nature climate change

Article

https://doi.org/10.1038/s41558-023-01696-3

Historical DNA reveals climate adaptation in an endangered songbird

Received: 22 November 2022

Accepted: 11 May 2023

Published online: 19 June 2023

Check for updates

Sheela P. Turbek[1] ✉, Christen Bossu[1], Christine Rayne[1], Cristian Gruppi[2], Barbara E. Kus[3], Mary Whitfield[4], Thomas B. Smith[2], Eben H. Paxton[5], Rachael A. Bay[6] & Kristen C. Ruegg[1]

To cope with climate change, species may shift their distributions or adapt in situ to changing environmental conditions. However, clear examples of genetic changes via adaptation are limited. We explore evolutionary responses to climate change in the endangered southwestern willow flycatcher (*Empidonax traillii extimus*) through whole-genome comparisons between historical specimens, collected from 1888 to 1909 near San Diego, California, United States, and contemporary individuals from across the breeding range. Genomic analyses revealed that introgression into San Diego increased adaptive potential over time and shifted genome-wide population structure towards that of neighbouring populations. In contrast, loci linked to climate (dew point temperature and precipitation) shifted away from neighbouring populations and in a direction consistent with adaptation to climate change in southern California. This research highlights the role of admixture in facilitating adaptive shifts through its impact on genome-wide genetic variation and represents one of the few studies to document climate adaptation in a wild population.

Climate change is threatening global biodiversity by disrupting food webs, altering habitat suitability and reducing the capacity of ecosystems to mitigate the effects of extreme weather events. To cope with environmental change, species may shift their distributions to track their ecological niche or adapt in situ to changing climate conditions[1]. Rapid evolution may play a key role in determining species persistence in the face of environmental change; nonetheless, adaptive capacity is rarely incorporated into models of climate-induced extinction risk[2]. Since the mid-twentieth century, a growing number of studies have documented phenotypic responses to rising global temperatures in climate-related traits, such as phenology, body size and thermal tolerance[3]. However, phenotypic plasticity appears to govern much of the variation in climate-related traits observed in natural populations[1,3,4] and clear examples of genetic changes via adaptation to altered environmental conditions are limited[3] (but see refs. 5–7).

The ability of a population to adapt to climate change depends on its adaptive potential[8,9] or the amount of genetic variation present in fitness-related traits. High levels of genetic drift in small populations can lead to the accumulation of deleterious alleles that negatively impact fitness and the loss of allelic diversity at functional genes (inbreeding depression), thereby compromising the ability of populations to adapt to environmental change. Natural hybridization or assisted gene flow can restore adaptive potential by introducing new genetic variation into a population[10–12], increasing the likelihood that the population will possess beneficial alleles that selection can act upon to adapt to changing climate conditions[8].

Recent advances in historical DNA sequencing have opened new avenues of research into evolutionary responses to past climate change[13]. We leverage historical and contemporary samples of the willow flycatcher (*Empidonax traillii*) to investigate genomic responses to

[1]Department of Biology, Colorado State University, Fort Collins, CO, USA. [2]Center for Tropical Research, Institute of the Environment and Sustainability and Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA, USA. [3]Western Ecological Research Center, San Diego Field Station, US Geological Survey, San Diego, CA, USA. [4]Southern Sierra Research Station, Weldon, CA, USA. [5]Pacific Island Ecosystems Research Center, US Geological Survey, Hawaii Volcanoes National Park, HI, USA. [6]Department of Evolution and Ecology, University of California Davis, Davis, CA, USA. ✉e-mail: sturbek@colostate.edu

https://doi.org/10.1038/s41558-023-01696-3

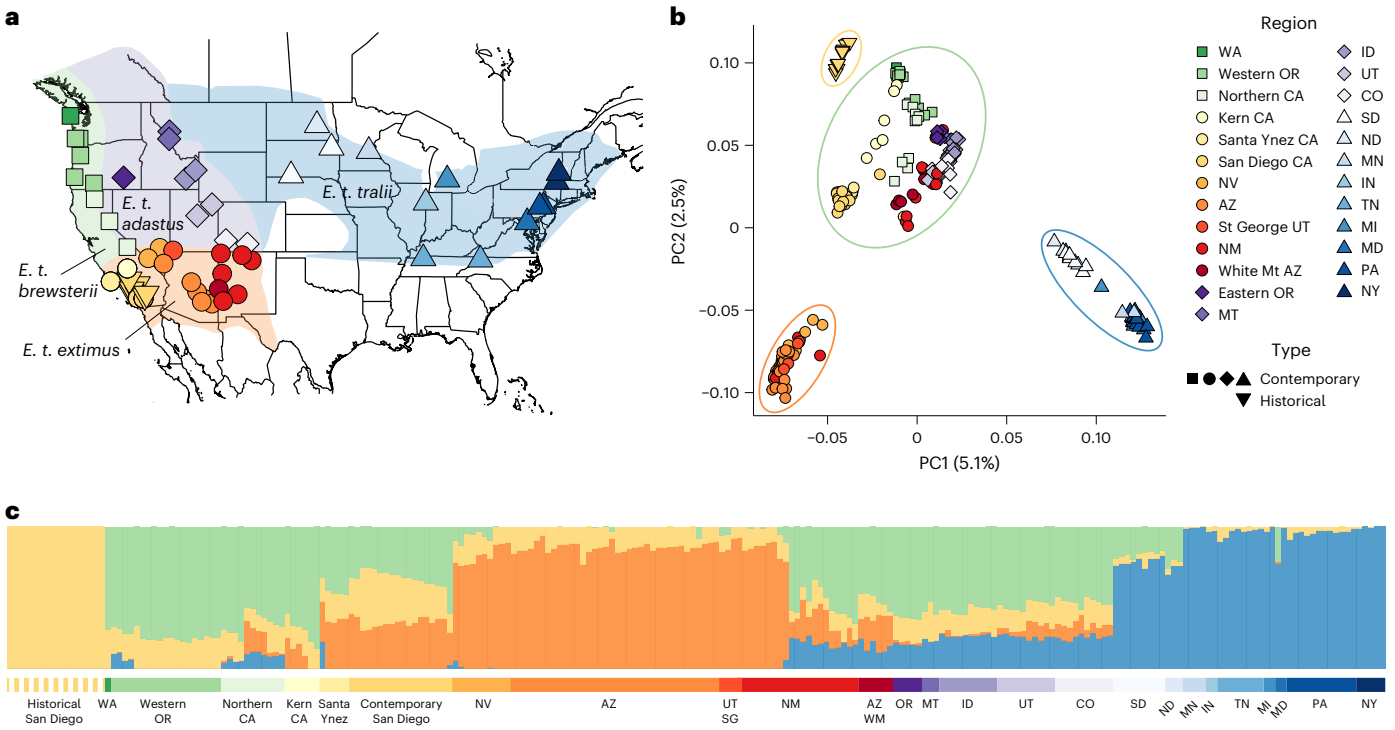


Fig. 1 | Sampling design and patterns of genomic differentiation from WGS data. a, Sampling locations and subspecies boundaries across the willow flycatcher breeding range[31]. Contemporary samples of the four subspecies are shown as squares, circles, diamonds and triangles, while upside-down triangles indicate historical samples. **b**, Genome-wide PCA (n = 238 individuals genotyped at 128,775 loci). Points are coloured by sampling location and correspond to the map in **a**. **c**, Individual admixture proportions of contemporary and historical samples for K = 4 (n = 238 individuals genotyped at 128,775 loci). Coloured bars along the x axis indicate sampling locations and correspond to the map in **a**. Two-letter codes are standard abbreviations for North American states.

climate change over the past century. The willow flycatcher is comprised of four subspecies that breed throughout the continental United States (Fig. 1a; Pacific northwest—*E. t. brewsteri*, interior west—*E. t. adastus*, east—*E. t. traillii* and southwest—*E. t. extimus*). The southwestern willow flycatcher, listed as federally endangered in 1995 due to severe population declines[14], primarily occupies the desert southwest, an area that has undergone extensive riparian habitat loss over the past century[14] and is predicted to experience more frequent heat waves as global temperatures continue to rise[15]. Previous work on the willow flycatcher identified genomic signatures of local adaptation to different climate regimes and predicted the degree of genomic vulnerability (the amount that populations must adapt to keep pace with climate change) across the breeding range[16]. Under a model of future climate change, genomic vulnerability is expected to be highest in the southwestern subspecies, raising concerns about climate-induced extinction[16].

To further explore evolutionary responses to climate change in the southwestern willow flycatcher, we focus specifically on San Diego, California, where the number of willow flycatcher breeding territories is declining despite widespread abundance before 1940[17,18]. California has warmed by an average of 1.4 °C since 1895[19]. Combined with an increase in the variability of precipitation patterns, this rise in temperature has led to the growing frequency and intensity of wildfires, droughts and other extreme weather events[19]. Using whole-genome sequencing (WGS) of 17 historical specimens, collected from 1888 to 1909 near San Diego and 221 contemporary individuals from across the willow flycatcher breeding range, we examine (1) shifts in population structure over time at putatively neutral and climate-linked loci, (2) changes in neutral genetic diversity, effective population size and adaptive potential between the historical and contemporary samples and (3) shifts in allele frequencies at climate-linked loci in response to changing climate conditions over the past century. We further validate the WGS results by genotyping 68 additional historical samples at a subset of markers designed to identify populations.

Genome-wide introgression into San Diego

Using WGS, we examined patterns of genomic differentiation between historical (n = 17) and contemporary (n = 18) willow flycatchers from San Diego and interpreted the observed patterns within the context of 221 contemporary individuals from across the willow flycatcher breeding range (Fig. 1a and Supplementary Tables 1 and 2). Principal component analysis (PCA) of the contemporary samples revealed three genomic clusters, corresponding to (1) the desert southwest, (2) the eastern subspecies and (3) remaining individuals from California, the Pacific northwest and the interior west (Fig. 1b; n = 128,775 loci, 238 individuals). While individuals sampled in San Diego in the late 1800s were more genetically distinct, allele frequencies in modern San Diego samples are now intermediate to those of the Pacific northwest and desert southwest (Fig. 1b). Mean fixation index (F_{ST}) between the historical and contemporary San Diego populations was 0.06, comparable to F_{ST} estimates between historical San Diego and neighbouring contemporary populations in the Pacific northwest (0.06) and desert southwest (0.05). Mapping of the sequence data to the zebra finch (*Taeniopygia guttata*) genome revealed that the F_{ST} outlier loci (F_{ST} > 0.38; 99th percentile of F_{ST} divergence) between the historical and contemporary individuals from San Diego were scattered throughout the genome (Extended Data Fig. 1).

