THESIS

GENOMICS-INFORMED CONSERVATION UNITS REVEAL SPATIAL VARIATION IN CLIMATE VULNERABILITY IN A MIGRATORY BIRD

Submitted by

Caitlin Vanessa Miller

Graduate Degree Program in Ecology

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Spring 2023

Master's Committee:

Advisor: Kristen Ruegg

Chris Funk
Kyle Horton

ABSTRACT

GENOMICS-INFORMED CONSERVATION UNITS REVEAL SPATIAL VARIATION IN CLIMATE VULNERABILITY IN A MIGRATORY BIRD

Identifying conservation units (CUs) in threatened species is critical for the preservation of adaptive capacity and evolutionary potential in the face of climate change. However, delineating CUs in highly mobile species remains a challenge due to high rates of gene flow and genetic signatures of isolation by distance. If CUs are delineated in highly mobile species, the CUs often lack key biological information about what populations have the most conservation need to guide management decisions. Here we implement a framework for rigorous CU identification in the Canada Warbler (*Cardellina canadensis*), a high-dispersal migratory bird species of conservation concern, and then integrate demographic modeling and genomic offset within a CU framework to guide conservation decisions. We find that whole-genome structure in this highly mobile species is primarily driven by putative adaptive variation. Identification of CUs across the breeding range revealed that Canada Warblers fall into two Evolutionary Significant Units (ESU), with three putative Adaptive Units (AUs) in the South, East and Northwest. Quantification of genomic offset within each AU reveals significant spatial variation in climate vulnerability, with the Northwestern AU being identified as the most vulnerable to future climate change. Alternatively, quantification of past population trends within each AU revealed the steepest population declines have occurred within the Eastern AU. Overall, we illustrate that genomics-informed CUs

provide a strong foundation for identifying current and potential future region-specific threats that can be used to manage highly mobile species in a rapidly changing world.

TABLE OF CONTENTS

Introduction

Recent estimates of biodiversity loss suggest that up to 60% of animal species are at risk of decline (Grooten & Almond, 2018), leading to an urgent need to identify and conserve threatened species. Conservation efforts to stop biodiversity loss focus on preserving biodiversity at the ecosystem, species, and genetic levels (Coates et al., 2018). While ecosystem and species level protections have historically been easier to quantify and administer, maintenance of genetic diversity is equally important for long term ecosystem viability (Exposito-Alonso et al., 2022; Ralls et al., 2018; Ruegg & Turbek, 2022). Species with low or declining genetic diversity are threatened by inbreeding depression (Frankham, 2003) and the loss of adaptive capacity (Thurman et al., 2020), which may lead to high extinction risk (Forester et al., 2022). With climate change further accelerating biodiversity loss across biological scales (Dale et al., 2001), it is increasingly important to maintain genetic diversity within vulnerable populations to allow them to adapt. However, the ability to identify populations most vulnerable to climate change and develop strategies for protecting them are not always straightforward.

Current strategies to protect populations or species based on genetic diversity often rely on the designation of intraspecific Conservation Units (CUs) to guide conservation and management decisions (Paetkau, 1999). There are many approaches to designating CUs, depending on the conservation priorities for the species. One of the most recognizable intraspecific CUs is the Evolutionarily Significant Unit (ESU). ESUs are generally designated as evolutionarily or ecologically distinct populations within a

broader species, though exact definitions may vary (de Guia & Saitoh, 2007; Moritz, 1994; Ryder, 1986; Waples, 1991). Generally ESUs are not legally protected units, but can have legal protected status when classified as distinct population segments (DPS) under the Endangered Species Act (ESA) in the USA or as designatable units (DU) under the Species at Risk Act in Canada (COSEWIC, 2015; USFW & NMFS, 1996; for a comparison of the designations see Waples et al., 2013). Alternately, Adaptive Units (AUs) are intraspecific groups that share similar adaptive traits and represent groups both within and across ESUs that are adapted to similar environments (Barbosa et al., 2018; Funk et al., 2012). Historically, CU designation was focused on ESUs and management units (MUs; smaller demographically independent populations inside ESUs), as AUs were difficult to define due to limited genomic tools needed to identify adaptive genetic markers (Luikart et al., 2003). However, with the advent of Next-Generation sequencing and landscape genomic methods it is now possible to identify putatively adaptive loci, how they are linked with environmental variation, and how these gene environment relationships may change with changing climate conditions. While each type of CU is important for preserving different aspects of genetic diversity, here we focus on considering both deeper genetic splits representing ESUs and adaptive differences that define AUs. Using ESUs and AUs together may illustrate region-specific variation in climate change response, as AUs may have different responses to climate change in light of locally adaptive differences, while different ESUs may have different response to climate change due to long-term isolation.

Identifying CU boundaries that rigorously integrate key biological information critical to conservation in the face of climate change is not always straightforward. One

issue that often arises but has not always been adequately dealt with (see Turbek et al 2023) is the need to identify CUs in highly mobile species. Establishing clear CU boundaries in organisms with high capacity for dispersal (e.g., migratory birds, bats, and many marine organisms) using genomics alone can be difficult because high levels of dispersal can lead to high gene flow between nearby populations. High gene flow can then result in a signature of isolation by distance- where increasing distance correlates with decreased genetic similarity- that makes it difficult to differentiate CUs despite clear genetic variation throughout a species' range (Kekkonen et al., 2011; Palumbi, 1994; Veith et al., 2004). Adding to this complexity, CUs that do not take genetic evolutionary potential into account may miss the opportunity to protect populations with the greatest conservation need. While metrics such as genomic offset are making it possible to estimate which populations will need to evolve most to keep pace with changing climate conditions (Capblancq et al., 2020; DeSaix et al., 2022; Fitzpatrick & Keller, 2015; Rellstab, 2021; Ruegg et al., 2018), such estimates have not yet been quantified within a CU framework. Here we implement a framework for rigorous CU identification in the Canada Warbler (*Cardellina canadensis*), a high-dispersal migratory bird species of conservation concern, and demonstrate how key biological information (e.g., genomic offset and population demography) can be integrated within a CU framework to guide conservation decisions at a region-specific level.

The Canada Warbler is a migratory songbird whose breeding range extends from Northwestern Canada to the Southeastern United States. Populations across the breeding range have declined 1.9% per year on average from 1966-2019 (Pardieck et al., 2020). Currently, Canada Warblers have federal protection in Canada under the

Species at Risk Act (COSEWIC, 2008) but are considered least concern under ICUN red-list designation (BirdLife International, 2021) partially due to their large range and heterogeneous declines. Previous genetic research, using eight microsatellite markers from 3 breeding sites, found that birds in the Southern portion of the range were genetically distinct from birds in the Eastern and Northwestern portions of the range, but the Eastern and Northwestern birds were not distinct from each other (Ferrari et al., 2018). Future conservation efforts would be bolstered by more genetic information about population structure within the species, whether declines have been focused in areas that contain unique genetic diversity, and which populations are likely to be most vulnerable to changing climate conditions.

Here we use whole genome resequencing to examine population structure across the Canada Warbler breeding range and identify putatively adaptive loci and neutral loci. We adopt the framework proposed by Turbek et al. (2023) to guide CU designation in highly mobile organisms. In addition, we quantify abundance and trend with demographic data, and genomic offset with adaptive loci both across the breeding range and within identified CUs to assess where management interventions would be most important. The resulting data provide a framework for integrating CU designations with estimates of genomic offset to improve our ability to identify and manage vulnerable populations in a changing world.

Methods

Sample collection and DNA extraction

To obtain create a reference genome, we captured a male Canada Warbler (record #SF12T03) on June 12th, 2019 in Rothrock State Forest, Pennsylvania (40.733402, -77.755214) to obtain blood for a high molecular weight (HMW) DNA sample. We affixed a standard USGS aluminum band (#284029445) and took standard measurements and photographs, then drew ~10uL of blood with a capillary tube from a brachial vein puncture. Using the blood, we extracted DNA using the Qiagen MagAttract HMW DNA Mini Kit (cat. no. 67563) with minor modifications to the standard protocol. We found that, likely because avian blood is nucleated such that there is a very high amount of DNA, the DNA became tightly bound to the beads, such that the standard elution protocol would not yield sufficient HMW DNA. Thus, we eluted in 200uL of water and left it on the mixer (at low speed) for approximately 1 hour.

After obtaining a sample for the reference genome, we collected samples from an additional 181 breeding adult Canada Warblers from across the breeding range in North America in collaboration with multiple university researchers and state and federal agencies (Supplemental Figure 1). For DNA extraction, we collected blood from 134 individuals (~80 µl), via brachial venipuncture and preserved it in Queen's lysis buffer and stored at room temperature. Blood (50-80 µl) was extracted using Qiagen DNeasy Blood and Tissue Kits (QIAGEN) and eluted into 100 µl of provided AE buffer. For the remaining 47 individuals, we collected tail feathers by pulling 2 tail feathers from each bird and storing feathers at -20C. We cut the calamus of one feather from the shaft and

extracted the calamus using the modified Qiagen DNeasy Blood and Tissue protocol (Schweizer et al., 2021). After DNA extraction, we quantified samples using Qubit dsDNA assay.

DNA sequencing

Using the HMW DNA obtained for the reference genome, we used 10X linked read sequencing to generate a whole-genome reference sequence of a Canada Warbler. Sequencing was part of "CanSeq150" project (https://www.cgen.ca/canseq150-project-list) in partnership with Birds Canada / Oiseaux Canada. 10X Genomics libraries were prepared at The Centre for Applied Genomics at The Hospital for Sick Children (Toronto, Canada) and libraries were sequenced on a HiSeq X machine (Illumina, San Diego) lane, with 150-bp paired-end reads.

After successfully sequencing our reference genome, we prepared the additional breeding samples for low coverage whole genome sequencing using a modified Nextera prep (Schweizer et al., 2021) with normalized DNA input. We sequenced samples in two libraries, 110 individuals on an Illumina HiSeq 4000 using paired end 150bp reads and 72 individuals on an Illumina NovaSeq 6000 using paired end 150bp reads. The 72 individuals on the NovaSeq were sequenced across multiple lanes to get to the targeted sequencing depth of 2-3X coverage per sample (for sequencing scheme, see Supplemental Table 1) and included replicates of 32 samples with lower than 1.5X coverage from the HiSeq 4000 run.

Bioinformatic processing

We assembled the reference genome using Supernova 2.1.1 (Weisenfeld et al. 2017) on the Pennsylvania State University's Institute for Computational Data Sciences'

Roar supercomputer. After genome assembly, we used Conda v4.13.0 (*Anaconda Documentation*, 2020) environments to manage bioinformatic packages on the RMACC Summit supercomputer managed jointly by Colorado State University and University of Colorado, Boulder. To process raw fastqs from the 181 individuals that underwent low-coverage whole genome sequencing, we used Trim Galore v0.6.7 (Krueger, 2012), a wrapper for cutadapt v1.18 (Martin, 2011), and FastQC v0.11.9 (Andrews, 2010) to trim any remaining Illumina adaptors in the fastqs. Next, based on recommendations for low coverage data generated with NovaSeq platforms (Lou & Therkildsen, 2022) we performed a sliding window cut of the 3' end of the reads to remove low quality tails, defined as 4 bases in a row with mean QUAL scores less than 20, using fastp v0.22.0 (Chen et al., 2018). We checked fastqs for quality using FastQC and MultiQC v1.0.dev0 (Ewels et al., 2016) before and after trimming reads.

After processing raw fastqs, we aligned samples to the Canada Warbler reference genome using Burrows-Wheeler Alignment software (bwa mem, bwa v0.7.17) (Li & Durbin, 2009). Then we added read group information using Picard v2.26.11 AddorReplaceReadGroups (*Picard Toolkit*, 2019) and marked duplicate reads using samtools v1.11 markdup (Danecek et al., 2021) before merging individuals with multiple bams. After merging bams, we checked sample coverage using bedtools v2.30.0 genomecov (Quinlan & Hall, 2010) and samples with less than 1X coverage were removed.

