Genome-wide signals of drift and local adaptation during rapid lineage divergence in a songbird

Running title: Drift and selection in a songbird radiation

Guillermo Friis,¹ Guillermo Fandos,² Amanda Zellmer,³ John McCormack,^{3,4} Brant Faircloth,⁵ Borja Milá¹

¹National Museum of Natural Sciences, Spanish National Research Council (CSIC), Madrid 28006, Spain; ²Department of Biodiversity, Ecology and Evolution, Complutense University of Madrid, Madrid 28040, Spain; ³Department of Biology, Occidental College, Los Angeles, CA 90041, USA; ⁴Moore Laboratory of Zoology and Department of Biology, Occidental College, Los Angeles, CA 90041, USA; ⁵Department of Biological Sciences and Museum of Natural Science, Louisiana State University, Baton Rouge, LA 70803, USA

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

Abstract

The formation of independent evolutionary lineages involves neutral and selective factors, and understanding their relative roles in population divergence is a fundamental goal of speciation research. Correlations between allele frequencies and environmental variability can reveal the role of selection, yet the relative contribution of drift can be difficult to establish. Recently diversified systems such as that of the Oregon junco (Aves: Emberizidae) of western North America provide ideal scenarios to apply genetic-environment association analyses (GEA) while controlling for population structure. Genome-wide SNP loci analyses revealed marked genetic structure consisting of differentiated populations in isolated, dry southern mountain ranges, and more admixed recently expanded populations in humid northern latitudes. We used correlations between genomic and environmental variance to test for three specific modes of evolutionary divergence: (i) drift in geographic isolation, (ii) differentiation along continuous selective gradients, and (iii) isolation by adaptation. We found evidence of strong drift in southern mountains, but also signals of local adaptation in several populations, driven by temperature, precipitation, elevation and vegetation, especially when controlling for population history. We identified numerous variants under selection scattered across the genome, suggesting that local adaptation can promote rapid differentiation over short periods when acting over multiple independent loci.

Key words: local adaptation, isolation by adaptation, drift, selective gradients, redundancy analysis, postglacial expansion

Introduction

Lineage diversification involves both selective and neutral factors, and elucidating their relative strengths and interactions in the process of evolutionary divergence is essential to understand the mechanisms underlying the early stages of speciation (Coyne and Orr 2004; Nosil 2012). Lineage differentiation can be driven by divergent natural selection (Darwin 1859; Coyne and Orr 2004), a mechanism that is at the core of ‘ecological speciation’ models, in which reproductive isolation arises as a by-product of cumulative, ecologically adaptive changes (Mayr 1947; Schluter 2000; Rundle and Nosil 2005). In turn, accumulation of genetic differences caused by drift in geographic isolation or in isolation-by-distance (IBD, Wright 1943; Wright 1946) has been proposed as a mode of divergence driven by neutral factors (Mayr 1954, 1963), a mechanism that can be particularly strong in populations of small effective size (e.g. Carson 1975; Templeton 1981; Uyeda et al. 2009). Selection and drift can also act jointly and even interact during evolutionary divergence in a number of ways. Geographic distance usually implies environmental differences that may drive adaptation to local conditions and ecological differentiation, even if populations are connected by moderate gene flow (Schluter 2000; Rundle and Nosil 2005). The evolution of local adaptation can in turn result in isolation-by-adaptation (IBA), a mode of divergence where adaptive changes lead to intrinsic barriers to gene flow, enabling genome-wide differentiation at both neutral and selected loci (Nosil et al. 2008; Funk et al. 2011). Consequently, geographic distance and ecological divergence may promote similar patterns of genetic diversity among populations, so that teasing apart the roles of neutral evolution and ecological adaptation in evolutionary diversification requires approaches that account for both environmental heterogeneity and neutral population

structure (Wang and Bradburd 2014; Frichot et al. 2015; Rellstab et al. 2015; Forester et al. 2016).

Our capacity to assess the relative roles of adaptation and neutral differentiation in driving population divergence has benefitted from our increased ability to survey genome-wide variation thanks to the development of next generation sequencing (NGS) techniques (McCormack et al. 2013; Faria et al. 2014). The increasingly large number of loci afforded by NGS provides improved resolution to detect neutral population structure and patterns of gene flow among differentiated lineages. In addition, highly differentiated loci identified as outliers in an F_{ST} distribution can be interpreted as potential targets of divergent selection standing out in a background of balanced or neutrally maintained genomic variation (Faria et al. 2014; Rellstab et al. 2015).

However, methods of outlier detection relying solely on allele frequencies are sensitive to the confounding effects of historical factors, such as past sudden changes in population size or strong drift in small populations that may result in high rates of false positives (Edmonds et al. 2004; Kawecki and Ebert 2004; Billiard et al. 2005; Christmas et al. 2016). Moreover, changes in allele frequencies due to local adaptation can sometimes go undetected by outlier analyses (Pritchard and Di Rienzo 2010; Bierne et al. 2011; Rellstab et al. 2015; Forester et al. 2017). Alternative approaches that integrate environmental parameters by identifying allele frequencies that correlate with ecological variability have proven useful to detect signals of adaptation, especially when selective forces are weak (Frichot et al. 2015; Rellstab et al. 2015; Forester et al. 2017). These methods, known as genetic-environment association (GEA, Hedrick et al. 1976; Mitton et al. 1977) analyses, have the potential to reveal genetic patterns of differentiation due to local adaptation while testing for the role of multiple, specific

environmental variables as drivers of selection (Forester et al. 2017). Importantly, GEA methods can correct for population history by controlling for general patterns of neutral genomic variation (Rellstab et al. 2015; Forester et al. 2016), allowing us to separate the respective effects of drift and selection in generating and maintaining variability. GEA analyses have greatly benefitted from the development of high-throughput sequencing techniques, resulting in a number of studies focusing on the genomic variability associated with environmental parameters in groups as diverse as plants (Lasky et al. 2012; Jones et al. 2013; De Kort et al. 2014; Nadeau et al. 2016; Sork et al. 2016), fungus (Ojeda Alayon et al. 2017), wolves (Forester et al. 2017) and birds (Manthey and Moyle 2015; Safran et al. 2016; Szulkin et al. 2016; Termignoni-García et al. 2017).

Recently diversified systems provide an ideal scenario for studying the relative roles of selective and neutral factors in incipient divergence and speciation. Specifically, GEA methods are particularly suitable when the system under study (i) is composed of closely related populations, among which the signals of selection are still recent and detectable; (ii) includes broad geographic distributions encompassing heterogeneous habitats across ecological clines (*i.e.* selective gradients) but also spatially discontinuous habitats so that adaptive and neutral divergence can be assessed in different spatial settings; (iii) shows large variability in the degree of geographical isolation among populations, from extensive gene flow to total isolation; and (iv) presents low variability in secondary sexual traits so that differential sexual selection can be ruled out as a major driver of population divergence.

The Oregon junco complex (*Junco hyemalis oregonus*) of western North America provides a particularly well suited system to carry out genome-environment association

analysis. The complex originated recently as part of the postglacial radiation of dark-eyed juncos across North America following a northward recolonization of the continent as ice sheets retreated after the last glacial maximum, c.a. 18,000 years ago (Milá et al. 2007; Friis et al. 2016; Milá et al. 2016). Among dark-eyed junco forms, the Oregon junco group presents the highest variability in terms of phenotype and ecological range, encompassing a broad latitudinal range from Baja California to Alaska (Fig. 1A). All forms of the Oregon junco share a characteristic dark hood, yet there is considerable population variation in plumage color, mainly of the hood, dorsum and flanks, and the complex has been traditionally divided into at least 7 subspecific forms (Dwight 1918; Miller 1941; Nolan et al. 2002), which include, from south to north: *townsendi*, from the San Pedro Mártir mountains in northern Baja California, Mexico; *pontilis*, distributed just north of *townsendi* in the Sierra Juárez mountains, also in Baja California, Mexico; *thurberi*, from the mountains of southern California and Sierra Nevada; *pinosus*, a coastal form from central California, predominantly distributed in the Santa Cruz mountains; *montanus*, distributed across the interior of Oregon, Washington and British Columbia; *shufeldti*, a more coastal form from Oregon and Washington; and *oreganus* from coastal British Columbia and southern Alaska (Miller 1941; Nolan et al. 2002; Fig. 1A, Fig. 2A, Table 1).

The diverse spatial configuration of populations and environmental variability across the Oregon junco distribution are critical aspects that will affect our capacity to disentangle the roles of adaptive and neutral factors in explaining genomic variance. Here we use a conceptual framework to classify into three main settings the distinct spatial scenarios observable in the Oregon junco with respect to gene flow and environmental variation. These include (i) geographically isolated populations in similar

habitats, as in the case of the Baja California *townsendi* and *pontilis* forms, where low levels of local adaptation and low rates of gene flow should result in limited adaptive divergence and high neutral divergence by drift (Fig. 2A); (ii) parapatric populations under divergent ecological conditions, as exemplified by the *pinosus* and *thurberi* forms in California, where divergence is expected to increase due to local adaptation, while geographic proximity and moderate gene flow should lead to intermediate levels of neutral differentiation by drift (Fig. 2B); and (iii) populations found along a continuous environmental gradient, as in the northernmost forms of Oregon junco (*thurberi*, *shufeldti*, *montanus* and *oreganus*) where neutral divergence is expected to be low due to high levels of gene flow, while local adaptation along the gradient may result in a pattern of high differentiation in adaptive variation (Fig. 2C).