Individual admixture proportions estimated with DyStruct showed a similar pattern to the genome-wide PCA (Fig. 1c). Although K = 4 and K = 5 had similar support (log likelihood = −2,482.325 and −2,482.174, respectively) (Supplementary Fig. 1), admixture proportions under K = 4 provided additional information about patterns of gene flow between San Diego and neighbouring populations (Fig. 1c and

https://doi.org/10.1038/s41558-023-01696-3

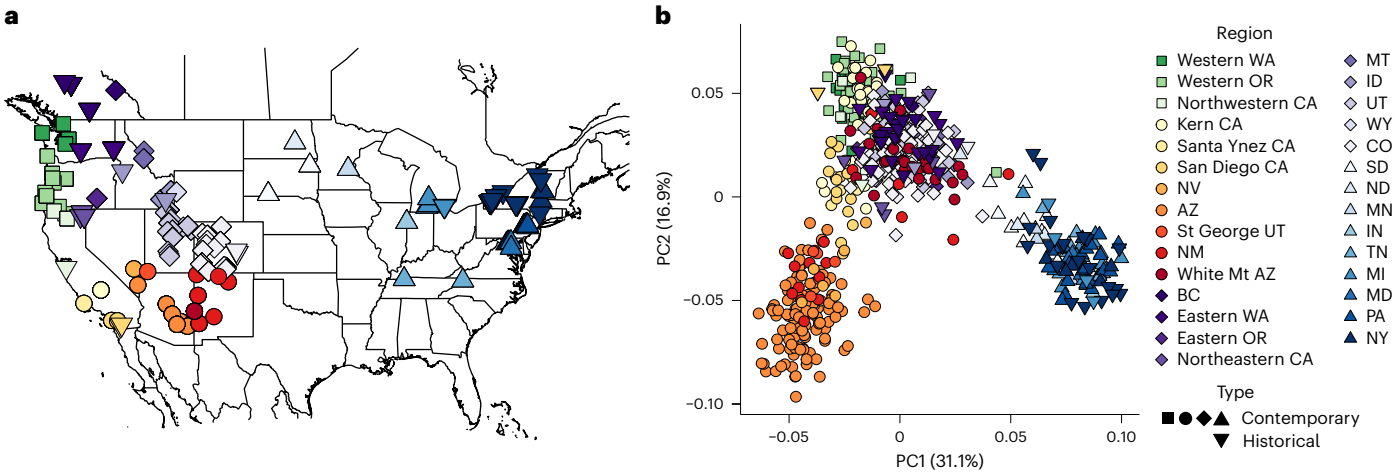


Fig. 2 | Geographic sampling and genomic differentiation from SNP genotyping data. a, Sampling locations of 68 historical (1901–1958) and 548 contemporary (1996–2016) breeding willow flycatcher samples included in the SNP genotyping analysis[31]. Contemporary samples of the four subspecies are shown as squares, circles, diamonds and triangles, while upside-down triangles indicate historical samples. **b**, PCA of the genome-wide covariance matrix (n = 616 individuals genotyped at 96 loci). Points are coloured by sampling location and correspond to the map in **a**.

Supplementary Figs. 2–8). Under K = 4, the contemporary San Diego population consisted of genetic variation from the Pacific northwest, the desert southwest and historical San Diego (Fig. 1c). Furthermore, D statistics revealed excess allele sharing between contemporary San Diego and neighbouring populations in the Pacific northwest and desert southwest relative to populations in the interior west and east, providing additional support for introgression between San Diego and neighbouring populations (Extended Data Table 1). Historical San Diego remained distinct under all values of K (Supplementary Figs. 2–8), potentially because contemporary populations have had the opportunity for admixture over the past century.

To evaluate whether population structure in the willow flycatcher has shifted outside of San Diego and verify the patterns observed in the WGS data, we genotyped additional historical (n = 68; 1901–1958) and contemporary (n = 548; 1996–2016) samples from across the breeding range at 96 single nucleotide polymorphisms (SNPs) pre-identified as providing high power for population assignment[20] (Fig. 2a and Supplementary Tables 2 and 3). PCA and STRUCTURE results for K = 2–7 revealed similar population structure in the past and present for geographically adjacent populations other than San Diego. Similar to the WGS results, historical San Diego clustered higher along PC2 (Fig. 2b) and formed a distinct genetic cluster from contemporary San Diego samples (Supplementary Figs. 9–14). Concordance between the two independent analyses supports the idea that allele frequency shifts in San Diego are not an artefact of historical sequencing methods.

We further investigated the impact of gene flow from surrounding populations on standing genetic variation within San Diego by calculating various population genomic parameters for the historical and contemporary samples. Our findings indicated that the shift in genome-wide population structure was accompanied by an increase in effective population size (N_e), a decrease in inbreeding (F) and an increase in neutral genetic diversity (observed heterozygosity and proportion of polymorphic loci) over time (Table 1). In contrast, the inbreeding coefficient based on runs of homozygosity (F_{ROH}) was slightly higher among contemporary samples, a pattern consistent with recent population declines.

Table 1 | Change in population genomic parameters over the past century

	Past (1888–1909)	Present (1996–2015)
Effective population size (N_e)	23.1–36.9	75.6–112.6
Inbreeding (F)	0.0587	−0.0327
Inbreeding (F_{ROH})	0.0104	0.0077
Runs of homozygosity	18 runs <1 Mb, 1 run 4–8 Mb	36 runs <1 Mb
Neutral genetic diversity:		
H_{obs} at non-climate-linked loci	0.0520	0.1816
Proportion of polymorphic loci	0.3620	0.8363
Adaptive potential:		
H_{obs} at climate-linked loci	0.1086	0.2868
Proportion of polymorphic loci	0.5000	0.9000

Change in effective population size (N_e), inbreeding (F and F_{ROH}) and genetic diversity (observed heterozygosity (H_{obs}) and proportion of polymorphic loci) at putatively neutral and climate-linked SNPs over time for willow flycatchers breeding in San Diego (historical, n=17; contemporary, n=18).

Increase in adaptive potential over time

To assess the impact of gene flow into San Diego on the amount of genetic variation present in climate-linked loci, we ran gradient forest with contemporary samples to identify climate variables most strongly associated with genomic variation. We confined this analysis to the southwestern subspecies to limit the confounding effect of population structure and identify environmental variables that might drive selection within the southwestern United States. The environmental variables that exhibited the strongest associations with genomic variation across the southwestern willow flycatcher breeding range were: monthly mean dew point temperature (tdmean; associated with moisture content in the air), monthly precipitation (precip) and monthly maximum temperature (tmax) (Supplementary Table 4 and Extended Data Fig. 2). The gradient forest model explained more variation than expected by chance; while 89,562 SNPs were associated with environment under the gradient forest model (mean R^2 = 0.188), only 41,154–68,644 SNPs were correlated under ten randomized environments (mean R^2 = 0.094–0.127).

Using the top three climate variables from gradient forest, we identified candidate SNPs linked to these environmental variables by carrying out two types of genotype–environment association (GEA) analyses. A partial redundancy analysis (RDA) was significant (P = 0.001) and identified 340 candidate SNPs associated with tdmean,

https://doi.org/10.1038/s41558-023-01696-3

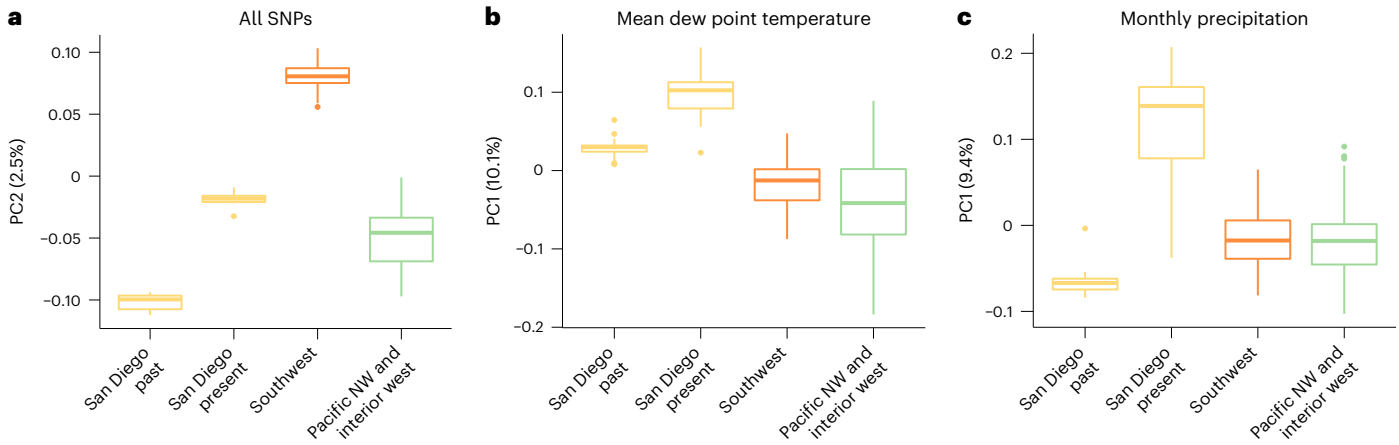


Fig. 3 | Genomic differentiation over time at climate-linked SNPs. **a**–**c**, Loadings from a PCA of all loci (n = 128,775) (**a**) and loci associated with mean dew point temperature (n = 104) (**b**) and precipitation (n = 72) (**c**) across 238 willow flycatcher samples. Loadings along PC2 (all loci) and PC1 (climate-linked loci) were extracted from historical and contemporary San Diego populations and the neighbouring genetic clusters in Fig. 1b. The boxplots show the median (central line), the interquartile range (IQR; box) and 1.5× IQR from the first and third quartiles (lower and upper whiskers).

338 SNPs associated with precip and 294 SNPs associated with tmax (Extended Data Figs. 3 and 4). The three significant axes explained 45%, 30% and 25%, respectively, of the variance. In contrast, latent factor mixed models (LFMM) detected 745 SNPs associated with tdmean, 536 SNPs linked to precip and 1,120 SNPs associated with tmax. The intersection of these two approaches retained 232 loci: 104 linked to tdmean, 72 associated with precip and 56 SNPs associated with tmax. Of these loci, 119 (51%) were located in non-coding regions, 3 (1.3%) were located in coding regions and the remaining 110 (47%) were located in intergenic regions. While the functions of these climate-linked loci are unknown, we further investigated their association with climate by generating a more restricted dataset that only contained SNPs located within 25 kilobases (kb) of genes related to thermal tolerance and bill morphology in birds (Supplementary Methods). This smaller dataset included 20 loci linked to tdmean, ten precipitation-associated SNPs and seven loci linked to tmax (Supplementary Table 5). Many of these genes are not well characterized; however, their detection in our dataset provides additional evidence that our identified candidate loci are important for climate adaptation. Similar to neutral genetic diversity, adaptive potential, quantified by measuring observed heterozygosity and proportion of polymorphic loci at the 232 identified climate-linked loci from the GEAs, was higher in the contemporary San Diego samples (Table 1).

