Ecological genomics predicts climate vulnerability in an endangered southwestern songbird

Kristen Ruegg,1,2,† Rachael A. Bay,1,3,4 Eric C. Anderson,3,4 James F. Saracco,5 Ryan J. Harrigan,1 Mary Whitfield,6 Eben H. Paxton7 and Thomas B. Smith1,8

Abstract

Few regions have been more severely impacted by climate change in the USA 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 can facilitate more precise predictions of climate impacts and help guide conservation in threatened and endangered groups.

Keywords

climate change, ecological genomics, genomic vulnerability, local adaptation.


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 favouring 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. Although intraspecific variation in climate tolerances may factor critically in the ability of species to move or adapt to environmental change, most modelling 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 USA, 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, specialisation, co-evolution, 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 may either need to adapt or 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 synchronise 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. Furthermore, this work has shown that such stress is not just physiological in nature, but is associated with differential expression in a suite of c. 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 USA (Fig. 1; Pacific Northwestern form, E. t. brewsteri; Western Central form, E. t. adastus; and Eastern form, E. t. trallii). 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 et al. 1999; McLeod et al. 2008). At the turn of the century, the southwestern willow flycatcher was described as common wherever its specialised 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, urbanisation, 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 potentially 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 mostly new 25 populations. To identify geographical regions where the mismatch between current and future genotype–environment relationships is predicted to be the greatest we used gradient forest modeling to calculate an index of genomic vulnerability (Bay et al. 2018). Lastly, we assessed the relationship between our estimates of genomic vulnerability and abundance across the range in order to assess which subspecies may be most vulnerable to future climate change.

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).

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 grey boxes represent populations used to validate gene–environment correlations. Lines represent currently recognised 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.
Two hundred and nineteen individuals from 24 populations were used to test for genome-wide genotype–environment correlations, while 274 individuals spanning 25 populations (eight replicate and 17 new populations) were used to validate a subset of significant genotype–environment correlations identified in the genome-wide analysis ($N_{\text{total indiv}} = 493; N_{\text{total pops}} = 41$; 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, USA).

### 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,
Hayward, CA), with adjustments. One ug of DNA was diluted in 100 μL of AE buffer and fragmented to an average insert size of c. 400 bp. The resulting library was sequenced on two lanes of an Illumina HiSeq2500 using 250 bp paired-end sequencing at the QB3 Vincent J. Coates Genomics Sequencing Laboratory, UC Berkeley. Two mate-pair libraries were also created, using 4 and 8 kb inserts and sequenced on one-third of a Illumina HiSeq 2500 lane, using 100 bp paired-end sequencing at the Huntsman Cancer Center at the University of Utah. The 250 bp 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 8 kb insert library to break the assembly at likely error regions. SSPACE scaffolding was repeated with \( k = 5 \) and scaffolds < 5kb 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 then aligned to the Zebra Finch genome (version 3.2.4) using the Scattering Climate Record Pathfinder project (QuickSCAT measure of surface moisture characteristics from the NASA Scatterometer Climate Record Pathfinder project (QuickSCAT).)

