec.16422

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

Rapid evolutionary divergence of a songbird population following recent colonization of an urban area

Guillermo Friis[1] | Jonathan W. Atwell[2] | Adam M. Fudickar[2] | Timothy J. Greives[3] | Pamela J. Yeh[4] | Trevor D. Price[5] | Ellen D. Ketterson[2] | Borja Milá[1]

[1]National Museum of Natural Sciences, Department of Biodiversity and Evolutionary Biology, Spanish National Research Council (CSIC), Madrid, Spain

[2]Department of Biology, Indiana University, Bloomington, Indiana, USA

[3]Department of Biological Sciences, North Dakota State University, Fargo, North Dakota, USA

[4]Department of Ecology and Evolutionary Biology, University of California Los Angeles, Los Angeles, California, USA

[5]Department of Ecology and Evolution, University of Chicago, Chicago, Illinois, USA

Correspondence
Guillermo Friis, Center for Genomics and Systems Biology, New York University — Abu Dhabi, PO Box 129188, Abu Dhabi, United Arab Emirates.
Email: gfm3@nyu.edu

Funding information
Ministerio de Ciencia e Innovación, Grant/Award Number: CGL-2011-25866 and CGL-2015-66381; USA National Science Foundation, Grant/Award Number: IOS-1257527

Handling Editor: Yanhua qu

Abstract

Colonization of a novel environment by a few individuals can lead to rapid evolutionary change, yet there is scarce evidence of the relative contributions of neutral and selective factors in promoting divergence during the early stages of colonization. Here 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 geographical origin of the colonization and the contributing factors remained uncertain. Our genome-wide single nucleotide polymorphism data set 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 neighbouring mountain population. Demographic inference recovered a separation from *pinosus* as recent as 20–32 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.

KEYWORDS
founder effects, genetic drift, *Junco*, local adaptation, rapid divergence, urban colonization

1 | 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 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 timescales 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 timescales 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., 2020).

A potential case of contemporary colonization is provided by a suburban population of the dark-eyed junco (*Junco hyemalis*), a small passerine that inhabits mixed-coniferous forests in North America (Miller, 1941; Nolan et al., 2002). The dark-eyed junco is composed of various subspecies of postglacial origin that diversified during the northward recolonization of the North American continent as ice sheets receded ~18,000 years ago (Friis et al., 2016; Milá et al., 2007), resulting in a set of phenotypically differentiated forms. In Western North America, the "Oregon junco" form of the dark-eyed junco is divided into seven geographically structured subspecies (Dwight, 1918; Friis et al., 2018; Miller, 1941; Nolan et al., 2002), including, from south to north: *J. h. townsendi* and *J. h. pontilis* in northern Baja California, Mexico; *J. h. thurberi* in the mountains of southern to northern California; *J. h. pinosus* in the coastal region of central California; *J. h. montanus* in the interior regions of Oregon, Washington and British Columbia; *J. h. shufeldti* in coastal regions of Oregon and Washington; and *J. h. oreganus* in coastal British Columbia and southern Alaska (Miller, 1941; Nolan et al., 2002) (Figure 1). 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 up to 2014 (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 from other populations (Fudickar et al., 2017; Unitt, 1984; Yeh, 2004).

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 (Figure 1). This population belongs to the subspecies *thurberi* and inhabits elevations above 1500 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 on UCSD juncos published to date have focused on the evolution of social signalling traits during colonization

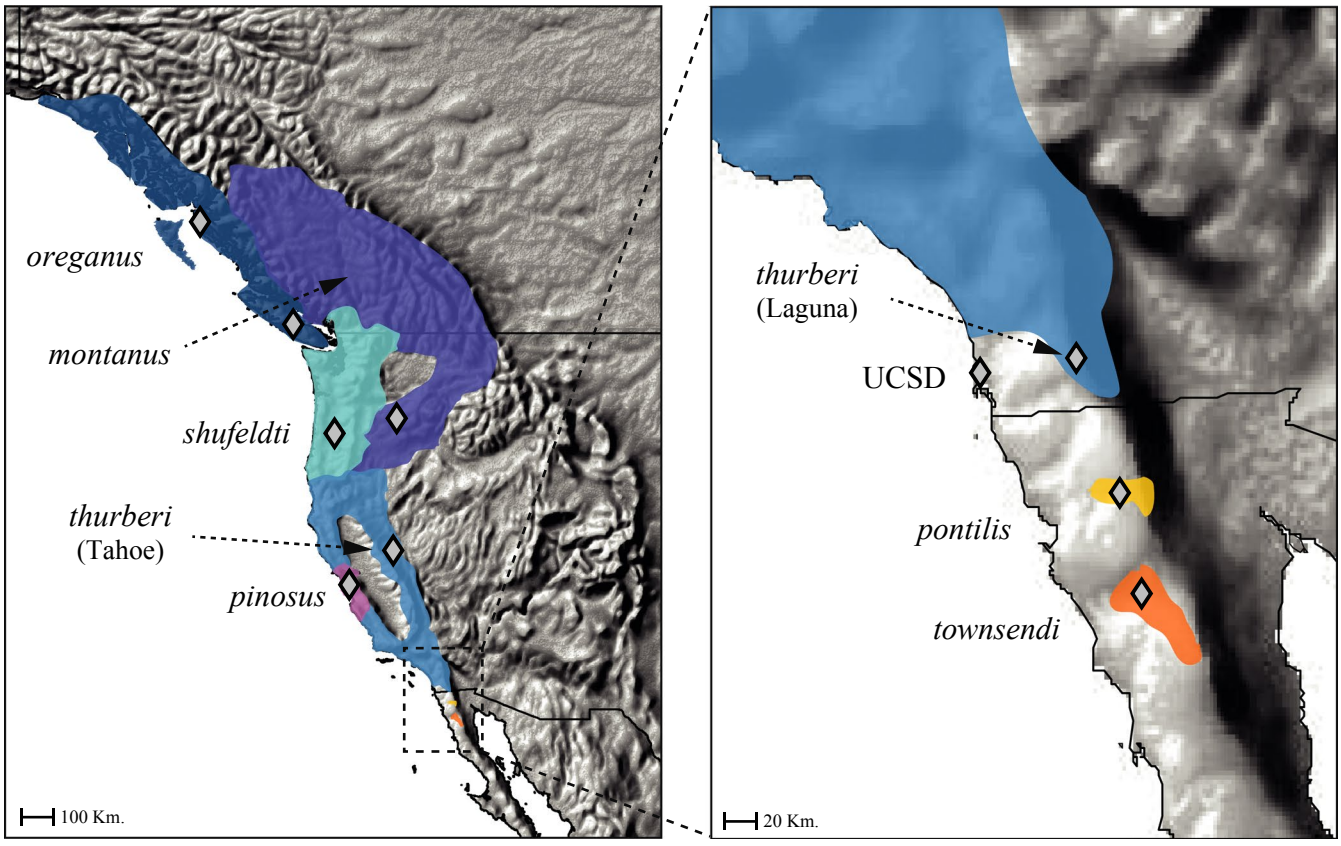

FIGURE 1 Geographic distribution of the Oregon junco subspecies. Breeding ranges (coloured areas) and sampling sites (diamonds) of all Oregon junco subspecies and samples included in this study. From north to south, sampling sites are located in the states of British Columbia (Canada); Oregon and California (USA); and Baja California (Mexico) [Colour figure can be viewed at wileyonlinelibrary.com]

of a novel environment (Price, Yeh & Harr, 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 (major histocompatibility complex) 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, 2014; Fudickar et al., 2017). Based on phenotypes of overwintering nonresident birds and proximity to potential sources, Yeh (2004) proposed that the population of UCSD juncos had probably 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 seven to 70 birds. Fudickar et al. (2017) used genome-wide single nucleotide polymorphism (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 postglacial radiation of the complex (Friis et al., 2016, 2018; Friis & Milá, 2020), suggesting that high-throughput SNP data may provide a level of phylogenetic resolution not afforded by genetic sampling of previous studies.

Here we use genome-wide SNP data and extensive geographical sampling of Oregon junco subspecies 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 ~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 *J. hyemalis* reference genome (Feng et al., 2020; Friis et al., 2018) to identify candidate loci under divergent selection.

2 | MATERIALS AND METHODS

2.1 | Population sampling

Oregon juncos were sampled across their breeding range using mist-nets and seed-baited walk-in traps during the breeding season, between April and June, and between 2004 and 2014. In addition to samples collected at UCSD in San Diego, California, we studied 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 (Figure 1). At UCSD, birds were sampled during both the breeding season and the nonbreeding season. Genetic and isotopic analyses were used to infer whether birds sampled in the nonbreeding season were local residents or wintering visitors breeding elsewhere (see below).

Birds were marked with uniquely numbered aluminium 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.