Shift in climate-linked loci within San Diego

When plotted in separate PCAs, the candidate loci associated with tdmean and precip, the top-ranking climate variables from the gradient forest analysis, shifted in a different direction relative to the rest of the genome over the past century (Fig. 3 and Extended Data Fig. 5). While the San Diego population shifted towards the desert southwest on a genome-wide level (Fig. 3a), SNPs associated with tdmean (Fig. 3b and Extended Data Fig. 5a) and precip (Fig. 3c and Extended Data Fig. 5b) shifted away from the Pacific northwest and desert southwest along PC1 over time. In contrast, loci associated with tmax did not shift from the past to the present along PC1 (Extended Data Fig. 5c). While we detected excess allele sharing at climate-linked loci between contemporary San Diego and neighbouring populations, this pattern was only significant when comparing the Pacific northwest and contemporary San Diego relative to populations in the east (Extended Data Table 1). D statistics for climate-linked loci should be interpreted with caution given the limited number of ABBA and BABA patterns detected (Extended Data Table 1).

Genomic shifts indicate adaptation to past climate change

To test whether the increase in adaptive potential and change in population structure at climate-linked SNPs reflect adaptation to climate change over the past century, we directly assessed how allele frequencies at the detected climate-associated loci have shifted within San Diego in response to the increase in tdmean (+1.41 °C), increase in precipitation (+0.36 mm) and decrease in tmax (−0.59 °C) in southern California since the late 1800s.

In the complete dataset of intersected SNPs associated with tdmean (n = 104), allele frequencies shifted in the expected direction (increased for positively associated major alleles and decreased for negatively associated major alleles) in 75 (72%) of 104 SNPs (Fig. 4a; one-tailed binomial test, $P < 0.0001$; allele frequency shift, 0.21 ± 0.16; mean ± s.d.). This pattern was also observed in the restricted dataset of SNPs located near genes involved in avian thermal tolerance (n = 20), which shifted in the expected manner in 16 (80%) of 20 SNPs (Extended Data Fig. 6a; one-tailed binomial test, P = 0.006).

The shift in monthly precipitation has resulted in a similar evolutionary response, with allele frequencies shifting in the expected direction in 51 (71%) of 72 SNPs (Fig. 4b; one-tailed binomial test, P = 0.0003; allele frequency shift, 0.24 ± 0.14; mean ± s.d.). Despite the small sample size in the restricted dataset of precipitation-associated loci (n = 10), allele frequencies shifted as expected in eight (80%) of ten SNPs (Extended Data Fig. 6b; one-tailed binomial test, P = 0.055).

Finally, allele frequencies have not shifted in a consistent direction over time in response to the decrease in tmax, the lowest-ranking climate variable from the gradient forest analysis. Allele frequencies only shifted in the expected manner (decreased for positively associated major alleles and increased for negatively associated major alleles) in 24 (43%) of the 56 associated SNPs (Supplementary Fig. 15; one-tailed binomial test, P = 0.89; allele frequency shift, 0.12 ± 0.11; mean ± s.d.). This lack of a clear shift was also evident in the restricted dataset (Extended Data Fig. 6c; one-tailed binomial test, P = 0.5, n = 7). That our broader set of climate-linked loci and restricted dataset of SNPs located near genes linked to thermal tolerance and bill morphology revealed similar patterns in each comparison suggests that the possible detection of false positives in our GEAs did not severely bias our results.

Discussion

We combine landscape genomic methods and WGS of historical and contemporary samples to examine the capacity for and extent of

https://doi.org/10.1038/s41558-023-01696-3

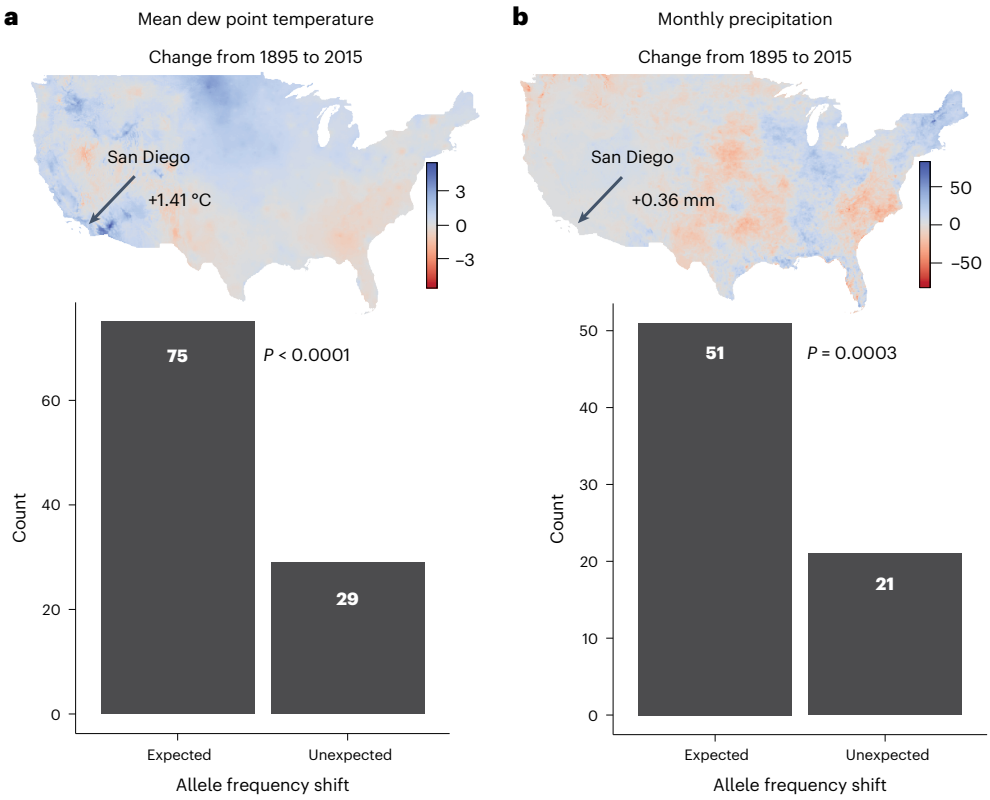


Fig. 4 | Allele frequency shifts at climate-linked loci within San Diego. **a**,**b**, Number of loci linked to mean dew point temperature ($n = 104$) (**a**) and precipitation (n = 72) (**b**) that shifted in the expected direction in willow flycatchers in response to the increase in dew point temperature and precipitation in southern California over the past century. The maps indicate how each climate variable has changed from 1895 to 2015 across the United States. The *P* values examine whether the observed proportion of loci that shifted in the expected direction (consistent with climate adaptation) was significantly greater than expected by chance (one-tailed binomial test).

evolutionary adaptation to climate change over the past century in the endangered southwestern willow flycatcher. We documented a shift in genome-wide population structure within southern California, with the San Diego population becoming more genetically similar to individuals in the Pacific northwest and desert southwest over time. This result was replicated via SNP genotyping of historical and contemporary populations across North America, which showed that, apart from San Diego, where historical and contemporary individuals cluster separately in the PCA, population structure has remained stable throughout the twentieth century. Introgression analyses indicated that San Diego probably experienced gene flow from the Pacific northwest and desert southwest in recent history, possibly following recorded population declines in the twentieth century[17]. The loss of critical riparian habitat, particularly in the southwestern United States, after the turn of the century may have led to increased dispersal and facilitated gene flow into southern California. Correspondingly, effective population size was lower and inbreeding (*F*) was higher in the historical San Diego population, as expected if admixture from neighbouring populations has increased adaptative potential. In contrast, the higher F_{ROH} estimate for contemporary San Diego is in keeping with recent population declines, as runs of homozygosity more accurately capture recent levels of inbreeding.

Gene flow is known to increase population fitness by introducing new genetic variation into populations that natural selection can act upon to adapt to changing environmental conditions[11,21]. Given the benefits of admixture for population viability, the human-mediated transfer of genetic material between populations (genetic rescue or assisted gene flow) is sometimes used as a conservation strategy to maintain evolutionary potential and prevent the loss of small, imperilled populations[10,22]. In San Diego, we documented an increase in genetic diversity at both neutral and climate-linked loci following gene flow from the Pacific northwest and desert southwest. Functional and neutral genetic variation are both necessary to mitigate inbreeding depression, facilitate genetic responses to new or changing environments and ensure population resilience[8,23]. By increasing neutral genetic diversity and adaptive potential within San Diego, admixture may have strengthened the capacity of this willow flycatcher population to adapt to environmental change[21,24–27].

Consistent with this idea, we found that allele frequencies at candidate loci linked to the two top-ranking climate variables from the gradient forest analysis—mean dew point temperature and precipitation—have shifted away from neighbouring populations and in a direction consistent with adaptation to climate change in southern California. Growing evidence indicates that natural selection can drive evolutionary change over just a few generations if there is existing genetic variation for a trait and the trait is under strong selection[3]. While the origin of these putatively adaptive alleles is uncertain, allele frequency shifts in the expected direction based on climate change without concurrent shifts across the genome support the idea that candidate SNPs associated with tdmean and precipitation are under selection and have not shifted over time due to admixture alone. This evolutionary response may have taken place through two non-mutually exclusive pathways. On one hand, the observed allele frequency shifts may have occurred through selection on standing genetic variation within San Diego. Allele frequencies at loci linked to either tdmean or precipitation ranged from 0 to 0.79 within the historical San Diego population, suggesting that standing variation did exist for many climate-linked loci. On the other hand, selection may have acted on climate-linked alleles that introgressed into

San Diego over the past century. *D* statistics revealed a significant pattern of excess allele sharing at climate-linked loci between contemporary San Diego and the Pacific northwest relative to the eastern subspecies. Thus, although these climate-linked alleles may have been neutral at the time of admixture, gene flow potentially introduced genetic variation needed for selection to shift allele frequencies in a direction consistent with climate adaptation. However, we acknowledge that this analysis only included two time points and further temporal sampling would increase our ability to robustly link the observed shifts in allele frequencies to climate change since the late 1800s.

