CCGP Whole Genome Resequencing Sample Submission Guidelines

This document describes the CCGP sample requirements and submission guidelines for whole genome resequencing (WGS).

For general inquiries about CCGP WGS, please contact Erin Toffelmier (etoff@ucla.edu). For inquiries regarding the CCGP Mini-core, please contact Andrew Tully (andrewtully@g.ucla.edu).

Table of Contents

General Guidelines 2
  Number of Samples per Species 2
  DNA Sequencing Facilities 2
  Sequencing Coverage 2
Data Sharing and Submission to CCGP 3
  Timeline 3
  Submitting your data to CCGP 3
Overview of WGS Library Prep and Sequencing Options 4
  Option 1. PI-generated libraries 4
  Option 2. Library generation at UC Core 5
  Option 3. CCGP Mini-Core 5
General Guidelines

Number of Samples per Species

- Please try to adhere to your funded sampling plan/excel spreadsheet
- Sampling schemes should maximize geographic coverage over the California distribution of the target taxa
- This should include ≤150 samples per species/genus
- Because we are focused on geographic breadth, population sampling is discouraged

DNA Sequencing Facilities

- A stipulation of the CCGP funding was that sequencing is performed at one of the University of California core labs
- The DNA Technologies Core at UC Davis, and the QB3 Genomics Core at UC Berkeley are familiar with the scope and goals of CCGP projects, but the use of other UC sequencing cores is acceptable

Sequencing Coverage

- Whole genome resequencing should aim for 10x sequencing depth, after adjusting for duplicate reads and organelle presence, with a likely outcome of between 8-12x after data filtering
  - duplicate reads may range from 10-20%, depending on the library prep type
  - organelles may account for up to 2% of DNA yield depending on species and tissue type
- One S4 NovaSeq lane (150bp paired end reads) yields 600-750 Gigabases/ 2-2.5 billion reads (PE). These are the specifications provided by Illumina; your sequencing core may achieve higher numbers (often up to 800-900 GB)
- Consult with a bioinformatician to confirm that you are targeting an appropriate level of coverage
Data Sharing and Submission to CCGP

All resequencing data generated as part of a CCGP award must be shared with CCGP and also be made publicly available.

Timeline

- All resequencing should be completed by Summer 2022
- All resequencing data should be shared with CCGP as it is produced. Sharing this data will not interfere with PI(s) ability to access or use this data.

Submitting your data to CCGP

I. Data generated by the CCGP Mini-Core will automatically be captured (WGS option 3 below).

II. For projects that do not utilize the Mini-Core, please see our WGS data ingest instructions.

Overview of WGS Library Prep and Sequencing Options

There are three basic options for generating whole genome resequencing libraries.

1. Extract DNA and generate your own libraries
2. UC core facility extracts DNA and/or generates libraries for you
3. CCGP Mini-Core extracts DNA and/or generates libraries for you

Option 1. PI-generated libraries

Extract DNA and generate your own libraries. Most PIs have indicated that will extract gDNA and generate their own libraries.

I. Library prep method: Library preparation method is up to individual labs to choose so that these data will fit in with your preferred workflows and be compatible with your non-CCGP projects.

II. Indexing: We strongly recommend dual indexing to avoid index hopping. See this site for more information.

There are two index options:

A. Preferred. Use indexes provided by UC Berkeley
   1. CCGP projects may submit partial lanes to UCD or UCB
2. UCD or UCB will pool with other compatible libraries to maximize lane efficiency and minimize sequencing cost per sample
3. Sequencing will be billed as a percentage of the lane used
4. Email gsl-fgl@berkeley.edu to request index plates from UCB (please include shipping address)
5. You must send these libraries to either UCB or UCD for sequencing
6. To submit these samples for sequencing:
   a) Follow the standard Illumina Novaseq submission protocols from either UCD or UCB
      (1) UCB: https://qb3.berkeley.edu/facility/genomics/
      (2) UCD: https://dnatech.genomecenter.ucdavis.edu/
   b) Note in the submission that the submission contains CCGP samples
   c) Include the index plate/well information for each sample

B. Use your own indexes
   1. Core sequencing facilities generally will not pool your samples with other projects if you use your own indexes because they cannot guarantee that indexes from different projects are compatible.
   2. You can maximize lane efficiency (and cost) if you completely fill out a lane
   3. You may sequence at any UC sequencing core

III. Pooling: Keep in mind that in most cases, Cores will charge a per sample fee for pooling libraries into a sequencing pool. Please contact the Core to confirm pricing and logistics.
   A. If you do choose to have the core pool your libraries, submit individual samples in tubes only (DO NOT submit samples in plates).

Option 2. Library generation at UC Core
Most UC genomics cores can extract DNA, generate whole genome libraries, and complete short read sequencing. Users can submit either tissues, or already extracted gDNA.

- We recommend that you contact your chosen Core prior to starting projects to confirm availability, timelines, tissue and DNA requirements, and cost.
- Generally, this option will mean that the libraries the core generates for you will be pooled and sequenced with other libraries that the Core makes, and you will be billed based on the percent of the lane that is allocated to your libraries.
- The QB3 Genomics Core at UC Berkeley, and the DNA Technologies Core at UC Davis are both familiar with the scope and goals of CCGP projects, but other UC cores should offer similar services.
- Follow the submission instructions laid out by the chosen core
● There is no need to specifically identify these projects as related to CCGP when you submit them.
● Remember, Cores often charge a lot for these services, and this expense comes from your budget.

**Option 3. CCGP Mini-Core**

For labs that choose to use it, CCGP offers at-cost DNA extraction and library preparation.

**Approach:**
- We can start with **high quality tissue or high quality extracted gDNA** that PIs provide
- DNA extractions will be performed robotically using standard DNA extraction kits
- Our current approach is that whole genome libraries will be made using the plexWell (seqWell) library preparation protocol, similar to Nextera library preparations. This may change as we work out final details.
  - For genomes smaller than 0.5 Gb, we will use the LP 384 kit.
  - For genomes larger than 0.5 Gb, we will use the WGS24 kit.
- Libraries will be pooled with other CCGP projects and submitted on your behalf to a UC Core.

**Submission:**
- To submit either tissue or gDNA to the Mini-Core, please begin by filling out and submitting this submission form: [https://airtable.com/shrwTLw4Z49ukKTmU](https://airtable.com/shrwTLw4Z49ukKTmU)
- Please refer to the MiniCore Submission Guidelines for specific tissue and DNA requirements.

**Billing:**
- For extraction and library preparation services, the Mini-Core **can only accept funds that were distributed as part of the original CCGP award**
- Sequencing services will be billed by the sequencing core for the portion of the lane(s) used (we will submit the libraries for sequencing on your behalf and include you in the process).

**Pricing:**

<table>
<thead>
<tr>
<th>Mini-Core</th>
<th>per sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA extraction from tissue + library preparation</td>
<td>$30</td>
</tr>
<tr>
<td>Library preparation from user supplied gDNA</td>
<td>$20</td>
</tr>
<tr>
<td>Service</td>
<td>Cost</td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Additional QC, if needed</td>
<td>$4</td>
</tr>
<tr>
<td>Additional extraction, if needed</td>
<td>$6</td>
</tr>
<tr>
<td>EDTA cleanup (SPRI based), if needed</td>
<td>$3</td>
</tr>
<tr>
<td>Troubleshooting</td>
<td>$5</td>
</tr>
</tbody>
</table>