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 and 1990, as well as vegetation indices (Carroll et al. 2004) (NDVI and NDV1std, 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; Table S1), scaling and centring each environmental variable to account for differences in magnitude, and then calculating pairwise Euclidean distances between sites. Mantel, Partial Mantel and multiple regression of distance matrices were used to test for associations between linearised \( F_{ST} (F_{ST}^2/1-F_{ST}) \) 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 randomised environments. Visualisation 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 summarise values. To visualise the different adaptive environments across the breeding range, colours were assigned based on the top three 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 & 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 eight 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 25 kb 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 eight climate variables and could be converted to SNPytype 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™ (ThermoFisher Scientific, USA) 96.96 IFC controller (Fluidigm Inc., San Francisco, CA, USA) following manufacturer guidelines. Nine individuals with > 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 et al. 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_r1a15.html) of route-level mean counts were 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 (c. 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 7791 scaffolds (contig N50 = 79 613 bp; scaffold N50 = 895 074 bp). 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 (Fig. S1), with less than 7.4% missing genotypes per SNP (mean = 2.3%), < 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 (Nazarenko et al. 2017), we retained all populations with a minimum of four (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 to 0.11 (Table S2). Mantel tests revealed high 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 ($r = 0.42$, $P = 1.8 \times 10^{-4}$) (Fig. S2A). Partial Mantel tests revealed the correlation between genetic and environmental distance remained significant after accounting for the relationship between genetic and geographic distance ($r = 0.42$, $P = 3 \times 10^{-4}$; Fig. S2). More genetic variation was explained by our gradient forest than those generated under randomised environments (Fig. S3). 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 randomised 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 visualise climate-associated allelic variation across the breeding range (Fig. 2a and 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 eight climatic variables (which explained 49% of the total variation) using Latent Factor Mixed Models (25) (Table 2, Table S3). 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 one SNP located on chromosome 16, Climate_20, shared among seven variables. The SNPs were broadly distributed across the genome and within 25 KB of 202 genes with a variety of functions (Table S3). We identified five genes (BRACA1, RND2, CIITA, ICOS and UBE2C) that were among the c. 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 five 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) (Table S4). Three of the five genes from the poultry thermal stress study were found to be outliers in association with BIO4, Temperature Seasonality (Fig. 3a). Furthermore, 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 \( P < 0.05 \); Table S5). In particular, we found a highly significant relationship between the Climate_20 SNP and seven of the eight 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. 3b, 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 region. Overall, highest genomic vulnerability occurred at sites with especially low abundance, resulting in a significant negative correlation between abundance and genetic vulnerability \((r = -0.18; P < 0.001; \text{d.f.} = 1382; \text{Fig. 4b, c})\). Abundance of southwestern willow flycatcher was low across sites and correlation between abundance and vulnerability for this subspecies was especially strong \((r = -0.49; P = 0.016; \text{d.f.} = 27)\) and weakest for the eastern subspecies region (traillii; \(r = -0.11; P < 0.001; \text{d.f.} = 957)\). 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, specialisation, co-evolution, etc.) or barriers to dispersal that may limit range shifts (Williams et al. 2008). In the case of the willow flycatcher, the capacity for range shifts may be restricted by the need to be proximate to specific water sources (Figgens & Finch 2015), making it important to incorporate the potential for adaptation into estimates of climate vulnerability. Here, we move beyond species distribution modelling to begin to identify populations that will need to adapt most to keep pace with climate change. 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 range. A comparison of the average genomic vulnerability across all currently recognised subspecies strongly supports the view that the endangered southwestern willow flycatcher is most vulnerable to 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 (Fig. S2). 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 harbour 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 and 3).

Figure 3 Candidate SNPs linked to temperature in the Willow flycatcher. (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 $P = 0.05$. Colours 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 seven of the eight top-ranked climate variables (except temperature seasonality shown in A above) in both the genome scan and validation results (Table S5) 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 colour coded from high frequency (red) to low (yellow).
In contrast, genotype–environment correlations in Interior Mountain West and Eastern populations, centre closer zero in the PCA (Fig. 2a), indicating a more moderate impact of climate variables underlying 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 and 107 SNPs significantly associated with each of the top eight environmental variables (Table S3). Independent validation of our top climate-associated SNPs in 274 new individuals from 24 populations revealed that eight 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 underscore the idea that genotype–environment associations that cannot be validated should be interpreted with caution. Highly significant 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 (Table S4). 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, Zhang et al. (2016) concluded that expression of these genes

**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 vs. 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.
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. Furthermore, 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 five 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 specialisation 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 the southwestern subspecies is at the greatest risk of climate-induced extinction. Our findings support the idea that the southwestern subspecies is at the greatest risk of climate-induced extinction.

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 within 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 T. Smith), a National Geographic grant WW-202R-17 (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.

AUTHOR CONTRIBUTIONS

K.R., R.A.B. and T.B.S. conceived 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 STATEMENT

The Willow flycatcher genome and annotations are available through NCBI (accession number: PWAB0000000) and population-level RAD-Seq data are available through NCBI’s Sequence Read Archive (http://www.ncbi.nlm.nih.gov/bioproject/453612).

REFERENCES


**SUPPORTING INFORMATION**

Additional Supporting Information may be found online in the supporting information tab for this article.

Editor, Tim Coulson
Manuscript received 9 November 2017
First decision made 11 December 2017
Second decision made 22 February 2018
Manuscript accepted 15 March 2018