We used the processed bams to call variants using GATK v4.2.5.0 HaplotypeCaller (McKenna et al., 2010) and BCFtools v1.15.1 mpileup (Danecek et al., 2021). Then, we stringently filtered the variant sets using BCFtools, allowing only

biallelic sites, a minor allele frequency of greater than 5%, QUAL score of greater than 30 and less than 10% missing across the 169 individuals. We intersected the filtered variant sets from bedtools and GATK to create a high-quality variant set to use for base quality score recalibration according to GATK current best practices. Using the intersected variants, we recalibrated the sample bams using GATK BaseRecalibrator and ApplyBQSR. With the recalibrated bams, we used HaplotypeCaller to call a recalibrated set of variants. Then we filtered the recalibrated variant set allowing only biallelic sites, a minor allele frequency of less than 5%, QUAL score of greater than 30 and less than 20% missing data across the 169 individuals.

Using the recalibrated, filtered variant set we performed an exploratory analysis using R (R Core Team, 2022) and the package srsStuff (Anderson, 2020) to produce single-read sampling principal components analysis (PCA) of whole genome structure. Instead of population clustering, we found significant platform effects (for example of platform effects on low coverage data, see Lou & Therkildsen, 2022). We removed platform-associated variants from the dataset and proceeded with the analysis once samples no longer clustered in platform groups by PCA (for full methods to remove platform effects, see Supplemental Methods).

ESU identification

We used Waples' (1991) definition of an ESU, populations that are reproductively isolated and that represent an important evolutionarily or ecologically distinct part of the species. To identify ESUs we used the criteria set out in Turbek et al. 2023 as a guide to delineate where reproductive isolation exists in a species with high gene flow. We decided to delineate ESUs based on population structure that was supported with 2 out

of 3 complementary, but different, approaches to finding breaks in genetic variation across the breeding range: PCA for a model-free approach, ADMIXTURE for a hierarchical model, and EEMS to model potential barriers to gene flow. We investigated genome structure using the filtered variant set after removing platform effects. We first used single-read sampling PCA to ascertain if there was population genetic structure. We used single-read sample PCA as uneven coverage across samples can be misinterpreted as population structure in low-coverage data (Lou & Therkildsen, 2022).

As called genotypes on low coverage data are low confidence and often result in missing data for any given SNP, we imputed missing genotypes using Beagle v4.1 (Browning et al., 2018) using the genotype probabilities from GATK. Using the imputed data, we then removed linked SNPs using linkage disequilibrium ($r > 0.5$) in PLINK v2.0 (Purcell et al., 2007) and further investigated the potential for population structure using the program ADMIXTURE (Alexander et al., 2009). We used 5 runs of ADMIXTURE with K values 1-6 with the full set of variants but different random seeds. In order to visualize the different values of K and identify the most supported value of K based on cross validation we used the R package pophelper (Francis, 2017).

To further investigate if structuring within the PCA or ADMIXTURE was due to a subtle barrier to gene flow, we used estimated effective migration surfaces (EEMS) to check for potential barriers to gene flow (Petkova et al., 2016). Using the imputed dataset and 200 demes to test for potential barriers to gene flow we mapped the estimated migration rates on the breeding range to visualize barriers across the range.

AU identification

Similarly to ESUs, we used the criteria set out in Turbek et al. 2023 as a guide to define how to delineate AUs. We defined AUs as breaks in genetic variation across the breeding range using only the adaptive loci set with 2 complementary methods: PCA for a model-free approach and ADMIXTURE for a hierarchical model. To select environmental variables, we used gradient forest (Ellis et al., 2012), an extension of random forest (Liaw et al., 2002), and 23 environmental variables potentially important to Canada Warbler breeding ecology based on previous research (Supplemental Table 3, Fick & Hijmans, 2017; Reitsma et al., 2020). While gradientForest has been previously used to identify adaptive loci, we decided to use it to inform our environmental variable choice but not identify loci as neutral population structure can confound gradientForest (Láruson et al., 2022). Environmental data were extracted from each of 16 sampling locations, excluding two sampling sites with fewer than 4 individuals. We used ANGSD to calculate the allele frequency for the included sampled sites from genotype likelihood estimation using only SNPs with a minor allele frequency of greater than 5% and removed SNPs that had missing data for any sampled site. Gradient forest was run with the R package gradientForest (Ellis et al., 2012) on 5 different subsets of 50,000 random SNPs using the environmental variables as predictors for the genomic data (ntree = 500, nbin = 101, corr.threshold = 0.5). To ensure that the models inferred from the data explained more than could be expected by random chance, we used 100 different randomizations of the data to create random models. Using these random models we compared the distribution of randomized r-squared values of the SNPs to the r-squared values for the 5 models inferred from the data and ensured the models inferred from the data were above than the 95th percentile

of the r-squared from the random models. We then chose the top 4 uncorrelated ($|r| < 0.75$) variables ranked as most important to explaining genetic variation shared across the 5 models inferred from the data. These environmental variables were used as a reduced variable set for the rest of the analyses: mean temperature of the warmest quarter (BIO10), precipitation of the wettest month (BIO13), precipitation seasonality (BIO15), and tree cover.

To identify putatively adaptive loci, we used two approaches, redundancy analysis (RDA) and Latent Factor Mixed Models (LFMM). LFMM is a univariate approach that controls for population structure with latent factors, while RDA is a multivariate constrained ordination approach that performs better at finding many loci of small effect (Forester et al., 2018). To account for population structure in our RDA, we generated spatial variables using Moran's Eigenvector Maps (MEMs) (Dray et al., 2006) using the R package adespatial v0.3-16 (Guénard & Legendre, 2022). Then we ran the RDA using the R package vegan with individual genotypes as the response and the reduced environmental variable set as the predictors, conditioned on the MEMs to account for underlying population structure and geographic distance. We selected loci that were above 3 standard deviations away from the mean.

We then used LFMM to find putatively adaptive loci by an alternate method. To account for population structure in our LFMM, we used K=3 as there was subtle structure within the dataset that would not be accounted for with the most likely K=1. We ran LFMM using the R package lfmm (Jumentier, 2021) with individual genotypes as the response and, as LFMM is a univariate test, used the first principal component of a PCA of the reduced environmental predictors to reduce the need for multiple corrections due

to multiple tests. Using LFMM best practices, we adjusted an initial genomic inflation factor of >2.5 to 1.0 and identified loci using a false discovery rate of 1%.

Once loci were identified using both RDA and LFMM, the union of loci discovered by both methods was used as our set of candidate adaptive loci. To identify population genetic structure among the putatively adaptive loci we used PCA and ADMIXTURE. The resulting posterior probabilities of genetic group membership estimated from ADMIXTURE were visualized as transparency levels of different colors overlaid and clipped to a map of the Canada Warbler breeding range using the R packages SP, RGDAL, and RASTER (Bivand et al. 2013, 2017; Hijmans 2017) creating a spatially-explicit map of adaptive groups.

Testing for isolation by distance versus isolation by environment

To determine if the geographically relevant clustering in the PCA and ADMIXTURE plot was a result of geography or environment, or both, we used a combination of mantel tests, partial mantel tests, and redundancy analysis variance partitioning. We generated pairwise FST comparisons using all loci between sites with at least 4 individuals, excluding two sampling sites with fewer than 4 individuals, using ANGSD v0.935 (Korneliussen et al., 2014). We calculated the site allele frequency (SAF) likelihoods for each sampled site from genotype likelihood estimation using only SNPs with a minor allele frequency of greater than 5% and less than 30% missingness within the sampling site. We then calculated the 2D site frequency spectrum (SFS) for each pair of sites and, with the per-site SAF files as priors, we estimated pairwise FST between each sampled site. Using this pairwise FST, we linearized FST ($\frac{FST}{1-FST}$) values for each pairwise comparison. We then calculated pairwise Euclidean distance between

each site's latitude and longitude using the R package sp (Pebesma & Bivand, 2005). To determine if genetic variation was more closely linked to environment or geography, we extracted environmental values for the reduced environmental variable set for each site's latitude and longitude. Then we centered each environmental variable to control for differences in absolute values of each variable, then calculated a pairwise environmental distance using the R package stats (R Core Team, 2022). We tested for isolation by distance and isolation by environment with Mantel tests in the R package vegan v2.6-2 (Oksanen et al., 2022) and partial Mantel tests conditioned on environmental distance and geographic distance, respectively.

Genomic offset analysis

Using the adaptive loci found with LFMM and RDA, we ran gradient forest (Ellis et al., 2012) and the reduced environmental variable set to generate a model of allele frequency turnover across the breeding range. We used this model as a baseline to predict expected allele frequencies in 2060-2080 using predicted environmental change under Shared-Socioeconomic Pathways 126 and 585 at 100,000 random points throughout the breeding range. We calculated the 'genomic offset' between current allele frequencies and predicted future allele frequencies using a Euclidean distance (Bay et al., 2018), as a measure of how much genetic change may be necessary to maintain current adaptive patterns (Capblancq et al., 2020; Fitzpatrick & Keller, 2015). Given the inherent uncertainty in predicting if or where range shifts will occur (Sofaer et al., 2018), we did not predict potential gene-environment associations or genomic offset outside of the current breeding range. Using the spatially-explicit map of adaptive groups, we created shapefiles of each of the putative adaptive groups identified across

the breeding range. We then extracted genomic offset values inside the boundaries of each adaptive group shapefile and calculated the median genomic offset within each adaptive group. We also extracted the genomic offset values across the entire breeding range and calculated the median genomic offset.

Demographic analysis

We estimated relative population size indices and 1968-2019 population trends for each of the three AUs and for all AUs combined based on the hierarchical over-dispersed Poisson model of Sauer et al. (2011) applied to Breeding Bird Survey data (Pardieck et al., 2020). While there are alternative data sources for estimating trends in migratory birds (e.g. eBird), BBS data is one of the longest temporal datasets and it has been shown that as species range increases trend estimate differences between data sources tends to decrease (Horns et al., 2018). The fixed strata effects in the model were defined based on the AUs, with BBS sampling points, called routes, assigned to AUs if they ever had a Canada Warbler detection on the route and if the coordinates of the route centroid were contained within the AU polygon boundary. In addition, routes with Canada Warbler detections that were outside of AU polygons but within a 50-km buffer of an AU boundary were assigned to the nearest AU. Population size indices were derived by summarizing posterior distributions of mean route-level counts weighted by AU area and proportions of routes with Canada Warbler detections (Sauer & Link, 2011). We estimated population size indices for each AU by summarizing posterior distributions over the most recent 5 years (2015-2019). We also derived estimates of population size indices in the year 2069, assuming the current estimated trend and annual variance remain constant, based on posterior samples from the model

at a time point 50 years in the future. Long-term trends for each AU and for the overall population (based on the summed population indices across AUs) were estimated as the geometric mean of yearly changes in population size from 1968-2019 (Sauer & Link, 2011). We implemented the BBS model with JAGS 4.3.1 (Plummer, 2003) via the jagsUI (Kellner & Meredith, 2021) package in R (R Core Team, 2022). We assigned vague prior distributions for all model parameters and hyperparameters. Posterior distributions were derived from 40,000 simulated values of four chains from the posterior distribution after an adaptive phase of 20,000 iterations and burn-in of 10,000 samples of the Gibbs sampler and thinning by 3. Markov chains were determined to have successfully converged based on $\hat{R}$<1.1 for posterior estimates of all parameters (Gelman & Hill, 2007).

Results

Bioinformatic processing

From the 10X sequencing libraries for genome assembly, we received 485.36 million reads. We produced an assembly with a "raw" coverage of 48.6X, a scaffold N50 size of 7.51 Mb, and genome size of 1.03 Gbp as estimated by *Supernova* for 3.04 x 10^3 scaffolds greater than 10 kb. The genome assembly is deposited at NCBI with accession number PRJNA689308.

Of the 181 samples from 18 different sites across the breeding range in our initial dataset (Supplemental Figure 1), 12 samples with less than 1X average coverage were removed as part of quality filtering, leaving 169 total samples (Supplemental Table 2). The median number of samples per site was 7.5 (range 3-22), with an average depth of coverage of 2.6X (range 1-22X). After filtering out low quality SNPs and indels, we found 672,053 variants. After filtering for platform effects (Supplemental Figure 2), we retained 654,226 SNPs.

ESU identification

Using PCA, we found there was subtle population structure throughout the breeding range (Figure 1A). Groups in the far Northwest and South clustered away from each other, with groups in the Eastern portion of the range falling between them, though overall variation explained was low- 0.94% and 0.90% on PCs 1 and 2 respectively. We filtered out variants in linkage disequilibrium and retained 451,571 SNPs, then assessed population structure for values of K 1-6 using ADMIXTURE (Supplemental Figure 4). Results from ADMIXTURE suggest that the most supported K was 1.