Methods for predicting climate vulnerability in wild populations often rely on the assumption that allele frequencies will shift over time in response to environmental change[28,29]. Our results help validate this widespread assumption and indicate that the possible influence of gene flow from neighbouring populations on adaptive potential should be taken into account when attempting to predict species persistence in the face of climate change. Overall, our findings are encouraging, as they suggest that wild populations may have some capacity to cope with climate change through adaptation over a century-long timescale. However, considering that climate change is occurring at an unprecedented rate and wild organisms face numerous other stressors that threaten their survival, including habitat destruction and poaching, this rate of adaptation may be insufficient to ensure population persistence. In the case of the southwestern willow flycatcher, for example, populations in coastal California continue to decline despite climate adaptation, probably due to insufficient habitat, exacerbated in some areas by brood parasitism[17,30]. Our work, which indicates that admixture following population decline increased evolutionary capacity, supports the idea that genetic rescue efforts aimed at maintaining genome-wide genetic diversity could serve as an important management tool to conserve vulnerable species in a changing world.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41558-023-01696-3.

References

1. Hoffmann, A. A. & Sgrò, C. M. Climate change and evolutionary adaptation. *Nature* **470**, 479–485 (2011).
2. Razgour, O. et al. Considering adaptive genetic variation in climate change vulnerability assessment reduces species range loss projections. *Proc. Natl Acad. Sci. USA* **116**, 10418–10423 (2019).
3. Catullo, R. A., Llewelyn, J., Phillips, B. L. & Moritz, C. C. The potential for rapid evolution under anthropogenic climate change. *Curr. Biol.* **29**, R996–R1007 (2019).
4. Merilä, J. & Hendry, A. P. Climate change, adaptation, and phenotypic plasticity: the problem and the evidence. *Evol. Appl* **7**, 1–14 (2014).
5. Bi, K. et al. Temporal genomic contrasts reveal rapid evolutionary responses in an alpine mammal during recent climate change. *PLoS Genet.* **15**, e1008119 (2019).
6. Geerts, A. N. et al. Rapid evolution of thermal tolerance in the water flea *Daphnia*. *Nat. Clim. Change* **5**, 665–668 (2015).
7. Umina, P. A., Weeks, A. R., Kearney, M. R., McKechnie, S. W. & Hoffmann, A. A. A rapid shift in a classic clinal pattern in *Drosophila* reflecting climate change. *Science* **308**, 691–693 (2005).
8. Kardos, M. et al. The crucial role of genome-wide genetic variation in conservation. *Proc. Natl Acad. Sci. USA* **118**, e2104642118 (2021).
9. Eizaguirre, C. & Baltazar-Soares, M. Evolutionary conservation—evaluating the adaptive potential of species. *Evol. Appl.* **7**, 963–967 (2014).
10. Aitken, S. N. & Whitlock, M. C. Assisted gene flow to facilitate local adaptation to climate change. *Annu. Rev. Ecol. Evol. Syst.* **44**, 367–388 (2013).
11. Bell, D. A. et al. The exciting potential and remaining uncertainties of genetic rescue. *Trends Ecol. Evol.* **34**, 1070–1079 (2019).
12. Chan, W. Y., Hoffmann, A. A. & van Oppen, M. J. H. Hybridization as a conservation management tool. *Conserv. Lett.* **12**, e12652 (2019).
13. Billerman, S. M. & Walsh, J. Historical DNA as a tool to address key questions in avian biology and evolution: a review of methods, challenges, applications, and future directions. *Mol. Ecol. Resour.* **19**, 1115–1130 (2019).
14. Sogge, M. K., Marshall, R. M., Sferra, S. J. & Tibbitts, T. J. *A Southwestern Willow Flycatcher Natural History Summary and Survey Protocol*. Technical Report NPS/NAUCPRS/NRTR-97/12 (USGS, 1997).
15. Diffenbaugh, N. S., Giorgi, F. & Pal, J. S. Climate change hotspots in the United States. *Geophys. Res. Lett.* **35**, L16709 (2008).
16. Ruegg, K. et al. Ecological genomics predicts climate vulnerability in an endangered southwestern songbird. *Ecol. Lett.* **21**, 1085–1096 (2018).
17. *Southwestern Willow Flycatcher* (Empidonax traillii extimus). *5-Year Review: Summary and Evaluation* (US Fish and Wildlife Service, 2014).
18. Unitt, P. *Empidonax traillii extimus*: an endangered subspecies. *West. Birds* **18**, 137–162 (1987).
19. *Indicators of Climate Change in California* 4th edn (California Environmental Protection Agency, OEHHA, 2022).
20. Ruegg, K. et al. Linking climate niches across seasons to assess population vulnerability in a migratory bird. *Glob. Change Biol.* **27**, 3519–3531 (2021).
21. Hamilton, J. A. & Miller, J. M. Adaptive introgression as a resource for management and genetic conservation in a changing climate. *Conserv. Biol.* **30**, 33–41 (2016).
22. Whiteley, A. R., Fitzpatrick, S. W., Funk, W. C. & Tallmon, D. A. Genetic rescue to the rescue. *Trends Ecol. Evol.* **30**, 42–49 (2015).
23. DeWoody, J. A., Harder, A. M., Mathur, S. & Willoughby, J. R. The long-standing significance of genetic diversity in conservation. *Mol. Ecol.* **30**, 4147–4154 (2021).
24. Lavergne, S. & Molofsky, J. Increased genetic variation and evolutionary potential drive the success of an invasive grass. *Proc. Natl Acad. Sci. USA* **104**, 3883–3888 (2007).
25. Nolte, A. W., Freyhof, J., Stemshorn, K. C. & Tautz, D. An invasive lineage of sculpins, *Cottus* sp. (Pisces, Teleostei) in the Rhine with new habitat adaptations has originated from hybridization between old phylogeographic groups. *Proc. R. Soc. B* **272**, 2379–2387 (2005).
26. Racimo, F., Sankararaman, S., Nielsen, R. & Huerta-Sánchez, E. Evidence for archaic adaptive introgression in humans. *Nat. Rev. Genet.* **16**, 359–371 (2015).
27. Becker, M. et al. Hybridization may facilitate in situ survival of endemic species through periods of climate change. *Nat. Clim. Change* **3**, 1039–1043 (2013).
28. Bay, R. A. et al. Genomic signals of selection predict climate-driven population declines in a migratory bird. *Science* **359**, 83–86 (2018).
29. Fitzpatrick, M. C. & Keller, S. R. Ecological genomics meets community-level modelling of biodiversity: mapping the genomic landscape of current and future environmental adaptation. *Ecol. Lett.* **18**, 1–16 (2015).

30. Kus, B. E., Beck, P. P. & Wells, J. M. Southwestern willow flycatcher populations in California: distribution, abundance, and potential for conservation. *Stud. Avian Biol.* **26**, 12–21 (2003).
31. *Bird Species Distribution Maps of the World* (BirdLife International & NatureServe, 2012).

Methods

Sample collection and DNA extraction

We compiled blood samples from 225 contemporary individuals of the willow flycatcher from a combination of previous studies, museum donations and field collections[16,32]. This contemporary dataset was collected from 51 locations across the breeding range, spanned 1996–2015 and included 18 individuals sampled near and within San Diego County, California from 2010 to 2011. In addition, we acquired either toepads or 2 × 4 mm^2 pieces of apterium skin from 17 historical specimens that were collected from 1888 to 1909 within 1° latitude and longitude of the contemporary San Diego samples and preserved as dried skins (Supplementary Table 2).

Library preparation and whole-genome sequencing

DNA from the contemporary samples was extracted using the Qiagen DNeasy Blood and Tissue Kit and quantified with the Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific). We used a modified version of Illumina's Nextera Library Preparation protocol to prepare WGS libraries and pooled the libraries by equal mass before sequencing. Ten of the historical samples were extracted in a separate dedicated laboratory using a phenol-chloroform protocol, while the remaining seven samples were extracted by Arbor Biosciences as part of their myReads NGS services. Single-stranded degraded DNA-specific library preparation and pooling of all 17 historical samples was carried out by Arbor Biosciences.

The resulting libraries from each library preparation method were sequenced on four lanes of an Illumina NovaSeq 6000 (three lanes for the contemporary samples and one lane for the historical samples) at the Center for Genomic and Computational Biology at Duke University. Fifteen contemporary samples that failed in the initial sequencing run were resequenced on a partial lane of an Illumina NovaSeq 6000 at Novogene Corporation.

Variant calling

To process the WGS data from the contemporary samples, we removed low-quality fragments and adaptors with Trimgalore v.0.6.6 using default parameters and paired-end validation (https://github.com/FelixKrueger/TrimGalore). We then used BWA v.0.7.3a mem to align the trimmed reads to an existing southwestern willow flycatcher genome assembly (NCBI BioProject PRJNA437496)[16,33]. We added read groups with Picard v.2.23.2 and sorted bam files, corrected mate information and marked duplicates with samtools v.1.13.0 (refs. 34,35).

Historical DNA is often subject to DNA fragmentation, low concentrations and postmortem damage[13]. As a result, we followed a different pipeline to address some of these challenges while processing the historical sequencing data. Following sequencing, we used SeqPrep2 (https://github.com/jeizenga/SeqPrep2) to remove adaptors and merge paired-end reads that had a minimum length of 28 base pairs (bp) and were overlapping by >19 bp. We then filtered out low-complexity reads with Prinseq-lite v.0.20.4 using the DUST approach and recommended complexity threshold of seven (ref. 36). The filtered historical data were aligned to the willow flycatcher reference genome using the aln (-l 1024) and samse commands in BWA. We used samtools to convert sam to bam files with a minimum mapping quality score of 30 and sort, remove duplicates and index the resulting bam files[34]. To account for DNA degradation in the historical samples, we also used mapDamage2 to rescale base quality scores in accordance with their probability of being damaged[37].

We estimated coverage for both contemporary and historical individuals with samtools and called variants by estimating genotype likelihoods (-doGlf) in ANGSD 0.939, given the low coverage of the historical sequencing data[38]. Genotype likelihoods were calculated using the samtools method (-GL 1) and BAQ computation[39] and we adjusted the mapping quality for excessive mismatches (-C 50). We inferred major and minor alleles from the genotype likelihoods (-doMajorMinor 1), discarded 'bad' reads (-remove_bads 1) and reads that did not map uniquely (-uniqueOnly 1) and only retained variable sites with a minor allele frequency >0.05 (-minMaf 0.05) and $P < 1 \times 10^{-6}$ (-SNP_pval 1e-6). We converted the BCF output from ANGSD to VCF format and removed indels and SNPs that were not biallelic.

We used vcftools v.0.1.16 to generate two final files for subsequent analyses[40]. For analyses using solely contemporary data (gradient forest), we filtered out sites with >25% missing data across all 242 samples (–max-missing 0.75) and removed loci with a maximum mean depth of 5 (–max-meanDP 5). This filtering process yielded 3,757,986 sites with an average of 8% missing data per individual across 225 contemporary samples. For analyses requiring data from both contemporary and historical samples, we identified sites in the historical dataset with >25% missing data and filtered out these sites for both contemporary and historical individuals. In addition, we removed loci with a maximum mean depth of 7.5 (–max-meanDP 7.5). This pipeline yielded 128,775 SNPs across 242 individuals with an average of 9% missing data per individual and comparable coverage in both the contemporary (3.28×) and historical (3.02×) datasets. By following the same filtering steps for historical and contemporary samples, subsetting the data to fewer loci with comparable levels of missing data and dedicating more sequencing effort towards historical samples, as recommend by ref. 41, we have aimed to minimize any biases resulting from the joint analysis of historical and modern DNA.

