

UCLA

UCLA Electronic Theses and Dissertations

Title

Genomic approaches to studying evolution and adaptation in birds

Permalink

<https://escholarship.org/uc/item/83x664c9>

Author

Nakashima, Whitney

Publication Date

2024

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Genomic approaches to studying evolution and adaptation in birds

A dissertation submitted in partial satisfaction
of the requirements for the degree Doctor of Philosophy
in Biology

by

Whitney L. E. Tsai Nakashima

2024

© Copyright by

Whitney L. E. Tsai Nakashima

2024

ABSTRACT OF THE DISSERTATION

Genomic approaches to studying evolution and adaptation in birds

by

Whitney L. E. Tsai Nakashima

Doctor of Philosophy in Biology

University of California, Los Angeles, 2024

Professor Thomas Bates Smith, Co-Chair

Professor Michael Edward Alfaro, Co-Chair

This doctoral dissertation focuses on genomic approaches to study evolution and adaptation in birds. Recent advances in next-generation sequencing (NGS) technology and analytical tools are making the production of genomic data more accessible than ever before. As such, the accumulation of genomic data is occurring at an ever-increasing pace. This genomic data harbors a trove of information waiting to be unlocked. For birds, an initiative called the Bird 10,000 Genomes Project strives to produce genomes for all living species of birds. This growing availability of bird genomes is fueling our ability to understand the genetic architecture

underlying evolutionary and adaptative processes. In my dissertation, I use genomic data and analytical tools to understand color evolution and genomic signals of climate adaptation in birds.

My first two chapters explore color and visual system evolution in birds. In the first chapter, I introduce the R package, *charisma*, for categorizing color from digital images for high-throughput analyses of color evolution. Current color categorization of digital images often entails manual classification of color for each image or a single estimate of colors for the entire group. I present *charisma* a novel efficient and standardized method to identify biologically relevant colors in thousands of images. In the second chapter, I explore the evolution of avian visual system sensitivity. I align and trim SWS1 opsin sequences to predict the visual sensitivity of 418 bird species across the bird tree of life. I estimate rates and reconstruct ancestral states of visual system sensitivity. I found that a violet-sensitive visual system is ancestral in birds and that an ultraviolet-sensitive visual system has evolved at least 18 times across the bird tree of life.

In the last two chapters of my dissertation, I present a reference genome and investigate climate adaptation in Yellow Warblers, *Setophaga petechia*. In my third chapter, I present a highly contiguous reference genome assembly for Yellow Warbler using HiFi long-read and Hi-C proximity sequencing technologies. I generated a 1.22 Gb assembly including 687 scaffolds with a contig N50 of 6.80 Mb, scaffold N50 of 21.18 Mb, and a BUSCO completeness score of 96.0%. This high-quality reference provides a key resource for understanding gene flow, divergence, and local adaptation and informing conservation management. In my fourth chapter, I investigate genomic signals of climate adaptation in Yellow Warblers in California. Climate change is an ongoing threat to biodiversity and species are being forced to respond or face extinction. This response is dependent on their ability to adapt to rapidly changing environments.

I use a whole genome sequencing approach to examine genomic signals of local adaptation in California breeding Yellow Warblers to identify populations most vulnerable to climate change. Despite low genetic structure in Yellow Warblers breeding in California, I identified unique genotype-environment associations and 2,972 putatively adaptive single-nucleotide polymorphisms across 137 individuals. This study highlights the importance of understanding neutral and adaptive genetic variation in bird populations.

The dissertation of Whitney L. E. Tsai Nakashima is approved.

Rachael A. Bay

Gregory F. Grether

Michael Edward Alfaro, Committee Co-Chair

Thomas Bates Smith, Committee Co-Chair

University of California, Los Angeles

2024

DEDICATION

Dedicated to my IIs and my rainbow III.

TABLE OF CONTENTS

ABSTRACT OF THE DISSERTATION.....	ii
DEDICATION.....	vi
LIST OF FIGURES.....	viii
LIST OF TABLES.....	ix
ACKNOWLEDGMENTS.....	x
VITA.....	xii
CHAPTER 1: <i>charisma</i> : an R package to determine discrete color classes for high-throughput analyses of color evolution.....	1
CHAPTER 2: Evolution of visual system sensitivity in birds.....	29
CHAPTER 3: A highly contiguous genome assembly for Yellow Warbler (<i>Setophaga petechia</i>).....	50
CHAPTER 4: Climate-associated genomic variation in California breeding Yellow Warbler (<i>Setophaga petechia</i>).....	62

LIST OF FIGURES

CHAPTER 1

Figure 1.....	12
Figure 2.....	13
Figure 3.....	14
Figure 4.....	15
Figure 5.....	16
Figure 6.....	17
Figure S1.....	24
Figure S2.....	25
Figure S3.....	26

CHAPTER 2

Figure 1.....	37
---------------	----

CHAPTER 3

Figure 1.....	52
Figure 2.....	55
Figure S1.....	61

CHAPTER 4

Figure 1.....	72
Figure 2.....	73
Figure 3.....	74

LIST OF TABLES

CHAPTER 1

Table 1.....	18
Table S1.....	27
Table S2.....	28

CHAPTER 2

Table 1.....	39
Table 2.....	39
Table S1.....	46

CHAPTER 3

Table 1.....	54
Table 2.....	56

CHAPTER 4

Table 1.....	75
Table 2.....	79

ACKNOWLEDGMENTS

To my family and friends for their love and support throughout this process.

To my co-advisors, committee members, and mentors for their guidance and support: Dr. Thomas Smith, Dr. Michael Alfaro, Dr. Rachael Bay, Dr. Gregory Grether, Dr. John E. McCormack, Dr. Allison J. Shultz, and Dr. Amanda J. Zellmer.

To the wonderful Alfaro Lab, Center for Tropical Research, and Moore Lab groups, the EEB community, and my field mates: Kelly Barr, Lucia Bolzoni, Russell Campbell, Jonathan Chang, Joey Curti, Christa Gomez, Cri Gruppi, Eliza Kirsch, Ryan Harrigan, Mark Juhn, Elizabeth Karan, Ren Larison, Eileen Mansoorian, Noah Medina, Marky Mutchler, Kevin Njabo, Alana Pizarro, Rosamari Orduna, Brenda Ramirez, Kelsey Reckling, Genaro Rodríguez Otero, Wendoly Rojas Abreu, Nicholas Ruffolo, Nick Russo, Shawn Schwartz, Yoonjin Shu, Ingrid Tello López, Ryan Terrill, Jen Tinsman, Patrice Tonnis, Tessa Villaseñor, Jenny Wong, Luz E. Zamudio Beltrán, and Virginia Zaunbrecher.

For Chapter 1, I would like to thank my co-first author, Shawn Schwartz and co-authors, Elizabeth A. Karan, Mark S. Juhn, Mackenzie Perillo, Allison J. Shultz, John E. McCormack, Thomas B. Smith, and Michael E. Alfaro.

For Chapter 2, I would like to thank my co-authors, Tania Romero, Tatiana Washington, John E. McCormack, Allison J. Shultz, Thomas B. Smith, and Michael E. Alfaro. Funding for this

chapter was provided by the Ecology and Evolutionary Biology Department at UCLA through the Lida Scott Brown Research Grant in Ornithology.

Please refer to the Acknowledgment section in Chapter 3 (reprint of published article).

For Chapter 4, I would like to thank my co-authors, Ryan J. Harrigan, John E. McCormack, Michael E. Alfaro, Rachael A. Bay, and Thomas B. Smith. I would also like to thank Kevin Burns at the San Diego Natural History Museum, Carla Cicero at the Museum of Vertebrate Zoology, and Marina Rodriguez and Kristen Ruegg for assistance in sample acquisition. Funding for this chapter was provided by the California Conservation Genomics Project, with funding provided to the University of California by the State of California, State Budget Act of 2019 [UC Award ID RSI-19-690224], the Ecology and Evolutionary Biology Department at UCLA through the Lida Scott Brown Research Grant in Ornithology, and the Pasadena Audubon Society.

Funding for my dissertation was provided by the Ecology and Evolutionary Biology Department at UCLA through the Lida Scott Brown Fellowship in Ornithology, the Graduate Division at UCLA through the Dissertation Year Fellowship, the Moore Laboratory of Zoology at Occidental College, and the National Science Foundation Graduate Research Fellowship Program (NSF DGE-1650604). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

VITA

EDUCATION

Occidental College, BA 2011 (Biology, minor in Studio Art, *cum laude*)

SELECTED PEER-REVIEWED PUBLICATIONS

- Tsai, WLE**, M Escalona, KL Garrett, RS Terrill, R Sahasrabudhe, O Nguyen, E Beraut, W Seligmann, CW Fairbairn, RJ Harrigan, JE McCormack, ME Alfaro, TB Smith, & RA Bay. 2024. A highly contiguous genome assembly for the Yellow Warbler (*Setophaga petechia*). *Journal of Heredity*, esae008. <https://doi.org/10.1093/jhered/esae008>.
- Ramirez, BR, RJ Freeland, A Muhlheim, AJ Zellmer, DA DeRaad, EJ Kirsch, MJ Mutchler, MB Secor, KR Reckling, ME Schedl, B Durham, **WLE Tsai**, RS Terrill, S Sannapareddy, AH Sivakumar, KL Garrett & McCormack, JE. 2024. Convergent niche shifts of endangered parrots (genus *Amazona*) during successful establishment in urban southern California. *Diversity and Distributions*, n/a(n/a), e13817. <https://doi.org/10.1111/ddi.13817>.
- Blackburn DC et al.† 2024. Increasing the impact of vertebrate scientific collections through 3D-imaging: the openVertebrate (oVert) Thematic Collections Network. *Bioscience*. <https://doi.org/10.1093/biosci/biad120>. (Part of oVert Project Team).
- JE McCormack, MM Hill, DA DeRaad, EJ Kirsch, KR Reckling, MJ Mutchler, BR Ramirez, RML Campbell, JF Salter, AK Pizarro, **WLE Tsai** & E Bonaccorso. 2023. An elevational shift facilitated the Mesoamerican diversification of Azure-hooded Jays (*Cyanolyca cucullata*) during the Great American Biotic Interchange. *Ecology & Evolution* 13(8): e10411.
- DeRaad, DA, EE Applewhite, **WLE Tsai**, RS Terrill, SE Kingston, MJ Braun, JE McCormack. 2022. Hybrid zone or hybrid lineage: a genomic reevaluation of Sibley's classic species conundrum in *Pipilo* towhees. *Evolution*, 77(3): 852-869. <https://doi.org/10.1093/evolut/qpac068>
- Salter, JF, PA Hosner, **WLE Tsai**, JE McCormack, EL Braun, RT Kimball, RT Brumfield, BC Faircloth. 2022. Historical specimens and the limits of subspecies phylogenomics in the New World quail (Odontophoridae). *Molecular Phylogenetics & Evolution*. <https://doi.org/10.1016/j.ympev.2022.107559>
- Tsai, WLE**, ME Schedl, JM Maley, JE McCormack. 2020. More than skin and bones: comparing extraction methods and alternative sources of DNA from avian museum specimens. *Molecular Ecology Resources*, 20(5): 1220-1227. <https://doi.org/10.1111/1755-0998.13077>
- Vinciguerra, NT, **WLE Tsai**, BC Faircloth, JE McCormack. 2019. Comparison of ultraconserved elements (UCEs) to microsatellite markers for the study of avian hybrid zones: a test in *Aphelocoma* jays. *BMC Research Notes*, 12:456.
- DeRaad, DA, JM Maley, **WLE Tsai**, JE McCormack. 2019. Phenotypic clines across an unstudied hybrid zone in Woodhouse's Scrub-Jay (*Aphelocoma woodhouseii*). *The Auk: Ornithological Advances*, 136(2):1-11.

Tsai, WLE, C Mota-Vargas, O Rojas-Soto, R Bhowmik, EY Liang, JM Maley, E Zarza, JE McCormack. 2019. Museum genomics reveals the speciation history of *Dendrortyx* wood-partridges in the Mesoamerican highlands. *Molecular Phylogenetics and Evolution*, 136:29-34.

RECENT GRANTS, AWARDS, & HONORS

- 2024 NSF Postdoctoral Research Fellowship in Biology. National Science Foundation.
La Kretz Postdoctoral Fellowship. La Kretz Center for California Conservation Science, UCLA.
- 2023 Dissertation Year Fellowship. Graduate Division, UCLA.
- 2022 UCLA Doctoral Student Travel Grant. AOS BC Conference.
“Genomic signals of climate adaptation in Yellow Warblers.” Graduate Student Research Award. Society of Systematic Biologists.
Travel Award in Support of Diversity and Inclusion. American Ornithological Society. AOS BC Conference.
K. Patricia Cross Future Leaders Award. Association of American Colleges and Universities.
- 2021 “Genomic signals of climate adaptation in Yellow Warblers.” Lida Scott Brown Research Grant. Department of Ecology and Evolutionary Biology, UCLA.
“Genomic signals of climate adaptation in Yellow Warblers.” Frank M. Chapman Grant. American Museum of Natural History.
“Genomic signals of climate adaptation in Yellow Warblers.” Hesse Research Award. American Ornithological Society.
“Genomic signals of climate adaptation in Southern California Yellow Warblers.” Research Grant. Pasadena Audubon Society.
- 2020 “Comparative evolution of visual system and color diversity in birds.” Lida Scott Brown Research Grant. Department of Ecology and Evolutionary Biology, UCLA.
“Population genomics of resident Yellow Warblers of Southeast Mexico.” Graduate Student Summer Research Grant. Latin American Institute, UCLA.
“Genomic signals of local climate adaptation in Yellow Warblers.” California Conservation Genomics Program. University of California. PIs: Rachael Bay, Thomas Smith.
- 2019 NSF Graduate Research Fellowship. National Science Foundation. NSF DGE-1650604.
Lida Scott Brown Fellowship. Department of Ecology and Evolutionary Biology, UCLA.

CHAPTER 1

charisma: an R package to determine discrete color classes for high-throughput analyses of color evolution

Abstract

Advances in digital imaging and software tools have provided increasingly accessible datasets and methods for analyzing color evolution. Despite the variety of computational packages available, most rely on manual identification for individual images, a single-color class estimation for a large set of images, or an overestimation of color classes. This limits the ability to analyze large datasets of images, decreases the accuracy of downstream analyses, and is not representative of biologically relevant color classes. Here, we present *charisma*, an R package designed to determine the number and proportions of distinct color classes in an image suitable for large-scale studies of biological organisms. We show that *charisma* can quickly and accurately classify every pixel in an image and validate classifications using known color swatches. We apply this method to avian color evolution by investigating how color evolves within a diverse and charismatic group of birds, tanagers in the subfamily Thraupinae. We find that *charisma* color classifications are consistent with those made by experts in the field. Our results demonstrate that using *charisma* to curate and call colors in images provides a standardized and reproducible framework for high-throughput color classification. It can be tailored for datasets collected from different sources to yield reproducible results and is designed to work with popular color analysis packages.

Introduction

Many of Earth's organisms' charismatic coloration and patterning have inspired biologists to study these traits for over a century (Darwin, 1981; Mayr et al., 1963). Animal colors and patterns have been shown to have ecologically important functions such as intra- and inter-specific communication in the form of sexual or social signaling, as well as provide crypsis,

advertisement, or mimicry (Cooney et al., 2019; Feller et al., 2017; Håstad et al., 2005; Irwin, 1994; Rabosky et al., 2016). The evolution of conspicuous coloration and patterning has historically been studied using genetic and observational approaches (Andersson & Amundsen, 1997; Barlow et al., 2018; Ehrlich et al., 1977; McMillan et al., 1999; Neudecker, 1989). Yet, advances in digital imaging and novel software tools have provided increasingly accessible datasets and reproducible methods for quantitative ecological and evolutionary studies of color and pattern (Berg et al., 2020; Chan et al., 2019; Endler, 2012; Endler et al., 2018; Hemingson et al., 2024; Maia et al., 2013; Valvo et al., 2020; Van Belleghem et al., 2018; Weller & Westneat, 2019).

Color categorization is one crucial task when analyzing color in a biological context; however, it is notoriously difficult. Despite the variety of computational packages available, most require users to input the number of dominant color classes (k) *a priori* to computing color metrics (Maia et al., 2019; Van Belleghem et al., 2018; Weller & Westneat, 2019). This involves a manual identification of k for individual images, a single k estimation for a large set of images, or a bespoke method for categorizing color (Alfaro et al., 2019; Delhey et al., 2023; Hemingson et al., 2019; Weller & Westneat, 2019). While these methods are suitable for a relatively small sample of organisms with low color variation across species or a single study, conducting large-scale analyses of many clades with high phenotypic diversity remains limited by both time and reproducibility.

To fill this gap, we introduce *charisma*, an R package designed to automatically determine the presence or absence of key color categories in organismal images that can be used across studies of broad taxa. The *charisma* package incorporates functions from the R package *recolorize* to determine k for organismal images (Weller et al., 2024). Here, we provide a flexible

and reproducible framework to assess discrete color classes in images suitable for large-scale studies of color and color pattern evolution that can be seamlessly integrated into popular downstream quantitative analysis workflows in R. We demonstrate the efficacy of our software by (1) validating *charisma* using images of known color classes, (2) applying *charisma* to standardized images of tanager museum specimens from the subfamily Thraupinae and comparing *charisma* classification of tanagers to expert color classification, and (3) analyzing color evolution in tanagers.

Description

The primary function of *charisma* is to classify the proportion of 10 discrete color categories in an image in a standardized, efficient, and reproducible way. For each pixel, the color hue, saturation, and value (HSV) are harvested and compared to a color look-up table (CLUT). The default CLUT assigns each pixel to ten discrete color classes: red, orange, yellow, green, blue, purple, brown, black, gray, and white. We tuned the default CLUT through an iterative process to have non-overlapping ranges for each color and optimized it using standardized, color-calibrated photos of bird museum specimens. We have structured *charisma* so that the CLUT is easily exchanged for one that is modified based on the needs of the user and the organisms they are studying. Images passed into the *charisma* function first undergo preprocessing and noise-reduction, facilitated by the R package *recolorize* (Weller et al., 2024). This step is implemented to reduce the impact of noisy colors and pixels have on downstream color classification in *charisma*. The color classification can be run automatically, or users can manually intervene by merging color categories, replacing colors, and/or using a threshold cutoff for colors with low proportions of pixels. The final *charisma* output (Figure 1, Table S1) for each image includes a diagnostic plot, the number of colors present ($k=1-10$), a table with presence and absence data for

each color class, a log of all manual interventions performed, and R objects that can easily be passed through existing evolutionary analysis pipelines, like those in the R package *pavo* or evolutionary rates analyses. *charisma* allows for a highly standardized and reproducible pipeline for obtaining color data from images of organisms for downstream evolutionary analyses.

Methods

Validation

We validated the performance of *charisma* and the accuracy of our CLUT by testing known color datasets. First, we obtained color grids for each of our ten colors from the images on the Wikipedia "Shades of [Color]" pages (Table S2). These images contain grids of 9-25 color squares representative of each color, which we used as input in the automated *charisma* pipeline. Next, we used color swatches from the PANTONE Matching System (PMS, X-Rite, Carlstadt, New Jersey), which provides a standardized color reproduction system. We extracted 901 PMS color swatches (<https://www.pantone-colours.com/>) and ran them through the automated *charisma* pipeline.

Imaging bird museum specimens

We illustrate the utility of *charisma* for evolutionary color analyses with images of tanagers in the subfamily Thraupinae (Family Thraupidae). Tanagers in this subfamily have been well-studied in molecular phylogenetics and color evolution (Burns et al., 2014, 2016; Price-Waldman et al., 2020; Shultz & Burns, 2017). Previous work has found that lineages within Thraupinae have the highest rates of plumage complexity evolution (Price-Waldman et al., 2020). The subfamily Thraupinae also contains the notably colorful genus *Tangara* making them an excellent group for testing *charisma*. We photographed 32 bird museum specimens at the Natural History Museum of Los Angeles County (Table 1). Specimens were photographed under

consistent conditions using a Nikon D70s with a Novoflex 35mm lens and Natural Lighting Naturesunlight 30-W full spectrum fluorescent bulbs. Each image included a color standard and RAW image files were white-balanced before processing. We were solely interested in plumage coloration and segmented images to remove the background and the specimen's bills, legs, tags, and cotton eyes for color analysis.

Bird coloration datasets

To test how well *charisma* categorized color in our bird museum specimen images, we created datasets by running the fully automated version of the package and a dataset where we manually adjusted the images using the *charisma* functions. For the fully automated datasets, we tested cutoff threshold values of 5%, 7.5%, and 10%, where any color with a proportion smaller than these values would be removed from the color classification. We also had a dataset of color classifications for all members of the family Thraupidae determined by an expert in tanager coloration (Shultz & Burns, 2017). We adjusted the expert color classifications to include only colors present in the side view of the specimen image. We also accounted for intraspecific and subspecific color variations in the specimens. To evaluate our color classification, we compared our automatic and manual color datasets against our expert datasets using a binary contingency table (Powers, 2020). Using our expert color dataset as the true colors of the birds, we classified each *charisma* color call as a true positive (hit), false negative (miss), false positive (false alarm), or true negative (correct rejection).

Evolutionary analyses

We used our datasets and a previously published tanager phylogeny (Burns et al., 2014) to explore rates of color evolution and to reconstruct ancestral color states. Additionally, because bird coloration is well-studied and the color-producing mechanisms and structures are reasonably

well-known in many species (Hill, 2006; Mason & Bowie, 2020; Stoddard & Osorio, 2019; Stoddard & Prum, 2011), we also estimate rates and reconstruct ancestral states for color producing mechanism in birds. We transformed our discrete color data to color-producing mechanism data by grouping our discrete colors by avian color-producing mechanism, removing white because this can either be produced by a lack of pigment or feather microstructure: melanin - black, brown, gray; carotenoid - red, orange, yellow; structural - green, blue, purple (Hill, 2006). We used the `fitDiscrete` function in the R package *geiger* to test the fit of two models of evolution: the equal rates (ER) model, which assumes equal rates of gains and losses of a trait, and the all rates different (ARD) model, which assumes different rates of gains and losses of a trait. We compared the natural log (ln) transition rates for each color and each dataset to determine the potential effects of differing datasets on the evolutionary analyses. We then selected the best model for each color and mechanism for our manual dataset using the sample size correct Akaike Information Criterion (AICc). Using the best fitting model, we reconstructed ancestral states for our manual dataset by estimating posterior probabilities and mapping them onto the tanager phylogeny using the R package *phytools* (Burnham et al., 1998; Revell, 2012).

Results and Discussion

Validation

Our color validations demonstrate the ability of *charisma* to identify colors in an image. Outputs from the Wikipedia color grids show that for each color grid, most *charisma* color calls return the color category of the grid (Figure 2). The most variation in color calls was present in the red and orange grids, with four colors being called for each. The color with the second highest proportion called for red and orange was brown. This highlights the difficulty of delineating

between these three color boundaries in HSV color space. We also show that *charisma* can quickly and accurately call colors for many colors from the PMS (Figure 3).

Comparison of color datasets

Our manual color dataset outperformed the automatic color datasets when compared to our expert color dataset (Figure 4, S1). While black was classified well across all automatic and manual datasets, brown and gray had high false alarm rates, and red and orange had high miss rates in our automatic datasets (Figures 1A, S1). The bases and tips of bird feathers, especially contour feathers, are often different colors, with the base generally being gray or drabber than the tip, which frequently has more highly pigmented or structural coloration (Terrill & Shultz, 2023). In bird museum specimens, overlapping feathers can be misaligned to create artifacts of gray or brown patches (Figure 1A). These color artifacts may contribute to over-representing brown and gray in the automatic color datasets. The colors red (2 instances) and orange (1 instance) were rare in this dataset. When they were present, they were represented by small patches. This may contribute to the high miss rates in the automatic datasets as the threshold values remove colors with a small proportion of pixels assigned to them in the image. The manual *charisma* classifications show an almost identical color profile to our expert color dataset as evidenced by the nearly 100% hit and correct rejection rate (Figure 4B). The only differences were a very low miss rate for blue and a very low false alarm rate for green.

Evolutionary analyses

We compared results from evolutionary rate analyses across our color datasets (Figure S2). We found that the automatic datasets revealed an elevated rate of evolution for gray across all models, providing evidence that these color artifacts in bird museum feathers due to misaligned feathers impact downstream evolutionary analyses. We also find slower rates of evolution of

blue and green in our automatic datasets, which may be due to the higher rate of *charisma* missing these color calls in the automated pipeline (Figure 4A, S1). We generally see high congruence in evolutionary rates between our expert and manual color datasets.

Here, we present the results of our evolutionary analyses using the manual color dataset. We tested two models of evolution on discrete color categories and colors grouped by color-producing mechanism. We excluded purple and the melanin-producing mechanism from our analyses because purple was absent in this dataset and melanin was ubiquitous across all species (Figure 5). When we compare color evolution using the ER model, brown shows an elevated rate of evolution compared to all other color categories, and black shows the slowest rate (Figure 6A). Melanin is a structurally robust pigment often deposited in the primary wing and tail feathers of birds to increase durability and resist abrasion in these high-use feathers (Bonser, 1995). This aligns with our finding that melanin coloration is highly conserved and indicates the need for structural integrity in bird feathers. For mechanisms, carotenoid coloration showed a higher rate of evolution than structural coloration. The ARD model shows that green and blue have elevated rates of gains and losses compared to all other colors (Figure 6B, 6C). Structural coloration showed a much higher gain rate compared to carotenoid coloration. Pigment-based color is much more widespread across the avian tree of life than structural color. However, where structural color is present, there is some evidence for color diversity accumulating faster than pigment-based color (Maia et al., 2016). The modularity of layering pigments and structure and the subtle nanometer changes that result in the production of significantly different colors may allow for this rapid evolution and accumulation of color diversity (Eliason et al., 2015; Maia et al., 2016).