2.2 | Genotyping-by-sequencing

Genomic DNA was extracted from blood and tissue samples using a Qiagen DNeasy kit (Qiagen). 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 nonbreeding season (referred to as "UCSD n.b.s." hereafter). In addition, we sequenced 128 dark-eyed juncos belonging to the following subspecies within the Oregon junco group (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) (Figure 1, Table 1; Table S1). GBS libraries were prepared and sequenced at Cornell University's Institute for Genomic Diversity, using the restriction enzyme *Pst*I 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 valid barcoded single-end reads 100 bp in length per lane.

2.3 | 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 40–90 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 no. PRJNA493001; for

TABLE 1 Oregon junco subspecies, number of genotyped individuals and population genetic diversity indices per locality

Subspecies	State	Localities	H_O	H_E	π	N (analysed)
oreganus	BC	Banks, Porcher and Susan Islands	0.177	0.181	0.188	16 (15)
shufeldti	OR	Willamette N. F.	0.180	0.179	0.188	12 (12)
montanus	OR	Wallowa N. F.	0.176	0.179	0.187	16 (14)
pinosus	CA	Santa Cruz Mts	0.180	0.177	0.185	14 (12)
thurberi (Tahoe)	CA	Tahoe	0.179	0.182	0.188	18 (18)
thurberi (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)
pontilis	BCN	Juarez Mts	0.173	0.171	0.179	13 (12)
townsendi	BCN	San Pedro Martir Mts	0.157	0.157	0.163	16 (16)
Total						174 (158)

Note: 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 as follows: 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. The abbreviation n.b.s denotes UCSD individuals sampled during the nonbreeding season.

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 variant call format (*vcf*). We retained biallelic SNPs and applied GATK generic hard-filtering recommendations consisting of 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 60% were excluded at this point. Using VCFTOOLS (Danecek et al., 2011), we then excluded those SNPs that fell oustide a range of coverage between 4 and 50 or with a genotyping phred quality score below 40. The resulting data set (hereafter referred to as "Full Dataset" in downstream analyses) consisted of 158 individuals (Table 1) and 1,448,676 SNPs with an average coverage of 4.25 and a missing data rate of 0.65.

2.4 | Population structure and genetic diversity analyses

To explore genome-wide population structure among Oregon junco subspecies, 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 .2 in windows of 10,000 bp. Using VCFTOOLS, positions with <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, Bazin & Blum, 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). In addition, to test for a potential impact of missing data rates in variation patterns based on SNPs, we repeated the analysis applying missing thresholds of 75% and 90%, keeping the remaining quality filters unaltered. The PCAs were 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, Stephens & Donnelly, 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. We implemented the "admixture" model and assumed the SNPs to be independent, as linkage disequilibrium filters were applied when constructing the SNP data set. Sampling locality was not used as a prior in the analysis (we did not use the *locprior* setting). 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 StAMPP R package (Pembleton, Cogan & Forster, 2013).

Associated *p*-values were generated using 10,000 bootstrap steps. We also used the program POPULATIONS from the STACKS pipeline (Rochette, Rivera-Colón & Catchen, 2019) to compute the observed (H_O) and expected heterozygosity (H_E), and nucleotide diversity (π) based on autosomal positions for each subspecies.

2.5 | 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 again excluded the UCSD n.b.s. individuals from the "Neutral Dataset." A maximum-likelihood phylogeny and a neighbour-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 subspecies 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-neighbour 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 100,000 burn-in and 100,000 sampling iterations, and a thinning interval of 1000 to produce a matrix of pairwise individual co-ancestry estimates, which were averaged over subspecies. For estimating 95% confidence intervals around pairwise co-ancestry values between UCSD and the different Oregon junco subspecies, we applied a nonparametric bootstrap procedure. Bootstrapping was carried out by splitting the SNP matrix into 1000 blocks and randomly combining them for each of the repetitions. A total of 100 analyses were run for confidence interval estimation.

2.6 | Breeding latitude of UCSD juncos captured during the nonbreeding season

Hydrogen isotopic ratios in bird feathers correspond to those found in the localities where the feathers were grown, and because juncos moult 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 nonbreeding season (UCSD n.b.s.) were local breeders or winter visitors breeding elsewhere, we tested for correlation between deuterium isotope ratio (δD) and posterior likelihoods of genetic population assignment. We obtained isotopic data from the 28 individuals sampled in UCSD over winter, to compare with 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), as described in Wunder, Jehl and Stricker (2012). δD values are reported in per mil notation (‰) relative to V-SMOW, using internal standards (−78‰ and −172‰ respectively) calibrated to CFS-CHS-BWB (Wassenaar & Hobson, 2003). Benzoic acid (δD = −61‰) and IAEA-CH-7 (δ^2H = −100‰) were also analysed within analytical sequences with a precision of < ±4‰. An SNP subset of the "Neutral Dataset" was generated for 26 of the 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 δD abundance.

2.7 | Demographic inference and molecular dating

The marked genetic divergence of the UCSD juncos with respect to neighbouring 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 1980s 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 ~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 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 site frequency spectra (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 (Figure 2a); postglacial population split followed by a recent founder event (Figure 2b); recent

population split without a founder event (Figure 2c); and postglacial population split without a founder event (Figure 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 Figure 2) was estimated within a range of 35–280 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 two and 280. Mutation rate was fixed at 2.21×10^{-9} in all cases (Nam et al., 2010; Smeds, Qvarnström & Ellegren, 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 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, 1973). For estimating 95% support limits for the parameters estimated under the best model, we followed the same bootstrap procedure as for RADPAINTER using blocks of 51 SNPs to produce 100 bootstrapped data sets.

2.8 | Genotype–environment association analysis

We used 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 PCA. RDA estimates the variance of a set of response variables that is 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 georeferenced 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). The selected variables 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: per cent 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 11 variables. To reduce the number of explanatory variables while retaining those maximizing differences between habitats at UCSD and

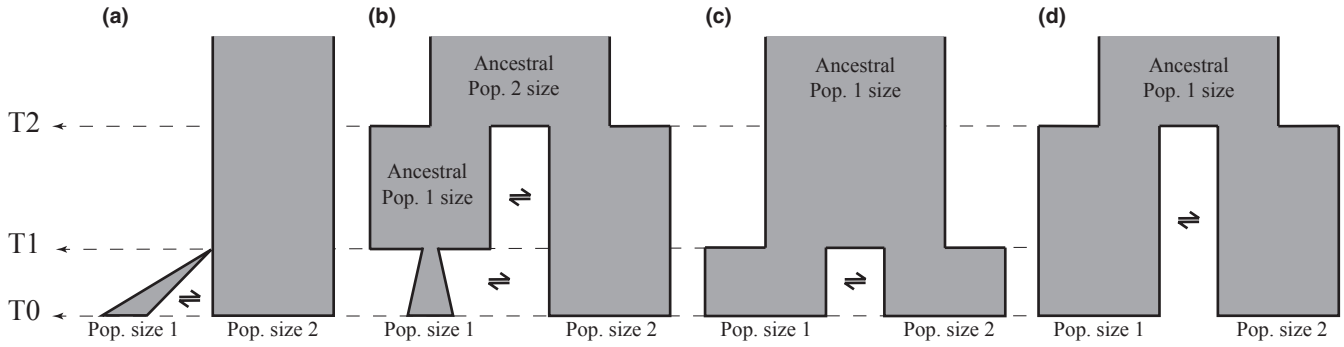

FIGURE 2 Demographic inference in UCSD juncos. Summary of the four demographic models and parameters implemented in FASTSIMCOAL2 for likelihood comparison using genome-wide SNP data. Branch width represents population size. Values of T correspond to the present time (T_0); a recent split/founder event about 20–100 generations ago (T_1); and a postglacial split about $5–20 \times 10^3$ generations ago (T_2)

elsewhere, we computed the sum of the pairwise Mahalanobis distances (Mahalanobis, 1936) between UCSD juncos and each of the other subspecies (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, so no variable presented a VIF > 2. The SNP data set 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 subspecies. The HWE filter was also applied for a final data set of 132,343 SNPs. Because RDA does not allow missing data, we replaced nongenotyped 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.

2.9 | Identification of candidate genes

To ask if signals of GEA 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 from the standard deviation from mean loading (Kamvar et al., 2017, tutorial by Forester in popgen.nescent.org/2018-03-27_RDA_GEA.html). We extracted 2000-bp regions flanking these SNPs (,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 subspecies 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 subspecies, as well as per locus between UCSD and its most closely related group (see Results).

All R analyses were carried out in RSTUDIO version 1.1.463 (R Studio Team, 2015) with R version 3.5.3 (R Core Team, 2019).

3 | RESULTS

3.1 | Neutral genetic structure among Oregon junco subspecies

A plot of PC1 against PC2 (11.5% and 8.7% of explained variance, respectively) revealed considerable differentiation among Oregon junco subspecies. 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 subspecies in Baja California, *pontilis* and *townsendi*. The subspecies *pinosus* also formed a separate cluster, although distance with respect to the remaining subspecies was less pronounced. Northern subspecies *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 (Figure 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 (Figure 3a, lower plot). The population structure signal in the PCA was robust, independently of the missing data filters applied (Figure S1).