Here we use the Oregon junco complex to study how geographic isolation, population history, and local ecological adaptation have driven population differentiation across the range, using extensive population sampling and genome-wide SNPs obtained from ‘genotyping by sequencing’ (GBS, Elshire et al. 2011). A draft consensus genome of junco has also been sequenced and assembled to be used as a reference. We first assess patterns of neutral genetic structure across the complex using selectively neutral SNPs, and we then look for correlations between environmental variables and allele frequencies across the Oregon junco distribution using redundancy analysis (RDA). We also use climatic variables to test for significant niche divergence while controlling for spatial autocorrelation. Finally, we map GBS sequences harboring significant outlier SNP loci to the zebra finch (*Taeniopygia guttata*) reference genome in order to recover the chromosomal position of polymorphic sites and explore how adaptive variation is distributed across the genome.

Materials and methods

Population sampling

Oregon junco populations were sampled across the geographical range of the species using mist nets in order to obtain biological samples for DNA extraction (Fig. 1A). Each captured individual was aged, sexed, and marked with a numbered aluminum band. A blood sample was collected by venipuncture of the sub-brachial vein and stored in Queen's lysis buffer (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, and with state and federal scientific collecting permits in the USA and Mexico. A high-quality tissue sample for whole-genome sequencing was obtained from a red-backed junco (*J. hyemalis dorsalis*) specimen (MLZ 69090) provided by the Moore Laboratory of Zoology, Occidental College, that was collected at Dude Mountain, Coconino County, Arizona. Genomic DNA was extracted from blood and tissue samples using a Qiagen DNeasy kit (Qiagen™, Valencia, CA).

Whole-genome sequencing and assembly

We assembled a draft dark-eyed junco genome by combining low-coverage genomes of eight different junco individuals, which we then used as a conspecific reference in the SNP calling process. Libraries for seven of the genomes were prepared with the Kapa Library Preparation Kit (Kapa Biosystems, Inc.) using TruSeq-style adapters (Faircloth and Glenn 2012). They were pooled after random shearing and individual barcoding and sequenced in a single lane of an Illumina HiSeq platform. The eighth genome was

sequenced at a higher coverage by means of two 101-bp paired-end shotgun libraries and two 101-bp mate-paired libraries with insert sizes of 8 Kb in length at Macrogen Inc. The TruSeq Nano DNA Kit (Illumina) was used for the preparation of the shotgun libraries, while the mate-paired libraries were prepared with Nextera Mate Pair Kit (Illumina). We used FASTQC (Andrews 2010) to evaluate the quality of the sequenced data, and quality filtering was carried out with NextClip (Leggett et al. 2013) in the case of the mate paired libraries. For the rest of them we used Trimmomatic (Bolger et al. 2014), applying a sliding window filtering approach with a size of 4 bp and a phred quality score threshold of 25. We also set a minimum length of 50 bp, below which reads were filtered out after trimming. We used the software SOAPDENOV02 (Luo et al. 2012) to perform the assembly. The average insert size for each library was estimated in a preliminary run, and we set a Kmer size of 27 and minimum edge coverage of 2. Gaps that emerged during the scaffolding process were removed with the GapCloser tool from SOAPDENOV02. Finally, we filtered out all the scaffolds shorter than 500 bp so the genome was functional as a mapping reference. The final assembled genome had 37,904 scaffolds with an N50 of 147,816 bases, a L50 of 1,951 scaffolds, a total size of 1.09 Gb, 17.5 Mb of missing sites, and an overall coverage of ~56X, as computed with VCFTOOLS version 0.1.13 (Danecek et al. 2011).

Genotyping-by-sequencing

We used genotyping-by-sequencing (Elshire *et al.* 2011) to obtain individual genotypes from 127 Oregon juncos belonging to the following subspecific taxa: *townsendi* (n=16), *pontilis* (n=16), *thurberi* (n=35), *pinosus* (n=16), *montanus* (n=16), *shufeldti* (n=12), *oreganus* (n=16) (Table 1). GBS libraries were prepared and sequenced at Cornell University's Institute for Genomic Diversity, using the restriction enzyme PstI for

digestion. Sequencing of the 127 individually-barcoded libraries was carried out in three different lanes (along with other 149 junco samples intended for other studies) of an Illumina HiSeq 2000, resulting in an average of 229.2 million good single-end reads 100 bp in length per lane. Samples of the same form were distributed among at least two of the lanes to avoid sequencing bias, except in the case of *shufeldti* individuals, which were all introduced in the last set of sequenced samples.

Alignment and variant calling

We evaluated GBS read quality using FASTQC after sorting them by individual with FASTXTOOLKIT (Gordon and Hannon 2010), and performed the trimming and filtering treatment using PRINSEQ (Schmieder and Edwards 2011). All resulting reads were 69 bp long and had a mean genotyping phred quality score of at least 30, with no positions below 20. The reads were then mapped against the assembled junco genome using the mem algorithm in the Burrows-Wheeler Aligner (BWA, Li and Durbin 2009). Two samples were excluded at this step due to too low read mapping (1 *thurberi* and 1 *montanus*). We used the Genome Analysis Toolkit (GATK, McKenna et al. 2010) version 3.6-0 to realign reads around indels using the IndelRealigner tool and then we applied the HaplotypeCaller tool to call the individual genotypes. We finally used the GenotypeGVCFs tool to gather all the per-sample GVCFs files generated in the previous step and produce a set of joint-called SNPs and indels in the variant call format (vcf) (GATK Best Practices, DePristo et al. 2011; Auwera et al. 2013). Because GBS data do not provide enough coverage for base quality score recalibration, we used VCFTOOLS to implement a ‘hard filtering’ process, customized for each of the downstream analyses (Table 2).

Genetic structure analyses

To explore genome-wide population structure among all Oregon junco forms, we ran a principal components analysis (PCA) based on SNP data. Using VCFTOOLS we retained all samples with less than 25% missing data after a ‘soft filtering’ (coverage range between 2 and 100, minimum phred quality score of 40), resulting in a dataset of 88 samples, and between 8 and 12 individuals per population (24 in the case of *thurberi*). We filtered out all the sites with any non-genotyped individuals or a minor allele frequency (MAF) below 0.02. We also applied a threshold for SNPs showing highly significant deviations from Hardy-Weinberg equilibrium (HWE) with a p-value of 10^{-4} to filter out false variants arisen by the alignment of paralogous loci, resulting in a matrix of 11,261 variants. We then excluded SNPs putatively under selection using BayeScan (Foll and Gaggiotti 2008). BayeScan computes per-SNP F_{ST} scores and decomposes them into a population-specific component shared by all loci that approximates population related effects, and a locus-specific component shared by all populations, which accounts for selection. The program compares two models of divergence, with and without selection, and assumes a departure from neutrality when the locus-specific component is necessary to explain a given diversity pattern (Foll 2012). We used BayeScan with default settings and a thinning interval size to 100 to ensure convergence. For each SNP we obtained the posterior probability for the selection model and the F_{ST} coefficient averaged over populations. For outlier detection and exclusion, we implemented a false discovery rate of 0.3. To filter out the SNPs under linkage disequilibrium (LD) we used the function `snpgdsLDpruning` from the `SNPrelate` package (Zheng 2012) in R Studio (R Studio Team 2015) version 1.0.136 with R (R Core Team 2013) version 3.2.2. We applied the correlation coefficient method with a threshold of 0.2 (`method = "corr", ld.threshold = 0.2`), resulting in a final

data matrix of 9,436 SNP loci (Table 2). We then used the function `snpGdsPCA` also available in `SNPrelate` to perform the PCA and obtain the eigenvectors to be plotted.

We examined population structure with `STRUCTURE` (Pritchard et al. 2000), using a smaller, more heavily filtered SNP data matrix to reduce the computational time of the analysis. Using `VCFTOOLS`, we retained the eight samples of each population (16 in the case of *thurberi*) with the lowest proportion of missing sites for a final number of 64 samples. We constructed a data matrix of biallelic SNPs excluding those out of a range of coverage between 2 and 100, or with a genotyping phred quality score below 70. Positions with less than 90% of individuals genotyped were removed from the data matrix, along with those presenting a MAF below 0.02. Once again, we implemented a threshold for SNPs showing highly significant deviations from Hardy-Weinberg equilibrium (HWE) with a p-value of 10^{-4} , and performed the filtering for non-neutral positions and linkage disequilibrium exactly as done for the PCA, to obtain a final data matrix of 34,367 SNP loci (Table 2). We converted the *vcf* file to `STRUCTURE` format using `PGDspider` (Lischer and Excoffier 2012) version 2.0.5.1. Bash scripts to perform the analyses were created with `STRAUTO` (Chhatre and Emerson 2016) and we ran the program five times per K, with K ranging from 1 to 10 after running a preliminary analysis to infer the lambda value. The burn-in was set to 50K iterations and the analysis ran for an additional 100K iterations. Similarity scores among runs and graphics were computed with `CLUMPAK` (Kopelman et al. 2015).