Using EEMS with the whole-genome dataset we found that there is a strong barrier to gene flow between the Southern population and the Eastern populations in the Pennsylvania/New York region. (Figure 1B).

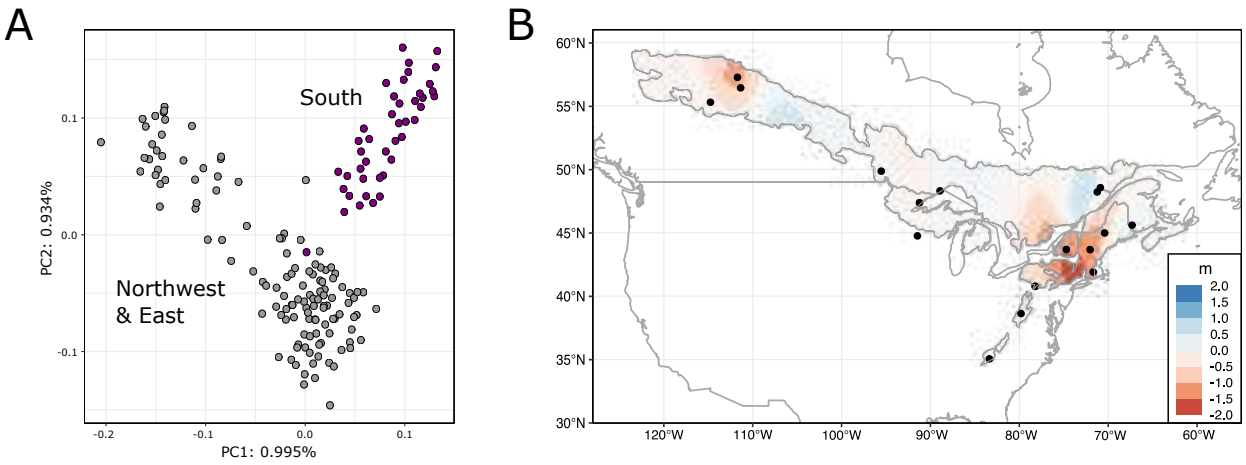


Figure 1: Population structure of Canada Warbler using whole-genome loci of 654,226 SNPs. A) Principal components analysis representing whole-genome structure. Northwestern and Eastern populations are in gray, while Southern populations are in purple. B) Estimated posterior mean migration rates on a log10 scale from EEMS. Areas with positive migration in blue are estimated to have greater gene flow than expected, while areas with negative migration in red are estimated to have less gene flow than expected. Transparency is scaled to reflect magnitude of estimated migration. Black outline reflects the breeding range.

AU identification

Using 5 random subsets of 50,000 SNPs in a gradient forest, we found more genetic variation was explained by the environment than by random chance. In all 5 cases, r-squared of the non-random model was above the 99th percentile of the randomized models (Supplemental Figure 5B). We selected the top four most highly ranked uncorrelated variables for a reduced set of variables to use in the rest of the analyses (Supplemental Figure 5A): mean temperature of the warmest quarter (BIO10),

precipitation of the wettest month (BIO13), precipitation seasonality (BIO15), and tree cover.

We found 4,832 SNPs associated with the environmental variables using a standard deviation of 3 using RDA (Figure 2). To complement the RDA approach, we used LFMM to identify putatively adaptive SNPs and found 9,212 SNPs associated with the PC1 environmental predictor using a false discovery rate of 1%. We took the unique SNPs found by both methods for a dataset of 11,441 SNPs.

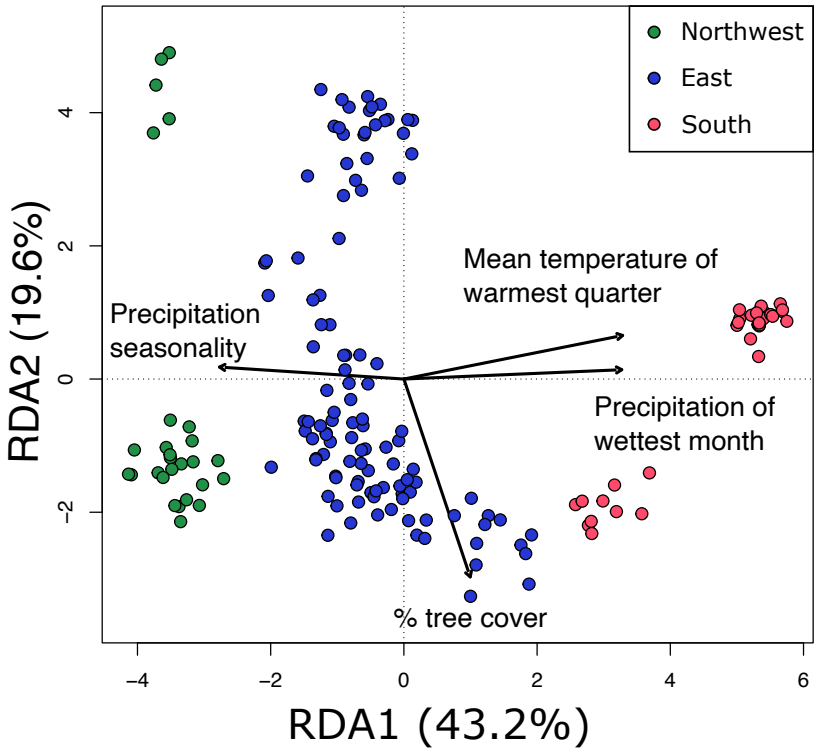


Figure 2: Principal component analysis of redundancy analysis axes 1 and 2. Colored points are individuals sampled with the Northwest AU in green, the Eastern AU in blue, and the Southern AU in pink. Arrows represent the magnitude and direction of environmental variables.

We used PCA with the putatively adaptive SNPs and found three potential clusters (Supplemental Figure 6). We used ADMIXTURE to assess population structure

for values of K 1 – 6 (Supplemental Figure 7) and found the best supported K was 3 (CV = 0.33822) for putatively adaptive loci (Figure 3A). We used the best supported K to assign individuals to putative AUs and mapped the regions onto the breeding ground in a spatially explicit map of adaptive variation (Figure 3B).

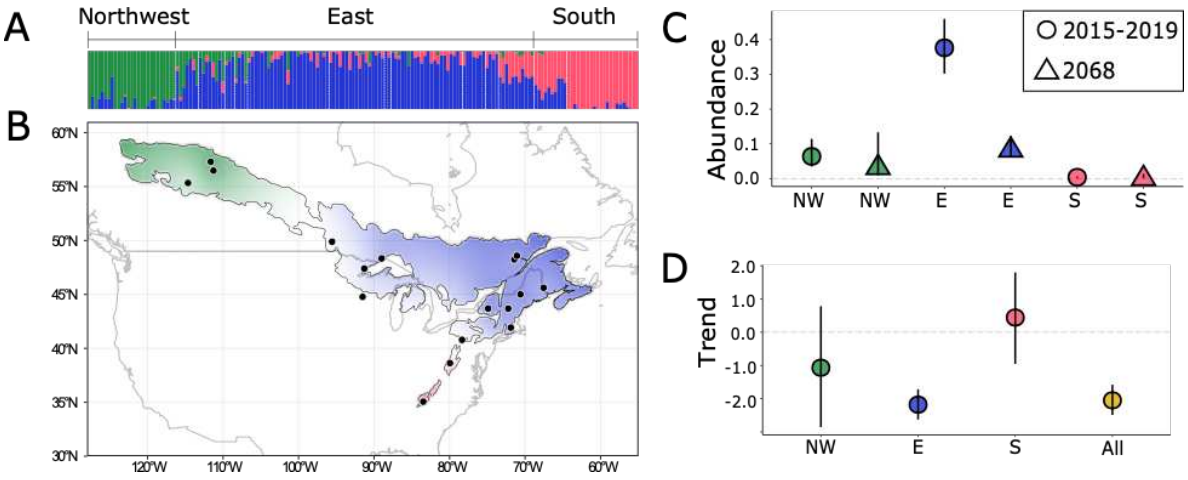


Figure 3: Putative adaptive units of Canada Warbler using adaptive loci. Colors are used to represent AUs, green for the Northwestern AU, blue for the Eastern AU, and pink for the Southern AU. A) Best supported ADMIXTURE plot of K of 3 using only putatively adaptive loci. B) Map of AU designations. Colors represent the AUs determined by ADMIXTURE groups, while points are sampled sites. Transparency is scaled to the predicted accuracy of assignment. C) Area-weighted median abundance estimates in each AU. Median estimates were calculated for 1968-2019 and a predicted median estimate for 2069. D) Estimated trend in area-weighted percent per year for each AU and range-wide calculated for 1968-2019.

Testing for isolation by distance or isolation by environment

Pairwise FST across all quality-filtered SNPs ranged from 0 to 0.02767 (Supplemental Table 4). Mantel tests revealed a strong correlation between environment and genetics (r = 0.5984, P = 0.001), as well as geography and genetics (r = 0.6699, P = 0.001). When we used a partial mantel test, the correlation between environment and genetics did not remain significant when accounting for geography (r =

0.2157, *P* = 0.1), but the correlation between geography and genetics remained significant when accounting for environment (*r* = 0.4256, *P* = 0.001).

Genomic offset analysis

Using the model of future climate under the emissions pathway in SSP585 that assumes the highest level of emissions pathways, genomic offset was predicted to be highest in the Northern-most sections of the breeding range (Figure 4A). When genomic offset was assessed by putative adaptive groups, the Northwestern group had the highest predicted genomic offset, followed by the Eastern and Southern groups, respectively (Figure 4B).

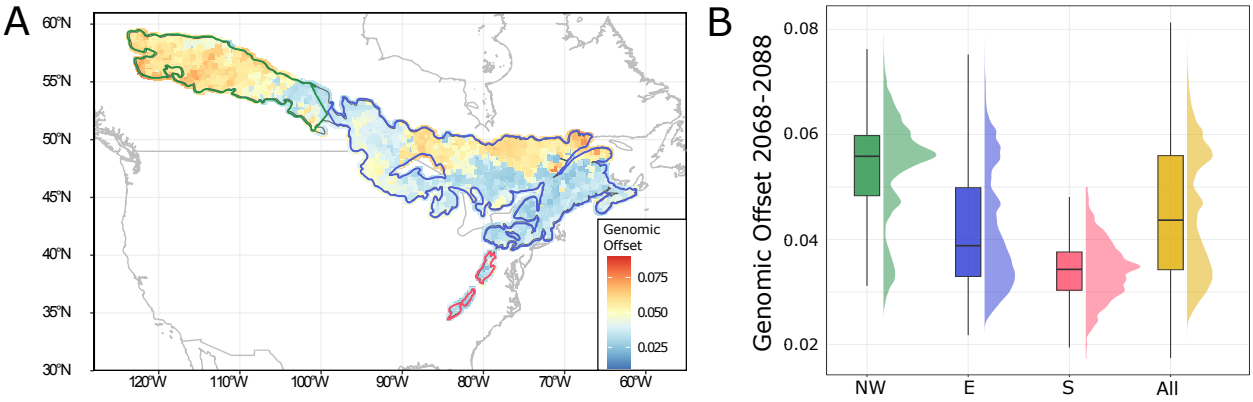


Figure 4: Predicted genomic offset across the Canada Warbler breeding range for 2068-2088 using SSP585. Colors are used to represent AUs, green for the Northwestern AU, blue for the Eastern AU, and pink for the Southern AU. A) Map of predicted genomic offset at 100,000 random points across the breeding range. Colored outlines represent the predicted AUs. B) Box plots and density curves of genomic offset values for each AU and the entire range.