The STRUCTURE analysis was consistent with PCA results and provided additional information on genetic differentiation among localities (Figure 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 $K = 4$ 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 subspecies *pinosus* formed a clearly differentiated cluster at $K = 5$. Birds from Laguna Mountains also appeared as a separate population at $K = 6$, again congruent 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*; Figure 3b).

Pairwise F_{ST} values among subspecies based on neutral loci were all highly significant (p-value range = [0–.005]) and generally congruent with the PCA and STRUCTURE results, with subspecies from Baja California and UCSD showing the highest degree of differentiation. The remaining subspecies 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 subspecies (Table 1).

3.2 | Breeding latitude of UCSD juncos sampled in the nonbreeding season

The linear regression between STRUCTURE scores for $K = 2$ and δD 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 (Figure S2). 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.

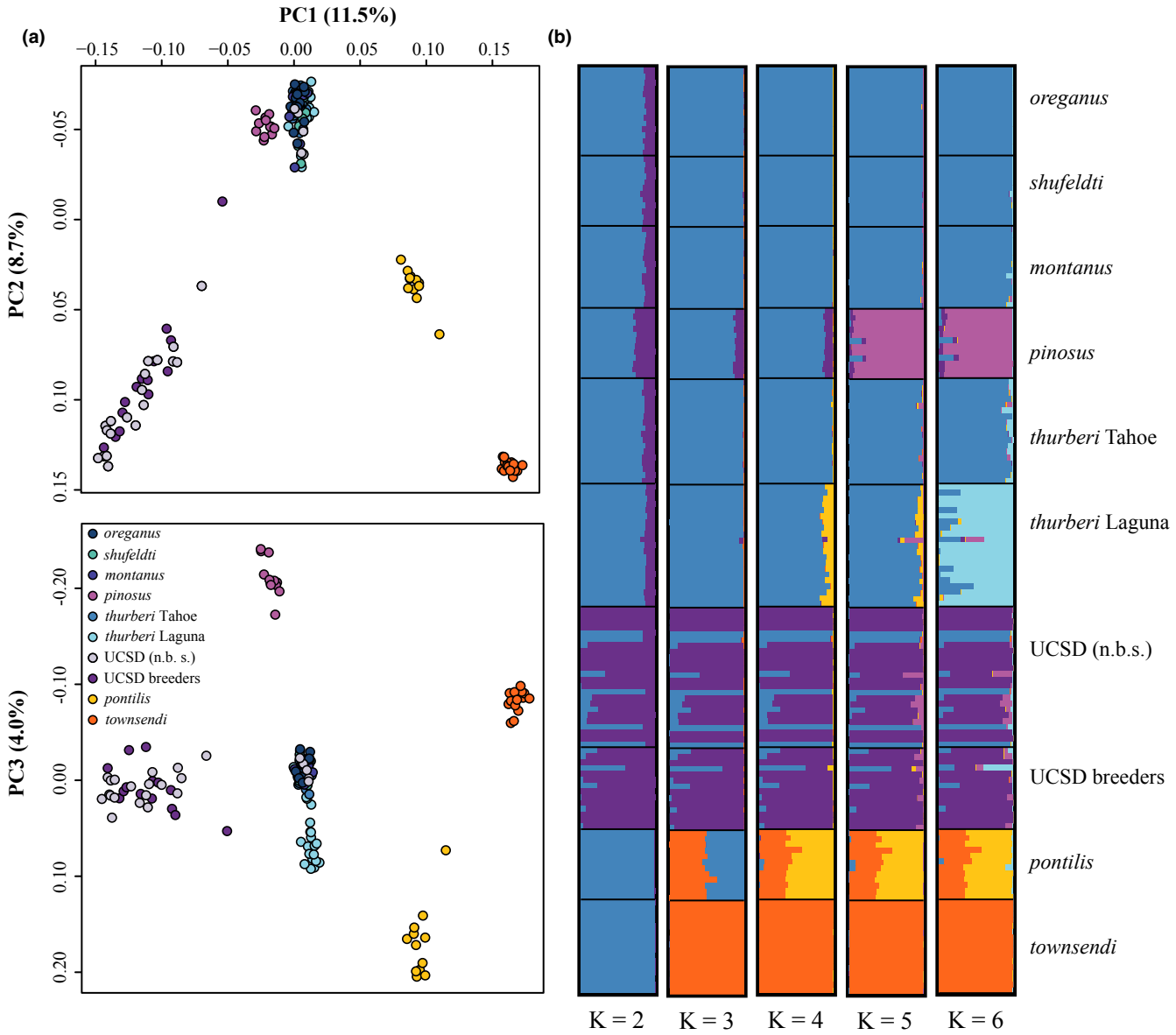

FIGURE 3 Neutral genetic structure of the Oregon junco subspecies. (a) Plot of axes PC1 against PC2 (above) and PC3 (below) from a PCA based on 21,879 selectively neutral genome-wide SNPs. Percentage of explained variation is shown in parentheses. Colours match those in Figure 1. (b) STRUCTURE analysis based on the same SNP set for $K = 2$–6. On each vertical plot, individuals are represented by horizontal bars, and colours correspond to the posterior assignment probabilities of belonging to each genetic cluster. The abbreviation "n.b.s" denotes UCSD individuals sampled during the nonbreeding season [Colour figure can be viewed at wileyonlinelibrary.com]

3.3 | Phylogenetic and co-ancestry analyses

Phylogenetic reconstructions based on maximum-likelihood and neighbour-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 (Figure 4). In turn, relatively low structure among northern Oregon juncos was detected in the analysis, in contrast to subspecies 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, Figure 3). In the neighbour-net, UCSD and *pinosus* also appear as sister groups, yet the degree of differentiation with respect to other subspecies is considerably less pronounced (Figure S3). Subspecies *townsendi* and *pontilis* appear as sister clades, highly divergent from the remaining junco subspecies. Southern *thurberi* individuals from Laguna Mountains cluster together as well in this analysis, while little differentiation is observed among northern subspecies.

The RADPAINTER genetic ancestry analysis consistently recovered high average values of intrapopulation co-ancestry in those subspecies showing marked differentiation and geographical isolation, namely *townsendi*, *pontilis*, *pinosus* and the UCSD population. UCSD

TABLE 2 Pairwise F_{ST} scores and associated significance 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 subspecies. $^{**}p < 1 \times 10^{-3}$; $^{*}p < 1 \times 10^{-2}$

	oreganus	*shufeldti*	*montanus*	*pinosus*	*thurberi* (Tahoe)	*thurberi* (Laguna)	UCSD	*pontilis*	*townsendi*
oreganus		0.004	0.041	0.012	0.025	−0.013	0.291**	0.191**	0.452**
shufeldti	0.003**		−0.008	0.009	−0.017	−0.011	0.343**	0.182**	0.451**
montanus	0.003**	0.002*		0.011	−0.006	0.008	0.340**	0.160**	0.420**
pinosus	0.020**	0.018**	0.018**		0.028	−0.003	0.206**	0.115**	0.411**
thurberi (Tahoe)	0.006**	0.002**	0.002**	0.017**		−0.003	0.342**	0.145**	0.406**
thurberi (Laguna)	0.013**	0.010**	0.009**	0.023**	0.008**		0.287**	0.172**	0.451**
UCSD	0.051**	0.047**	0.047**	0.050**	0.046**	0.050**		0.275**	0.540**
pontilis	0.036**	0.033**	0.032**	0.048**	0.031**	0.033**	0.073**		0.163**
townsendi	0.077**	0.075**	0.073**	0.087**	0.071**	0.074**	0.112**	0.045	

