

UCLA

Annual Reports

Title

California Conservation Genomics Project Second Year Annual Report

Permalink

<https://escholarship.org/uc/item/4209w7hx>

Authors

Shaffer, Howard B
Toffelmier, Erin

Publication Date

2021-05-31



California Conservation Genomics Project Second Year Annual Report May 2021

Overview

One million species are at risk of extinction worldwide, many of which are found in—or restricted to—California. Climate change is stretching species and habitats to their breaking points. To address this crisis, the California Conservation Genomics Project (CCGP) is engaging leading experts in genomics and conservation science to rapidly sample and study the genetics of 235 species that span the breadth of California's native species and ecosystems. The State of California provided \$10 million to UCOP to administer this work. Researchers and experts from all 10 University of California campuses are involved in this research partnership, which is managed by UCLA.

Key Accomplishments and Outcomes

Reference Genome Generation Progress

During Year 2, we have made substantial progress towards our goal of generating reference genomes for each of the roughly 150 CCGP projects. Towards this goal, we have received sample submissions for nearly 70% of our species which are working their way through the DNA extraction and data generation phases, while approximately 20% are already in some stage of the data assembly phase. We have focused on the development of an efficient and consistent pipeline for processing these data and generating reference genome assemblies-- the final, publishable product that comprises a basic tool for all future work on the species. To meet our goal of having all of the reference genomes completed by the end of calendar-year 2021, our reference genome team has worked closely with the four UC Core Facilities that play a role in data generation, and consulted with industry experts and other genome consortia, such as the Earth BioGenome Project, to use their existing pipelines and resources, as appropriate, for our genome assembly workflow. Given the volume of data that we are generating in a short amount of time, the UC Davis DNA Technologies Core has acquired a new PacBio Sequel IIe Long Read DNA Sequencing System dedicated exclusively to the CCGP's needs for the next year; the purchase of this new sequencer, at a very attractive price, was only possible because we could guarantee the purchase of chemistry and supplies at a very high level. Finally, to manage the influx of sample submissions and keep PIs and project team members informed of where their species are in the pipeline, we have set up a species tracking page on the CCGP website. The page displays the overall status of sample submissions as well as detailed progress at each UC facility. Together, these accomplishments will allow us to generate high quality reference genomes for approximately 150 species that will serve as a valuable resource

for the 75 supported projects as well as future research in landscape and conservation genomics.

Establishment of the CCGP Mini-Core at UCLA

As of early 2021, the CCGP Mini-Core is fully operational and equipped to handle DNA isolations and library preparations for whole genome resequencing for the full range of taxonomic groups, including vertebrates, invertebrates, and terrestrial and aquatic plants, that the CCGP supports. This completely new facility includes custom-tailored methods and protocols, specially designed laboratory kits, and a liquid handling robot for automation, all of which is overseen by our dedicated lab technician. Based on feedback from CCGP projects, we expect that the facility will process about half of all CCGP samples. As projects finalize their sample collections and submissions, we have received 712 of the 9900 expected samples (~7%) and expect to be able process upwards of 300 samples per week given our novel, streamlined workflow.

Setbacks or Challenges

The COVID-19 pandemic has continued to impact the CCGP and the research groups it supports in various ways. Meetings and workshops have continued virtually, and while we have become accustomed to this medium, it has inevitably delayed the pace of decision making. With the UCLA campus navigating various tiers of closure, our team has worked to adapt to each new development in guidelines by altering procedures and protocols. The CCGP's funded projects have almost invariably reported various challenges to sample collection. Teams have experienced reduced or delayed field sampling efforts due to COVID-19 (e.g. fewer personnel to sample per trip, difficulties in permit acquisition), but also due to the extreme drought across the state (e.g. fewer individuals, especially plant and insect species, at sampling locations). Despite these obstacles, we have worked to streamline pipelines for the samples we have received and are continuing to strive to meet our yearly goals. However, we anticipate that we will require a fourth year (that is, a one-year extension) of the CCGP to complete its primary goals.