We found that the ER model was the best-fitting model for every color and color-producing mechanism except for orange. Using the best-fitting model for each color and mechanism, we reconstructed ancestral states and mapped them across the phylogeny (Figure 5, S3). We found that most ancestral nodes had black, brown, yellow, green, and blue color states, and these were also the most common colors in our dataset. Black, white, brown, and gray are the most common colors across the bird tree of life (Delhey et al., 2023) and differences in our findings are likely due to the uniquely colorful nature of birds in the subfamily Thraupinae. The ancestral state reconstruction demonstrates that rates of evolution are driven mainly by losses of colors across internal branches of the tree with only a single gain of white on an internal node. Orange fits the ARD model best with a slightly elevated rate of gain than loss (Figure 6B, 6C). However, this result for orange is mainly driven by the uncertainty of the presence of orange at the root of the phylogeny (Figure 5, 6). Interpreting results for orange and red is not recommended because these colors are only present in two tanagers in our phylogeny. A more comprehensive sampling of tanagers outside of the subfamily Thraupinae could resolve this.

Conclusions

The R package *charisma* provides a standardized, reproducible, and flexible framework for classifying color from digital images. We integrated *charisma* with the existing color and color pattern analyses R packages, *recolorize* and *pavo*. Here, we used color-calibrated images of bird museum specimens; however, we built functionality to adjust the default CLUT based on user needs and their specific image dataset. These datasets could include existing standardized photos like J.E. Randall's images of fish (<http://pbs.bishopmuseum.org/images/JER/>) or datasets of non-standardized images like those culled from iNaturalist (*iNaturalist*, n.d.) or ebird (Sullivan et al., 2009). We found that artifacts from misaligned feathers in our bird museum specimen image

dataset led to the overrepresentation of brown and gray in our automatic color datasets. Thus, we recommend using *charisma* to manually adjust image colors to fit the biology of the study system. The outputs of *charisma* allow for a seamless transition to downstream evolutionary analyses, which we demonstrated on tanagers in the subfamily Thraupinae. Ultimately, we present *charisma* as a solution for high-throughput organismal color analyses.

Figures

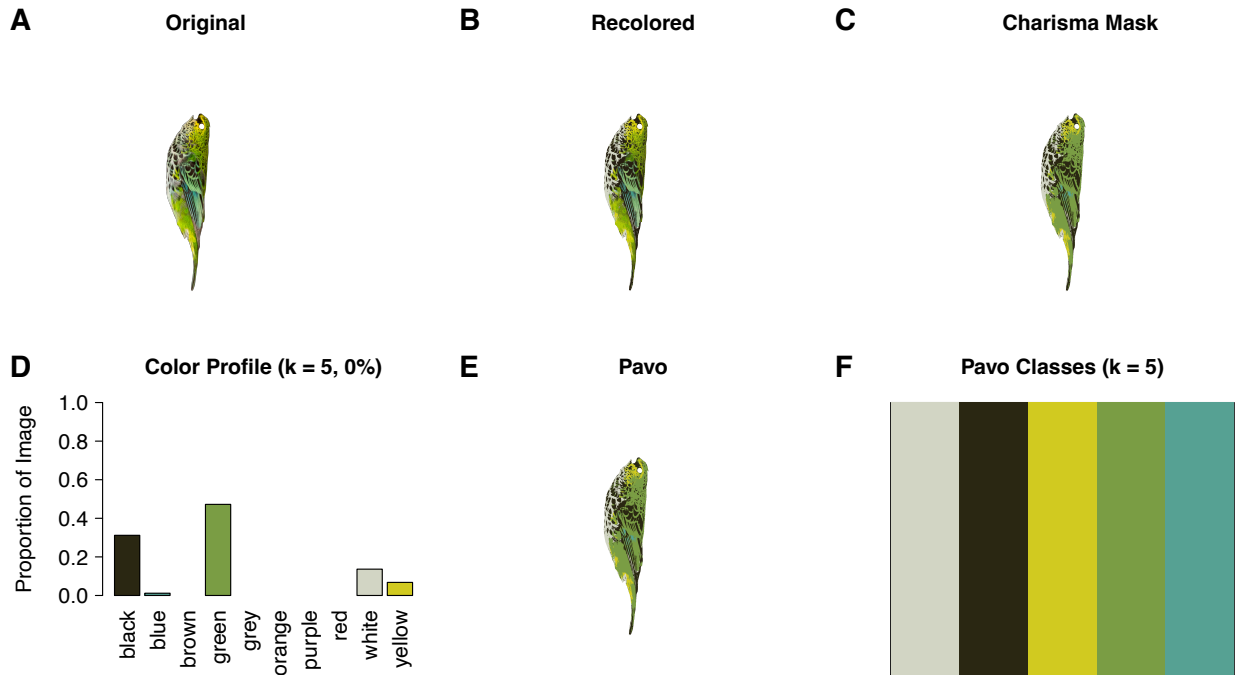


Figure 1. Example of a diagnostic plot output from *charisma* for the Speckled Tanager, *Ixothraupis guttata*. (A) The original color-calibrated input image was segmented to remove the bills, legs, tags, and cotton eyes. (B) Preprocessed image run through the R package *recolorize*. (C) *charisma* mask with manual color adjustments. (D) Plot showing the proportion of pixels in the image for each color category present. (E) Input image for analyses in the R package *pavo*. (F) Color classes output by *charisma* with color swatches for each color class.

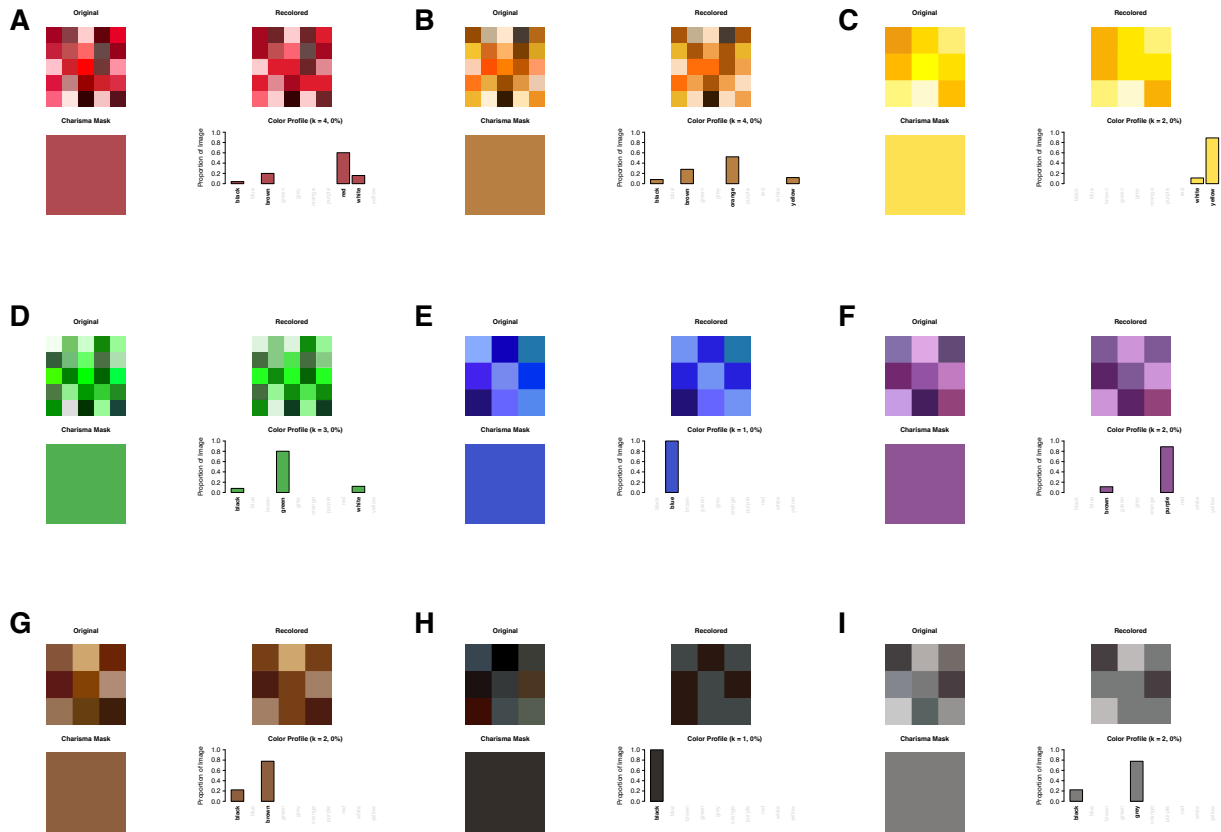


Figure 2. *charisma* outputs for Wikipedia color grids. Each color output contains four plots: the original color grid, the color grid following *recolorize* preprocessing, the average color across the entire grid, and a plot of the proportion of *charisma* colors called for each discrete color category. (A) Red, (B) Orange, (C) Yellow, (D) Green, (E) Blue, (F) Purple, (G) Brown, (H) Black, (I) Gray.

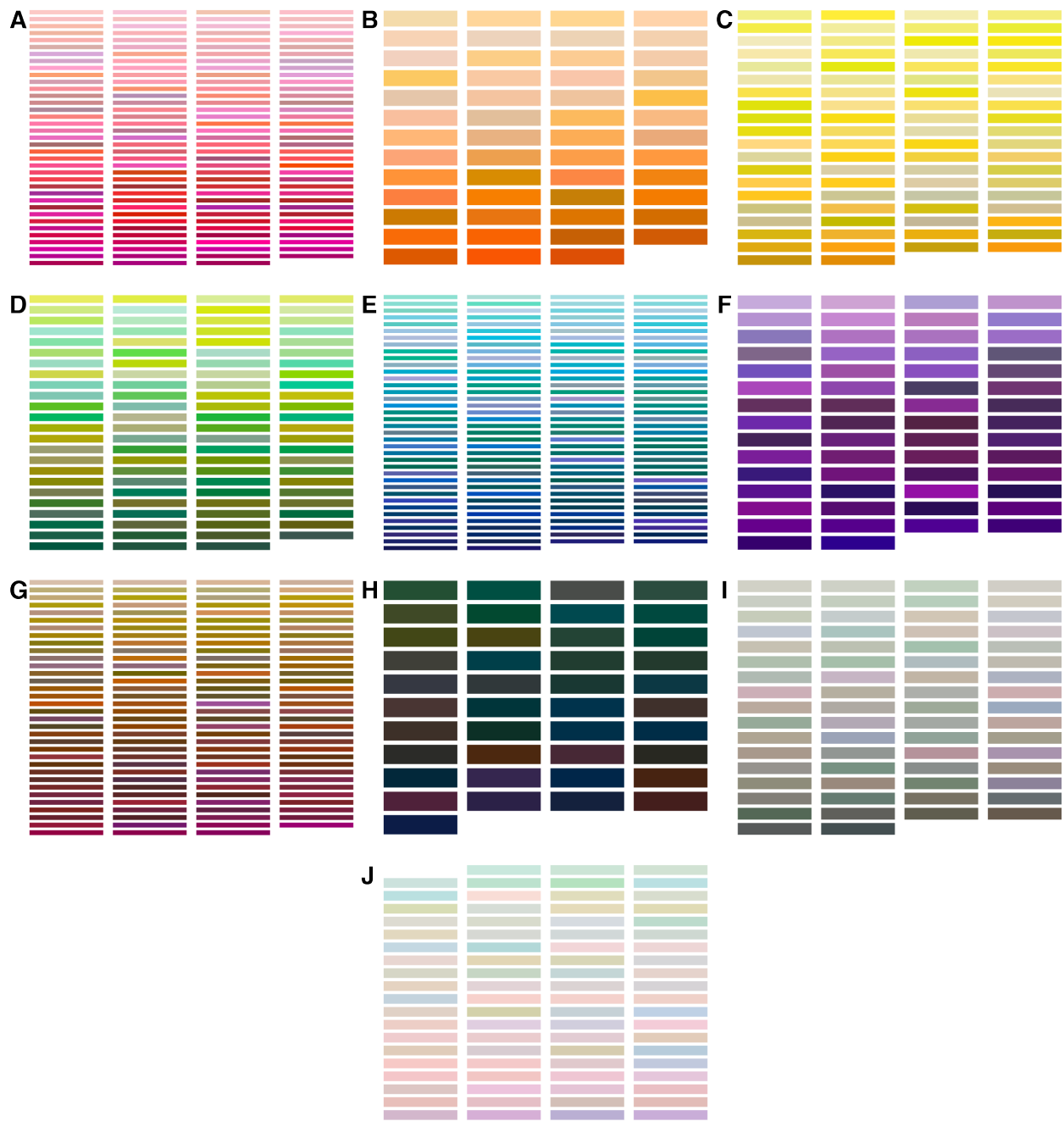


Figure 3. PANTONE Matching System color swatches sorted by *charisma* color call. (A) Red, (B) Orange, (C) Yellow, (D) Green, (E) Blue, (F) Purple, (G) Brown, (H) Black, (I) Gray, (J) White.

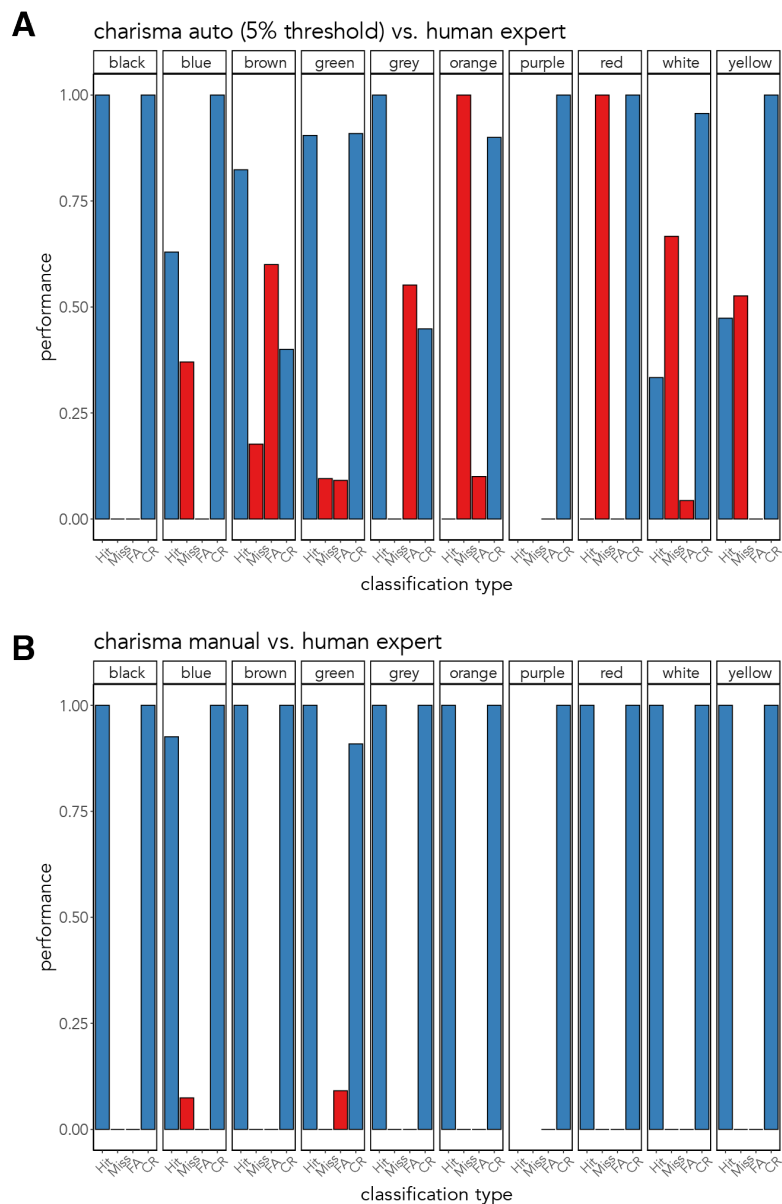
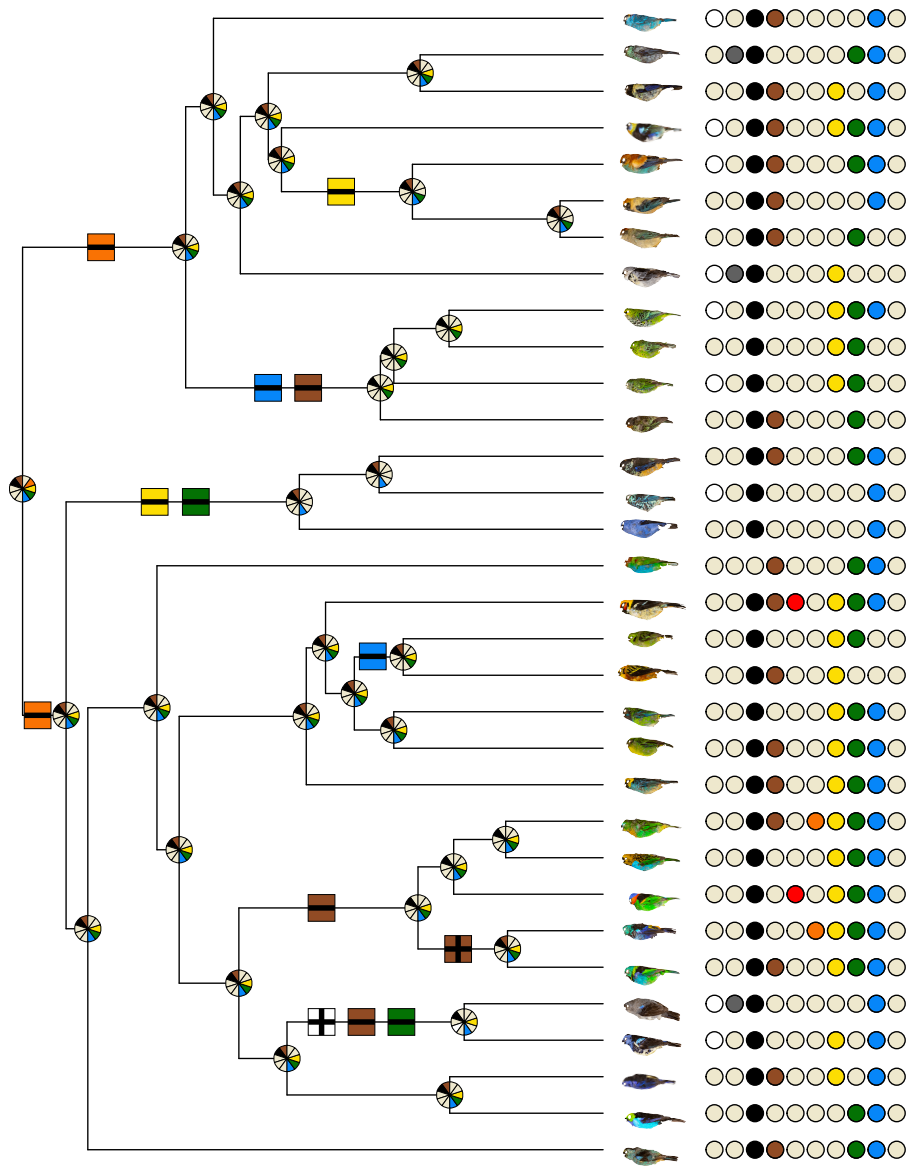


Figure 4. Plotted binary contingency table analyses comparing (A) the automated color dataset generated using *charisma* with a cutoff threshold of 5% to the expert color dataset and (B) the manual color dataset generated using *charisma* to the expert color dataset. Blue bars represent correct color calls or non-calls, and red bars represent false color calls or non-calls. Hit – true positive, Miss – false negative, FA (false alarm) – false positive, CR (correct rejection) – true negative.





 Gain or loss of color at internal branches
 Color wheel at each node with ancestral color states (≥50% posterior probability) filled in

Figure 5. Colors mapped on the tanager phylogeny.

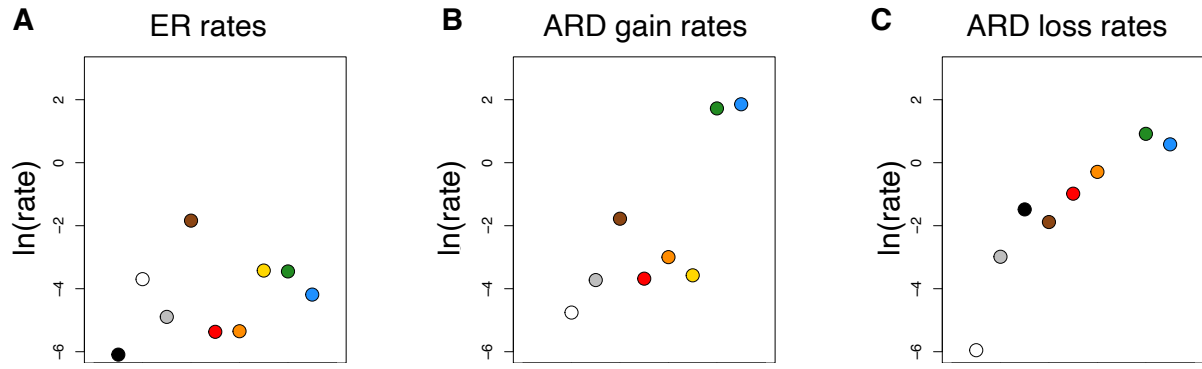


Figure 6. Plotted evolutionary rates for each discrete color category: (A) ER, (B) ARD gain, (C) ARD loss.

Tables

Table 1. Bird museum specimen image list.

Museum Identifier	Genus	Species	Common name
LACM 59610	<i>Tangara</i>	<i>arthus</i>	Golden Tanager
LACM 39216	<i>Stilpnia</i>	<i>cayana</i>	Burnished-buff Tanager
LACM 59272	<i>Tangara</i>	<i>chilensis</i>	Paradise Tanager
LACM 59652	<i>Tangara</i>	<i>cyanocephala</i>	Red-necked Tanager
LACM 37716	<i>Stilpnia</i>	<i>cyanoptera</i>	Black-headed Tanager
LACM 27866	<i>Tangara</i>	<i>cyanovenstris</i>	Gilt-edged Tanager
LACM 28775	<i>Tangara</i>	<i>desmaresti</i>	Brassy-breasted Tanager
LACM 16290	<i>Tangara</i>	<i>dowii</i>	Spangle-cheeked Tanager
LACM 60421	<i>Tangara</i>	<i>fastuosa</i>	Seven-colored Tanager
LACM 60422	<i>Tangara</i>	<i>florida</i>	Emerald Tanager
LACM 4843	<i>Ixothraupis</i>	<i>guttata</i>	Speckled Tanager
LACM 43661	<i>Tangara</i>	<i>gyrola</i>	Bay-headed Tanager
LACM 29385	<i>Stilpnia</i>	<i>heinei</i>	Black-capped Tanager
LACM 30414	<i>Tangara</i>	<i>inornata</i>	Plain-colored Tanager
LACM 37463	<i>Tangara</i>	<i>johannae</i>	Blue-whiskered Tanager
LACM 59219	<i>Tangara</i>	<i>labradorides</i>	Metallic-green Tanager
LACM 4841	<i>Stilpnia</i>	<i>larvata</i>	Golden-headed Tanager
LACM 29473	<i>Tangara</i>	<i>lavinia</i>	Rufous-winged Tanager
LACM 32721	<i>Tangara</i>	<i>mexicana</i>	Turquoise Tanager
LACM 40998	<i>Tangara</i>	<i>nigroviridis</i>	Beryl-spangled Tanager
LACM 37481	<i>Poecilostreptus</i>	<i>palmeri</i>	Gray-and-gold Tanager
LACM 29400	<i>Tangara</i>	<i>parzudakii</i>	Flame-faced Tanager
LACM 53462	<i>Stilpnia</i>	<i>preciosa</i>	Chestnut-backed Tanager
LACM 43685	<i>Ixothraupis</i>	<i>punctata</i>	Spotted Tanager
LACM 34880	<i>Chalcothraupis</i>	<i>ruficervix</i>	Golden-naped Tanager
LACM 16611	<i>Ixothraupis</i>	<i>rufigula</i>	Rufous-throated Tanager
LACM 50757	<i>Tangara</i>	<i>schrankii</i>	Green-and-gold Tanager
LACM 53515	<i>Tangara</i>	<i>seledon</i>	Green-headed Tanager
LACM 29453	<i>Tangara</i>	<i>vassori</i>	Blue-and-black Tanager
LACM 43655	<i>Tangara</i>	<i>velia</i>	Opal-rumped Tanager
LACM 36850	<i>Stilpnia</i>	<i>vitriolina</i>	Scrub Tanager
LACM 33255	<i>Tangara</i>	<i>xanthocephala</i>	Saffron-crowned Tanager
LACM 50758	<i>Ixothraupis</i>	<i>xanthogastra</i>	Yellow-bellied Tanager

References

- Alfaro, M. E., Karan, E. A., Schwartz, S. T., & Shultz, A. J. (2019). The evolution of color pattern in butterflyfishes (chaetodontidae). *Integrative and Comparative Biology*, *59*(3), 604–615.
- Andersson, S., & Amundsen, T. (1997). Ultraviolet colour vision and ornamentation in bluethroats. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *264*(1388), 1587–1591.
- Barlow, A., Cahill, J. A., Hartmann, S., Theunert, C., Xenikoudakis, G., Fortes, G. G., Paijmans, J. L., Rabeder, G., Frischauf, C., Grandal-d'Anglade, A., et al. (2018). Partial genomic survival of cave bears in living brown bears. *Nature Ecology & Evolution*, *2*(10), 1563–1570.
- Berg, C. P. van den, Troscianko, J., Endler, J. A., Marshall, N. J., & Cheney, K. L. (2020). Quantitative colour pattern analysis (QCPA): A comprehensive framework for the analysis of colour patterns in nature. *Methods in Ecology and Evolution*, *11*(2), 316–332.
- Bonser, R. H. (1995). Melanin and the abrasion resistance of feathers. *The Condor*, *97*(2), 590–591.
- Burnham, K. P., Anderson, D. R., Burnham, K. P., & Anderson, D. R. (1998). *Practical use of the information-theoretic approach*. Springer.
- Burns, K. J., Shultz, A. J., Title, P. O., Mason, N. A., Barker, F. K., Klicka, J., Lanyon, S. M., & Lovette, I. J. (2014). Phylogenetics and diversification of tanagers (passeriformes: Thraupidae), the largest radiation of neotropical songbirds. *Molecular Phylogenetics and Evolution*, *75*, 41–77.