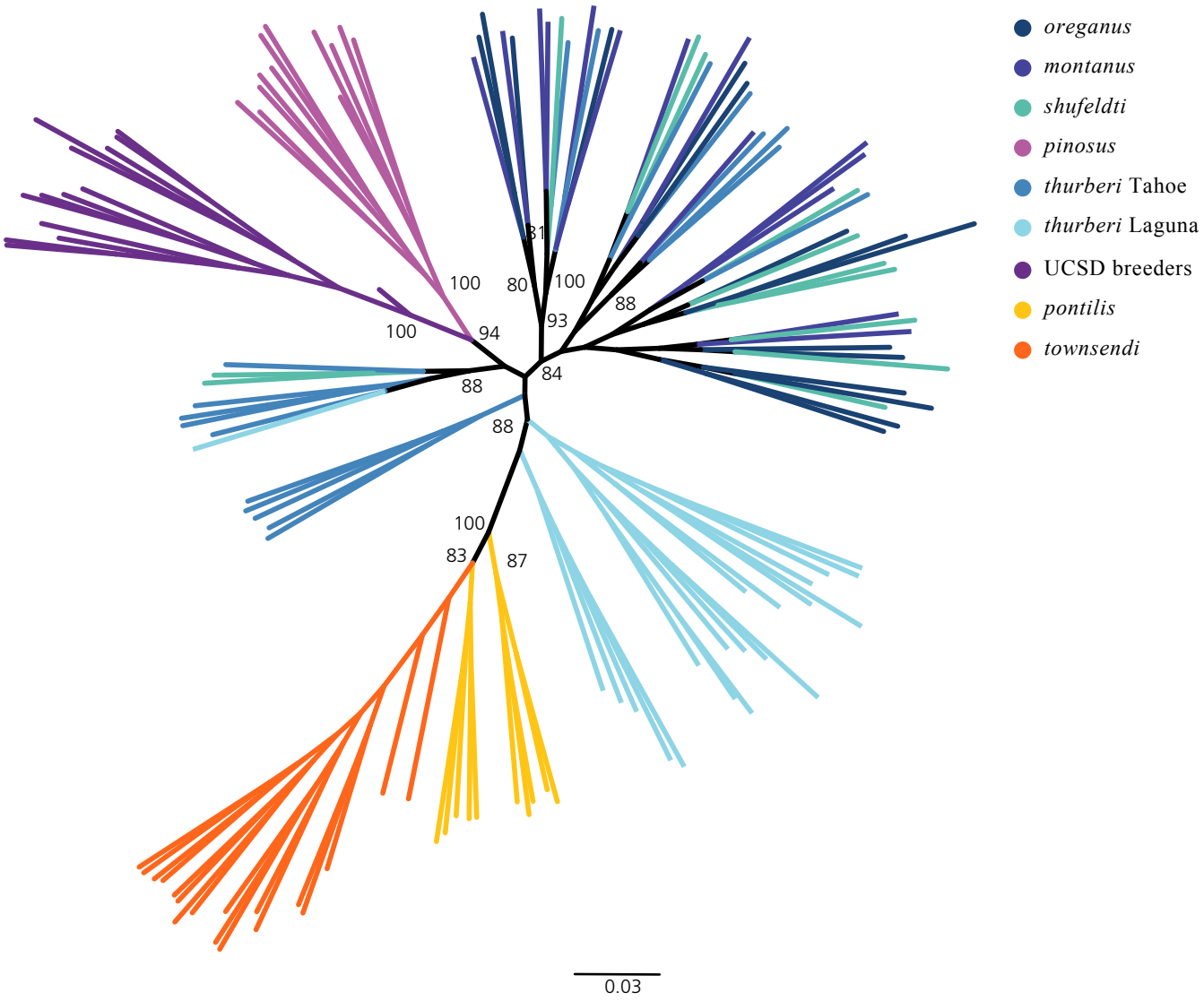

FIGURE 4 Phylogenetic relationships among Oregon junco subspecies. Unrooted maximum-likelihood phylogeny based on 14,175 selectively neutral SNP loci, excluding UCSD individuals not sampled during the breeding season. The scale bar represents 0.03 base substitutions per 1000 nucleotide positions. Bootstrap values ≥80% are shown. Branch colours correspond to those in Figure 1 [Colour figure can be viewed at wileyonlinelibrary.com]

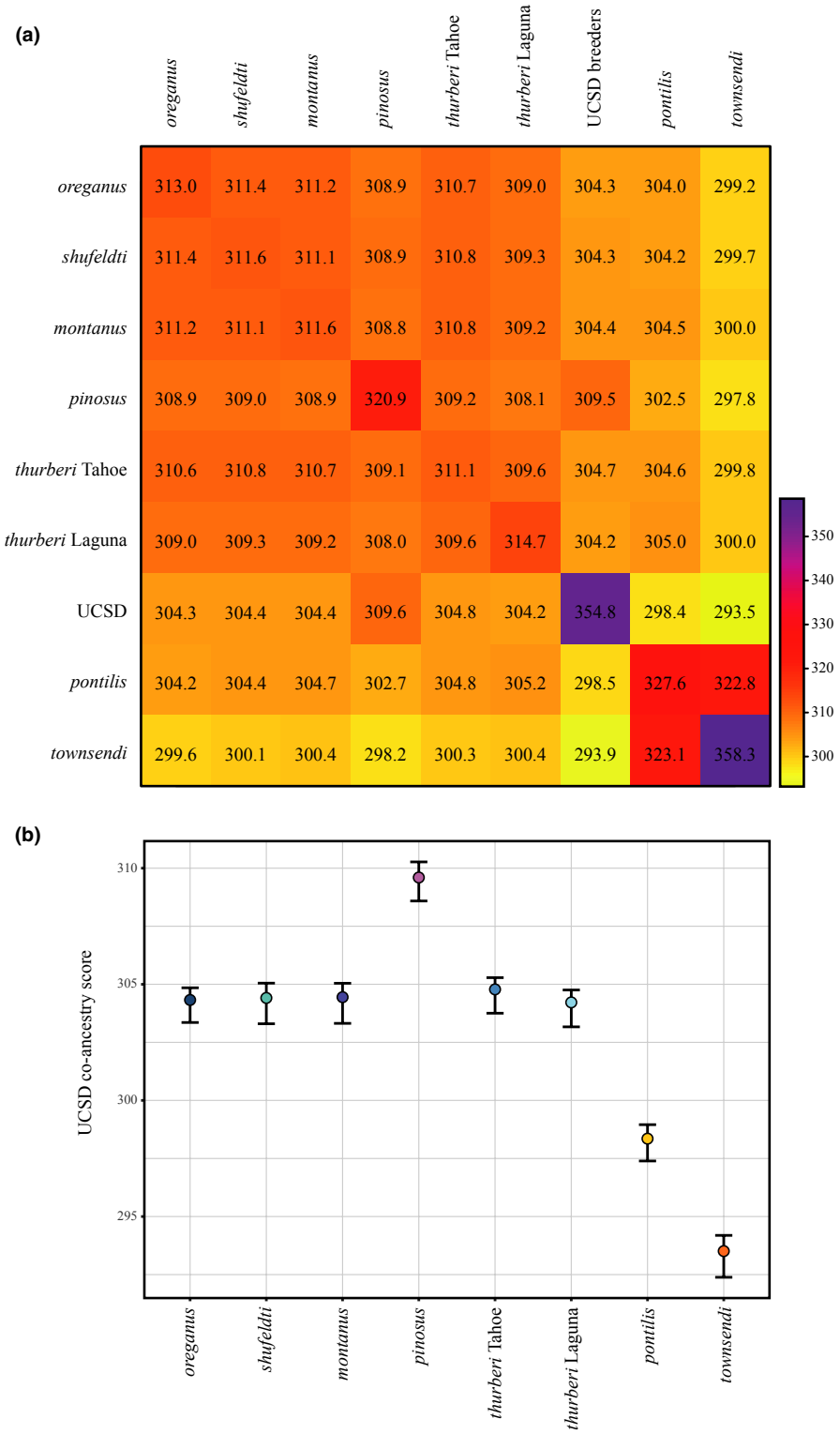

FIGURE 5 Co-ancestry among Oregon junco populations. (a) Individual-based pairwise co-ancestry scores based on 41,231 SNPs, computed with RADPAINTER and averaged over Oregon junco subspecies. (b) Co-ancestry scores for UCSD individuals and each of the other Oregon junco populations with 95% confidence intervals obtained by bootstrap [Colour figure can be viewed at wileyonlinelibrary.com]

individuals showed highest co-ancestry values with respect to *pinosus* individuals (Figure 5a), a pattern that was significant according to a bootstrap analysis (Figure 5b).

3.4 | Demographic inference

Eight different models were compared to identify the most likely scenario under which UCSD juncos diverged from its closest relative, the subspecies *pinosus*, using FASTSIMCOAL2. AIC revealed that the most likely scenario was the one consistent with a recent population split due to a founder event (Figure 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% confidence interval [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

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

Model	Δ Likelihood	AIC	Δ AIC
Model A–IM	67,732.987	939,268.006	0.000
Model B–SI	67,945.960	940,250.783	982.777
Model B–IM	67,956.734	940,308.399	1,040.393
Model A–SI	68,033.065	940,645.916	1,377.910
Model D–SI	68,106.265	940,983.015	1,715.009
Model D–IM	68,107.213	940,991.381	1,723.374
Model C–IM	68,260.534	941,697.450	2,429.444
Model C–IM	68,312.247	941,931.597	2,663.591

Model A parameters	Estimate	CI 95%
N_e UCSD	252	215–274
N_e *pinosus*	5,859,774	5,857,240–5,865,925
N_e colonizers	2	2–3
Time of divergence	44	30–47
MIG *pinosus* → UCSD	6.02×10^{-2}	5.87×10^{-2}–8.03×10^{-2}
MIG UCSD → *pinosus*	2.01×10^{-6}	9.38×10^{-10}–2.57×10^{-5}
UCSD growth rate	0.159	0.150–0.244

Note: Models tested correspond to those in Figure 2. IM and SI refer to "isolation with migration" and "strict isolation," respectively. Δ 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.

