

JGI resources

Yanbin Yin
Spring 2013

Homework 1

- SRA dataset
- Display options

Homework assignment 4

1. Explore the fungal genomics database of JGI
2. Find how many *Aspergillus* genomes are sequenced
3. Download the predicted protein sequence data set of all the *Aspergillus* genomes (choose Annotation -> Filtered models -> Proteins) and upload the files to glu
4. Create a folder called *aspergillus* in your `home@glu` and move all the files to that folder (refer to page 5 and 9 in <http://cys.bios.niu.edu/yyin/teach/PBB/Yin-chapter1.pdf>)

Write a report (in **word or ppt**) to include all the operations and screen shots including the command lines used in step 4.

Due on Feb 26 (send by email)

Office hour:

Tue, Thu and Fri 2-4pm, MO325A

Or email: yyin@niu.edu

Outline

- Intro to JGI genome resources
 - Plant genomes
 - Prokaryotic genomes
 - Fungal genomes
 - Metagenomes
- Hands on practice
- Intro to comparative genomics

The **mission** of the U.S. Department of Energy (DOE) Joint Genome Institute (JGI) is to **advance genomics** in support of the DOE missions related to **clean energy** generation and **environmental** characterization and cleanup.

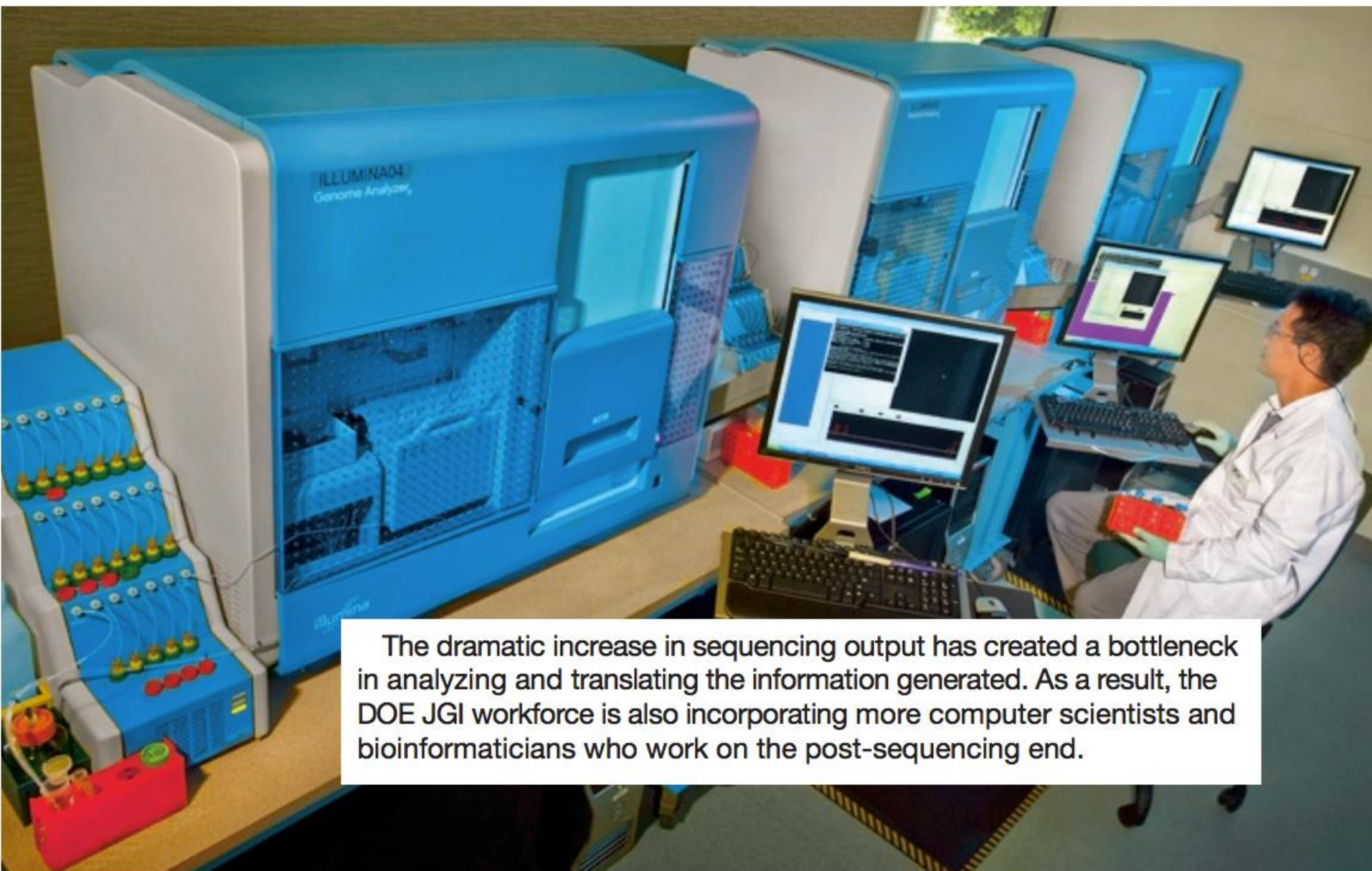
Supported by the DOE Office of Science, the DOE JGI unites the expertise of five **national laboratories**—Lawrence Berkeley, Lawrence Livermore, Los Alamos, Oak Ridge, and Pacific Northwest—along with the HudsonAlpha Institute for Biotechnology.

