Article Type: Letters

Ecological Genomics Predicts Climate Vulnerability in an Endangered Southwestern Songbird

Kristen Ruegg[1, 2*], Rachael A. Bay[1,3*], Eric C. Anderson[3], James F. Saracco[4], Ryan J. Harrigan[1], Mary Whitfield[5], Eben H. Paxton[6], Thomas B. Smith[1,7]

* *Authors contributed equally.*

[1] Center for Tropical Research, Institute for the Environment and Sustainability, University of California Los Angeles, Los Angeles, CA 90095

[2] Department of Evolution and Ecology, University of California Santa Cruz, Santa Cruz, CA 95060

[3]Department of Evolution and Ecology, University of California Davis, One Shields Ave, Davis, CA 95616

[3] Southwest Fisheries Science Center, National Marine Fisheries Service, 110 Shaffer Road, Santa Cruz, CA 95060, USA

[4] The Institute for Bird Populations, PO Box 1346, Point Reyes Station, CA 94956, USA

[5] Southern Sierra Research Station, P.O. Box 1316, Weldon, CA, 932883, USA

[6] U.S. Geological Survey Pacific Island Ecosystems Research Center, Hawaii Volcano National Park, HI 96718

[7] Department of Ecology and Evolutionary Biology, University of California, Los Angeles, 621 Charles E. Young Drive South, Los Angeles, California 90095, USA.

Corresponding author: Kristen Ruegg
Department of Evolution and Ecology
University of California Santa Cruz
Santa Cruz, CA 95060
kruegg@ucsc.edu

Running title: Climate vulnerability in an endangered songbird

Keywords: local adaptation, climate change, genomic vulnerability, ecological genomics

Type of Article: Letter

Number of words in Abstract: 150

Number of words in main body: 5003

Number of references: 62

Number of Figs: 4

Number of tables: 2

Number of boxes: 0

Author Contributions: K.R., R.A.B., and T.B.S. conceived of the study; R.A.B. assembled and annotated the genome; R.B., K.R., E.C.A., J.F.S., and R.J.H. contributed to the population genetic, BBS and landscape genetic analyses; M.W. and E.H.P. contributed samples and biological expertise; K.R. wrote the paper with contribution from all authors.

Data Accessibility: The Willow flycatcher genome and annotations are available through DRYAD [DRYAD NUMBER] and population-level RAD-Seq data are available through NCBI's Sequence Read Archive [SRA NUMBER].

Few regions have been more severely impacted by climate change in the United States than the Desert Southwest. Here we use ecological genomics to assess the potential for adaptation to rising global temperatures in a widespread songbird, the willow flycatcher (*Empidonax traillii*), and find the endangered desert southwestern subspecies (*E. t. extimus*) most vulnerable to future climate change. Highly significant correlations between present abundance and estimates of genomic vulnerability – the mismatch between current and predicted future genotype-environment relationships – indicate small, fragmented populations of the Southwestern willow flycatcher will have to adapt most to keep pace with climate change. Links between climate-associated genotypes and genes important to thermal tolerance in birds provide a potential mechanism for adaptation to temperature extremes. Our results demonstrate that the incorporation of genotype-environment relationships into landscape-scale models of climate vulnerability facilitates more precise predictions of climate impacts that can guide conservation in threatened and endangered groups.

Introduction

The effects of climate change on biodiversity are forecast to be one of the leading causes of extinction over the next century (Dawson *et al.* 2011; Warren *et al.* 2013; Pacifici *et al.* 2015; Urban 2015). Evidence of climate-induced local extinctions are now widespread among plant and animal species (Sinervo *et al.* 2010; Wiens 2016) and the velocity of climate change impacts in desert biomes is predicted to be among the fastest (Loarie *et al.* 2009). Recent climate change has altered community

composition by favoring generalist taxa over habitat specialists and rare species (Menéndez *et al.* 2006; Estrada *et al.* 2016), but the ability to measure climate impacts below the species level is often lacking. Fine-scale estimates of vulnerability to climate change require an understanding of both the capacity for populations to shift their ranges to track climate conditions, as well as their capacity to tolerate climate alterations *in situ* via phenotypic plasticity or adaptation. Despite the fact that intraspecific variation in climate tolerances may factor critically in the ability of species to move or adapt to environmental change, most modeling efforts ignore local adaptation. However, genomic tools are facilitating assessments of local adaptation in non-model species with increasing reliability (Savolainen *et al.* 2013) and such information can be used to improve climate vulnerability estimates. Here we combine genome-wide sequencing with environmental data to improve predictions of how genotype-environment relationships may be disrupted by future environmental change in an endangered songbird native to the Desert Southwest of the United States, the Southwestern willow flycatcher.

Until recently, assessing species vulnerability to climate change focused largely on using current range-climate associations to predict distributions under models of future climate (Parmesan & Yohe 2003; Pacifici *et al.* 2015). However, complex biotic interactions (competition, specialization, coevolution, etc.) and or limits to dispersal imposed by physical barriers may limit range shifts, making it important to understand a species' potential to adapt to climate change *in situ* (Williams *et al.* 2008). Methodologies in the field of ecological genomics have provided tools to help incorporate information on local adaptation into climate vulnerability models by identifying regions where climate-induced selective pressure will be highest (Fitzpatrick & Keller 2015), but such methods have yet to be widely implemented. These approaches calculate the difference between current genotype-environment relationships and those predicted under future climate change to identify the geographic regions of greatest mismatch. More specifically, they can be used to ask, "How much would allele frequencies across the range have to change to keep pace with projected changes in climate?". In the absence of a range shift, populations in regions where the mismatch is greatest will either need to adapt or may suffer population declines, as was recently shown in the North American songbird, the Yellow warbler (*Setophaga petechia*) (Bay *et al.* 2018).

Few regions in North America will be more severely impacted by temperature extremes than the desert Southwest (Diffenbaugh *et al.* 2008; Hsiang *et al.* 2017). While most large-scale analyses of climate impacts in birds have focused on changes in geographic ranges or shifts in migratory phenology to better synchronize arrival times with earlier spring onset (Both & Visser 2001; Both *et al.* 2006; Stephens *et al.* 2016), these changes will do little to offset the impact of summer heat waves in desert regions. Recent

work suggests that small desert passerines, in particular, will experience higher rates of mortality due to dehydration and hyperthermia as the frequency of extreme temperature events increases (Albright *et al.* 2017). In addition, work in poultry has shown that high temperatures can cause heart strain, or in some cases heart failure, as birds attempt to dissipate heat through increased blood circulation. Further, this work has shown that such stress is not just physiological in nature, but is associated with differential expression in a suite of ~300 genes (Zhang *et al.* 2017). Based on these studies, we predict that genes important to thermal cooling will be under strong selection in small desert passerines as the frequency of heat waves increases.

The endangered Southwestern willow flycatcher provides an example of a desert passerine for which a better understanding of climate vulnerability has important implications for its conservation. This desert subspecies is one of four subspecies within the willow flycatcher whose combined ranges span the continental United States (Fig. 1; Pacific Northwestern form, *E. t. brewsteri*; Western Central form, *E. t. adastus*; and Eastern form, *E. t. traillii*). The presence of the Southwestern willow flycatcher in particular is associated with riparian woodlands along streams and waterways (Sedgwick 2000) and such habitats are thought to provide important refuges from temperature extremes (Chen 1999; McLeod *et al.* 2008). At the turn of the century, the Southwestern willow flycatcher was described as common wherever its specialized habitat existed (Grinnell & Miller 1944), but by 1995 when it was listed under the Endangered Species Act, the number of known breeding pairs had been reduced to between 300 and 500 (Unitt 1987; Sogge *et al.* 1997). Population declines have been attributed to loss of riparian habitats in the Southwest following dam-building, water diversions, groundwater pumping, urbanization, agricultural development, and livestock grazing (Service 2002), but the role that climate change may have played in declines is unknown. Some researchers have questioned the subspecies designation of the Southwestern willow flycatcher, suggesting that it is a peripheral population of an otherwise widespread species with no evidence for ecological distinctiveness (Zink 2015), although this suggestion has been questioned (Theimer *et al.* 2016). Here we use ecological genomics to investigate the potential for ecological distinctiveness within the willow flycatcher as well as the potential role of rising global temperatures on its future persistence.