(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 = [5.86×10^6–5.87×10^6]). The analysis also recovered a narrow, limited rate of gene migration from *pinosus* to UCSD ($MIG_{pin \rightarrow UCSD} = 6.02 \times 10^{-2}$, 95% CI = [$5.87 \times 10^{-2}$–$8.03 \times 10^{-2}$]). The range of migration rate from UCSD to *pinosus* varied by several orders of magnitude and was close to zero, revealing little impact of the parameter in the SFS ($MIG_{UCSD \rightarrow pin} = 2.01 \times 10^{-6}$, 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 SFS 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 these results should be interpreted with caution.

3.5 | Genotype association analysis

In the RDA, four explanatory environmental variables accounted for 7.65% ($R^2 = .0765$) of the variability in the SNP data set, and the model was highly significant ($p = .001$). 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 (Figure 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 subspecies, mostly differentiated along RDA1. Subspecies *townsendi* and *pontilis* showed negative correlations with vegetation greenness, with the former also loading heavily on elevation. The subspecies *pinosus* presented general associations of the same direction as UCSD, yet much less pronounced. The remaining, more boreal subspecies 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 (Figure 6).

3.6 | 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. 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; Rakvács et al., 2019);

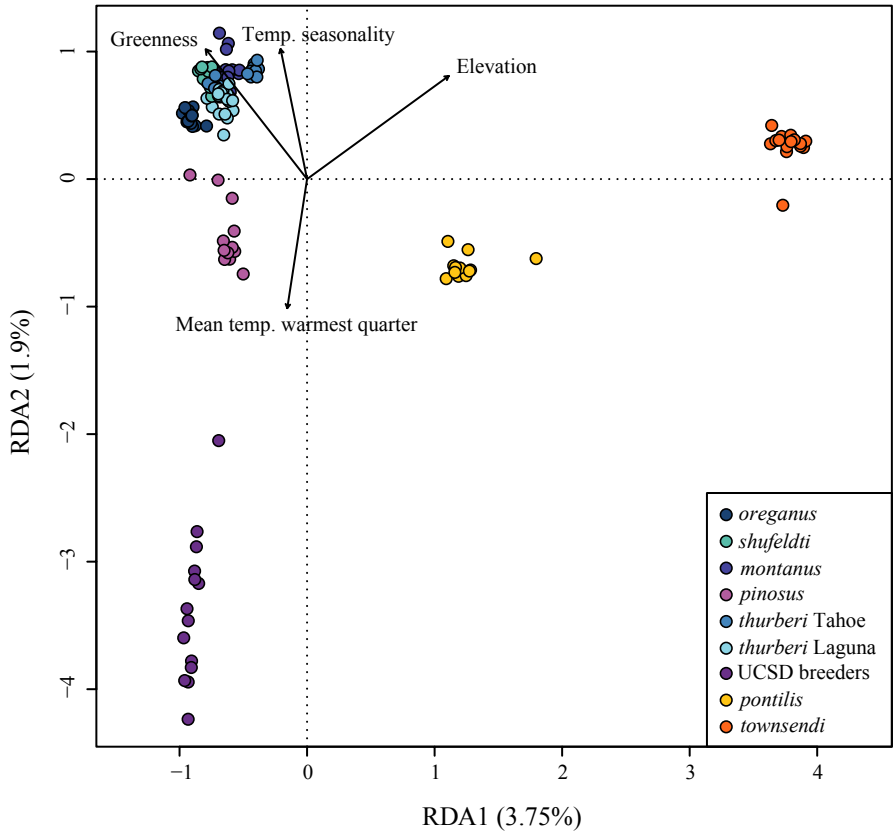

FIGURE 6 Signals of selection in UCSD juncos. Redundancy analysis (RDA) between 132,343 SNP response variables including both neutral and putatively adaptive positions, and four explanatory variables corresponding to environmental factors (mean temperature of the warmest quarter, temperature seasonality, greenness and elevation) for Oregon junco subspecies. Points represent the projection of individual genotypes on the first two RDA axes. Marker colours correspond to those on the range map in Figure 1. The explanatory variables are shown within the space defined by RDA1 and RDA2 by labelled 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. Explained variance is shown in parentheses for each axis [Colour figure can be viewed at wileyonlinelibrary.com]

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 behaviour 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. However, only comparisons involving UCSD juncos and the populations from the south of the distribution of *pontilis* and *townsendi* were significant (*p*-values range = [0–2 – 10^{-4}]), while comparisons among boreal lineages *oreganus*, *shufeldti*, *montanus* and *thurberi* yielded nonsignificant or marginally significant values (*p*-values range = [.02–.95]), congruent with the GEA patterns observed in the RDA. UCSD showed the highest pairwise F_{ST} scores with isolated subspecies *townsendi* (0.54) and *pontilis* (0.28), an average of 0.3 with northern subspecies, 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 (0.26–0.51) in the genes TPOR, ABCB6, FANCM, KCNQ4 and IOV7, with F_{ST} values for the rest of the genes ranging from 0 to 0.2 (Table S2).

4 | DISCUSSION

4.1 | Marked genetic differentiation of the UCSD juncos

Our results confirmed the striking differentiation of the urban population of Oregon juncos inhabiting the UCSD campus. Recovered levels of divergence for the UCSD residents were comparable to those of long-established, geographically isolated subspecies such as *townsendi* and *pontilis* from Baja California sky islands, and in sharp contrast to the low genetic structure found among more northern subspecies. 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 subspecies was detected. Juncos have become established in other suburban and natural areas of southern California in recent years (including the campus at UCLA; Abolins-Abols et al., 2016), 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 subspecies, 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. Even when time of divergence was allowed to vary between 20 and 60 generations, parameter estimates recovered a time of divergence highly consistent with the time at which the juncos were first observed breeding at UCSD in the 1980s. The 95% CI was narrow (20–32 generations), lending support to the recent founder-event hypothesis. The inferred number of founders was as low as two to three individuals, below the seven to 70 individuals previously estimated using microsatellite markers (Rasner et al., 2004). The analysis also favoured 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 4-year period, although immigrants from other populations besides *pinosus* may be included in that estimate (see below). Regardless, 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 δD in feathers and population assignment likelihoods in UCSD individuals sampled during the nonbreeding 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.

4.2 | 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, Grant & Petren, 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 a 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 recovered in our analysis suggests that severe founder effects and genetic drift are involved in the differentiation of the UCSD juncos (Kolbe et al., 2012), a scenario also congruent with the low levels of genetic diversity detected at UCSD.

We also found evidence of local adaptation in promoting evolutionary divergence between UCSD juncos and closely related Oregon junco subspecies. An RDA based on the most dissimilar set of our georeferenced 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. RDA applied to GEA analysis has proven particularly effective in detecting covarying sets of loci under weak and recent/contemporary selection (Forester et al., 2018), yet a number of caveats are worth discussing. First, the ecological parameters used as explanatory variables were distinct for the UCSD campus with respect to native ranges, but may not encompass key aspects of the urban environment (e.g., unsuitable habitat due to human structures, contamination, human-processed foods, mechanical noise or other distressing stimuli) with which they may, in turn, present spatial correlation (Szulkin, Garroway, et al., 2020). These and other covarying parameters not included in our model could be involved in the patterns of adaptive variation described here, yet go undetected by our analysis. Second, much of the recovered correlation signal is likely to be due to population-history-related effects (Capblancq & Forester, 2021; Forester et al., 2016; Frichot et al., 2015; Wang & Bradburd, 2014). In RDA models, covariables can be included to control for the effects of neutral population structure and other confounding factors. However, it has been reported that adding covariables approximating population structure effects often results in a substantial loss of power to detect signals of selection, and can even increase the false positive rate (FPR) in identification of candidate loci. Indeed, in a series of comparative analyses using simulated data in a range of different demographic scenarios, Forester et al. (2018) found that controlling for neutral structure led to a slight increase of the FPR for most of the cases, and in those in which the FPR improved, the true positive rate (TPR) was reduced by up to 92%. We therefore consider that in the case of the UCSD juncos, including covariables accounting for population structure in our analysis would be overly conservative and potentially detrimental, given the high levels of neutral differentiation and the arguably subtle effects of selection acting over a period as limited as 30 generations. Functionally annotated genes linked to the RDA outliers identified as candidate loci also support a role for selective pressures in shaping allele frequencies in the newly colonized habitat.

Among the functions associated with 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 urban 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 signalling, 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 & Wiktelius, 2011). A significantly higher minimum frequency in the vocal signalling of the UCSD juncos with respect to natural habitats has 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.

4.3 | Geographical origin of the UCSD juncos

Phylogenetic and co-ancestry analyses strongly supported the subspecies *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 probable 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 probably 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 subspecies over the last ~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 divergence of allele frequencies established during colonization may be resistant to decay by gene exchange under conditions of rapid population growth following a founder event (Boileau et al., 1992; Lombal et al., 2020). Moreover, moderate admixture can also contribute to colonization success by minimizing deleterious effects of reduced intrapopulation diversity or by introducing favourable alleles and enhancing evolutionary potential (Keller & Taylor, 2010). Evolutionary studies are increasingly recognizing the contributions of introgression to adaptation and gene sorting during 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 overwintering 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 subspecies other than *pinosus*, although its relevance remains to be determined.