- Burns, K. J., Unitt, P., & Mason, N. A. (2016). A genus-level classification of the family thraupidae (class aves: Order passeriformes). *Zootaxa*, 4088(3), 329–354.
- Chan, I. Z., Stevens, M., & Todd, P. A. (2019). PAT-GEOM: A software package for the analysis of animal patterns. *Methods in Ecology and Evolution*, 10(4), 591–600.
- Cooney, C. R., Varley, Z. K., Nouri, L. O., Moody, C. J., Jardine, M. D., & Thomas, G. H. (2019). Sexual selection predicts the rate and direction of colour divergence in a large avian radiation. *Nature Communications*, 10(1), 1–9.
- Darwin, C. (1981). *The descent of man, and selections in relation to sex*.
- Delhey, K., Valcu, M., Muck, C., Dale, J., & Kempenaers, B. (2023). Evolutionary predictors of the specific colors of birds. *Proceedings of the National Academy of Sciences*, 120(34), e2217692120.
- Ehrlich, P., Talbot, F., Russell, B., & Anderson, G. (1977). The behaviour of chaetodontid fishes with special reference to lorenz’“poster colouration” hypothesis. *Journal of Zoology*, 183(2), 213–228.
- Eliason, C. M., Maia, R., & Shawkey, M. D. (2015). Modular color evolution facilitated by a complex nanostructure in birds. *Evolution*, 69(2), 357–367.
- Endler, J. A. (2012). A framework for analysing colour pattern geometry: Adjacent colours. *Biological Journal of the Linnean Society*, 107(2), 233–253.
- Endler, J. A., Cole, G. L., & Kranz, A. M. (2018). Boundary strength analysis: Combining colour pattern geometry and coloured patch visual properties for use in predicting behaviour and fitness. *Methods in Ecology and Evolution*, 9(12), 2334–2348.
- Feller, K. D., Jordan, T. M., Wilby, D., & Roberts, N. W. (2017). Selection of the intrinsic polarization properties of animal optical materials creates enhanced structural reflectivity

- and camouflage. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1724), 20160336.
- Håstad, O., Victorsson, J., & Ödeen, A. (2005). Differences in color vision make passerines less conspicuous in the eyes of their predators. *Proceedings of the National Academy of Sciences*, 102(18), 6391–6394.
- Hemingson, C. R., Cowman, P. F., & Bellwood, D. R. (2024). Analysing biological colour patterns from digital images: An introduction to the current toolbox. *Ecology and Evolution*, 14(3), e11045.
- Hemingson, C. R., Cowman, P. F., Hodge, J. R., & Bellwood, D. R. (2019). Colour pattern divergence in reef fish species is rapid and driven by both range overlap and symmetry. *Ecology Letters*, 22(1), 190–199.
- Hill, G. E. (2006). *Bird coloration, volume 2: Function and evolution*. Harvard University Press.
- iNaturalist*. (n.d.). <https://www.inaturalist.org>.
- Irwin, R. E. (1994). The evolution of plumage dichromatism in the new world blackbirds: Social selection on female brightness. *The American Naturalist*, 144(6), 890–907.
- Maia, R., Eliason, C. M., Bitton, P.-P., Doucet, S. M., & Shawkey, M. D. (2013). Pavo: An R package for the analysis, visualization and organization of spectral data. *Methods in Ecology and Evolution*, 4(10), 906–913.
- Maia, R., Gruson, H., Endler, J. A., & White, T. E. (2019). Pavo 2: New tools for the spectral and spatial analysis of colour in R. *Methods in Ecology and Evolution*, 10(7), 1097–1107.
- Maia, R., Rubenstein, D. R., & Shawkey, M. D. (2016). Selection, constraint, and the evolution of coloration in African starlings. *Evolution*, 70(5), 1064–1079.

- Mason, N. A., & Bowie, R. C. (2020). Plumage patterns: Ecological functions, evolutionary origins, and advances in quantification. *The Auk*, *137*(4), ukaa060.
- Mayr, E. et al. (1963). Animal species and evolution. *Animal Species and Evolution*.
- McMillan, W. O., Weigt, L. A., & Palumbi, S. R. (1999). Color pattern evolution, assortative mating, and genetic differentiation in brightly colored butterflyfishes (chaetodontidae). *Evolution*, *53*(1), 247–260.
- Neudecker, S. (1989). Eye camouflage and false eyespots: Chaetodontid responses to predators. In *The butterflyfishes: Success on the coral reef* (pp. 143–158). Springer.
- Powers, D. M. (2020). Evaluation: From precision, recall and f-measure to ROC, informedness, markedness and correlation. *arXiv Preprint arXiv:2010.16061*.
- Price-Waldman, R. M., Shultz, A. J., & Burns, K. J. (2020). Speciation rates are correlated with changes in plumage color complexity in the largest family of songbirds. *Evolution*, *74*(6), 1155–1169.
- Rabosky, A. R. D., Cox, C. L., Rabosky, D. L., Title, P. O., Holmes, I. A., Feldman, A., & McGuire, J. A. (2016). Coral snakes predict the evolution of mimicry across new world snakes. *Nature Communications*, *7*(1), 1–9.
- Revell, L. J. (2012). Phytools: An r package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, *2*, 217–223.
- Shultz, A. J., & Burns, K. J. (2017). The role of sexual and natural selection in shaping patterns of sexual dichromatism in the largest family of songbirds (aves: thraupidae). *Evolution*, *71*(4), 1061–1074.
- Stoddard, M. C., & Osorio, D. (2019). Animal coloration patterns: Linking spatial vision to quantitative analysis. *The American Naturalist*, *193*(2), 164–186.

- Stoddard, M. C., & Prum, R. O. (2011). How colorful are birds? Evolution of the avian plumage color gamut. *Behavioral Ecology*, *22*(5), 1042–1052.
- Sullivan, B. L., Wood, C. L., Iliff, M. J., Bonney, R. E., Fink, D., & Kelling, S. (2009). eBird: A citizen-based bird observation network in the biological sciences. *Biological Conservation*, *142*(10), 2282–2292.
- Terrill, R. S., & Shultz, A. J. (2023). Feather function and the evolution of birds. *Biological Reviews*, *98*(2), 540–566.
- Valvo, J. J., Rodd, F. H., Aponte, J. D., Daniel, M., Dwinell, K., Houle, D., & Hughes, K. A. (2020). Colormesh: A novel method for quantifying variation in complex color patterns. *bioRxiv*.
- Van Belleghem, S. M., Papa, R., Ortiz-Zuazaga, H., Hendrickx, F., Jiggins, C. D., Owen McMillan, W., & Counterman, B. A. (2018). Patternize: An r package for quantifying colour pattern variation. *Methods in Ecology and Evolution*, *9*(2), 390–398.
- Weller, H. I., Hiller, A. E., Lord, N. P., & Van Belleghem, S. M. (2024). Recolorize: An r package for flexible colour segmentation of biological images. *Ecology Letters*, *27*(2), e14378.
- Weller, H. I., & Westneat, M. W. (2019). Quantitative color profiling of digital images with earth mover's distance using the r package colordistance. *PeerJ*, *7*, e6398.

Supplemental Figures

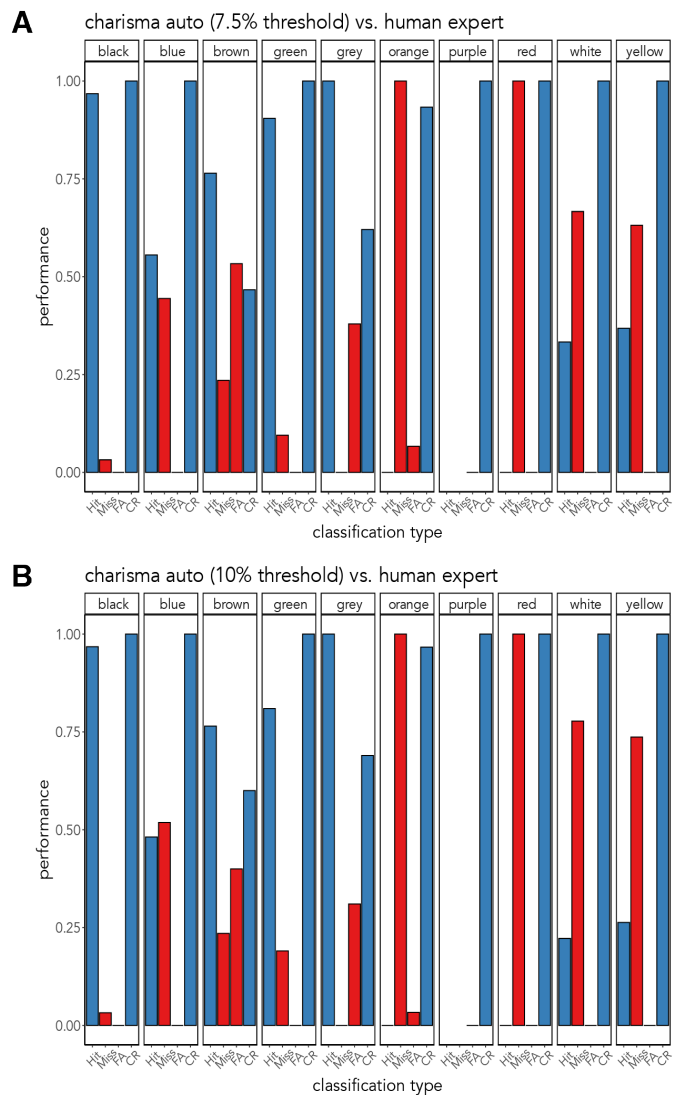


Figure S1. Plotted binary contingency table analyses comparing (A) the automated color dataset generated using *charisma* with a cutoff threshold of 7.5% to the expert color dataset, and (B) the automated color dataset generated using *charisma* with a cutoff threshold of 10% to the expert color dataset. Blue bars represent correct color calls or non-calls, and red bars represent false color calls or non-calls. Hit – true positive, Miss – false negative, FA (false alarm) – false positive, CR (correct rejection) – true negative.

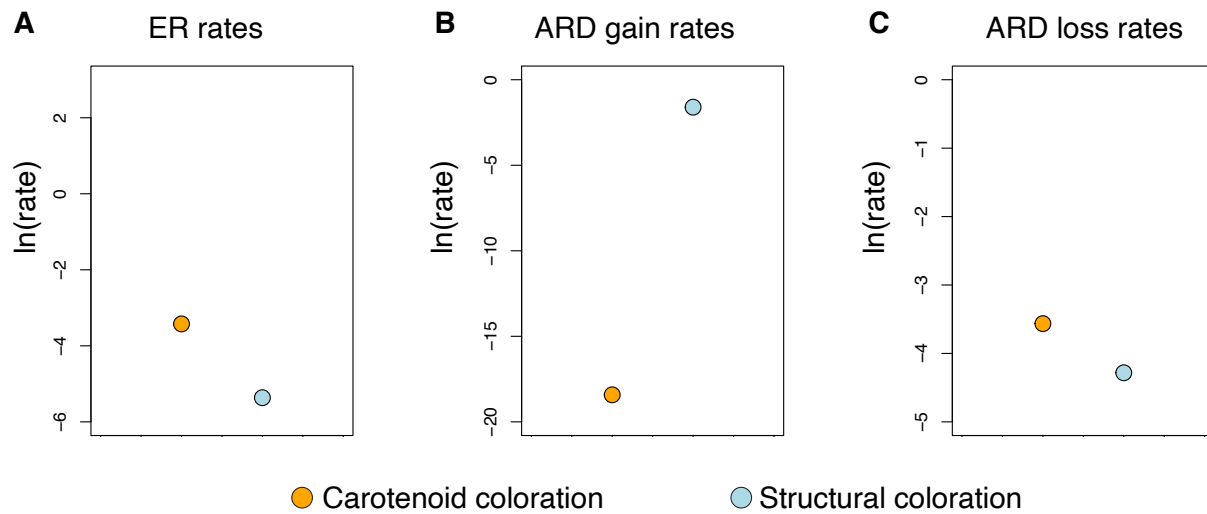


Figure S3. Plotted evolutionary rates for carotenoid and structural color producing mechanism:

(A) ER, (B) ARD gain, (C) ARD loss.

Supplemental Tables

Table S1. Example output matrix from running the manual *charisma* pipeline on bird museum specimens.

file_name	black	white	grey	brown	red	orange	yellow	green	blue	purple	k
Tangara_arthus_LACM59610_charisma_06-19-2024_11.49.17	1	0	0	1	0	0	1	0	0	0	3
Tangara_cayana_LACM39216_charisma_06-19-2024_12.07.11	1	0	0	1	0	0	0	0	1	0	3
Tangara_chilensis_LACM59272_charisma2_06-19-2024_15.42.40	1	0	0	0	0	0	0	1	1	0	3
Tangara_cyanocephala_LACM59652_charisma2_06-19-2024_16.04.56	1	0	0	0	1	0	1	1	1	0	5
Tangara_cyanooptera_LACM37716_charisma2_06-19-2024_16.19.49	1	0	0	1	0	0	1	0	1	0	4
Tangara_cyanoventris_LACM27866_charisma_06-19-2024_16.22.03	1	0	0	0	0	0	1	1	1	0	4
Tangara_desmaresti_LACM28775_charisma_06-19-2024_16.49.22	1	0	0	1	0	1	1	1	1	0	6
Tangara_dowii_LACM16290_charisma2_06-21-2024_09.49.31	1	0	0	1	0	0	0	1	1	0	4
Tangara_fastuosa_LACM60421_charisma_06-21-2024_09.52.39	1	0	0	0	0	1	1	1	1	0	5
Tangara_florida_LACM60422_charisma_06-21-2024_09.58.17	1	0	0	0	0	0	1	1	0	0	3
Tangara_guttata_LACM4843_charisma2_06-21-2024_10.23.06	1	1	0	0	0	0	1	1	1	0	5
Tangara_gyrola_LACM43661_charisma_06-21-2024_10.32.56	0	0	0	1	0	0	0	1	1	0	3
Tangara_heinei_LACM29385_charisma2_06-21-2024_11.03.17	1	0	1	0	0	0	0	1	1	0	4
Tangara_inornata_LACM30414_charisma2_06-21-2024_11.21.44	1	1	1	0	0	0	0	0	1	0	4
Tangara_johannae_LACM37463_charisma2_06-21-2024_15.24.38	1	0	0	0	0	0	1	1	1	0	4
Tangara_labradorides_LACM59124_charisma_06-24-2024_06.29.17	1	0	0	1	0	0	0	1	1	0	4
Tangara_larvata_LACM4843_charisma2_06-24-2024_06.50.28	1	1	0	1	0	0	1	1	1	0	6
Tangara_mexicana_LACM32721_charisma_06-24-2024_06.58.13	1	1	0	0	0	0	1	0	1	0	4
Tangara_nigroviridis_LACM40998_charisma_06-24-2024_07.09.04	1	1	0	0	0	0	0	0	1	0	3
Tangara_palmeri_LACM37481_charisma_06-24-2024_07.20.06	1	1	1	0	0	0	1	0	0	0	4
Tangara_parzudakii_LACM29400_charisma2_06-24-2024_07.51.41	1	0	0	1	1	0	1	1	1	0	6
Tangara_preciosa_LACM53462_charisma2_06-24-2024_14.56.31	1	1	0	1	0	0	0	1	1	0	5
Tangara_punctata_LACM43658_charisma_06-24-2024_15.07.18	1	1	0	0	0	0	1	1	0	0	4
Tangara_ruficervix_LACM34880_charisma_06-24-2024_16.06.32	1	1	0	1	0	0	0	0	1	0	4
Tangara_rufigula_LACM16611_charisma_06-24-2024_16.16.36	1	0	0	1	0	0	0	1	0	0	3
Tangara_schrankii_LACM50757_charisma_06-24-2024_16.19.53	1	0	0	1	0	0	1	1	1	0	5
Tangara_seledon_LACM53515_charisma_06-24-2024_16.22.53	1	0	0	1	0	0	1	1	1	0	5
Tangara_vassorii_LACM29453_charisma_06-24-2024_16.31.38	1	0	0	0	0	0	0	0	1	0	2
Tangara_velia_LACM43655_charisma_06-24-2024_16.37.57	1	0	0	1	0	0	1	0	1	0	4
Tangara_vitriolina_LACM36850_charisma_06-24-2024_16.41.25	1	0	0	1	0	0	0	1	0	0	3
Tangara_xanthocephala_LACM33255_charisma_06-24-2024_16.43.24	1	0	0	1	0	0	1	1	1	0	5
Tangara_xanthogastra_LACM50758_charisma_06-24-2024_16.50.08	1	0	0	0	0	0	1	1	0	0	3

Table S2. Citations and sources for Wikipedia images for *charisma* validation.

Color	Title	Author(s)	License	Source
Black	"Shades of Black"	Tony Mach	CC0	https://commons.wikimedia.org/w/index.php?curid=31335075
Blue	"Shades of the color blue"	Booyabazooka	Public Domain	https://commons.wikimedia.org/w/index.php?curid=935752
Brown	"Brown icon"	Booyabazooka, Mizunoryu, Badseed, Jacobolus	Public Domain	https://commons.wikimedia.org/w/index.php?curid=10511178
Green	"Color icon green"	Booyabazooka	CC BY-SA 3.0	https://commons.wikimedia.org/w/index.php?curid=3364583
Gray	"Gray icon"	Mizunoryu, Badseed, Jacobolus	Public Domain	https://commons.wikimedia.org/w/index.php?curid=10509947
Orange	"Brown and orange squares as color sample"	Booyabazooka	Public Domain	https://commons.wikimedia.org/w/index.php?curid=3364585
Purple	"Purple icon"	Mizunoryu, Badseed, Jacobolus	Public Domain	https://commons.wikimedia.org/w/index.php?curid=10509867
Red	"5x5 color square for red"	ThunderBrine	CC BY-SA 4.0	https://commons.wikimedia.org/w/index.php?curid=110954343
White	"Color icon white"	Badseed	Public Domain	https://commons.wikimedia.org/w/index.php?curid=3529704
Yellow	"Color icon yellow"	Badseed	Public Domain	https://commons.wikimedia.org/w/index.php?curid=3512367

CHAPTER 2

Evolution of visual system sensitivity in birds

Abstract

Color vision in birds is considered highly conserved evolutionarily and can be categorized into two main classes: violet-sensitive (VS) and ultraviolet-sensitive (UVS). The visual system sensitivity classes can be inferred for most bird species by sequencing the short-wavelength-sensitive opsin (SWS1). Due to previous limitations in sequencing costs and the availability of genomic data for many bird species, analyses exploring the phylogenetic distribution of visual system sensitivity have been limited to less than 100 species. Here, we use genomic data and a resolved phylogeny of all bird species to conduct the most comprehensive investigation into visual system sensitivity evolution in birds to date. We aligned and trimmed SWS1 sequences to the amino acid tuning sites that predict lambda max for each species to infer visual system sensitivity. We tested three models of evolution and determined that a model where transition rates from VS to UVS are different from transition rates from UVS to VS. We recovered SWS1 sequences from 418 species, including at least one representative from 85% of all avian orders and 67% of all avian families. We found that the most recent common ancestor of birds likely had a VS visual system. Our analyses show that transitions from a VS to UVS have occurred at least 18 times in the evolutionary history of birds and that there are few transitions from UVS to VS.

Introduction

The visual system in vertebrates is highly conserved, but subtle changes and spectral fine-tuning of the system can lead to variations in visual perception between organisms (Bowmaker, 2008). The avian eye uses a lens to focus light on a retina containing two photoreceptor cell types. Rods detect dim light and function in peripheral and night vision, while the cones are the

photoreceptors that respond to bright light and are used in color vision. Each cone has an associated visual pigment and the protein component of this pigment, called an opsin, determines its spectral sensitivity to different wavelengths of light (Hart & Hunt, 2007; Yokoyama et al., 2000). Birds possess tetrachromatic color vision with cones that are long, medium, short, and ultraviolet wavelength sensitive (Hart, 2001; Osorio & Vorobyev, 2008). The addition of a fourth single-cone type in their eyes facilitates this expanded color range, and the visual system can either be classified as violet-sensitive (VS) or ultraviolet-sensitive (UVS) depending on the amino acid sequence of the short-wavelength-sensitive opsin (SWS1).

The amino acid sequence of the SWS1 opsin in birds has been shown to accurately predict a VS and UVS visual system (Borges et al., 2015; Carvalho et al., 2007; Hauser et al., 2014; Ödeen & Håstad, 2003; Yokoyama et al., 2000). This shift in visual system sensitivity is determined by the amino acids at sites 86, 90, and 93 in the SWS1 opsin sequence (Shi & Yokoyama, 2003; Wilkie et al., 2000; Yokoyama et al., 2000). This provides a simple prediction of bird visual system sensitivity and makes for a powerful tool for understanding the evolution of avian color perception (Hauser et al., 2014).

The increasing availability of genomic data and a resolved phylogeny of all bird species allows for large-scale comparative studies of the evolution of the avian visual system (Feng et al., 2020; Jetz et al., 2012; Prum et al., 2015). While previous studies have investigated visual system sensitivity across the avian tree, sampling has been limited to one or a few individuals in an order or family even though recent evidence suggests that shifts between VS and UVS can occur at this level (Borges et al., 2015; Friedman & Remeš, 2015; Ödeen et al., 2011). By scanning the SWS1 opsin of all genome-enabled bird species for evidence of VS or UVS system,

we conduct the most comprehensive assessment of bird visual system sensitivity and reveal how it has evolved across the avian tree of life.

Methods

Predicting visual system sensitivity

To predict the visual system sensitivity of each bird species, we queried the National Center for Biotechnology and Information (NCBI) for all available whole genome and SWS1 opsin sequences. We used Geneious Prime and MEGA X (Kumar et al., 2018) to align and trim SWS1 opsin sequences to nucleotide positions 252-282. We translated the DNA sequences to amino acids 84-94. We then used the spectral tuning sites at amino acids 86, 90, and 93 to predict lambda max values and estimate a VS or UVS visual system for each species (Table S1).

Evolution of visual system sensitivity

We used our SWS1 opsin dataset and a previously published time-calibrated phylogeny of birds (Jetz et al., 2012) to map visual system sensitivity across the avian tree of life. We used the fitDiscrete function in the R package *geiger* to test the fit of three different models of evolution: equal rates (ER), all rates different (ARD), and symmetric (SYM) (Pennell et al., 2014). We compared models using sample size corrected Akaike Information Criterion (AICc). We also tested the following transformations to test models where rates vary across the tree: no transformation, Early-burst (EB), lambda, kappa, and delta (Harmon et al., 2010; Pagel, 1999; Pennell et al., 2014). We selected and used the best-fitting model to reconstruct ancestral states by performing 100 stochastic character maps using a Bayesian approach with Markov Chain Monte Carlo (MCMC) in the R package *phytools* (Bollback, 2006; Burnham & Anderson, 2003).

Results

Predicting visual system sensitivity

We queried 610 genome and SWS1 sequences for the SWS1 amino acid tuning sites. We eliminated Bobolink (*Dolichonyx oryxivorus*) from our analyses because previous work showed a conflict between having a VS or UVS visual system either because the Bobolink expresses two SWS1 opsin pigments or because it has a mutated SWS2 opsin pigment that closely resembles the VS SWS1 opsin (Beason & Loew, 2008). Removing sequences of the same species resulted in genome sequences from 128 species and SWS1 opsin sequences from 377 species (505 total). We recovered the tuning sites for 430 species (85%) with data for 53 species (13%) coming from genome sequences and 374 species (87%) from SWS1 opsin sequences. Removing species for which we could not predict visual system sensitivity, we were left with a final dataset of 418 species. This dataset included visual system sensitivities for 35 of 41 bird orders (85%), 169 of 251 bird families (67%), and 418 of 11,017 bird species of the world (4%).