To investigate potential genomic signals of local adaptation in the willow flycatcher, we tested for significant genotype-environment correlations using 105,000 SNP markers from 219 individuals spanning 24 populations across the breeding range (Fig. 1; Table 1). To identify the genomic locations of climate-associated SNPs in relation to genes and gene regions important to adaptation under climate change, we also assembled and annotated the first willow flycatcher genome. Significant genotype-environment

correlations for a subset of loci were further validated by genotyping an additional 274 individuals spanning 25 populations. To identify geographical regions where climate-induced selective pressure is predicted to be greatest under future climate change, we used gradient forest modeling and calculated an index of genomic vulnerability - defined as the mismatch between current and future genotype-environment relationships. We compared genomic vulnerability across the four subspecies and examined relationships between abundance and genomic vulnerability in order to understand which geographical regions will be most severely impacted by climate-induced selective pressure.

Materials and Methods

Sample Collection and DNA extraction

We compiled a collection of 493 willow flycatcher blood or tissue samples from 41 locations across the breeding range using a combination of samples from previous studies, museum donations and new field collections (Paxton 2000). 219 individuals from 24 populations were used to test for genome-wide genotype-environment correlations, while 274 individuals spanning 25 populations (8 replicate and 17 new populations) were used to validate a subset of significant genotype-environment correlations identified in the genome-wide analysis (N_{total_indiv} = 493; N_{total_pops} = 41; Table 1; Fig. 1). The willow flycatcher range map and associated subspecies boundaries was taken from the most current United States Geological Survey map used for willow flycatcher surveys (Sogge *et al.* 1997). Samples within one degree latitude and longitude and with no more than 10% difference in any environmental variable (as indicated by our environmental analysis, below) were lumped into a single population. DNA was purified using the Qiagen™ DNeasy Blood and Tissue extraction kit and quantified using the Qubit® dsDNA HS Assay kit (Thermo Fisher Scientific).

Genome Sequencing, Assembly, and Annotation

The genomic DNA library was created using a single Southwestern willow flycatcher individual from Roosevelt Lake, AZ and the Illumina TruSeq DNA PCR-Free LT kit (Illumina), with adjustments. One ug of DNA was diluted in 100 ul of AE buffer and fragmented to an average insert size of ~400bp. The resulting library was sequenced on two lanes of an Illumina HiSeq2500 using 250bp paired-end sequencing at the QB3 Vincent J. Coates Genomics Sequencing Laboratory, UC Berkeley. Two mate-pair libraries were also created, using 4kb and 8kb inserts and sequenced on one-third of a 100bp paired-end Illumina HiSeq 2500 lane at the Huntsman Cancer Center at the University of Utah. The 250bp paired

end reads were used to assemble contigs with the Discovar DeNovo assembler from the Broad Institute (http://www.broadinstitute.org), discarding contigs less than 1000 bp in length. Mate pair reads were trimmed and separated from paired end reads using NxTrim (O'Connell *et al.* 2015) and contigs were scaffolded with SSPACE (overlap requirement k=3) (Boetzer *et al.* 2010) using both paired end and mate pair libraries. We used reapr (Hunt *et al.* 2013) and mapping of the 8kb insert library to break the assembly at likely error regions. SSPACE scaffolding was repeated with k=5 and scaffolds <5kbp were discarded for the final assembly.

For annotation purposes, repetitive regions were replaced with N's using RepeatMasker (-species birds) (Tarailo-Graovac & Chen 2009). For annotation, we used two different *ab initio* gene predictions within the MAKER pipeline (Cantarel *et al.* 2008): SNAP and AUGUSTUS. SNAP was trained iteratively using Zebra Finch cDNA and protein sequences downloaded from Ensembl and AUGUSTUS was run using the available chicken training dataset. We used Interproscan (Zdobnov & Apweiler 2001) to add Pfam protein annotation and gene ontology (GO) terms and identified 15,489 genes. Scaffolds were aligned to the Zebra Finch genome (version 3.2.4) using the software promer, part of the MUMmer package (Delcher *et al.* 2003). After alignment, we retained the longest consistent alignment (-q) for each chromosome while filtering for similarity (-i 50) and alignment length (-l 500). We then determined the location of the longest alignment for each scaffold and ordered scaffolds accordingly for visualization purposes.

SNP Discovery and SNP Filtering

Genome scans on 219 individuals were conducted following the BestRAD library preparation protocol with some modifications (Ali *et al.* 2016). 100ng of DNA was digested using the SbfI restriction enzyme (New England Biolabs, NEB) and fragments were ligated with SbfI adapters prepared with biotinylated ends. Adapter-ligated samples were pooled and cleaned using 1X Agencourt® AMPure XP beads (Beckman Coulter). All DNA fragments were sheared to an average length of 400bp and adapter-ligated fragments were bound to M-280 streptavidin magnetic Dynabeads (Life Technologies). Blunt end repair and ligation of NEBNext Adapters was performed using the Illumina NEBNext Ultra DNA Library Prep Kit (New England Biolabs, NEB) and Agencourt® AMPure XP beads (Beckman Coulter) were used to size select an average of 500bp fragments. The final library was cleaned and run on a Bioanalyzer at the UCLA Technology Center for Genomics & Bioinformatics to check for the size distribution and the absence of contaminants. Two libraries, each comprised of 96 individuals, were initially sequenced in two lanes of 100bp paired end reads on an Illumina HiSeq 2500 at the UC Davis Genome Center. In a

third lane, 69 individuals with low coverage from the first two libraries were re-sequenced and an additional 27 individuals were sequenced.

The program Stacks (Catchen *et al.* 2013) was used to demultiplex, filter and trim adapters from the data with the process_radtags function and to remove duplicate read pairs using the clone_filter function. Reads were mapped to our genome assembly using bowtie2 (Langmead & Salzberg 2012) and the Haplotype Caller in the Genome Analysis Toolkit was used to identify single nucleotide polymorphisms (SNPs), following best practices from the Broad Institute (http://www.broadinstitute.org). Finally, we discarded low quality and rare variants (genotype quality<30; depth<8; minor allele frequency<0.01), as well as indels and non-biallelic SNPs using vcftools (Danecek *et al.* 2011). We used the R package genoscapeRtools (DOI: 10.5281/zenodo.848279) to visualize the tradeoff between discarding SNPs with low coverage and discarding individuals with missing genotypes in order to determine the final number of SNPs and individuals retained (SI Fig. 1).

Environmental data

For each sampling location, we obtained environmental data from publicly available databases. These 25 variables included 19 climate variables downloaded from WorldClim (Hijmans *et al.* 2005) which represented average climate between the years 1960-1990, as well as vegetation indices (Carroll *et al.* 2004) (NDVI and NDVIstd, average for the year 2003), Tree Cover (Sexton *et al.* 2013) and elevation data from the Global Land Cover Facility (http://www.landcover.org) and a measure of surface moisture characteristics from the NASA Scatterometer Climate Record Pathfinder project (QuickSCAT mean and standard deviation, downloaded from scp.byu.edu).

Assessing the role of geography and environment

To assess the relative contributions of geography and the environment to genetic divergence in the willow flycatcher, we compared genetic, environmental, and geographic distance matrices and used multiple tests designed to account for spatial autocorrelation. For locations with > 4 individuals (Table 1), we calculated pairwise F_{ST} across all quality-filtered SNPs using the R package SNPrelate (Zheng *et al.* 2012) and pairwise geographic distances from longitude and latitude using the R package geosphere (Hijmans *et al.* 2012). We then calculated environmental distance between each pair of sites by removing highly correlated climate variables (Pearson's r>0.7; Table 2; SI Table 1), scaling and centering each environmental variable to account for differences in magnitude, and then calculating pairwise Euclidean differences between sites. Mantel, Partial Mantel, and multiple regression of distance matrices were used

to test for associations between linearized F_{ST} (F_{ST}/1-F_{ST}) and genetic and environmental distance after accounting for geographic distance.