Located in Walnut Creek, California, the DOE JGI is operated by the University of California for the U.S. Department of Energy and the facility provides integrated **high-throughput sequencing and computational analysis** that enable systems-based scientific approaches to these challenges

<http://www.jgi.doe.gov/News/>

created in 1997





The dramatic increase in sequencing output has created a bottleneck in analyzing and translating the information generated. As a result, the DOE JGI workforce is also incorporating more computer scientists and bioinformaticians who work on the post-sequencing end.

DOE Science Drivers

Bioenergy, Carbon Cycling, Biogeochemistry

Experimental Data Generation

High-Throughput DNA Sequencing

DNA Synthesis

Experimental Functional Annotation

Single-Cell Genomics

Sample Preparation

Biological Data Interpretation

Assembling Sequence Data

Functional Discovery and Annotation

High Performance Computing-Based Annotation

Multidimensional Data Integration

User Interactions

Community Building and Coordination

Interactive Data Platforms

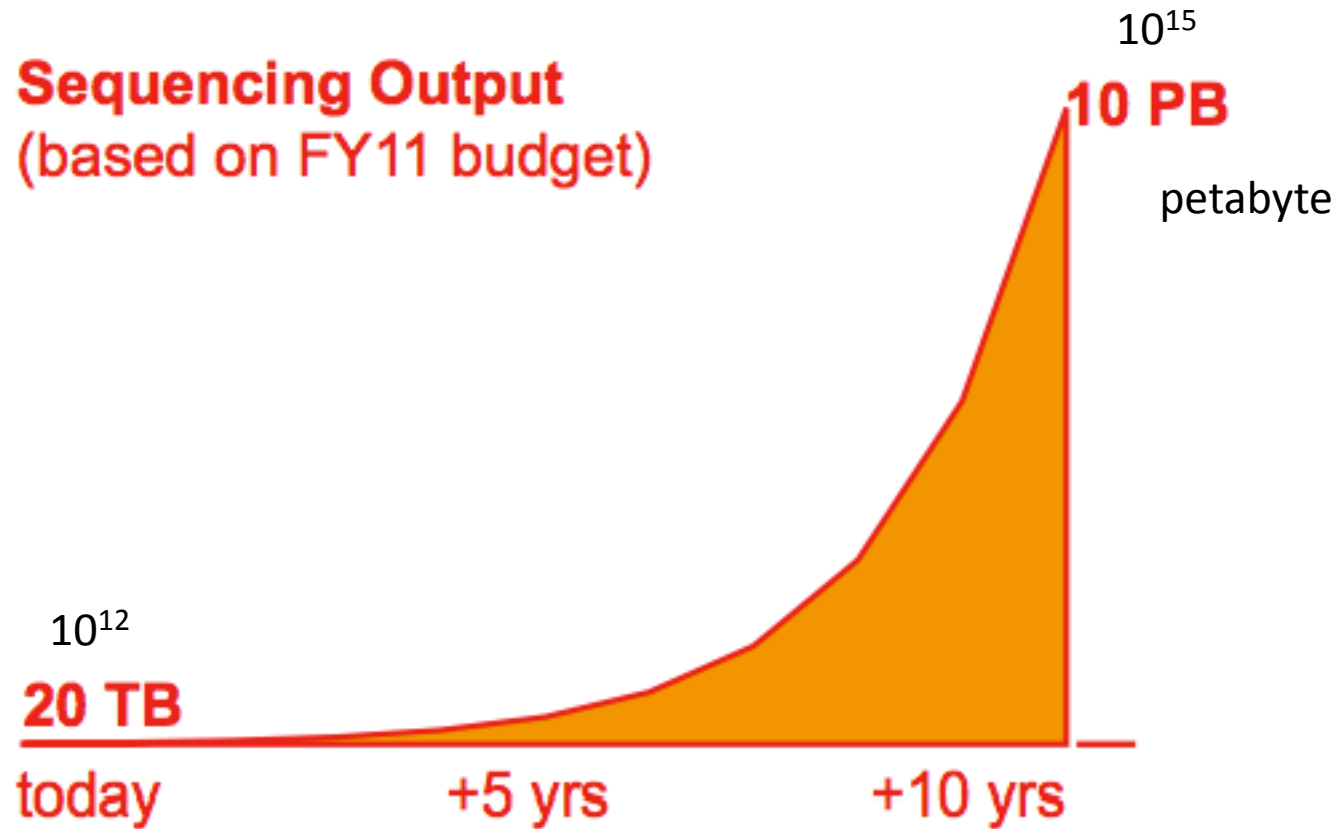
Standards and Support of Data Exchange

Onsite User Interactions

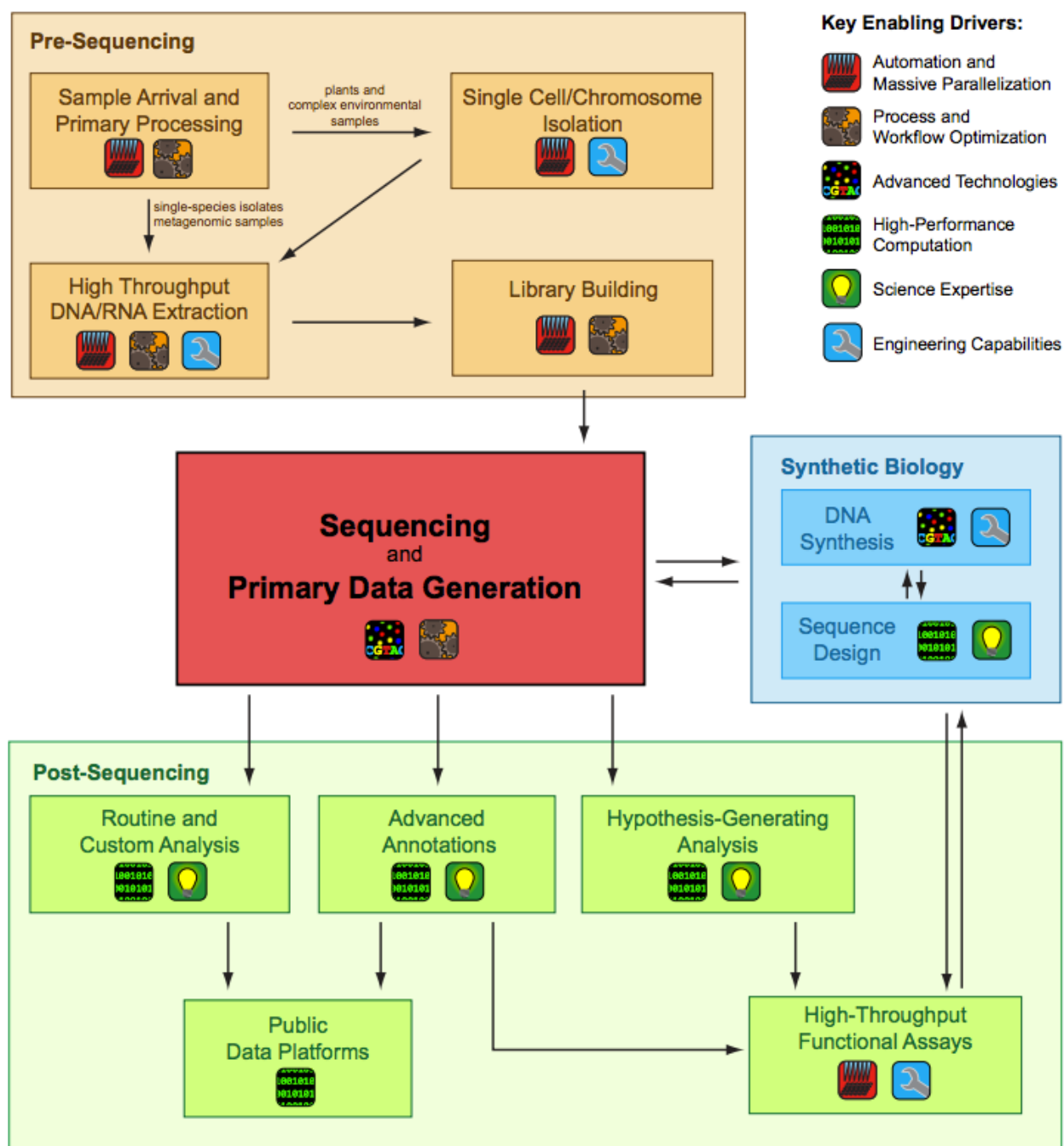
User Training

Employing an evolving set of cutting-edge experimental and computational technologies, the DOE JGI next-generation genome science user facility will empower users to perform studies at a scale and complexity far exceeding the capabilities of any individual laboratory.

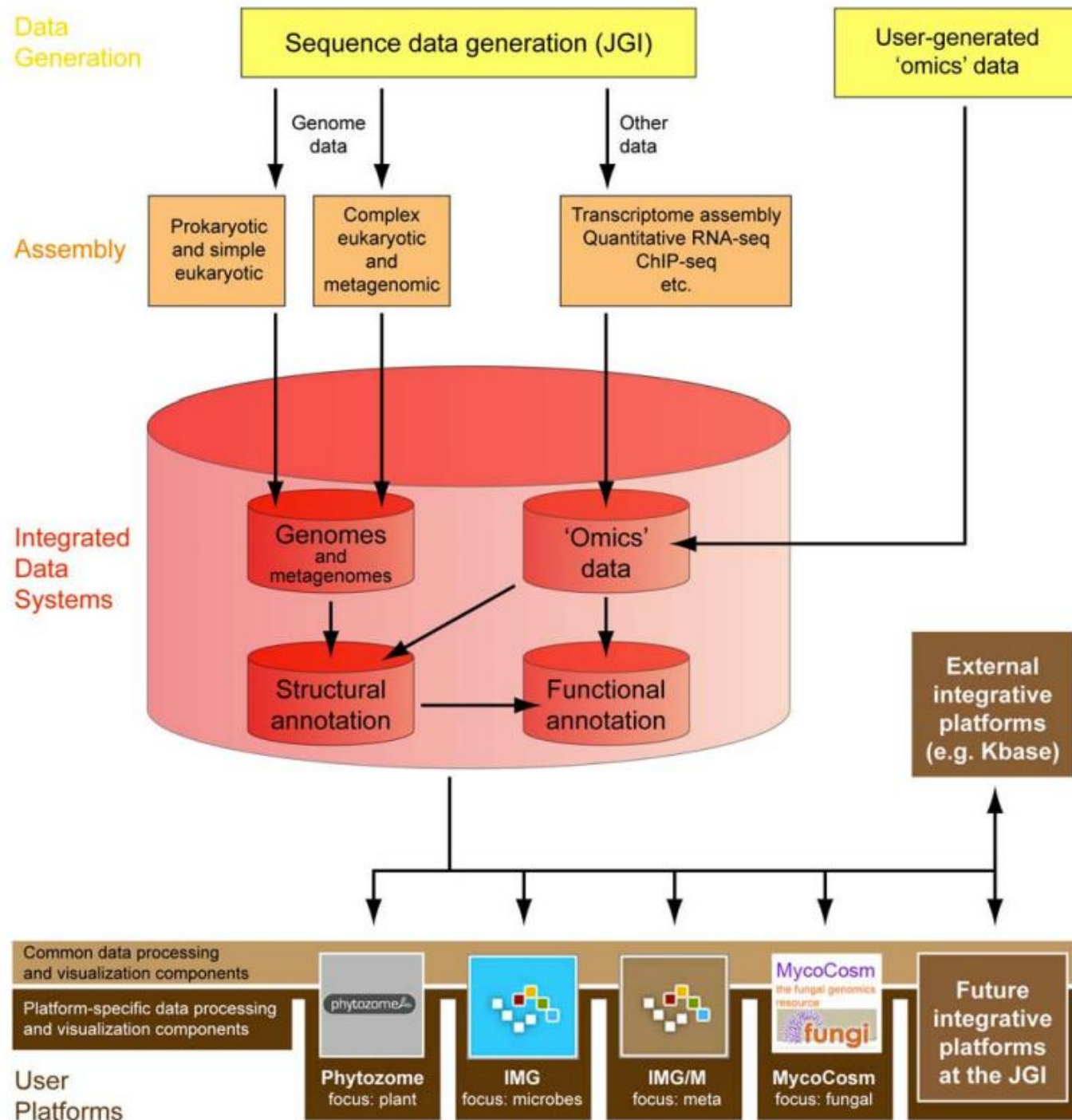
Sequencing Output
(based on FY11 budget)