The subspecies *pinosus*, to which the UCSD population is most closely related, is a nonmigratory population 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* is 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 the chances of adaptive alleles such as those we have inferred from entering into the UCSD population.

5 | 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 additional fundamental aspects of the evolutionary history of the UCSD population have been established here based on genomic data: (i) the UCSD junco population is most closely related to the coastal subspecies *pinosus*, a group showing ecological and phenotypic similarities, implying it is likely to have been the initial colonist; (ii) environmental factors played a role in driving rapid divergence at functional loci, some of which may be involved in adaptations to an urban 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 signalling traits (Price et al., 2008; Yeh, 2004), and shifts in a range of behavioural and life-history traits (Price et al., 2008; Yeh & Price, 2004) were thought to be due to rapid evolutionary change, as were associated differences in hormones and behaviour (Atwell et al., 2012, 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 need to be qualified 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 previous explanations may have overestimated the role of rapid evolutionary change in plumage variation. Further, previous 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 the selection reported here is associated more strongly with the anthropogenic environment at UCSD rather than a 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 valuable help with field sampling. Thanks to Paul C. Paquet, Chris Darimont and the Raincoast Conservation Foundation for making possible the sampling in British Columbia. We are also grateful to three anonymous reviewers for their comments and suggestions. Funding was provided by grants CGL-2011-25866 and CGL-2015-66381 from Spain's Ministerio de Ciencia e Innovación to B.M., and USA National Science Foundation IOS-1257527 to T.J.G.

CONFLICT OF INTERESTS

The authors have no conflict of interests to declare.

AUTHOR CONTRIBUTIONS

GF and BM designed the study; GF, BM, JWA, AMF, TJM and PJY carried out field sampling; GF and JWA generated the genomic data; GF analyzed the genomic data; AMF and EDK generated the isotopic data; GF and BM wrote the manuscript with input from all co-authors.

DATA AVAILABILITY STATEMENT

Tne Junco hyemalis reference genome is deposited at DDBJ/ENA/GenBank under accession QZWM00000000. The SNP "Full Dataset", isotopic and georeferenced data sets are deposited in Dryad, as well as the model files for the demographic analysis with FASTSIMCOAL2 and R scripts for redundancy analyses (DOI: https://doi.org/10.5061/dryad.gf1vhhmpv).

ORCID

Guillermo Friis https://orcid.org/0000-0002-0731-6468
Trevor D. Price https://orcid.org/0000-0002-4423-8999
Borja Milá https://orcid.org/0000-0002-6446-0079

REFERENCES

Abbott, R., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J. E., Bierne, N., Boughman, J., Brelsford, A., Buerkle, C. A., Buggs, R., Butlin, R. K., Dieckmann, U., Eroukhmanoff, F., Grill, A., Cahan, S. H., Hermansen, J. S., Hewitt, G., Hudson, A. G., Jiggins, C., … Zinner, D. (2013). Hybridization and speciation. *Journal of Evolutionary Biology*, *26*(2), 229–246. https://doi.org/10.1111/j.1420-9101.2012.02599.x

Abolins-Abols, M., Hope, S. F., & Ketterson, E. D. (2016). Effect of acute stressor on reproductive behavior differs between urban and rural birds. *Ecology and Evolution*, *6*(18), 6546–6555.

Akaike, H. (1973). Information theory as an extension of the maximum likelihood principle. In *Paper presented at the Second International Symposium on Information Theory* (pp. 276–281). Akademiai Kiado.

Andrews, S. (2010). *FastQC: A quality control tool for high throughput sequence data*. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Atwell, J. W., Cardoso, G. C., Whittaker, D. J., Campbell-Nelson, S., Robertson, K. W., & Ketterson, E. D. (2012). Boldness behavior and stress physiology in a novel urban environment suggest rapid correlated evolutionary adaptation. *Behavioral Ecology*, *23*(5), 960–969.

Atwell, J. W., Cardoso, G. C., Whittaker, D. J., Price, T. D., & Ketterson, E. D. (2014). Hormonal, behavioral, and life-history traits exhibit correlated shifts in relation to population establishment in a novel environment. *The American Naturalist*, *184*(6), E147–E160.

Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K. V., Altshuler, D., Gabriel, S., & DePristo, M. A. (2013). From FastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. *Current Protocols in Bioinformatics*, *43*, 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*(3), 435–438.

Beichman, A. C., Huerta-Sanchez, E., & Lohmueller, K. E. (2018). Using genomic data to infer historic population dynamics of nonmodel organisms. *Annual Review of Ecology, Evolution, and Systematics*,

49(1), 433–456. https://doi.org/10.1146/annurev-ecolsys-110617-062431

Boileau, M. G., Herbert, P. N. D., & Schwartz, S. S. (1992). Non-equilibrium gene frequency divergence: Persistent founder events in natural populations. *Journal of Evolutionary Biology*, *5*, 25–39.

Boortz, K. A., Syring, K. E., Dai, C., Pound, L. D., Oeser, J. K., Jacobson, D. A., Wang, J.-C., McGuinness, O. P., Powers, A. C., & O'Brien, R. M. (2016). G6PC2 modulates fasting blood glucose in male mice in response to stress. *Endocrinology*, *157*(8), 3002–3008.

Borcard, D., Gillet, F., & Legendre, P. (2011). Chapter 6. Canonical ordination. In R. Gentleman, K. Hornik, & G. G. Parmigiani (Eds.), *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*(2), 255–265. https://doi.org/10.1093/molbev/msh018

Burger, J., & Gochfeld, M. (1988). Effects of lead on growth in young herring gulls (*Larus argentatus*). *Journal of Toxicology and Environmental Health*, *25*(2), 227–236.

Campbell-Staton, S. C., Winchell, K. M., Rochette, N. C., Fredette, J., Maayan, I., Schweizer, R. M., & Catchen, J. (2020). Parallel selection on thermal physiology facilitates repeated adaptation of city lizards to urban heat islands. *Nature Ecology & Evolution*, *4*(4), 652–658.

Capblancq, T., & Forester, B. R. (2021). Redundancy analysis: A Swiss Army Knife for landscape genomics. *Methods in Ecology and Evolution*, *12*(12), 2298–2309. https://doi.org/10.1111/2041-210X.13722

Cardoso, G. C., & Atwell, J. W. (2010). Directional cultural change by modification and replacement of memes. *Evolution*, *65*(1), 295–300.

Cardoso, G. C., & Atwell, J. W. (2011). On the relation between loudness and the increased song frequency of urban birds. *Animal Behaviour*, *82*(4), 831–836.

Carson, H. L. (1975). The genetics of speciation at the diploid level. *The American Naturalist*, *109*(965), 83–92.

Carson, H. L., & Kaneshiro, K. Y. (1976). Drosophila of Hawaii: Systematics and ecological genetics. *Annual Review of Ecology and Systematics*, *7*(1), 311–345.

Chamberlain, C., Blum, J., Holmes, R. T., Feng, X., Sherry, T., & Graves, G. R. (1996). The use of isotope tracers for identifying populations of migratory birds. *Oecologia*, *109*(1), 132–141.

Charlesworth, B. (2009). Effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics*, *10*(3), 195–205.

Chen, Y., Shenkar, N., Ni, P., Lin, Y., Li, S., & Zhan, A. (2018). Rapid microevolution during recent range expansion to harsh environments. *BMC Evolutionary Biology*, *18*(1), 187.

Chhatre, V., & Emerson, K. (2016). *StrAuto: Automation and parallelization of STRUCTURE analysis*. http://strauto.popgen.org

Colautti, R. I., & Lau, J. A. (2015). Contemporary evolution during invasion: Evidence for differentiation, natural selection, and local adaptation. *Molecular Ecology*, *24*(9), 1999–2017.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158.

Danecek, P., & McCarthy, S. A. (2017). BCFtools/csq: Haplotype-aware variant consequences. *Bioinformatics*, *33*(13), 2037–2039.

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., Philippakis, A. A., del Angel, G., Rivas, M. A., Hanna, M., McKenna, A., Fennell, T. J., Kernytsky, A. M., Sivachenko, A. Y., Cibulskis, K., Gabriel, S. B., Altshuler, D., & Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, *43*(5), 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. *Bulletin of the American Museum of Natural History*, *38*(9), 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*(4), 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*(1), 23–34. https://doi.org/10.1080/09670874.2010.499611

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One*, *6*(5), e19379. https://doi.org/10.1371/journal.pone.0019379

Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C., & Foll, M. (2013). Robust demographic inference from genomic and SNP data. *Plos Genetics*, *9*(10), e1003905.