Gradient forest prediction of genomic mismatch

We identified the environmental variables that best explained genetic variation using gradient forest analysis with the R package gradientForest (Ellis *et al.* 2012). Because rare alleles are more likely to yield false positives, we only used SNPs with minor allele frequency >10%. The gradient forest analysis (ntree=500, nbin=201, corr.threshold=0.5) provided a ranked list based on the relative predictive power of all environmental variables (Table 2). To ensure that our model was explaining more variation than we would expect by chance, we compared the number of SNPs with positive R^2 and the mean R^2 across these 'predictive' loci (those with positive R^2) to 10 runs with randomized environments. Visualization of the gradient forest model across the range of the willow flycatcher (Buschke *et al.* 2016), was done by generating and extracting uncorrelated BIOCLIM values for 100,000 random points. The final gradient forest model was used to predict the genomic composition from uncorrelated environmental variables for each random point (Table 2). Principal components analysis (PCA) was used to summarize values. To visualize the different adaptive environments across the breeding range, colors were assigned based on the top 3 principal components axes, as recommended by the authors (Ellis *et al.* 2012).

We extended the gradient forest analysis to predict "genomic vulnerability" using the method presented by Fitzpatrick and Keller (2015). Here, "genomic vulnerability" (termed "genetic offset" by Fitzpatrick and Keller) is a measure of the mismatch between genotype and future predicted environment using associations across current gradients as a baseline. We used the baseline gradient forest model calculated using current BIOCLIM values to predict genomes under future environmental conditions (based on RCP 2.6 2050 projections) at the same 100,000 random points. The Euclidean distance between these weighted current and predicted values is what we refer to as "genomic vulnerability" (Bay *et al.* 2018).

Identification of SNPs as candidates for environmental selection

To identify SNPs (with minor allele frequency >0.1) that were most highly associated with the top environmental variables while accounting for underlying population structure, we used Latent Factor Mixed Models (LFMM) (Frichot *et al.* 2013). For each of the top 8 environmental variables from the gradient forest analysis, we ran five separate MCMC runs with a latent factor of K=4, based on the number of reported subspecies and previous morphological and genetic analysis based upon neutral markers (Paxton 2000). P-values from all five runs were combined and adjusted for multiple tests using a false discovery rate (FDR) correction. We annotated each significant SNP with genes within 25kb

upstream or downstream which we assume is within the distances before which LD should break down (Backstrom *et al.* 2006).

Validation of climate associated SNPs

To validate genotype-environment correlations identified in the LFMM analysis, we genotyped the top ranking 18 SNPS that were significantly associated with the top 8 climate variables and could be converted to SNPtype Assays in an additional 274 breeding individuals from 25 locations. DNA was extracted from feather samples using the KingFisher™ Cell and Tissue DNA Kit and SNP genotyping was performed on the Fluidigm™ 96.96 IFC controller following manufacturer guidelines. Nine individuals with greater then 8% of missing data were removed from downstream analysis and final allele frequencies were calculated for each SNP at each location. Standard linear regression was used to test for significant associations between climate and allele frequency (FDR-corrected p-value <0.05).

Association between genomic vulnerability and abundance

To assess the relationship between genomic vulnerability and abundance and determine which subspecies may be most vulnerable to future climate change, we correlated estimates of genomic vulnerability with willow flycatcher relative abundance from the North American Breeding Bird Survey (BBS) for 2011-2015, including all sites where the species was detected at least once during the history of the survey (Pardieck 2017). In order to associate the two datasets, vector-based BBS relative abundance estimates derived from inverse-distance weighting interpolation (2010-15; Sauer et al. 2017; https://www.mbr-pwrc.usgs.gov/bbs/shape_ra15.html) of route-level mean counts was converted to raster format with grid resolution of approximately 15 × 15 km. We then extracted values of relative abundance and genomic vulnerability for grid cells including BBS routes using bilinear interpolation (Hijmans 2015). For cells with BBS routes where detections had been recorded, but for which model-based estimates of abundance were not available due to low abundance and isolation from other sites with detections, we assigned mean count values (~ 9% of routes; mean count = 0.06). Significant differences in genomic vulnerability between subspecies were assessed using boxplots with 95% confidence intervals around median vulnerability scores (Chambers *et al.* 1983).

Results

Genome assembly, SNP discovery, and SNP/population filtering

The final Southwestern willow flycatcher genome assembly was 1.2 Gb in length and consisted of 7,791 scaffolds (contig N50=79,613bp; scaffold N50=895,074bp). In total, we identified 6,355,061 SNPs across the genome. Discarding low quality SNPs and low coverage individuals resulted in a final set of

105,000 SNPs and 175 individuals (SI Fig. 1), with less than 7.4% missing genotypes per SNP (mean=2.3%), less than 15.6% missing SNPs per individual (mean=2.3%), and minor allele frequency greater than 1%. Because F_{ST} is robust to low sample size when a large number of SNPs are employed (Nazareno *et al.* 2017), we retained all populations with a minimum of 4 (mean = 8) individuals for analysis based upon F_{ST} (distance matrix comparisons), resulting in a final dataset of 168 individuals from 22 sampling locations. Alternatively, to avoid bias associated with low sample size in analyses requiring estimates of allele frequency (Gradient Forest and LFMM), we used only populations with a minimum of six individuals (average = 10), resulting in a final dataset of 136 individuals from 14 sampling locations (Fig. 1; Table 1).

Assessing the role of geography and environment in shaping genetic structure

Pairwise F_{ST} across all quality-filtered SNPs ranged from 0 - 0.11 (SI Table 2). Mantel tests revealed highly significant correlations between genetic and geographic distance ($r = 0.70$, $P = 1 \times 10^{-6}$), genetic and environmental distance ($r = 0.56$, $P = 1 \times 10^{-6}$), and geographic and environmental distance ($r = 0.42$, $p = 1.8 \times 10^{-4}$) (SI Fig. 2A). Partial Mantel tests revealed the correlation between genetic and environmental distance remained significant after accounting for the relationship between genetic and geographic distance ($r = 0.42$, $p = 3 \times 10^{-4}$; SI Fig. 2) and both geographic and environmental distances were significant in a multiple regression of distance matrices (MRM: R^2=0.59; geography $P = 1 \times 10^{-5}$; environment $P = 3 \times 10^{-5}$).

Gradient forest mapping of genotype environment correlations

More genetic variation was explained by our gradient forest than those generated under randomized environments (SI Fig. 3). A total of 9015 SNPs were correlated with environment with mean R^2=0.18, compared to a mean R^2 of 0.13-0.15 across 3489-5633 SNPs for randomized data. We used gradient forest models to identify which climate and vegetation variables were most important in structuring genetic variation in the willow flycatcher and visualize climate-associated allelic variation across the breeding range (Fig. 2 A & B). Seven temperature variables and one precipitation variable were most strongly correlated with genetic variation across the breeding range of willow flycatchers (Table 2). Mapping principal components of gradient forest output revealed putative signals of local adaptation across the US Southwest, the East, the Inter-Mountain West, and the Pacific Northwest geographic regions (Fig. 2C).

Identification of candidate SNPs for environmental selection

To investigate genomic regions potentially involved in climate adaptation, we identified genomic regions associated with the top 8 climatic variables (which explained 49% of the total variation) using Latent Factor Mixed Models (*25*) (Table 2, SI Table 3). We found 77, 100, 104, 97, 97, 58, 107 and 70 SNPs significantly associated with BIO11, BIO10, BIO5, BIO1, BIO6, BIO9, BIO4 and BIO17 respectively (FDR-corrected $p<0.05$), with 1 SNP located on chromosome 16, Climate_20, shared among 7 variables. The SNPs were broadly distributed across the genome and within 25 KB of 202 genes with a variety of functions (SI Table 3). We identified 5 genes (BRACA1, RND2, CIITA, ICOS, and UBE2C) that were among the ~300 genes found to be differentially expressed in an RNA-seq analysis of thermal tolerance in chickens (Zhang *et al.* 2017), two of which were physically linked (BRACA1 and RND2), and an additional 5 genes (Ecel1, SLC23A2, NOX4, PIRT, and GR1N1) with GO terms related to other aspects of thermal tolerance, including respiratory system process, oxidative stress, and response to heat (Rimoldi *et al.* 2015) (SI Table 4). Three of the 5 genes from the poultry thermal stress study as well as 3 of the genes with potentially relevant GO terms were found to be outliers in association with BIO6 (Fig. 3A). Further, targeted genotyping using Fluidigm SNPtype assays for 18 of the top candidate SNPs in an additional 274 birds from 24 locations validated climate associations in 8/18 SNPs (FDR-corrected $p<0.05$; SI Table 5). In particular, we found a highly significant relationship between the Climate_20 SNP and 7 of the 8 top ranked climate variables in both the genome scan and validation results. While no link between Climate_20 and genes linked to thermal tolerance in birds was found, the highly significant relationship between this SNP and climate variables reflective of the intensity of summer heat waves, such as Mean Temperature of the Warmest Quarter (BIO10), suggests a potential role for this region in climate adaptation (Fig. 3 B, C, and D).