Multicampus collaborative components and activities

Additional Proposals Funded

The funding decisions made in the first year of this project represent a substantial amount of California's taxonomic diversity. Following an internal review by the CCGP leadership and the CCGP Scientific Executive Committee (SEC), we determined that several important taxonomic groups were missing from the scope of the supported projects. We identified groups that, in addition to being underrepresented in the current taxonomic set, have endangered taxa that are restricted to California, or are otherwise of great interest, and would benefit from concerted conservation genomics work. These included butterflies, beetles, succulents/cacti, freshwater fishes, black widow spiders, and scorpions. We initiated a second call for proposals focusing on these identified

groups in March 2021. Following reviews conducted internally by the SEC, five new proposals were funded, including 14 new species, led by faculty from UC San Diego, UC Riverside, UC Berkeley, and UC Los Angeles. A sixth proposal, led by UC Berkeley, is still under consideration.

CCGP PI Genomics Webinar (Dec. 7 and 10, 2020)

CCGP conducted an update and planning webinar to deliver guidance on the next major data generation phase: Whole Genome Resequencing. The webinar covered the three options for data generation, including the protocol for submitting material to the UCLA Mini-Core as well as options for non-Mini-Core submissions. CCGP also provided a brief update on the progress of the generation of reference genomes.

Establishment of the Bioinformatics and Landscape Genomics Working Groups

We have begun to assemble the working groups which will direct the Bioinformatics and Landscape Genetics working groups. These two groups, along with a third dedicated to Policy and Applications, will comprise the core analysis groups for CCGP. We have appointed Russel Corbett-Detig, assistant professor in the Department of Biomolecular Engineering at UC Santa Cruz, to lead the Bioinformatics team. Dr. Corbett-Detig, along with a dedicated postdoctoral scholar bioinformatician and several advisory members from the CCGP community, will develop and implement standardized and efficient pipelines for processing ~18,000 resequenced genomes from across CCGP projects. We have appointed Ian Wang, associate professor in the Department of Environmental Science, Policy, and Management at UC Berkeley to lead the Landscape Genomics team. Dr. Wang, a dedicated landscape genomics postdoctoral scholar, and several advisory members from the CCGP community, will work to develop and execute analytical pipelines that generate data sets and mapping tools that will be used by the Policy and Applications Working group to present our results to decision-makers as the project progresses.

Other important outcomes

Launch of the CCGP Website

CCGP has created a public website (www.ccgproject.org). Its primary goals are to 1) provide relevant information about the project to both the general public and specialists, 2) host resources for CCGP researchers and team members, and 3) provide a platform for policy makers and land management agencies to access products and resources for conservation planning. Information about CCGP includes general background on the project, its goals, and who is involved (e.g. leadership organization, researchers, and CCGP partners) while showcasing the vast diversity of species that will be used to influence land use management and conservation actions. The shared resources for CCGP collaborators include information on sample requirements and submission guidelines for generating reference genomes and whole genome resequencing data as well as tracking the status of sequencing progress for their

samples. Material and recordings from previous planning meetings and webinars are also publicly available. The platform for conservation planning on the website will host a collection of interactive landscape genomic maps and other tools and resources. This website is considered a resource in and of itself and is actively updated with several aspects of project progress, including data generation, relevant announcements from collaborators (e.g. research and public engagement opportunities), and links to where CCGP has been featured in the press. Overall, this website is a centralized location for rapid communications and sharing of deliverables that will grow and adapt along with the needs of CCGP and its partners.

New Key Team Members

We have hired several new personnel which comprise the core administrative and technical staff for CCGP. These new positions have helped us to ensure that the various efforts of CCGP are carried out efficiently and that deliverables meet our high-quality standards. We describe each of these team members and their contributions to the project below.

