

CCGP Genetic Data Release Policies

Whole Genome Resequencing Data

Whole genome resequencing data (FASTQ files and associated metadata) will be uploaded to the NCBI Short Read Archive (SRA). Because of the scale of the CCGP sequencing effort, and to help PIs with the onerous task of data uploading, we will begin this process upon receipt of sequencing data from CCGP projects. However, WGS short read data will be **embargoed and remain private** until either the PI chooses to release the data, or shortly before the planned submission of the first synthetic CCGP publication, whichever comes first.

The expected submission date of this first data-driven synthetic manuscript is the second quarter of 2023. Consistent with the data policies of most journals, sequencing data must be publicly available by the time publication of any paper that uses those data, whether it is from the PI lab or in the form of a large CCGP synthesis paper. We will adhere to these policies regardless of whether resequencing samples are run through the CCGP Mini-Core or individual PI labs. Please contact CCGP Director Brad Shaffer (brad.shaffer@ucla.edu) to discuss any questions regarding WGS embargoes.

Draft HiFi Reference Assembly

Upon assembly completion, CCGP will make the assembly files (hosted on CCGP servers), and the quality metrics available via password protected FTP download. Project PIs will be notified of the availability and provided with download instructions. These draft assemblies are meant as an intermediate resource for CCGP projects, and because of their incomplete nature, are not intended for submission to any public genome repository (e.g., NCBI's Genome Database, or GenomeArk). As reference genomes are finalized with NCBI, these temporary files will be removed from our servers and PIs will be provided with permanent links to the NCBI database.

Completed Reference Genomes

Completed reference genome assemblies will be publicly released through NCBI's Genome Database, following their standard data validation procedure. Our intention is to help PIs publish a genome release note in the Journal of Heredity (JOH) as soon as the assembly is finished, and make these reference genomes available for general use immediately.

In some cases, we may grant an exception to this public release timeline and allow the reference to be embargoed for a few months if the principal investigator makes a compelling case that public release would seriously hinder their research program. Note that an embargoed genome cannot go through NCBI genome annotation, nor can the release paper in JOH be published. Please contact CCGP Director Brad Shaffer (brad.shaffer@ucla.edu) to discuss reference genome embargoes.

RNAseq Data

RNAseq data generated through CCGP will be uploaded to NCBI and publicly released as soon as it is available. We will request that NCBI start annotation as soon as a reference genome is finalized in the NCBI genome database and transcriptomes are complete. PIs may request that we embargo their RNAseq data until the relevant reference genome is finalized or if they wish to complete annotation independently. PIs will be notified before data upload, and may request that RNAseq data files are embargoed at that time. However, in order for genome annotations to proceed through NCBI, all data (the final reference assembly and the RNAseq data) must be public. Please contact CCGP Director Brad Shaffer (brad.shaffer@ucla.edu) to discuss RNAseq embargoes.