Demographic analysis

We used 819 BBS routes from years 1968-2019 to estimate the breeding range had a declining trend of -2.05% per year (CI -2.49% – -1.58%). We split the breeding range into the 3 putative adaptive units, with 28 routes in the Northwestern AU, 748 routes in the Eastern AU, and 33 routes in the Southern AU. We found that the Eastern

AU had the highest area-weighted abundance of 0.3763 (CI 0.3020 – 0.4598), followed by the Northwest AU with 0.0637 (CI 0.0363 – 0.1141), and then the Southern AU with 0.0037 (CI 0.0022 – 0.0059) (Figure 2C) when using data from 2015-2019. Trends in abundance from 1968-2019 in the Southern AU (0.4416%; CI -0.9532% – 1.7961%) and Northwestern AU (-1.0682%; CI -2.8545% – 0.7831%) were not clearly positive or negative, but the Eastern AU (-2.1803%; CI -2.6276% – -1.7166%) had a strongly negative trend (Figure 3D). Predicted area-weighted abundance 50 years in the future in 2069 based on extending current trend estimates was highly variable for both the Southern AU (0.0048; CI 0.0017 – 0.0130) and Northwestern AU (0.0363; CI 0.0098 – 0.1332) making it unclear if there will be declines or increases, but the Eastern AU (0.0866; CI 0.0609 – 0.1222) was predicted to decline steeply (Figure 3C).

Discussion

While preserving genetic diversity is important for the maintenance of current and future adaptive potential, defining CUs in high dispersal species remains a challenge. Here, we demonstrate the value of a genomics-informed approach for identifying CUs in a highly mobile species by identifying CUs in the Canada Warbler, a migratory songbird with a large range and heterogeneous declines. We found that overall genomic differentiation in Canada Warblers was low, with little population structure across the breeding range. In contrast, population structure at putatively adaptive loci was associated with significant differences in abundance, trend, and potential vulnerability to a changing climate as determined by estimates of genomic offset in the future. Overall, our results point more generally to the conclusion that genomics-informed CUs provide a powerful tool for incorporating genetic diversity into management in a changing world.

ESU identification

Designating ESUs based on genomics can be challenging in species that remain highly connected through immigration and emigration across their range. ESU definitions have varied (reviewed in Fraser & Bernatchez, 2001), but using the criteria from the Waples' (1991) definition of an ESU, populations that are reproductively isolated and that represent an important evolutionarily or ecologically distinct part of the species can be considered ESUs. Here we use a method proposed by Turbek et al (2023) for defining CU's in highly mobile species and find that Canada Warblers fall into two ESUs. Specifically, population genomic analysis of all loci revealed two distinct groups on PCA, but ADMIXTURE had strongest support for a single ESU (K= 1) with

strong isolation by distance across most of the range. However, EEMS predicts a barrier to gene flow that exists between the two groups seen on PCA. This barrier to gene flow identified by EEMS in the Northeast was concordant with a previously documented habitat transition for Canada Warbler from lower elevation breeding sites in the North to higher elevation breeding sites found only in Southern Appalachian Mountains (Howell, 1910). From a species-specific perspective, these results suggest that while gene flow is high across the range, populations on either side of the identified barrier are ecologically different and have reduced gene flow compared to the rest of the range resulting in two ESUs by our criteria. More generally, these results illustrate the utility of programs like EEMS for identifying significant breakpoints in gene flow in species with high dispersal, even when isolation by distance appears to be the primary genetic pattern.

AU identification

AUs are a relatively recent addition to CU delineation, as it has only recently been possible to sequence the large amount of genetic data necessary to identify putatively adaptive loci (Funk et al., 2012). For conservation and management of species with high dispersal, the utility of AUs lies in finding groups that share adaptive differences that may not have strong genetic structure otherwise (de Guia & Saitoh, 2007; Whitlock, 2014). While neutral genetic diversity may change incrementally across the landscape, larger regions of similar environmental variables may lead to selection for different adaptive loci (Ackerman et al., 2013; Jackson et al., 2020; Vincent et al., 2013). Here we analyze population genetic structure at putatively adaptive loci and find support for three distinct AUs within the Canada warbler, a Northwest, an East, and a

Southern AU, with evidence of admixture between the three. Further, redundancy analyses revealed that genetic variation in Southern AU was associated with warmer mean temperatures, the Eastern AU was associated with higher amounts of precipitation during the wettest month, and the Northwestern AU was associated with high seasonality of precipitation. Because each AU is associated with distinct environmental parameters, identifying AUs on the breeding range provides a strong foundation for analyzing how past and future environmental change may influence population trends within and between ecologically distinct regions.

While it has been recently suggested that whole genome structure could be used as proxy for adaptive variation without the need to identify putatively adaptive loci (Fernandez-Fournier et al., 2021), here we find that analyzing population structure at putatively adaptive loci separately allowed us to identify AUs that may otherwise have been overlooked. Specifically, when all loci were analyzed together, the best supported K-value was one, but when putatively adaptive loci were analyzed separately, the best supported K-value was three. These results were robust to randomizing training and test sets, suggesting that they are not a result of ascertainment bias (Anderson, 2020). Overall, the difference in the population structure results between the all loci and adaptive loci analyses is likely because strong isolation by distance at neutral loci swamp signatures of population structure at adaptative loci when all loci are analyzed together. Overall, our results support the idea that in highly mobile species with high gene flow, putatively adaptive variation may be the strongest signal of genetic differentiation (Yeaman & Whitlock, 2011).

Identifying threats with genomics-informed CUs

The fact that highly mobile species like birds, bats, and fish often exhibit continuous genetic variation across space, but their exposure to anthropogenic threats is often highly discontinuous, has historically posed a challenge to their conservation and management (Kekkonen et al., 2011; Palumbi, 1994; Veith et al., 2004). Here we use genomics to identify CUs in the Canada Warbler and find that separating the species into ESUs and AUs provides a strong foundation for understanding past population declines and assessing vulnerability to future environmental threats. Overall, the identification of two ESUs with multiple AUs nested between them suggests that while gene flow between regions may help with recovery from regional population declines, all individuals will not have optimal fitness in all environments. This may be important if translocations are being considered as an option for recovery of declining populations as previous work has shown translocating individuals from locally adapted populations into dissimilar environments can result in poor fitness and failure to meet management goals (Frankham et al., 2011; Weeks et al., 2011). Beyond translocations, our analysis of past trends, current abundance, and future vulnerability to climate change within the ESU and AUs separately highlights the importance of considering both region and timescale when attempting to assess threat status across a species range.

The potential contrast between past, present, and future management priorities for the Canada Warbler is perhaps most apparent when we compare the potential threats to each AU separately. Overall, our analysis of past population declines across the ESU and by each AU separately revealed that while the single species ESU has been declining since 1968, declines in the Eastern AU have been most pronounced. In

particular, if the rate of past population declines continues into the future, then the Eastern AU is predicted to lose 77% of its current abundance by 2068 and thus may warrant high priority for conservation. In contrast, if we focus only on the current abundance and the existence of a genome-wide barrier to gene flow, then the isolated, lowest abundance Southern AU may warrant special attention. Finally, while the method is still being validated, genomic offset provides a metric for predicting future vulnerability to climate based on the mismatch between current and future gene-environment relationships (Capblancq et al., 2020; Rellstab et al., 2021). Based on this metric, the Northwest population is predicted to have the most trouble adapting to future climate change, with Eastern and then the Southern AU being decreasingly vulnerable, respectively. High vulnerability to climate change in the Northwest may be related to the predicted faster than average warming at Northern latitudes and shifting precipitation regimes (Newton et al., 2021; Rantanen et al., 2022). While ultimately management recommendations will depend upon the timescale being considered, one strategy for the Canada Warbler may be to focus on current declines (e.g. Eastern AU), while monitoring, or taking preventative action, in areas that have yet to decline steeply (e.g. Northwestern AU and Southern populations). Overall, our work demonstrates the utility of a genomics-informed CU approach for assessing past, present and future conservation threats across heterogeneous landscapes.

Conclusion

Here, we adopt a newly proposed framework for designating conservation units from Turbek, et al. 2023 and apply it to the Canada Warbler to delineate genomics-informed CUs. We identified that whole-genome structure in this highly mobile species was low and seemed to be driven primarily by adaptive variation. We identified that overall Canada Warblers could be considered a single ESU and have three putative AUs. In addition, we show that identifying multiple genomics-informed conservation units can reveal spatial variation in both current declines and climate vulnerability in a species with high dispersal. Using multiple conservation units may clarify what areas need protection, or need monitoring, to preserve genetic diversity in a highly mobile species with low overall genetic differentiation.

References

Ackerman, M. W., Templin, W. D., Seeb, J. E., & Seeb, L. W. (2013). Landscape heterogeneity and local adaptation define the spatial genetic structure of Pacific salmon in a pristine environment. *Conservation Genetics*, *14*(2), 483–498. https://doi.org/10.1007/s10592-012-0401-7

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*(9), 1655–1664. https://doi.org/10.1101/gr.094052.109

Anaconda documentation. (2020). Anaconda Software Distribution. https://docs.anaconda.com/

Anderson, E. C. (2020). *srsStuff: R package for inference by sampling single reads from sequencing data*. https://github.com/eriqande/srsStuff

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

Barbosa, S., Mestre, F., White, T. A., Paupério, J., Alves, P. C., & Searle, J. B. (2018). Integrative approaches to guide conservation decisions: Using genomics to define conservation units and functional corridors. *Molecular Ecology*, *27*(17), 3452–3465. https://doi.org/10.1111/mec.14806

Bay, R. A., Harrigan, R. J., Underwood, V. L., Gibbs, H. L., Smith, T. B., & Ruegg, K. (2018). Genomic signals of selection predict climate-driven population declines in a migratory bird. *Science*, *359*(6371), 83–86. https://doi.org/10.1126/science.aan4380

BirdLife International. (2021). *Cardellina canadensis. The IUCN Red List of Threatened Species*. https://dx.doi.org/10.2305/IUCN.UK.2021-3.RLTS.T22721882A137213211.en.

Browning, B. L., Zhou, Y., & Browning, S. R. (2018). A One-Penny Imputed Genome from Next-Generation Reference Panels. *The American Journal of Human Genetics*, *103*(3), 338–348. https://doi.org/10.1016/j.ajhg.2018.07.015

Capblancq, T., Fitzpatrick, M. C., Bay, R. A., Exposito-Alonso, M., & Keller, S. R. (2020). Genomic Prediction of (Mal)Adaptation Across Current and Future Climatic Landscapes. *Annual Review of Ecology, Evolution, and Systematics*, *51*(1), null. https://doi.org/10.1146/annurev-ecolsys-020720-042553

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, *34*(17), i884–i890. https://doi.org/10.1093/bioinformatics/bty560

Coates, D. J., Byrne, M., & Moritz, C. (2018). Genetic Diversity and Conservation Units: Dealing With the Species-Population Continuum in the Age of Genomics. *Frontiers in Ecology and Evolution*, *6*. https://doi.org/10.3389/fevo.2018.00165

COSEWIC. (2008). *COSEWIC assessment and status report on the Canada Warbler Wilsonia Canadensis in Canada. Committee on the Status of Endangered Wildlife in Canada* [Assessments;research]. https://www.canada.ca/en/environment-climate-change/services/species-risk-public-registry/cosewic-assessments-status-reports/canada-warbler.html

COSEWIC. (2015). *Guidelines for Recognizing Designatable Units*. https://www.cosewic.ca/index.php/en-ca/reports/preparing-status-reports/guidelines-recognizing-designatable-units.html

Dale, V. H., Joyce, L. A., McNulty, S., Neilson, R. P., Ayres, M. P., Flannigan, M. D., Hanson, P. J., Irland, L. C., Lugo, A. E., Peterson, C. J., Simberloff, D., Swanson, F. J., Stocks, B. J., & Wotton, B. M. (2001). Climate Change and Forest Disturbances: Climate change can affect forests by altering the frequency, intensity, duration, and timing of fire, drought, introduced species, insect and pathogen outbreaks, hurricanes, windstorms, ice storms, or landslides. *BioScience*, *51*(9), 723–734. https://doi.org/10.1641/0006-3568(2001)051[0723:CCAFD]2.0.CO;2

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., & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, *10*(2), giab008. https://doi.org/10.1093/gigascience/giab008

de Guia, A. P. O., & Saitoh, T. (2007). The gap between the concept and definitions in the Evolutionarily Significant Unit: The need to integrate neutral genetic variation and adaptive variation. *Ecological Research*, *22*(4), 604–612. https://doi.org/10.1007/s11284-006-0059-z

DeSaix, M. G., George, T. L., Seglund, A. E., Spellman, G. M., Zavaleta, E. S., & Ruegg, K. C. (2022). Forecasting climate change response in an alpine specialist songbird reveals the importance of considering novel climate. *Diversity and Distributions*, *28*(10), 2239–2254. https://doi.org/10.1111/ddi.13628

