

# **UCLA**

## **Annual Reports**

### **Title**

California Conservation Genomics Project Third Year Annual Report

### **Permalink**

<https://escholarship.org/uc/item/70x23229>

### **Authors**

Shaffer, Howard B  
Toffelmier, Erin

### **Publication Date**

2022-05-01



## California Conservation Genomics Project Third Year Annual Report May 2022

### Overview

The California Conservation Genomics Project (CCGP) is a state-wide initiative led by UCLA to generate the most comprehensive multispecies genomic dataset ever assembled to support conservation of the state's biodiversity in response to climate change and other stressors. We have gathered leading experts in genomics and conservation science to study patterns of genetic diversity and divergence in 233 species, sampled from every ecoregion in the state. Although managed by UCLA, the CCGP consortium includes representatives from all 10 University of California campuses as well as several state and federal agencies and conservation-focused organizations. This includes 70 UC-based principal investigators and over 100 additional researchers (graduate students, postdoctoral scholars, and other collaborators) that are working on 153 interrelated projects. During the third year of the CCGP, substantial progress has been made to achieve its goals of whole genome resequencing of thousands of individuals, generating high quality reference genomes for every project, building teams and pipelines to analyze data, publicizing progress, and reaching out to decision-makers across the state.

### Key Accomplishments and Outcomes

#### *Whole Genome Sequencing Progress*

The primary goal of the CCGP is to generate whole genome sequencing data (WGS) for 233 species from 148 genera totaling nearly 22,000 individuals. This massive genomic data set will be used in conjunction with CCGP-produced reference genomes to quantify, analyze, and map spatial patterns of variation across the state. Our project investigators have collected well over half of the planned samples and sequencing is complete for nearly 20% of individuals. During this year, our Bioinformatics team has used several complete data sets to develop and refine the genetic variant calling pipeline which can now be applied to all remaining genomic data as it is generated.

#### *Reference Genome Generation Progress*

An additional goal of the CCGP is to generate high quality reference genomes for 148 of our target species, including a representative of every genus in the project. These reference genomes are essential for fully utilizing the whole genome resequencing data that comprise our main data set. Through the efforts of our network of primary

investigators, we have collected high-quality tissue samples required for 142 species, and CCGP PIs are currently working hard to source new sample material for the few remaining species. We have completed the reference genomes for 35% of the species and have generated the partial or complete raw assembly data for an additional 30% of species. We expect that the data generation and genome assembly for all remaining to species will be completed by the end of 2022. Our goal is for each of these genomes to have their own Genome Resource article in the Journal of Heredity (JOH), a top scientific journal in the field of conservation genomics. We have worked with the journal editors and its publisher, Oxford University Press, to prepare article templates and streamline the submission process for our CCGP PIs. During this year of the award, two of these papers have been accepted for publication, two are in revision, and many others are near submission.

## **Multicampus Collaborative Components and Activities**

### *Expansion of the Bioinformatics and Landscape Genomics Teams and Pipeline Development*

The CCGP team has grown to keep pace with data generation. We hired a postdoctoral scholar, Dr. Erik Enbody, to join the bioinformatics team, headquartered at UC Santa Cruz under the leadership of Professor Russell Corbett-Detig. Dr. Enbody contributes to the development, deployment, and maintenance of protocols for whole genome sequence data alignment, filtration, and variant calling. The Bioinformatics team has made tremendous progress in the development of this pipeline, which will ultimately be used to generate final genomic data sets for all 22,000 samples. During the last year, this team has optimized the pipeline with cloud computing and new software to speed up our data processing time per sample 127-fold. This will both accelerate our overall progress, and greatly reduce our bioinformatic costs (over twenty-fold reduction in per-sample cost). This optimization ultimately will allow us to achieve fast, accurate, and reproducible results.