Whole-genome population structure

We assessed changes in genome-wide population structure over time by running a PCA in PCAngsd 1.10 with 128,775 SNPs and all historical and contemporary samples[42]. PCAngsd works with genotype likelihoods and thus incorporates uncertainty due to unobserved genotypes in low-coverage WGS data. Four contemporary individuals had genomic PC1 scores that were outside the expected range, probably due to contamination or the mislabelling of samples, and were excluded from further analyses. In addition, we estimated individual admixture proportions from the whole-genome data by running DyStruct v.1.1.0 with 25 independent repetitions of $K = 2–8$ (ref. 43). Unlike other model-based clustering programmes, DyStruct can infer shared genetic ancestry in time series data by explicitly incorporating temporal dynamics in allele frequency change. To assess support for different values of K, we held out a random subset of loci (5%) during training and subsequently compared the conditional log likelihoods outputted by the programme[43]. We calculated F_{ST} on a site-by-site basis and over 25 kb windows with vcftools and aligned scaffolds containing F_{ST} outliers (>99th percentile of F_{ST} divergence) to the zebra finch assembly (Taeniopygia_guttata-3.2.4) using the Satsuma synteny model from Satsuma2, assigning scaffolds to the chromosome with the top hit[44]. We then used the R package qqman[45] in R v.4.1.2 (ref. 46) to generate a Manhattan plot of genomic differentiation between historical and contemporary San Diego samples.

Introgression analysis

We used the R package admixr to estimate D statistics (ABBA-BABA method) with AdmixTools v.7.0.2 (refs. 47,48). Specifically, we quantified the number of derived alleles shared between the contemporary San Diego population and individuals in the Pacific northwest, desert southwest, interior west and east in different triplet combinations, with two alder flycatcher (*E. alnorum*) individuals comprising the outgroup (Supplementary Methods). In the absence of introgression, the focal population (contemporary San Diego) should share an equal number of derived alleles ($D = 0$) with individuals in the other two populations. However, past gene flow between San Diego and other populations will result in an excess of shared derived alleles (D value significantly different from 0). We calculated the significance of all D statistics using block jack-knifing to determine a Z-score distribution, where an absolute value of $Z > 3$ is equivalent to $P \sim 0.001$. We carried out this

analysis separately with all 128,775 SNPs, as well as with the 232 identified climate-linked SNPs discussed below (see section on Identification of candidate climate-linked loci).

SNP genotyping

We compiled toepads or skin punches from 68 additional museum specimens, collected from 1901 to 1958 from 19 locations across the breeding range of the willow flycatcher (Supplementary Table 2) and blood or feather samples from 548 contemporary samples from previous studies, museum donations and field collections. One historical individual and 182 contemporary birds in the SNP genotyping dataset were also included in the whole-genome dataset, while the remaining 366 contemporary individuals were new samples, collected from 36 additional locations across the willow flycatcher breeding range. DNA was extracted from the historical and contemporary samples as detailed above and all individuals were genotyped at 96 SNPs previously selected as informative for assigning individuals to genetically identifiable populations[20] using SNPtype Assays (Fluidigm) on a Fluidigm 96.96 IFC controller (Supplementary Table 3). All individuals had <20% missing data except for two historical samples from eastern Washington state, which were successfully genotyped at 79% of SNPs. To assess changes in population structure over time across the willow flycatcher breeding range, we ran a PCA with plink v.1.9 (ref. 49) and five independent runs of $K = 2–7$ with STRUCTURE 2.3.4 (burnin = 50,000, numreps = 150,000) to calculate individual admixture proportions for the contemporary and historical samples[50].

Population genomic parameters

We evaluated how standard population genomic parameters have changed over the past century in San Diego by estimating current and historical effective population sizes (N_e) from the whole-genome data with the LD method in NeEstimator2 (ref. 51). In addition, we used plink v.1.9 to calculate the F inbreeding coefficient[49] and the R package detectRUNS[52] to calculate F_{ROH} for both time periods. We identified runs of homozygosity in the historical and contemporary datasets using the consecutive runs method with the following parameters: minSNP = 50, maxGap = 10^6, minLengthBps = 10,000, maxOppRun = 2 and maxMissRun = 1. For the neutral (non-climate-linked) and putatively adaptive (climate-linked) loci, we estimated past and present observed heterozygosity using the R package hierfstat[53] and proportion of polymorphic loci (for loci with <5% missing data) within San Diego using Arlequin v.3.5.2.2 (refs. 54).

Gradient forest

To run the gradient forest analysis, we lumped contemporary samples collected within 1° latitude and longitude of one another into a single population and used ANGSD to estimate allele frequencies for each contemporary population containing at least four individuals (-minInd 4) from genotype probabilities (-doMaf 4), using the reference to determine the major allele (-doMajorMinor 4). This analysis yielded allele frequencies at 3,757,986 loci for 11 sites spanning the breeding range of the southwestern willow flycatcher. In addition, we downloaded monthly climate data from 1895 to the present from the Prism Climate Group (https://prism.oregonstate.edu). The climate dataset included seven environmental variables: precipitation, minimum temperature, mean temperature, maximum temperature, mean dew point temperature, minimum vapour pressure deficit and maximum vapour pressure deficit at a 4 km resolution. The climate variables were calculated on a daily basis and either summed (for precipitation) or averaged over all days in the month (for the remaining six variables). For each of the 11 sites with allele frequency data, we calculated the average value of each climate variable for the months of June and July, when willow flycatchers are stationary on the breeding grounds, over the time period from 1996 to 2015, when our contemporary samples were collected. If samples collected within 1° were lumped together, we extracted climate data for each individual sampling location and averaged the resulting values for each of the 11 sites.

We used the R package gradientForest to generate our gradient forest model (ntree = 500, nbin = 01, corr.threshold = 0.5) with 11 populations and a random subset of 500,000 loci[55]. This analysis generated a ranked list of environmental variables on the basis of their relative predictive power to explain observed genomic variation across the southwestern willow flycatcher breeding range. To assess whether our model explained more variation than expected by chance, we repeated this analysis ten times with random permutations of our environmental variables. We compared the number of SNPs with positive R^2 (predictive loci) and the mean R^2 of these predictive loci between our model and the ten runs with randomized environments (as in ref. 16). To visualize the results of our gradient forest model, we extracted environmental data from 100,000 random points across the breeding range of the southwestern willow flycatcher and predicted the genomic composition at each random point from our top three environmental variables (monthly mean dew point temperature, monthly precipitation and monthly maximum temperature). We summarized values with a PCA, assigning colours on the basis of the top three axes of variation to facilitate visualization of adaptive environments[55].

Identification of candidate climate-linked loci

We identified genomic loci in the contemporary and historical datasets that were strongly associated with the top-ranking uncorrelated climate variables (Pearson's $r < 0.7$) from our gradient forest analysis by carrying out two types of GEA analyses that account for underlying population structure: RDA and LFMM[56]. Both GEA methods require complete or imputed SNP data. We therefore used Beagle v.4.1 to impute genotypes for 221 contemporary samples from genotype likelihoods, excluding the 628 scaffolds that only contained a single SNP[57]. For each individual sampling location, we averaged climate data for our three top-ranking environmental variables for June and July over the time period from 1996 to 2015 (as above).

We ran a partial RDA with imputed genotypes at 128,147 loci across 221 contemporary individuals using the R package vegan[58], removing the effect of individual genomic PC1 scores to control for population structure. To test the significance of the model, we ran the anova.cca function with 999 permutations. Individual SNPs with loadings from the three significant constrained axes that were >3 s.d. from the mean (two-tailed $P = 0.0027$) were identified as candidate climate-linked SNPs[59]. We removed duplicate detections and determined the predictor variable that was most strongly correlated with each putatively adaptive SNP.

For the LFMMs, we ran five separate MCMC runs with three latent factors ($K = 3$) for each top-ranking climate variable using imputed genotypes at 128,147 loci for 221 contemporary individuals and the lfmm function in the R package LEA[60]. We ran each run for 6,000 cycles (-i 6000) with a burnin period of 3,000 (-b 3000). The number of latent factors was determined by the minimum cross-entropy criterion calculated for $K = 2–8$ with the LEA package and the number of discrete clusters in the PCA. We used the median to combine Z-scores from the five runs and adjusted P values for multiple tests by setting the expected false discovery rate (FDR) level to $q = 10\%$.

We obtained our final set of candidate climate-linked loci by taking the intersection of these two GEA approaches, retaining putatively adaptive SNPs that were detected in both the LFMM and RDA analyses. To further investigate our climate-linked SNPs, we used the programme Integrative Genomics Viewer to count the number of candidate SNPs located within coding and non-coding regions of genes[61]. In addition, we compiled a list of candidate genes linked to avian thermal stress and/or bill morphology (related to insect availability and heat dissipation) from the literature, as well as genes with gene ontology (GO) terms associated with aspects of thermal tolerance (response to heat and cold, temperature homoeostasis, respiratory system processes and response

to oxidative stress; Supplementary Methods). Following the methods of ref. 16, we used bedtools v.2.30.0 intersect[62] to annotate each SNP in the intersected dataset with genes located within 25 kb upstream or downstream, which is thought to be within the distance before which linkage disequilibrium should break down[63] and searched for the resulting genes in our list of published candidates and genes with relevant GO terms. This process allowed us to generate a more restricted list of candidate SNPs (n = 37) located near genes known to play a role in avian thermal tolerance and/or bill morphology[64–67] and therefore probably involved in climate adaptation (Supplementary Methods).

Climate adaptation over time

To determine whether allele frequencies at our identified climate-linked SNPs have shifted in response to climate change, we calculated the average value of each top climate variable for June and July from 1895 to 1909, the time period over which the historical samples were collected and assessed how each environmental variable has changed in San Diego from the past to the present (averaged over 1996–2015). In addition, we estimated historical allele frequencies within San Diego at each candidate climate-linked locus with ANGSD and calculated the change in allele frequency between the historical and contemporary samples. Finally, we used the contemporary dataset to calculate Pearson's correlation coefficient between the major allele at each candidate SNP and the associated climate variable to determine whether an increase or decrease in allele frequency at each locus was consistent with climate adaptation over time. For example, if precipitation has increased in San Diego over the past century, an increase in allele frequencies at major alleles that are positively associated with precipitation and a decrease in allele frequencies at negatively associated major alleles would be consistent with climate adaptation.