Dray, S., Legendre, P., & Peres-Neto, P. R. (2006). Spatial modelling: A comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecological Modelling*, *196*(3), 483–493. https://doi.org/10.1016/j.ecolmodel.2006.02.015

Ellis, N., Smith, S. J., & Pitcher, C. R. (2012). Gradient forests: Calculating importance gradients on physical predictors. *Ecology*, *93*(1), 156–168. https://doi.org/10.1890/11-0252.1

Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, *32*(19), 3047–3048. https://doi.org/10.1093/bioinformatics/btw354

Exposito-Alonso, M., Booker, T. R., Czech, L., Gillespie, L., Hateley, S., Kyriazis, C. C., Lang, P. L. M., Leventhal, L., Nogues-Bravo, D., Pagowski, V., Ruffley, M., Spence, J. P., Toro Arana, S. E., Weiß, C. L., & Zess, E. (2022). Genetic diversity loss in the Anthropocene. *Science*, *377*(6613), 1431–1435. https://doi.org/10.1126/science.abn5642

Fernandez-Fournier, P., Lewthwaite, J. M. M., & Mooers, A. Ø. (2021). Do We Need to Identify Adaptive Genetic Variation When Prioritizing Populations for Conservation? *Conservation Genetics*, *22*(2), 205–216. https://doi.org/10.1007/s10592-020-01327-w

Ferrari, B., Shamblin, B., Chandler, R., Tumas, H., Hache, S., Reitsma, L., & Nairn, C. (2018). Canada Warbler (Cardellina canadensis): Novel molecular markers and a preliminary analysis of genetic diversity and structure. *Avian Conservation and Ecology*, *13*(1). https://doi.org/10.5751/ACE-01176-130108

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, *37*(12), 4302–4315. https://doi.org/10.1002/joc.5086

Fitzpatrick, M. C., & Keller, S. R. (2015). Ecological genomics meets community-level modelling of biodiversity: Mapping the genomic landscape of current and future environmental adaptation. *Ecology Letters*, *18*(1), 1–16. https://doi.org/10.1111/ele.12376

Forester, B. R., Beever, E. A., Darst, C., Szymanski, J., & Funk, W. C. (2022). Linking evolutionary potential to extinction risk: Applications and future directions. *Frontiers in Ecology and the Environment*. https://doi.org/10.1002/fee.2552

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

Francis, R. M. (2017). pophelper: An R package and web app to analyse and visualize population structure. *Molecular Ecology Resources*, *17*(1), 27–32. https://doi.org/10.1111/1755-0998.12509

Frankham, R. (2003). Genetics and conservation biology. *Comptes Rendus Biologies*, *326*, 22–29. https://doi.org/10.1016/S1631-0691(03)00023-4

Frankham, R., Ballou, J. D., Eldridge, M. D. B., Lacy, R. C., Ralls, K., Dudash, M. R., & Fenster, C. B. (2011). Predicting the Probability of Outbreeding Depression. *Conservation Biology*, *25*(3), 465–475. https://doi.org/10.1111/j.1523-1739.2011.01662.x

Fraser, D. J., & Bernatchez, L. (2001). Adaptive evolutionary conservation: Towards a unified concept for defining conservation units. *Molecular Ecology*, *10*(12), 2741–2752. https://doi.org/10.1046/j.0962-1083.2001.01411.x

Funk, W. C., Lovich, R. E., Hohenlohe, P. A., Hofman, C. A., Morrison, S. A., Sillett, T. S., Ghalambor, C. K., Maldonado, J. E., Rick, T. C., Day, M. D., Polato, N. R., Fitzpatrick, S. W., Coonan, T. J., Crooks, K. R., Dillon, A., Garcelon, D. K., King, J. L., Boser, C. L., Gould, N., & Andelt, W. F. (2016). Adaptive divergence despite strong genetic drift: Genomic analysis of the evolutionary mechanisms causing genetic differentiation in the island fox (Urocyon littoralis). *Molecular Ecology*, *25*(10), 2176–2194. https://doi.org/10.1111/mec.13605

Funk, W. C., McKay, J. K., Hohenlohe, P. A., & Allendorf, F. W. (2012). Harnessing genomics for delineating conservation units. *Trends in Ecology & Evolution*, *27*(9), 489–496. https://doi.org/10.1016/j.tree.2012.05.012

Gelman, Andrew., & Hill, J. (2007). Data analysis using regression and multilevel/hierarchical models. In *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.

Grooten, M., & Almond, R. E. A. (2018). *Living Planet Report—2018: Aiming Higher.* World Wildlife Fund. https://www.worldwildlife.org/pages/living-planet-report-2018?_hsenc=p2ANqtz-9DJ-2Fh_gaupLvz8xfsFchP9DBn_QUFDUj0kJd-3saW9lSZzAIgyEZQXoQrOrOtXTDwB2IQtHoWcp_XPv8qmftM4T5FA

Guénard, G., & Legendre, P. (2022). Hierarchical Clustering with Contiguity Constraint in R. *Journal of Statistical Software*, *103*(7), 1–26. https://doi.org/10.18637/jss.v103.i07

Horns, J. J., Adler, F. R., & Şekercioğlu, Ç. H. (2018). Using opportunistic citizen science data to estimate avian population trends. *Biological Conservation*, *221*, 151–159. https://doi.org/10.1016/j.biocon.2018.02.027

Howell, A. H. (1910). Notes on the Summer Birds of Kentucky and Tennessee. *The Auk*, *27*(3), 295–304. https://doi.org/10.2307/4071313

Jackson, J. M., Pimsler, M. L., Oyen, K. J., Strange, J. P., Dillon, M. E., & Lozier, J. D. (2020). Local adaptation across a complex bioclimatic landscape in two montane bumble bee species. *Molecular Ecology*, *29*(5), 920–939. https://doi.org/10.1111/mec.15376

Jumentier, B. (2021). *lfmm: Latent Factor Mixed Models*. https://CRAN.R-project.org/package=lfmm

Kekkonen, J., Seppä, P., Hanski, I. K., Jensen, H., Väisänen, R. A., & Brommer, J. E. (2011). Low genetic differentiation in a sedentary bird: House sparrow population genetics in a contiguous landscape. *Heredity*, *106*(1), Article 1. https://doi.org/10.1038/hdy.2010.32

Kellner, K., & Meredith, M. (2021). *jagsUI: A Wrapper Around "rjags" to Streamline "JAGS" Analyses* (1.5.2). https://CRAN.R-project.org/package=jagsUI

Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics*, *15*(1), 356. https://doi.org/10.1186/s12859-014-0356-4

Krueger, F. (2012). *TrimGalore: A wrapper around Cutadapt and FastQC to consistently apply adapter and quality trimming to FastQ files, with extra functionality for RRBS data*. https://github.com/FelixKrueger/TrimGalore

Láruson, Á. J., Fitzpatrick, M. C., Keller, S. R., Haller, B. C., & Lotterhos, K. E. (2022). Seeing the forest for the trees: Assessing genetic offset predictions from gradient forest. *Evolutionary Applications*, *15*(3), 403–416. https://doi.org/10.1111/eva.13354

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Liaw, A., Wiener, M., & others. (2002). Classification and regression by randomForest. *R News*, *2*(3), 18–22.

Lou, R. N., & Therkildsen, N. O. (2022). Batch effects in population genomic studies with low-coverage whole genome sequencing data: Causes, detection and mitigation. *Molecular Ecology Resources*, *22*(5), 1678–1692. https://doi.org/10.1111/1755-0998.13559

Luikart, G., England, P. R., Tallmon, D., Jordan, S., & Taberlet, P. (2003). The power and promise of population genomics: From genotyping to genome typing. *Nature Reviews Genetics*, *4*(12), Article 12. https://doi.org/10.1038/nrg1226

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, *17*(1), Article 1. https://doi.org/10.14806/ej.17.1.200

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. https://doi.org/10.1101/gr.107524.110

Moritz, C. (1994). Defining 'Evolutionarily Significant Units' for conservation. *Trends in Ecology & Evolution*, *9*(10), 373–375. https://doi.org/10.1016/0169-5347(94)90057-4

Newton, B. W., Farjad, B., & Orwin, J. F. (2021). Spatial and Temporal Shifts in Historic and Future Temperature and Precipitation Patterns Related to Snow Accumulation and Melt Regimes in Alberta, Canada. *Water*, *13*(8), Article 8. https://doi.org/10.3390/w13081013

Oksanen, J., Simpson, G. L., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., Solymos, P., Stevens, M. H. H., Szoecs, E., Wagner, H., Barbour, M., Bedward, M., Bolker, B., Borcard, D., Carvalho, G., Chirico, M., Caceres, M. D., Durand, S., … Weedon, J. (2022). *vegan: Community Ecology Package*. https://CRAN.R-project.org/package=vegan

Paetkau, D. (1999). Using Genetics to Identify Intraspecific Conservation Units: A Critique of Current Methods. *Conservation Biology*, *13*(6), 1507–1509. https://doi.org/10.1046/j.1523-1739.1999.98507.x

Palumbi, S. R. (1994). Genetic Divergence, Reproductive Isolation, and Marine Speciation. *Annual Review of Ecology and Systematics*, *25*(1), 547–572. https://doi.org/10.1146/annurev.es.25.110194.002555

Pardieck, K. L., David, Z. J., Lutmerding, M., Aponte, V., & Hudson, M.-A. R. (2020). *North American Breeding Bird Survey Dataset 1966—2019, version 2019.0* [Data set]. U.S. Geological Survey. https://doi.org/10.5066/P9J6QUF6

Pebesma, E. J., & Bivand, R. S. (2005). Classes and methods for spatial data in R. *R News*, *5*(2), 9–13.

Petkova, D., Novembre, J., & Stephens, M. (2016). Visualizing spatial population structure with estimated effective migration surfaces. *Nature Genetics*, *48*(1), 94–100. https://doi.org/10.1038/ng.3464

Picard Toolkit. (2019). [Java]. Broad Institute. https://github.com/broadinstitute/picard (Original work published 2014)

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. *Working Papers*.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *American Journal of Human Genetics*, *81*(3), 559–575.

QIAGEN. (n.d.). *Kit ID: 69504*. Retrieved September 23, 2022, from https://www.qiagen.com/us/products/discovery-and-translational-research/dna-rna-purification/dna-purification/genomic-dna/dneasy-blood-and-tissue-kit/

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), 841–842. https://doi.org/10.1093/bioinformatics/btq033

R Core Team. (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.r-project.org/

Ralls, K., Ballou, J. D., Dudash, M. R., Eldridge, M. D. B., Fenster, C. B., Lacy, R. C., Sunnucks, P., & Frankham, R. (2018). Call for a Paradigm Shift in the Genetic Management of Fragmented Populations. *Conservation Letters*, *11*(2), e12412. https://doi.org/10.1111/conl.12412

Rantanen, M., Karpechko, A. Y., Lipponen, A., Nordling, K., Hyvärinen, O., Ruosteenoja, K., Vihma, T., & Laaksonen, A. (2022). The Arctic has warmed nearly four times faster than

the globe since 1979. *Communications Earth & Environment*, *3*(1), Article 1. https://doi.org/10.1038/s43247-022-00498-3

Reitsma, L. R., Hallworth, M. T., McMahon, M., & Conway, C. J. (2020). Canada Warbler (Cardellina canadensis), version 2.0. *Birds of the World*. https://doi.org/10.2173/bow.canwar.02

Rellstab, C. (2021). Genomics helps to predict maladaptation to climate change. *Nature Climate Change*, *11*(2), Article 2. https://doi.org/10.1038/s41558-020-00964-w

Rellstab, C., Dauphin, B., & Exposito-Alonso, M. (2021). Prospects and limitations of genomic offset in conservation management. *Evolutionary Applications*, *14*(5), 1202–1212. https://doi.org/10.1111/eva.13205

Ruegg, K., Bay, R. A., Anderson, E. C., Saracco, J. F., Harrigan, R. J., Whitfield, M., Paxton, E. H., & Smith, T. B. (2018). Ecological genomics predicts climate vulnerability in an endangered southwestern songbird. *Ecology Letters*, *21*(7), 1085–1096. https://doi.org/10.1111/ele.12977

Ruegg, K., & Turbek, S. (2022). Estimating global genetic diversity loss. *Science*, *377*(6613), 1384–1385. https://doi.org/10.1126/science.add0007

Ryder, O. A. (1986). Species conservation and systematics: The dilemma of subspecies. *Trends in Ecology & Evolution*, *1*(1), 9–10. https://doi.org/10.1016/0169-5347(86)90059-5

Sauer, J. R., & Link, W. A. (2011). Analysis of the North American Breeding Bird Survey Using Hierarchical Models. *The Auk*, *128*(1), 87–98. https://doi.org/10.1525/auk.2010.09220

Schweizer, T. M., DeSaix, M. G., & Ruegg, K. C. (2021). *LI-Seq: A Cost-Effective, Low Input DNA method for Whole Genome Library Preparation* [Preprint]. Genetics. https://doi.org/10.1101/2021.07.06.451326

Sofaer, H. R., Jarnevich, C. S., & Flather, C. H. (2018). Misleading prioritizations from modelling range shifts under climate change. *Global Ecology and Biogeography*, *27*(6), 658–666. https://doi.org/10.1111/geb.12726

Thurman, L. L., Stein, B. A., Beever, E. A., Foden, W., Geange, S. R., Green, N., Gross, J. E., Lawrence, D. J., LeDee, O., Olden, J. D., Thompson, L. M., & Young, B. E. (2020). Persist in place or shift in space? Evaluating the adaptive capacity of species to climate change. *Frontiers in Ecology and the Environment*, *18*(9), 520–528. https://doi.org/10.1002/fee.2253

USFW, & NMFS. (1996, February 7). *Policy Regarding the Recognition of Distinct Vertebrate Population Segments Under the Endangered Species Act*. Federal Register. https://www.federalregister.gov/documents/1996/02/07/96-2639/policy-regarding-the-recognition-of-distinct-vertebrate-population-segments-under-the-endangered

Veith, M., Beer, N., Kiefer, A., Johannesen, J., & Seitz, A. (2004). The role of swarming sites for maintaining gene flow in the brown long-eared bat (Plecotus auritus). *Heredity*, *93*(4), Article 4. https://doi.org/10.1038/sj.hdy.6800509

Vincent, B., Dionne, M., Kent, M. P., Lien, S., & Bernatchez, L. (2013). Landscape Genomics in Atlantic Salmon (salmo Salar): Searching for Gene–Environment Interactions Driving Local Adaptation. *Evolution*, *67*(12), 3469–3487. https://doi.org/10.1111/evo.12139

Waples, R. S. (1991). Pacific Salmon, Oncorhynchus spp., and the Definition of "Species" Under the Endangered Species Act. *Marine Fisheries Review*, *53*(3), Article 3.

Waples, R. S., Nammack, M., Cochrane, J. F., & Hutchings, J. A. (2013). A Tale of Two Acts: Endangered Species Listing Practices in Canada and the United States. *BioScience*, *63*(9), 723–734. https://doi.org/10.1525/bio.2013.63.9.8

Weeks, A. R., Sgro, C. M., Young, A. G., Frankham, R., Mitchell, N. J., Miller, K. A., Byrne, M., Coates, D. J., Eldridge, M. D. B., Sunnucks, P., Breed, M. F., James, E. A., & Hoffmann, A. A. (2011). Assessing the benefits and risks of translocations in changing environments: A genetic perspective. *Evolutionary Applications*, *4*(6), 709–725. https://doi.org/10.1111/j.1752-4571.2011.00192.x

Whitlock, R. (2014). Relationships between adaptive and neutral genetic diversity and ecological structure and functioning: A meta-analysis. *The Journal of Ecology*, *102*(4), 857–872. https://doi.org/10.1111/1365-2745.12240

Yeaman, S., & Whitlock, M. C. (2011). The Genetic Architecture of Adaptation Under Migration–Selection Balance. *Evolution*, *65*(7), 1897–1911. https://doi.org/10.1111/j.1558-5646.2011.01269.x

Appendix A: Supplemental Methods

To remove platform effects, we chose to remove variants in two steps. First, we used 17 samples that had >1X coverage on both platforms and split each sample into a HiSeq-derived bam and NovaSeq-derived bam. We used VCFtools v0.1.16 (Danecek et al., 2011) to calculate the pairwise FST between the variants found in the HiSeq-derived bams and the variants found in the NovaSeq-derived bams. We predicted since both groups were composed of the same samples on different platforms, the highest FST values would be associated with platform effects. We tested removing variants identified with FST values above the 80th, 85th, 90th, 95th, and 99th percentiles from the entire dataset. We identified the 85th percentile as the threshold where PCs 1 and 2 did not appear to have platform clustering (Supplemental Figure 2).

After our initial filtering, we found additional platform specific variants using redundancy analysis (Supplemental Figure 3). During our initial round of redundancy analysis, we used the environmental variables identified by gradient forest (see Methods) and included a variable for platform. We identified variants strongly associated with platform with a standard deviation of 3. We decided to lower the threshold to a lower standard deviation of 1.5 to find as many potential platform-associated variants as possible. We identified 138,921 variants and removed them from our dataset. Following removal, we re-ran the redundancy analysis using platform as a variable and confirmed platform-associated variants were no longer identified at a threshold standard deviation of 3. We then removed the variants associated with

platform from the entire dataset leaving 738,660 variants. We then quality filtered and continued with analysis.

Appendix B: Supplemental Tables

Supplemental Table 1: Sequencing scheme. Library preparation is noted on the left. Platform and number of times sequenced per platform is noted with gray shading.

Library preparation	HiSeq 2500	NovaSeq 4000	NovaSeq 4000	NovaSeq 4000
A	78			
A & B	32			
B		22		
B		14		
B		35		

Supplemental Table 2: Sites samples for low-coverage whole genome sequencing. Total number of sequenced samples, QC-passed sample, latitude, and longitude for each site. Estimated ESU, AU, and demographic group for each site.

Site #	Sampling location	Lat	Long	# sampled	# passed QC	ESU	AU
1	Fort MacKay, AB, Canada	55.285575	-114.770935	8	8	1	Northwest
2	Fort McMurray, AB, Canada	57.24806	-111.73444	14	14	1	Northwest
3	Slave Lake, AB, Canada	56.42	-111.37528	10	5	1	Northwest
4	Rennie, MB, Canada	49.854417	-95.549994	7	7	1	East
5	Finland, MN, USA	44.982778	-70.416389	8	8	1	East
6	Thunder Bay, ON, Canada	47.366667	-91.25	5	3	1	East
7	Eau Claire, WI, USA	45.594306	-67.325582	5	4	1	East
8	Laterrière, QC, Canada	35.06	-83.38	17	17	1	East
9	St-Fulgence, QC, Canada	43.67	-72.05	8	8	1	East
10	McAdam, NB, USA	43.67737443	-74.7249777	17	17	1	East
11	Dallas, ME, USA	48.301667	-88.935	3	2	1	East
12	Canaan, NH, USA	40.777297	-78.210148	22	22	1	East
13	Moose River, NY, USA	40.776047	-78.231172	7	7	1	East
14	Sprague Farm, RI, USA	48.21209882	-71.2426141	4	4	1	East
15	Bright Run, PA, USA	48.55242	-70.89314	5	5	1	East
16	Wolf Run, PA, USA	41.910554	-71.70924	6	6	1	East
17	Richwood, WV, USA	44.7525	-91.480833	13	10	1	South
18	Otto, NC, USA	38.628555	-79.825683	22	22	1	South
			Total	181	169		

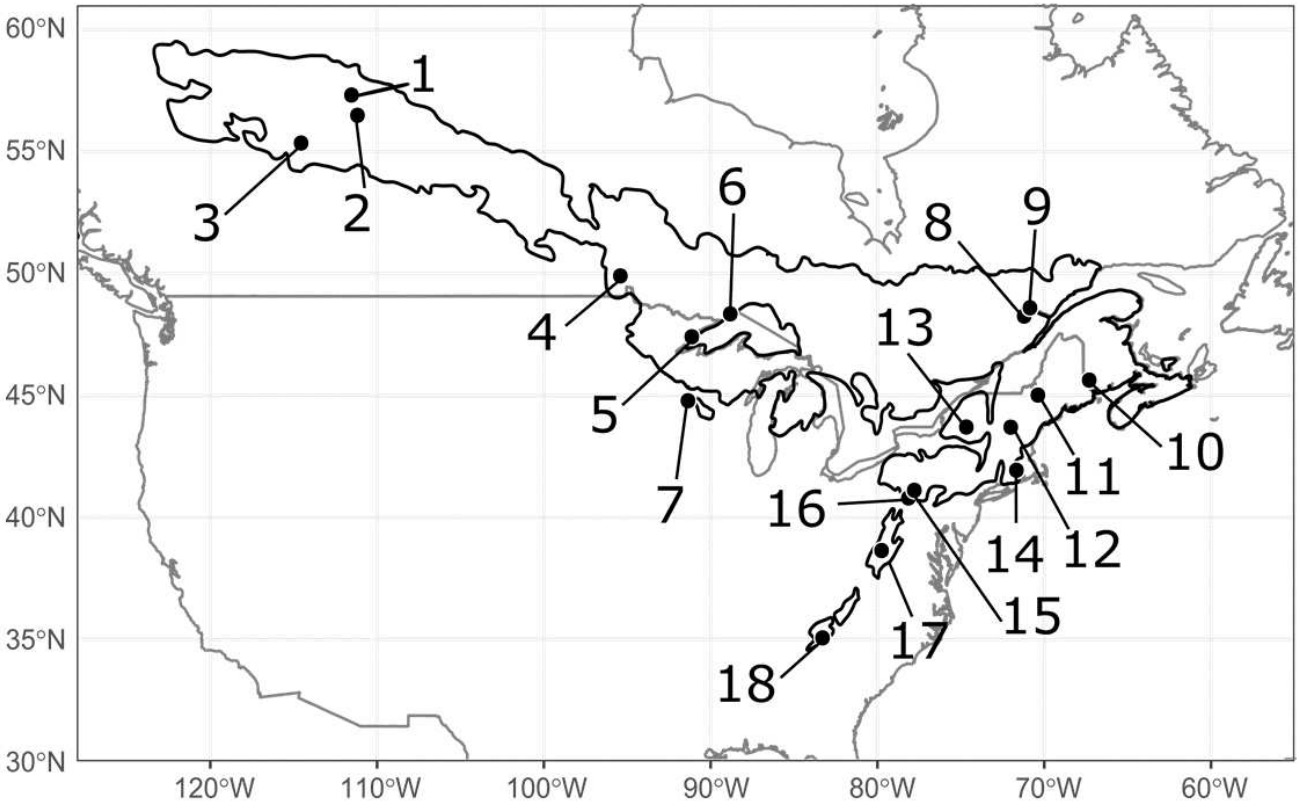
Supplemental Table 3: Raster names and environmental variable associated with raster for initial gradientForest.