Once the genomic raw data are processed by the bioinformatics team, the Landscape Genomics team will integrate it with environmental data to perform meaningful state-wide landscape-based analyses, including the production of maps of high and low vulnerability regions to climate change. To aid these efforts, we have hired postdoctoral scholar Dr. Anne Chambers to work with Professor Ian Wang at UC Berkeley. Dr. Chambers is responsible for developing, deploying, and maintaining a pipeline for landscape genetic analyses that involves statistical and computational modeling of GIS and genomic data, data visualization, and applications to species conservation. To aid in the development of this pipeline, we have assembled a working group comprised of UC experts in landscape analyses to develop and refine our analytical approaches so that they are applicable across taxonomic groups and habitat types. Additionally, we have begun working with state policy makers and stakeholders to determine the best avenues for the dissemination of our data and analytical products.

## *CCGP Fall WGS Webinar*

On November 1<sup>st</sup> and 2<sup>nd</sup>, 2021, the CCGP hosted an all-hands virtual meeting focused on whole genome resequencing and the next phases of the CCGP. We also took the opportunity to hear from PIs and team members on their progress and challenges, as well as to discuss resolutions and provide additional direction on optimizing progress. We shared updates from our bioinformatics team and provided guidance on genomic data organization and policies. We held four breakout sessions to discuss challenges relating to field work and bench work, separately for plants and animals. These discussions provided valuable insights on project challenges and successes. To help facilitate further progress, we explored ways PIs could expedite data generation. We also made several Slack channels available for fostering discussions, troubleshooting, and collaboration.

## **Other Important Outcomes**

### *Community Engagement*

The CCGP sends a bimonthly newsletter to the conservation science community. The newsletter highlights both CCGP-relevant research progress as well as conservation-related news that may be of interest to the broader community.

On May 11, 2022, the CCGP collaborated with seqWell, the company who provides some of our essential molecular lab supplies, to put on a webinar titled “Conserving California: Applying Whole Genome Sequencing in the California Conservation Genomics Project (CCGP).” In this webinar, we shared how high-throughput technologies can help move conservation genomics from a single species to a multi-species scale. This webinar was hosted on the leading scientific social networking website, Labroots. The live webinar had over 70 registrants and the recorded version is available for unlimited on-demand viewing.

## **Outline of Next Year’s Research Activities and Plans**

The next year’s research activity will focus on completing the data generation and analysis components of this project. Our Reference Genome team will have completed approximately 40% of the planned reference genomes by the current end date and will complete the remaining 60% midway through the extension period. Several aspects of our project were negatively impacted by COVID-19, and we will be working to complete those aspects of the project. Understandably, landscape sampling for many species was our most heavily impacted activity. Most projects have at least 50% of their samples in hand. We anticipate that most teams will complete field work in the next 3 months. Currently, 20% of species have had lab work completed and sequencing data available. We anticipate having all sequencing data available by halfway through next

year. Our Bioinformatics team currently has data for over 2000 completely sequenced individuals in hand and has developed a pipeline to process these data efficiently using cloud computing. They will continue to process data in the coming year as is it submitted and aim to complete this component of data generation during next year. Finally, we have assembled the Landscape Genomics team who will use outputs of the Bioinformatics team to generate final products for conservation management, including landscape maps of genomic variation. The Landscape Genomics team has already made significant progress in developing their analytical pipelines and we anticipate that these will be ready for deployment in the next 6 months. We expect that the timeline outlined herein will allow us the needed time to finalize all datasets and pipelines leading to the generation of our proposed deliverables.