Evolution of visual system sensitivity in birds

Our analysis of visual system sensitivities from 418 species indicates that the ARD model best fits our data, where gains and losses differ in rate (Table 1). The ARD model reveals that the evolution from a VS to a UVS visual system occurs faster than from a UVS to a VS visual system. The ARD kappa transformation, where character divergence is related to the number of speciation events between two species, best fits our data (Table 2). However, interpretation under the kappa transformation is difficult because we have incomplete sampling across our tree (67% of bird families and 4% of all bird species). Thus, we used the ARD model without transformation to reconstruct ancestral states.

Discussion

We found that a VS visual system is likely to be ancestral in birds (Figure 4). Our results indicate that UVS may have evolved independently in the visual system of birds up to 18 times, which is

more than previous analyses of visual system sensitivity in birds (Borges et al., 2015; Ödeen & Håstad, 2013). The bird families in which we recovered visual system sensitivity for the most species included Laridae (28 of 100 species, 12 of 23 genera), Maluridae (16 of 32 species, 4 of 6 genera), Parulidae (16 of 115 species, 7 of 18 genera), Thraupidae (13 of 384 species, 13 of 107 genera), and Ptilonorhynchidae (12 of 27 species, 7 of 8 genera). The UVS visual system has evolved in all species sampled from the Casuaridae, Tinamidae, Pteroclididae, Trogonidae, Momotidae, Strigopidae, Cacatuidae, Psittacidae, Menuridae, Orthonychidae, Callaeidae, Melanocharitidae, and Vireonidae families. Only three families have both VS and UVS visual systems present: Laridae, Maluridae, and Thamnophilidae.

Across the Charadriiformes, the UVS visual system is rare. However, our results indicate two independent evolutionary events of UVS in the Laridae family. We find the evolution of UVS in the White Tern, *Gygis alba*, and another in the group containing noddies and gulls represented by 16 species in our phylogeny. The retinas of birds are at risk of photodamage due to exposure to UV radiation and birds with VS visual systems filter UV more effectively than those with UVS visual systems (Carvalho et al., 2011). In general, UV radiation at the water's surface is high, increasing the risk of photodamage, which may be why UVS is rare in seabirds (Losey et al., 1999). Although fish are known to exhibit UV coloration (Losey et al., 1999), many seabirds feed on this prey, indicating that the UVS visual system is not necessary to locate fish. Therefore, birds in the family Laridae may have evolved the UVS visual system to adapt to their omnivorous foraging habits and to enable foraging in both terrestrial and coastal areas (Håstad et al., 2005).

The avian order Caprimulgiformes contains four families of mostly nocturnal birds (Caprimulgidae, Nyctibiidae, Steatornithidae, and Aegothelidae) as well as swifts (Apodidae)

and hummingbirds (Trochilidae). Birds in this order proved difficult to recover amino acid tuning site data from genetic sequences. Of 17 sequences, we recovered tuning sites and predicted visual system sensitivity for just six species (35%) in this order. In previous studies, only partial sequences of the SWS1 opsin were recovered, revealing a VS visual system in all species in this order (Feng et al., 2020; Ödeen & Håstad, 2010). However, multiple studies have shown that hummingbirds can perceive and distinguish UV colors well (Herrera et al., 2008; Stoddard et al., 2020). One possibility is that these partial sequences may reflect a mutated SWS2 opsin instead of a true SWS1 opsin and that birds in this order lost the SWS1 opsin (Beason & Loew, 2008). Owls are also known to be missing the SWS1 opsin and given that hummingbirds may have evolved from a nocturnal ancestor, it is possible that they lost this visual pigment and have evolved a novel pathway to achieve UVS (Feng et al., 2020; Höglund et al., 2019). This would not be the first time hummingbirds developed a novel sensory pathway, as this has also been shown in the sweet taste receptor in hummingbirds (Toda et al., 2021).

In Passeriformes, UVS has evolved at least once in the Suboscines and up to ten times within the Oscines. In the Suboscine antbird family, *Thamnophilidae*, the Black-crowned Antshrike, *Thamnophilus atrinucha*, was the only individual to exhibit a UVS visual system. An earlier study of antbirds found them to have only a VS visual system and hypothesized a link between low UV-reflecting plumage and VS visual system (Seddon et al., 2010). However, our study agrees with a more recent study showing the presence of the UVS visual system in this group (Dell’Aglia et al., 2018). Further investigation into the antbirds and the Suboscines may reveal more transitions between the VS and UVS visual systems. In Oscine passerines, UVS has evolved multiple times and the fairy-wrens in the genus *Malurus* (family *Maluridae*) present an interesting case in which the UVS visual system has evolved up to four times in a single genus.

This is the only example of visual system sensitivity shifting within a single genus to date and has been associated with shifts in plumage changes in this genus (Ödeen et al., 2012). However, although birds in the Infraorder Passerides experience similar shifts in plumage coloration from drab to colorful, the UVS visual system is ubiquitous across this Infraorder suggesting another driver of visual system sensitivity fixation in this group.

Continued genomic investigation into the visual system sensitivities of all birds will likely result in the discovery of more independent evolutions of the UVS visual system and cases where VS and UVS visual systems are present within families and even genera. Using results from this study with the growing body of literature on color diversity in birds will allow for future studies investigating the link between the visual system sensitivity, color production, and life history of birds.

Figures

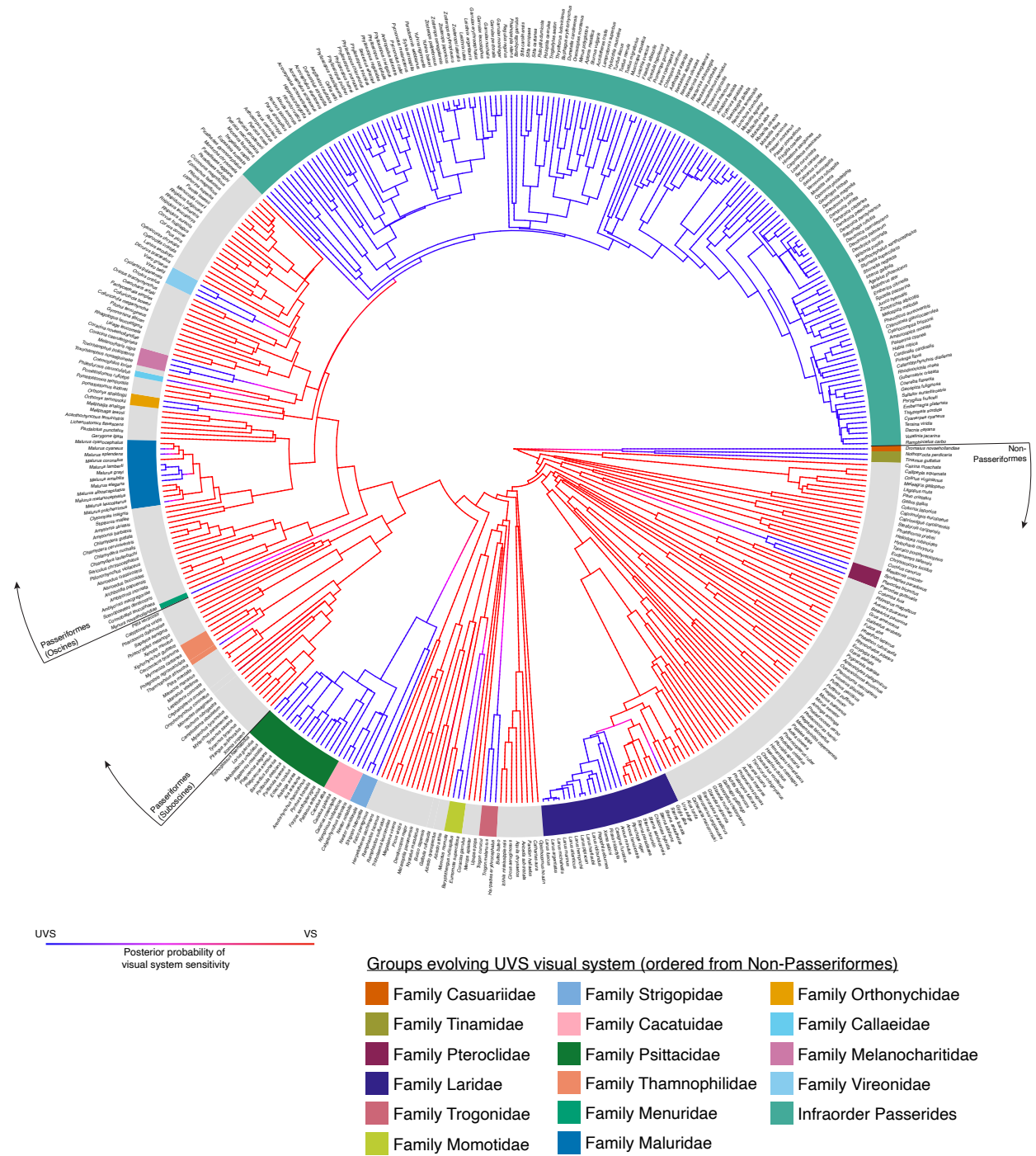


Figure 1. Phylogenetic distribution of VS and UVS visual systems across the bird tree of life.

The branch color represents the Bayesian posterior probability of VS (red) or UVS (blue) visual

system. Groups that have evolved a UVS visual system are highlighted, and the key is ordered starting with the Family Casuariidae at the beginning of the non-Passeriformes.

Tables

Table 1. Rates and AICc weights for the Equal Rates, All Rates Different, and Symmetric models of evolution.

	ER	ARD	SYM
VS-UVS Rate	0.003245	0.004772	0.003245
UVS-VS Rate	0.003245	5.263E-17	0.003245
AICC weight	0.01	0.98	0.01

Table 2. Rates and AICc weights for the All Rates Different model of evolution with transformations: Early Burst, lambda, kappa, and delta.

	ARD	ARD (Early Burst)	ARD (lambda)	ARD (kappa)	ARD (delta)
VS-UVS Rate	0.004772	0.000265	0.00426	0.0238	0.004101
UVS-VS Rate	5.263E-17	4.438E-17	3.39E-17	1.48E-15	2.61E-17
AICC weight	0.05	0.02	0.04	0.86	0.02
AICC weight (without kappa transformation)	0.37	0.17	0.31	NA	0.15

Literature Cited

- Beason, R. C., & Loew, E. R. (2008). Visual pigment and oil droplet characteristics of the bobolink (*Dolichonyx oryzivorus*), a new world migratory bird. *Vision Research*, *48*(1), 1–8. <https://doi.org/10.1016/j.visres.2007.10.006>
- Bollback, J. P. (2006). SIMMAP: Stochastic character mapping of discrete traits on phylogenies. *BMC Bioinformatics*, *7*(1), 88. <https://doi.org/10.1186/1471-2105-7-88>
- Borges, R., Khan, I., Johnson, W. E., Gilbert, M. T. P., Zhang, G., Jarvis, E. D., O'Brien, S. J., & Antunes, A. (2015). Gene loss, adaptive evolution and the co-evolution of plumage coloration genes with opsins in birds. *BMC Genomics*, *16*(1), 751. <https://doi.org/10.1186/s12864-015-1924-3>
- Bowmaker, J. K. (2008). Evolution of vertebrate visual pigments. *Vision Research*, *48*(20), 2022–2041. <https://doi.org/10.1016/j.visres.2008.03.025>
- Burnham, K. P., & Anderson, D. R. (2003). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer Science & Business Media.
- Carvalho, L. S., Cowing, J. A., Wilkie, S. E., Bowmaker, J. K., & Hunt, D. M. (2007). The Molecular Evolution of Avian Ultraviolet- and Violet-Sensitive Visual Pigments. *Molecular Biology and Evolution*, *24*(8), 1843–1852. <https://doi.org/10.1093/molbev/msm109>
- Carvalho, L. S., Knott, B., Berg, M. L., Bennett, A. T. D., & Hunt, D. M. (2011). Ultraviolet-sensitive vision in long-lived birds. *Proceedings of the Royal Society B: Biological Sciences*, *278*(1702), 107–114. <https://doi.org/10.1098/rspb.2010.1100>

- Dell’Aglío, D. D., Troscianko, J., McMillan, W. O., Stevens, M., & Jiggins, C. D. (2018). The appearance of mimetic *Heliconius* butterflies to predators and conspecifics. *Evolution*, 72(10), 2156–2166. <https://doi.org/10.1111/evo.13583>
- Feng, S., Stiller, J., Deng, Y., Armstrong, J., Fang, Q., Reeve, A. H., Xie, D., Chen, G., Guo, C., Faircloth, B. C., Petersen, B., Wang, Z., Zhou, Q., Diekhans, M., Chen, W., Andreu-Sánchez, S., Margaryan, A., Howard, J. T., Parent, C., ... Zhang, G. (2020). Dense sampling of bird diversity increases power of comparative genomics. *Nature*, 587(7833), Article 7833. <https://doi.org/10.1038/s41586-020-2873-9>
- Friedman, N. R., & Remeš, V. (2015). Rapid evolution of elaborate male coloration is driven by visual system in Australian fairy-wrens (Maluridae). *Journal of Evolutionary Biology*, 28(12), 2125–2135. <https://doi.org/10.1111/jeb.12737>
- Harmon, L. J., Losos, J. B., Jonathan Davies, T., Gillespie, R. G., Gittleman, J. L., Bryan Jennings, W., Kozak, K. H., McPeck, M. A., Moreno-Roark, F., Near, T. J., Purvis, A., Ricklefs, R. E., Schluter, D., Schulte II, J. A., Seehausen, O., Sidlauskas, B. L., Torres-Carvajal, O., Weir, J. T., & Mooers, A. Ø. (2010). Early Bursts of Body Size and Shape Evolution Are Rare in Comparative Data. *Evolution*, 64(8), 2385–2396. <https://doi.org/10.1111/j.1558-5646.2010.01025.x>
- Hart, N. S. (2001). The Visual Ecology of Avian Photoreceptors. *Progress in Retinal and Eye Research*, 20(5), 675–703. [https://doi.org/10.1016/S1350-9462\(01\)00009-X](https://doi.org/10.1016/S1350-9462(01)00009-X)
- Hart, N. S., & Hunt, D. M. (2007). Avian Visual Pigments: Characteristics, Spectral Tuning, and Evolution. *The American Naturalist*, 169(S1), S7–S26. <https://doi.org/10.1086/510141>

- Håstad, O., Ernstdotter, E., & Ödeen, A. (2005). Ultraviolet vision and foraging in dip and plunge diving birds. *Biology Letters*, *1*(3), 306–309.
<https://doi.org/10.1098/rsbl.2005.0320>
- Hauser, F. E., Hazel, I. van, & Chang, B. S. W. (2014). Spectral tuning in vertebrate short wavelength-sensitive 1 (SWS1) visual pigments: Can wavelength sensitivity be inferred from sequence data? *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, *322*(7), 529–539. <https://doi.org/10.1002/jez.b.22576>
- Herrera, G., Zagal, J. C., Diaz, M., Fernández, M. J., Vielma, A., Cure, M., Martinez, J., Bozinovic, F., & Palacios, A. G. (2008). Spectral sensitivities of photoreceptors and their role in colour discrimination in the green-backed firecrown hummingbird (*Sephanoides sephaniodes*). *Journal of Comparative Physiology A*, *194*(9), 785.
<https://doi.org/10.1007/s00359-008-0349-8>
- Höglund, J., Mitkus, M., Olsson, P., Lind, O., Drews, A., Bloch, N. I., Kelber, A., & Strandh, M. (2019). Owls lack UV-sensitive cone opsin and red oil droplets, but see UV light at night: Retinal transcriptomes and ocular media transmittance. *Vision Research*, *158*, 109–119.
<https://doi.org/10.1016/j.visres.2019.02.005>
- Jetz, W., Thomas, G. H., Joy, J. B., Hartmann, K., & Mooers, A. O. (2012). The global diversity of birds in space and time. *Nature*, *491*(7424), 444–448.
<https://doi.org/10.1038/nature11631>
- Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular Biology and Evolution*, *35*(6), 1547–1549. <https://doi.org/10.1093/molbev/msy096>

- Losey, G. S., Cronin, T. W., Goldsmith, T. H., Hyde, D., Marshall, N. J., & McFarland, W. N. (1999). The UV visual world of fishes: A review. *Journal of Fish Biology*, *54*(5), 921–943. <https://doi.org/10.1111/j.1095-8649.1999.tb00848.x>
- Ödeen, A., & Håstad, O. (2003). Complex Distribution of Avian Color Vision Systems Revealed by Sequencing the SWS1 Opsin from Total DNA. *Molecular Biology and Evolution*, *20*(6), 855–861. <https://doi.org/10.1093/molbev/msg108>
- Ödeen, A., & Håstad, O. (2010). Pollinating birds differ in spectral sensitivity. *Journal of Comparative Physiology A*, *196*(2), 91–96. <https://doi.org/10.1007/s00359-009-0474-z>
- Ödeen, A., & Håstad, O. (2013). The phylogenetic distribution of ultraviolet sensitivity in birds. *BMC Evolutionary Biology*, *13*(1), 36. <https://doi.org/10.1186/1471-2148-13-36>
- Ödeen, A., Håstad, O., & Alström, P. (2011). Evolution of ultraviolet vision in the largest avian radiation—The passerines. *BMC Evolutionary Biology*, *11*(1), 313. <https://doi.org/10.1186/1471-2148-11-313>
- Ödeen, A., Pruett-Jones, S., Driskell, A. C., Armenta, J. K., & Håstad, O. (2012). Multiple shifts between violet and ultraviolet vision in a family of passerine birds with associated changes in plumage coloration. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1732), 1269–1276. <https://doi.org/10.1098/rspb.2011.1777>
- Osorio, D., & Vorobyev, M. (2008). A review of the evolution of animal colour vision and visual communication signals. *Vision Research*, *48*(20), 2042–2051. <https://doi.org/10.1016/j.visres.2008.06.018>
- Pagel, M. (1999). Inferring the historical patterns of biological evolution. *Nature*, *401*(6756), 877–884. <https://doi.org/10.1038/44766>

- Pennell, M. W., Eastman, J. M., Slater, G. J., Brown, J. W., Uyeda, J. C., FitzJohn, R. G., Alfaro, M. E., & Harmon, L. J. (2014). geiger v2.0: An expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. *Bioinformatics*, *30*(15), 2216–2218. <https://doi.org/10.1093/bioinformatics/btu181>
- Prum, R. O., Berv, J. S., Dornburg, A., Field, D. J., Townsend, J. P., Lemmon, E. M., & Lemmon, A. R. (2015). A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature*, *526*(7574), 569–573. <https://doi.org/10.1038/nature15697>
- Seddon, N., Tobias, J. A., Eaton, M., & Ödeen, A. (2010). Human Vision can Provide a Valid Proxy for Avian Perception of Sexual Dichromatism. *The Auk*, *127*(2), 283–292. <https://doi.org/10.1525/auk.2009.09070>
- Shi, Y., & Yokoyama, S. (2003). Molecular analysis of the evolutionary significance of ultraviolet vision in vertebrates. *Proceedings of the National Academy of Sciences*, *100*(14), 8308–8313. <https://doi.org/10.1073/pnas.1532535100>
- Stoddard, M. C., Eyster, H. N., Hogan, B. G., Morris, D. H., Soucy, E. R., & Inouye, D. W. (2020). Wild hummingbirds discriminate nonspectral colors. *Proceedings of the National Academy of Sciences*. <https://doi.org/10.1073/pnas.1919377117>
- Toda, Y., Ko, M.-C., Liang, Q., Miller, E. T., Rico-Guevara, A., Nakagita, T., Sakakibara, A., Uemura, K., Sackton, T., Hayakawa, T., Sin, S. Y. W., Ishimaru, Y., Misaka, T., Oteiza, P., Crall, J., Edwards, S. V., Buttemer, W., Matsumura, S., & Baldwin, M. W. (2021). Early origin of sweet perception in the songbird radiation. *Science*, *373*(6551), 226–231. <https://doi.org/10.1126/science.abf6505>

Wilkie, S. E., Robinson, P. R., Cronin, T. W., Poopalasundaram, S., Bowmaker, J. K., & Hunt, D. M. (2000). Spectral Tuning of Avian Violet- and Ultraviolet-Sensitive Visual

Pigments. *Biochemistry*, 39(27), 7895–7901. <https://doi.org/10.1021/bi992776m>

Yokoyama, S., Radlwimmer, F. B., & Blow, N. S. (2000). Ultraviolet pigments in birds evolved from violet pigments by a single amino acid change. *Proceedings of the National Academy of Sciences*, 97(13), 7366–7371. <https://doi.org/10.1073/pnas.97.13.7366>

Supplemental Materials

Table S1. Amino acid tuning sites for predicting VS and UVS visual system.

86	90	93	Visual System	References
A	C	T	UVS	Bowmaker et al., 1997
				Hart & Bennett, 2000, Hart & Cuthill, 2007, Das et al., 1999, Beason & Loew, 2008, Bowmaker et al., 1997, Hart et al., 1998
C	C	T	UVS	
F	C	M	UVS	Hart et al., 2016
F	C	M	UVS	Ödeen & Håstad, 2003
F	S	T	UVS	Hunt et al., 2007
I	C	I	UVS	Ödeen & Håstad, 2003
M	C	T	UVS	Maier & Bowmaker, 1993
A	S	T	VS	Beason & Loew, 2008
C	S	T	VS	Carvalho et al., 2007
C	S	T	VS	Porter et al., 2015
C	S	I	VS	Ödeen et al., 2012
S	S	V	VS	Hart et al., 1998, Okano et al., 1992, Hart, 2002
S	S	I	VS	Jane & Bowmaker, 1988
				Bowmaker et al., 1997, Hart, 2004, Bowmaker & Martin, 1985
S	S	T	VS	
T	S	T	VS	Håstad et al., 2005

Supplemental references

Beason, R. C., & Loew, E. R. (2008). Visual pigment and oil droplet characteristics of the bobolink (*Dolichonyx oryzivorus*), a new world migratory bird. *Vision Research*, 48(1), 1–8. <https://doi.org/10.1016/j.visres.2007.10.006>

Bowmaker, J. K., Heath, L. A., Wilkie, S. E., & Hunt, D. M. (1997). Visual Pigments and Oil Droplets from Six Classes of Photoreceptor in the Retinas of Birds. 12.