Feng, S., Stiller, J., Deng, Y., Armstrong, J., Fang, Q. I., Reeve, A. H., Xie, D., Chen, G., Guo, C., Faircloth, B. C., Petersen, B., Wang, Z., Zhou, Q. I., Diekhans, M., Chen, W., Andreu-Sánchez, S., Margaryan, A., Howard, J. T., Parent, C., … Zhang, G. (2020). Dense sampling of bird diversity increases power of comparative genomics. *Nature*, *587*(7833), 252–257. https://doi.org/10.1038/s41586-020-2873-9

Forester, B. R., Jones, M. R., Joost, S., Landguth, E. L., & Lasky, J. R. (2016). Detecting spatial genetic signatures of local adaptation in heterogeneous landscapes. *Molecular Ecology*, *25*(1), 104–120.

Forester, B. R., Lasky, J. R., Wagner, H. H., & Urban, D. L. (2018). Comparing methods for detecting multilocus adaptation with multivariate genotype-environment associations. *Molecular Ecology*, *27*(9), 2215–2233. https://doi.org/10.1111/mec.14584

Frichot, E., Schoville, S. D., de Villemereuil, P., Gaggiotti, O. E., & François, O. (2015). Detecting adaptive evolution based on association with ecological gradients: orientation matters! *Heredity*, *115*(1), 22–28.

Friis, G., Aleixandre, P., Rodríguez-Estrella, R., Navarro-Sigüenza, A. G., & Milá, B. (2016). Rapid postglacial diversification and long-term stasis within the songbird genus Junco: Phylogeographic and phylogenomic evidence. *Molecular Ecology*, *25*(24), 6175–6195.

Friis, G., Fandos, G., Zellmer, A. J., McCormack, J. E., Faircloth, B. C., & Milá, B. (2018). Genome-wide signals of drift and local adaptation during rapid lineage divergence in a songbird. *Molecular Ecology*, *27*(24), 5137–5153.

Friis, G., & Milá, B. (2020). Change in sexual signalling traits outruns morphological divergence across an ecological gradient in the postglacial radiation of the songbird genus Junco. *Journal of Evolutionary Biology*, *33*(9), 1276–1293.

Fudickar, A. M., Greives, T. J., Abolins-Abols, M., Atwell, J. W., Meddle, S. L., Friis, G., Stricker, C. A., & Ketterson, E. D. (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*(5), 1174–1179.

Grant, P. R., Grant, B. R., & Petren, K. (2001). A population founded by a single pair of individuals: establishment, expansion, and evolution. *Genetica*, *112–113*, 359–382.

Guo, L., Costanzo-Garvey, D. L., Smith, D. R., Neilsen, B. K., MacDonald, R. G., & Lewis, R. E. (2017). Kinase Suppressor of Ras 2 (KSR2) expression in the brain regulates energy balance and glucose homeostasis. *Molecular Metabolism*, *6*(2), 194–205.

Hendry, A. P., & Kinnison, M. T. (1999). Perspective: the pace of modern life: Measuring rates of contemporary microevolution. *Evolution*, *53*(6), 1637–1653.

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & 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*(2), 88–93. https://doi.org/10.1080/00071668.2018.1564240

Huey, R. B., Gilchrist, G. W., Carlson, M. L., Berrigan, D., & Serra, L. (2000). Rapid evolution of a geographic cline in size in an introduced fly. *Science*, *287*(5451), 308–309.

Jensen, A. J., Hansen, L. P., Johnsen, B. O., & 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*(1), 22.

Johnson, M. T., & Munshi-South, J. (2017). Evolution of life in urban environments. *Science*, *358*(6363), eaam8327.

Kamvar, Z. N., López-Uribe, M. M., Coughlan, S., Grünwald, N. J., Lapp, H., & Manel, S. (2017). Developing educational resources for population genetics in R: An open and collaborative approach. *Molecular Ecology Resources*, *17*(1), 120–128.

Keller, S., & Taylor, D. (2010). Genomic admixture increases fitness during a biological invasion. *Journal of Evolutionary Biology*, *23*(8), 1720–1731.

Kharkovets, T., Hardelin, J.-P., Safieddine, S., Schweizer, M., El-Amraoui, A., Petit, C., & Jentsch, T. J. (2000). KCNQ4, a K+ 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 of the United States of America*, *97*(8), 4333–4338.

Kliber, A., & Eckert, C. G. (2005). Interaction between fournder effect and selection during biological invasion in an aquatic plant. *Evolution*, *59*(9), 1900–1913.

Kolbe, J. J., Leal, M., Schoener, T. W., Spiller, D. A., & Losos, J. B. (2012). Founder effects persist despite adaptive differentiation: A field experiment with lizards. *Science*, *335*(6072), 1086–1089.

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, *15*(5), 1179–1191.

Koskinen, M. T., Haugen, T. O., & Primmer, C. R. (2002). Contemporary fisherian life-history evolution in small salmonid populations. *Nature*, *419*(6909), 826–830.

Krueger, F. (2015). *Trim Galore!: A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files*. https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/

Lamichhaney, S., Berglund, J., Almén, M. S., Maqbool, K., Grabherr, M., Martinez-Barrio, A., Promerová, M., Rubin, C.-J., Wang, C., Zamani, N., Grant, B. R., Grant, P. R., Webster, M. T., & Andersson, L. (2015). Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature*, *518*(7539), 371–375. https://doi.org/10.1038/nature14181

Lande, R. (1980). Genetic variation and phenotypic evolution during allopatric speciation. *The American Naturalist*, *116*(4), 463–479.

Lange, M., Kaynak, B., Forster, U. B., Tönjes, M., Fischer, J. J., Grimm, C., Schlesinger, J., Just, S., Dunkel, I., Krueger, T., Mebus, S., Lehrach, H., Lurz, R., Gobom, J., Rottbauer, W., Abdelilah-Seyfried, S., & Sperling, S. (2008). Regulation of muscle development by DPF3, a novel histone acetylation and methylation reader of the BAF chromatin remodeling complex. *Genes & Development*, *22*(17), 2370–2384.

Legendre, P., & Legendre, L. (1998). Numerical ecology: Second English edition. In *Developments in environmental modelling* (p. 20).

Lerner, H. R. L., Meyer, M., James, H. F., Hofreiter, M., & Fleischer, R. C. (2011). Multilocus resolution of phylogeny and timescale in the extant adaptive radiation of Hawaiian honeycreepers. *Current Biology*, *21*, 1838–1844.

Lescak, E. A., Bassham, S. L., Catchen, J., Gelmond, O., Sherbick, M. L., von Hippel, F. A., & Cresko, W. A. (2015). Evolution of stickleback in 50 years on earthquake-uplifted islands. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(52), E7204–E7212.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760.

Lischer, H. E. L., & Excoffier, L. (2012). PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, *28*(2), 298–299. https://doi.org/10.1093/bioinformatics/btr642

Liu, Z., Li, S., Wang, W., Xu, D., Murphy, R. W., & Shi, P. (2011). Parallel evolution of KCNQ4 in echolocating bats. *PLoS One*, *6*(10), e26618.

Lombal, A. J., O'dwyer, J. E., Friesen, V., Woehler, E. J., & Burridge, C. P. (2020). Identifying mechanisms of genetic differentiation among populations in Vagile species: Historical factors dominate genetic differentiation in seabirds. *Biological Reviews*, *95*(3), 625–651.

Luu, K., Bazin, E., & Blum, M. G. (2017). pcadapt: An R package to perform genome scans for selection based on principal component analysis. *Molecular Ecology Resources*, *17*(1), 67–77.

Mahalanobis, P. C. (1936). On the generalized distance in statistics. *National Institute of Science of India*, *2*, 49–55.

Malinsky, M., Trucchi, E., Lawson, D. J., & Falush, D. (2018). RADpainter and fineRADstructure: Population inference from RADseq data. *Molecular Biology and Evolution*, *35*(5), 1284–1290.

Mathys, B. A., & Lockwood, J. L. (2011). Contemporary morphological diversification of passerine birds introduced to the Hawaiian archipelago. *Proceedings of the Royal Society B: Biological Sciences*, *278*(1716), 2392–2400.

Matschinsky, F. M. (2005). Glucokinase, glucose homeostasis, and diabetes mellitus. *Current Diabetes Reports*, *5*(3), 171–176.

Mayr, E. (1963). *Animal species and evolution*. Belknap Press.

McCaskie, G. (1986). Wood warblers to finches. *American Birds*, *40*, 1257.

McCoy, R. C., Garud, N. R., Kelley, J. L., Boggs, C. L., & Petrov, D. A. (2014). Genomic inference accurately predicts the timing and severity of a recent bottleneck in a nonmodel insect population. *Molecular Ecology*, *23*(1), 136–150.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303.

Milá, B., McCormack, J. E., Castaneda, G., Wayne, R. K., & Smith, T. B. (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*(1626), 2653–2660. https://doi.org/10.1098/rspb.2007.0852

Miller, A. (1941). Speciation in the avian genus *Junco*. *University of California Publications in Zoology*, *44*(3), 173–434.