## **List of Talks**

*CCGP Director Professor Brad Shaffer:*

Vertebrate Genomes Project/Earth BioGenome Project, April 2022; virtual

UC Cooperative Extension workshop on post-fire recovery, February 2022; virtual

Plant and Animal Genome (PAG) Conference XXIX

Lead talk in workshop on Population and Conservation genomics, January 2022; Snowbird, UT/virtual

UNLV Guest lecturer, Cons. Genetics course, November 2021; virtual

California Conservation Genomics Project, November 2021; virtual

University of South Florida, November 2021; virtual

American Genetic Association Symposium - Conservation Genomics: Current Applications and Future Directions, October 2021; virtual

Biodiversity Genomics 2021: "Sequencing genomes across the planet", September 2021; virtual

*CCGP Staff:*

Institute of Bioinformatics, University of Georgia, "Harnessing genomics to characterize and conserve biodiversity" (invited seminar)  
April 2022

HHMI Virtual Science Meeting "Deconstructing and Decoding the Genome" (poster/short talk), April 2022

Vertebrate Genomes Project/Earth BioGenome Project, April 2022; virtual

Molecular Biology, Genetics, and Genomics at UCLA, ULCA Graduate Student Recruitments, Feb 2022; virtual

American Genetic Association Symposium - Conservation Genomics: Current Applications and Future Directions, October 2021 (poster)

## Publications

Shaffer, H. B., E. Toffelmier, R. B. Corbett-Detig, M. Escalona, B. Erickson, P. Fiedler, M. Gold, R. J. Harrigan, S. Hodges, T. K. Luckau, C. Miller, D. R. Oliveira, K. E. Shaffer, B. Shapiro, V. L. Sork, and I. J. Wang. Landscape genomics to enable conservation actions: The California Conservation Genomics Project. *Journal of Heredity*, 3 April 2022 DOI: 10.1093/jhered/esac020

Yi Huang, Merly Escalona, Glen Morrison, Mohan P A Marimuthu, Oanh Nguyen, Erin Toffelmier, H Bradley Shaffer, Amy Litt. Reference Genome Assembly of the Big Berry Manzanita (*Arctostaphylos glauca*). *Journal of Heredity*, 2021;, esab071, <https://doi.org/10.1093/jhered/esab071>

Brian D. Todd, Thomas S. Jenkinson, Merly Escalona, Eric Beraut, Oanh Nguyen, Ruta Sahasrabudhe, Peter A. Scott, Erin Toffelmier, Ian J. Wang, H. B. Shaffer. 2022. Reference genome of the northwestern pond turtle, *Actinemys marmorata*. *Journal of Heredity*

Giacomo Bernardi, Jason A. Toy, Merly Escalona, Mohan P. A. Marimuthu, Ruta Sahasrabudhe, Oanh Nguyen, Samuel Sacco, Eric Beraut, Erin Toffelmier, Courtney Miller, H. Bradley Shaffer. Reference genome of the Black Surfperch, *Embiotoca jacksoni* (Embiotocidae, Perciformes), a California kelp forest fish that lacks a pelagic larval stage. *Journal of Heredity* (In review)

Meixi Lin, Merly Escalona, Ruta Sahasrabudhe, Oanh Nguyen, Eric Beraut, Michael R. Buchalski, Robert K. Wayne. A Reference Genome Assembly of the Bobcat, *Lynx rufus*. *Journal of Heredity* (In review)

Lewin, H. A. (plus 84 co-authors). The Earth BioGenome Project 2020: Starting the clock. 2022. *Proceedings of the National Academy of Sciences, USA*, 119 :1-7 (No. 4 e2115635118) <https://doi.org/10.1073/pnas.2115635118>

Beninde, J., E. M. Toffelmier, A. Andreas, C. Nishioka, M. Slay, A. Soto, J. P. Bueno, G. Gonzalez, H. V. Pham, M. Posta, J. L. Pace, and H. B. Shaffer. 2022. CaliPopGen: A genetic and life history database for the fauna and flora of California. *Scientific Data*, in press.

MacDonald, Z.G., Shaffer, H. B., Sperling, F.A.H. 2022. Impacts of land use and climate change on natural populations: the butterfly perspective. Chapter 8 In: Cork, S., Whiteside, D. (eds.) Case Studies in Eco Health. 5m Publishing, Sheffield, U.K. in press.