



## CALIFORNIA CONSERVATION GENOMICS PROJECT

---

### Dear CCGP Investigators,

As we transition from the data collection and processing phases to the data analysis phase, we would like to share some important information and reminders regarding **CCGP's organization on NCBI** and the two data components of the CCGP: **whole genome resequencing** and **reference genome assemblies**.

---

### CCGP on NCBI

We are in the process of establishing [CCGP's organizational framework on the NCBI](#). Currently, the CCGP (Umbrella BioProject [PRJNA720569](#)) contains genus-level Umbrella BioProjects. In general, these genus-level Umbrella BioProjects will contain the reference genome and whole genome resequencing data for all of the species within that genus.

We are now setting up BioProjects for all the CCGP species and **require that PI(s) or project team members complete the following brief survey for all of the species within their project (approx. 3-5 min/species)**. Please return these surveys by **5pm PDT October 22nd**. We will prioritize reference genome assemblies for species with completed BioProject information.

[Please click here to complete species BioProject survey](#)

---

### Whole Genome Resequencing

#### *Data Submission to the CCGP*

Based on PI feedback, we expect that about half of all CCGP WGS data will be generated outside of the CCGP Mini-Core. This means that it is your responsibility to share raw resequencing data with the CCGP Bioinformatics team. In addition, all projects are required to provide key metadata for all of their resequencing samples.

If your project has already, or plans to generate data outside of the CCGP Mini-Core, **please review** our [Whole Genome Resequencing Data Ingest and Management Plan](#) for more information on how to share raw sequencing data and associated metadata. Projects generating sequence data through the Mini-Core do not need to submit data independently.

If you have any questions about this data sharing process, please contact our Data Wrangler, [Cade Mirchandani](#).

Please note that available WGS data is one of the factors used to prioritize species in the reference genome assembly pipeline.

### *Submission of Partial Data Sets*

We understand that this has been a difficult year for a variety of reasons. Because of this, we expect that some groups may not have been able to complete field sampling or laboratory work according to our original timelines. Projects may continue field work into the Spring of 2022 and should plan on completing laboratory work and sequencing within three months of completing field work.

However, we would like to emphasize that projects that already have samples in hand should be working to generate and submit sequencing data regardless of whether they have completed sample sets. Projects planning on utilizing the CCGP Mini-Core can submit sample sets of roughly half or more of their planned samples. We encourage submission of partial sequencing data sets to help move the project forward.

### *Supplemental Sequencing Grants*

We have received feedback that for some projects, the original sequencing budget will not provide enough sequencing depth to meet CCGP project goals. We are pleased to offer **small supplemental sequencing grants** to cover this gap. Our project-wide goal is to achieve ~10x genome coverage for approximately 150 resequenced genomes for each species project. If your current sequencing budget will yield less than roughly 10x coverage per genome, we encourage you to submit an application for additional funds. For budget calculations, please use the \$7 per Gbp sequencing rate (Novaseq 4000) available through QB3 Genomics at UC Berkeley.

Please see [here](#) for more information on library prep and sequencing options.

Acceptable reasons for a supplemental request include:

- The original budget incorrectly estimated sequencing need or miscalculated total cost.
- The genome size estimate used to calculate the original sequencing need was incorrect.
- The effective genome coverage will drop below 8x after removing duplicates.

Many projects have inquired about additional sequencing effort to adjust for mitochondrial or chloroplast genomes present in the sample. However, our genome experts have indicated that organellar reads generally comprise less than 2% of all WGS reads, and so will very minimally impact nuclear genome coverage. Thus, this is almost certainly not a sufficient justification for requesting additional sequencing funds.

Similarly, there has been concern over sequence duplication rates. We set the 10x coverage goal with the understanding that there would be variation across species, samples and library preparations, and thus expect that the ultimate project-wide coverage will be something less than 10x. If you observe duplication rates that drop coverage to below 8x, consider applying for additional funding.

At this time, we cannot provide supplemental funds to support personnel. This is purely to help out with sequencing costs. Under most circumstances, we expect this will be relatively modest, and not more than 5-10% of the original request.

For questions regarding Supplemental Sequencing Grants, please contact [Erin Toffelmier](#).

To request supplemental funds, please fill out this request form:

---

## Reference Genomes

### *Completed assemblies on NCBI*

- As with the resequencing data, reference genome assemblies and their accompanying metadata will be uploaded to NCBI as soon as the assembly is complete. For this to happen, we will require PI(s) to provide various pieces of information, one part of which is the species BioProject survey above.

### *Draft HiFi genomes*

- HiFi-based non-curated contig-level assemblies generated through our automated pipelines will be made available to CCGP researchers immediately upon completion. **These draft assemblies do not constitute the final reference genome assembly**, but can be utilized by researchers for preliminary analyses, protocol and pipeline development, and collaborative efforts.
- More information can be found [here](#).

### *Genome Release papers in the Journal of Heredity (JOH)*

- Once your reference genome assembly is uploaded to NCBI, the accompanying JOH Genome Release article should be submitted. More information on this process can be found [here](#).

---

## Summary of General CCGP Data Resources

All of the information regarding the CCGP's data and data resources, including all of the CCGP's data release policies can be found on the CCGP website [here](#).

Thank you as always for your participation and collaboration. Please don't hesitate to reach out with any questions regarding the above information.

**Best,  
The CCGP Team**

P.S. Please don't forget to register for the upcoming CCGP Whole Genome Resequencing Webinar ([Session A](#) or [Session B](#)).

---

**California Conservation Genomics Project**

[www.ccgproject.org](http://www.ccgproject.org)

[Unsubscribe](#)



mailer lite