<http://www.jgi.doe.gov/whoweare/10-Year-JGI-Strategic-Vision.pdf>



Sequencing as a core capability at the DOE JGI is tightly linked with other advanced capabilities of a next-generation genome science user facility.



Other major sequencing centers

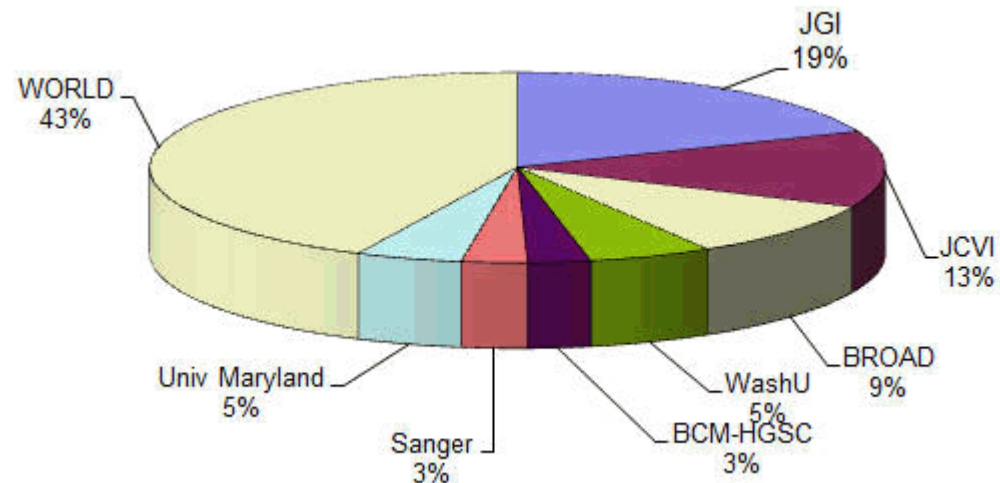
Broad Institute

Washington U Genome Center

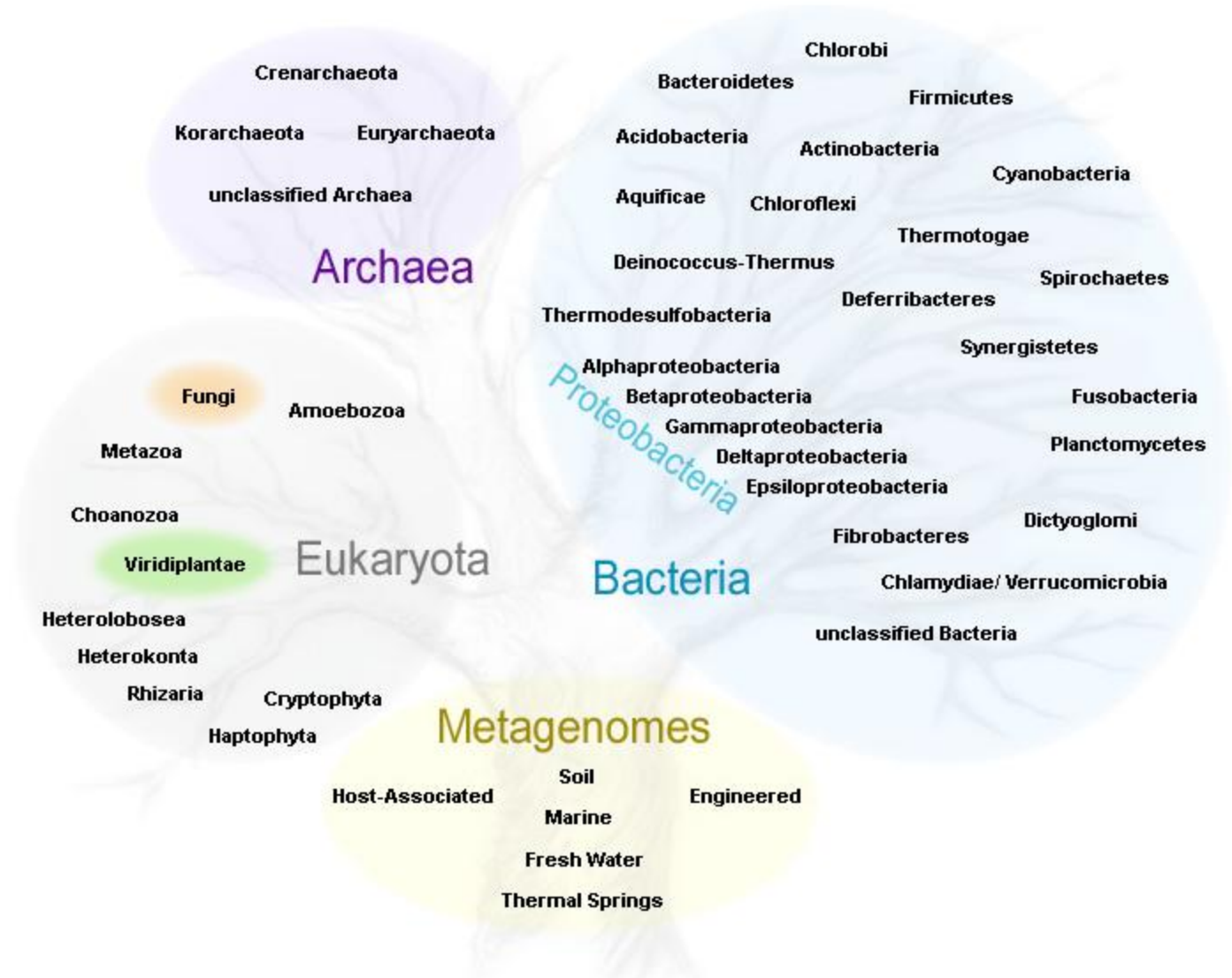
J. Craig Venter Institute (former TIGR)