Prediction of genomic mismatch and association between vulnerability and abundance

Under a model of future climate change, genomic vulnerability was predicted to be highest in the southern part of the willow flycatcher range (Fig. 4A), corresponding to the range of the Southwestern willow flycatcher subspecies range. Overall, highest genomic vulnerability occurred at sites with especially low abundance, resulting in a significant negative correlation between abundance and genetic vulnerability ($r = -0.18$; $P < 0.001$; $df = 1382$; Fig. 4B, C). Abundance of southwestern willow flycatcher was low across sites and correlation between abundance and vulnerability for this subspecies was especially strong ($r = -0.49$; $P = 0.016$; $df = 27$) and weakest for the eastern subspecies region (*traillii*; $r = -0.11$; $P < 0.001$; $df = 957$). While there were regions of high and low genomic vulnerability across the range, the southwestern willow flycatcher subspecies had the highest overall median genomic vulnerability score (Fig. 4D).

Discussion

Climate envelope models are widely used to predict future species distributions (Parmesan & Yohe 2003; Pacifici *et al.* 2015), but such models do not account for complex biotic interactions (competition, specialization, coevolution, etc.) or barriers to dispersal that may limit range shifts (Williams *et al.* 2008). Here we show evidence for local adaptation in the willow flycatcher, supporting the idea that climate vulnerability estimates based on a single species distribution model across the entire North American breeding range (Zink 2015) could potentially result in misplaced conservation efforts. Here we move beyond species distribution modeling to identify populations that will need to adapt most to keep pace with climate change - a critically important question for the endangered Southwestern willow flycatcher whose dispersal is known to be limited by proximity to water sources (Friggens & Finch 2015). By calculating the difference between current genotype-environment relationships and those predicted under future climate change, we identify regions of highest vulnerability in the southern part of the willow flycatcher range. A comparison of the average genomic vulnerability across all currently recognized subspecies strongly supports the view that allele frequencies in the endangered Southwestern willow flycatcher will have to evolve most to keep pace with climate change. Significant correlations between estimates of genomic vulnerability and abundance from Breeding Bird Survey data confirm that already rare populations in the Southwest and throughout the range have the highest genomic vulnerability, suggesting that climate change may have already had an impact on population declines in regions at the edge of the species niche. Our results demonstrate how the incorporation of genotype-environment relationships into models of climate vulnerability can improve predictions of climate-induced impacts below the species level.

Assessing the extent of intraspecific variation in climate tolerances is an important first step towards understanding species vulnerability to climate change. Here we investigate the relationship between genetic, geographic and environmental distance in the willow flycatcher and find consistent support for the conclusion both geography and environment are important to genetic divergence in the willow flycatcher (SI Fig. 2). Mapping putatively adaptive genetic variation using gradient forest-transformed climate variables supports the idea that the Pacific Northwest, the Southwest, the East, and the Inter-Mountain West harbor unique genotype-environment correlations. More specifically, our results support the idea that high maximum temperatures during the warmest month (BIO5) are important to genotype-environment correlations in the Southwest, while genotype-environment relationships in the Pacific Northwest are driven by environmental variables such as precipitation during the driest quarter (BIO17) and mean temperatures during the coldest quarter (Figs. 2 & 3). In contrast, genotype-environment

correlations in Inter-Mountain West and Eastern populations, center closer zero in the PCA (Fig. 2A), indicating a more moderate impact of climate variables underling climate adaptation in this area. In sum, our results support the idea that genotype-environment correlations in the willow flycatcher are complex, involving multiple environmental variables and genomic regions and such information can be used to help refine estimates of future climate vulnerability.

Adaptation to local environments often occurs through natural selection acting on a large number of loci, each with a small effect on phenotype (Orr 2005). Here we identify putative loci important to local adaptation in the willow flycatcher, after accounting for underlying population structure, and find between 58 – 107 SNPs significantly associated with each of the top 8 environmental variables (SI Table 3). Independent validation of our top climate-associated SNPs in 274 new individuals from 24 populations revealed that 8 of our top 18 loci were likely robust to Type 1 error. While such error is a problem common to all association studies (McCarthy *et al.* 2008), the high number of false positives in our data underscores the idea that genotype-environment associations that cannot be validated should be interpreted with caution. Highly significant associations between Climate_20 and 7 of our 8 top-ranked environmental variables in both the genome scan and validation datasets provides the strongest evidence for local adaptation across the willow flycatcher genome (Fig. 3). While no associations between Climate_20 and genes known to be important to thermal tolerance in birds were identified, the relationship between allele frequency variation in this SNP and Mean Temperature of the Warmest Quarter (BIO5) suggests a potential role for this region in adaptation to temperature extremes. Overall, our results are in keeping with the idea that willow flycatchers exhibit region-specific genotype-climate associations that should be considered when assessing the capacity for endangered populations of the Southwestern willow flycatcher to shift their range in response to rising global temperatures.

While genotype-environment correlations have been noted across a variety of plant and animal systems, the mechanisms behind such local adaptation remain less well understood. Recent work on birds supports the idea that exposure to high temperatures can result in dehydration and heat stress related mortality (Albright *et al.* 2017; Zhang *et al.* 2017). As a first step towards understanding the genomic basis of adaptation to temperature in the willow flycatcher, we identify genes within 25KB of our top-ranking climate-associated SNPs (SI Table 4). Our strongest evidence for genes and gene regions that may be important to climate adaptation in this species comes from the overlap between five genes in our panel (BRACA1, RND2, CIITA, ICOS, and UBE2C) and those that were also found to be differentially expressed in a thermal tolerance study in poultry (Zhang *et al.* 2017). More specifically, Zang et al (2016) concluded that expression of these genes was linked to the dissipation of heat through increased heart pumping and blood circulation in smaller breeds of chickens. These results are consistent with the recent

work by Albright et al (2017) who found that small passerines in the Desert Southwest were particularly prone to mortality resulting from the failure to maintain body temperatures below lethal limits. While more research is needed, it is possible that physiological pathways responsible for overheating are related to those involved in interspecific adaptation to temperature extremes. Further, while limited gene annotation information for non-model organisms makes us cautious about placing significance on GO term analyses (Stein 2001), we also note the presence of four genes (Ecel1, SLC23A2, NOX4, PIRT, and GRIN1) with GO terms related to heat stress, thermal tolerance, and oxidative stress. Future efforts will focus on validating gene environment correlations at putative heat stress related loci as well as investigating the extent to which the genes identified here may serve as a mechanism for adaptation to temperature extremes in the willow flycatcher.

Desert ecosystems are home to some of the world's rarest species, many of which are already threatened by climate change (Loarie *et al.* 2009). Methods for assessing climate change impacts that rely on single species distribution models may overlook the importance of local adaptation in the ability of populations to respond to environmental shifts, potentially leading to misplaced conservation efforts. The US Fish and Wildlife Service was considering removing the Southwestern willow flycatcher from the endangered species list, in part because of a single species distribution model that showed no evidence of habitat specialization across the range. Here we annotate the first willow flycatcher genome and use population-level, genome-wide sequencing to show that willow flycatchers are not a single homogenous group, but a composite of locally adapted populations with specific genotype-environment relationships related to differences in temperature extremes. Clear evidence for local adaptation across the range highlights the need for management efforts below the species level if locally adapted populations are to be conserved. Estimates of the mismatch between current genotype-environment correlations and those predicted under future climate indicate that the Southwestern subspecies is at the greatest risk of climate-induced extinction. Our findings support the idea that protection or enhancement of riparian thermal refuges (Chen 1999) within regions of lower genomic vulnerability in the desert Southwest may be the most effective strategy for conserving remaining populations of flycatchers by buffering them from temperature extremes.