Redundancy analysis and variance partition

When applying GEA methods there are two main potentially confounding effects related to neutral factors: (i) structure among populations derived from strong drift in

isolation may result in genetic patterns similar to those related to adaptive divergence; and (ii) demographic expansions along latitudinal axes may create gradients of allele frequencies at neutral loci correlated with latitude, that in turn would correlate with any environmental variable that changes with latitude, mimicking a pattern of selective sweep and local adaptation (Excoffier and Ray 2008; Excoffier et al. 2009; Rellstab et al. 2015; Forester et al. 2016). Redundancy analysis (Van Den Wollenberg 1977; Legendre and Legendre 1998) is a canonical ordination method that allows computing the variance of a set of response variables explained by a number of constraining or explanatory variables. In addition, partial RDA enables computation of this shared variance between two sets of variables while conditioning or holding constant the effects of a third set of covariables. Here we used RDA and partial RDA as implemented in the R package *vegan* (Oksanen et al. 2016) to explore the associations between genetic variability and environmental data. Ecological data were obtained from 7 of the 19 variables available in the BioClim database (Hijmans et al. 2005), specifically chosen in accordance to their relevance to junco ecology (Miller 1941; Nolan et al. 2002). They measured mean temperature and precipitation over the year (BIO1 and BIO12); mean temperature and precipitation over the warmest quarter (BIO10 and BIO18), which corresponds to the birds' breeding season; isothermality, referring to how the range of day-to-night temperature differs from the range of summer-to-winter, where a value of 100 indicates equality between them; and seasonality of temperature and precipitation (BIO4 and BIO15). We also included three vegetation variables from the Moderate Resolution Imaging Spectroradiometer (MODIS) satellites as available in <https://modis.gsfc.nasa.gov>: percent tree cover (TREE), Normalized Difference Vegetation Index (NDVI, a measure of canopy greenness), and NDVI's annual standard deviation (std_NDVI). Finally, we included

the high-quality elevation data provided by the NASA Shuttle Radar Topographic Mission (ELEV), downloadable from <http://www2.jpl.nasa.gov/srtm> for a total of eleven variables (Table 3). All ecological variables were centered and standardized. Following Blanchet et al. (2008), we implemented a forward selection method using the forward.sel function from the R-package packfor (Dray et al. 2009) to reduce the number of variables in the model. This procedure applies two stopping criteria: a significance level for each tested variable, which we set at 0.01; and a maximum limit for global adjusted R^2 , equal to the adjusted R^2 of the RDA model including all initial variables. In doing so we prevent inflation of the overall type I error and of the amount of explained variance. After this, we excluded those retained variables with a variance inflation factor (VIF) over 10 (Borcard et al. 2011) to avoid high collinearity. Despite signs of low orthogonality observable among variables in the partial RDA (especially among BIO18, TREE and NDVI, see Results) we chose not to exclude more variables or to apply dimension-reduction treatments like PCA to the environmental space of variables so as to assess their specific and relative contributions to differentiation patterns (McCormack et al. 2010) and discuss signals of adaptation with higher confidence. The final selected ecological variables were used as explanatory variables in two RDAs, with and without subtracting the effects of neutral processes, which were approximated by the first two principal components of the PCA of population structure based on selectively neutral loci (see above). As response variables, we used the same SNP dataset used for the PCA, but excluding LD and neutrality filters. SNP data were coded as counts of the alternative allele for each position (*i.e.*, 0, 1 or 2 copies) with VCFTOOLS and transformed following Patterson et al. (2006). Statistical significance was obtained using a permutation-based procedure with 10,000 permutations, assuming $\alpha = 0.01$. We also used variance partitioning as implemented in the vegan R-package to

estimate (i) the total proportion of genomic variation explained by ecological variables alone; (ii) by neutral structure alone; and (iii) the effects of both sets of variables.

Finally, we repeated the whole RDA treatment for a subset of 87 SNPs identified as selectively divergent by BayeScan, with no conditional treatment. The analyses were conducted in R version 3.2.2 (code included in Appendix I, Supplementary Material).

Niche divergence tests

To further explore patterns of ecological divergence in the Oregon Junco, we tested for niche divergence applying the method developed by McCormack et al. (2010), a method that allows us to examine each environmental variable separately. To avoid a loss of statistical power due to multiple analyses, we conducted three specific comparisons of forms presenting different patterns of genetic divergence and geographical settings, including: (i) *townsendi* and southern *thurberi* forms, in order to estimate niche divergence between geographically isolated, genetically differentiated forms; (ii) the ecologically divergent *pinosus* with the parapatric northern *thurberi* form, to further test a possible case of isolation-by-adaptation; and (iii) northern and southern populations of *thurberi*, as conspecific extremes of a potential adaptive gradient. We used occurrence points from our own georeferenced field sampling records, and this set of occurrence records was further revised to avoid spatial autocorrelation and to match the spatial resolution of environmental variables (1-km grid). Our final dataset comprised 80 localities: *pinosus* (n = 14), *thurberi* north (n = 26), *thurberi* south (n = 19), and *townsendi* (n = 21). We decided to improve quality (geographic accuracy) vs. quantity (number of occurrence records), by using fewer data but with higher spatial accuracy (Engler et al. 2004). To generate a background dataset for each population, we drew 1000 random points from a background representing the geographic range of each junco

population. In order to select an appropriately sized area for the niche divergence tests, we included accessible habitats according to the dispersal ability of each population (Soberon and Peterson 2005). We generated background samples from a 100-km “buffer zone” around known occurrences (Warren et al. 2008). For populations with small ranges or small dispersal ability (*thurberi* south, *pontilis* and *townsendi*) we restricted the buffer zone to 10 km to reduce spatial inaccuracies in the null distribution (Barve et al. 2011), after testing different buffer sizes to test the robustness in delimiting accessible areas for juncos. Next, we extracted the environmental data (same as the data used for the RDA, see above) for both occurrence points and random background points from within the geographic range of each junco form. Niche divergence and conservatism were tested by comparing the observed environmental differences among forms against a null model of background divergence (generated by calculating the difference between background points using a bootstrapping approach and 1000 resamples) for each environmental variable using a two-tailed test. We conducted all the analyses in R 3.2.2.

Genome scans

We performed genome scans for different Oregon junco forms using BayeScan. In order to obtain the chromosomal positions of the SNPs, we mapped the GBS reads against the zebra finch (*Taeniopygia guttata*) genome v87 available in Ensembl (Yates et al. 2016), applying the same set of tools and parameters as for mapping against the junco genome. Using the same set of samples as in the PCA and the RDA, we conducted the analysis for all forms together; for *townsendi* against *pontilis*; and for *townsendi* against all *thurberi* (see Table 2 for final dataset sizes). For each of these matrices, we retained only biallelic SNPs with coverage between 2 and 100 and a

genotyping phred quality score over 40. Positions with less than 25% of the individuals genotyped were removed from each data matrix, along with those presenting a MAF below 0.05. Once again, we implemented a p-value threshold for HWE of 10^{-4} to filter out false variants arisen by the alignment of paralogous loci. We ran BayeScan with the same settings used for filtering out SNPs under selection in population structure analyses, but implemented a more conservative 10% FDR for outlier detection. Genome scan plots were conducted in R 3.2.2 using the package qqman (Turner 2014).

Results

Neutral genetic structure

The plot of the first two principal components from the PCA revealed four distinct clusters in the Oregon junco group. The most differentiated groups were *townsendi* and *pontilis* from Baja California, which formed two highly divergent clusters with respect to each other and to other populations. A third, highly differentiated group corresponded to *pinosus* from coastal California, showing less differentiation than the Baja California forms with respect to a fourth cluster, which included all the remaining forms in the PCA (Fig. 1B). Within this fourth cluster, southern *thurberi* individuals presented certain degree of differentiation from the rest of the forms, a pattern more conspicuous when plotting the third and fourth components, which also revealed a slight signal of divergence in the *oreganus* form (Fig. S1 in Supplementary Information).

The STRUCTURE results were generally congruent with the PCA. The $K = 2$ plot recovered *townsendi* as an independent population, which shared a considerable amount

of variance with *pontilis*. In the analysis for $K = 3$, *pontilis* separated, and $K = 4$ identified the same four main clusters seen in the PC1 vs. PC2 plot (Fig. 1B), revealing *pinosus* as an independent genetic cluster. The plot for $K = 5$ clearly captured the differentiation of the southern *thurberi* form in a fifth cluster, and in $K=6$, *oreganus* appears as an independent northern group with all individuals from northern *thurberi* and especially *montanus* and *shufeldti* forms showing an increasing probability of assignment to the *oreganus* cluster from south to north (Fig. 2).

Patterns of divergence in Nei's distances and F_{ST} values among forms were highly congruent with previous results, with *townsendi*, *pontilis* and *pinosus* showing the highest values for both indices, and northern forms showing lower levels of differentiation, especially between southern and northern *thurberi* individuals (Table S1).

Forward selection of explanatory variables

Out of eleven potentially relevant ecological variables for juncos, six were retained after the forward selection method intended for excluding non-significant effects. Retained variables included isothermality (BIO3), mean temperature of the warmest quarter (BIO10), mean precipitation of the warmest quarter (BIO18), vegetation cover (TREE), greenness (NDVI), and elevation (ELEV) (Table 3). None of these variables was excluded due to excessive correlation as VIF values were below 10 (maximum recovered VIF = 5.75).

Redundancy analysis and variance partition

All six ecological variables retained in the forward selection method were included as explanatory variables in the RDA and the partial RDA models. RDA computes, in successive order, a series of axes that are linear combinations of the explanatory variables, and that best explain the variation in the matrix of response variables (Borcard et al. 2011). Six RDA axes (named RDA1 to RDA6 hereafter, ordered by the amount of variance explained by each one, as reflected by the adjusted R^2) explained 6.26% of the total genetic variance in the non-conditioned model, and 1.18% when removing the effects of neutral genetic structure (Table 4). The amount of explained variance increased to 36.61% when using only BayeScan outliers as response variables (Table 4, Table S2). The permutation tests for the RDA models yielded a p-value below 0.001 in all three analyses, confirming the significance of the constraining variable effects.

Loadings of ecological explanatory variables on each of the axes varied across the three different RDA models (Fig 4, Table 4, Table S2). In the non-conditioned RDA, the RDA1 axis had a large negative contribution of TREE and NDVI, and loaded positively on elevation (Fig. 4A). The RDA2 axis loaded mostly on BIO3 and BIO10. The plot of per-individual projections on these two axes revealed a pattern generally similar to the PCA. The forms *townsendi*, *pontilis* and to a lesser extent *pinosus*, showed distinctive high values of correlation with both axes. The remaining forms showed similar correlation patterns with respect to RDA1, with southern *thurberi* individuals showing a clear association with RDA2 (Fig. 4A).

