

Downloading IMG sequence and annotation data

[General Requirements.](#)

[Downloading a single dataset.](#)

[Non-JGI Generated Datasets.](#)

[JGI-Generated datasets.](#)

[“IMG Data” directory content.](#)

[Datasets not available for download from the JGI Genome Portal.](#)

[Batch download of multiple datasets.](#)

[Opening downloaded tarballs.](#)

[Special case downloads \(IMG/VR, IMG collections, etc.\).](#)

General Requirements.

Note that downloading of any data via IMG is available only in IMG/MER, i. e. it requires signing in with your IMG account. You can create an IMG account following the “Login / Sign-on” link on the top of IMG Home Page as shown below and follow the instructions therein.

JGI **IMG/M** INTEGRATED MICROBIAL GENOMES & MICROBIOMES

JGI HOME CONTACT US **LOGIN / SIGN-ON**

Home IMG/M Find Genomes Find Genes Find Functions Compare Genomes OMICS My IMG Help

COVID-19 Notice: Due to the emerging COVID-19 pandemic, JGI will not be accepting or processing any physical samples because of reduced onsite staffing until further notice. **IMG** is still accepting bioinformatic data sets via [IMG Submission Site](#)

NEW IMG Webinars [Read more...](#)

Integrated Microbial Genomes and Microbiomes

The **mission** of the Integrated Microbial Genomes & Microbiomes (IMG/M) system is to support the annotation, analysis and distribution of microbial genome and microbiome datasets sequenced at DOE's Joint Genome Institute (JGI).

IMG/M is also open to scientists worldwide for the annotation, analysis, and distribution of their own genome and microbiome datasets, as long as they agree with the IMG/M data release policy and follow the metadata requirements for integrating data into IMG/M (see IMG/M submission site).

If you use **IMG** web resources or data to assist in research publications or proposals, please cite [Chen et al., 2018 \(PMID: 30289528\)](#).

[Review Data Usage Policy](#) [Submit Your Data](#)

Announcements

New Gene Search tool with Advanced Search capabilities. See [Gene Search User Guide](#).

IMG Content

Genes	60,410,857,733
Bases	18,322,913,393,420
Scaffolds	55,035,196,475

[IMG Statistics](#)

IMG & GOLD Citations

Downloading a single dataset.

Download links are provided on the respective Genome Details and Metagenome Details pages as shown below.

Bradyrhizobium symbiodeficiens 85S1MB

[Add to Genome Cart](#) [Browse Genome](#) [BLAST Genome](#) [Download Data](#)

About Genome

- [Overview](#)
- [Statistics](#)
- [Genes](#)

Overview

Study Name (Proposal Name)	Bradyrhizobium sp. 85S1MB genome sequencing
Organism Name	Bradyrhizobium symbiodeficiens 85S1MB
Taxon ID	2849076700
IMG Submission ID	225233
NCBI Taxon ID	1404367
GOLD ID in IMG Database	Study ID: Gs0143638 Project ID: Gp0443445
GOLD Analysis Project Id	Ga0398464
GOLD Analysis Project Type	Genome Analysis (Isolate)
Submission Type	Primary
JGI Analysis Project Type	Genome Analysis (Isolate)
External Links	JGI Portal
Lineage	Bacteria ; Proteobacteria ; Alphaproteobacteria ; Rhizobiales ; Bradyrhizobiaceae ; Bradyrhizobium ; Bradyrhizobium symbiodeficiens
Sequencing Status	Finished
Sequencing Center	Ottawa Research and Development Center
IMG Release/Pipeline Version	IMG Annotation Pipeline v.5.0.9
Comment	

The link will redirect you to the JGI Genome Portal page. Depending on whether the dataset has been generated by the JGI or not, the page will look differently.

Non-JGI Generated Datasets.

Below is an example of a Genome Portal page for non-JGI sequenced dataset. Note the Funding information saying that the project was not sequenced at the JGI. You should click on the "Download" tab.

JGI  **Genome Portal** JGI HOME GENOME PORTAL LOGOUT NATALIA.IVANOVA(NNIVANOVA@LBL.GOV) (0 items)

Bradyrhizobium symbiodeficiens 85S1MB

INFO **DOWNLOAD** IMG HELP!

 On April 22, 2020 the JGI computer systems will be undergoing maintenance and access to certain files and tools will be affected. Sorry for the inconvenience.