For each climate variable, we carried out a one-tailed binomial test to evaluate whether the observed proportion of loci that shifted in a direction consistent with adaptation to climate change was significantly greater than expected by chance. We carried out this analysis with our broader set of intersected SNPs (n = 232), as well as our restricted set of loci located near genes involved in thermal tolerance and/or bill morphology (n = 37), to verify that the patterns we observed in the broader dataset were not driven by the presence of false positives detected in our GEAs.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The data generated for this study are available in the NCBI Sequence Read Archive (BioProject PRJNA957938) and Dryad Digital Repository (https://datadryad.org/stash/share/qlafA02RwIPjoptjefUaKbrMszEsMJWP8a_H9-e24MY)[68].

Code availability

Code associated with this study can be accessed on GitHub (https://github.com/sturbek/WIFL-Climate-Adaptation)[69].

References

32. Paxton, E. H. *Molecular Genetic Structuring and Demographic History of the Willow Flycatcher*, Empidonax traillii (Northern Arizona Univ., 2000).
33. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
34. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
35. *Picard Toolkit* (Broad Institute, 2018); http://broadinstitute.github.io/picard/
36. Schmieder, R. & Edwards, R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **27**, 863–864 (2011).
37. Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F. & Orlando, L. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29**, 1682–1684 (2013).
38. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: analysis of next generation sequencing data. *BMC Bioinf.* **15**, 356 (2014).
39. Li, H. Improving SNP discovery by base alignment quality. *Bioinformatics* **27**, 1157–1158 (2011).
40. Danecek, P. et al. The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
41. Raxworthy, C. J. & Smith, B. T. Mining museums for historical DNA: advances and challenges in museomics. *Trends Ecol. Evol.* **36**, 1049–1060 (2021).
42. Meisner, J. & Albrechtsen, A. Inferring population structure and admixture proportions in low-depth NGS data. *Genetics* **210**, 719–731 (2018).
43. Joseph, T. A. & Pe'er, I. Inference of population structure from time-series genotype data. *Am. J. Hum. Genet.* **105**, 317–333 (2019).
44. Grabherr, M. G. et al. Genome-wide synteny through highly sensitive sequence alignment: Satsuma. *Bioinformatics* **26**, 1145–1151 (2010).
45. Turner, S. D. qqman: an R package for visualizing GWAS results using QQ and manhattan plots. *J. Open Source Softw.* **3**, 731 (2018).
46. R Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2020).
47. Patterson, N. et al. Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012).
48. Petr, M., Vernot, B. & Kelso, J. admixr—R package for reproducible analyses using ADMIXTOOLS. *Bioinformatics* **35**, 3194–3195 (2019).
49. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
50. Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959 (2000).
51. Do, C. et al. NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (N_e) from genetic data. *Mol. Ecol. Resour.* **14**, 209–214 (2014).
52. Biscarini, F., Cozzi, P., Gaspa, G. & Marras, G. detectRUNS: Detect runs of homozygosity and runs of heterozygosity in diploid genomes. R package version 0.9.6 (2019).
53. Goudet, J. hierfstat, a package for R to compute and test hierarchical *F*-statistics. *Mol. Ecol. Notes* **5**, 184–186 (2005).
54. Excoffier, L. & Lischer, H. E. L. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* **10**, 564–567 (2010).
55. Ellis, N., Smith, S. J. & Pitcher, C. R. Gradient forests: calculating importance gradients on physical predictors. *Ecology* **93**, 156–168 (2012).
56. Forester, B. R., Lasky, J. R., Wagner, H. H. & Urban, D. L. Comparing methods for detecting multilocus adaptation with multivariate genotype–environment associations. *Mol. Ecol.* **27**, 2215–2233 (2018).
57. Browning, B. L., Zhou, Y. & Browning, S. R. A one-penny imputed genome from next-generation reference panels. *Am. J. Hum. Genet.* **103**, 338–348 (2018).
58. Oksanen, J. et al. vegan: Community ecology package. R package version 2.6-4 (2022).
59. Forester, B. R., Jones, M. R., Joost, S., Landguth, E. L. & Lasky, J. R. Detecting spatial genetic signatures of local adaptation in heterogeneous landscapes. *Mol. Ecol.* **25**, 104–120 (2016).

60. Frichot, E. & François, O. LEA: an R package for landscape and ecological association studies. *Methods Ecol. Evol.* **6**, 925–929 (2015).
61. Robinson, J. T. et al. Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
62. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
63. Backström, N., Qvarnström, A., Gustafsson, L. & Ellegren, H. Levels of linkage disequilibrium in a wild bird population. *Biol. Lett.* **2**, 435–438 (2006).
64. Vitorino Carvalho, A. et al. Embryonic thermal manipulation impacts the postnatal transcriptome response of heat-challenged Japanese quails. *BMC Genom.* **22**, 488 (2021).
65. Kumar, H. et al. Transcriptome of chicken liver tissues reveals the candidate genes and pathways responsible for adaptation into two different climatic conditions. *Animals* **9**, 1076 (2019).
66. Zhang, J., Schmidt, C. J. & Lamont, S. J. Transcriptome analysis reveals potential mechanisms underlying differential heart development in fast- and slow-growing broilers under heat stress. *BMC Genom.* **18**, 295 (2017).
67. Wang, Y. et al. Liver transcriptome responses to heat stress and Newcastle disease virus infection in genetically distinct chicken inbred lines. *Genes* **11**, 1067 (2020).
68. Turbek, S. T. et al. Data for 'Historical DNA reveals climate adaptation in an endangered songbird'. Dryad https://datadryad.org/stash/share/qlafA02RwIPjoptjefUaKbrMszEsMJWP8a_H9-e24MY (2023).
69. Turbek, S. T. et al. Code for 'Historical DNA reveals climate adaptation in an endangered songbird'. GitHub https://github.com/sturbek/WIFL-Climate-Adaptation (2023).

Acknowledgements

This work was supported by a National Science Foundation (NSF) Postdoctoral Research Fellowship in Biology (2208881) to S.T. and an NSF CAREER grant (1942313) to K.R. We thank B. Forester for contributing funding for sequencing. We thank all of the individuals who contributed genetic samples, including T. Kita, B. Keith, R. Taylor and S. Birks. The following museums generously provided historical samples for the analyses presented in this study: the University of California Berkeley Museum of Vertebrate Zoology, California Academy of Sciences, Los Angeles County Natural History Museum, American Museum of Natural History, Buffalo Museum of Science, Charles R. Conner Museum, Cornell University Museum of Vertebrates, University of Kansas Biodiversity Institute & Natural History Museum, New York State Museum, Royal Ontario Museum, University of Colorado Museum of Natural History and the University of Michigan Museum of Zoology. All samples were collected under federal, state and IACUC permits necessary for the research. Any use of trade, firm or product names is for descriptive purposes only and does not imply endorsement by the US government.

Author contributions

S.T. and K.R. designed the study. B.K., M.W. and E.P. contributed samples for genomic analysis. R.B. and K.R. obtained museum samples. C.R. and C.G. generated the genomic data. S.T. and C.B. analysed the genomic data. K.R. funded the study. K.R. and T.S. provided logistical support. S.T. wrote the paper with edits from all authors.

Competing interests

The authors declare no competing interests.

Additional information

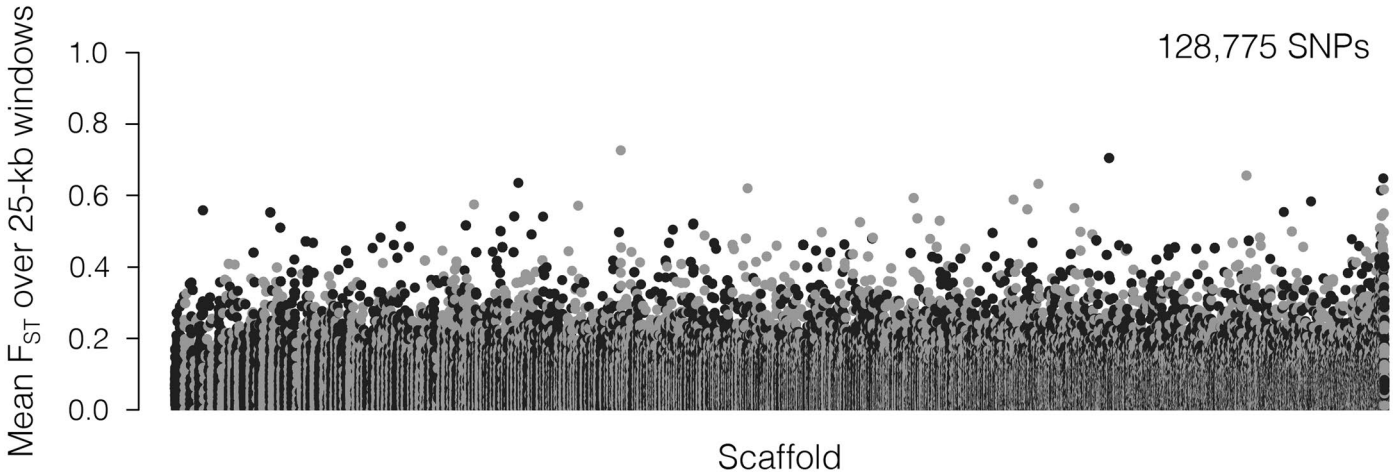
Extended data is available for this paper at https://doi.org/10.1038/s41558-023-01696-3.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41558-023-01696-3.

Correspondence and requests for materials should be addressed to Sheela P. Turbek.

