

A quick reference guide to metagenome sequencing at JGI

Alicia Clum and Emiley Eloë-Fadrosh

January 2022

The U.S. Department of Energy Joint Genome Institute is a DOE Office of Science User Facility that provides sequencing and analysis to scientists from all over the world studying carbon cycling, bioenergy and biogeochemistry. These services are primarily provided through several competitive grant calls: the annual Community Science Program (CSP) call, a twice annual CSP New Investigator call and the "Facilities Integrating Collaborations for User Science" (FICUS) calls. Further details on each can be found on JGI's [User Program Info page](#). Each of these calls places caps on the total amount of sequencing that can be requested. The purpose of this guide is to describe the standard products the Metagenome Program offers and how to estimate the amount of sequencing needed. Metatranscriptomes are not covered as part of this reference guide.

We offer several standard product types for metagenomes detailed in Table 1. In addition to our standard products, we have experimental sequencing products that we offer on a provisional basis. It is best to consult with JGI staff during the proposal writing process if you are interested in using experimental products. Many of the standard products are sequenced using Illumina NovaSeq sequencers. In 2021, we introduced an option to apply for long-read (PacBio) sequencing for up to 4 samples during calls. We encourage you to focus on sample quality and take advantage of [JGI guidance and suggestions on sample prep](#) as sample quality is a key factor in project success. In our experience working with tens of thousands of samples, the largest user controllable factor contributing to a successful project delivered on time, is high quality starting material. By contrast, samples which are low quality, contaminated, or low concentration typically produce inferior results and result in project delays.

Table 1: JGI Metagenome Program product offerings

Product	Sequencing mode	Read length (bp)	Library creation (amount)	Target depth (Gb)	Example sample types
Metagenome Minimal Draft	Illumina NovaSeq S4 paired end	150	KAPA HyperPrep (200 ng)	11	Bioreactor, hot springs
Metagenome Standard Draft	Illumina NovaSeq S4 paired end	150	KAPA HyperPrep (200 ng)	45	Ocean, soil, rhizosphere
Metagenome Improved Draft	PacBio Sequel IIe/CCS	Kbp	PacBio 6-10kb (2000 ng)	4	Samples to generate reference metagenome-assembled genomes (MAGs)
Metagenome Viral Standard Draft	Illumina NovaSeq S4 paired end	150	Swift 1S Plus (1 ng)	11	Samples expected to be mostly viral
Stable Isotope Probing (SIP) Metagenome Draft	Illumina NovaSeq S4 paired end	150	Nextera (1 ng)	5	Samples from a stable isotope incubation experiment
Cell Enrichment Metagenome Draft	Illumina NovaSeq S4 paired end	150	Nextera (1 ng)	2	Microcolony, microbial aggregate

Table 1: The modes and typical yields of standard sequencing products from the Metagenome Program. Some common environments are also listed in the table. The exact allocation of samples is done collaboratively with JGI staff after proposals have been accepted. This guide should help you propose a reasonable sample allocation for your project based on our normal workflows. More detailed information on submitting samples is available on at JGI's [sample submission page](#).

For a detailed analysis on how sequencing depth for common habitats impacts assembly and binning results using coverage estimates with the tool Nonpareil 3 [1] please see [this JGI technical note](#). For complex samples, we do not recommend sequencing below the Metagenome Standard Draft recommendation of 45 Gb.

In addition to the above metagenome sequencing products offered, the JGI provides analysis support for combined metagenome assemblies that computationally combine sequencing data from multiple samples where it makes sense biologically. JGI staff have been able to assemble up to 25 terabytes of metagenome data using a distributed assembler called MetaHipMer. Combined metagenome assemblies can increase assembly contiguity and improve the number of metagenome-assembled genomes (MAGs) recovered, but may produce poor results for individual genomes when introducing strain heterogeneity[2]. Please contact your JGI Project Manager if you are interested in this analysis offering.

References

- [1] Rodriguez-R et al. “Nonpareil 3: Fast Estimation of Metagenomic Coverage and Sequence Diversity”. In: *mSystems* (June 2018). URL: <https://journals.asm.org/doi/10.1128/mSystems.00039-18>.
- [2] Steven Hofmeyr et al. “Terabase-scale metagenome coassembly with MetaHipMer”. In: *Scientific Reports* (Dec. 2020). URL: <https://www.nature.com/articles/s41598-020-67416-5>.

Frequently Asked Questions

What is the difference between a standard Illumina library and Nextera? Our standard Illumina library requires 100-200 nanograms of material depending on if the library is made in a tube or plate, and does not use amplification. We use Nextera as an alternative protocol for low biomass samples which requires as little as 1 nanogram of input material and requires PCR amplification. For Metagenome Standard Drafts and Metagenome Minimal Drafts we use a default of 9 cycles of PCR for Nextera libraries with a maximum of 20 cycles. Stable Isotope Probing (SIP) Metagenome Drafts and Cell Enrichment Metagenome Drafts use 12 cycles of PCR.

What is the difference between metagenome and viral metagenome sequencing?

For projects where we expect a significant viral component we use a different library preparation which captures both single and double stranded DNA. These samples are also typically less complex so are also sequenced to a shallower depth.

What is SIP metagenome sequencing?

Isotopically labelled (ie ^{13}C) nutrients are added to samples which are subsequently fractionated using a density gradient. A metagenome library is prepared and sequenced for each fraction. This experimental approach is used to identify active groups of organisms within a community. The fractionation, library preparation, sequencing, and quantitative analyses are performed at the JGI as a new capability for users. For more information, please refer to our [webinar](#).

What if I am expecting prokaryotes and eukaryotes in my metagenome samples?

The JGI metagenome assembly workflow is agnostic to what organism’s genomes are ultimately assembled into contigs, however the gene calling and annotation components are specific to prokaryotes. The JGI has a limited set of reference eukaryote genomes that are used in the reference database for contig-level taxonomic assignments that are available through the Integrated Microbial Genomes & Microbiomes (IMG/M). Please contact your JGI Project Manager if you interested in additional support for identification and analysis of eukaryotes in your metagenome data.

What if I have synthetic spike-ins I want to use? The JGI strongly discourages the submission of samples that include internal standards, either synthetic or known template material, to avoid unwanted propagation of these sequences in public repositories. While the JGI has implemented robust quality-control measures to ensure the metagenome and metatranscriptome sequence data produced is high quality, we do not currently have processes in place to detect and remove an expanded set of synthetic or known template material that can be applied across the diverse set of environmental samples the JGI handles. If your study design or planned sample submission requires the use of custom internal standards, please be sure to discuss with your JGI Project Manager to determine a suitable solution.