Due to the emerging COVID-19 pandemic, JGI will not be accepting or processing any samples because of reduced onsite staffing until further notice. Please [read more](#)

Portal name: Bradyrhizobium symbiodeficiens 85S1MB
Project PI: [Jon Bertsch](#)
Release Date: 2020-01-19

Name
 Genus/species/strain/isolate: Bradyrhizobium/Bradyrhizobium symbiodeficiens//

Data Submission
 NCBI Tax ID: [1404367](#)

Contacts
 JGI: microbial@lists.jgi-psf.org

Funding
 This project was not sequenced at the JGI.

Groups

This link will lead you to the page below.

Bradyrhizobium symbiodeficiens 85S1MB

INFO **DOWNLOAD** IMG HELP!

 On April 22, 2020 the JGI computer systems will be undergoing maintenance and access to certain files and tools will be affected. Sorry for the inconvenience.

Due to the emerging COVID-19 pandemic, JGI will not be accepting or processing any samples because of reduced onsite staffing until further notice. Please [read more](#)

Portal name: Bradyrhizobium symbiodeficiens 85S1MB
Project PI: [Jon Bertsch](#)
Release Date: 2020-01-19

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:

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

Even though this is not a JGI-generated dataset and no notification of the PI is required, you still have to agree with the JGI Data Utilization policy by clicking on the "Agree" button, which will open a page like the one below.

 Open Downloads as XML - the document tree is shown within XML file

 Please be aware that **restore requests** sourced **from tape** and **of more than 1 file** (including Globus downloads) will be delayed until 8PM.

Download via Globus (v.2)

Download Selected Files ExpandAll CollapseAll Rescan Organize By File Type

Please keep in mind that downloading **tape files** () can take a few minutes.

- IMG_2849076700
 - IMG Data**
 - Metagenome Report Tables

You can select an entire directory or open the tree to navigate to the specific file as shown below.

Download Selected Files ExpandAll CollapseAll Rescan Organize By File Type

Please keep in mind that downloading **tape files** () can take a few minutes. Legend:   

- IMG_2849076700
 - IMG Data
 - Bradyrhizobium symbiodeficiens* 85S1MB: 2849076700.tar.gz 6 MB; Fri Jan 10 13:13:02 PST 2020 Cache Expires: Sat May 02 09:47:59 PDT 2020  
 - Bradyrhizobium symbiodeficiens* 85S1MB: Ga0398464_annotation_config.yaml 2 KB; Tue Dec 17 14:03:03 PST 2019  
 - Bradyrhizobium symbiodeficiens* 85S1MB: Ga0398464_cath_funfam.gff 1 MB; Tue Dec 17 14:03:03 PST 2019  
 - Bradyrhizobium symbiodeficiens* 85S1MB: Ga0398464_cleavage_sites.gff 92 KB; Tue Dec 17 14:03:03 PST 2019  
 - Bradyrhizobium symbiodeficiens* 85S1MB: Ga0398464_cog.gff 1 MB; Tue Dec 17 14:03:02 PST 2019  
 - Bradyrhizobium symbiodeficiens* 85S1MB: Ga0398464_combined_logs.txt 16 KB; Tue Dec 17 14:03:04 PST 2019  
 - Bradyrhizobium symbiodeficiens* 85S1MB: Ga0398464_contig_names_mapping.tsv 24 bytes; Tue Dec 17 14:02:59 PST 2019  
 - Bradyrhizobium symbiodeficiens* 85S1MB: Ga0398464_contigs.fna 6 MB; Tue Dec 17 14:03:03 PST 2019  
 - Bradyrhizobium symbiodeficiens* 85S1MB: Ga0398464_ec.tsv 139 KB; Tue Dec 17 14:03:01 PST 2019  
 - Bradyrhizobium symbiodeficiens* 85S1MB: Ga0398464_functional_annotation.gff 2 MB; Tue Dec 17 14:03:03 PST 2019  
 - Bradyrhizobium symbiodeficiens* 85S1MB: Ga0398464_gene_phylogeny.tsv 1 MB; Tue Dec 17 14:03:01 PST 2019  
 - Bradyrhizobium symbiodeficiens* 85S1MB: Ga0398464_genemark.gff 752 KB; Tue Dec 17 14:03:00 PST 2019  
 - Bradyrhizobium symbiodeficiens* 85S1MB: Ga0398464_genemark_genes.fna 6 MB; Tue Dec 17 14:03:00 PST 2019  
 - Bradyrhizobium symbiodeficiens* 85S1MB: Ga0398464_genemark_proteins.faa 2 MB; Tue Dec 17 14:03:00 PST 2019  
 - Bradyrhizobium symbiodeficiens* 85S1MB: Ga0398464_genes.fna 6 MB; Tue Dec 17 14:03:00 PST 2019  
 - Bradyrhizobium symbiodeficiens* 85S1MB: Ga0398464_imgap.info 436 bytes; Tue Dec 17 14:03:03 PST 2019  
 - Bradyrhizobium symbiodeficiens* 85S1MB: Ga0398464_ko.tsv 247 KB; Tue Dec 17 14:03:01 PST 2019  
 - Bradyrhizobium symbiodeficiens* 85S1MB: Ga0398464_ko_ec.gff 848 KB; Tue Dec 17 14:03:01 PST 2019  
 - Bradyrhizobium symbiodeficiens* 85S1MB: Ga0398464_pfam.gff 1 MB; Tue Dec 17 14:03:02 PST 2019  