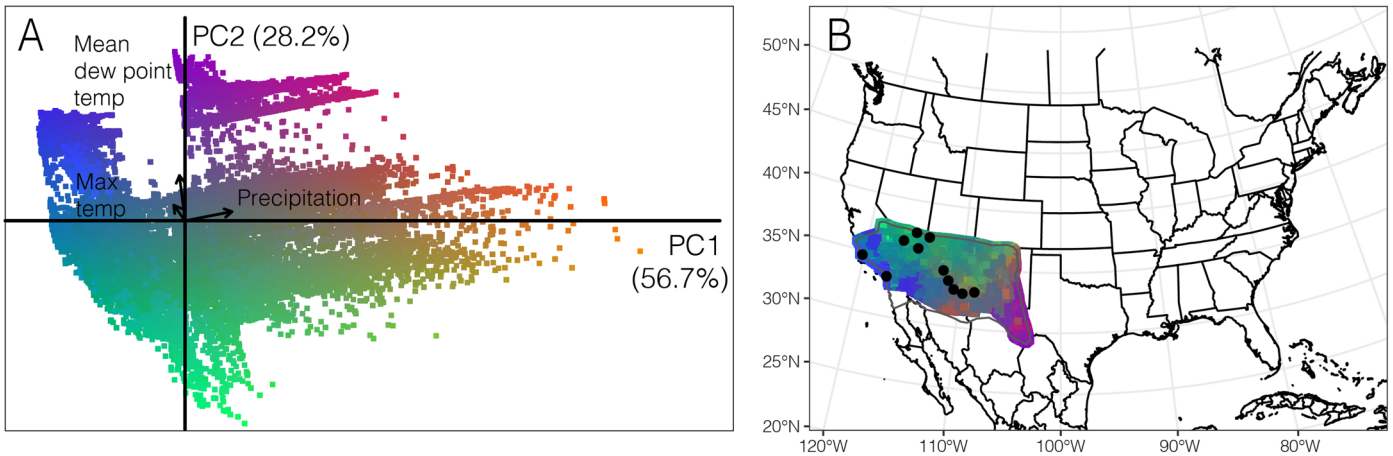
Peer review information *Nature Climate Change* thanks Fumin Lei, Orly Razgour and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

https://doi.org/10.1038/s41558-023-01696-3



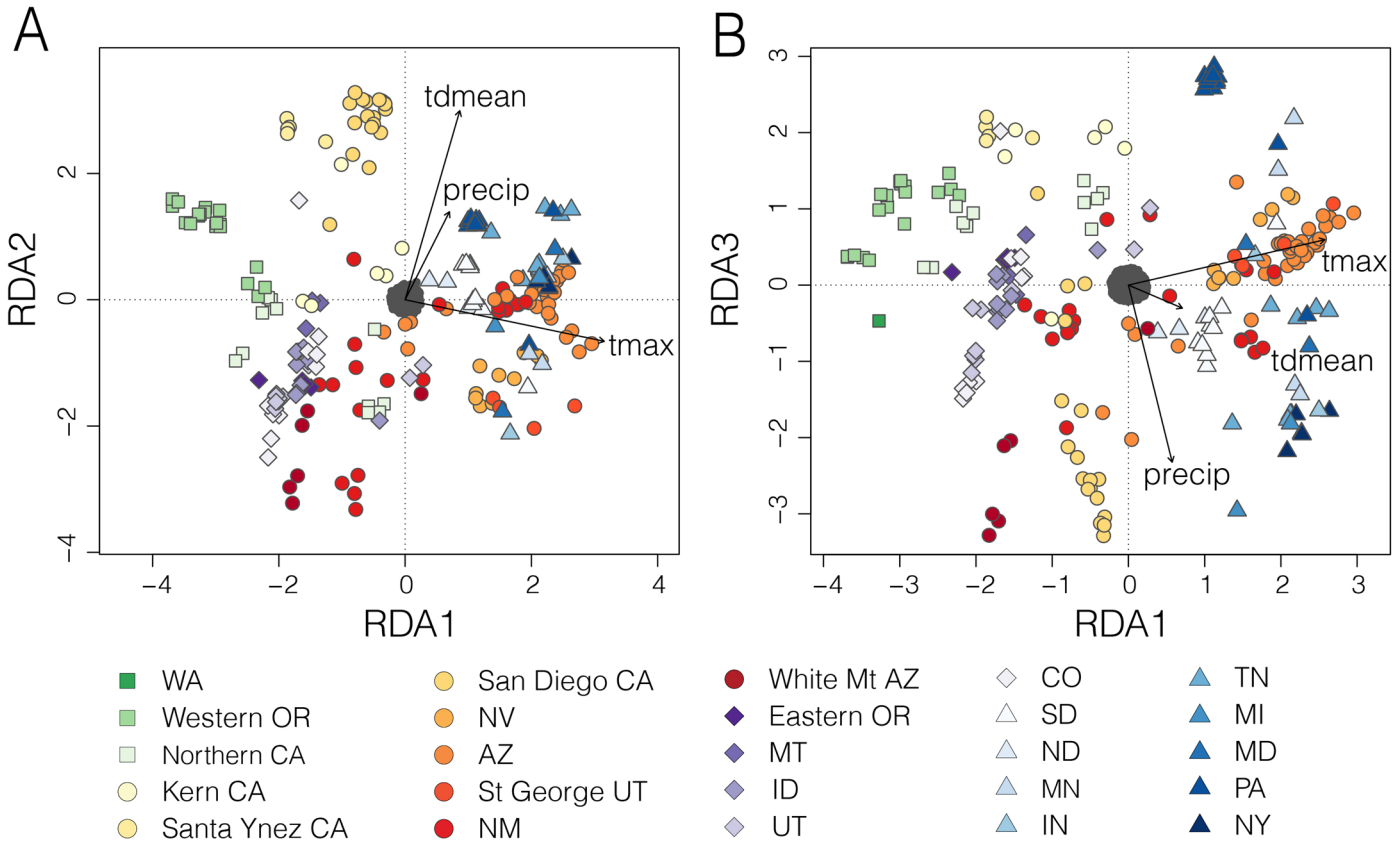
Extended Data Fig. 1 | Genomic differentiation between historical and contemporary willow flycatchers in San Diego. Manhattan plot showing mean F_{ST}, averaged over 25-kb windows, between historical (n = 17) and contemporary (n = 18) individuals breeding in San Diego. Colours (black and grey) alternate between scaffolds in the willow flycatcher reference genome.

https://doi.org/10.1038/s41558-023-01696-3



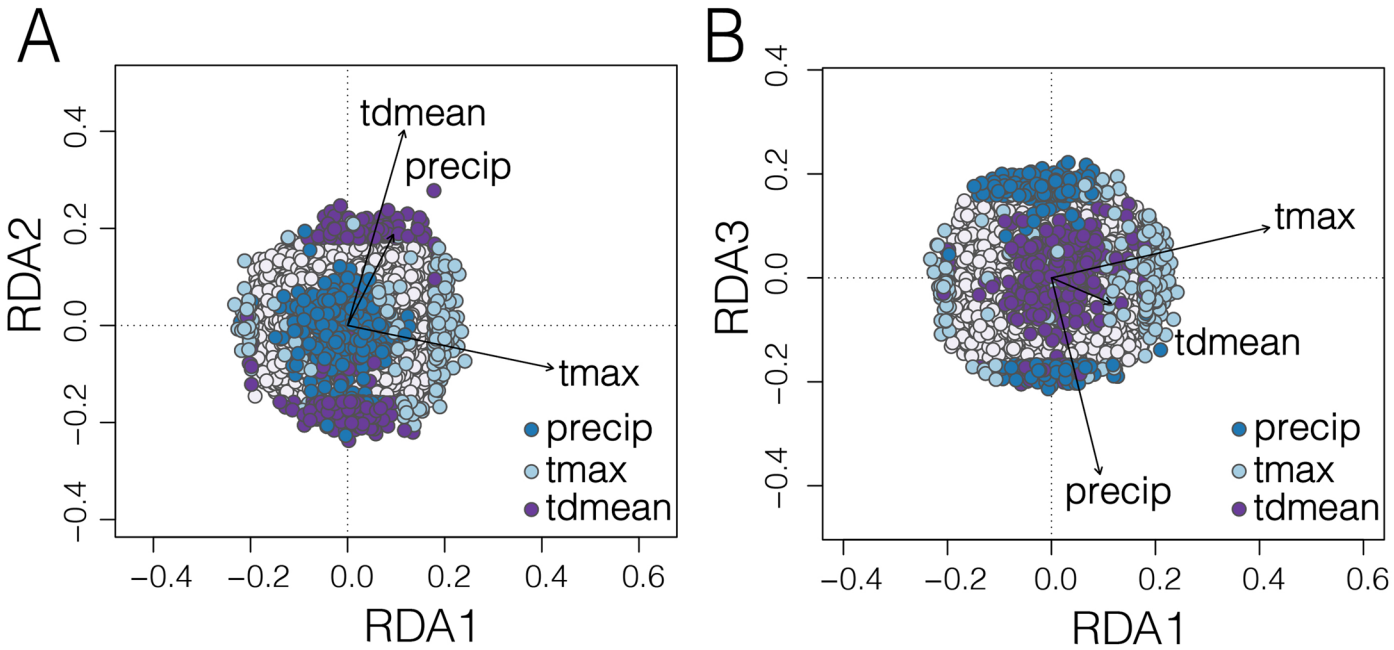
Extended Data Fig. 2 | Results from the gradient forest analysis of 500,000 loci and 11 contemporary populations across the breeding range of the southwestern willow flycatcher. (A) PCA of the top three climate variables associated with genomic variation. The arrows indicate the loadings of the top-ranking environmental variables. Colours are based on the predicted genomic composition from the top climate variables at 100,000 random points. **(B)** Gradient forest-transformed environmental variables from the PCA mapped across the breeding range of the southwestern willow flycatcher. Sampling locations are indicated with black circles.

https://doi.org/10.1038/s41558-023-01696-3



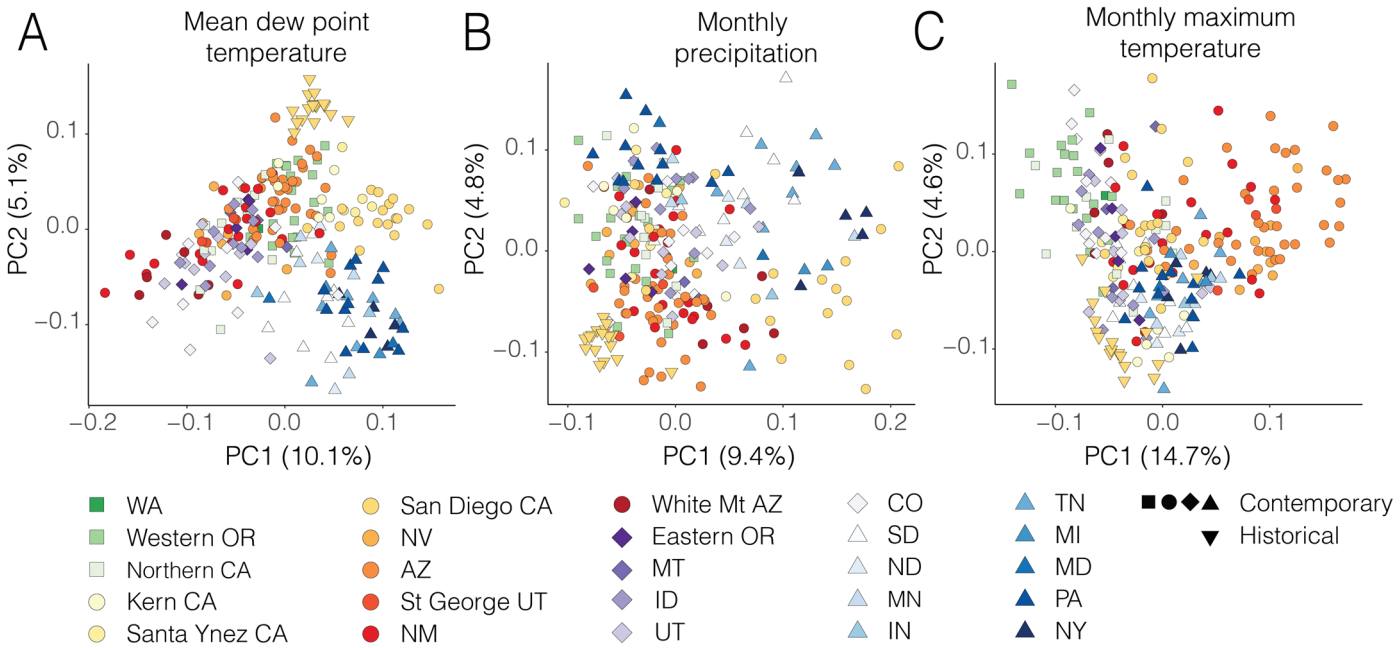
Extended Data Fig. 3 | Redundancy analysis (RDA) showing the relationship between genomic variation and top-ranking climate variables for 221 contemporary willow flycatchers breeding across the United States. The loadings on **(A)** axes one and two and **(B)** axes one and three for monthly mean dew point temperature (tdmean), monthly precipitation (precip), and monthly maximum temperature (tmax) are shown. Points are coloured by sampling location and the grey cloud in the center of each plot shows the 128,147 SNPs included in the analysis. The four subspecies of willow flycatcher are shown as different shapes.