Acknowledgements: We thank the many individuals who contributed genetic samples, including T. Kita, B. Kus, R. Taylor, M. Fylling and many MAPs (Monitoring Avian Productivity and Survivorship) station operators with in the Institute for Bird Populations Network. This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant ACI-1548562. We thank the Vincent J. Coates Genomics Sequencing Laboratory at the University of

California, Berkeley as well as the UC Davis Genome Center for their help with the sequencing. This work was made possible by a generous gift from J. Ellis as well as an NSF Postdoctoral Fellowship (to R. Bay), a California Energy Commission grant EPC-15-043 (to K. Ruegg) and donation from First Solar Incorporated. Any use of trade, product, or firm names in this publication does not imply endorsement by the U.S. Government.

Literature Cited

Albright, T.P., Mutiibwa, D., Gerson, A.R., Smith, E.K., Talbot, W.A., O'Neill, J.J. *et al.* (2017). Mapping evaporative water loss in desert passerines reveals an expanding threat of lethal dehydration. *Proceedings of the National Academy of Sciences*, 201613625.

Ali, O.A., O'Rourke, S.M., Amish, S.J., Meek, M.H., Luikart, G., Jeffres, C. *et al.* (2016). RAD capture (Rapture): flexible and efficient sequence-based genotyping. *Genetics*, 202, 389-400.

Backstrom, N., Ovarnstrom, A., Gustafsson, L. & Ellegren, H. (2006). Levels of linkage disequilibrium in a wild bird population. *Biol Letters*, 2, 435-438.

Bay, R., Harrigan, R.J., Underwood, V.L., Gibbs, H.L., Smith, T.B. & Ruegg, K. (2018). Genomic signals of selection predict climate-driven population declines. *Science*, 83-89.

Boetzer, M., Henkel, C.V., Jansen, H.J., Butler, D. & Pirovano, W. (2010). Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics*, 27, 578-579.

Both, C., Bouwhuis, S., Lessells, C. & Visser, M.E. (2006). Climate change and population declines in a long-distance migratory bird. *Nature*, 441, 81-83.

Both, C. & Visser, M.E. (2001). Adjustment to climate change is constrained by arrival date in a long-distance migrant bird. *Nature*, 411, 296-298.

Buschke, F.T., Brendonck, L. & Vanschoenwinkel, B. (2016). Adding energy gradients and long -distance dispersal to a neutral model improves predictions of Madagascan bird diversity. *Ecology and Evolution*, 6, 6919-6929.

497 Cantarel, B.L., Korf, I., Robb, S.M., Parra, G., Ross, E., Moore, B. *et al.* (2008). MAKER: an
498 easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res.*,
499 18, 188-196.

500 Carroll, M., DiMiceli, C., Sohlberg, R. & Townshend, J. (2004). 250m MODIS normalized
501 difference vegetation index. *University of Maryland, College Park, Maryland*.

502 Catchen, J., Hohenlohe, P.A., Bassham, S., Amores, A. & Cresko, W.A. (2013). Stacks: an
503 analysis tool set for population genomics. *Molecular ecology*, 22, 3124-3140.

504 Chambers, J.M., Cleveland, W.S., Kleiner, B. & Tukey, P.A. (1983). *Graphical methods for data*
505 *analysis*. Wadsworth Belmont, CA.

506 Chen, J., Saunders, S.C., Crow, T.R., Naiman, R.J., Brosofske, K.D., Mroz, G.D., Brookshire,
507 B.L. and Franklin, J.F. (1999). Microclimate in forest ecosystem and landscape ecology:
508 variations in local climate can be used to monitor and compare the effects of different
509 management regimes. *Bioscience*, 49, 288-297.

510 Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A. *et al.* (2011). The
511 variant call format and VCFtools. *Bioinformatics*, 27, 2156-2158.

512 Dawson, T.P., Jackson, S.T., House, J.I., Prentice, I.C. & Mace, G.M. (2011). Beyond
513 predictions: biodiversity conservation in a changing climate. *Science*, 332, 53-58.

514 Delcher, A.L., Salzberg, S.L. & Phillippy, A.M. (2003). Using MUMmer to identify similar
515 regions in large sequence sets. *Current Protocols in Bioinformatics*, 10.13. 11-10.13. 18.

516 Diffenbaugh, N.S., Giorgi, F. & Pal, J.S. (2008). Climate change hotspots in the United States.
517 *Geophysical Research Letters*, 35.

518 Ellis, N., Smith, S.J. & Pitcher, C.R. (2012). Gradient forests: calculating importance gradients
519 on physical predictors. *Ecology*, 93, 156-168.

520 Estrada, A., Morales-Castilla, I., Caplat, P. & Early, R. (2016). Usefulness of species traits in
521 predicting range shifts. *Trends in ecology & evolution*, 31, 190-203.

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, 1-16.

Frichot, E., Schoville, S.D., Bouchard, G. & François, O. (2013). Testing for associations between loci and environmental gradients using latent factor mixed models. *Molecular biology and evolution*, 30, 1687-1699.

Friggens, M.M. & Finch, D.M. (2015). Implications of Climate Change for Bird Conservation in the Southwestern US under Three Alternative Futures. *Plos One*, 10.

Grinnell, J. & Miller, A.H. (1944). *The distribution of the birds of California*. The Cooper Ornithological Society, Berkeley, CA.

Hijmans, R.J. (2015). Geographic Data Analysis and Modeling. R package version 2.5-8.

Hijmans, R.J., Cameron, S.E., L., P.J., Jones, P.G. & Jarvis, A. (2005). Very high resolution interpolated global terrestrial climate surfaces. *International Journal of Climatology*, 25, 1965-1978.

Hijmans, R.J., Williams, E. & Vennes, C. (2012). geosphere: Spherical Trigonometry. R package version 1.2–28. *CRAN. R-project. org/package= geosphere*.

Hsiang, S., Kopp, R., Jina, A., Rising, J., Delgado, M., Mohan, S. *et al.* (2017). Estimating economic damage from climate change in the United States. *Science*, 356, 1362-1369.

Hunt, M., Kikuchi, T., Sanders, M., Newbold, C., Berriman, M. & Otto, T.D. (2013). REAPR: a universal tool for genome assembly evaluation. *Genome biology*, 14, R47.

Langmead, B. & Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Meth.*, 9, 357-359.

Loarie, S.R., Duffy, P.B., Hamilton, H., Asner, G.P., Field, C.B. & Ackerly, D.D. (2009). The velocity of climate change. *Nature*, 462, 1052.

McCarthy, M.I., Abecasis, G.R., Cardon, L.R., Goldstein, D.B., Little, J., Ioannidis, J.P. *et al.* (2008). Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat. Rev. Genet.*, 9, 356.

McLeod, M.A., Koronkiewicz, T.J., Brown, B.T., Langeberg, W.J. & Carothers, S.W. (2008). Southwestern Willow Flycatcher surveys, demography, and ecology along the lower Colorado River and tributaries, 2003–2007. Five-year summary report submitted to U.S. Bureau of Reclamation, Boulder City, Nevada. SWCA Environmental Consultants, Flagstaff, Arizona.

Menéndez, R., Megías, A.G., Hill, J.K., Braschler, B., Willis, S.G., Collingham, Y. *et al.* (2006). Species richness changes lag behind climate change. *Proceedings of the Royal Society of London B: Biological Sciences*, 273, 1465-1470.

Nazareno, A.G., Bemmels, J.B., Dick, C.W. & Lohmann, L.G. (2017). Minimum sample sizes for population genomics: an empirical study from an Amazonian plant species. *Molecular ecology resources*.

O'Connell, J., Schulz-Trieglaff, O., Carlson, E., Hims, M.M., Gormley, N.A. & Cox, A.J. (2015). NxTrim: optimized trimming of Illumina mate pair reads. *Bioinformatics*, 31, 2035-2037.