Then click on “Download Selected Files” to download the data.

JGI-Generated datasets.

The link on the dataset Details page will lead to a JGI Genome Portal page, which looks like the one below. Note that a PI and proposal name is provided, as well as other relevant metadata.

INFO **DOWNLOAD** STATUS REPORT HELP!

On April 22, 2020 the JGI computer systems will be undergoing maintenance and access to certain files and tools will be affected. Sorry for the inconvenience.

Due to the emerging COVID-19 pandemic, JGI will not be accepting or processing any samples because of reduced onsite staffing until further notice. Please [read more](#)

Project name: Bradyrhizobium centrosematis SEMIA 431 (*Project ID: 1244128*)
Product: Microbial Annotated Minimal Draft Isolate
Proposal Name: Core and pangenomes of soil and plant-associated prokaryotes (*Proposal ID: [502937](#)*)
Project PI: Camila Gazolla Volpiano
User Program: CSP
Program Year: 2017
Scientific Program: Microbial
Genome Portal:  2842045827
Related Projects: FD 1244128; SP [1244323](#); AP [1244129](#)
Release Date: 2019-03-14

JGI Data Utilization policy

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

For JGI-generated data you need to follow the JGI Data Usage Policy and you may need to request the PI's permission to use the data. After agreeing to follow the JGI Data Usage Policy, a page similar to the screenshot below will appear. Note that for the JGI-generated datasets a lot more different directories with different data types are available. IMG sequence and annotation data is stored in the directory "IMG Data" (highlighted). Please refer to the JGI Genome Portal documentation and Help for information on other types of data available for JGI-generated datasets.

Bradyrhizobium centrosematis SEMIA 431

INFO **DOWNLOAD** STATUS REPORT HELP!

On April 22, 2020 the JGI computer systems will be undergoing maintenance and access to certain files and tools will be affected. Sorry for the inconvenience.

Due to the emerging COVID-19 pandemic, JGI will not be accepting or processing any samples because of reduced onsite staffing until further notice. Please [read more](#)

 Open Downloads as XML - the document tree is shown within XML file

 Please be aware that restore requests sourced from tape and of more than 1 file (including Globus downloads) will be delayed until 8PM.

Download via Globus (v.2)

[Download Selected Files](#) [ExpandAll](#) [CollapseAll](#) [Rescan](#) Organize By File Type

Please keep in mind that downloading tape files () can take a few minutes. Legend: "a", "c", "d", "l", "n", "p", "q", "r", "s", "t", "h"   

- BracenSEMIA431_FD
 - Filtered Raw Data
 - IMG Data**
 - Metagenome Report Tables
 - QC and Genome Assembly
 - Raw Data
 - Sequencing QC Reports

Again, you can select some or all directories, or open the "IMG Data" directory to select individual files, then click on "Download Selected Files" to start downloading.