https://doi.org/10.1038/s41558-023-01696-3



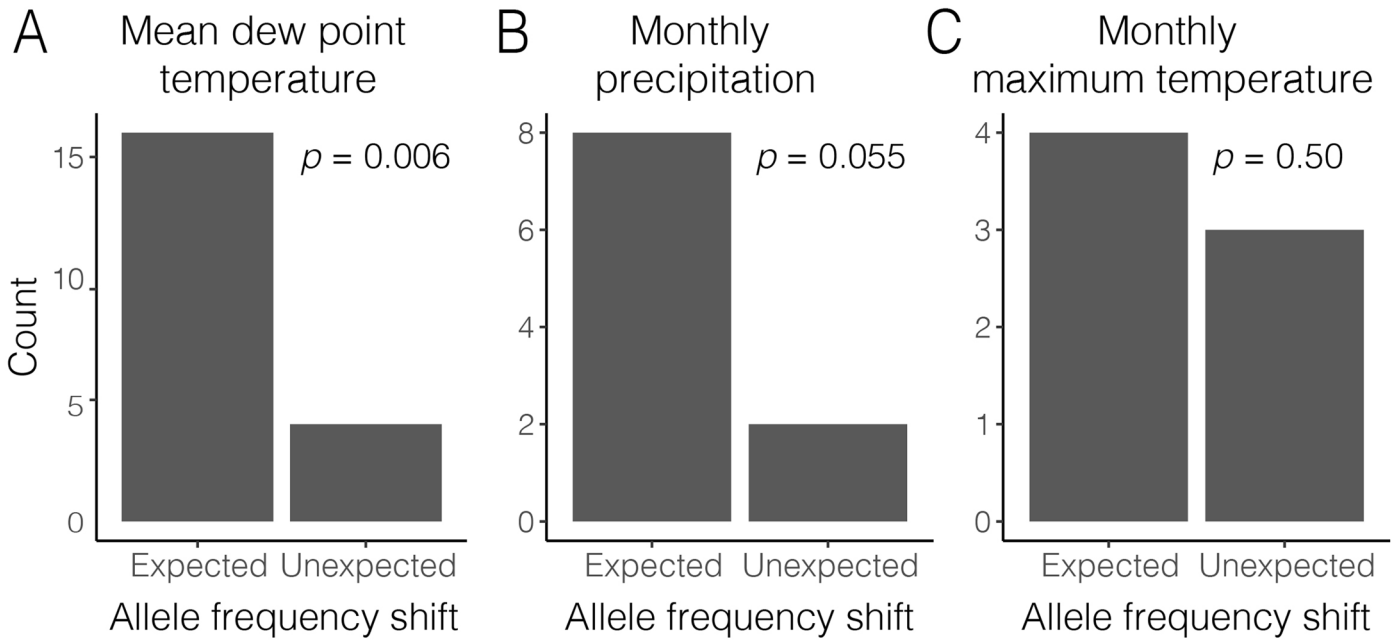
Extended Data Fig. 4 | Redundancy analysis (RDA) showing the candidate SNPs associated with top-ranking climate variables for 221 contemporary willow flycatchers breeding across the United States. The loadings of 128,147 SNPs on **(A)** axes one and two and **(B)** axes one and three and their relationship with monthly mean dew point temperature (tdmean), monthly precipitation (precip), and monthly maximum temperature (tmax) are shown. Candidate SNPs (with loadings greater than three standard deviations from the mean) are coloured according to their associations, while non-candidate SNPs are shown in light grey.

https://doi.org/10.1038/s41558-023-01696-3



Extended Data Fig. 5 | Shifts in patterns of genomic differentiation over time in willow flycatchers breeding across the United States at climate-linked SNPs. Principal component analysis (PCA) of loci associated with **(A)** monthly mean dew point temperature (n = 104), **(B)** monthly precipitation (n = 72), and **(C)** monthly maximum temperature (n = 56) across 238 contemporary and historical samples of the willow flycatcher. Details as in Fig. 1. Contemporary samples of the four subspecies are shown as squares, circles, diamonds, and triangles, while upside-down triangles indicate historical samples.

https://doi.org/10.1038/s41558-023-01696-3



Extended Data Fig. 6 | Allele frequency shifts within San Diego at climate-linked loci located near genes associated with avian thermal tolerance and bill morphology. Number of loci linked to **(A)** monthly mean dew point temperature (n = 20), **(B)** monthly precipitation (n = 10), and **(C)** monthly maximum temperature (n = 7) in willow flycatchers breeding in San Diego that shifted in a direction consistent with climate adaptation. The *p*-values show the results for one-tailed binomial tests examining whether the observed proportion of loci that shifted in the expected direction (*that is*, consistent with climate adaptation) was significantly greater than expected by chance.

https://doi.org/10.1038/s41558-023-01696-3

Extended Data Table 1 | Patterns of introgression between willow flycatchers in San Diego and neighbouring populations

Test	P1	P2	P3	O	D	stderr	Z-score	BABA	ABBA	nsnps
All loci	SD Present	IW	PNW	Alder	0.061	0.00164	**37.1**	4213	3729	128582
All loci	SD Present	IW	SW	Alder	0.092	0.00169	**54.6**	4407	3664	128582
All loci	SD Present	East	PNW	Alder	0.196	0.00266	**73.7**	4665	3137	128582
All loci	SD Present	East	SW	Alder	0.219	0.00291	**75.3**	4747	3040	128582
Climate-linked loci	SD Present	IW	PNW	Alder	0.0105	0.0372	0.282	16	15	232
Climate-linked loci	SD Present	IW	SW	Alder	0.034	0.0383	0.887	16	15	232
Climate-linked loci	SD Present	East	PNW	Alder	0.132	0.0418	**3.17**	17	13	232
Climate-linked loci	SD Present	East	SW	Alder	0.034	0.0383	0.887	16	15	232

D statistics quantifying the number of derived alleles shared between the contemporary San Diego population (SD Present) and individuals in the Pacific Northwest (PNW), Desert Southwest (SW), Interior West (IW), and East (East) for all loci (n=128,775) and climate-linked loci (n=232). We included the alder flycatcher as an outgroup in the test topology (((P1, P2), P3), O). Significant *D* statistics (*Z*-score > 3) are highlighted in bold.

nature portfolio

Corresponding author(s): Sheela Turbek

Last updated by author(s): Apr 25, 2023

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used
Data analysis	Trimgalore 0.6.6, BWA 0.7.3a, Picard 2.23.2, samtools 1.13.0, SeqPrep2, Prinseq-lite 0.20.4, mapDamage2, ANGSD 0.939, vcftools 0.1.16, PCAngsd 1.10, DyStruct 1.1.0, Satsuma2, qqman, R version 4.1.2, AdmixTools 7.0.2, admixr, plink 1.9, STRUCTURE 2.3.4, NeEstimator2, detectRUNS, hierfstat, Arlequin 3.5.2.2, gradientForest, Beagle 4.1, vegan, LEA, bedtools 2.30.0, Genetic Data Conversion package, and Integrative Genomics Viewer (IGV). Code associated with this study can be accessed on GitHub (https://github.com/sturbek/WIFL-Climate-Adaptation).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

We aligned our whole-genome sequencing data to an existing southwestern willow flycatcher genome assembly (NCBI BioProject: PRJNA437496) and used published whole-genome sequence data (NCBI SRA: PRJNA692615) for the introgression analyses. The environmental data used in this study were obtained from the Prism Climate Group. The data associated with this study are available in the Dryad Digital Repository (https://datadryad.org/stash/share/qlafA02RwIPjoptjefUaKbrMszEsMJWP8a_H9-e24MY) and NCBI Sequence Read Archive (BioProject: PRJNA957938).

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	not applicable
Population characteristics	not applicable
Recruitment	not applicable
Ethics oversight	not applicable

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	This study compares genomic sequencing data from willow flycatchers collected near San Diego, California from 1888-1909 to contemporary individuals collected across the breeding range from 1996-2015.
Research sample	The research sample includes 17 historical individuals from San Diego and 221 contemporary individuals of the willow flycatcher, 18 of which were sampled in San Diego from 2010-2011. In addition, a separate analysis was performed on 68 additional museum specimens, collected from 1901-1958 from 19 locations across the breeding range of the willow flycatcher, and 548 contemporary individuals that spanned the breeding range.
Sampling strategy	Sample sizes were chosen based off of the availability of southwestern willow flycatcher museum specimens collected in the late 1800s. We included a comparable number of San Diego samples from the past and present in order to track changes in allele frequencies over time.
Data collection	This study used previously collected samples.
Timing and spatial scale	This study used previously collected historical (1888-1958) and contemporary (1996-2016) samples.
Data exclusions	Four contemporary individuals had biologically unrealistic PC1 scores in the genome-wide principal component analysis, potentially due to contamination or the mislabeling of samples, and were excluded from further analyses.
Reproducibility	This study does not include experimental data.
Randomization	Randomization was performed when estimating individual admixture proportions using DyStruct and when assessing whether our gradient forest model explained more variation than expected by chance.
Blinding	Blinding was not relevant to this study.

Did the study involve field work? ☐ Yes ☒ No

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging