Widespread gene flow following range expansion in Anna's hummingbird

Nicole Adams[1], Ruta Bandivadekar[1], C.J. Battey[2], Michael Clark[1], Kevin Epperly[3], Kristen Ruegg[4], Lisa Tell[1], and Rachael Bay[1]

[1]UC Davis
[2]Myriad Genetics
[3]University of Washington
[4]Colorado State University System

October 6, 2022

Abstract

Anthropogenic changes have altered the historical distributions of many North American taxa. As environments shift, ecological and evolutionary processes can combine in complex ways to either stimulate or inhibit range expansion. Here we examine the role of evolution in a rapid range expansion whose ecological context has been well-documented, Anna's Hummingbird (*Calypte anna*). Previous work suggests that the *C. anna* range expansion is the result of an ecological release facilitated by human-mediated environmental changes, where access to new food sources have allowed further filling of the abiotic niche. We examine the role of gene flow and adaptation during range expansion from their native California north into Canada and east into New Mexico and Texas, USA. Using low coverage whole genome sequencing we found high genetic diversity, low divergence, and little evidence of selection on the northern and eastern expansion fronts. Additionally, there are few (if any) limits to gene flow across the native and expanded range. The lack of selective signals between core and expanded ranges could reflect i) an absence of novel selection pressure in the extended range (supporting the ecological release hypothesis), ii) swamping of adaptive variation due to high gene flow, or iii) limitations of genome scans for detecting small shifts in allele frequencies across many loci. Nevertheless, our results provide an example where strong selection is not apparent during a rapid, contemporary range shift.

Introduction

The study of species' geographic limits encompasses some of the most fundamental processes in ecology and evolution including dispersal, gene flow, and adaptation. Species' ranges can be limited by abiotic factors such as precipitation, day length, and soil chemistry as well as by biotic factors such as population density, interspecific interactions, and predator-prey relationships. Anthropogenic-induced changes in these factors can therefore result in rapid shifts in species abundance and distribution . For example, a staggering decline in North American bird abundance since 1970 has been documented and is potentially attributed to increased agriculture, urbanization, habitat loss, and climate change . However, not all environmental shifts result in population decline — some species have also been able to adapt to these changes or track optimal conditions, especially moving poleward and to higher elevation . In fact, a number of North American birds appear to be shifting their ranges northward due in part to warmer temperatures and land use changes . Contemporary shifts in species' ranges provide an opportunity to examine the factors defining range limits in real time.

While the ecological causes of range expansions are often well documented, a complex combination of evolutionary processes also contribute to the push and pull of species distributions. A suite of evolutionary processes can promote range expansion including spatial sorting, natural selection, and genetic drift. Spatial

sorting followed by assortative mating among successful colonizers can shift phenotypic traits associated with expansion, such as dispersal abilities, which in turn can lead to further colonization . Selection on life history traits that increase the reproductive rate can also promote range expansion, as can adaptation to novel environments . Large-effect mutations may increase expansion potential , but limits to genetic variance also affect the capacity for expansion . In contrast, evolutionary processes can also slow or inhibit range expansion. Small population sizes and serial bottlenecks at the range edge lead to strong genetic drift, which decrease genetic variation and expansion potential . Density-dependent dynamics such as the Allee effect limit population growth and therefore expansion .

In combination, the theoretical and empirical literature show us that the same evolutionary processes can have conflicting effects on range expansion outcomes depending on the context . For example, allele surfing, the fixation of alleles along an expansion front, can lead to greater expansion potential if the fixed alleles are neutral or beneficial, but the fixation of deleterious alleles can reduce fitness at the edge — a phenomenon known as expansion load — reducing the expansion potential . Broadly, reduced gene flow from the species' core to the range edge can decrease genetic diversity and thus adaptive potential. However, high gene flow can lead to either increased genetic diversity and higher evolutionary potential or a propagation of maladaptive alleles from the species core that can limit local adaptation at the edges . Understanding and predicting the dynamics of range expansions therefore requires an understanding of gene flow, genetic diversity, and adaptive divergence across the species range paired with a knowledge of the ecological context in which the range expansion is occurring.

A recent and dramatic range expansion in Anna's hummingbird (*Calypte anna)* provides an ideal system to examine the evolutionary processes associated with rapid range expansion . The historical range of *C. anna* is central and southern California, USA and northwestern Mexico. By leveraging community science (Project FeederWatch, Christmas Bird Count) and museum data, previous studies showed a northern and eastern expansion starting around 1940 . Currently, *C. anna* can be found as far north as British Colombia, Canada and southern Alaska, USA and as far east as western Texas, USA. Human habitation and climate change appear to be the drivers of the expansion. In the expanded ranges *C. anna* individuals were more likely to colonize areas with higher housing density and were more likely to visit bird feeders compared to those in the historical range . However, like many North American migratory birds, they may also experience mortality associated with urban settings such as window collisions and encounters with domesticated animals . Increases in minimum winter temperatures were also shown to facilitate the expansion . However, Battey (2019) suggested that the range expansion was largely driven by an 'ecological release' facilitated by introduced plants and supplemental feeding, and that *C. anna* 's climate niche had previously existed in the expanded ranges.

Little is known about genetic variation and population structure in *C. anna* beyond one study that showed low divergence between three California populations and sparse anecdotes of long-distance dispersal in *C. anna* . Further, the genetic makeup of populations in the expanded regions and whether they are adapting to the novel environments has not yet been explored. Here we present the first species-wide genomics study of *C. anna* , testing evolutionary hypotheses regarding their distribution. If range expansion is the result of spatial expansion of the historical fundamental niche linked to an increase in the presence of feeders and ornamental plants (Battey 2019), adaptation may not be necessary to facilitate expansion at the leading edge. However, many abiotic and biotic conditions vary across the range so although adaptation might not be required for expansion, it may still occur as the result of moving into a new environment. The degree to which adaptation occurs and can be detected in our data will depend on a number of factors, including gene flow across the range, the strength of selection, and the genetic architecture of traits under selection. The two leading edges (northern and eastern) allow us to compare these pseudo-replicate expansions, rare in most studies, to answer questions specifically about adaptation, gene flow, and genetic diversity across the native and expanded range, and broadly add to our understanding of eco-evolutionary dynamics in natural populations.

Materials and Methods

Sampling

We collected blood (N=178) and tissue (N=160) samples from live hummingbirds and carcasses for a total of 338 unique hummingbirds across the historical Central and Southern California range and the expanded ranges in Northern California, Washington, Arizona, and Nevada (Fig. 1). Hummingbirds were trapped using previously published methods by a federally permitted hummingbird bander (Tell; US Geological Survey Bird Banding Laboratory Permit #23947). When possible, female birds were examined for evidence of an enlarged oviduct or the presence of an egg for inclusion in the study. Blood was collected (20-30µl, less than 1% body weight) via a toenail clip (N=166;), and placed in Queen's lysis buffer , or blood was collected on an FTA card (N=12). All collection methods were approved by the University of California, Davis IACUC (F#20355). In addition to samples taken from live birds in the field, we added specimens from the Burke Museum collection at the University of Washington and carcasses from wildlife centers for nestlings or fledglings that did not survive the rehabilitation process (see Supplemental file, ANHU_metadata4ms.xlsx, for sample details). For native range samples, we used females and males collected between February and August when many of the sampled birds showed evidence of breeding, however due to low sample availability we used females and males collected throughout the year in the expanded regions. Sampling outside of the breeding season, however, did not appear to affect our results (see Results).

DNA extraction and species identification

Whole genomic DNA was extracted from muscle tissue (for museum collection birds or carcasses from rehabilitation centers), 100-150µl of blood stored in lysis buffer, or 2 to 3 blood spots from a blood collection card using the DNeasy Blood & Tissue Kit (Qiagen). The following modifications to the extraction protocol were used: samples were incubated overnight at 56ºC, the sample was passed over the spin column twice prior to washing, an extra column drying step was taken (14000rpm for 3min), and DNA was eluted in 200µl AE buffer heated to 56ºC. Whole genomic DNA was quantified using a Qubit Fluorometer (Thermo Fisher Scientific) and the quality of DNA was assessed using a 2% agarose gel.

Because nestling and fledgling hummingbirds are difficult to identify to species, we used molecular methods to determine which of the juvenile samples represented *C. anna* and thus could be used in our study. To identify nestlings and fledglings from the wildlife centers as *C. anna* we used Sanger sequencing to sequence 32 unknown individuals and 30 known samples from Anna's (*C. anna*) and other hummingbird species likely to be collected in the region: Costa's (*C. costae*), Allen's (*Selasphorus sasin*), Calliope (*S. calliope*), and Rufous (*S. rufus*) Hummingbirds. We amplified part of the NADH dehydrogenase subunit 2 (ND2) gene using H6313 and L5219 primers , cleaned the products using an ExoSAP protocol, then sequenced them at UCDNA Sequencing Facility at the University of California, Davis. We trimmed and aligned the resulting sequences and used the Neighbor-joining method to build the tree in Geneious v. 9.1.7 (http://www.geneious.com/). We visualized the phylogenetic tree with FigTree (http://tree.bio.ed.ac.uk/software/figtree/) using a Black-chinned Hummingbird sample to root the tree. Black-chinned Hummingbird was used because it was outside of the *Selasphorus* and *Calypte*subclades between which we were identifying unknown samples. ND2 is a marker often used to create hummingbird phylogenies , and in our work has consistently separated individuals that are *C. anna* from other hummingbird species found in California.

Library preparation, and whole genome sequencing

We used a modified library preparation based on Illumina's Nextera protocol to sequence entire genomes of 283 birds. To start, genomic DNA was standardized to 3ng/µl then underwent a tagmentation step using TDE1 enzyme and buffer (Illumina). Dual combination Nextera indexes (Illumina) were then added to tagged DNA fragments followed by a booster PCR using the Kapa HiFi Kit (Kapa Biosystems). Libraries were then bead cleaned and single size selected to remove fragments $<$ 320bp using AMPure XP Beads (Beckman Coulter) and quantified using a Qubit Fluorometer (Thermo Fisher Scientific). All libraries were pooled equimolarly then visualized with a Bioanalyzer (Agilent). The pooled libraries were further size selected to 320-500bp fragments using Ampure XP Beads (Beckman Coulter). A subset of samples (N=40) was size selected using Blue Pippin (Sage Science; University of California Davis Genome Center). The final

libraries were sequenced on an Illumina HiSeq 4000 as 150bp paired-end reads and the resulting sequences were demultiplexed by Novogene (Sacramento, CA, USA). The samples were sequenced across 7 lanes to target 2.5X coverage.

Data processing

Adapters and low-quality reads were trimmed using Trimmomatic or Trim Galore! (a wrapper around Cutadapt , accessible at http://www.bioinformatics.babraham.ac.uk/projects/trim_galore). Each sample was aligned to the *C. anna* reference genome, GCA_003957555.2 using bwa then sorted and indexed using Samtools . For individuals sequenced across two lanes, bam files were merged using Samtools . For all samples, duplicate reads were marked with MarkDuplicates from Picard Tools (http://broadinstitute.github.io/picard). For a subset of samples (N=40), duplicate reads were removed using FastUniq prior to mapping.

Single nucleotide polymorphisms (SNPs), were identified, and genotype likelihoods were estimated using the ANGSD tool accessed through ngsTools . For the parameters used in ANGSD see Table S1. Potentially related samples were identified with NGSrelate , using the rab metric which calculates pairwise relatedness based on . For pairs of related samples (rab>0.45), one individual of each pair was removed.

Population structure

Population structure was analyzed using principal components as well as hierarchical clustering analysis. A covariance matrix was calculated using PCAngsd and then we used RStudio v. 1.3.1093 (RStudio Team, 2018) with R v. 3.6.0 (R Core Team, 2019) to conduct eigenvector decomposition and created plots comparing principal components axes (PCs). We used clustering in NGSadmix to infer the "best" number of populations and estimate ancestry proportions. We ran NGSadmix five times each with population numbers (*i.e.* K values) ranging from one to six. We used the Evanno method implemented in CLUMPAK to determine the best fit K value (accessed at http://clumpak.tau.ac.il/bestK.html).

Genetic variation and gene flow

To estimate nucleotide diversity, we first grouped samples into six regions (Fig. 1A) based on geography and expansion history : Washington (WAS), Humboldt California (HUM), Bay Area California (BAY), Central Valley California (VAL), Pacific coast of southern California (PAC), and eastern expansion samples (EAS). Since the number of samples can affect estimates of genetic diversity, we downsampled each population to the lowest sample size (N=13) by randomly selecting that number of individuals from each population for downstream calculations. A folded site frequency spectrum (SFS) was generated for each downsampled population by generating a site allele frequency file using ANGSD (for parameter details see Table S1) from which an SFS is estimated using realSFS -fold . Finally, each SFS was used as a prior (-pest) to estimate diversity statistics (-doTheta) in ANGSD. We estimated pairwise divergence between samples grouped by county for counties that had at least five individuals using ANGSD (for parameter details see Table S1) and realSFS (fst stats) on polymorphic sites. We estimated global heterozygosity per individual for five to ten individuals per county (Table S2) using ANGSD (for parameter details see Table S1) and realSFS (parameters: -fold 1) to create site frequency spectra. To assess the direction of gene flow among the defined populations we calculated a directionality index, ψ (Peter and Slatkin 2013). First, we created pairwise 2D SFSs using the site allele frequency files created for each population SFS. Then we calculated ψ using equation 1b from , which detects mismatches between pairwise site frequency spectra indicative of successive founder events and thus identifies geographic origins and directionality of expansions.

Selection

We tested for both local and species-wide genomic signals of selection associated with the recent range expansion in *C. anna* . We looked for potential genomic regions under selection in the expanded range using an F_{ST} outlier approach. F_{ST}outliers are a common metric for identifying selection. Peaks of significantly different allele frequencies between populations at close loci are often an indication of potential selection . In this case, we compared the northern (WAS) and eastern (EAS) expansion regions to their nearest native range regions, Central California (BAY) and Southern California (PAC), respectively. We used the pFst tool

in VCFlib (https://github.com/vcflib/vcflib) after creating a BCF file using ANGSD (-dobcf) and converting it to a VCF file with BCFtools accessed through Samtools. The pFst tool uses a likelihood ratio test to detect allele frequency differences between populations.

While the expectation for the magnitude and direction of gene flow is unknown in *C. anna,* largely due to enigmatic movement patterns, a previous study suggested high gene flow between three California populations . Another California hummingbird, Allen's Hummingbird (*S. sasin*), was found to have high geneflow among the mainland populations, potentially indicative of high overall levels of mobility in hummingbirds. If gene flow in *C. anna* is extremely high, we might expect signatures of selection caused by exposure to novel selective agents during range expansion to be present across the entire species rather than divergent between populations. We therefore used all samples to test for the presence of recent selective sweeps using SweeD v. 3.2.1 . We first estimated minor allele frequencies at polymorphic sites using ANGSD (for parameter details see Table S1). We converted these into the required allele count input for SweeD by multiplying the minor allele frequency by the number of individuals sequenced for each site and rounding to the nearest integer. All sites were considered folded. We ran SweeD separately for each chromosome, with a grid equal to the length of the chromosome divided by 5000 (so that we tested every 5kb).

Results

Species identification

We used Sanger sequencing to identify *C. anna* from a set of 32 unknown nestlings and fledglings using 30 individuals of known species identity for comparison. We removed 14 of the 62 ND2 sequences due to low quality or short sequence length, including 10 samples of unknown species. Within known samples, *C. anna* samples formed a monophyletic group allowing us to reliably identify other *C. anna*in our unknown samples (Fig. S1). Of the 22 remaining unknown hummingbird samples, 18 were identified as *C. anna* (Fig. S1) and nine of those had location data and were therefore used for whole genome sequencing.

Whole genome sequencing

We received no data for one sample and despite being sequenced across two lanes we received only one demultiplexed fastq file for two samples, both of which were in sequencing pool ANHU_003. Across 283 individuals, our sequencing runs produced 5.4 billion short reads with more than 99% of samples having 90% or more of reads with a quality score of Q>30. On average, 98.3% of the sequence reads mapped to the reference genome per individual, and individual coverage ranged from 0X to 4.7X with an average of 2.2X. We removed individuals (N = 35) from the dataset that exhibited any of the following: samples that failed to sequence, indicated by a very low number of raw reads (< 1000), samples that mapped poorly to the reference genome (< 50%), and samples that had low individual coverage (< 1.0X). We also removed five outliers in an initial PCA, which we believe may have been misidentified. We identified two pairs of potentially related individuals and removed one individual from each pair. The remaining 241 individuals had on average 98.7% of the sequence reads mapped to the reference genome, and their coverage ranged from 1.0X to 4.45X with an average of 2.5X. The number of loci used for analyses ranged from 22,902 for the PCA (SNPs present in all individuals) to 934,225,517 (all base pairs with sufficient coverage) to calculate theta (Table S3).

High gene flow across the range

We found no evidence of barriers to gene flow across the native and expanded range of *C. anna* . Overall, genetic divergence was low species-wide, with pairwise county divergence (F_{ST}) ranging from 0–0.01 (Fig. S2). We detected no pattern of isolation by distance in pairwise F_{ST} between counties (Fig. 1B; Mantel Test $P = 0.53$). The admixture analyses showed that the ancestral groups were dispersed relatively evenly across all populations. Although the optimal number of ancestral groups was K=2 (Fig. 1C), there was no clear geographic structuring at either K=2 or higher values of K (Fig. S3), suggesting only one major genetic group with no barriers to gene flow. We found no geographic signal based on the first three principal components axes although none of the PC axes explained much more than 1% of the variance (Fig. 1D). There was

a correlation between sequencing pool and PC1 (Fig. S4A), although it explained only a small amount of the variance similar to the other PC axes, potentially highlighting the absence of other factors structuring genetic variation across the range but also reinforcing the need for consideration of sequencing artifacts in next-generation sequencing. To investigate if a larger SNP dataset would clarify any population structure, we reran the PCA with more permissive filtering (-minInd 49, instead of -minInd 241 (all individuals)), which resulted in 9.5 million SNPs. After identifying eight more pairs of potentially related individuals (rab = 0.30–0.56), one individual from each pair was excluded from the PCA (TL016, ANHU_323, ANHU_267, AN86237, UW113545, UW119089, UCK89810), except for the pair UCK89662 and UCP71084 in which both were excluded as they were outliers in an initial PCA. The resulting PCA did not indicate spatial structuring, consistent with the result from the PCA with fewer SNPs (Fig. S4B). We also conducted a PCA with only females collected between February and April (N = 72, after PCA outliers were removed) to ensure only breeding birds were examined, which also showed no genetic structure (Fig. S4C).

No evidence of selection

We found no evidence for either local adaptation or global selection that might be linked to the range expansion. For local adaptation we used F_{ST} to compare the expanded regions (Washington State and the eastern populations) to their nearest native range region (the Bay Area of California and the Pacific coast of southern California), but no obvious peaks stood out for either comparison (Fig. 2A–B). Chromosomal mean pF_{ST}s were $p = 0.7 \pm 0.2$ and no SNPs were significant after Bonferroni or false discovery rate corrections. Based on the most significant 1% of p-values from pFst, we found 6,079 SNPs that were shared between the two expanded region comparisons, only slightly higher than the expected number of high F_{ST} shared SNPs (5,599 SNPs). An alternative to divergence due to spatially varying selection is that with high gene flow homogenizing genetic diversity, selection from resources and environments in the expanded range would affect allele frequencies in the entire species. However, we also found no evidence of selective sweeps when analyzing the site frequency spectrum generated from all samples (Fig. 2C). In fact, the composite likelihood ratio (CLR) statistics representing the test for selective sweeps were all below 2.0, much lower than standard cutoffs found in many studies .

Evidence for a species core

Despite the lack of population structure or signals of selection, we did find expected core-edge patterns of genetic diversity. We found higher nucleotide diversity (pairwise theta) in regions in the native range compared to the expanded ranges (Fig. 3A), consistent with classic core-edge expectations, though the magnitude of these differences is small. The Pacific Coast region, in the native range, had the highest average nucleotide diversity at 1.88×10^{-3} while the eastern and Washington state expanded regions had the lowest values at 1.74×10^{-3} and 1.77×10^{-3}, respectively. Decreased diversity in the expanded regions did not appear to be driven by increased relatedness among sampled individuals (Fig. S5A). However, some counties in the northern and eastern expanded regions appeared to have lower heterozygosity, which could contribute to the observed pattern (Fig. S5B). Using the directionality index, ψ, calculated pairwise across six broad regions, we identified the greater San Francisco Bay Area and the California Central Valley regions as sources of gene flow and the expanded range in Washington state as a majority sink of gene flow (Fig. 3B, Table S4). However, all of the ψ values were below 0.05 (absolute value), which is far lower than the cutoff used in the original paper, an absolute value of 3, and those seen in other studies using this method . The non-significant directionality pairwise comparisons are consistent with our high levels of gene flow.

Discussion

The redistribution of species globally has ignited interest in and urgency for understanding eco-evolutionary dynamics of range shifts. Foundational theoretical work has provided a basis for understanding the interplay between stochastic and deterministic forces in facilitating or hindering range expansion. In many cases, range expansions are expected to lead to higher divergence and lower genetic diversity at the expanding edge compared to the core due to small population sizes, serial bottlenecks, reduced gene flow, and selection pressures . This mixture of neutral and adaptive processes across the expansion axis often leads to spatial

structuring and patterns of isolation by distance . Termed a "pulled wave", the founders at the range edge pull the expansion forward through increased dispersal and reproduction that stratifies demes . However, the opposite pattern of maintained/increased genetic variance at expansion fronts has also been reported in several empirical studies . Conceptual frameworks term these cases "pushed waves," where genetic variation is maintained at the range edge due to gene flow from the range core and potentially positive density dependence, novel interspecific competition or environmental stress that result in less genetic sorting at the edge . These varying outcomes suggest that the interplay between neutral and selective evolutionary processes create variation in range expansion outcomes giving rise to more nuanced approaches to eco-evolutionary dynamics. . In the well-documented contemporary range expansion in *C. anna* , we show patterns largely consistent with expectations of a "pushed wave" expansion; gene flow is high throughout the entire range and we find no strong divergence in allele frequencies between the core of the range and expansion fronts. We do find reduced genetic diversity at expansion fronts, which are characteristic of pulled waves, but the magnitude of the reduction in genetic diversity is small. Together our evidence highlights both the complexity of rapid range shifts in natural populations and potential limitations of genomic data in investigating eco-evolutionary phenomena.

Evidenced by low species-wide genetic divergence and a lack of spatial structuring or signals of selection at the expansion edges or species-wide, we show that *C. anna* has few, if any, limits to gene flow. These results support a previous genetic study in *C. anna* that found no genetic structure among three California populations using mitochondrial DNA . The preservation of genetic diversity across the expanded ranges is consistent with recent, rapid range expansions characterized by a short time frame, growing population sizes, and multiple independent expansion fronts. Our results align with the well-documented rapid range expansion in *C. anna* over the past 80 years . While movement details remain enigmatic, *C. anna* has a broad diet, relatively large territories, and some seasonal migration that was documented based on abundance data , all of which could contribute to high gene flow in this system. Regardless of the mechanism, we observe no population structure and little evidence for increased differentiation at the expansion fronts. Long-distance dispersal, especially from the core, has been shown to preserve genetic diversity in other taxa . This result is often seen in highly mobile species and recent invasions. Examples of high gene flow within species in newly colonized territories include the following: invasive Indo-Pacific lionfish (*Pterois volitans*) in the Caribbean , European starlings (*Sturnus vulgaris*) in South Africa and North America , Pyrenean rocket (*Sisymbriumaustriacum*) in the Meuse River Basin , and round gobies (*Neogobius melanostomus*) in the Great Lake tributaries . The similarities with colonizing species expansions (e.g., propagule and dispersal pressure, novel biotic and abiotic interactions) underscore the emerging work viewing range shifts and expansions of native species, especially those caused by climate change, through the lens of invasion biology . By incorporating theory from invasion biology including assessing the potential impacts, positive and negative, of colonizing species on novel environments, we can gain a more holistic understanding of range expansion.

Range expansions often expose species to novel environments containing new combinations of biotic and abiotic interactions that can coincide with niche shifts, expansions, or unfilling . Previous modeling showed that the expanded regions fell within *C. anna's* fundamental climatic niche prior to the range expansion, suggesting that previous range limits were defined by the presence of resources . Although it is possible that selection was overlooked due to the limitations of genome scans (see below), the lack of selective signatures between the core and expanded ranges identified here aligns with the previous hypothesis that the range expansion in *C. anna* could be the result of an ecological release facilitated by human-mediated landscapes. This hypothesis states that introduced plants and supplemental feeding have allowed *C. anna* to fill out its existing climate niche even in the expanded regions. While other ecological factors induced by urbanization and climate change could also be aiding the expansion, a similar pattern of spatial expansion and ecological release associated with supplemental feeding has been documented in Eurasian Blackcap warbler . Together, these studies provide evidence for the role of local anthropogenic alterations of the landscape shaping broadscale shifts in species' ranges.

High gene flow can be maladaptive at the expansion edge and inhibited selection during pushed wave ex-

pansions often slows and/or prevents further range expansion , such as the southeastern invasion edge of cane toads in Australia that is thought to be limited by cold temperatures . Despite this paradigm, C. *anna* expansions do not appear to be hindered by gene flow from the range core or lack of selection at the expanding edge, adding to the growing literature that gene flow does not limit species ranges or their range expansions (Kottler, Dickman, Sexton, Emery, & Franks, 2021). This again supports the hypothesis that the *C. anna* range was potentially limited by resource availability and thus adaptation may not be required for expansion. Recently, a growing breeding population of *C. anna* has been found in Idaho , an area predicted to be suitable for *C. anna* . Southeast Washington state and sections of Utah are also predicted to have suitable habitat for *C. anna* , potentially suggesting that the expansion will continue in the coming decades.

Despite high gene flow and lack of genomic signatures of selection, we do find very subtle evidence of classic core-edge patterns of genetic diversity. While we did not detect structure in the PCA or admixture analysis, we observed lower nucleotide diversity at both expansion fronts. This result could indicate that mating is not completely random across the entire range despite high gene flow . Increased relatedness among individuals due to small populations sizes could drive the decrease, however this does not appear to be the cause in *C. anna* . Alternatively, lower genetic diversity could be a result of decreased observed individual heterozygosity, which is what we observed. The loss of heterozygosity at the expansion front could be caused by genetic drift, specifically in response to population bottlenecks or allele surfing or by selective sweeps in the expanded ranges, although we were not able to detect any with our genome scans. Relatedly, many of the northern expansion front samples were collected in earlier years (2000 vs late 2010s) which may represent a time point closer to a founding bottleneck before more birds dispersed from the core, a pattern previously suggested in the invasive Indo-Pacific lionfish . However, this decreased heterozygosity does not appear to be strong and consistent enough to result in signatures of divergence and selection between the range core and expansion fronts.