One of the primary CCGP outputs will be high quality reference genomes. We have hired an experienced postdoctoral scholar, Dr. Merly Escalona, to develop our CCGP reference genome assembly pipeline, process and assemble reference genomes for all funded projects, deliver data to CCGP, work with project PIs and research teams to finalize reference genomes, submit them and relevant supporting material to NCBI (the National Center for Biotechnology Information) for public use. As the genome assembly coordinator, Dr. Escalona's additional responsibilities include attending CCGP project meetings, coordinating genome assembly meetings, and assisting teams with preparing manuscripts that describe each genome assembly. To support these efforts, we also enlisted the help of an experienced genomicist, Dr. Stephen Richards, to consult as a temporary reference genome project manager. Dr. Richards helped us establish communication between the four UC labs (UCLA, UCD, UCSC, and UCB) that play a role in processing reference genome samples, helped to coordinate the collection of tissue samples from all of the supported projects, and helped develop a CCGP reference genome generation roadmap to guide the completion of the reference genome generation process. He finished his six-month contract in March 2021.

A second primary output of CCGP will be ~19,000 resequenced whole genomes, spread across all CCGP projects. This number of samples and the sheer volume of data that will be generated through the efforts of all the CCGP sub-projects constitute enormous organizational and technological challenges. We have hired a data manager, Mr. Cade Mirchandani, to develop, deploy, and maintain data management protocols for our data use and archiving. In collaboration with Dr. Corbett-Detig, he will develop bioinformatics workflows to confirm the quality of the sequence data and work with the individual sample submitting labs to confirm the accuracy and the quality of sample metadata. Mr. Mirchandani will be working at UC Santa Cruz.

We have hired a dedicated laboratory technician, Mr. Daniel Oliviera, to manage and operate the CCGP Mini-Core. As described in the progress report above, the CCGP Mini-Core is our dedicated facility to prepare samples for both whole genome resequencing and RNA sequencing for reference genome assembly. Mr. Oliviera has set up the operating space for the core, and has designed, adapted, and tested new protocols to maximize the Mini-Core's ability to manage a vast variety of sample types and challenges. He, and the Mini-Core that he runs, are at UC Los Angeles.

We have hired an intercampus project and fund manager, Dr. Courtney Miller, who has a background in conservation biology and experience working with multi-institutional conservation projects. One of the key challenges of scale for CCGP is administrative and technical coordination between CCGP PIs, their institutions, and our key partners. Dr. Miller coordinates and manages the reference genome generation process by communicating with PIs regarding sample submissions and maintaining web platforms to track CCGP progress. As project manager, she also handles various aspects of communications support including website development and maintenance, helps organize CCGP planning webinars and meetings, and works with some of UCLA's environmental communications personnel to feature CCGP in public-facing/outreach efforts. Dr. Miller also serves as our dedicated CCGP Fund Manager. As fund manager, she processes purchasing orders pertaining to lab supplies and services, distributes CCGP funding to PIs across UC campuses, reconciles accounts by tracking and verifying expenses, and organizes financial reports.

List of Talks

CCGP Director Professor Brad Shaffer:

Wellcome Trust Biodiversity Genomics (plenary in Conservation session), October 2020
California Department of Fish and Wildlife Conservation Lecture, October 2020
City of Los Angeles Biodiversity symposium, November 2020
Museum of Vertebrate Zoology, UC Berkeley, November 2020
Santa Ynez Valley Natural History Society, February 2021
US Fish and Wildlife Service seminar, Ventura office, March 2021
Northern California Science Writers Association April, 2021

CCGP Reference Genome Team Lead:

Vertebrate Genome project/Earth BioGenome Weekly meeting (invited talk organized by Adam Phillippy NHGRI - Merly Escalona, March 2021

CCGP Scientific Staff:

EEB Recruitment Weekend 2021, "Molecular Biology, Genetics, and Genomics at UCLA" - Ryan J. Harrigan, April 2021

Grand Challenges URSP Lecture, "Conservation of Ecological Process in California" – Ryan J. Harrigan, April 2021

CCGP Projects:

PAGBio (Plant and Animal Genomics Day for PacBio) - Chloe Orland, January 2021

Dovetail Genomics' "Genomics of Animals and Plants" Conference - Beth Shapiro, January 2021

State of Biodiversity Symposium (run by the SD Natural History Museum) - Hollis Woodard, April 2021

Long-term Genetic Diversity Monitoring Workshop, ETH Zurich - Victoria Sork, April 2021

Departmental Series of Biology, San Diego State University - Victoria Sork, May 2021