Download via Globus (v.2)

Download Selected Files ExpandAll CollapseAll Rescan Organize By File Type

Please keep in mind that downloading tape files () can take a few minutes. Legend: "a", "c", "d", "f", "h", "n", "p", "q", "r", "s", "t", "h" MDS

- BracenSEMAI431_FD
 - Filtered Raw Data
 - IMG Data
 - Bradyrhizobium centrosematis SEMIA 431: Project: 1244129, 2842045827.tar.gz 7 MB; Thu Nov 14 11:17:40 PST 2019 Cache Expires: Sat May 02 09:47:59 PDT 2020
 - Bradyrhizobium centrosematis SEMIA 431: Project: 1244129, Ga0394309_annotation_config.yaml 2 KB; Fri Sep 13 20:27:34 PDT 2019
 - Bradyrhizobium centrosematis SEMIA 431: Project: 1244129, Ga0394309_cath_funfam.gff 2 MB; Sun Sep 15 20:17:50 PDT 2019
 - Bradyrhizobium centrosematis SEMIA 431: Project: 1244129, Ga0394309_cleavage_sites.gff 105 KB; Sun Sep 15 20:20:27 PDT 2019
 - Bradyrhizobium centrosematis SEMIA 431: Project: 1244129, Ga0394309_cog.gff 1 MB; Sun Sep 15 17:12:20 PDT 2019
 - Bradyrhizobium centrosematis SEMIA 431: Project: 1244129, Ga0394309_combined_logs.txt 15 KB; Mon Sep 16 16:39:25 PDT 2019
 - Bradyrhizobium centrosematis SEMIA 431: Project: 1244129, Ga0394309_contig_names_mapping.tsv 486 bytes; Sun Sep 15 15:12:19 PDT 2019
 - Bradyrhizobium centrosematis SEMIA 431: Project: 1244129, Ga0394309_contigs.fna 7 MB; Sun Sep 15 15:12:19 PDT 2019 Cache Expires: Sun Jun 21 16:35:51 PDT 2020
 - Bradyrhizobium centrosematis SEMIA 431: Project: 1244129, Ga0394309_ec.tsv 149 KB; Sun Sep 15 17:02:59 PDT 2019
 - Bradyrhizobium centrosematis SEMIA 431: Project: 1244129, Ga0394309_functional_annotation.gff 2 MB; Sun Sep 15 20:24:16 PDT 2019 Cache Expires: Sun Jun 21 16:35:51 PDT 2020
 - Bradyrhizobium centrosematis SEMIA 431: Project: 1244129, Ga0394309_gene_phylogeny.tsv 1 MB; Sun Sep 15 17:02:59 PDT 2019
 - Bradyrhizobium centrosematis SEMIA 431: Project: 1244129, Ga0394309_genemark.gff 842 KB; Sun Sep 15 16:19:53 PDT 2019
 - Bradyrhizobium centrosematis SEMIA 431: Project: 1244129, Ga0394309_genemark_genes.fna 7 MB; Sun Sep 15 16:19:54 PDT 2019
 - Bradyrhizobium centrosematis SEMIA 431: Project: 1244129, Ga0394309_genemark_proteins.faa 2 MB; Sun Sep 15 16:19:54 PDT 2019
 - Bradyrhizobium centrosematis SEMIA 431: Project: 1244129, Ga0394309_genes.fna 7 MB; Sun Sep 15 16:19:55 PDT 2019
 - Bradyrhizobium centrosematis SEMIA 431: Project: 1244129, Ga0394309_imgap.info 436 bytes; Sun Sep 15 20:24:16 PDT 2019

“IMG Data” directory content.

Although in general this directory contains all IMG sequence and annotation files, depending on the version of the IMG pipeline, which was used to annotate this particular dataset, the content of “IMG Data” directory is different. The information about the version of the pipeline used to annotate the dataset of interest can be found on the respective Details page as shown below.

Study Name (Proposal Name)	Genomic Encyclopedia of Type Strains, Phase IV (KMG-V): Genome sequencing to study the core and pangenomes of soil and plant-associated prokaryotes
Organism Name	Bradyrhizobium centrosematis SEMIA 431
Taxon ID	2842045827
IMG Submission ID	221602
NCBI Taxon ID	1300039
GOLD ID in IMG Database	Study ID: Gs0129091 Project ID: Gp0436719
GOLD Analysis Project Id	Ga0394309
GOLD Analysis Project Type	Genome Analysis (Isolate)
Submission Type	Primary
JGI Analysis Product Name	Microbial Minimal Draft Isolate Genome
JGI Analysis Project Type	Genome Analysis (Isolate)
External Links	JGI Portal
Lineage	Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Bradyrhizobiales; Bradyrhizobiaceae; Bradyrhizobium; Bradyrhizobium centrosematis
Sequencing Status	Permanent Draft
Sequencing Center	DOE Joint Genome Institute (JGI)
IMG Release/Pipeline Version	IMG Annotation Pipeline v.5.0.3
Comment	
Release Date	2019-11-12
Add Date	2019-11-12
Modified Date	
High Quality	Yes
Is Public	Yes
Genome Completeness %	
Annotation Method	Annotation System: IMG (IMGAP v5.0.3) Gene calling program: GeneMark.hmm-2 v1.05
Project Information	
Culture Type	Isolate
Cultured	Yes
Ecosystem	Host-associated