In the partial RDA, RDA1 had a large contribution of BIO3, while RDA2 loaded mostly on BIO10 and to a lesser extent, BIO18 and TREE (Fig. 4B). Plotting these first two

RDA axes revealed patterns of genetic correlation especially related to the first RDA axis for *pinosus*, which consequently presented the strongest association with isothermality. The individuals from the southern population of *thurberi* showed a more pronounced negative association with RDA2 than in the non-conditioned RDA, further evidence of a positive correlation with mean temperature of the warmest quarter. Northern forms of the Oregon junco (northern *thurberi*, *shufeldti*, *montanus* and *oreganus*) also separated along the second RDA axis, with the northernmost *oreganus* showing the strongest association with the particularly conspicuous mean precipitation of the warmest quarter gradient, a pattern that was not visible in the non-conditioned RDA. *Townsendi* from Baja California occupied positions closer to the origin of coordinates, suggesting a lower association between environmental and genetic variance (Fig. 4B).

In the RDA based on outlier loci potentially under selection, the first axis showed moderate negative contributions from TREE and NDVI, and also a positive contribution of ELEV (Fig. 4C). Variance in BIO18 was almost entirely captured by RDA2, which also had a relatively high, negative contribution from BIO3. The plot showed a pattern of correlation between *pontilis* and *townsendi* along RDA1, while genetic variance in *oreganus* appeared strongly associated with the gradient of mean precipitation of the warmest quarter along RDA2. The rest of Oregon junco forms (and one atypical *oreganus* individual) were distributed in an opposite fashion, with small differences in their patterns of correlation with environmental variability captured in the second axis of the RDA (Fig. 4C).

The variance partition analysis showed that climate and neutral structure together explained 7.17% of the total genetic variability (fractions A+B+C, Fig. 5). Since variable sets are not orthogonal, a 5.08% of variation was explained jointly by the environmental data and the first two components of the PCA based on neutral genetic positions (fraction B, Fig. 5). As recovered in the partial RDA, environmental variables alone explained 1.18% of the total variance (fraction A, Fig. 5), while the non-overlapping fraction of neutral genetic structure explained 0.91% of the variability in the SNP dataset (fraction C, Fig. 5). The p-value computed through the 10,000-step permutation test for each individual fraction was below 0.001 in all cases, thus confirming the significant effects of both variable sets.

Niche divergence tests

We tested for niche divergence and conservatism on each of the environmental variables. We found significant niche divergence between *pinosus* and northern *thurberi* for three of the six environmental variables analyzed (isothermality, mean precipitation of the warmest quarter and elevation; Table 5). When considering northern *thurberi* vs. southern *thurberi*, we found significant divergence for mean temperature of the warmest quarter and conservatism for isothermality (Table 5). NDVI was the only variable that exhibited significant divergence between *townsendi* and *thurberi* south (Table 5).

Genome scans

The BayeScan survey comparing all Oregon junco forms together detected 32 SNPs potentially under divergent selection, and 5 significant SNPs potentially under balancing selection. In the two pairwise comparisons *townsendi* vs. *pontilis* and *townsendi* vs. *thurberi*, 20 and 30 significant SNPs with signs of having diverged under

selection were detected, respectively, and no significant SNPs under balancing selection were found in either case. SNPs potentially differentiated under divergent selection appeared distributed across the genome in all comparisons, without obvious signs of heterogeneity among regions (Fig. 6). Chromosomes 1B and 16 harbored no SNPs so they are not shown in the plot.

Discussion

Neutral population structure and local adaptation explains genomic variance among Oregon junco forms

Our results reveal that both neutral and selective factors have played a role in driving divergence among Oregon junco populations, and that the relative contributions of geographic isolation and environment-driven selection are not uniform across the distribution range of the complex. Environmental variables explained 1.18% of genomic variation when controlling for population structure, and environment and neutral structure together accounted for 7.17% of the variability in the 11,256 SNP matrix. The remaining 92.83% of the variance corresponds potentially to loci under balancing selection or selective pressures not represented in our ecological variables, and to neutral variation shared by all Oregon junco forms because of their close relatedness and/or gene flow among them. The amount of variance explained solely by environmental variables in our study was comparable to the values reported in studies applying RDA to detect specific correlations between genomic variation and a given set of potentially correlated variables, as shown in plants (e.g. Lasky et al. 2012; Vincent et al. 2013; De Kort et al. 2014) or other avian species (Safran et al. 2016; Szulkin et al.

2016). Previous studies on birds have used simple spatial variables such as geographic distance to control for the effects of spatial autocorrelation (Safran et al. 2016; Szulkin et al. 2016). Here, we controlled for genome-wide patterns of neutral variation by subtracting the variance captured by the first two PCs of a PCA based on neutral genome-wide SNPs, a method which should better account for population history and structure, including changes in effective population size, geographic isolation and related effects (Forester et al. 2016). Since spatial autocorrelation is usually reflected by neutral genetic structure, we did not include spatial covariates to avoid over-conditioning the model (Rellstab et al. 2015). Given that only a small fraction of the surveyed genome is expected to be related to genes coding for climatic adaptation or linked to them (Meirmans 2015), a significant 1.18% of association between genomic variation and environmental variability in the conditioned (partial) RDA over only 11,261 genome-wide distributed SNPs is a compelling signal of local adaptation.

Genetic-environment association patterns in the diversification of the Oregon junco

The RDA revealed a number of strikingly different patterns of covariation between genetic variance and ecological variables likely to have played a role in Oregon junco diversification, especially when the effects of population history were removed. The forms *pontilis* and *townsendi* from Baja California, markedly isolated in terms of geography and neutral genetic variability, presented a low genetic-environment association when controlling for population history. This suggests that the differentiation between *townsendi* and *pontilis* is due largely to geographic isolation, in this case caused by unsuitable desert habitat in the lowlands surrounding their respective mountain ranges, a pattern also known as isolation by resistance (IBR, McRae and Beier 2007). In this scenario, our results suggest that differentiation is

caused by drift under conditions of small population size and reduced gene flow, rather than divergent selection due to local adaptation, fitting the classic allopatric speciation model (Mayr 1942, 1963; Coyne and Orr 2004). This hypothesis is consistent with the niche divergence test comparing *townsendi* with southern *thurberi*, for which all tested variables but NDVI showed no signal of divergence beyond expectations based on background divergence.

The form *pinosus* showed considerable neutral genetic structure and a conspicuous pattern of genetic-environment association in both non-conditioned and partial RDA. When controlling for population structure, *pinosus* individuals showed high positive correlation values with isothermality, while correlating negatively with elevation. Indeed, *pinosus* presents the highest isothermality values, and the second lowest elevation after *oreganus*, reflecting a tolerance for low elevation conditions that are absent in neighboring *thurberi*, a pattern also recovered in the niche divergence test. Unlike other differentiated forms like *pontilis* and *townsendi*, *pinosus* does not show high geographic isolation, and zones of intergradation with *thurberi* have been described (Miller 1941). Neutral population divergence despite the absence of geographic barriers to gene flow along with signs of local adaptation is a pattern consistent with isolation-by-adaptation, where barriers to gene flow may have arisen as individuals adapted to the distinct habitat of the coastal mountains of California. Niche distinctiveness and the genetic-environment association pattern of this form is thus congruent with a combination of warm latitude, low elevation and coastal influence that has seemingly resulted in the adaptive differentiation of *pinosus* from the rest of the Oregon junco taxa. As a result, differentiation by drift may have led to positive correlations between adaptive and neutral genetic divergence (Nosil et al. 2008).

The southern *thurberi* individuals from Mount Laguna showed high overlap in terms of neutral genetic structure with northern *thurberi* and other boreal forms, and only slight differences in their genetic-environment association patterns when no controls for confounding factors were implemented. The Mount Laguna site represents the southernmost tip of the *thurberi* range in Southern California, which extends northward and reaches Oregon, forming a relatively continuous distribution (Miller 1941; Nolan et al. 2002), suggesting potentially high gene flow. However, the partial RDA revealed a distinctive pattern of high correlation with the mean temperature of the warmest quarter for the southern *thurberi* juncos, differentiating them from the rest of Oregon forms. They also correlated negatively with the mean precipitation of the warmest quarter. This pattern seems congruent with the habitat of Mount Laguna, and in general with the southern inland range of Oregon juncos, quite arid during summer but subject to snowfalls in winter due to the high elevations (Miller 1941), contrasting sharply with the more climatically moderate coastal and northern populations. The limited neutral genetic structure between *thurberi* range extremes but considerable differentiation in the genetic-environment association patterns is consistent with a process of local adaptation across a selective gradient (Forester et al. 2016), in which selection is the prominent evolutionary force driving differentiation (Haldane 1948; Slatkin 1973; Nagylaki 1975; Felsenstein 1976) while neutral alleles may move freely across space. The niche divergence test comparing southern and northern *thurberi* populations was significant for mean temperature of the warmest quarter.

The boreal Oregon junco forms including *oreganus*, *montanus*, *shufeldti*, and *thurberi* individuals from Tahoe, California, presented a more conspicuous pattern of local

adaptation along a shallow selective gradient. These forms showed very low neutral genetic structure or differences in ecological covariances in the non-conditioned RDA, yet showed an increasing signal of association following their latitudinal distribution in the partial RDA. A strong association pattern emerged especially for mean precipitation of the warmest quarter and for correlated environmental variables of tree cover and greenness, matching quite precisely their latitudinal distribution along a gradient of increasing humidity and vegetation cover. The ecology-related differences in genetic variance, consistent with the taxonomic classification of these forms, is especially relevant considering the relative phenotypic similarity of these taxa, and their apparent intergradation (Miller 1941; Nolan et al. 2002).