Orr, H.A. (2005). The genetic theory of adaptation: a brief history. *Nat. Rev. Genet.*, 6, 119-127.

Pacifici, M., Foden, W.B., Visconti, P., Watson, J.E., Butchart, S.H., Kovacs, K.M. *et al.* (2015). Assessing species vulnerability to climate change. *Nature Climate Change*, 5, 215-224.

Pardieck, K.L., D.J. Ziolkowski Jr., M. Lutmerding, K. Campbell and M.-A.R. Hudson (2017). North American Breeding Bird Survey

Parmesan, C. & Yohe, G. (2003). A globally coherent fingerprint of climate change impacts across natural systems. *Nature*, 421, 37-42.

Paxton, E.H. (2000). Molecular genetic structuring and demographic history of the Willow Flycatcher (*Empidonax traillii*). Northern Arizona University Flagstaff, AZ, USA.

570 Rimoldi, S., Lasagna, E., Sarti, F.M., Marelli, S.P., Cozzi, M.C., Bernardini, G. *et al.* (2015).
571 Expression profile of six stress-related genes and productive performances of fast and slow
572 growing broiler strains reared under heat stress conditions. *Meta gene*, 6, 17-25.

573 Savolainen, O., Lascoux, M. & Merilä, J. (2013). Ecological genomics of local adaptation. *Nat.*
574 *Rev. Genet.*, 14, 807-820.

575 Sedgwick, J.A. (2000). Willow flycatcher (Empidonax traillii). In: *The Birds of North America*
576 (ed. Rodewald, PG). Cornell Lab of Ornithology Ithaca.

577 Service, U.S.F.a.W. (2002). Final Recovery Plan Southwestern Willow Flycatcher (Empidonax
578 traillii extimus). (ed. Service, USFaW) Albuquerque, New Mexico, p. 210.

579 Sexton, J.O., Song, X.-P., Feng, M., Noojipady, P., Anand, A., Huang, C. *et al.* (2013). Global,
580 30-m resolution continuous fields of tree cover: Landsat-based rescaling of MODIS vegetation
581 continuous fields with lidar-based estimates of error. *International Journal of Digital Earth*, 6,
582 427-448.

583 Sinervo, B., Mendez-De-La-Cruz, F., Miles, D.B., Heulin, B., Bastiaans, E., Villagrán-Santa
584 Cruz, M. *et al.* (2010). Erosion of lizard diversity by climate change and altered thermal niches.
585 *Science*, 328, 894-899.

586 Sogge, M.K., Marshall, R.M., Sferra, S.J. & Tibbitts, T.J. (1997). A southwestern willow
587 flycatcher natural history summary and survey protocol. Technical Report
588 NPS/NAUCPRS/NRTR-97/12. USGS Colorado Plateau Research Station, Northern Arizona
589 University, Flagstaff, AZ.

590 Stein, L. (2001). Genome annotation: from sequence to biology. *Nat. Rev. Genet.*, 2, 493.

591 Stephens, P.A., Mason, L.R., Green, R.E., Gregory, R.D., Sauer, J.R., Alison, J. *et al.* (2016).
592 Consistent response of bird populations to climate change on two continents. *Science*, 352, 84-
593 87.

Tarailo-Graovac, M. & Chen, N. (2009). Using RepeatMasker to identify repetitive elements in genomic sequences. *Current protocols in bioinformatics*, 4.10. 11-14.10. 14.

Theimer, T.C., Smith, A.D., Mahoney, S.M. & Ironside, K.E. (2016). Available data support protection of the Southwestern Willow Flycatcher Under the Endangered Species Act. *The Condor*, 118, 289-299.

Unitt, P. (1987). Empidonax traillii extimus: An endangered subspecies. *West. Birds*, 18, 137-162.

Urban, M.C. (2015). Accelerating extinction risk from climate change. *Science*, 348, 571-573.

Warren, R., VanDerWal, J., Price, J., Welbergen, J., Atkinson, I., Ramirez-Villegas, J. *et al.* (2013). Quantifying the benefit of early climate change mitigation in avoiding biodiversity loss. *Nature Climate Change*, 3, 678-682.

Wiens, J.J. (2016). Climate-related local extinctions are already widespread among plant and animal species. *PLoS Biol.*, 14, e2001104.

Williams, S.E., Shoo, L.P., Isaac, J.L., Hoffmann, A.A. & Langham, G. (2008). Towards an integrated framework for assessing the vulnerability of species to climate change. *PLoS Biol.*, 6, 2621-2626.

Zdobnov, E.M. & Apweiler, R. (2001). InterProScan–an integration platform for the signature-recognition methods in InterPro. *Bioinformatics*, 17, 847-848.

Zhang, J., Schmidt, C.J. & Lamont, S.J. (2017). Transcriptome analysis reveals potential mechanisms underlying differential heart development in fast-and slow-growing broilers under heat stress. *BMC Genomics*, 18, 295.

Zheng, X., Levine, D., Shen, J., Gogarten, S.M., Laurie, C. & Weir, B.S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, 28, 3326-3328.

Zink, R.M. (2015). Genetics, morphology, and ecological niche modeling do not support the subspecies status of the endangered Willow Flycatcher (*Empidonax traillii extimus*). *The Condor*, 117, 78-86.

Table 1. Sample location information. $N_{RAD\_nofilter}$ = number of idividuals for genome-wide RAD dataset before filtering for read depth and missing data , NRAD_filter = number of remaining post-filtering, Nvalidation = number of individuals in the SNP validation dataset.

Location	Latitude	Longitude	$N_{RAD\_}$	N_{RAD_filter}	$N_{validation}$

			nofilter		
East Pima, AZ	32.83	-109.7	_	_	8
San Pedro/Gila River confluence, AZ	32.98	-110.77	18	14	17
West Fort Ditch, NM	33.04	-108.54	_	_	11
San Carlos Reservation, AZ	33.2	-110.44	_	_	30
San Diego, CA	33.28	-117.37	14	4	6
Roosevelt Lake, AZ	33.77	-111.24	20	10	18
White Mountains, AZ	34.00	-109.00	15	13	15
Camp Verde, AZ	34.56	-111.84	_	_	17
Santa Ynez River, CA	34.62	-120.18	_	_	8
Zuni/Nutria Diversion Reservation, NM	35.24	-108.64	_	_	8
South Fork Kern River, CA	35.66	-118.46	20	13	11
Southern Ute Reservation, CO	37.12	-107.59	_	_	6
Pahranagat Lake NWR, NV	37.32	-115.13	_	_	6
Owen's River at Bishop, CA	37.41	-118.48	_	_	12
Alamosa National Wildlife Refuge, CO	37.5	-106	_	_	17
Beaver Creek, CO	37.68	-108.38	_	_	6
Clear Creek, CO	37.79	-108.24	_	_	8
Baltimore Area, MD	39.4	-76.99	_	_	8
Escalante State Wildlife Area, CO	39.47	-106.37	_	_	13
Fish Creek, UT	39.78	-111.20	14	11	_
Rio Blanco Lake, CO	40.09	-108.21	_	_	7
Orefield, PA	40.66	-75.67	21	21	_
White River Confl. to the Green River, UT	40.67	-109.68	7	6	_
Willow Slew, IN	40.98	-87.53	4	4	_
Bigelow Meadows, CA	41.26	-121.88	7	6	_
Agusta, MI	42.3	-85.32	_	_	9
Mink Creek, ID	42.75	-112.39	6	6	_
Malheur NWR, OR	42.83	-118.87	7	6	_
FCTC-SABO, MI	42.84	-85.30	4	4	6
Jones Creek, OR	43.04	-123.97	10	10	_
Little White River Rec. Area, SD	43.17	-101.53	4	4	6
Black Creek, NY	43.38	-73.91	6	4	_
Fall Creek 2, ID	43.43	-111.40	7	7	_
Marion Forks, OR	44.37	-122.02	_	_	14
Finley NWR, OR	44.41	-123.35	3	0	_
Priem Road, OR	44.78	-123.38	7	6	_
Elm Creek, MN	45.13	-93.45	4	0	6
Waubay NWR, SD	45.40	-97.33	4	4	_
Hamon Memorial, MT	45.95	-114.13	5	4	_
Carbondale (Edgwick), WA	47.09	-122.05	8	7	_
Fork clearcut, WA	47.97	-124.40	4	4	_
Total			219	168	273

Table 2. Environmental variables used in the gradient forest analysis, ordered by ranked importance of variables and the cumulative contribution of each variable. The top eight environmental variables represent 49% of the total.