The main differences are between the genomes and metagenomes annotated by the IMG pipeline v.5+ and genomes and metagenomes annotated by the earlier versions of the pipeline. The details for all of them can be found in the IMG Help menu as shown below.

Home IMG/MER Find Genomes Find Genes Find Functions Compare Genomes OMICS Workspace My IMG **Help**

Due to the emerging COVID-19 pandemic, JGI will not be accepting or processing any physical samples because of reduced onsite staffing until further notice. **IMG is still accepting bioinformatic data sets via [IMG Submission Site](#)**

NEW IMG Webinars [Read more...](#)

About IMG

- **IMG News**
- [IMG Mission](#)
- [FAQ](#) [Submission FAQ](#)
- [Related Links](#)
- [Credits](#)
- [Education](#)
- [Citation \(PMID: 30289528\)](#)
- [IMG System Requirements](#)
- [MGM Workshop](#)
- [IMG Webinar Sign Up](#) [Past IMG Webinar Videos](#)
- [JGI/IMG Communication Policy](#)

Documents & User Guides

- [Pipeline V.5.0.0](#) pre-version 5 genome annotation
- [Genome Annotation SOP](#)
- [Metagenome SOP](#) pre-version 5 metagenome annotation
- [Metagenome Replacement Notice](#)
- [Single Cell Data Decontamination](#)
- [IMG M Addendum](#)
- [IMG ER Tutorial](#)
- [IMG Virus Help](#)
- Site Maps:
 - [Menu](#) [Analysis Cart](#) [Components](#)
- User Manual, see above site map links
- [User Interface Map](#)

Download **Contact Us & Other**

File names in the directories generally correspond to the annotation types described in these documents. In addition, a tarball of all IMG annotations (a file named .tar.gz) is provided, which includes a README file describing file contents and column definitions.

Datasets not available for download from the JGI Genome Portal.

Many older genomes in IMG were not re-annotated with the IMG annotation pipeline, i. e. their annotations are generally the same as those available in GenBank. For these genomes the “Download” link goes to NCBI page like the one shown below.

NCBI Taxonomy Browser

Search for _____ as complete name lock Go Clear

Display 3 levels using filter: none

Bradyrhizobium diazoefficiens USDA 110

Taxonomy ID: 224911 (for references in articles please use NCBI:txid224911)

current name
Bradyrhizobium diazoefficiens USDA 110

NCBI BLAST name: **a-proteobacteria**
Rank: **no rank**
Genetic code: [Translation table 11 \(Bacterial, Archaeal and Plant Plastid\)](#)
Other names:
heterotypic synonym
Bradyrhizobium japonicum USDA 110

Lineage(full)
cellular organisms; Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Bradyrhizobiaceae; Bradyrhizobium; Bradyrhizobium diazoefficiens

Entrez records		
Database name	Direct links	Links from type
Nucleotide	9,468	35
Protein	27,957	-
Structure	41	-
Genome	1	-
Popset	7	-
GEO Datasets	222	-
PubMed Central	30	-
Gene	17,686	-
SRA Experiments	14	-
GEO Profiles	94,840	-
Identical Protein Groups	11,831	-
Bio Project	26	-
Bio Sample	18	4
Bio Systems	253	1
Assembly	2	-

In some cases Genome Details pages include the “External Links” section with link-outs to GenBank or RefSeq pages via contig/scaffold-specific accessions, like shown below.

Bradyrhizobium japonicum USDA 110

[Add to Genome Cart](#) [Browse Genome](#) [BLAST Genome](#) [Download Data](#)

About Genome

- [Overview](#)
- [Statistics](#)
- [Genes](#)

Overview

Study Name (Proposal Name)	Bradyrhizobium japonicum USDA 110
Organism Name	Bradyrhizobium japonicum USDA 110
Taxon ID	637000038
NCBI Taxon ID	224911
GOLD ID in IMG Database	Study ID: Gs0002473 Project ID: Gp0000660
GOLD Analysis Project Id	Ga0027865
GOLD Analysis Project Type	Genome Analysis (Isolate)
Submission Type	Primary
JGI Analysis Project Type	Genome Analysis (Isolate)
External Links	PUBMED:1259727 ; NCBI/RefSeq:NC_004463 ; NCBI/Genbank:224911 ; GreenGenes:137054
Lineage	Bacteria ; Proteobacteria ; Alphaproteobacteria ; Rhizobiales ; Bradyrhizobiaceae ; Bradyrhizobium ; Bradyrhizobium diazoefficiens
Sequencing Status	Finished
Sequencing Center	Kazusa DNA Research Institute
IMG Release/Pipeline Version	IMG/W 2.0
Comment	Genome replaced by a new version processed using IMG's annotation pipelines. New IMG Identifier: 2623620631
Release Date	2006-12-01
Add Date	2006-10-02

You can download the data following the links on these pages.

For batch downloads via Genome Cart a link in the email will include information about the genomes that could not be included in the download bundle because the data is not available in the JGI Genome Portal. An example of such a message is shown below. You can use IMG

taxon oids provided on this page to retrieve the genomes from NCBI using one of the ways described above.

Download request completed.

[Download \(668MB, tar file\)](#)

The data will be stored on the server for 14 days.

Data for the following genomes are not available:
637000038, 640427103, 640427104

Note that for some private submissions that were later released by the submitter (i. e. non-JGI generated data that wasn't loaded from one of NCBI sources) the submitter opted not to provide the data for download via the JGI Genome Portal. For such datasets no data download is available.

Batch download of multiple datasets.

Batch downloads of multiple datasets are enabled via IMG Genome Cart. The count of genomes in your Genome Cart is displayed above the IMG main menu as shown below; clicking on this link will lead you to the Genome Cart. Use IMG Genome Search and other tools to add datasets to your Genome Cart and table configuration and filtering options therein to refine your genome selection.

Quick Genome Search: Hi Natalia Ivanova (JGI SSO) 24825

My Analysis Carts**: **85 Genomes** | 0 Scaffolds | 0 Functions | 0 Genes | 2 Genome Search History | 0 Gene Search History

Home IMG/MER Find Genomes Find Genes Find Functions Compare Genomes OMICS Workspa

NEW IMG Webinars [Read more...](#)

IMG Content	Datasets	JGI	All
Bacteria	19813	95238	
Archaea	892	2965	
Eukarya	371	751	
Plasmids	3	1208	
Viruses	1216	9804	
Genome Fragments	0	89	
Metagenome	13073	24611	
Cell Enrichments	2302	2352	
Single Particle Sorts	5433	5806	
Metatranscriptome	4069	6305	
Total Datasets		149192	
My Private Datasets		22853	
Last Datasets Added On:			

The **Integrated Microbial Genomes (IMG)** system serves as a community resource for analysis and annotation of genome and metagenome datasets in a comprehensive comparative context. The **IMG data warehouse** integrates genome and metagenome datasets provided by IMG users with a comprehensive set of publicly available genome and metagenome datasets.

IMG/MER provides users with tools ([IMG/MER UI Map](#)) for analyzing their private (password protected access) genome datasets and/or metagenome datasets in the context of all public (free access) genome and metagenome datasets in IMG. ([Nucleic Acids Research, January 2019](#))

If you use **IMG web resources or data to assist in research publications or proposals** please cite [Chen et al., 2019 \(PMID: 30289528\)](#).

In the Genome Cart you can select some or all of the datasets, go to the tab “Upload & Export & Save” (highlighted below) and click on “Download Genomes” button.

Genome Cart

Only a maximum of 20000 genomes can be in cart.

85 genome(s) in cart

Genomes in Cart **Upload & Export & Save**

[Group Genome Cart by Phyla](#)

hint: Scaffolds will not be added into cart for very large genomes. Only scaffolds (assembled data only) of selected MER-FS genomes can be added into cart.

85 of 85 rows selected

Filter column: Domain Filter text:

Page 1 of 1 << first < prev 1 next > last >> All

Column Selector

Select	Domain	Sequencing Status	Study Name	Genome Name / Sample Name	Sequencing Center	IMG Genome ID	Genome Size * assembled	Gene Count * assembled
<input checked="" type="checkbox"/>	Bacteria	Permanent Draft	Genomic Encyclopedia of Type Strains, Phase IV (KMG-V): Genome sequencing to study the core and pan-genomes of	Bradyrhizobium centrosematis SEMIA 424	DOE Joint Genome Institute (JGI)	2842038055	8002051	7771

This link will redirect your request to the JGI Genome Portal, which will generate a download bundle (a .tar file) of IMG annotation and sequence tarballs (the same that can be downloaded

from individual dataset portal pages). Follow the instructions on this page. When your download bundle is generated, you will receive an email with the link to the download page.

Opening downloaded tarballs.

In many cases the downloaded file is a tarball of multiple files, either compressed (extension .tar.gz) or uncompressed (extension .tar). Unix users have native utilities that they can use to extract this kind of an archive. Windows users may need to download [7-zip](#) or similar uncompression utility to extract files from this archive.

7-zip is available for download at <http://www.7-zip.org/> .

Contents of the download archive may vary based on the availability of data in IMG, but in general the following files are made accessible:

- Nucleotide sequence (assembled) of a genome or metagenome in multi-fasta format
- README.txt file describing the format for each file
- Gene sequences in multi-fasta format
- Aminoacid sequence file of protein-coding genes in a multi-fasta format
- Intergenic regions in multi-fasta format
- Gene Information file in GFF3 format (Strict conformance is not guaranteed, esp. in type and attributes fields)
- COG hits in tab-delimited format
- KOG hits in tab-delimited format
- Pfam hits in tab-delimited format
- TIGRFam hits in tab-delimited format
- InterPro hits or hits to a subset of InterPro databases (CATH-FunFam, SFam, SMART) in tab-delimited format
- Transmembrane helices in tab-delimited format
- KEGG Orthology (KO) and EC annotation in tab-delimited format
- Signal peptide annotation in tab-delimited format
- External references information in tab-delimited format

Special case downloads (IMG/VR, IMG collections, etc.).

In addition to downloads of individual datasets and custom-selected sets of genomes and/or metagenomes, IMG provides downloads of pre-packaged large datasets containing viral contigs from IMG/VR, and other IMG collections. Below is an example of a special case download of IMG/VR data.

IMG Viral Content

Viral Datasets	
Isolate Viruses (VIGs)	8392
UVIGs (Uncultivated Viral Genomes)	693 1569

Viral Operational Taxonomic Units (vOTUs)	
Viral Clusters	1399
Viral Singletons	0

With Host	
Isolate Viruses (VIGs)	6707
UVIGs: Spacer Hit	288 300
UVIGs: Total	693 1569

Quality	
Metagenomes UVIGs	
Concatemer_artifact	5 5
Unsure (Ns)	1 1
High-quality	124 151
Genome fragment	621 1412

Viral/Spacer BLAST

Download VPF Models
Download IMG/VR Database

The **IMG/VR** system (<http://nar.oxfordjournals.org/content/early/2016/10/30/nar.gkw1030>) serves as a starting point for the sequence analysis of viral fragments derived from metagenomic samples. Virus detection methods and host assignment approaches in IMG/VR are fully described by Paez-Espino et al. in *Nature*, 2016 "Uncovering Earth's virome" and in *Nature Protocols*, 2017 "Nontargeted virus sequence discovery pipeline and virus clustering for metagenomic data".

Ecosystems ?

[Show Human Body Sites](#)

739759 viral scaffolds
Some projects maybe rejected via Geo Maps because of bad location coordinates.
Map pins represent location counts. Some pins may have multiple genomes.
Map pins are grouped into [clusters](#) and clusters themselves into larger clusters.

hint: For any given genome at a location on the map, you may access the [list of scaffolds](#) that belong to a virus by clicking on a map pin and selecting the [count](#) next to the genome of interest for that location. The [total count](#) of viral scaffolds for a location is displayed in the label and tooltip of a map pin e.g. Arctic Ocean [3].

Ecosystem:

A link to Hungate collection portal:

<https://genome.jgi.doe.gov/portal/HungateCollection/HungateCollection.info.html>

A link to a collection of 1003 GEBA genomes:

<https://genome.jgi.doe.gov/portal/geba1003/geba1003.info.html>