The seemingly contradicting observations of decreased genetic diversity in the absence of signals of selection or structure could have several biological or technical explanations. One possibility is that the expansion is too recent for the detection of significant divergence in range edge populations and more differentiation may develop over time. Further, our low-coverage approach and moderate sample size may not have the power to detect multiple small shifts in allele frequencies across loci that could lead to adaptive evolution, a common issue with genome-wide scans for polygenic traits (Kemper, Saxton, Bolormaa, Hayes, & Goddard, 2014; Pritchard, Pickrell, & Coop, 2010). Reduced diversity in the expanded regions could therefore reflect this weak genome-wide selection that was not detected at any single SNP. Alternatively, environment-mediated trait differences may be plastic - perhaps looking at gene expression or plasticity would provide a more timeframe-appropriate picture of how these birds are responding to the novel environments at the expansion edges. For example, there is widespread use of torpor in Trochilidae and *C. anna* is no exception. In fact, *C. anna* were found to increase their use of torpor in cold temperatures . Plastic behavioral changes at the expansion edges, such as increased use of torpor, could account for how these birds are surviving colder northern temperatures without seeing genetic changes, and warrant further study. Finally, our samples represent a single snapshot in time thus cannot rule out that selection is happening at the range edges, but not enough time has passed to shift allele frequencies that would be detected in the current study, especially in the face of high gene flow.

Anthropogenic influences are changing the genetic landscape through shifting species ranges. Much of the recent focus has been on the role of climate change in facilitating range shifts and the likely eco-evolutionary dynamics of these phenomena. However, this study demonstrates that not all expanding species respond in predicted ways, in fact, not all human-induced range expansions show obvious signatures of evolution. Further work is needed to confirm these results and test the stability of our conclusions over time. For example, using museum specimens to understand the genetic landscape in *C. anna* before the expansions could confirm past gene flow or illuminate if increased urbanization is decreasing genetic diversity, increasing homogenization, or favoring certain alleles. Additionally, while we focused on the northern and eastern expansions, sampling individuals from what might be the "trailing edge" in Mexico would further our understanding of whether climate or resources are defining species range limits in *C. anna* . This study

contributes to the growing literature on the consequences of human-mediated range expansions by adding empirical evidence that eco-evolutionary dynamics are not one-size fits all.

Acknowledgments

Thank you to Teia Schweizer for preparing libraries for sequencing, Christen Bossu for bioinformatic counsel, and Aya Brantley for the illustration of *C. anna* in Fig. 1. Nestlings and fledglings that did not survive rehabilitation were provided by Liberty Wildlife (Arizona, USA), Wetlands and Wildlife Care Center (California, USA), Lindsay Wildlife Experience (California, USA), and Santa Barbara Wildlife Care Network (California, USA). Thanks to John Klicka and the Burke Museum for supplying samples from the northern expanded region. This project was supported by funding from CalTrans 65A0686 (USDOT 69A3551747114) to R.A.B and L.A.T, and NSF Rules of Life EAGER 1837940 to R.A.B and K.R. This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1548562.

References

Data Accessibility Statement

Raw sequence reads and metadata are available on NCBI's Sequence Read Archive digital repository (DOI), and R scripts are available on GitHub (LINK).

Author Contributions

LAT and RAB conceived of the study and secured funding. RRB, LAT, and KE supplied the samples. MWC conducted the lab work and analysis for nestling identification. NEA conducted the DNA extractions, some of the library preparation, and the sequence processing. KR supervised some of the library preparation. NEA and RAB conducted the analyses with input from CJB. NEA and RAB wrote the manuscript with contributions from all authors.