Variable	Definition	GF Rank	Cumulative Contribution
BIO11 *	Mean Temperature of Coldest Quarter	8.03E-04	7.66
BIO10	Mean Temperature of Warmest Quarter	6.71E-04	14.40
BIO1	Annual Mean Temperature	6.41E-04	21.05
BIO5*	Max Temperature of Warmest Month	6.40E-04	27.47
BIO6	Min Temperature of Coldest Month	5.79E-04	32.90
BIO4*	Temperature Seasonality (standard deviation *100)	5.20E-04	38.30
BIO9	Mean Temperature of Driest Quarter	4.91E-04	43.64
BIO17*	Precipitation of Driest Quarter	4.78E-04	48.76
NDVI_Mean	Vegetation Indicies	4.50E-04	53.18
BIO15	Precipitation Seasonality (Coefficient of Variation)	4.28E-04	57.39
BIO7	Temperature Annual Range (BIO5-BIO6)	3.75E-04	61.41
TreeCover	Tree Cover	3.72E-04	65.41
BIO14	Precipitation of Driest Month	3.64E-04	69.36
BIO16	Precipitation of Wettest Quarter	3.09E-04	72.75
BIO19	Precipitation of Coldest Quarter	2.96E-04	76.04
BIO2*	Mean Diurnal Range (Mean of monthly (max temp - min temp))	2.90E-04	79.31
BIO8*	Mean Temperature of Wettest Quarter	2.87E-04	82.49
BIO13	Precipitation of Wettest Month	2.82E-04	85.63
STM	Elevation	2.21E-04	88.36
BIO12	Annual Precipitation	2.12E-04	90.98
BIO3	Isothermality (BIO2/BIO7) (* 100)	2.07E-04	93.54
QuickScat	Surface moisture characteristics	2.02E-04	95.87
BIO18	Precipitation of Warmest Quarter	1.92E-04	98.17
NDVI_StDev	Vegetation Indicies	1.81E-04	100.00

** Top ranked, uncorrelated climate variables used for Gradient Forest mapping and distance matrix comparison analyses. These variables were selected by moving down the list of ranked importance for the full model and discarding*

variables highly correlated (Pearson's r>7) with a variable of higher importance.

647

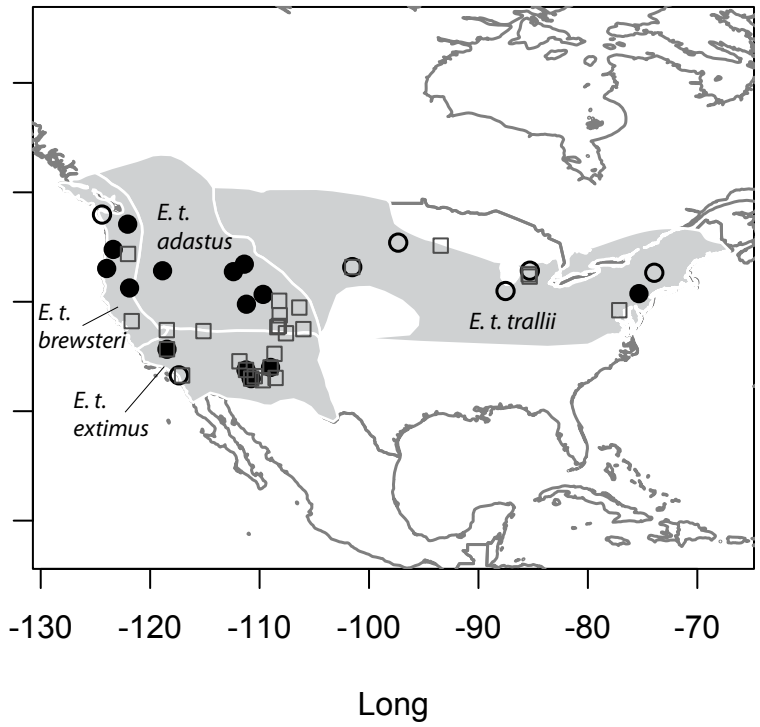

Figure 1. Willow Flycatcher Range Map and Sampling. Open and closed circles represent the data used in distance matrix comparison tests, while only populations represented by closed circles were used in the Gradient Forest analysis. Open gray boxes represent populations used to validate gene-environment correlations. Lines represent currently recognized subspecies boundaries according to Sogge et al (1997). *E. t. brewsteri* = Pacific Coastal, *E.t. adastus* = Interior West, *E. t. trallii* = East, and *E. t. extimus* = Southwest.

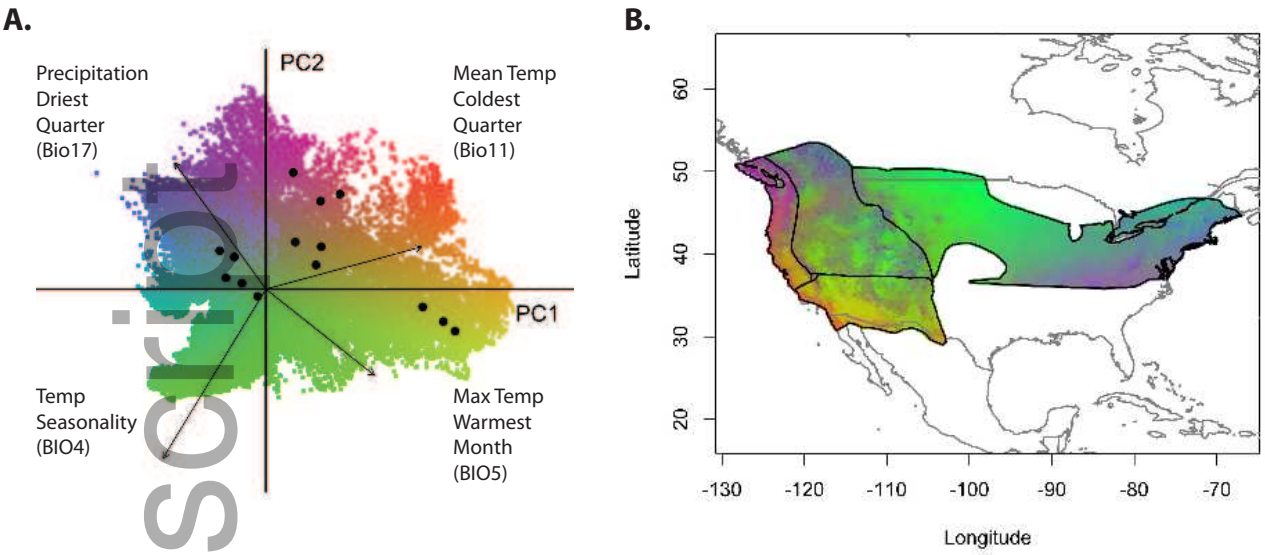

Figure 2. Mapping gene-environment correlations across the willow flycatcher breeding range. A) Principal components analysis of gradient forest-transformed climate variables. Black dots represent the PC scores associated with the sampling locations, while colors are based upon modeled gene-environment correlations from 100,000 random points across the breeding range. Arrows show the loadings of the top ranked uncorrelated environmental variables. B) Gradient forest-transformed climate variables from the PCA mapped to geography support climate adaptation across the breeding range. Black lines designating approximate subspecies locations support the idea that while subspecies are adapted to distinct ecological regions, climate adaption is complex.