Millien, V. (2006). Morphological evolution is accelerated among island mammals. *PLOS Biology*, *4*(10), 1863–1868.

Murray, K. D., & Borevitz, J. O. (2017). Axe: Rapid, competitive sequence read demultiplexing using a trie. bioRxiv, 160606.

Nam, K., Mugal, C., Nabholz, B., Schielzeth, H., Wolf, J. B. W., Backström, N., Künstner, A., Balakrishnan, C. N., Heger, A., Ponting, C. P., Clayton, D. F., & Ellegren, H. (2010). Molecular evolution of genes in avian genomes. *Genome Biology*, *11*(6), R68.

Nguyen, L.-T., Schmidt, H. A., Von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, *32*(1), 268–274.

Nolan, V. J., Ketterson, E. D., Cristol, D. A., Rogers, C. M., Clotfelter, E. D., Titus, R. C., & Snajdr, E. (2002). Dark-eyed Junco (*Junco hyemalis*). In A. Poole, & F. Gill (Eds.), *The Birds of North America*, Vol. *716*. The Birds of North America Inc.

O'Brien, R. M. (2013). Moving on from GWAS: functional studies on the G6PC2 gene implicated in the regulation of fasting blood glucose. *Current Diabetes Reports*, *13*(6), 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. https://doi.org/10.1016/j.ygcen.2017.06.021

Pembleton, L. W., Cogan, N. O., & Forster, J. W. (2013). StAMPP: An R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. *Molecular Ecology Resources*, *13*(5), 946–952.

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 P. R. Grant, & B. R. Grant (Eds.), *In search of the causes of evolution: From field observations to mechanisms* (pp. 251–281).

Price, T. D., Yeh, P. J., & 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, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945–959.

R Core Team (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Retrieved from http://www.R-project.org

R Studio Team (2015). *RStudio: Integrated development for R*. R Studio, Inc.

Rakvács, Z., Kucsma, N., Gera, M., Igriczi, B., Kiss, K., Barna, J., Kovács, D., Vellai, T., Bencs, L., Reisecker, J. M., Szoboszlai, N., & Szakács, G. (2019). The human ABCB6 protein is the functional homologue of HMT-1 proteins mediating cadmium detoxification. *Cellular and Molecular Life Sciences*, *76*(20), 4131–4144.

Rasner, C. A., Yeh, P. J., Eggert, L. S., Hunt, K. E., Woodruff, D. S., & Price, T. D. (2004). Genetic and morphological evolution following a founder event in the dark-eyed junco, *Junco hyemalis thurberi*. *Molecular Ecology*, *13*, 671–681.

Reichard, D. G., Atwell, J. W., Pandit, M. M., Cardoso, G. C., Price, T. D., & Ketterson, E. D. (2020). Urban birdsongs: Higher minimum song frequency of an urban colonist persists in a common garden experiment. *Animal Behaviour*, *170*, 33–41.

Reznick, D. N., & Ghalambor, C. K. (2001). The population ecology of contemporary adaptations: what empirical studies reveal about the conditions that promote adaptive evolution. *Genetica*, *112–113*, 183–198.

Rochette, N. C., Rivera-Colón, A. G., & Catchen, J. M. (2019). Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. *Molecular Ecology*, *28*(21), 4737–4754.

Rosado, I. V., Niedzwiedz, W., Alpi, A. F., & Patel, K. J. (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*(13), 4360–4370. https://doi.org/10.1093/nar/gkp365

Rundle, H. D., & Nosil, P. (2005). Ecological speciation. *Ecology Letters*, *8*(3), 336–352. https://doi.org/10.1111/j.1461-0248.2004.00715.x

Salmón, P., Jacobs, A., Ahrén, D., Biard, C., Dingemanse, N. J., Dominoni, D. M., Helm, B., Lundberg, M., Senar, J. C., Sprau, P., Visser, M. E., & Isaksson, C. (2021). Continent-wide genomic signatures of adaptation to urbanisation in a songbird across Europe. *Nature Communications*, *12*(1), 1–14.

Sato, A., Tichy, H., O'hUigin, C., Grant, P. R., Grant, B. R., & Klein, J. (2001). On the origin of Darwin's finches. *Molecular Biology and Evolution*, *18*(3), 299–311. https://doi.org/10.1093/oxfordjournals.molbev.a003806

Schluter, D. (2000). *The ecology of adaptive radiation*. Oxford University Press.

Seehausen, O. (2006). African cichlid fish: A model system in adaptive radiation research. *Proceedings of the Royal Society B*, *273*, 1987–1998.

Sendell-Price, A. T., Ruegg, K. C., & Clegg, S. M. (2020). Rapid morphological divergence following a human-mediated introduction: The role of drift and directional selection. *Heredity*, *124*(4), 535–549.

Seutin, G., White, B. N., & Boag, P. T. (1991). Preservation of avian blood and tissue samples for DNA analyses. *Canadian Journal of Zoology*, *69*(1), 82–90.

Simpson, G. (1953). In L. C. Dunn (Ed.), *The major features of evolution*. Columbia University Press.

Slabbekoorn, H., Yeh, P., & Hunt, K. (2007). Sound transmission and song divergence: A comparison of urban and forest acoustics. *The Condor*, *109*(1), 67–78.

Slatkin, M. (1987). Gene flow and the geographic structure of animal populations. *Science*, *236*, 787–792.

Smeds, L., Qvarnström, A., & Ellegren, H. (2016). Direct estimate of the rate of germline mutation in a bird. *Genome Research*, *26*(9), 1211–1218.

Szulkin, M., Garroway, C. J., Corsini, M., Kotarba, A. Z., Dominoni, D., Munshi-South, J., & Charmantier, A. (2020). How to quantify urbanisation when testing for urban evolution? *Urban Evolutionary Biology*, *13*. https://doi.org/10.1093/oso/9780198836841.001.0001

Szulkin, M., Munshi-South, J., & Charmantier, A. (2020). *Urban evolutionary biology*. Oxford University Press.

Takekawa, J. Y., 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*(1), 60–70.

Tatusova, T. A., & Madden, T. L. (1999). BLAST 2 Sequences, a new tool for comparing protein and nucleotide sequences. *FEMS Microbiology Letters*, *174*(2), 247–250.

Templeton, A. R. (1980). The theory of speciation via the founder principle. *Genetics*, *94*(4), 1011–1038.

Unitt, P. (1984). *The birds of San Diego county*, Vol. *13*. San Diego Society of Natural History.

Van Den Wollenberg, A. L. (1977). Redundancy analysis an alternative for canonical correlation analysis. *Psychometrika*, *42*(2), 207–219.

Wang, I. J., & Bradburd, G. S. (2014). Isolation by environment. *Molecular Ecology*, *23*(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*(3), 211–217.

Wegener, J. E., Pita-Aquino, J. N., Atutubo, J., Moreno, A., & Kolbe, J. J. (2019). Hybridization and rapid differentiation after secondary contact between the native green anole (*Anolis carolinensis*) and the introduced green anole (*Anolis porcatus*). *Ecology and Evolution*, *9*(7), 4138–4148.

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, *38*, 1358–1370.

Wessel, A., Hoch, H., Asche, M., von Rintelen, T., Stelbrink, B., Heck, V., Stone, F. D., & Howarth, F. G. (2013). Founder effects initiated rapid species radiation in Hawaiian cave planthoppers. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(23), 9391–9396.

Whittaker, D. J., Dapper, A. L., Peterson, M. P., Atwell, J. W., & Ketterson, E. D. (2012). Maintenance of MHC Class IIB diversity in a recently established songbird population. *Journal of Avian Biology*, *43*(2), 109–118.

Wright, S. (1931). Evolution in Mendelian populations. *Genetics*, *16*(2), 97.

Wu, Y., Liang, H., Zhang, H., Pi, J. S., Pan, A. L., Shen, J., Pu, Y. J., & Du, J. P. (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*(2), 281–286. https://doi.org/10.1590/1806-9061-2017-0665

Wunder, M. B., Jehl, J. R. Jr, & Stricker, C. A. (2012). The early bird gets the shrimp: Confronting assumptions of isotopic equilibrium and homogeneity in a wild bird population. *Journal of Animal Ecology*, *81*(6), 1223–1232.

Yeh, P. J. (2004). Rapid evolution of a sexually selected trait following population establishment in a novel habitat. *Evolution*, *58*(1), 166–174.
Yeh, P. J., & Price, T. D. (2004). Adaptive phenotypic plasticity and the successful colonization of a novel environment. *The American Naturalist*, *164*(4), 531–542.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Friis, G., Atwell, J. W., Fudickar, A. M., Greives, T. J., Yeh, P. J., Price, T. D., Ketterson, E. D., & Milá, B. (2022). Rapid evolutionary divergence of a songbird population following recent colonization of an urban area. *Molecular Ecology*, *31*, 2625–2643. https://doi.org/10.1111/mec.16422