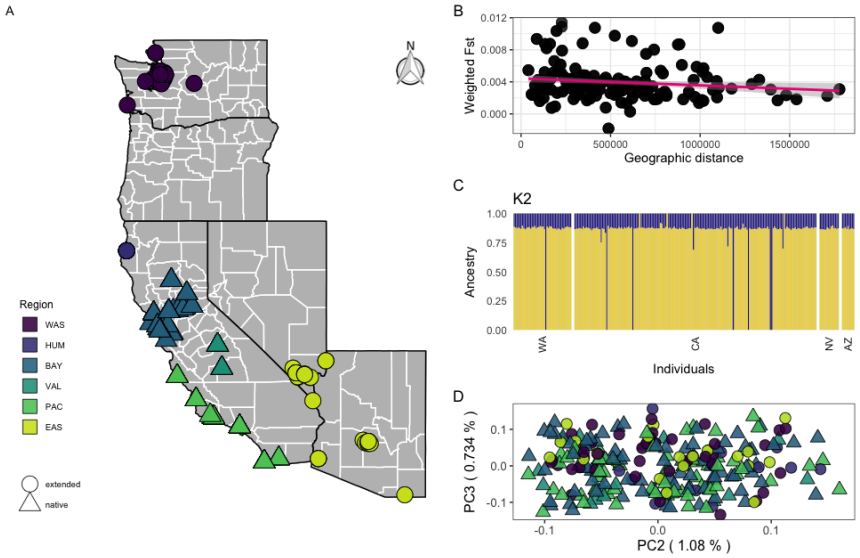

Figure 1. A) Map of *C. anna* samples grouped by region with native (triangle) or expanded (circle) range indicated by shape. The regions are defined as follows: Washington state (WAS), Humboldt County (HUM), San Francisco Bay Area (BAY), California Central Valley (VAL), Pacific Coast (PAC), and the eastern expanded region (EAS). **B)** Genetic divergence (weighted F_{ST}) showing little relationship with increasing pairwise county geographic distance ($y = 0.004 - 8.6E^{-10}x$)**C)** Hierarchical clustering analysis for 2 (K=2) ancestral groups, which was found to be the optimal number of groups. For additional ancestral groups (K=3–5) see Fig. S3. **D)** The second and third components from the principal component analysis (PCA) in which samples are colored by region (see A) and the shapes indicate if the samples are from the native (circle) or expanded (triangle) ranges. The first axis appeared to be driven by sequencing pool (see Fig. S4A).

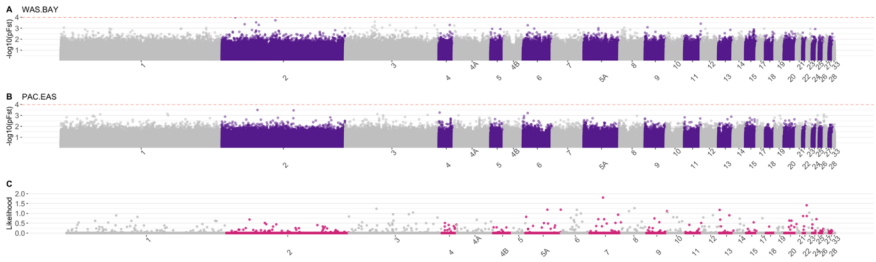

Figure 2. A-B) Significance values (-log10 p-values) from the likelihood ratio test to detect allele frequency differences between WAS and BAY (A) and between PAC and EAS (B) regions plotted across the genome. The red dashed line is a 0.0001 significance threshold. There were no significant SNPs after Bonferroni or false discovery rate corrections. **C)** Selection likelihood values resulting from the selective sweep analysis plotted across the genome. Chromosomes are numbered along the x-axis.

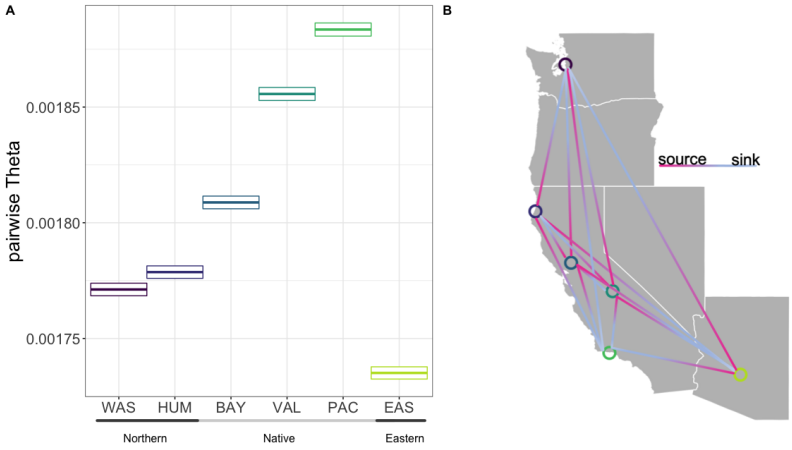

Figure 3. **A)** Mean and standard error of pairwise theta (a genetic diversity measure calculated as tP/nSites) for each region.**B)** Inferred gene flow for each pairwise regional comparison based on directionality index, ψ, indicated as source (pink) or sink (light blue). Point colors in both panels correspond to the region and match those in Fig. 1. Note that none of the ψ values are significant.