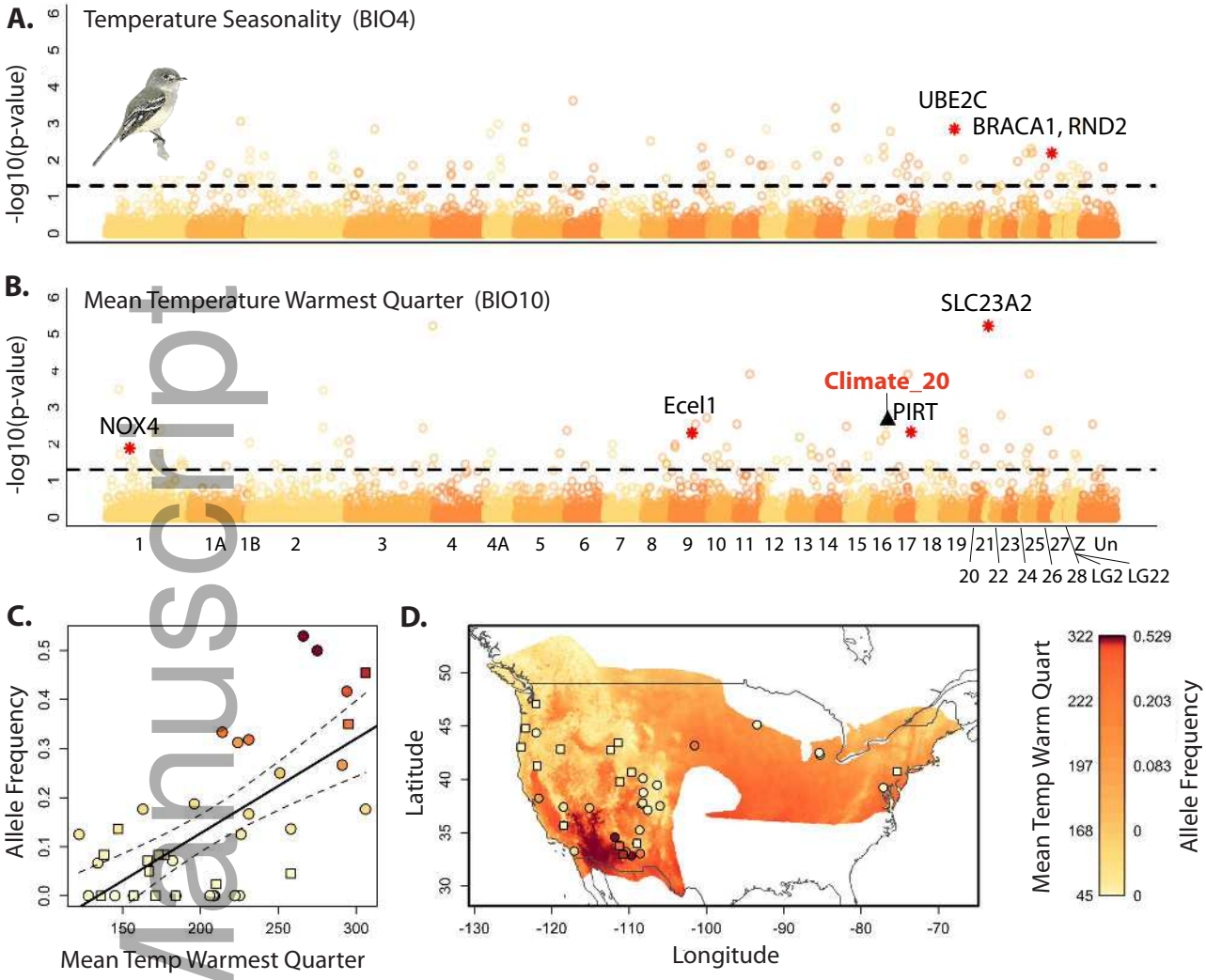

Figure 3. Candidate SNPs linked to temperature in the Willow flycatchers. A) Manhattan plot showing the FDR-corrected significance level for SNPs associated with Temperature Seasonality (BIO4) and B) Mean Temperature of the Warmest Quarter (BIO10). Dashed line represents p=0.05. Colors distinguish different chromosomes. Candidate genes linked to thermal tolerance in birds are highlighted by red stars and denoted with gene names, while Climate_20, the SNP validated in B and C below, is denoted by a black triangle. No link between Climate_20 and genes linked to thermal tolerance in birds was found, but the highly significant relationship between this SNP and 7 of the 8 top ranked climate variables (except temperature seasonality shown in A above) in both the genome scan and validation results (SI Table 5) suggest a potential role for this region in climate adaptation. C) Relationship between Climate_20 and mean temperature of the warmest quarter in genome scan and SNP validation datasets. The allele frequencies from the original genome scan data are denoted by squares, while allele frequencies based upon the validation set are denoted by circles. D) The association between Mean Temperature of the Warmest Quarter (BIO10) and Climate_20 across geographic space, with population allele frequencies color coded from high frequency (red) to low (yellow).

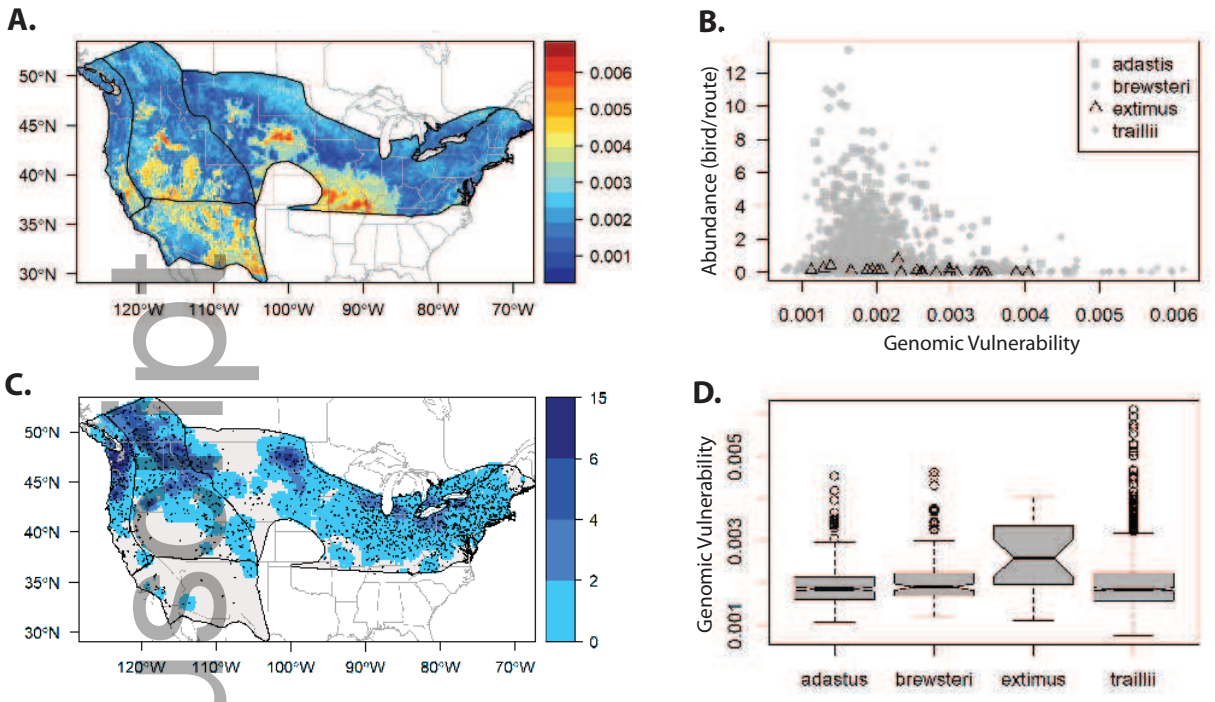

Figure 4. Genomic Vulnerability and abundance in the Willow Flycatcher. A) Map of genomic vulnerability across the Willow Flycatcher breeding range. Red = high genomic vulnerability, blue = low genomic vulnerability, lines indicate subspecies boundaries. B) Genomic Vulnerability versus abundance based upon the estimated mean number of birds/ route in 2011-2015 Breeding Bird Survey. C) Estimates of relative abundance from the BBS based on inverse-distance weighting interpolation. Points indicate the BBS routes where Willow Flycatchers have been recorded. Points in the grey areas fall in regions where abundance was too low or distant from other detection routes to be included in the BBS spatial model. D) Quantile box plots of the median Genomic Vulnerability broken down by subspecies. Open circles represent outliers.