Bowmaker, J. K., & Martin, G. R. (1985). Visual pigments and oil droplets in the penguin, *Spheniscus humboldti*. *Journal of Comparative Physiology A*, 156(1), 71–77. <https://doi.org/10.1007/BF00610668>

- Carvalho, L. S., Cowing, J. A., Wilkie, S. E., Bowmaker, J. K., & Hunt, D. M. (2007). The Molecular Evolution of Avian Ultraviolet- and Violet-Sensitive Visual Pigments. *Molecular Biology and Evolution*, 24(8), 1843–1852. <https://doi.org/10.1093/molbev/msm109>
- Das, D., Wilkie, S. E., Hunt, D. M., & Bowmaker, J. K. (1999). Visual pigments and oil droplets in the retina of a passerine bird, the canary *Serinus canaria*: Microspectrophotometry and opsin sequences. *Vision Research*, 39(17), 2801–2815. [https://doi.org/10.1016/S0042-6989\(99\)00023-1](https://doi.org/10.1016/S0042-6989(99)00023-1)
- Hart, N. S. (2002). Vision in the peafowl. *Journal of Experimental Biology*, 205(24), 3925–3935. <https://doi.org/10.1242/jeb.205.24.3925>
- Hart, N. S. (2004). Microspectrophotometry of visual pigments and oil droplets in a marine bird, the wedge-tailed shearwater *Puffinus pacificus*: Topographic variations in photoreceptor spectral characteristics. *Journal of Experimental Biology*, 207(7), 1229–1240. <https://doi.org/10.1242/jeb.00857>
- Hart, N. S., & Bennett, A. T. D. (2000). Visual pigments, oil droplets, ocular media and cone photoreceptor distribution in two species of passerine bird: The blue tit (*Parus caeruleus* L.) and the blackbird (*Turdus merula* L.). 13.
- Hart, N. S., & Cuthill, I. C. (2000). Visual pigments, cone oil droplets and ocular media in four species of estrildid *in*ch. . . B, 14.
- Hart, N. S., Mountford, J. K., Davies, W. I. L., Collin, S. P., & Hunt, D. M. (2016). Visual pigments in a palaeognath bird, the emu *Dromaius novaehollandiae*: Implications for spectral sensitivity and the origin of ultraviolet vision. *Proceedings of the Royal Society B: Biological Sciences*, 283(1834), 20161063. <https://doi.org/10.1098/rspb.2016.1063>

- Hart, N. S., Partridge, J. C., & Cuthill, I. C. (1998). Visual pigments, oil droplets, oil droplets and cone photoreceptor distribution in the European Starling (*Sturnus vulgaris*). *Journal of Experimental Biology*, 201(9), 1433–1446. <https://doi.org/10.1242/jeb.201.9.1433>
- Håstad, O., Ernstdotter, E., & Ödeen, A. (2005). Ultraviolet vision and foraging in dip and plunge diving birds. *Biology Letters*, 1(3), 306–309. <https://doi.org/10.1098/rsbl.2005.0320>
- Hunt, D. M., Carvalho, L. S., Cowing, J. A., Parry, J. W. L., Wilkie, S. E., Davies, W. L., & Bowmaker, J. K. (2007). Spectral Tuning of Shortwave-sensitive Visual Pigments in Vertebrates†. *Photochemistry and Photobiology*, 83(2), 303–310. <https://doi.org/10.1562/2006-06-27-IR-952>
- Jane, S. D., & Bowmaker, J. K. (1988). Tetrachromatic colour vision in the duck (*Anas platyrhynchos* L.): microspectrophotometry of visual pigments and oil droplets. *Journal of Comparative Physiology A*, 162(2), 225–235. <https://doi.org/10.1007/BF00606087>
- Maier, E. J., & Bowmaker, J. K. (1993). Colour vision in the passeriform bird, *Leiothrix lutea*: Correlation of visual pigment absorbance and oil droplet transmission with spectral sensitivity. *Journal of Comparative Physiology A*, 172(3), 295–301. <https://doi.org/10.1007/BF00216611>
- Ödeen, A., & Håstad, O. (2003). Complex Distribution of Avian Color Vision Systems Revealed by Sequencing the SWS1 Opsin from Total DNA. *Molecular Biology and Evolution*, 20(6), 855–861. <https://doi.org/10.1093/molbev/msg108>
- Ödeen, A., Pruett-Jones, S., Driskell, A. C., Armenta, J. K., & Håstad, O. (2012). Multiple shifts between violet and ultraviolet vision in a family of passerine birds with associated

changes in plumage coloration. *Proceedings of the Royal Society B: Biological Sciences*, 279(1732), 1269–1276. <https://doi.org/10.1098/rspb.2011.1777>

Okano, T., Kojima, D., Fukada, Y., Shichida, Y., & Yoshizawa, T. (1992). Primary structures of chicken cone visual pigments: Vertebrate rhodopsins have evolved out of cone visual pigments. *Proceedings of the National Academy of Sciences*, 89(13), 5932–5936. <https://doi.org/10.1073/pnas.89.13.5932>

Porter, M. L., Kingston, A. C. N., McCready, R., Cameron, E. G., Hofmann, C. M., Suarez, L., Olsen, G. H., Cronin, T. W., & Robinson, P. R. (2014). Characterization of visual pigments, oil droplets, lens and cornea in the whooping crane *Grus americana*. *Journal of Experimental Biology*, 217(21), 3883–3890. <https://doi.org/10.1242/jeb.108456>

CHAPTER 3

A highly contiguous genome assembly for Yellow Warbler (*Setophaga petechia*)



Genome Resources

A highly contiguous genome assembly for the Yellow Warbler (*Setophaga petechia*)

Whitney L.E. Tsai^{1,2,*}, Merly Escalona³, Kimball L. Garrett⁴, Ryan S. Terrill², Ruta Sahasrabudhe⁵, Oanh Nguyen⁵, Eric Beraut⁶, William Seligmann⁶, Colin W. Fairbairn⁶, Ryan J. Harrigan¹, John E. McCormack², Michael E. Alfaro¹, Thomas B. Smith¹ and Rachael A. Bay⁷

¹Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA 90095, United States,

²Moore Laboratory of Zoology, Biology Department, Occidental College, Los Angeles, CA 90041, United States,

³Department of Biomolecular Engineering, University of California, Santa Cruz, CA 95064, United States,

⁴Ornithology Department, Natural History Museum of Los Angeles County, Los Angeles, CA 90007, United States,

⁵DNA Technologies and Expression Analysis Core Laboratory, Genome Center, University of California, Davis, CA 95616, United States,

⁶Department of Ecology and Evolutionary Biology, University of California, Santa Cruz, CA 95064, United States,

⁷Department of Evolution and Ecology, University of California, Davis, CA 95616, United States

*Address correspondence to W. L. E. Tsai at the address above, or email: whitney.le.tsai@gmail.com

Corresponding Editor: Alexander Suh

Abstract

The Yellow Warbler (*Setophaga petechia*) is a small songbird in the wood-warbler family (Parulidae) that exhibits phenotypic and ecological differences across a widespread distribution and is important to California's riparian habitat conservation. Here, we present a high-quality de novo genome assembly of a vouchered female Yellow Warbler from southern California. Using HiFi long-read and Omni-C proximity sequencing technologies, we generated a 1.22 Gb assembly including 687 scaffolds with a contig N50 of 6.80 Mb, scaffold N50 of 21.18 Mb, and a BUSCO completeness score of 96.0%. This highly contiguous genome assembly provides an essential resource for understanding the history of gene flow, divergence, and local adaptation in Yellow Warblers and can inform conservation management of this charismatic bird species.

Key words: California Conservation Genomics Project, Parulidae

Introduction

The Yellow Warbler (*Setophaga petechia*) is a widespread songbird species distributed from Alaska to northern South America (Fig. 1). The species complex comprises up to 43 subspecies in four distinct subspecies groups that display notable diversity in phenotype and ecology across their range (Browning 1994; Klein and Brown 1994; Wilson and Holberton 2004; Salgado-Ortiz et al. 2008). This phenotypic diversity and the presence of both migratory and resident populations have encouraged investigation into the history of adaptation, divergence, and gene flow in this species (Gibbs et al. 2000; Milot et al. 2000; Chaves et al. 2012; Chavarria-Pizarro et al. 2019; Machkour-M'Rabet et al. 2023). Additionally, as a widespread migratory bird species, the Yellow Warbler inhabits variable environmental conditions across its range, allowing for the investigation into the influence of climate on geographic variation and genomic capacity to adapt to climate change (Bay et al. 2018; Chen et al. 2022; DeSaix et al. 2022).

In California, Yellow Warblers are listed as a Species of Special Concern (Shuford et al. 2008) and have experienced notable

declines over the last 50 years (Sauer et al. 2014). Previous genomic work indicates that the inability to adapt to climate change may play a role in population declines in California (Bay et al. 2018). California wetlands and riparian corridors are crucial stopover and breeding habitats for Yellow Warblers and other species of migratory birds. In the last century, 90% to 95% of historic wetland and riparian habitats have been lost, and those that remain are threatened by development and climate change (Dahl 1990; Krueper 1996; Poff et al. 2012). As indicators of healthy riparian habitat, understanding how California Yellow Warbler populations adapt to dramatic changes in their environment will inform conservation action and help mitigate habitat loss in other vulnerable and threatened riparian species, like the California Red-legged Frog (*Rana draytonii*), the Riparian Brush Rabbit (*Sylvilagus bachmani riparius*), and the Valley Elderberry Longhorn Beetle (*Desmocerus californicus dimorphus*) (Collinge et al. 2001; Davidson et al. 2001; Heath and Ballard 2003; Phillips et al. 2005).

The evolutionary and conservation genomics studies needed to address these questions increasingly rely on low-coverage,

Received December 22, 2023; Accepted February 16, 2024

© The American Genetic Association. 2024.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

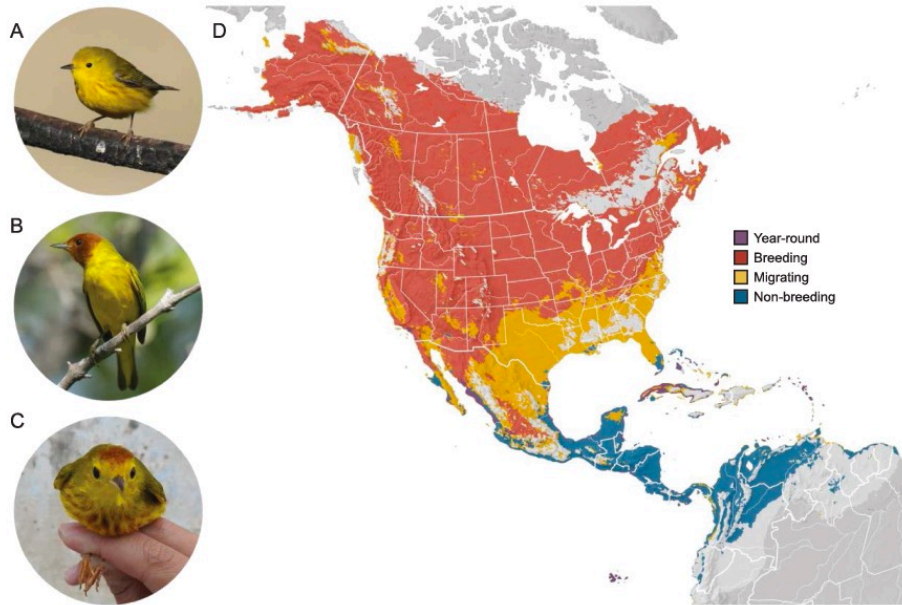


Fig. 1. Geographic variation and distribution of Yellow Warblers (*Setophaga petechia*). A) The Northern (*aestiva*) group includes migratory subspecies with chestnut streaking on the breast. Northern subspecies breed in North America and winter in Central and northern South America. Photo taken by R. S. Terrill at Piute Ponds, Los Angeles, CA, USA. B) The Mangrove (*erithachoides*) group includes resident subspecies with a characteristic chestnut head. Mangrove subspecies inhabit mangroves along the coasts of Central and northern South America year-round. Photo taken by R. S. Terrill on Isla Holbox, Quintana Roo, MX. C) The Galapagos (*aureola*) and Golden (*petechia*) subspecies groups include resident subspecies with a chestnut cap and thick breast streaking except for *S. p. ruficapilla* from Martinique which exhibits the Mangrove phenotype. Populations of the Galapagos subspecies are found on the Galapagos Islands and Cocos Island off Costa Rica and Golden subspecies are found on the islands of the Caribbean. Photo taken by W. L. E. Tsai on Isla Cozumel, Quintana Roo, MX. D) Map of species distributional abundance (Fink et al. 2022). Shaded colors indicate seasonal shifts in distributions: year-round (purple), breeding (red), migrating (yellow), and non-breeding (blue).

whole genome sequencing (WGS), which requires a high-quality reference genome for alignment. Reference genome assemblies provide a map of the structural features and organization of the genome and the choice of reference genome assembly for WGS studies can impact evolutionary inferences like demographic history and genetic diversity (Gopalakrishnan et al. 2017). Currently, there are four genome assemblies generated with short-read sequencing technology for the genus *Setophaga*. There is one Yellow-rumped Warbler (*S. coronata*) chromosome-level assembly (Toews et al. 2016), two Kirtland's Warbler (*S. kirtlandii*) scaffold-level assemblies (Feng et al. 2020), and the existing draft genome assembly for Yellow Warbler has a length of 1.26 Gb, a total of 18,414 scaffolds, and a scaffold N50 491.7 kb (Bay et al. 2018). The use of an interspecific reference genome assembly can lead to many errors and biases, including lower mapping ability (especially in regions with higher evolutionary rates) and inaccurate gene order (Prasad et al. 2022). The high number and relatively short scaffold length of the existing Yellow Warbler genome assembly could hinder the identification of structural variants often maintained between and within species and are important in adaptive evolution, speciation, and generating morphological diversity (Lamichhaney et al. 2016; Wellenreuther and Bernatchez 2018; Mérot et al. 2020). Additionally, reference genome assemblies generated solely from short-read sequencing technology fail to resolve lengths and placement

of repeat regions, such as transposable elements or telomeres, leading to gaps in avian genome assemblies (Peona et al. 2021). This highlights the need for a high-quality, species-specific reference genome for WGS studies.

Here, we present a new genome assembly for the Yellow Warbler generated as part of the California Conservation Genomics Project (CCGP) consortium (Shaffer et al. 2022). We used high-molecular-weight (HMW) genomic DNA (gDNA) extracted from a vouchered, female bird collected in California and leveraged Pacific Biosciences (PacBio) HiFi long-read and Dovetail Genomics Omni-C proximity sequencing technologies. This produced a high-quality genome assembly that will allow us to better understand evolutionary processes like phenotypic variation and migration and conduct conservation genomics studies to inform conservation initiatives.

Methods

Biological materials

We sampled heart, liver, muscle, and other tissues from a female Yellow Warbler collected using mist nets near Stephen Sorensen Park (34.60549°N, 117.8306°W) in Los Angeles County, California on 25 September 2020. This migrant Yellow Warbler can presumably be assigned to *Setophaga petechia brewsteri* based on collection date and locality (Browning 1994) and was

collected with approval from the following entities: California Department of Fish and Wildlife Scientific Collecting Permit (#SC-000939), US Fish and Wildlife Services Scientific Collecting Permit (MB708062-0), and US Geological Survey Banding Permit (22804-B). Tissue samples were retrieved and flash-frozen in liquid nitrogen, and the first muscle tissues were frozen within 2 min of specimen collection. A voucher specimen and tissue are deposited at the Natural History Museum of Los Angeles (LACM Bird #122168, KLG4550, LAF9440). Additional tissues for this individual are housed in the CCGP tissue repository at the University of California, Los Angeles under identification YEWA_CCGP3.

Nucleic acid extraction, library preparation, and sequencing

We extracted HMW gDNA from 30 mg of flash-frozen heart tissue. We homogenized the tissue by grinding it in a mortar and pestle in liquid nitrogen. We lysed the homogenized tissue at room temperature overnight with 2 ml of lysis buffer containing 100 mM NaCl, 10 mM Tris-HCl pH 8.0, 25 mM EDTA, 0.5% (w/v) SDS, and 100 µg/ml Proteinase K. We treated the lysate with 20 µg/ml RNase at 37 °C for 30 min. We cleaned the lysate with equal volumes of phenol/chloroform using phase lock gels (Quantabio, MA; Cat # 2302830). We precipitated the DNA from the cleaned lysate by adding 0.4× volume of 5M ammonium acetate and 3× volume of ice-cold ethanol. We washed the pellet twice with 70% ethanol and resuspended it in elution buffer (10 mM Tris, pH 8.0). We measured DNA purity using absorbance ratios (260/280 = 1.87 and 260/230 = 2.29) using a NanoDrop ND-1000 spectrophotometer. We quantified DNA yield (30 µg) using a Qubit 2.0 Fluorometer (Thermo Fisher Scientific, MA). We verified HMW gDNA integrity on a Femto pulse system (Agilent Technologies, CA), where 80% of the DNA was found in fragments above 120 kb.

According to the manufacturer's instructions, we constructed the HiFi Single Molecule, Real-Time (SMRT) library using SMRTbell Express Template Prep Kit v2.0 (PacBio, CA; Cat. #100-938-900). We sheared HMW gDNA to a target DNA size distribution between 15 and 20 kb and concentrated it using 0.45× of AMPure PB beads (PacBio; Cat. #100-265-900). We performed the enzymatic incubations as follows: removal of single-strand overhangs at 37 °C for 15 min, DNA damage repair at 37 °C for 30 min, end repair at 20 °C for 10 min, A-tailing at 65 °C for 30 min, ligation of overhang adapter v3 at 20 °C for 60 min, ligase inactivation at 65 °C for 10 min, and nuclease treatment at 37 °C for 1 h. We purified and concentrated the library with 0.45× Ampure PB beads for size selection to collect fragments greater than 7 to 9 kb using the BluePippin/PippinHT system (Sage Science, MA; Cat #BLF7510/HPE7510). The HiFi library averaged 15 to 20 kb. It was sequenced at UC Davis DNA Technologies Core (Davis, CA) using two 8M SMRT cells, Sequel II sequencing chemistry 2.0, and 30-h movies each on a PacBio Sequel II sequencer.

We used the Omni-C™ Kit (Dovetail Genomics, CA) for Omni-C proximity sequencing according to the manufacturer's protocol with slight modifications. First, we ground muscle tissue (Sample YEWA_CCGP3; LACM Bird #122168, KLG4550, LAF9440) with a mortar and pestle while cooled with liquid nitrogen. Subsequently, chromatin was fixed in place in the nucleus. We passed the suspended chromatin solution

through 100 µm and 40 µm cell strainers to remove large debris. We digested fixed chromatin under various conditions of DNase I until a suitable fragment length distribution of DNA molecules was obtained. We repaired chromatin ends, ligated a biotinylated bridge adapter, and performed proximity ligation of adapter-containing ends. After proximity ligation, crosslinks were reversed, and the DNA was purified from proteins. We treated purified DNA to remove biotin that was not internal to ligated fragments. We generated a next-generation sequencing library using an NEB Ultra II DNA Library Prep kit (New England Biolabs, MA) with an Illumina-compatible y-adapter. Then, we captured biotin-containing fragments using streptavidin beads. We split the post-capture product into two replicates before PCR enrichment to preserve library complexity, with each replicate receiving unique dual indices. The library was sequenced at the Vincent J. Coates Genomics Sequencing Lab (Berkeley, CA) on an Illumina NovaSeq 6000 PE150 (Illumina, CA). Based on a genome size of 1.26 Gb (Bay et al. 2018), we targeted 126 million base pair reads (100 million read pairs per Gb genome size).

Nuclear genome assembly

We assembled the Yellow Warbler genome following the CCGP assembly pipeline Version 4.0, as outlined in Table 1, which lists the tools and non-default parameters used in the assembly. The pipeline uses PacBio HiFi reads and Omni-C data to produce high-quality and highly contiguous genome assemblies, minimizing manual curation. We removed remnant adapter sequences from the PacBio HiFi dataset using HiFiAdapterFilt (Sim et al. 2022). Then, we obtained the initial phased diploid assembly using HiFiasm (Cheng et al. 2022) with the filtered PacBio HiFi reads and the Omni-C dataset. We aligned the Omni-C data to both assemblies following the Arima Genomics Mapping Pipeline (https://github.com/ArimaGenomics/mapping_pipeline) and then scaffolded both assemblies with SALSA (Ghurye et al. 2017, 2019).

We generated Omni-C contact maps for both assemblies by aligning the Omni-C data with BWA-MEM (Li 2013), identified ligation junctions, and generated Omni-C pairs using pairtools (Goloborodko et al. 2018). We generated a multi-resolution Omni-C matrix with a cooler (Abdennur and Mirny 2020) and balanced it with hicExplorer (Ramírez et al. 2018). We used HiGlass [Version 2.1.11] (Kerpedjiev et al. 2018) and the PretextSuite (<https://github.com/wtsi-hpag/PretextView>; <https://github.com/wtsi-hpag/PretextMap>; <https://github.com/wtsi-hpag/PretextSnapshot>) to visualize the contact maps and then we checked the contact maps for major misassemblies. In detail, if we identified a strong off-diagonal signal and a lack of signal in the consecutive genomic region in the proximity of a join made by the scaffold, we dissolved it by breaking the scaffolds at the coordinates of the join. After this process, no further manual joins were made. Some remaining gaps (joins generated by the scaffold) were closed using the PacBio HiFi reads and YAGCloser (<https://github.com/merlyescalona/yagcloser>). Finally, we checked for contamination using the BlobToolKit Framework (Challis et al. 2020). Given the similar contiguity metrics and fragmentation of both assemblies, we decided to tag the assemblies as primary and alternate, where primary is the one that overall is more complete, has better BUSCO (Benchmarking Universal Single-Copy Orthologs) scores and better k-mer completeness.

Table 1. Assembly pipeline and software used for assembly of the Yellow Warbler genome.

Purpose	Software ^a	Version
Assembly		
Adapters	HiFiAdapterFilt	Commit 64d1c7b
K-mer counting	Meryl ($k = 21$)	1
Estimation of genome size and heterozygosity	GenomeScope	2
De novo assembly (contiging)	HiFiasm (Hi-C Mode, -primary, output p_ctg.hap1, p_ctg.hap2)	0.16.1-r375
Scaffolding		
Omni-C data alignment	Arima Genomics Mapping Pipeline	Commit 2e74ea4
Omni-C Scaffolding	SALSA (-DNASE, -i 20, -p yes)	2
Gap closing	YAGCloscr (-mins 2 -f 20 -mcc 2 -prt 0.25 -cft 0.2 -pld 0.2)	Commit 0e34c3b
Omni-C contact map generation		
Short-read alignment	BWA-MEM (-SSP)	0.7.17-r1188
SAM/BAM processing	samtools	1.11
SAM/BAM filtering	pairtools	0.3.0
Pairs indexing	pairix	0.3.7
Matrix generation	cooler	0.8.10
Matrix balancing	hicExplorer (hicCorrectmatrix correct --filterThreshold -2 4)	3.6
Contact map visualization	HiGlass	2.1.11
	PretextView	0.1.4
	PretextView	0.1.5
	PretextViewSnapshot	0.0.3
Genome quality assessment		
Basic assembly metrics	QUAST (--est-ref-size)	5.0.2
Assembly completeness	BUSCO (-m geno, -l aves)	5.0.0
	Merqury	2020-01-29
Contamination screening		
Local alignment tool	BLAST+ (-db nt, -outfmt '6 qseqid staxids bitscore std', -max_target_seqs 1, -max_hsps 1, -evalue 1e-25)	2.1
General contamination screening	BlobToolKit (PacBlo HiFi Coverage, NCBI Taxa ID = 123631, BUSCOdB = aves)	2.3.3

Software citations are listed in the text
^aOptions detailed for non-default parameters.

Genome assembly assessment

We generated k-mer counts from the PacBio HiFi reads using meryl (<https://github.com/marbl/meryl>). The k-mer database was then used in GenomeScope2.0 (Ranallo-Benavidez et al. 2020) to estimate genome features, including genome size, heterozygosity, and overall repeat content. To obtain general contiguity metrics, we ran QUAST (Gurevich et al. 2013). To evaluate genome quality and functional completeness, we used BUSCO (Manni et al. 2021) with the Aves ortholog database (aves_odb10) containing 8,338 genes. Base level accuracy (QV) and k-mer completeness were assessed using the previously generated meryl database and merqury (Rhie et al. 2020). We further estimated genome assembly accuracy via BUSCO gene set frameshift analysis using the pipeline described in Korchach et al. (2017). Measurements of the size of the phased blocks are based on the size of the contigs generated by HiFiasm on HiC mode. We follow the quality metric nomenclature established by Rhie et al. (2021), with the genome quality code $x.y.P.Q.C$, where $x = \log_{10}[\text{contig NG50}]$; $y = \log_{10}[\text{scaffold NG50}]$; $P = \log_{10}[\text{phased block NG50}]$; $Q = \text{Phred base accuracy QV (quality value)}$; $C = \% \text{ genome represented by the first "n" scaffolds, following a known karyotype of } 2n = 80 \text{ for Yellow Warbler [Bird Chromosome$

Database, Chromosome number data V3.0/2022—(Hobart 1991; Degrandi et al. 2020)]. Quality metrics for the notation were calculated on the primary assembly (bSetPet1.0.p).

Mitochondrial genome assembly

We assembled the mitochondrial genome of Yellow Warbler from the PacBio HiFi reads using the reference-guided pipeline MitoHiFi (Allio et al. 2020; Uliano-Silva et al. 2021). We used the mitochondrial sequence of Kirtland's Warbler (NCBI:NC_051027.1) as the starting reference sequence. After completion of the nuclear genome, we searched for matches of the resulting mitochondrial assembly sequence in the nuclear genome assembly using BLAST+ (Camacho et al. 2009) and filtered out contigs and scaffolds from the nuclear genome with a percentage of sequence identity >99% and size smaller than the mitochondrial assembly sequence.

Results

Sequencing data

The Omni-C and PacBio HiFi sequencing libraries generated 85.3 million read pairs and 2.7 million reads, respectively.

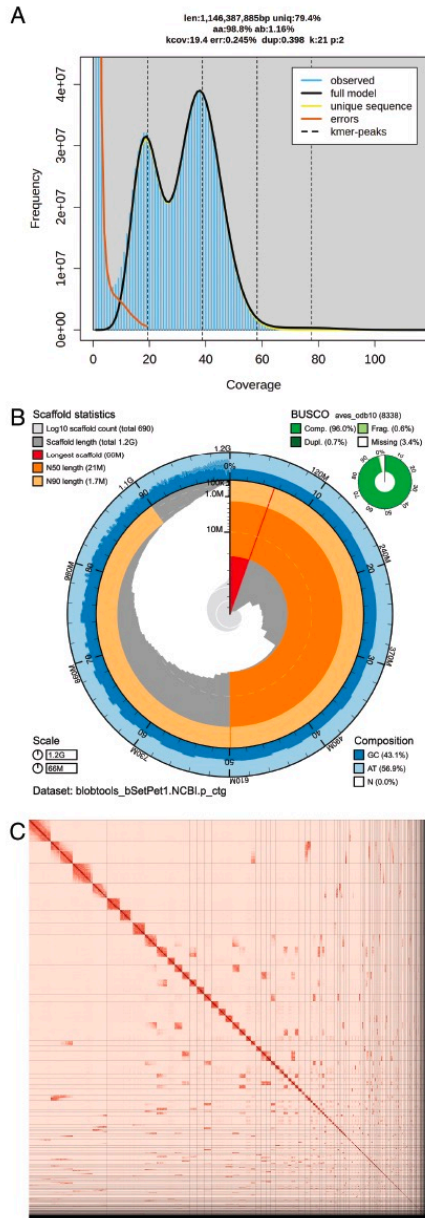


Fig. 2. Visual overview of genome assembly metrics. A) K-mer spectra output generated from PacBio HiFi data without adapters using GenomScope2.0. The bimodal pattern observed corresponds to a diploid genome. K-mers covered at lower coverage and lower frequency correspond to differences between haplotypes, whereas the higher coverage and higher frequency k-mers correspond to the similarities between haplotypes. B) BlobToolKit Snail plot showing a graphical representation of the quality metrics presented in Table 2 for the *Setophaga petechia* primary assembly (bSetPet1.0.p). The plot circle represents the full size of the assembly. From the inside-out, the

central plot covers scaffold and length-related metrics. The central light gray spiral shows the cumulative scaffold count with a white line at each order of magnitude. The red line represents the size of the longest scaffold; all other scaffolds are arranged in size-order moving clockwise around the plot and drawn in gray starting from the outside of the central plot. Dark and light orange arcs show the scaffold N50 and scaffold N90 values. The outer light and dark blue ring show the mean, maximum, and minimum GC vs. AT content at 0.1% intervals (Challis et al. 2020). C) Omni-C contact map for the primary genome assembly generated with PretextSnapshot. Omni-C contact maps translate proximity of genomic regions in 3D space to contiguous linear organization. Each cell in the contact map corresponds to sequencing data supporting the linkage (or join) between two such regions. Scaffolds are separated by black lines and higher density corresponds to higher levels of fragmentation.