GEA methods present a number of limitations, including potentially high rates of type I error (see Supplementary Materials for details). Here, rather than detecting specific loci under selection, we aimed to explore how selection and neutral processes shape the variability in Oregon juncos, but the risk of finding false significant associations between genetic variance and ecological parameters persists. To further test the environmental associations revealed in this study, we implemented a highly conservative approach using only BayeScan outliers as response variables in the model. The non-conditioned RDA based on 87 SNP loci identified by BayeScan as potential targets of divergent selection yielded relatively lower resolution than the partial RDA. BayeScan has been shown to produce relatively few false positives, but it is also a conservative approach, the sensitivity of which decreases with selection strength (Narum and Hess 2011). The RDA suggests that BayeScan correctly identified outliers related to low temperatures and high precipitation for *oreganus* samples, a pattern congruent with the habitat and with the outcomes of previous analyses for this form. It

also detected highly differentiated positions in *pontilis* and *townsendi* that correlate with RDA1, but in this case associations with specific environmental variables were lower, and the pattern disappears in the RDA based on the entire SNP dataset when correcting for population structure. This may suggest that BayeScan failed to exclude the effects of demographic history, or in turn, that controlling for the genetic variance captured by the PCA was overly conservative. BayeScan was also less successful in detecting adaptive divergence in *pinosus*, and especially in northern Oregon junco forms, where selection may be weaker or have acted during a shorter period. Nevertheless, the outlier SNP dataset explained 36.61% of the total climatic variability, a considerable amount compared with the full SNP data RDA models, indicating a good fit of the retained outliers to the linear regression on the environmental parameters.

Interactions among environment, geography and demographic processes result in three different modes of divergence within the Oregon junco lineage

The Oregon junco is one of the six phenotypically and genetically differentiated dark-eyed junco taxa evolved during a northward expansion from Central America after the last glacial maximum, c.a. 18,000 years ago (Milá et al. 2007; Friis et al. 2016). Similar postglacial expansions have been reported for many other bird species (Seutin et al. 1995; Milá et al. 2000; 2006; Hansson et al. 2008; Malpica and Ornelas 2014; Alvarez et al. 2015). However, the population structure documented in this study reveals a variety of different spatial, selective and demographic factors not previously documented in other avian taxa. In light of our genetic-environment association analyses, the patterns recovered by the PCA and STRUCTURE analysis reveal at least three different effects of geography and demographic history interacting to varying degrees with selection in the process of Oregon junco diversification. First, the IBR

666 pattern of differentiation presented by *pontilis* and *townsendi* may suggest that these
667 forms are peripheral remnants of an original, broader distribution of the Oregon juncos,
668 thereafter isolated in ‘sky islands’ of Baja California and diverging predominantly by
669 drift. Indeed, in his thorough monograph on the geographic variation in juncos, Alden
670 Miller (1941) had perceptively suggested early on that the habitat of Oregon juncos
671 from Baja California did not seem to account for their phenotypic differentiation from
672 Californian forms, and that their distinctive traits appeared to be predominantly
673 “historical” (p. 306). The spatial configuration and recovered patterns of neutral and
674 adaptive divergence for *pontilis* and *townsendi* fit the first model of neutral divergence
675 of isolated population in approximately similar habitats (scenario A in Fig. 2). Second,
676 the IBA pattern found in *pinosus* suggests that the current area of intergradation
677 corresponds to a secondary contact zone that formed after diverging in relative isolation,
678 maybe linked to the ancient coastal closed-cone pine forest that has allegedly
679 diminished since the Pleistocene (Miller 1941). The mode of divergence between
680 parapatric populations in different habitats (scenario B in Fig. 2) has been only partially
681 fulfilled, since local adaptation seems to have resulted in reduced levels of gene flow,
682 leading to increased neutral genome-wide differentiation (Nosil et al. 2008; Funk et al.
683 2011; Flaxman et al. 2014). Third, the geographic continuum represented by *thurberi*,
684 *shufeldti*, *montanus* and *oreganus* is also captured in the STRUCTURE analysis by a
685 gradual signal of differentiation following a latitudinal distribution, suggesting that
686 ongoing gene flow may occur among forms. Combined with the signal of increasing
687 environmental association recovered in the partial RDA, these outcomes are consistent
688 with a process of differentiation driven by local adaptation along a selective gradient in
689 the direction of the northward expansion, fulfilling the third hypothesized mode of
690 divergence for these forms of Oregon junco (scenario C in Fig. 2). Interestingly,

ecological association approaches have been shown to perform better along clines of selection where demographic expansions align with the gradient of ecological variables, usually related to latitude (Frichot et al. 2015), as is the case in the *Junco* system.

A relevant aspect of the marked population structure found among Oregon junco forms is that it is based on a relatively small subset of genome-wide SNPs randomly sampled from across the genome, representing a genomic fraction not greater than 0.2% of the total of 1.2 Gb. The clear signal of divergence mediated by environmental factors recovered also in the RDA indicates that divergence may have taken place at the level of the entire genome, suggesting the role of multiple selective pressures during local adaptation along the latitudinally broad and heterogeneous distribution of the Oregon juncos. The presence of outliers potentially under positive selection scattered across the genome seems to support this hypothesis of selection-driven genome-wide divergence, rather than widespread drift among isolated populations. Other examples of such patterns of genomic differentiation due to divergent selection at early (e.g. Parchman et al. 2013; Brawand et al. 2014; Egan et al. 2015) and intermediate (e.g. Riesch et al. 2017) stages of speciation have been reported recently, contrasting with proposed models of speciation initiated by divergent selection in a few, localized genes involved in reproductive isolation (e.g. Nadeau et al. 2012; Poelstra et al. 2014).

Conclusion

Our analyses reveal the role of both local adaptation and demography in driving rapid diversification during the northward recolonization of western North America by the Oregon junco. The combined effects of a demographic expansion along a selective

gradient with a heterogeneous landscape of environmental variability have resulted in a striking array of divergence modes within a single lineage, from isolated forms in Baja California that have differentiated largely by drift in isolated ‘sky islands’, to adaptive diversification along selective gradients with no obvious geographic barriers to gene flow. There is also a compelling example of isolation by adaptation in the case of *pinosus*, where ecological barriers to gene flow seem to maintain its divergence with respect to nearby forms. Genome-wide patterns of divergence indicate that Oregon junco diversification has been driven by multiple ecological factors acting on many loci across the genome, and suggests that selection may promote local adaptation in short periods of time, highlighting the role of adaptive divergence in the early stages of the speciation process. Future analyses of dense sequencing and functional gene characterization will be necessary to further identify adaptive changes promoting barriers to gene flow and reveal the genomic architecture of rapid diversification.

Acknowledgements

We thank Pau Aleixandre, Jonathan Atwell, Elena Berg, Setefilla Buenavista, Steve Burns, Jatziri Calderón, Adrián Gutiérrez, Alfonsina Hernández, Fritz Hertel, Ellen Ketterson, Adán Oliveras de Ita, César Ríos, Sahid Robles, Vicente Rodríguez, Whitney Tsai, Rich Van Buskirk and Alvar Veiga for invaluable help in the field and lab. We are also grateful to Yoann Anselmetti, Clémentine Francois, Rachel Johnston, Etienne Kornobis, Benoit Nabholz, Sergio Nigenda, Jacqueline Robinson, Marjolaine Rouselle and Robert K. Wayne for their assistance with bioinformatic analyses. We also thank Laura Barrios and Luis M. Carrascal for reviewing the statistical analyses and providing

orientation. Funding was provided by grant CGL-2011-25866 from the Spanish
Ministry of Science and Innovation to BM.

Data Accessibility

Genomic data will be deposited in Dryad in short.

Author Contributions

G. Friis and BM designed the study and carried out field sampling; G. Friis, JM, and BF
generated and analyzed genomic data; G. Friis, G. Fandos and AZ generated and
analyzed environmental data; G. Friis and BM wrote the manuscript with input from all
co-authors.

Table 1. Oregon junco forms and number of genotyped individuals per locality. State abbreviations are the following: British Columbia (BC) in Canada; Oregon (OR), California (CA) in the USA; Baja California Norte (BCN) in Mexico.

<i>Form</i>	<i>State</i>	<i>Localities</i>	<i>Sequenced</i>
<i>oreganus</i>	BC	Banks Island, Porcher Island, Susan Island	16
<i>shufeldti</i>	OR	Willamette N.F.	12
<i>montanus</i>	OR	Wallowa N.F.	16
<i>northern thurberi</i>	CA	Tahoe	18
<i>southern thurberi</i>	CA	Mount Laguna	17
<i>pinosus</i>	CA	Santa Cruz Mountains	16
<i>pontilis</i>	BCN	Sierra Juárez	16
<i>townsendi</i>	BCN	Sierra San Pedro Mártir	16
Total			127

Table 2. SNP data matrices used in each analyses. General filters included a depth range from 2 to 100 and a p-value for Hardy-Weinberg deviation of 10^{-4} .

Analysis	Number of samples	Min. phred score	MAF threshold	Number of SNPs	Allowed missing data
STRUCTURE	64	70	0.02	34,367	10%
PCA	88	40	0.02	9,436	0%
RDA:					
On all <i>loci</i>	88	40	0.02	11,261	0%
On BayeScan outliers	88	40	0.02	87	0%
Genome Scans:					
All lineages	88	40	0.02	29,868	75%
<i>townsendi</i> vs. <i>pontilis</i>	24	40	0.05	22,773	75%
<i>townsendi</i> vs. <i>thurberi</i>	24	40	0.05	22,516	75%

Table 3. Set of environmental variables included in the initial stepforward selection method. Significant variables retained by the method are shown in bold.

Environmental variable	Description
BIO1	Annual mean temperature
BIO3	Isothermality
BIO4	Temperature seasonality (standard deviation *100)
BIO10	Mean temperature of the warmest quarter
BIO12	Annual precipitation
BIO15	Precipitation seasonality (coefficient of variation)
BIO18	Precipitation of the warmest quarter
NDVI	Normalized Difference Vegetation Index (greenness)
NDVI SD	Annual NDVI standard deviation (greenness seasonality)
TREE	Percent tree cover
ELEV	Elevation from the NASA Shuttle Radar Topographic Mission

Table 4. RDA loads of the constraining variables in the first two axes and their explained variance for each one of the RDA models. The total variance explained by the full model (adjusted R^2 for the resultant six axes) is shown for each analysis. In all of the three analyses, the p-value for the full models was below 0.001. See Table 3 for variable definitions.

	Non conditioned RDA		Partial RDA		RDA of BayeScan outliers	
Total explained variance (adjusted R^2)	6.26%		1.18%		36.61%	
Variable	RDA1	RDA2	RDA1	RDA2	RDA1	RDA2
BIO3	0.124	0.239	0.839	-0.427	0.214	0.763
BIO10	0.108	-0.610	0.359	-0.679	0.221	0.374
BIO18	0.052	0.021	-0.452	0.501	-0.031	-0.994
ELEV	0.646	-0.031	-0.324	-0.301	0.629	0.420
NDVI	-0.641	0.588	-0.320	0.199	-0.731	-0.277
TREE	-0.710	0.320	-0.270	0.326	-0.776	-0.423
Variance						
Eigenvalue	1304.262	567.705	481.320	350.714	63.655	12.657
Proportion explained	0.025	0.011	0.003	0.002	0.250	0.050
Cumulative proportion	0.025	0.037	0.003	0.005	0.250	0.300

Table 5. Results from the niche divergence test for *pinosus* vs. northern *thurberi*, northern vs. southern *thurberi* and southern *thurberi* vs. *townsendi*. Variables showing significant divergence (Diverged) or conservatism (Conserved) are shown in bold (p-value < 0.05). e.d.b.b.: expected divergence based on background.

	Variable	Result	p-value	Observed mean difference	Background mean difference
<i>pinosus</i> vs.	TREE	e.d.b.b.	0.636	3.95	7.17
northern <i>thurberi</i>	BIO10	e.d.b.b.	0.718	20.40	17.34
	BIO3	Diverged	0.000	16.78	9.45
	BIO18	Diverged	0.000	41.92	27.24
	ELEV	Diverged	0.032	1329.81	978.26
	NDVI	e.d.b.b.	0.550	192.67	428.18
northern vs.	TREE	e.d.b.b.	0.580	2.11	4.52
southern <i>thurberi</i>	BIO10	Diverged	0.030	41.85	23.87
	BIO3	Conserved	0.002	1.37	2.87
	BIO18	e.d.b.b.	0.860	15.02	14.25
	ELEV	e.d.b.b.	0.474	202.60	140.29
	NDVI	e.d.b.b.	0.144	258.89	1019.97
southern <i>thurberi</i>	TREE	e.d.b.b.	0.536	15.01	12.25
vs. <i>townsendi</i>	BIO10	e.d.b.b.	0.412	47.88	43.03
	BIO3	e.d.b.b.	0.472	2.47	2.89
	BIO18	e.d.b.b.	0.188	74.77	67.26
	ELEV	e.d.b.b.	0.786	750.22	717.21
	NDVI	Diverged	0.012	1597.62	574.28

784 **Figure 1.** Geographic distribution and neutral genetic structure of the Oregon junco
785 forms. (A) Breeding ranges of all the Oregon junco forms. (B) Genetic structure of
786 Oregon junco forms based on a principal components analysis of independent,
787 selectively neutral genome-wide SNPs.

788
789
790

Figure 2. Expectations for neutral and adaptive divergence under different environmental and spatial configurations found across the Oregon junco distribution. (A) Geographically isolated populations in similar habitats. (B) Parapatric populations in ecologically divergent habitats. (C) Population continuum across a selective gradient.

Figure 3. Genetic structure of the Oregon junco forms based on 34,367 selectively neutral genome-wide SNPs using the program STRUCTURE. Each horizontal bar corresponds to an individual, with colors corresponding to posterior assignment probabilities to each of a number of genetic clusters (K). Colors correspond approximately to those in Fig. 2A.

Figure 4. Genetic-environment association analyses in the Oregon junco. Points represent the projection of individual genotypes on the first two RDA axes. Marker colors correspond to those on the range map on Fig. 2A. The explanatory variables are shown within the space defined by RDA1 and RDA2 by labeled vectors. Their contribution to each axis is represented by the length of their orthogonal projections over the scale bars along top and right sides of the graphs. Arrows indicate the direction of the gradient of variation for the corresponding environmental parameter. The value for each sample point on each explanatory variable can be obtained by an orthogonal projection on the corresponding plotted vector. (A) First two RDA axes of a non-conditioned RDA based on 11,261 SNPs. (B) First two RDA axes of a partial RDA based on 11,261 SNPs conditioned by neutral genetic structure, approximated by the first two PCs of a PCA based on neutral markers. (C) First two RDA axes of a non-conditioned RDA based on 87 SNP outliers identified in a BayeScan analysis.

819

820 **Figure 5.** Plot of the fractions of the genetic variability in Oregon samples explained by
 821 (A) environmental variables alone; (B) the overlap of both environmental variables and
 822 neutral structure; and (C) neutral genetic structure alone; and the unexplained genetic
 823 variability (residuals). P-values computed through a 1000-step permutation test for the
 824 fractions A, B and C, were below 0.001 in all cases.
 825

826
 827

Figure 6. Plot of per-SNP posterior probability of divergence mediated by selection (shown as $1 - Q$ -value as computed by BayeScan) in (A) all Oregon junco forms together; (B) *townsendi* against *pontilis* and (C) *townsendi* against all *thurberi*. Loci above the dotted line (in red) are those below a false discovery rate of 10%.

References

- Alvarez, S., J. F. Salter, J. E. McCormack, and B. Milá. 2015. Speciation in mountain refugia: phylogeography and demographic history of the pine siskin and black-capped siskin complex. *Journal of Avian Biology* 47:335-345.
- Andrews, S. 2010. FastQC: a quality control tool for high throughput sequence data.
- Auwer, G. A., M. O. Carneiro, C. Hartl, R. Poplin, G. del Angel, A. Levy-Moonshine, T. Jordan, K. Shakir, D. Roazen, and J. Thibault. 2013. From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. *Current protocols in bioinformatics*:11.10. 11-11.10. 33.
- Barve, N., V. Barve, A. Jiménez-Valverde, A. Lira-Noriega, S. P. Maher, A. T. Peterson, J. Soberón, and F. Villalobos. 2011. The crucial role of the accessible area in ecological niche modeling and species distribution modeling. *Ecological Modelling* 222:1810-1819.
- Bierne, N., J. Welch, E. Loire, F. Bonhomme, and P. David. 2011. The coupling hypothesis: why genome scans may fail to map local adaptation genes. *Molecular Ecology* 20:2044-2072.
- Billiard, S., T. Lenormand, and S. Gavrillets. 2005. Evolution of migration under kin selection and local adaptation. *Evolution* 59:13-23.
- Blanchet, F. G., P. Legendre, and D. Borcard. 2008. Forward selection of explanatory variables. *Ecology* 89:2623-2632.
- Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*:btu170.
- Borcard, D., F. Gillet, and P. Legendre. 2011. Chapter 6. Canonical Ordination. Pp. 153-226. *Numerical Ecology with R*. Springer.

859 Brawand, D., C. E. Wagner, Y. I. Li, M. Malinsky, I. Keller, S. Fan, O. Simakov, A. Y.
860 Ng, Z. W. Lim, and E. Bezault. 2014. The genomic substrate for adaptive
861 radiation in African cichlid fish. *Nature* 513:375.

862 Carson, H. L. 1975. The genetics of speciation at the diploid level. *The American*
863 *Naturalist* 109:83-92.

864 Chhatre, V. and K. Emerson. 2016. StrAuto: Automation and parallelization of
865 STRUCTURE analysis. See <http://strauto.popgen.org>.

866 Christmas, M. J., E. Biffin, M. F. Breed, and A. J. Lowe. 2016. Finding needles in a
867 genomic haystack: targeted capture identifies clear signatures of selection in a
868 nonmodel plant species. *Molecular Ecology* 25:4216-4233.

869 Coyne, J. A. and H. A. Orr. 2004. *Speciation*. Sinauer Associates, Inc., Sunderland,
870 Massachusetts.

871 Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E.
872 Handsaker, G. Lunter, G. T. Marth, and S. T. Sherry. 2011. The variant call
873 format and VCFtools. *Bioinformatics* 27:2156-2158.

874 Darwin, C. 1859. *The origin of species by means of natural selection, or the*
875 *preservation of favoured races in the struggle for life*. John Murray, London.

876 De Kort, H., K. Vandepitte, H. H. Bruun, D. Closset-Kopp, O. Honnay, and J. Mergeay.
877 2014. Landscape genomics and a common garden trial reveal adaptive
878 differentiation to temperature across Europe in the tree species *Alnus glutinosa*.
879 *Molecular ecology* 23:4709-4721.

880 DePristo, M. A., E. Banks, R. Poplin, K. V. Garimella, J. R. Maguire, C. Hartl, A. A.
881 Philippakis, G. Del Angel, M. A. Rivas, and M. Hanna. 2011. A framework for
882 variation discovery and genotyping using next-generation DNA sequencing data.
883 *Anglais* 43:491-498.

884 Dray, S., P. Legendre, and G. Blanchet. 2009. packfor: Forward Selection with
885 permutation. R package version 0.0-7/r58.

886 Dwight, J. 1918. The geographical distribution of color and of other variable characters
887 in the genus *Junco*: a new aspect of specific and subspecific values. Bull. Am
888 Mus. Nat. Hist. 38:269-309.

889 Edmonds, C. A., A. S. Lillie, and L. L. Cavalli-Sforza. 2004. Mutations arising in the
890 wave front of an expanding population. Proceedings of the National Academy of
891 Sciences of the United States of America 101:975-979.

892 Egan, S. P., G. J. Ragland, L. Assour, T. H. Powell, G. R. Hood, S. Emrich, P. Nosil,
893 and J. L. Feder. 2015. Experimental evidence of genome-wide impact of
894 ecological selection during early stages of speciation-with-gene-flow. Ecology
895 letters 18:817-825.

896 Elshire, R. J., J. C. Glaubitz, Q. Sun, J. A. Poland, K. Kawamoto, E. S. Buckler, and S.
897 E. Mitchell. 2011. A Robust, Simple Genotyping-by-Sequencing (GBS)
898 Approach for High Diversity Species. PLoS ONE 6:e19379.

899 Engler, R., A. Guisan, and L. Rechsteiner. 2004. An improved approach for predicting
900 the distribution of rare and endangered species from occurrence and pseudo-
901 absence data. Journal of applied ecology 41:263-274.

902 Excoffier, L., T. Hofer, and M. Foll. 2009. Detecting loci under selection in a
903 hierarchically structured population. Heredity 103:285-298.

904 Excoffier, L. and N. Ray. 2008. Surfing during population expansions promotes genetic
905 revolutions and structuration. Trends in ecology & evolution 23:347-351.

906 Faircloth, B. C. and T. C. Glenn. 2012. Not all sequence tags are created equal:
907 designing and validating sequence identification tags robust to indels. PloS one
908 7:e42543.

909 Faria, R., S. Renaut, J. Galindo, C. Pinho, J. Melo-Ferreira, M. Melo, F. Jones, W.
910 Salzburger, D. Schluter, and R. Butlin. 2014. Advances in Ecological
911 Speciation: an integrative approach. *Molecular ecology* 23:513-521.

912 Felsenstein, J. 1976. The theoretical population genetics of variable selection and
913 migration. *Annual review of genetics* 10:253-280.

914 Flaxman, S. M., A. C. Wacholder, J. L. Feder, and P. Nosil. 2014. Theoretical models
915 of the influence of genomic architecture on the dynamics of speciation.
916 *Molecular ecology* 23:4074-4088.

917 Foll, M. 2012. Bayescan v2. 1 user manual. *Ecology* 20:1450-1462.

918 Foll, M. and O. E. Gaggiotti. 2008. A genome scan method to identify selected loci
919 appropriate for both dominant and codominant markers: A Bayesian perspective.
920 *Genetics* 180:977-993.

921 Forester, B. R., M. R. Jones, S. Joost, E. L. Landguth, and J. R. Lasky. 2016. Detecting
922 spatial genetic signatures of local adaptation in heterogeneous landscapes.
923 *Molecular ecology* 25:104-120.

924 Forester, B. R., J. R. Lasky, H. H. Wagner, and D. L. Urban. 2017. Comparing methods
925 for detecting multilocus adaptation with multivariate genotype-environment
926 associations. *bioRxiv*.

927 Frichot, E., S. D. Schoville, P. de Villemereuil, O. E. Gaggiotti, and O. François. 2015.
928 Detecting adaptive evolution based on association with ecological gradients:
929 orientation matters! *Heredity* 115:22-28.

930 Friis, G., P. Aleixandre, R. Rodríguez-Estrella, A. G. Navarro-Sigüenza, and B. Milá.
931 2016. Rapid postglacial diversification and long-term stasis within the songbird
932 genus *Junco*: phylogeographic and phylogenomic evidence. *Molecular Ecology*
933 25:6175-6195.

934 Funk, D. J., S. P. Egan, and P. Nosil. 2011. Isolation by adaptation in *Neochlamisus* leaf
935 beetles: host-related selection promotes neutral genomic divergence. *Molecular*
936 *Ecology* 20:4671-4682.

937 Gordon, A. and G. Hannon. 2010. Fastx-toolkit. FASTQ/A short-reads preprocessing
938 tools (unpublished) http://hannonlab.cshl.edu/fastx_toolkit.

939 Haldane, J. B. 1948. The theory of a cline. *Journal of genetics* 48:277-284.

940 Hansson, B., D. Hasselquist, M. Tarka, P. Zehindjiev, and S. Bensch. 2008. Postglacial
941 colonisation patterns and the role of isolation and expansion in driving
942 diversification in a passerine bird. *PLoS One* 3:e2794.

943 Hedrick, P. W., M. E. Ginevan, and E. P. Ewing. 1976. Genetic polymorphism in
944 heterogeneous environments. *Annual review of Ecology and Systematics* 7:1-32.

945 Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis. 2005. Very high
946 resolution interpolated climate surfaces for global land areas. *International*
947 *Journal of Climatology* 25:1965-1978.

948 Jones, M. R., B. R. Forester, A. I. Teufel, R. V. Adams, D. N. Anstett, B. A. Goodrich,
949 E. L. Landguth, S. Joost, and S. Manel. 2013. Integrating landscape genomics
950 and spatially explicit approaches to detect loci under selection in clinal
951 populations. *Evolution* 67:3455-3468.

952 Kawecki, T. J. and D. Ebert. 2004. Conceptual issues in local adaptation. *Ecology*
953 *letters* 7:1225-1241.

954 Kopelman, N. M., J. Mayzel, M. Jakobsson, N. A. Rosenberg, and I. Mayrose. 2015.
955 Clumpak: a program for identifying clustering modes and packaging population
956 structure inferences across K. *Molecular ecology resources* 15:1179-1191.

957 Lasky, J. R., D. L. Des Marais, J. McKAY, J. H. Richards, T. E. Juenger, and T. H.
958 Keitt. 2012. Characterizing genomic variation of *Arabidopsis thaliana*: the roles
959 of geography and climate. *Molecular Ecology* 21:5512-5529.

960 Legendre, P. and L. Legendre. 1998. Numerical ecology: second English edition.
961 Developments in environmental modelling 20.

962 Leggett, R. M., B. J. Clavijo, L. Clissold, M. D. Clark, and M. Caccamo. 2013.
963 NextClip: an analysis and read preparation tool for Nextera Long Mate Pair
964 libraries. *Bioinformatics*:btt702.

965 Li, H. and R. Durbin. 2009. Fast and accurate short read alignment with Burrows–
966 Wheeler transform. *Bioinformatics* 25:1754-1760.

967 Lischer, H. E. L. and L. Excoffier. 2012. PGDSpider: an automated data conversion tool
968 for connecting population genetics and genomics programs. *Bioinformatics*
969 28:298-299.

970 Luo, R., B. Liu, Y. Xie, Z. Li, W. Huang, J. Yuan, G. He, Y. Chen, Q. Pan, and Y. Liu.
971 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de
972 novo assembler. *Gigascience* 1:18.

973 Malpica, A. and J. F. Ornelas. 2014. Postglacial northward expansion and genetic
974 differentiation between migratory and sedentary populations of the broad-tailed
975 hummingbird (*Selasphorus platycercus*). *Molecular Ecology* 23:435-452.

976 Manthey, J. D. and R. G. Moyle. 2015. Isolation by environment in white-breasted
977 nuthatches (*Sitta carolinensis*) of the Madrean Archipelago sky islands: a
978 landscape genomics approach. *Molecular Ecology* 24:3628-3638.

979 Mayr, E. 1942. Systematics and the origin of species. Columbia Univ. Press, New York.

980 Mayr, E. 1947. Ecological factors in speciation. *Evolution*:263-288.

981 Mayr, E. 1954. Change of genetic environment and evolution.

982 Mayr, E. 1963. Animal species and evolution. Belknap Press, Cambridge, MA.

983 McCormack, J. E., S. M. Hird, A. J. Zellmer, B. C. Carstens, and R. T. Brumfield. 2013.

984 Applications of next-generation sequencing to phylogeography and

985 phylogenetics. *Molecular Phylogenetics and Evolution* 66:526-538.

986 McCormack, J. E., A. J. Zellmer, and L. L. Knowles. 2010. Does Niche Divergence

987 Accompany Allopatric Divergence in Aphelocoma Jays as Predicted Under

988 Ecological Speciation?: Insights from Tests with Niche Models. *Evolution*

989 64:1231-1244.

990 McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K.

991 Garimella, D. Altshuler, S. Gabriel, and M. Daly. 2010. The Genome Analysis

992 Toolkit: a MapReduce framework for analyzing next-generation DNA

993 sequencing data. *Genome research* 20:1297-1303.

994 McRae, B. H. and P. Beier. 2007. Circuit theory predicts gene flow in plant and animal

995 populations. *Proceedings of the National Academy of Sciences* 104:19885-

996 19890.

997 Meirmans, P. G. 2015. Seven common mistakes in population genetics and how to

998 avoid them. *Molecular Ecology* 24:3223-3231.

999 Milá, B., P. Aleixandre, S. Alvarez-Nordström, J. McCormack, E. Ketterson, and J.

1000 Atwell. 2016. More than meets the eye: Lineage diversity and evolutionary

1001 history of Dark-eyed and Yellow-eyed juncos. *Snowbird: Integrative Biology*

1002 and Evolutionary Diversity in the Junco (ED Ketterson and JW Atwell, Editors).

1003 University of Chicago Press, Chicago, Illinois, USA:179-198.

1004 Milá, B., D. J. Girman, M. Kimura, and T. B. Smith. 2000. Genetic evidence for the

1005 effect of a postglacial population expansion on the phylogeography of a North

1006 American songbird. Proceedings of the Royal Society Biological Sciences Series
1007 B 267:1033-1040.

1008 Milá, B., J. E. McCormack, G. Castaneda, R. K. Wayne, and T. B. Smith. 2007. Recent
1009 postglacial range expansion drives the rapid diversification of a songbird lineage
1010 in the genus *Junco*. Proceedings of the Royal Society B-Biological Sciences
1011 274:2653-2660.

1012 Milá, B., T. B. Smith, and R. K. Wayne. 2006. Postglacial population expansion drives
1013 the evolution of long-distance migration in a songbird. *Evolution* 60:2403-2409.

1014 Miller, A. 1941. Speciation in the avian genus *Junco*. University of California
1015 Publications in Zoology 44:173-434.

1016 Mitton, J., Y. Linhart, J. Hamrick, and J. Beckman. 1977. Observations on the genetic
1017 structure and mating system of ponderosa pine in the Colorado Front Range.
1018 *Theoretical and Applied Genetics* 51:5-13.

1019 Nadeau, N. J., A. Whibley, R. T. Jones, J. W. Davey, K. K. Dasmahapatra, S. W.
1020 Baxter, M. A. Quail, M. Joron, R. H. ffrench-Constant, M. L. Blaxter, J. Mallet,
1021 and C. D. Jiggins. 2012. Genomic islands of divergence in hybridizing
1022 *Heliconius* butterflies identified by large-scale targeted sequencing.
1023 *Philosophical Transactions of the Royal Society B: Biological Sciences*
1024 367:343-353.

1025 Nadeau, S., P. G. Meirmans, S. N. Aitken, K. Ritland, and N. Isabel. 2016. The
1026 challenge of separating signatures of local adaptation from those of isolation by
1027 distance and colonization history: The case of two white pines. *Ecology and*
1028 *evolution* 6:8649-8664.

1029 Nagylaki, T. 1975. Conditions for the existence of clines. *Genetics* 80:595-615.

1030 Narum, S. R. and J. E. Hess. 2011. Comparison of FST outlier tests for SNP loci under
1031 selection. *Molecular Ecology Resources* 11:184-194.

1032 Nolan, V. J., E. D. Ketterson, D. A. Cristol, C. M. Rogers, E. D. Clotfelter, R. C. Titus,
1033 S. J. Schoech, and E. Snajdr. 2002. Dark-eyed Junco (*Junco hyemalis*) in A.
1034 Poole, and F. Gill, eds. *The Birds of North America*. The Birds of North
1035 America, Inc., Philadelphia, Pennsylvania.

1036 Nosil, P. 2012. *Ecological Speciation*. Oxford University Press.

1037 Nosil, P., S. P. Egan, and D. J. Funk. 2008. Heterogeneous genomic differentiation
1038 between walking-stick ecotypes: “isolation by adaptation” and multiple roles for
1039 divergent selection. *Evolution* 62:316-336.

1040 Ojeda Alayon, D. I., C. K. Tsui, N. Feau, A. Capron, B. Dhillon, Y. Zhang, S.
1041 Massoumi Alamouti, C. K. Boone, A. L. Carroll, and J. E. Cooke. 2017. Genetic
1042 and genomic evidence of niche partitioning and adaptive radiation in mountain
1043 pine beetle fungal symbionts. *Molecular Ecology* 26:2077-2091.

1044 Oksanen, J., F. Blanchet, R. Kindt, P. Legendre, and R. O’Hara. 2016. *Vegan*:
1045 community ecology package. *R Packag.* 2.3-5.

1046 Parchman, T., Z. Gompert, M. Braun, R. Brumfield, D. McDonald, J. Uy, G. Zhang, E.
1047 Jarvis, B. Schlinger, and C. Buerkle. 2013. The genomic consequences of
1048 adaptive divergence and reproductive isolation between species of manakins.
1049 *Molecular Ecology* 22:3304-3317.

1050 Patterson, N., A. L. Price, and D. Reich. 2006. Population structure and eigenanalysis.
1051 *PLoS genet* 2:e190.

1052 Poelstra, J. W., N. Vijay, C. M. Bossu, H. Lantz, B. Ryll, I. Müller, V. Baglione, P.
1053 Unneberg, M. Wikelski, M. G. Grabherr, and J. B. W. Wolf. 2014. The genomic

1054 landscape underlying phenotypic integrity in the face of gene flow in crows.
1055 Science 344:1410-1414.

1056 Pritchard, J. K. and A. Di Rienzo. 2010. Adaptation—not by sweeps alone. Nature
1057 Reviews Genetics 11:665-667.

1058 Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure
1059 using multilocus genotype data. Genetics 155:945-959.

1060 R_Core_Team. 2013. R: a language and environment for statistical computing. R
1061 Foundation for Statistical Computing, Vienna, Austria.

1062 R_Studio_Team. 2015. RStudio: Integrated Development for R. R Studio, Inc., Boston,
1063 MA.

1064 Rellstab, C., F. Gugerli, A. J. Eckert, A. M. Hancock, and R. Holderegger. 2015. A
1065 practical guide to environmental association analysis in landscape genomics.
1066 Molecular Ecology 24:4348-4370.

1067 Riesch, R., M. Muschick, D. Lindtke, R. Villoutreix, A. A. Comeault, T. E. Farkas, K.
1068 Lucek, E. Hellen, V. Soria-Carrasco, and S. R. Dennis. 2017. Transitions
1069 between phases of genomic differentiation during stick-insect speciation. Nature
1070 Ecology & Evolution 1:0082.

1071 Rundle, H. and P. Nosil. 2005. Ecological speciation. Ecology Letters 8:336-352.

1072 Safran, R., E. Scordato, M. Wilkins, J. K. Hubbard, B. Jenkins, T. Albrecht, S.
1073 Flaxman, H. Karaardıç, Y. Vortman, and A. Lotem. 2016. Genome-wide
1074 differentiation in closely related populations: the roles of selection and
1075 geographic isolation. Molecular Ecology 25:3865-3883.

1076 Schluter, D. 2000. The ecology of adaptive radiation. Oxford University Press, Oxford.

1077 Schmieder, R. and R. Edwards. 2011. Quality control and preprocessing of
1078 metagenomic datasets. Bioinformatics 27:863-864.

1079 Seutin, G. 1991. Preservation of avian blood and tissue samples for DNA analyses.
1080 Canadian Journal of Zoology 69:82-90.

1081 Seutin, G., L. M. Ratcliffe, and P. T. Boag. 1995. Mitochondrial DNA homogeneity in
1082 the phenotypically diverse redpoll finch complex (Aves: Carduelinae: *Carduelis*
1083 *flammea-hornemanni*). Evolution 49:962-973.

1084 Slatkin, M. 1973. Gene flow and selection in a cline. Genetics 75:733-756.

1085 Soberon, J. and A. T. Peterson. 2005. Interpretation of models of fundamental
1086 ecological niches and species' distributional areas.

1087 Sork, V. L., K. Squire, P. F. Gugger, S. E. Steele, E. D. Levy, and A. J. Eckert. 2016.
1088 Landscape genomic analysis of candidate genes for climate adaptation in a
1089 California endemic oak, *Quercus lobata*. American journal of botany 103:33-46.

1090 Szulkin, M., P. A. Gagnaire, N. Bierne, and A. Charmantier. 2016. Population genomic
1091 footprints of fine-scale differentiation between habitats in Mediterranean blue
1092 tits. Molecular ecology 25:542-558.

1093 Templeton, A. R. 1981. Mechanisms of speciation-a population genetic approach.
1094 Annual review of Ecology and Systematics 12:23-48.

1095 Termignoni-García, F., J. P. Jaramillo-Correa, J. Chablé-Santos, M. Liu, A. J. Shultz, S.
1096 V. Edwards, and P. Escalante Pliego. 2017. Genomic footprints of adaptation in
1097 a cooperatively breeding tropical bird across a vegetation gradient. Molecular
1098 Ecology.

1099 Turner, S. D. 2014. qqman: an R package for visualizing GWAS results using QQ and
1100 manhattan plots. bioRxiv:005165.

1101 Uyeda, J. C., S. J. Arnold, P. A. Hohenlohe, and L. S. Mead. 2009. Drift promotes
1102 speciation by sexual selection. Evolution 63:583-594.

1103 Van Den Wollenberg, A. L. 1977. Redundancy analysis an alternative for canonical
1104 correlation analysis. *Psychometrika* 42:207-219.

1105 Vincent, B., M. Dionne, M. P. Kent, S. Lien, and L. Bernatchez. 2013. Landscape
1106 genomics in Atlantic salmon (*Salmo salar*): searching for gene–environment
1107 interactions driving local adaptation. *Evolution* 67:3469-3487.

1108 Wang, I. J. and G. S. Bradburd. 2014. Isolation by environment. *Molecular Ecology*
1109 23:5649-5662.

1110 Warren, D. L., R. E. Glor, and M. Turelli. 2008. Environmental Niche Equivalency
1111 Versus Conservatism: Quantitative Approaches to Niche Evolution. *Evolution*
1112 62:2868-2883.

1113 Wright, S. 1943. Isolation by distance. *Genetics*. 28:114-138.

1114 Wright, S. 1946. Isolation by distance under diverse systems of mating. *Genetics* 31:39.

1115 Yates, A., W. Akanni, M. R. Amode, D. Barrell, K. Billis, D. Carvalho-Silva, C.
1116 Cummins, P. Clapham, S. Fitzgerald, and L. Gil. 2016. Ensembl 2016. *Nucleic*
1117 *acids research*:gkv1157.

1118 Zheng, X. 2012. SNPRelate: parallel computing toolset for genome-wide association
1119 studies. R package version 95.

1120

1121