Raster name	Environmental attribute
BIO1	Annual Mean Temperature
BIO2	Mean Diurnal Range
BIO3	Isothermality
BIO4	Temperature Seasonality
BIO5	Max Temperature of the Warmest Month
BIO6	Min Temperature of the Warmest Month
BIO7	Temperature Annual Range
BIO8	Mean Temperature of the Wettest Quarter
BIO9	Mean Temperature of the Driest Quarter
BIO10	Mean Temperature of the Warmest Quarter
BIO11	Mean Temperature of the Coldest Quarter
BIO12	Annual Precipitation
BIO13	Precipitation of the Wettest Month
BIO14	Precipitation of the Driest Month
BIO15	Precipitation Seasonality
BIO16	Precipitation of the Wettest Quarter
BIO17	Precipitation of the Driest Quarter
BIO18	Precipitation of the Warmest Quarter
BIO19	Precipitation of the Coldest Quarter
ndvi	Mean Normalized Difference Vegetation Index
ndvistd	Standard Deviation of Normalized Difference Vegetation Index
SRTM	Elevation
cawaHII	Human Influence Index
tree	Percent of tree cover

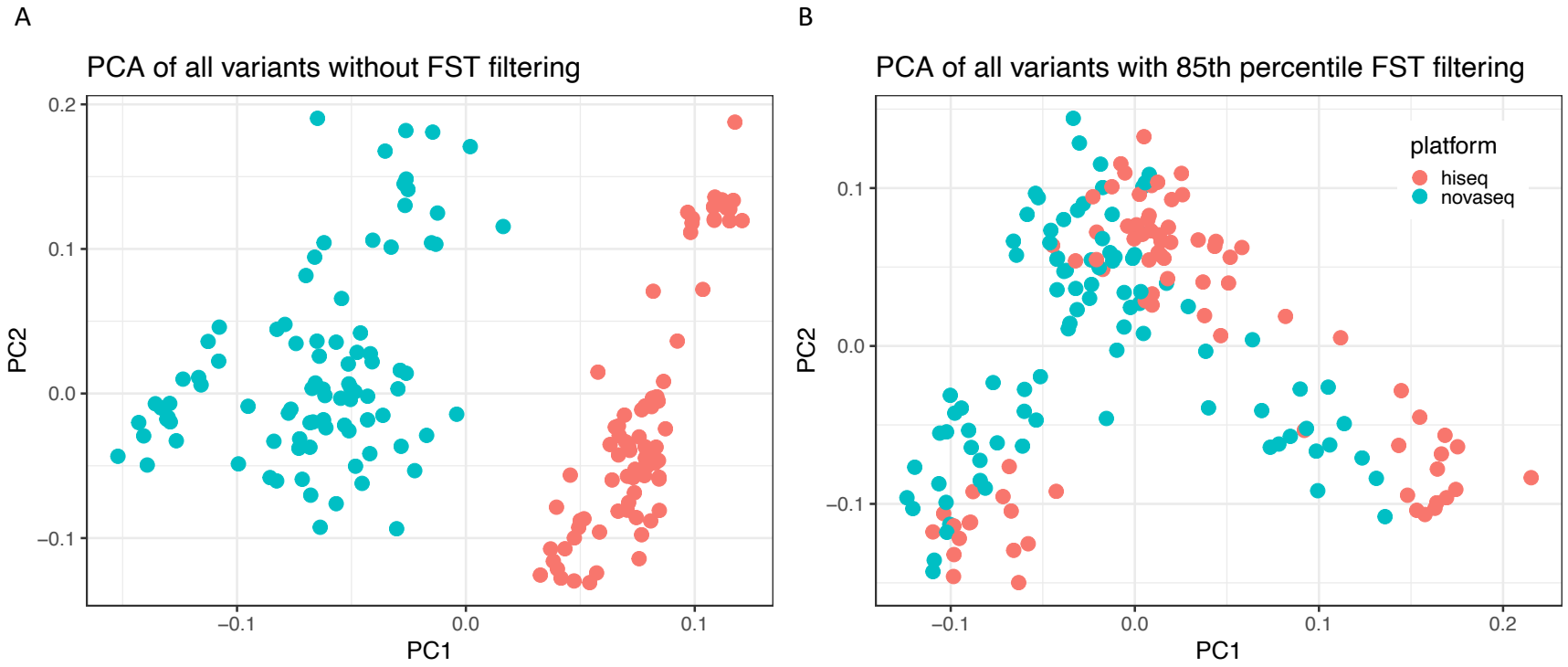
Supplemental Table 4: Pairwise F_{st} between all sampled sites with at least 4 samples passing QC. Thunder Bay, ON and Dallas, ME were excluded due to low numbers of samples passing QC.

	AB.FortMacKay	AB.FortMcMurray	AB.SlaveLake	MB.Rennie	MN.Finland	NB.McAdam	NC.Otto	NH.Canaan	NY.MooseRiver	PA.BrightRun	PA.WolfRun	QC.Laterriere	QC.StFulgence	RI.SpragueFarm	WI.EauClaire	WV.Richwood
AB.FortMacKay	0	0	0.005108	0.006259	0.004026	0.015394	0.020104	0.014726	0.0134	0.011279	0.017842	0.012713	0.009739	0.013203	0.007537	0.015925
AB.FortMcMurray		0	0.009515	0.007755	0.008048	0.017631	0.021249	0.014039	0.013791	0.015273	0.019668	0.013091	0.011961	0.017606	0.009058	0.021645
AB.SlaveLake			0	0.007569	0.008461	0.017477	0.02767	0.020399	0.020134	0.010516	0.017611	0.018663	0.014743	0.017855	0.003699	0.018948
MB.Rennie				0	0.000196	0.007373	0.017115	0.010175	0.0091	0.002893	0.005317	0.007523	0.004246	0.009963	0	0.011148
MN.Finland					0	0.004008	0.01339	0.005181	0.004212	0.003259	9.27E-03	2.62E-03	0	0.002378	0.00016	0.007824
NB.McAdam						0	0.018067	0.002867	0.004639	0.00631	0.009759	0.002045	0.000881	0.00526	0.005576	0.010129
NC.Otto							0	0.013843	0.014914	0.010293	0.015688	0.014683	0.012926	0.016738	0.018222	0.006327
NH.Canaan								0	0.001569	0.009585	0.014306	0.00027	0.001196	0.004181	0.010022	0.013285
NY.MooseRiver									0	0.010323	0.014883	0.001	0	0.002573	0.009547	0.015473
PA.BrightRun										0	0	0.009508	0.004697	0.010282	0.000613	0
PA.WolfRun											0	0.01371	0.009462	0.016318	0.001153	0.003596
QC.Laterriere												0	0.000411	0.003871	0.00747	0.014013
QC.StFulgence													0	0.002787	0.003805	0.008259
RI.SpragueFarm														0	0.007291	0.013522
WI.EauClaire															0	0.010193
WV.Richwood																0

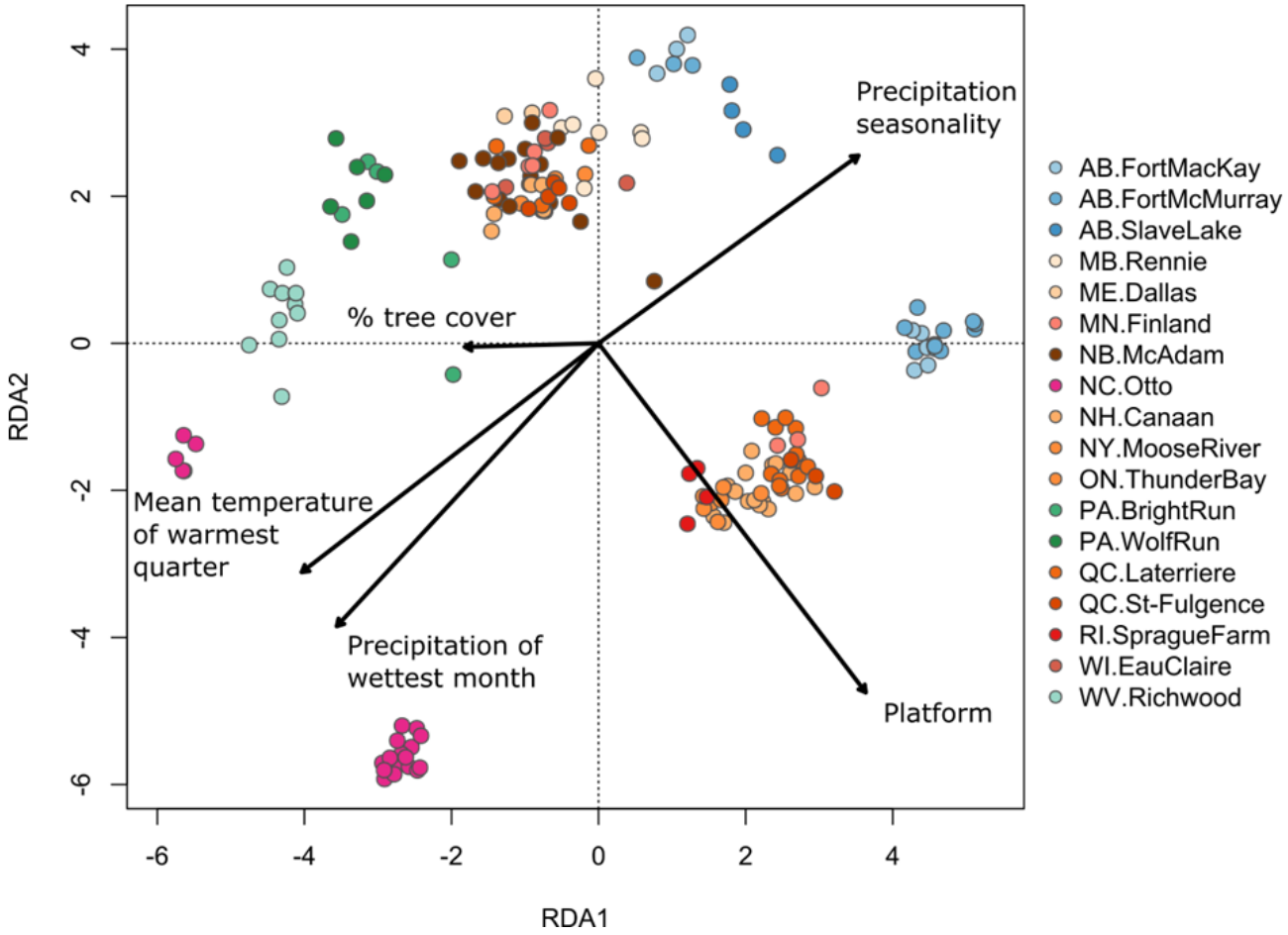
Appendix C: Supplemental Figures



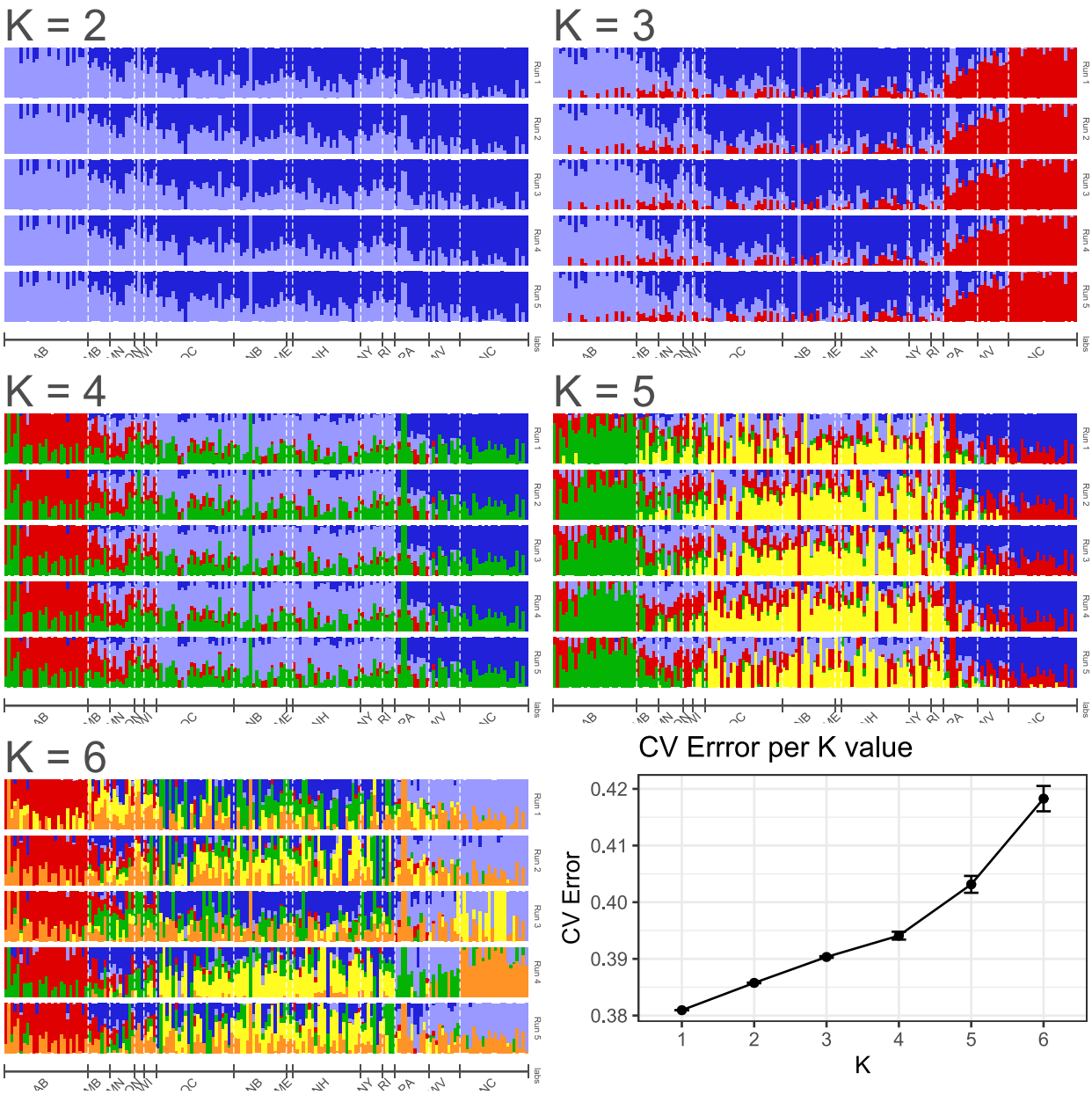
Supplemental Figure 1: Canada Warbler sampling map. Breeding range is outlined in black, sampled sites are denoted with black points. Each site has been numbered.