The latter yielded 40.87-fold coverage (N50 read length 17,523 bp; minimum read length 41 bp; mean read length 17,110 bp; maximum read length of 54,497 bp). Based on PacBio HiFi reads, we estimated a genome assembly size of 1.14 Gb, 79.39% sequence uniqueness (20.61% repeat content), 0.245% sequencing error rate, and 1.16% nucleotide heterozygosity rate using Genomescope2.0. The k-mer spectrum based on PacBio HiFi reads shows (Fig. 2A) a bimodal distribution with two major peaks at 19- and 39-fold coverage, where peaks correspond to heterozygous and homozygous states of a diploid species.

Nuclear genome assembly

The final assembly consists of two haplotypes tagged as primary and alternate (bSetPet1.0.p and bSetPet1.0.a). Both genome assembly sizes are similar but not equal to the estimated value from Genomescope2.0 (Fig. 2A). The primary assembly (bSetPet1.0.p) consists of 687 scaffolds spanning 1.22 Gb with contig N50 of 6.8 Mb, scaffold N50 of 21.18 Mb, longest contig of 53.52 Mb, and largest scaffold of 66.28 Mb. The alternate assembly (bSetPet1.0.a) consists of 530 scaffolds, spanning 1.24 Gb with contig N50 of 8.3Mb, scaffold N50 of 21.18 Mb, largest contig 40.02 Mb and largest scaffold of 74.56 Mb. The Omni-C contact maps suggest highly contiguous primary and alternate assemblies (Fig. 2C and Supplementary Fig. S1B). The primary assembly has a BUSCO completeness score of 96.0% using the Aves gene set, a per-base quality (QV) of 62.34, a k-mer completeness of 84.95, and a frameshift indel QV of 41.54. In comparison, the alternate assembly has a BUSCO completeness score of 93.5% using the same gene set, a per-base quality (QV) of 62.79, a k-mer completeness of 81.57, and a frameshift indel QV of 40.43.

During manual curation, we identified 13 misassemblies requiring breaking nine joins on the primary assembly and four on the alternate assembly. We were able to close a total of five gaps, three on the primary and two on the alternate assembly. We removed two contigs, one per assembly, corresponding to mitochondrial contaminants. Detailed assembly statistics are reported in Table 2, and a graphical representation of the primary assembly in Fig. 2B (see Supplementary Fig. S1A for the alternate assembly). We have deposited both assemblies on NCBI (see Table 2 and Data Availability for details).

Mitochondrial genome assembly

We assembled a mitochondrial genome with MitoHiFi. The final mitochondrial assembly has a size of 16,809 bp. The base composition of the final assembly version is A = 30.19%,

Table 2. Sequencing and assembly statistics and accession information for the primary and alternate assemblies of the Yellow Warbler (*Setophaga petechia*) genome.

Bio Projects and Vouchers	CCGP NCBI BioProject		PRJNA720569			
	Genera NCBI BioProject		PRJNA765861			
	Species NCBI BioProject		PRJNA777222			
	NCBI BioSample		SAMN29044059, SAMN29044060			
	Specimen identification		LACM:Birds122168			
	NCBI Genome accessions		Primary Alternate			
	Assembly accession	JANCRA000000000	JANCRB000000000			
	Genome sequences	GCA_024362935.1	GCA_024372515.1			
Genome sequence	PacBio HiFi reads	Run	1 PACBIO_SMRT (Sequel II) run: 2.7M spots, 46.9G bases, 35.6Gb downloads			
		Accession	SRX16742538			
	Omni-C Illumina reads	Run	1 ILLUMINA (Illumina NovaSeq 6000) run: 85.3M spots, 25.8G bases, 8.6Gb			
		Accession	SRX16742539, SRX16742540			
Genome Assembly Quality Metrics	Assembly identifier (Quality code ^a)	bSetPet1(6.7.P6.Q62.C)				
	HiFi Read coverage ^b	40.87X				
		Primary	Alternate			
	Number of contigs	971	776			
	Contig N50 (bp)	68,07,045	83,68,636			
	Contig NG50 ^b	72,19,428	89,24,963			
	Longest Contigs	5,35,26,829	4,00,27,624			
	Number of scaffolds	687	530			
	Scaffold N50	2,11,88,473	2,11,88,473			
	Scaffold NG50 ^b	2,17,69,140	2,04,09,353			
	Largest scaffold	6,62,88,485	7,45,62,066			
	Size of final assembly	1,22,23,85,128	1,24,97,65,916			
	Phased block NG50 ^b	73,91,252	93,25,426			
	Gaps per Gbp (# Gaps)	232(284)	197(246)			
	Indel QV (Frame shift)	41.54557	40.4344848			
	Base pair QV	62.3497	62.7988			
			Full assembly = 62.5709			
	k-mer completeness	84.9555	81.57			
			Full assembly = 99.2811			
	BUSCO completeness (aves) <i>n</i> =	C ^c	S ^c	D ^c	F ^c	M ^c
P ^d		96.00%	95.30%	0.70%	0.60%	3.40%
A ^d		93.50%	92.50%	1.00%	0.60%	5.90%
Organelles	1 complete mitochondrial sequence		CM044545.1			

^aAssembly quality code *x.y.P.Q.C* derived notation, from Rhie et al. (2021). *x* = log10[contig NG50]; *y* = log10[scaffold NG50]; *P* = log10 [phased block NG50]; *Q* = Phred base accuracy QV (Quality value); *C* = % genome represented by the first "*n*" scaffolds, following a known karyotype for *S. petechia* of *2n* = 80 (Bird Chromosome Database, Chromosome number data V3.0/2022; Hobart 1991; Degrandi et al. 2020). Quality code for all the assembly denoted by primary assembly (bSetPet1.0.p).

^bRead coverage and NGx statistics have been calculated based on the estimated genome size of 1.14 Gb.

^cBUSCO Scores. Complete BUSCOs (C). Complete and single-copy BUSCOs (S). Complete and duplicated BUSCOs (D). Fragmented BUSCOs (F). Missing BUSCOs (M).

^d(P)primary and (A)lternate assembly values.

C = 31.77%, *G* = 14.19%, *T* = 23.85%, and consists of 22 unique transfer RNAs and 13 protein-coding genes.

Discussion

Here, we present a highly contiguous genome assembly for the Yellow Warbler with two pseudo haplotypes. Our genome

assemblies meet thresholds for proposed quality standards for vertebrate and avian genomes (Jarvis 2016; Kapusta and Suh 2017; Rhie et al. 2021). Compared to the existing *Setophaga* genomes, the primary Yellow Warbler genome assembly presented here has the highest BUSCO completeness (96.0% of Aves orthologs present) and the highest contig N50 (6.8 Mb). Although the Yellow-rumped and Kirtland's

Warbler genome assemblies have higher scaffold N50 values, our Yellow Warbler genome assembly has the fewest gaps greater than 5 N's (284 compared to 49K to 67K in other *Setophaga* genome assemblies), which highlights the improvement gained when using long-read sequencing technology in combination with short reads for more contiguous and complete genomes.

The reference genome presented here provides an essential resource for evolutionary research and conservation efforts in California and beyond. Future range-wide genomic analyses will facilitate investigations into the history of gene flow and divergence between the various subspecies groups in this complex (Browning 1994; Chaves et al. 2012; Machkour-M'Rabet et al. 2023). This system-wide genomic context lends itself to investigations into the genetic basis underlying both phenotypic diversity and the evolution of migration (Toews et al. 2016; Franchini et al. 2017; Delmore et al. 2020; Aguilon et al. 2021; Caballero-López et al. 2022).

Future landscape genomic analyses investigating environmental associations with genomic variation could identify loci important for local adaptation in this widespread species (Bay et al. 2018; Forester et al. 2018; Chen et al. 2022). Using this framework with future climate models will allow for predictions of how Yellow Warblers may adapt to future climate change and identify both populations that are likely to persist in and vulnerable to future climate change regimes, which will guide local conservation implementation (Fitzpatrick and Keller 2015; Shaffer et al. 2022). This will be especially important for California populations experiencing population declines and dwindling breeding habitat, which could benefit from direct conservation and management efforts (Heath and Ballard 2003; Shuford et al. 2008). Overall, the Yellow Warbler genome presented here provides a key resource for investigating phenotypic and ecological evolution and conservation in this charismatic migratory bird species.

Supplementary material

Supplementary material is available at *Journal of Heredity* Journal online.

Acknowledgments

PacBio Sequel II library prep and sequencing was carried out at the DNA Technologies and Expression Analysis Cores at the UC Davis Genome Center, supported by NIH Shared Instrumentation Grant 1S10OD010786-01. Deep sequencing of Omni-C libraries used the Novaseq S4 sequencing platforms at the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by NIH S10 OD018174 Instrumentation Grant. We thank the staff at the UC Davis DNA Technologies and Expression Analysis Cores and the UC Santa Cruz Paleogenomics Laboratory for their diligence and dedication to generating high-quality sequence data. We thank Maeve Secor for help with fieldwork; Tara Luckau, Dr. Courtney Miller, and Dr. Erin Toffelmier for help with coordination and sample submission.

Funding

This work was supported by the California Conservation Genomics Project, with funding provided to the University of California by the State of California, State Budget Act of

2019 [UC Award ID RSI-19-690224]. WLET was supported by the University of California, Los Angeles, Department of Ecology and Evolutionary Biology, Lida Scott Brown Fellowship; and the National Science Foundation, Graduate Research Fellowship [DGE-2034835]. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Data availability

Data generated for this study are available under NCBI BioProject PRJNA777222. Raw sequencing data for individual with voucher LACM:122168 (NCBI BioSamples SAMN29044059, SAMN29044060) are deposited in the NCBI Short Read Archive (SRA) under SRX16742538 for PacBio HiFi sequencing data, and SRX16742539 and SRX16742540 for the Omni-C Illumina sequencing data. GenBank accessions for both primary and alternate assemblies are GCA_024362935.1 and GCA_024372515.1; and for genome sequences JANCRA000000000 and JANCRB000000000. The GenBank organelle genome assembly for the mitochondrial genome is CM044545.1. Assembly scripts and other data for the analyses presented can be found at the following GitHub repository: www.github.com/ccgproject/ccg_assembly.

References

- Abdennur N, Mirny LA. Cooler: scalable storage for Hi-C data and other genomically labeled arrays. *Bioinformatics*. 2020;36:311–316. doi:10.1093/bioinformatics/btz540
- Aguillon SM, Walsh J, Lovette IJ. Extensive hybridization reveals multiple coloration genes underlying a complex plumage phenotype. *Proc Biol Sci*. 2021;288:20201805. doi:10.1098/rspb.2020.1805
- Allio R, Schomaker-Bastos A, Romiguier J, Prosdociimi F, Nabholz B, Delsuc F. MitoFinder: efficient automated large-scale extraction of mitochondrial data in target enrichment phylogenomics. *Mol Ecol Resour*. 2020;20:892–905. doi:10.1111/1755-0998.13160
- Bay RA, Harrigan RJ, Underwood VL, Gibbs HL, Smith TB, Ruegg K. Genomic signals of selection predict climate-driven population declines in a migratory bird. *Science*. 2018;359:83–86. doi:10.1126/science.aan4380
- Browning MR. A taxonomic review of *Dendroica petechia* (Yellow Warbler) (Aves: Parulinae). *Proc Biol Soc Wash*. 1994;107:27–51.
- Caballero-López V, Lundberg M, Sokolovskis K, Bensch S. Transposable elements mark a repeat-rich region associated with migratory phenotypes of willow warblers (*Phylloscopus trochilus*). *Mol Ecol*. 2022;31:1128–1141. doi:10.1111/mec.16292
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. *BMC Bioinf*. 2009;10:421. doi:10.1186/1471-2105-10-421
- Challis R, Richards E, Rajan J, Cochrane G, Blaxter M. BlobToolKit—interactive quality assessment of genome assemblies. *G3 Genes Genomes Genetics*. 2020;10:1361–1374. doi:10.1534/g3.119.400908
- Chavarría-Pizarro T, Gómez JP, Ungvári-Martin J, Bay R, Miyamoto MM, Kimball R. Strong phenotypic divergence in spite of low genetic structure in the endemic Mangrove Warbler subspecies (*Setophaga petechia xanthotera*) of Costa Rica. *Ecol Evol*. 2019;9:13902–13918. doi:10.1002/ecc3.5826
- Chaves JA, Parker PG, Smith TB. Origin and population history of a recent colonizer, the yellow warbler in Galápagos and Cocos Islands. *J Evol Biol*. 2012;25:509–521. doi:10.1111/j.1420-9101.2011.02447.x
- Chen Y, Jiang Z, Fan P, Ericson PGP, Song G, Luo X, Lei F, Qu Y. The combination of genomic offset and niche modelling provides insights into climate change-driven vulnerability. *Nat Commun*. 2022;13:1. doi:10.1038/s41467-022-32546-z

- Cheng H, Jarvis ED, Fedrigo O, Koepfli K-P, Urban L, Gemmill NJ, Li H. Haplotype-resolved assembly of diploid genomes without parental data. *Nat Biotechnol*. 2022;40:1332–1335. doi:10.1038/s41587-022-01261-x
- Collinge SK, Holyoak M, Barr CB, Marty JT. Riparian habitat fragmentation and population persistence of the threatened valley elderberry longhorn beetle in central California. *Biol Conserv*. 2001;100:103–113. doi:10.1016/S0006-3207(00)00211-1
- Dahl, T. E. (1990). Wetlands losses in the United States, 1780's to 1980's. Report to the Congress (PB-91-169284/XAB). St. Petersburg, FL (USA): National Wetlands Inventory. [accessed 2020 Feb 21]. <https://www.osti.gov/biblio/5527872-wetlands-losses-united-states-report-congress>
- Davidson C, Bradley Shaffer H, Jennings MR. Declines of the California Red-Legged Frog: climate, Uv-B, habitat, and pesticides hypotheses. *Ecol Appl*. 2001;11:464–479. doi:10.1890/1051-0761(2001)011[0464:DOTCRL]2.0.CO;2
- Degrandi TM, Barcellos SA, Costa AL, Garnero ADV, Hass I, Gunski RJ. Introducing the bird chromosome database: an overview of cytogenetic studies in birds. *Cytogenet Genome Res*. 2020;160:199–205. doi:10.1159/000507768
- Delmore K, Illera JC, Pérez-Tris J, Segelbacher G, Lugo Ramos JS, Durieux G, Ishigohoka J, Liedvogel M. The evolutionary history and genomics of European blackcap migration. *eLife*. 2020;9:e54462. doi:10.7554/eLife.54462
- DeSaix MG, George TL, Seglund AE, Spellman GM, Zavaleta ES, Ruegg KC. Forecasting climate change response in an alpine specialist songbird reveals the importance of considering novel climate. *Divers Distrib*. 2022;28:2239–2254. doi:10.1111/ddi.13628
- Feng S, Stiller J, Deng Y, Armstrong J, Fang Q, Reeve AH, Xie D, Chen G, Guo C, Faircloth BC, et al. Dense sampling of bird diversity increases power of comparative genomics. *Nature*. 2020;59:7833. doi:10.1038/s41586-020-2873-9
- Fink D, Auer T, Johnston A, Strimas-Mackey M, Ligocki S, Robinson O, Hochachka W, Jaromczyk L, Rodewald A, Wood C, et al. *eBird Status and Trends, Data Version: 2021*; Released: 2022. Ithaca, NY: Cornell Lab of Ornithology; 2022. doi:10.2173/ebirdst.2021
- Fitzpatrick MC, Keller SR. Ecological genomics meets community-level modelling of biodiversity: mapping the genomic landscape of current and future environmental adaptation. *Ecol Lett*. 2015;18:1–16. doi:10.1111/ele.12376
- Forester BR, Lasky JR, Wagner HH, Urban DL. Comparing methods for detecting multilocus adaptation with multivariate genotype–environment associations. *Mol Ecol*. 2018;27:2215–2233. doi:10.1111/mec.14584
- Franchini P, Irisarri I, Fudickar A, Schmidt A, Meyer A, Wikelski M, Partecke J. Animal tracking meets migration genomics: transcriptomic analysis of a partially migratory bird species. *Mol Ecol*. 2017;26:3204–3216. doi:10.1111/mec.14108
- Ghurye J, Pop M, Koren S, Bickhart D, Chin C-S. Scaffolding of long read assemblies using long range contact information. *BMC Genomics*. 2017;18:527. doi:10.1186/s12864-017-3879-z
- Ghurye J, Rhie A, Walenz BP, Schmitt A, Selvaraj S, Pop M, Phillippy AM, Koren S. Integrating Hi-C links with assembly graphs for chromosome-scale assembly. *PLoS Comput Biol*. 2019;15:e1007273. doi:10.1371/journal.pcbi.1007273
- Gibbs HL, Dawson RJG, Hobson KA. Limited differentiation in microsatellite DNA variation among northern populations of the yellow warbler: evidence for male-biased gene flow? *Mol Ecol*. 2000;9:2137–2147. doi:10.1046/j.1365-294X.2000.01136.x
- Goloborodko A, Abdennur N, Venev S, hbranda, & gfuldenberg. (2018). *mirnylab/paritools: V0.2.0* [Computer software]. Zenodo. doi:10.5281/zenodo.1490831
- Gopalakrishnan S, Samaniego Castruita JA, Sinding M-HS, Kuderna LFK, Räikkönen J, Petersen B, Sicheritz-Ponten T, Larson G, Orlando L, Marques-Bonet T, et al. The wolf reference genome sequence (*Canis lupus lupus*) and its implications for *Canis* spp. population genomics. *BMC Genomics*. 2017;18:495. doi:10.1186/s12864-017-3883-3
- Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics*. 2013;29:1072–1075. doi:10.1093/bioinformatics/btt086
- Heath SK, Ballard G. Patterns of breeding songbird diversity and occurrence in Riparian habitats of the Eastern Sierra Nevada. In *California riparian systems: processes and floodplain management, ecology and restoration. Riparian Habitat and Floodplains Conference Proceedings, Riparian Habitat Joint Venture, Sacramento, CA, 2003*, (pp. 21–34).
- Hobart HH. Comparative karyology in nine-primaried oscines (Aves); 1991. [accessed 2022 Oct 25]. <https://repository.arizona.edu/handle/10150/185492>
- Jarvis ED. Perspectives from the Avian Phylogenomics Project: questions that can be answered with sequencing all genomes of a vertebrate class. *Annu Rev Anim Biosci*. 2016;4:45–59. doi:10.1146/annurev-animal-021815-111216
- Kapusta A, Suh A. Evolution of bird genomes—a transposon's-eye view. *Ann N Y Acad Sci*. 2017;1389:164–185. doi:10.1111/nyas.13295
- Kerpedjiev P, Abdennur N, Lekschas F, McCallum C, Dinkla K, Strobel H, Lubner JM, Ouellette SB, Azhir A, Kumar N, et al. HiGlass: web-based visual exploration and analysis of genome interaction maps. *Genome Biol*. 2018;19:125. doi:10.1186/s13059-018-1486-1
- Klein NK, Brown WM. Intraspecific molecular phylogeny in the yellow warbler (*Dendroica petechia*), and implications for Avian Biogeography in the West Indies. *Evolution*. 1994;48:1914–1932. doi:10.2307/2410517
- Korlach J, Gedman G, Kingan SB, Chin C-S, Howard JT, Audet J-N, Cantin L, Jarvis ED. De novo PacBio long-read and phased avian genome assemblies correct and add to reference genes generated with intermediate and short reads. *GigaScience*. 2017;6:gix085. doi:10.1093/gigascience/gix085
- Krueper DJ. Effects of livestock management on Southwestern riparian ecosystems. In: Shaw DW, Finch DM, editors. *Tech Coords. Desired Future Conditions for Southwestern Riparian Ecosystems: Bringing Interests and Concerns Together*. 1995 Sept. 18–22, 1995; Albuquerque, NM. *General Technical Report RM-GTR-272*. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Forest and Range Experiment Station; 1996. p. 272, 281–301.
- Lamichanay S, Han F, Berglund J, Wang C, Almén MS, Webster MT, Grant BR, Grant PR, Andersson L. A beak size locus in Darwin's finches facilitated character displacement during a drought. *Science*. 2016;352:470–474. doi:10.1126/science.aad8786
- Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*. 2013:1303.3997. <https://doi.org/10.48550/arXiv.1303.3997>.
- Machkour-M'Rabet S, Santamaria-Rivero W, Dzib-Chay A, Cristiani LT, MacKinnon-Haskins B. Multi-character approach reveals a new mangrove population of the Yellow Warbler complex, *Setophaga petechia*, on Cozumel Island, Mexico. *PLoS One*. 2023;18:e0287425. doi:10.1371/journal.pone.0287425
- Manni M, Berkeley MR, Seppy M, Simão FA, Zdobnov EM. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol Biol Evol*. 2021;38:4647–4654. doi:10.1093/molbev/msab199
- Mérot C, Oomen RA, Tigano A, Wellenreuther M. A roadmap for understanding the evolutionary significance of structural genomic variation. *Trends Ecol Evol*. 2020;35:561–572. doi:10.1016/j.tree.2020.03.002
- Milot E, Gibbs HL, Hobson KA. Phylogeography and genetic structure of northern populations of the yellow warbler (*Dendroica petechia*). *Mol Ecol*. 2000;9:667–681. doi:10.1046/j.1365-294x.2000.00897.x
- Peona V, Blom MPK, Xu L, Burri R, Sullivan S, Bunikis I, Liachko I, Haryoko T, Jönsson KA, Zhou Q, et al. Identifying the causes and

- consequences of assembly gaps using a multiplatform genome assembly of a bird-of-paradise. *Mol Ecol Resour.* 2021;21:263–286. doi:10.1111/1755-0998.13252
- Phillips SE, Hamilton LP, Kelly PA. Assessment of habitat conditions for the Riparian Brush Rabbit on the San Joaquin River National Wildlife Refuge, California. *Endangered Species Recovery Program.* 2005.
- Poff B, Koestner KA, Neary DG, Merritt D. Threats to western United States riparian ecosystems: a bibliography (RMRS-GTR-269). U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station; 2012. doi:10.2737/RMRS-GTR-269
- Prasad A, Lorenzen ED, Westbury MV. Evaluating the role of reference-genome phylogenetic distance on evolutionary inference. *Mol Ecol Resour.* 2022;22:45–55. doi:10.1111/1755-0998.13457
- Ramírez F, Bhardwaj V, Arrigoni L, Lam KC, Grüning BA, Villavecchia J, Habermann B, Akhtar A, Manke T. High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nat Commun.* 2018;9:1. doi:10.1038/s41467-017-02525-w
- Ranallo-Benavidez TR, Jaron KS, Schatz MC. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat Commun.* 2020;11:1. doi:10.1038/s41467-020-14998-3
- Rhie A, McCarthy SA, Fedrigo O, Damas J, Formenti G, Koren S, Uliano-Silva M, Chow W, Fungtammasan A, Gedman GL, et al. 2020. Towards complete and error-free genome assemblies of all vertebrate species. *Nature.* 2021;592:737–746.
- Rhie A, McCarthy SA, Fedrigo O, Damas J, Formenti G, Koren S, Uliano-Silva M, Chow W, Fungtammasan A, Kim J, et al. Towards complete and error-free genome assemblies of all vertebrate species. *Nature.* 2021;592:7856. doi:10.1038/s41586-021-03451-0
- Salgado-Ortiz J, Marra PP, Sillett TS, Robertson RJ. Breeding ecology of the Mangrove Warbler (*Dendroica petechia bryanti*) and comparative life history of the yellow warbler subspecies complex. *Auk.* 2008;125:402–410. doi:10.1525/auk.2008.07012
- Sauer JR, Hines JE, Fallon JE, Pardiek KL. The North American Breeding Bird Survey, results and analysis 1966–2012 (Version 02.19). U.S. Geological Survey Patuxent Wildlife Research Center; 2014.
- Shaffer HB, Toffelmier E, Corbett-Detig RB, Escalona M, Erickson B, Fiedler P, Gold M, Harrigan RJ, Hodges S, Luckau TK, et al. Landscape genomics to enable conservation actions: the California conservation genomics project. *J Hered.* 2022;113:577–588. doi:10.1093/jhered/esac020
- Shuford WD, Gardali T, Western Field Ornithologists, California, and Department of Fish and Game. California bird species of special concern: a ranked assessment of species, subspecies, and distinct populations of birds of immediate conservation concern in California. Camarillo, CA, USA: Western Field Ornithologists; California Department of Fish and Game; 2008. <http://books.google.com/books?id=9INFAQAIAAJ>
- Sim SB, Corpuz RL, Simmonds TJ, Geib SM. HiFiAdapterFilter, a memory efficient read processing pipeline, prevents occurrence of adapter sequence in PacBio HiFi reads and their negative impacts on genome assembly. *BMC Genomics.* 2022;23:157. doi:10.1186/s12864-022-08375-1
- Toews DPL, Taylor SA, Vallender R, Brelsford A, Butcher BG, Messer PW, Lovette IJ. Plumage genes and little else distinguish the genomes of hybridizing warblers. *Curr Biol.* 2016;26:2313–2318. doi:10.1016/j.cub.2016.06.034
- Uliano-Silva M, Nunes JGF, Krashennikova K, McCarthy SA. *marcelauliano/MitoHiFi: MitoHiFi_v2.0* [Computer software]. Zenodo; 2021. doi:10.5281/zenodo.5205678
- Wellenreuther M, Bernatchez L. Eco-evolutionary genomics of chromosomal inversions. *Trends Ecol Evol.* 2018;33:427–440. doi:10.1016/j.tree.2018.04.002
- Wilson CM, Holberton RL. Individual risk versus immediate reproductive success: a basis for latitudinal differences in the adrenocortical response to stress in yellow warblers (*Dendroica petechia*). *Auk.* 2004;122:378–1249. doi:10.1093/auk/121.4.1238

Supplementary materials for

A highly contiguous genome assembly for the Yellow Warbler (*Setophaga petechia*)

Whitney L. E. Tsai*, Merly Escalona, Kimball L. Garrett, Ryan S. Terrill, Ruta Sahasrabudhe,
Oanh Nguyen, Eric Beraut, William Seligmann, Colin W. Fairbairn, Ryan J. Harrigan, John E.
McCormack, Michael E. Alfaro, Thomas B. Smith, Rachael A. Bay

**Corresponding author. Email: whitney.le.tsai@gmail.com*

This PDF file includes:

Figure S1

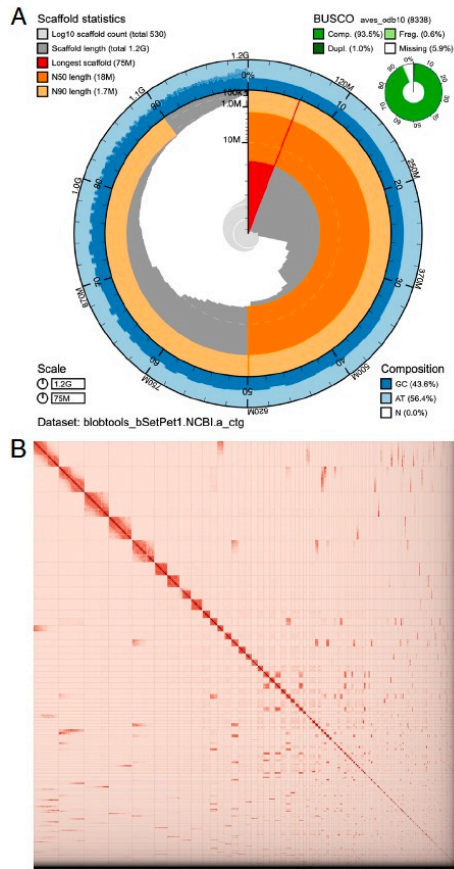


Figure S1. Visual overview of genome assembly metrics for alternate assembly (bSetPet1.0.a). A) BlobToolKit Snail plot showing a graphical representation of the quality metrics presented in Table 2. The plot circle represents the full size of the assembly. From the inside-out, the central plot covers scaffold and length-related metrics. The central light gray spiral shows the cumulative scaffold count with a white line at each order of magnitude. The red line represents the size of the longest scaffold; all other scaffolds are arranged in size-order moving clockwise around the plot and drawn in gray starting from the outside of the central plot. Dark and light orange arcs show the scaffold N50 and scaffold N90 values. The outer light and dark blue ring show the mean, maximum, and minimum GC versus AT content at 0.1% intervals (Challis et al. 2020). B) Omni-C contact map for the alternate genome assembly generated with PretextSnapshot. Omni-C contact maps translate proximity of genomic regions in 3D space to contiguous linear organization. Each cell in the contact map corresponds to sequencing data supporting the linkage (or join) between 2 such regions. Scaffolds are separated by black lines and higher density corresponds to higher levels of fragmentation.

CHAPTER 4

Climate-associated genomic variation in California breeding Yellow Warblers (*Setophaga petechia*)

Abstract

The protection of wildlife populations and species in the face of climate change requires an understanding of their responses to environmental changes to inform conservation decisions. Recent work integrating environmental and genomic data across the breeding range of Yellow Warblers found that the inability to adapt to climate change might already play a role in Yellow Warbler declines in California. However, due to the broad scale of this study, we know less about how climate-associated genetic variation is distributed across smaller geographic areas in California, which is crucial for conserving populations. Here, we performed low-coverage whole genome sequencing of 137 California breeding Yellow Warblers and analyzed their population genetic structure and genotype-environmental associations. We found little evidence for population genetic structure in Yellow Warblers breeding in California. Despite this low genetic structure, we identified 2,972 putatively adaptive SNPs associated with climate with the top environmental variables related to precipitation and vegetation. When we map these associations across California, we identify populations harboring unique genotype-environment associations that can be prioritized for conservation. These findings highlight the importance of understanding both neutral and adaptive genetic variation when considering conservation prioritization.

Introduction

The positive impact of conservation decisions relies on our ability to be informed by predictions of species' responses to environmental change to mitigate the effects of unprecedented climate change affecting the planet's wildlife populations. Drastic variations in temperature and precipitation can create a mismatch between the environment and an organism's climatic

tolerance. In response, species can adapt to new climate regimes, shift their range to more suitable environments (Pinsky et al., 2013), experience population declines, or face extinction (Urban, 2015). This response is determined by species' ability to adapt to rapidly changing environments.

Evolutionary adaptation plays an important role in species persistence and has recently been incorporated into predictive models to understand the genomic potential of species to adapt to changing environments (Capblancq et al., 2020; Fitzpatrick & Keller, 2015). Integrating environmental and genomic data into future species distribution models can approximate the genomic capacity of species to adapt to changing environments. Widely distributed species that experience variable environmental and selective pressures are more likely to harbor the genomic variation needed for climate adaptation (Bay et al., 2018; DeSaix et al., 2022; Razgour et al., 2019). Indeed, these range-wide studies confirm intraspecific genomic variation associated with climate, which may provide the raw material needed for adaptation under climate change regimes. However, the resolution of these studies is typically not at the geographic scale local managers need to make informed conservation decisions.

California wetlands and riparian corridors are crucial stopover and breeding habitat for migratory birds. They also provide important ecosystem services like water quality improvement, flood protection, and groundwater recharge (Naiman et al., 2010). In the last century, 90-95% of historic wetlands and riparian habitats have been lost, and those that remain are threatened by development and climate change (Dahl, 1990; Krueper, 1996; Poff et al., 2012). Statewide, several types of riparian habitats exist with variable environmental conditions, however, we do not know how putatively adaptive genetic variation is distributed across these smaller environmental gradients and riparian fragments, which are particularly crucial for

management. Thus, a fine-scale approach to studying this variation is needed to better inform statewide conservation decisions.

The Yellow Warbler (*Setophaga petechia*) is a widespread migratory songbird species that breeds in the riparian habitats of California. The breeding range of this species extends from Alaska to northern Mexico and the wintering range from Mexico to northern South America. The Yellow Warbler is an ideal species to study climate-associated genetic variation in California riparian ecosystems because it is a focal riparian species, obligate riparian breeder, and because conservation management for this species will benefit and protect other riparian species and habitats (Dybala et al., 2017; Heath & Ballard, 2003; Shuford et al., 2008). Although many populations of Yellow Warblers throughout North America are stable, populations in California have experienced declines over the last half century, resulting in their listing as a Species of Special Concern in California (Shuford et al., 2008). Previous work has shown that the inability to adapt to climate change might play a role in these California declines (Bay et al., 2018). While range-wide climate-associated genetic variation has been identified, it is unclear how much putatively adaptive variation exists in California and how it is distributed across the landscape. As California is the region where Yellow Warblers are experiencing the most severe declines, understanding the distribution of climate-associated genetic variation across the state is crucial for conserving these populations. Here, we use whole genome sequencing of Yellow Warblers breeding in California to (1) investigate population genetic structure in California, (2) identify environmental variables associated with genetic variation, and (3) map adaptive environments in Yellow Warblers across California.

Methods

Sampling and low-coverage whole genome sequencing

We collected new sequence data for 31 blood, 106 feather, and eight tissue samples for 145 Yellow Warblers sampled across California (Figure 1, Table 1). Samples were collected during the breeding season (May-August) between 1996 and 2020. We extracted DNA from all samples using the Qiagen DNeasy Blood & Tissue Kit. For feather samples, we cut the tip of the calamus for extraction. We modified the protocol for maximum yield by adding DTT, incubating pre-warmed Buffer AE on the column filter prior to elution, and increasing elution volume. For blood and tissue samples, we followed the manufacturer's protocol. We prepared whole genome libraries using a modified Illumina Nextera XT protocol optimized for low-coverage whole genome sequencing from low-concentration samples (Schweizer et al., 2021). We performed a Dual-SPRI bead cleanup on pools of 5-8 samples to restrict the fragment distribution of libraries to 300-600bp. We quantified samples using a Qubit Fluorometer and Bioanalyzer and pooled all samples at equimolar ratios in a single pool. We sequenced the pooled libraries on three lanes of an Illumina NovaSeq 6000 PE150.

Mapping and variant calling and filtering

We performed read mapping and called single-nucleotide polymorphisms (SNPs) using the snpArcher workflow (Mirchandani et al., 2024). Briefly, we trimmed adapter sequences using *fastp* (Chen et al., 2018) and aligned sequencing reads to a Yellow Warbler reference genome (Tsai et al., 2024) using *bwa mem* (Li & Durbin, 2009). We called individual variants and performed joint genotyping to produce a multi-sample VCF (variant call format) file using Sention Haplotyper and Genotyper (Kendig et al., 2019). We filtered out indels and non-biallelic SNPs and discarded low-quality variants using *vcftools* with the following filters: genotype quality<30, depth<5, and missing data <10%. We further filtered our SNP dataset using a custom

filtering and visualization pipeline using the R packages *SNPfiltR* (DeRaad, 2022) and *vcfR* (Knaus & Grünwald, 2017).

Population genetic structure

We pruned SNPs for linkage disequilibrium (LD; $ld.threshold = 0.2$) and used the R package *adegenet* (Jombart, 2008) to run principal components analysis (PCA) and ADMIXTURE (Alexander et al., 2009) to estimate ancestry and characterize potential population genetic structure. We executed ADMIXTURE for k-values 1-10 and used the cross-validation procedure to determine the best k. For visualization of genetic analyses and running environmental analyses, we grouped individuals into 10 populations based on California ecoregions (Cleland et al., 2007).

Environmental variables

We obtained environmental data from publicly available databases for each sampling location (Table 2). We included a total of 33 environmental variables, including climate variables from WorldClim (Hijmans et al., 2005), physical variables related to vegetation, land cover, and geography from the Global Land Cover Facility (Sexton et al., 2013; USGS, 2018), and variables related to human use or influence (de Sherbinin et al., 2002; Elvidge et al., 2017).

Genotype-environment associations

We filtered our SNP dataset to retain only SNPs with minor allele frequency $>10\%$ because rare alleles are more likely to result in false positives. We ran 100 bootstrapped trees on our filtered SNPs across our 10 populations in California using the R package *randomForest* (Breiman, 2001). Then, we tested which environmental variables best explained genetic variation in our dataset using a gradient forest analysis run in the R package *gradientForest* (Ellis et al., 2012). To visualize the gradient forest model for Yellow Warblers breeding in California, we generated

100,000 random points across the state and extracted BIOCLIM values for each point. We used the gradient forest model to transform environmental variables into genetic importance values and used a PCA to summarize these values. To visualize the different adaptive environments across California, we assigned colors based on the top three principal components, where similar colors correspond to similarity in genomic association to environmental variables.

Results

Sequencing, variant calling, and variant filtering

Our NovaSeq lanes output 8.5 billion read pairs with an average of 58.5 million reads per sample and an average depth of coverage of 7x. We identified 104,929,592 SNPs across the genome. Discarding low-quality SNPs and low-coverage individuals resulted in a final dataset of 473,912 SNPs and 137 individuals across our 10 populations.

Population genetic structure

Following LD pruning, we were left with a dataset of 44,601 SNPs for running our population genetic analyses. The PCA revealed little to no clustering structure based on population or collecting locality (Figure 2A). Along the PC1 axis, some samples from the Central California Coast, Mojave Desert, and Southern California Coast are differentiated. Two samples were separated from all other samples on PC2 and were located about 100 miles apart in adjacent ecoregions: Klamath Mountains and Southern Cascades. However, these patterns only hold for some individuals from each ecoregion and are not representative of all samples from each population. The cross-validation procedure for ADMIXTURE indicated $k=1$ as the best k -value for our dataset. Additionally, plotting genetic clusters for $k=2$ across our 10 populations in California suggests no evidence for population genetic structure (Figure 2B, 2C).

Genotype-environment associations

After filtering SNPs, we ran our gradient forest analysis for 36,748 SNPs across 10 populations in California. The gradient forest analysis revealed 2,972 SNPs associated with our 33 environmental variables and provided a ranked list of all variables based on the relative predictive power (Figure 3A). The top eight environmental variables related to precipitation and physical land properties like vegetation and latitude: (1) precipitation of the driest quarter, (2) canopy tree cover, (3) normalized difference in vegetation index (NDVI), (4) annual precipitation, (5) latitude, (6) precipitation of the coldest quarter, (7) precipitation seasonality, and (8) mean temperature of the wettest quarter.

After removing points with no associated environmental variables (e.g. points in bodies of water or too close to the coastline), we visualized environmentally associated genetic variation across 96,280 random points in California (Figure 3B, 3C). We found strong differences in genomic variation associated with climate across different habitat types in California. This includes differences between montane regions, the Central Valley, and the southeastern deserts. Additionally, the eastern edge of the Sierra Nevada Mountains may harbor unique genomic variation associated with climate.

Discussion

Using neutral and adaptive SNPs, our population genetic analyses suggest high gene flow among Yellow Warblers breeding in California. As a long-distance migratory bird species, these results suggest that these populations have the ability to change their breeding sites annually. However, previous work has shown Yellow Warblers to have high breeding site fidelity (Knopf & Sedgwick, 1992; Studd & Robertson, 1989). Three subspecies of Yellow Warbler have been shown to occur in California: *S. p. brewsteri* along the coast, *S. p. morcomi* in the interior, and *S. p. sonarana* in the Sonoran Desert (Browning, 1994). However, *S. p. brewsteri* and *S. p.*

morcomi are not consistently distinguishable (Patten et al., 2003), a finding that our genomic results support. *S. p. sonorana* only breeds in California along the lower Colorado River on the California-Arizona border and is of particular conservation concern (Shuford et al., 2008). This historically small breeding range has been significantly reduced, and our recent sampling of the region may not have captured these breeding individuals. Including more extensive sampling and historical samples of this subspecies in future genetic studies could help to determine genetic differentiation in this priority species.

We found evidence of unique genetic variation associated with climate in California breeding Yellow Warblers by investigating a smaller dataset of potentially adaptive SNPs. The top variables associated with genetic variation related to precipitation, vegetation, and latitude (Figure 3A). As in previous findings, our results indicated that adaptation to precipitation may be important for Yellow Warblers (Bay et al., 2018). Additionally, we find that adaptation to vegetation may be specific to Yellow Warblers breeding in California. Riparian habitats are characterized by specific vegetation composition and structure and have particular water requirements (Stromberg & Patten, 1990). As an obligate riparian breeder in California, nest success depends on the availability of riparian-associated species like alder, willow, and cottonwoods (Dybala et al., 2017; Shuford et al., 2008).

We show that analyzing selective SNPs reveals patterns of genetic variation not present when doing traditional population genetic analyses based mostly on neutral genetic diversity. Our results reveal unique genetic variation associated with climate across Yellow Warblers breeding in California (Figure 3B, 3C). While we see unique genotype-environment associations in the Central Valley and Southeastern Desert, the regions exhibit fairly uniform associations represented by similar colors from the PCA. In comparison, regions like the Klamath Mountains,

Southern Cascades, and the eastern edge of the Sierra Nevada Mountains harbor variation in environmental associations, which may provide more tolerance to future climate threats (Bay et al., 2018; DeSaix et al., 2022). Our results also demonstrate that increasing sampling across a finer scale can provide more resolution and uncover unique associations between genomes and environment that were not present when investigating broader scales (Bay et al., 2018; Ruegg et al., 2018). This higher resolution could be important for on-the-ground conservation actions.

Traditional measures to prioritize conservation resources include approaches that are based on current metrics of an area: population abundance or diversity, phylogenetic diversity, species gains or losses, or more recently genetic diversity, variation, or uniqueness. While useful, these measures cannot inform conservation for the future, and do not assess the ability of species to adapt to rapidly changing climate conditions or the ability to survive under future climate change regimes. It has become increasingly clear that genomic analyses are important to wildlife conservation (Cassin-Sackett et al., 2019; Fiedler et al., 2022; Larison et al., 2021; Romanov et al., 2009; Shaffer et al., 2022; Smith et al., 2020). Using genomic resources to understand how species adapt to climate change today can help us predict where they will most likely persist (Capblancq et al., 2020; DeSaix et al., 2022; Fitzpatrick & Keller, 2015). Leveraging these methods and more traditional measures to prioritize species conservation can help protect organisms today and ensure they have the genetic variation necessary to adapt to future climate conditions.

Figures

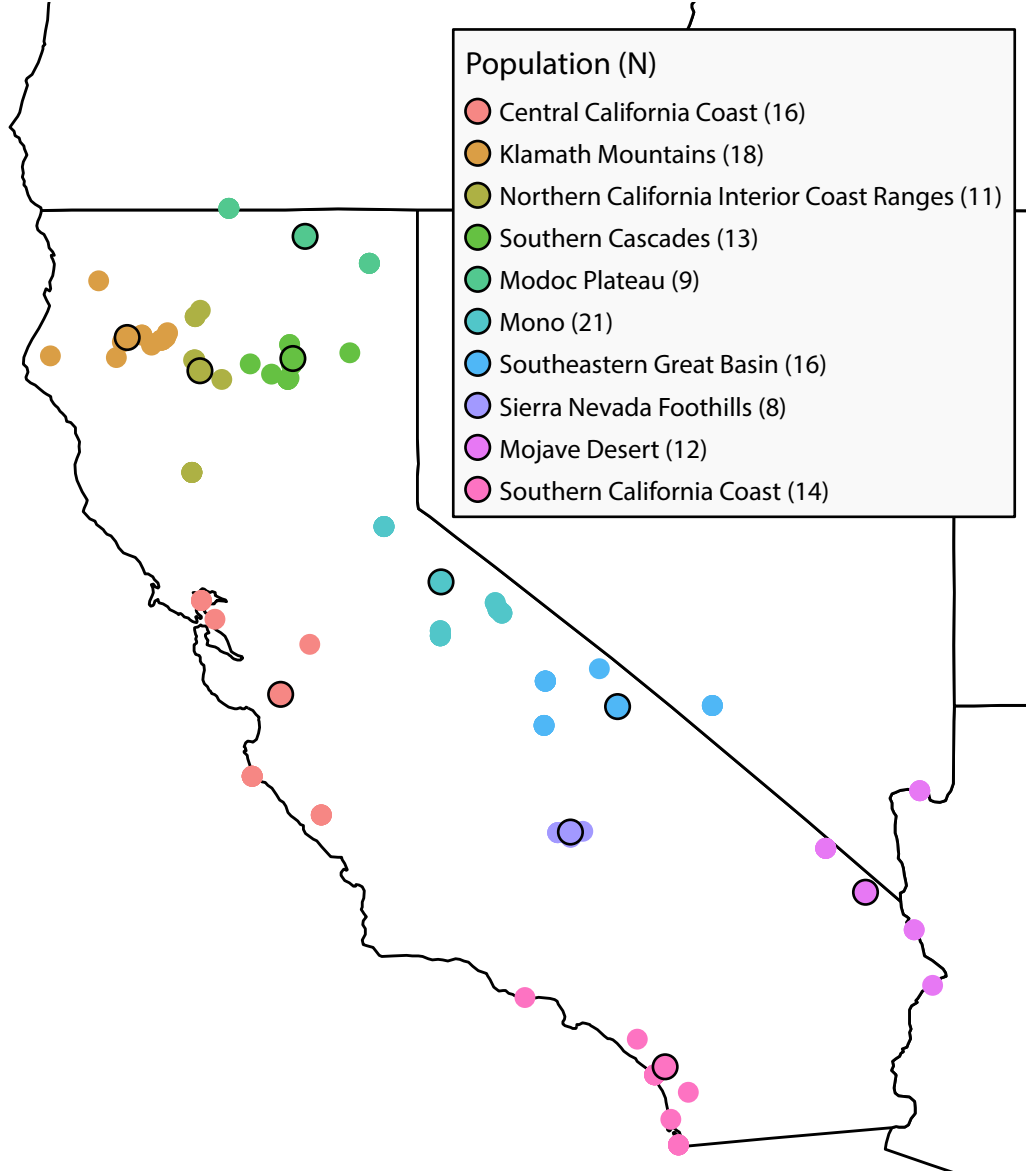


Figure 1. Map of California Yellow Warbler samples ($N = 137$). Details about sample localities and populations are available in Table 1. Dots outlined in black indicate the centroid for each population. For the Mojave Desert population, the Nevada samples were excluded from the centroid calculation to ensure the centroid remained in California.

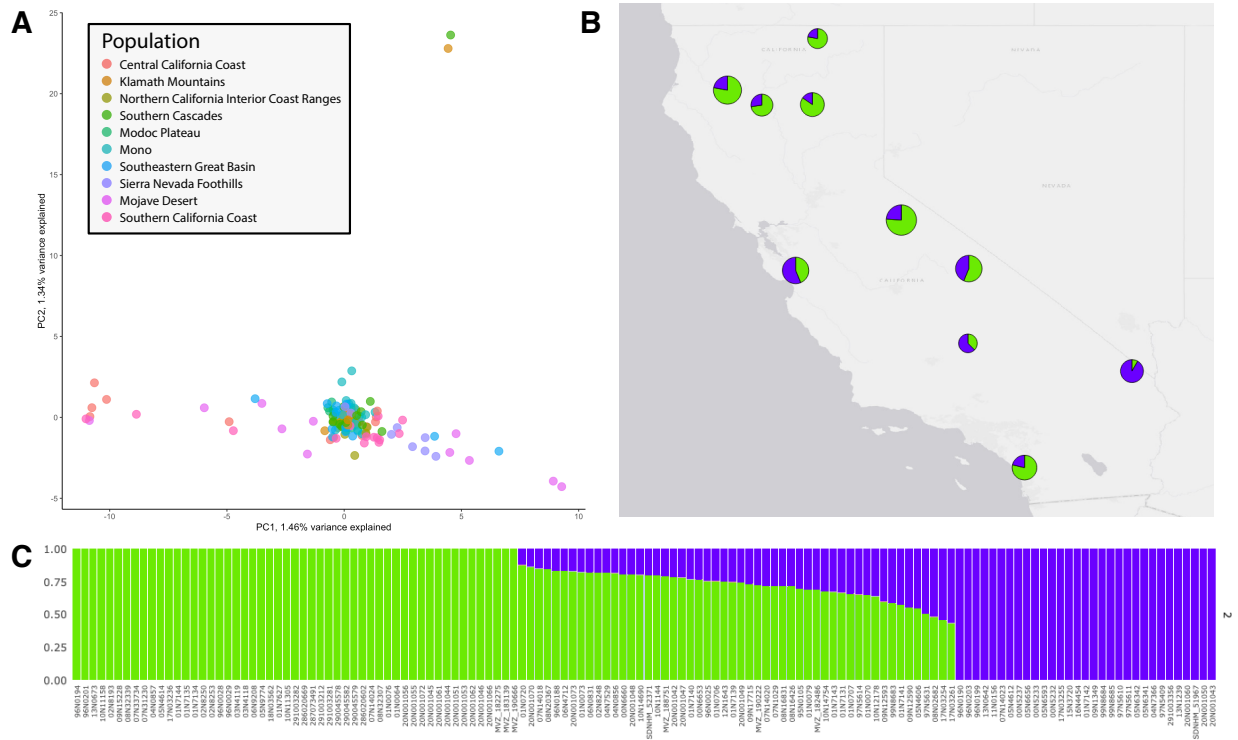


Figure 2. (A) PCA plot of PC1 and PC2 for 137 Yellow Warblers using 44,601 SNPs with loadings marked on axes. (B) Mapping of $k=2$ genetic clusters for ADMIXTURE plot. Pie charts represent the percentage of individuals assigning to each genetic cluster. (C) ADMIXTURE plot for $k=2$.

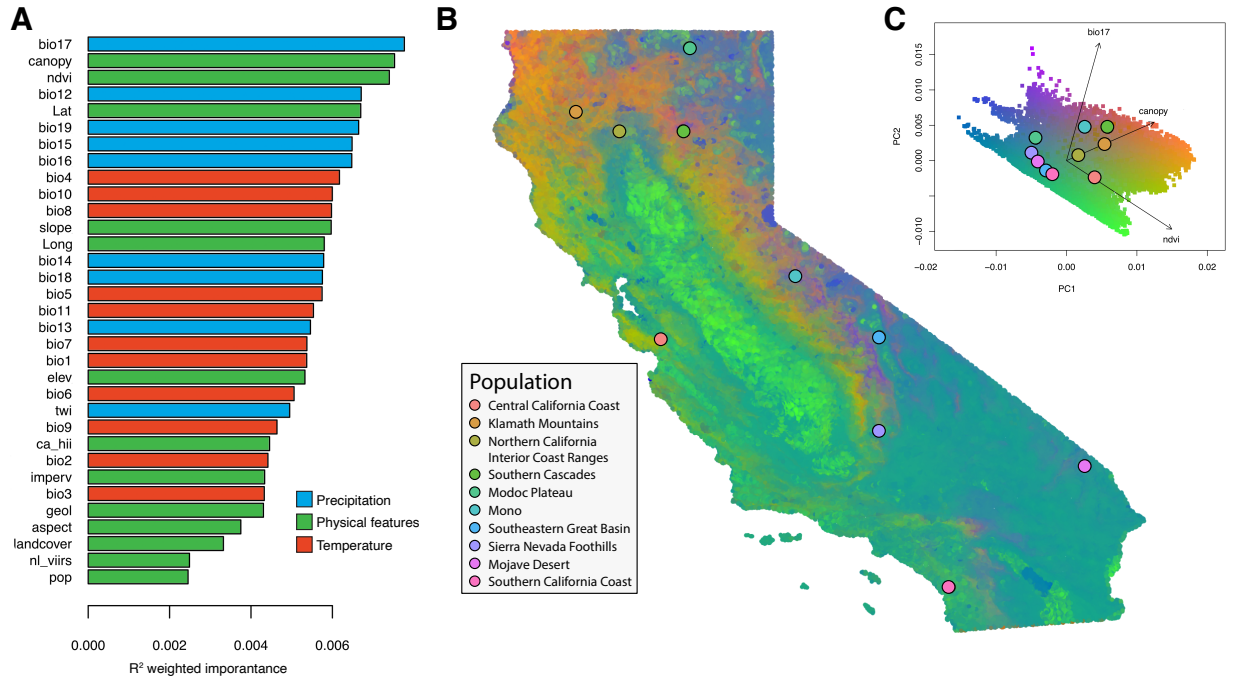


Figure 3. (A) Gradient forest analysis output of ranked importance of environmental variables. (B) Map of gradient forest transformed environmental variables from the PCA (C) representing environmental adaptation in California. Dots represent population centers. (C) PCA of PC1 and PC2 for gradient forest transformed climate variables. Dots represent PC scores associated with populations and colors are associated with genotype-environment correlations for 96,280 random points across California. Arrows show the direction and magnitude of association with adaptation to the top three ranked environmental variables.

Tables

Table 1. Sampling locations of California breeding Yellow Warbler individuals used for low-coverage whole genome sequencing and analyses.

Sample ID	Ecoregion based population	Latitude	Longitude
96N0188	Central California Coast	36.285556	-121.841944
96N0190	Central California Coast	36.285556	-121.841944
96N0194	Central California Coast	36.285556	-121.841944
96N0199	Central California Coast	36.285556	-121.841944
96N0201	Central California Coast	36.285556	-121.841944
96N0203	Central California Coast	36.285556	-121.841944
MVZ_183139	Central California Coast	37.872066	-122.255028
10N11158	Central California Coast	35.894167	-121.073611
11N0156	Central California Coast	35.894167	-121.073611
12N1643	Central California Coast	35.894167	-121.073611
13N0642	Central California Coast	35.894167	-121.073611
13N0673	Central California Coast	35.894167	-121.073611
02N8193	Central California Coast	37.618643	-121.201948
99N8683	Central California Coast	38.06152	-122.405227
99N8684	Central California Coast	38.06152	-122.405227
99N8685	Central California Coast	38.06152	-122.405227
07N14020	Klamath Mountains	40.517222	-123.350556
06N4712	Klamath Mountains	40.643333	-122.959722
07N14023	Klamath Mountains	40.6575	-122.959722
09N15228	Klamath Mountains	40.674722	-123.283889
07N33734	Klamath Mountains	40.693333	-122.855833
08N16831	Klamath Mountains	40.693333	-122.855833
20N001066	Klamath Mountains	40.6956	-122.8404
20N001070	Klamath Mountains	40.6956	-122.8404
20N001071	Klamath Mountains	40.6956	-122.8404
07N14018	Klamath Mountains	40.720833	-122.804444
07N31230	Klamath Mountains	40.720833	-122.804444
08N16426	Klamath Mountains	40.745278	-123.066389
08N32339	Klamath Mountains	40.745278	-123.066389
20N001072	Klamath Mountains	40.7601	-122.786
20N001073	Klamath Mountains	40.7685	-122.7818
95N0105	Klamath Mountains	41.293056	-123.546944
09N1349	Klamath Mountains	40.5342	-124.083
MVZ_190222	Northern California Interior Coast Ranges	40.29614	-122.17913

97N5610	Northern California Interior Coast Ranges	39.356631	-122.511355
97N5611	Northern California Interior Coast Ranges	39.356631	-122.511355
97N5614	Northern California Interior Coast Ranges	39.356631	-122.511355
97N5616	Northern California Interior Coast Ranges	39.356631	-122.511355
97N5631	Northern California Interior Coast Ranges	39.356631	-122.511355
02N8253	Northern California Interior Coast Ranges	40.4915	-122.4928
02N8248	Northern California Interior Coast Ranges	40.493611	-122.478611
02N8250	Northern California Interior Coast Ranges	40.493611	-122.478611
96N0028	Northern California Interior Coast Ranges	40.931	-122.4751
96N0029	Northern California Interior Coast Ranges	40.931	-122.4751
96N0025	Northern California Interior Coast Ranges	40.9954	-122.418
01N0064	Southern Cascades	40.298085	-121.4455
01N0070	Southern Cascades	40.298085	-121.4455
01N0073	Southern Cascades	40.298085	-121.4455
01N0076	Southern Cascades	40.298085	-121.4455
01N0079	Southern Cascades	40.298085	-121.4455
MVZ_188751	Southern Cascades	40.3096634	-121.4323141
MVZ_182486	Southern Cascades	40.3498	-121.6267
01N0706	Southern Cascades	40.444497	-121.4033
01N0707	Southern Cascades	40.444497	-121.4033
MVZ_182275	Southern Cascades	40.565523	-120.756963
07N14024	Southern Cascades	40.64862	-121.425621
08N32307	Southern Cascades	40.64862	-121.425621
MVZ_190666	Southern Cascades	40.45349	-121.86354
04N0856	Modoc Plateau	41.47	-120.54
04N0857	Modoc Plateau	41.47	-120.54
05N4606	Modoc Plateau	41.47	-120.54
05N4612	Modoc Plateau	41.47	-120.54
05N4614	Modoc Plateau	41.47	-120.54
06N0831	Modoc Plateau	42.025556	-122.100278
07N1029	Modoc Plateau	42.025556	-122.100278
08N20367	Modoc Plateau	42.025556	-122.100278
10N1144	Modoc Plateau	42.025556	-122.100278
01N7131	Mono	37.93444444	-119.0677778
01N7134	Mono	37.93444444	-119.0677778
01N7135	Mono	37.93444444	-119.0677778
01N7139	Mono	37.973056	-119.111111
01N7140	Mono	37.973056	-119.111111
01N7141	Mono	37.973056	-119.111111
01N7142	Mono	37.973056	-119.111111

01N7143	Mono	38.04	-119.141944
01N7144	Mono	38.04	-119.141944
00N6660	Mono	37.704444	-119.751944
05N6341	Mono	37.704444	-119.751944
05N6342	Mono	37.704444	-119.751944
10N14754	Mono	37.704444	-119.751944
00N6653	Mono	37.755556	-119.751944
04N7529	Mono	37.755556	-119.751944
09N17715	Mono	37.755556	-119.751944
20N001055	Mono	38.8086	-120.378
20N001056	Mono	38.8086	-120.378
20N001060	Mono	38.8096	-120.3772
20N001061	Mono	38.8096	-120.3772
20N001062	Mono	38.8096	-120.3772
08N02682	Southeastern Great Basin	37	-116.73
09N12590	Southeastern Great Basin	37	-116.73
09N12593	Southeastern Great Basin	37	-116.73
10N12178	Southeastern Great Basin	37	-116.73
20N001049	Southeastern Great Basin	37.246	-118.58607
20N001050	Southeastern Great Basin	37.2462	-118.5855
20N001051	Southeastern Great Basin	37.2467	-118.5844
20N001048	Southeastern Great Basin	37.2468	-118.58617
20N001046	Southeastern Great Basin	37.2478	-118.5866
20N001047	Southeastern Great Basin	37.2478	-118.5866
20N001053	Southeastern Great Basin	37.2488	-118.5826
97N5409	Southeastern Great Basin	37.3723	-117.9852
01N0720	Southeastern Great Basin	36.799167	-118.599444
03N4118	Southeastern Great Basin	36.799167	-118.599444
03N4119	Southeastern Great Basin	36.799167	-118.599444
10N14690	Southeastern Great Basin	36.799167	-118.599444
20N001042	Sierra Nevada Foothills	35.7151	-118.4478
20N001044	Sierra Nevada Foothills	35.7151	-118.4478
20N001045	Sierra Nevada Foothills	35.7151	-118.4478
20N001043	Sierra Nevada Foothills	35.717	-118.434
01N7627	Sierra Nevada Foothills	35.728889	-118.169167
18N03562	Sierra Nevada Foothills	35.668	-118.305
05N9774	Sierra Nevada Foothills	35.673889	-118.299444
06N9208	Sierra Nevada Foothills	35.673889	-118.299444
05N6593	Mojave Desert	34.733056	-114.488333
05N6656	Mojave Desert	34.733056	-114.488333

15N3720	Mojave Desert	35.556584	-115.470638
16N4454	Mojave Desert	35.556584	-115.470638
17N03254	Mojave Desert	35.556584	-115.470638
17N03255	Mojave Desert	35.556584	-115.470638
17N03261	Mojave Desert	35.556584	-115.470638
00N5232	Mojave Desert	36.138854	-114.426125
00N5233	Mojave Desert	36.138854	-114.426125
00N5237	Mojave Desert	36.138854	-114.426125
04N7366	Mojave Desert	34.171944	-114.283333
286020669	Southern California Coast	32.558727	-117.106403
290045578	Southern California Coast	32.558727	-117.106403
290045579	Southern California Coast	32.558727	-117.106403
290045582	Southern California Coast	32.558727	-117.106403
291003281	Southern California Coast	32.558727	-117.106403
SDNHM_52371	Southern California Coast	32.8176194	-117.1913356
SDNHM_51967	Southern California Coast	33.09098	-116.99688
286020602	Southern California Coast	33.267029	-117.37134
287073491	Southern California Coast	33.267029	-117.37134
291003212	Southern California Coast	33.267029	-117.37134
291003282	Southern California Coast	33.267029	-117.37134
291003356	Southern California Coast	33.267029	-117.37134
10N11305	Southern California Coast	33.627778	-117.563889
13N1239	Southern California Coast	34.048611	-118.812222

Table 2. Environmental variables used for gradient forest analyses.

Abbreviation	Environmental variable description
BIO1	Annual Mean Temperature
BIO2	Mean Diurnal Range
BIO3	Isothermality
BIO4	Temperature Seasonality
BIO5	Max Temperature of the Warmest Month
BIO6	Min Temperature of the Warmest Month
BIO7	Temperature Annual Range
BIO8	Mean Temperature of the Wettest Quarter
BIO9	Mean Temperature of the Driest Quarter
BIO10	Mean Temperature of the Warmest Quarter
BIO11	Mean Temperature of the Coldest Quarter
BIO12	Annual Precipitation
BIO13	Precipitation of the Wettest Month
BIO14	Precipitation of the Driest Month
BIO15	Precipitation Seasonality
BIO16	Precipitation of the Wettest Quarter
BIO17	Precipitation of the Driest Quarter
BIO18	Precipitation of the Warmest Quarter
BIO19	Precipitation of the Coldest Quarter
CANOPY	Canopy Tree Cover
NDVI	Normalized Difference in Vegetation Index
LAT	Latitude Coordinates in decimal degrees, WGS84 Projection
SLOPE	Slope of Land in degrees
LONG	Longitude Coordinates in decimal degrees, WGS84 Projection
ELEV	Elevation in meters
TWI	Topographic Wetness Index
CA_HII	Human Influence Index
IMPERV	Impervious Structures
GEOL	Geology of California
ASPECT	Direction of Slope in degrees
LANDCOVER	Land Cover Layer
NI_VIIRS	Nighttime Lights
POP	Human Population Size in people/square kilometer

References

- Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*(9), 1655–1664.
<https://doi.org/10.1101/gr.094052.109>
- Bay, R. A., Harrigan, R. J., Underwood, V. L., Gibbs, H. L., Smith, T. B., & Ruegg, K. (2018). Genomic signals of selection predict climate-driven population declines in a migratory bird. *Science*, *359*(6371), 83–86. <https://doi.org/10.1126/science.aan4380>
- Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32.
<https://doi.org/10.1023/A:1010933404324>
- Browning, M. R. (1994). A taxonomic review of *Dendroica petechia* (Yellow Warbler) (Aves: Parulinae). *PROCEEDINGS OF THE BIOLOGICAL SOCIETY OF WASHINGTON*, *107*(1), 27–51.
- Capblancq, T., Fitzpatrick, M. C., Bay, R. A., Exposito-Alonso, M., & Keller, S. R. (2020). Genomic Prediction of (Mal)Adaptation Across Current and Future Climatic Landscapes. *Annual Review of Ecology, Evolution, and Systematics*, *51*(1), null.
<https://doi.org/10.1146/annurev-ecolsys-020720-042553>
- Cassin-Sackett, L., Welch, A. J., Venkatraman, M. X., Callicrate, T. E., & Fleischer, R. C. (2019). The Contribution of Genomics to Bird Conservation. In R. H. S. Kraus (Ed.), *Avian Genomics in Ecology and Evolution: From the Lab into the Wild* (pp. 295–330). Springer International Publishing. https://doi.org/10.1007/978-3-030-16477-5_10
- Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, *34*(17), i884–i890.
<https://doi.org/10.1093/bioinformatics/bty560>

- Cleland, D. T., Freeouf, J. A., Keys, J. E., Jr., Nowacki, G. J., Carpenter, C., & McNab, W. H. (2007). *Ecological Subregions: Sections and Subsections of the Conterminous United States* (Gen. Tech. WO-76). U.S. Department of Agriculture, Forest Service.
- Dahl, T. E. (1990). *Wetlands losses in the United States, 1780's to 1980's. Report to the Congress* (PB-91-169284/XAB). National Wetlands Inventory, St. Petersburg, FL (USA). <https://www.osti.gov/biblio/5527872-wetlands-losses-united-states-report-congress>
- de Sherbinin, A., Balk, D., Yager, K., Jaiteh, M., Pozzi, F., Giri, C., & Wannebo, A. (2002). *A CIESIN Thematic Guide to Social Science Applications of Remote Sensing*. Center for International Earth Science Information Network (CIESIN), Columbia University.
- DeRaad, D. A. (2022). snpfiltr: An R package for interactive and reproducible SNP filtering. *Molecular Ecology Resources*, 22(6), 2443–2453. <https://doi.org/10.1111/1755-0998.13618>
- DeSaix, M. G., George, T. L., Seglund, A. E., Spellman, G. M., Zavaleta, E. S., & Ruegg, K. C. (2022). Forecasting climate change response in an alpine specialist songbird reveals the importance of considering novel climate. *Diversity and Distributions*, 28(10), 2239–2254. <https://doi.org/10.1111/ddi.13628>
- Dybala, K. E., Clipperton, N., Gardali, T., Golet, G. H., Kelsey, R., Lorenzato, S., Melcer, J., Seavy, N. E., Silveira, J. G., & Yarris, G. S. (2017). Population and Habitat Objectives for Avian Conservation in California's Central Valley Riparian Ecosystems. *San Francisco Estuary and Watershed Science*, 15(1). <https://doi.org/10.15447/sfews.2017v15iss1art5>

- Ellis, N., Smith, S. J., & Pitcher, C. R. (2012). Gradient forests: Calculating importance gradients on physical predictors. *Ecology*, *93*(1), 156–168. <https://doi.org/10.1890/11-0252.1>
- Elvidge, C. D., Baugh, K., Zhizhin, M., Hsu, F. C., & Ghosh, T. (2017). VIIRS night-time lights. *International Journal of Remote Sensing*, *38*(21), 5860–5879. <https://doi.org/10.1080/01431161.2017.1342050>
- Fiedler, P. L., Erickson, B., Esgro, M., Gold, M., Hull, J. M., Norris, J. M., Shapiro, B., Westphal, M., Toffelmier, E., & Shaffer, H. B. (2022). Seizing the moment: The opportunity and relevance of the California Conservation Genomics Project to state and federal conservation policy. *Journal of Heredity*, *113*(6), 589–596. <https://doi.org/10.1093/jhered/esac046>
- Fitzpatrick, M. C., & Keller, S. R. (2015). Ecological genomics meets community-level modelling of biodiversity: Mapping the genomic landscape of current and future environmental adaptation. *Ecology Letters*, *18*(1), 1–16. <https://doi.org/10.1111/ele.12376>
- Heath, S. K., & Ballard, G. (2003). *Patterns of Breeding Songbird Diversity and Occurrence in Riparian Habitats of the Eastern Sierra Nevada*.
- Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, *25*(15), 1965–1978. <https://doi.org/10.1002/joc.1276>
- Jombart, T. (2008). adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403–1405. <https://doi.org/10.1093/bioinformatics/btn129>
- Kendig, K. I., Baheti, S., Bockol, M. A., Drucker, T. M., Hart, S. N., Heldenbrand, J. R., Hernaez, M., Hudson, M. E., Kalmbach, M. T., Klee, E. W., Mattson, N. R., Ross, C. A.,

- Taschuk, M., Wieben, E. D., Wiepert, M., Wildman, D. E., & Mainzer, L. S. (2019). Sentieon DNaseq Variant Calling Workflow Demonstrates Strong Computational Performance and Accuracy. *Frontiers in Genetics, 10*.
<https://doi.org/10.3389/fgene.2019.00736>
- Knaus, B. J., & Grünwald, N. J. (2017). vcfr: A package to manipulate and visualize variant call format data in R. *Molecular Ecology Resources, 17*(1), 44–53.
<https://doi.org/10.1111/1755-0998.12549>
- Knopf, F. L., & Sedgwick, J. A. (1992). An Experimental Study of Nest-Site Selection by Yellow Warblers. *The Condor, 94*(3), 734–742. <https://doi.org/10.2307/1369258>
- Krueper, D. J. (1996). Effects of livestock management on Southwestern riparian ecosystems. *In: Shaw, Douglas W.; Finch, Deborah M., Tech Coords. Desired Future Conditions for Southwestern Riparian Ecosystems: Bringing Interests and Concerns Together. 1995 Sept. 18-22, 1995; Albuquerque, NM. General Technical Report RM-GTR-272. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Forest and Range Experiment Station. p. 281-301., 272, 281–301.*
- Larison, B., Lindsay, A. R., Bossu, C., Sorenson, M. D., Kaplan, J. D., Evers, D. C., Paruk, J., DaCosta, J. M., Smith, T. B., & Ruegg, K. (2021). Leveraging genomics to understand threats to migratory birds. *Evolutionary Applications, 14*(6), 1646–1658.
<https://doi.org/10.1111/eva.13231>
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England), 25*(14), 1754–1760.
<https://doi.org/10.1093/bioinformatics/btp324>

- Mirchandani, C. D., Shultz, A. J., Thomas, G. W. C., Smith, S. J., Baylis, M., Arnold, B., Corbett-Detig, R., Enbody, E., & Sackton, T. B. (2024). A Fast, Reproducible, High-throughput Variant Calling Workflow for Population Genomics. *Molecular Biology and Evolution*, *41*(1), msad270. <https://doi.org/10.1093/molbev/msad270>
- Naiman, R. J., Decamps, H., & McClain, M. E. (2010). *Riparia: Ecology, Conservation, and Management of Streamside Communities*. Elsevier.
- Patten, M. A., McCaskie, G., & Unitt, P. (2003). *Birds of the Salton Sea: Status, Biogeography, and Ecology* (1st ed.). University of California Press.
<https://www.jstor.org/stable/10.1525/j.ctt1pnrqq>
- Pinsky, M. L., Worm, B., Fogarty, M. J., Sarmiento, J. L., & Levin, S. A. (2013). Marine Taxa Track Local Climate Velocities. *Science*, *341*(6151), 1239–1242.
<https://doi.org/10.1126/science.1239352>
- Poff, B., Koestner, K. A., Neary, D. G., & Merritt, D. (2012). *Threats to western United States riparian ecosystems: A bibliography* (RMRS-GTR-269). U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. <https://doi.org/10.2737/RMRS-GTR-269>
- Razgour, O., Forester, B., Taggart, J. B., Bekaert, M., Juste, J., Ibáñez, C., Puechmaille, S. J., Novella-Fernandez, R., Alberdi, A., & Manel, S. (2019). Considering adaptive genetic variation in climate change vulnerability assessment reduces species range loss projections. *Proceedings of the National Academy of Sciences*, *116*(21), 10418–10423.
<https://doi.org/10.1073/pnas.1820663116>
- Romanov, M. N., Tuttle, E. M., Houck, M. L., Modi, W. S., Chemnick, L. G., Korody, M. L., Mork, E. M. S., Otten, C. A., Renner, T., Jones, K. C., Dandekar, S., Papp, J. C., Da, Y.,

- Green, E. D., Magrini, V., Hickenbotham, M. T., Glasscock, J., McGrath, S., Mardis, E. R., ... NISC Comparative Sequencing Program. (2009). The value of avian genomics to the conservation of wildlife. *BMC Genomics*, *10*(2), S10. <https://doi.org/10.1186/1471-2164-10-S2-S10>
- Ruegg, K., Bay, R. A., Anderson, E. C., Saracco, J. F., Harrigan, R. J., Whitfield, M., Paxton, E. H., & Smith, T. B. (2018). Ecological genomics predicts climate vulnerability in an endangered southwestern songbird. *Ecology Letters*, *21*(7), 1085–1096. <https://doi.org/10.1111/ele.12977>
- Schweizer, T. M., DeSaix, M. G., & Ruegg, K. C. (2021). *LI-Seq: A Cost-Effective, Low Input DNA method for Whole Genome Library Preparation* [Preprint]. *Genetics*. <https://doi.org/10.1101/2021.07.06.451326>
- Sexton, J. O., Song, X.-P., Feng, M., Noojipady, P., Anand, A., Huang, C., Kim, D.-H., Collins, K. M., Channan, S., DiMiceli, C., & Townshend, J. R. (2013). Global, 30-m resolution continuous fields of tree cover: Landsat-based rescaling of MODIS vegetation continuous fields with lidar-based estimates of error. *International Journal of Digital Earth*, *6*(5), 427–448. <https://doi.org/10.1080/17538947.2013.786146>
- Shaffer, H. B., Toffelmier, E., Corbett-Detig, R. B., Escalona, M., Erickson, B., Fiedler, P., Gold, M., Harrigan, R. J., Hodges, S., Luckau, T. K., Miller, C., Oliveira, D. R., Shaffer, K. E., Shapiro, B., Sork, V. L., & Wang, I. J. (2022). Landscape Genomics to Enable Conservation Actions: The California Conservation Genomics Project. *Journal of Heredity*, *113*(6), 577–588. <https://doi.org/10.1093/jhered/esac020>
- Shuford, W. D., Gardali, T., Western Field Ornithologists, California, & Department of Fish and Game. (2008). *California bird species of special concern: A ranked assessment of*

- species, subspecies, and distinct populations of birds of immediate conservation concern in California*. Western Field Ornithologists ; California Dept. of Fish and Game.
<http://books.google.com/books?id=9INFAQAIAAJ>
- Smith, T. B., Fuller, T. L., Zhen, Y., Zaunbrecher, V., Thomassen, H. A., Njabo, K., Anthony, N. M., Gonder, M. K., Buermann, W., Larison, B., Ruegg, K., & Harrigan, R. J. (2020). Genomic vulnerability and socio-economic threats under climate change in an African rainforest bird. *Evolutionary Applications*, *n/a(n/a)*. <https://doi.org/10.1111/eva.13193>
- Stromberg, J. C., & Patten, D. T. (1990). Riparian vegetation instream flow requirements: A case study from a diverted stream in the Eastern Sierra Nevada, California, USA. *Environmental Management*, *14*(2), 185–194. <https://doi.org/10.1007/BF02394035>
- Studd, M. V., & Robertson, R. J. (1989). Influence of age and territory quality on the reproductive behaviour of male Yellow Warblers. *Canadian Journal of Zoology*, *67*(2), 268–273. <https://doi.org/10.1139/z89-039>
- Tsai, W. L. E., Escalona, M., Garrett, K. L., Terrill, R. S., Sahasrabudhe, R., Nguyen, O., Beraut, E., Seligmann, W., Fairbairn, C. W., Harrigan, R. J., McCormack, J. E., Alfaro, M. E., Smith, T. B., & Bay, R. A. (2024). *A highly contiguous genome assembly for the Yellow Warbler (Setophaga petechia)* (p. 2024.01.10.575107). bioRxiv.
<https://doi.org/10.1101/2024.01.10.575107>
- Urban, M. C. (2015). Accelerating extinction risk from climate change. *Science*, *348*(6234), 571–573. <https://doi.org/10.1126/science.aaa4984>
- USGS. (2018). *Shuttle Radar Topography Mission 1 Arc-Second Global* [Dataset]. Global Land Cover Facility. [https://doi.org/\(DOI\) number: /10.5066/F7PR7TFT](https://doi.org/(DOI) number: /10.5066/F7PR7TFT)