



JGI Legacy Data Policies

JGI Data Release Policy

The JGI Standard Data Release Policy for sequencing and metabolomics is immediate public release for all user projects accepted through the Community Science Program (CSP), Facilities Integrating Collaborations for User Science (FICUS) initiative, Bioenergy Research Centers (BRCs) program or as Director's Science projects. The data released includes the following:

- Raw sequence data: fastq or other file formats as appropriate are deposited to NCBI's Sequence Read Archive (SRA) after completion of standard analysis as described in the [product offerings](#).
- Genome projects: assemblies and annotations are posted to JGI portals as soon as each genome project is complete. Microbial sequences (raw data and assembled contigs) are released after assembly and annotation in IMG. Fungal and plant sequences are released after assembly and annotation with RNA-seq data in MycoCosm and Phytozome, respectively.
- Functional genomics data (e.g. resequencing and transcriptomes): analysis files, including alignments to reference genomes, are released upon completion of standard deliverables as described in the [product offerings](#).
- Improved assemblies and annotations and associated data will be publicly released as soon as complete.
- Metabolomics projects: raw data will be released on the JGI Genome Portal upon completion of standard deliverables

The JGI Standard Release Policy for DNA synthesis projects requires that detailed sequence information for constructs be made publicly available through the Inventory of Composable Elements (ICE) repository two years after construct delivery for all user projects accepted through the Community Science Program (CSP), Facilities Integrating Collaborations for User Science (FICUS) initiative, Bioenergy Research Centers (BRCs) program or as Director's Discretion projects.

This JGI Standard Data Release policy is effective for all user proposals/projects approved January 1, 2019 and later. **All data are subject to the data usage policies described below.**

JGI Data Utilization Policy

Data produced and released to the public by the JGI is subject to a Data Usage Policy.

As a public service, the data from sequencing projects are made available to the community by the Department of Energy Joint Genome Institute (JGI) before their publication in the scientific literature. The purpose of this policy is to balance the imperative of DOE and JGI that the data from its sequencing projects be made available to the scientific community as soon as possible with the reasonable expectation that the principal investigators will publish the primary report without concerns about preemption by other groups that did not participate in the effort.

The Data Usage Policy for data generated under proposals accepted before November 2018 is:

Update May 20, 2021

In support of open science and as part of DOE's ongoing effort to harmonize data policies, data and analysis products generated under proposals accepted before November 30th 2018 are now **required to be public without use-restrictions at the latest two years from now, by May 31st, 2023**. Use-restrictions will lift earlier than May 2023 if the data is published, or at the written request of a PI. When data is unrestricted it will be available without use-restrictions on JGI portals and will be publicly released to appropriate community repositories including NCBI (SRA, GenBank).

For proposals that are not yet completed and for which data generation is ongoing, data and analysis products that are generated after May 31st, 2021 will be subject to the JGI Data Utilization Policy below that was previously only applied to data from proposals accepted after November 2018. This means that this data will be use-restricted for two years after public release of the individual data or analysis product to JGI portal(s), until publication by the data-generating PI, or at the written request of a PI, whichever comes first. See below for complete policy terms.

Original Utilization Policy

By accessing these data, you agree not to publish any articles containing analyses of genes or genomic data prior to publication by the principal investigators of its comprehensive analysis without the consent of project's principal investigator. These restrictions will be lifted on the publication of the whole genome or comparable description. Scientists must

contact the principal investigator and JGI about their intentions to include any data from this project in a publication prior to publication of principal investigator's primary report in order to ensure that their publication does not compete directly with planned publications (e.g. reserved analyses) of the JGI and collaborators.

The principal investigator, listed on the relevant portal page, is the point of contact and arbiter regarding publication plans. Reserved analyses include but are not limited to the identification and analysis of complete (whole genome) sets of genomic features such as genes and gene expression levels, gene families, gene clusters, regulatory elements, repeat structures, GC content, etc., and whole-genome comparisons of regions of evolutionary conservation or change. Manually annotated genes within the Genome Portal are also reserved. Studies of any type on the reserved data sets that are not in direct competition with those planned by the JGI and its collaborators may also be undertaken after an agreement with principal investigator and JGI. Interested parties must contact the principal collaborator and JGI to discuss such possibilities.

If these data are used for publication, the following acknowledgment should be included: "These sequence data were produced by the US Department of Energy Joint Genome Institute <http://www.jgi.doe.gov/> in collaboration with the user community." We request that you notify us upon publication so that this information can be included in the final annotation.

The data may be freely downloaded and used by all who respect the restrictions in the previous paragraphs. While still in waiting period, the assembly and raw sequence reads should not be redistributed or repackaged without permission from the JGI. When such permission has been obtained, any redistribution of the data during this period should carry this notice: "The Joint Genome Institute provides these data in good faith, but makes no warranty, expressed or implied, nor assumes any legal liability or responsibility for any purpose for which the data are used." and users should be directed to this release policy page. Once moved to unreserved status, the data are freely available for any subsequent use.

The Data Usage Policy for data generated under proposals accepted after November 2018 is:

By accessing JGI data, you agree not to publish any articles containing analyses of genes or genomic data prior to publication by the principal investigators of its comprehensive analysis without the consent of the project's principal investigator(s). These restrictions will be lifted upon **publication(s) of the dataset** or **two years after the public release of this data**, whichever is first. Scientists are expected to contact the principal investigator about their intentions to include any data from this project in a publication prior to publication of the primary report in order to ensure that their publication does not compete directly with planned publications (e.g. reserved analyses) of the principal investigators.

The principal investigator, listed on the relevant portal page, is the point of contact regarding publication plans. Reserved analyses include but are not limited to the identification and analysis of complete (whole genome) sets of genomic features such as genes and gene expression levels, gene families, gene clusters, regulatory elements, repeat structures, etc., and whole-genome comparisons of regions of evolutionary conservation or change. Studies of any type on the reserved data sets that are not in direct competition with those planned by the principal investigators may also be undertaken after agreement with the principal investigator. Interested parties must contact the principal investigator (or JGI if the PI cannot be reached) to discuss such possibilities.

If these data are used for publication, the following acknowledgment should be included: 'These sequence data were produced by the US Department of Energy Joint Genome Institute <http://www.jgi.doe.gov/> in collaboration with the user community.' We also request that you appropriately cite any JGI resources used for analysis (such as IMG, Phytozome or Mycocosm) and that you notify us upon publication.

The data may be freely downloaded and used by all who respect the restrictions in the previous paragraphs, and the same restrictions apply to internal use by JGI staff and affiliates. Prior to full data release, the assembly and raw sequence reads must not be redistributed or repackaged without permission from the JGI. When such permission has been obtained, any redistribution of the data during this period must carry this notice: 'The Joint Genome Institute provides these data in good faith, but makes no warranty, expressed or implied, nor assumes any legal liability or responsibility for any purpose for which the data are used.' and users should be directed to this release policy page. Once moved to unreserved status (after publication or two years after public data release), the data are freely available for any subsequent use, but the source must still be cited.