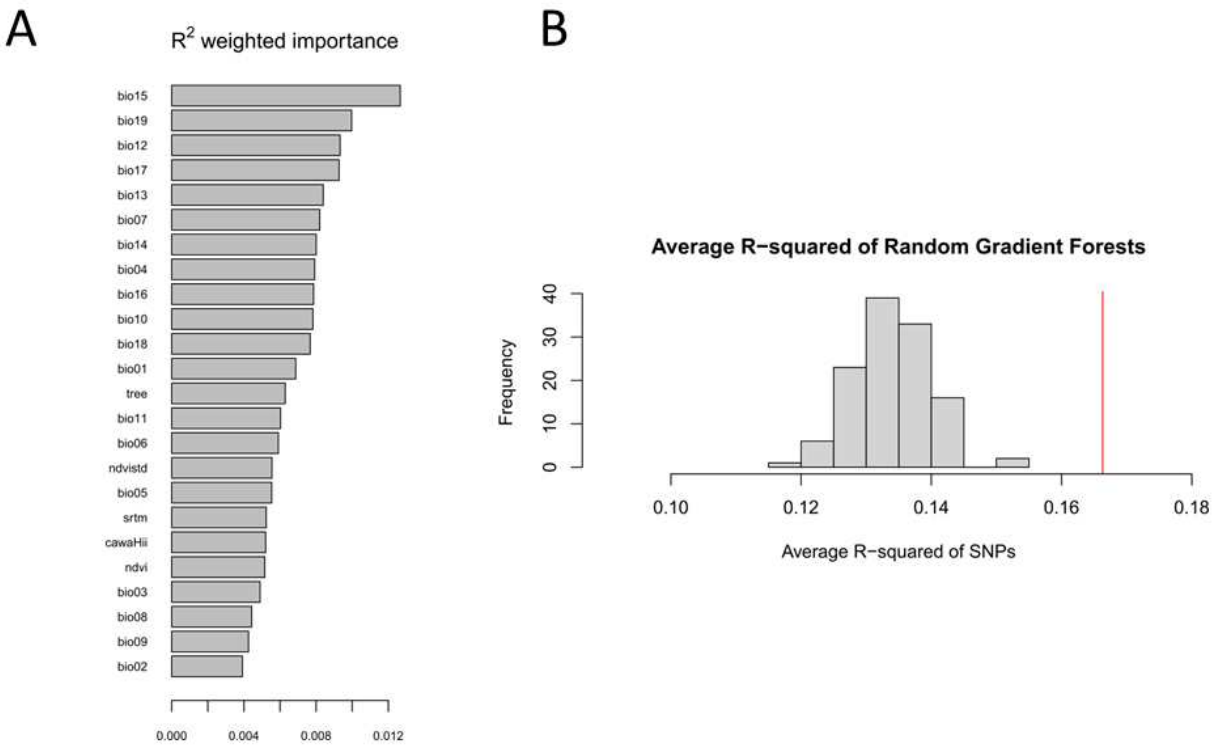
Supplemental Figure 2: Principal components analyses of whole-genome variation. A) PCA of initial variant set. Note the clustering based on platform on PC1. B) PCA after filtering variants at the 85th FST percentile.



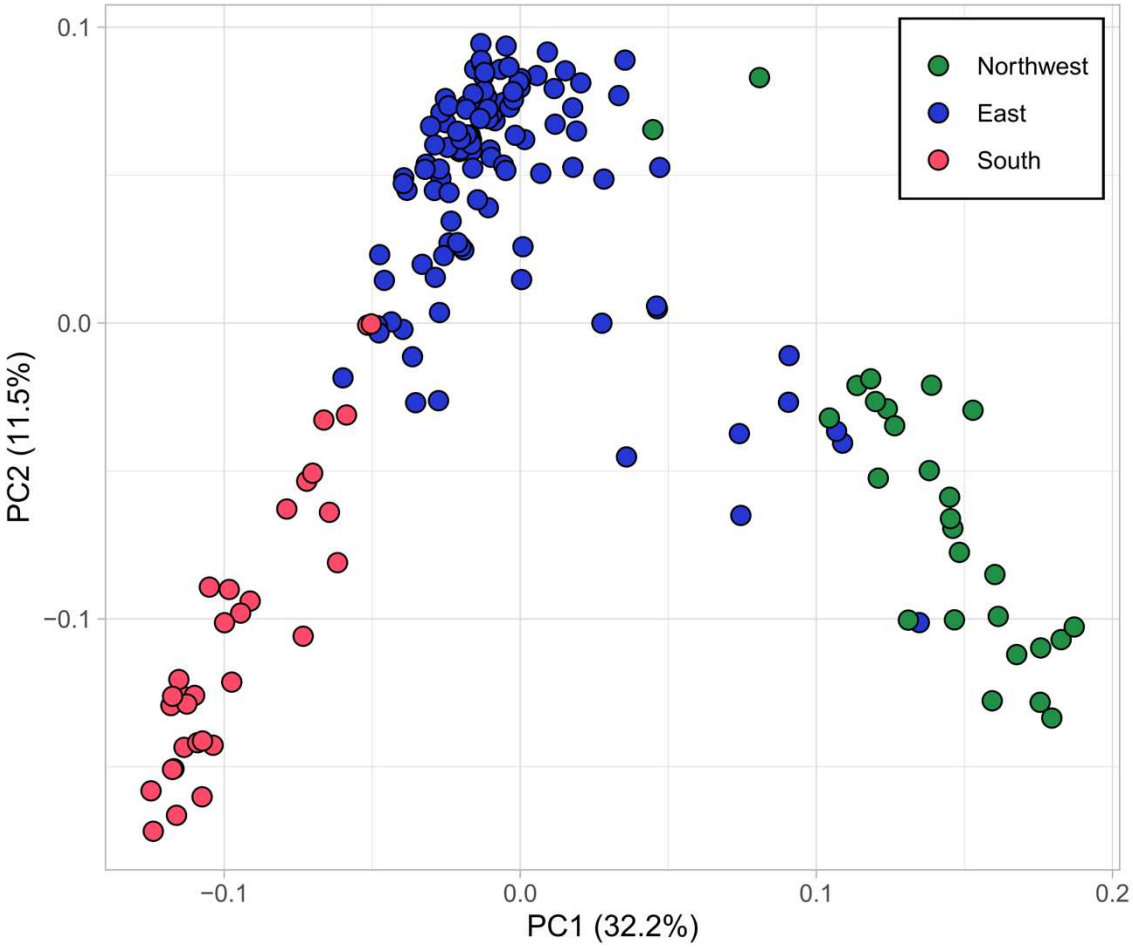
Supplemental Figure 3: Redundancy analysis to identify putatively adaptive loci using the environmental variables and platform. Notice the separation along the platform axis.



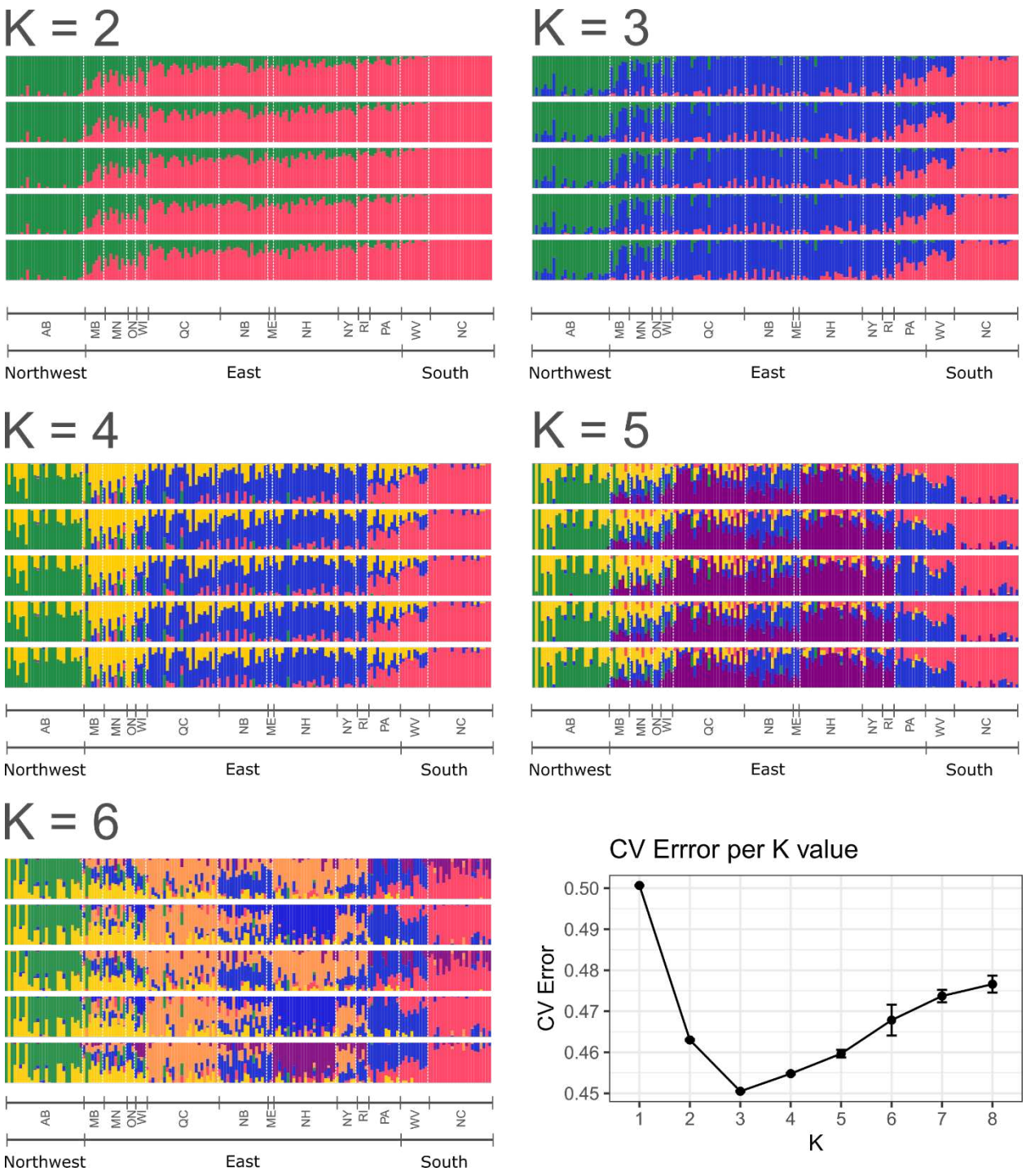
Supplemental Figure 4: ADMIXTURE plots of whole-genome loci (543,816 SNPs) for K's 2-6. Each K 1-6 was repeated in 5 runs with different random seeds. Cross validation error and standard error are plotted for K's 1-6. The lowest CV error is at K = 1.



Supplemental Figure 5: gradientForest models for selecting environmental variables. A) Ranked R^2 contribution of 23 environmental variables for the gradientforest model. B) Histogram of 100 randomized model R^2 values, non-randomized model noted with a red line.



Supplemental Figure 6: Principal components analysis of putatively adaptive loci (11,441 SNPs).



Supplemental Figure 7: ADMIXTURE plots of putatively adaptive loci (11,441 SNPs) for K 1-6. Each K 1-6 was repeated in 5 runs with different random seeds. Cross validation error and standard error are plotted for K's 1-6. The lowest CV error is at K = 3.