Sanger Institute

Beijing Genomics Institute

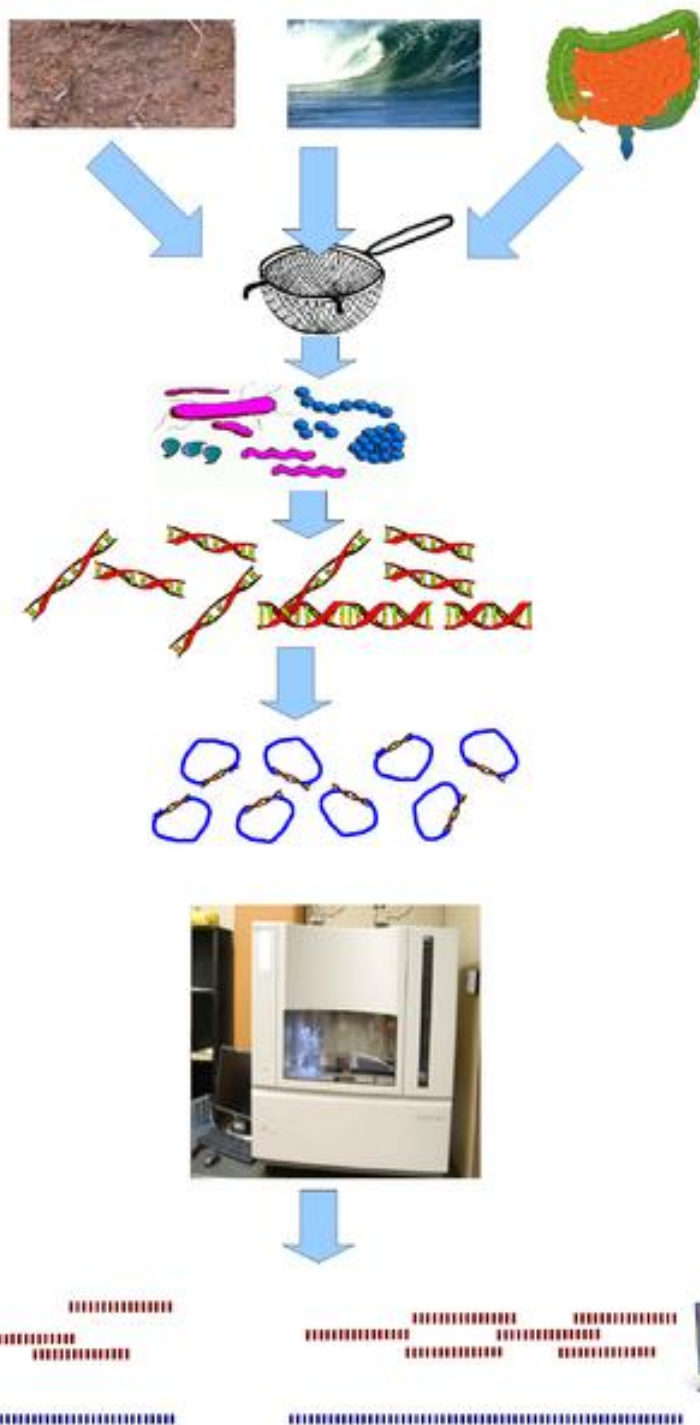


<http://genome.jgi.doe.gov/>



"Tree of Life" drawing by Leila Hornick, copyright 2005

TREE OF LIFE: To use the tree navigation: click a branch name and select the system for the organism of interest.



A

Metagenomics

(A) Sampling from habitat;

B

(B) filtering particles, typically by size;

C

(C) DNA extraction and lysis;

D

(D) cloning and library;

E

(E) sequence the clones;

(F) sequence assembly

F

Hands on practice: the Genome OnLine Database

Genomes OnLine Database



Version 4.0

Genomes Online Database

[Home](#)

Last update: 2013-02-08
Total # of genomes: **21571**

[Home](#)

[Genome Map](#)
[Genome Earth](#)
[Search](#)
[News](#)
[Statistics](#)
[Team](#)
[Reference](#)
[Contact](#)



Welcome to the Genomes OnLine Database

GOLD: Genomes Online Database, is a World Wide Web resource for comprehensive access to information regarding genome and metagenome sequencing projects, and their associated metadata, around the world.

Metagenomes	Isolate Genomes	Genome Distribution
Classification <ul style="list-style-type: none">Studies: 369Samples: 2350	Complete Projects: 4129 Incomplete Projects: 17402 Targeted Projects: 1087	<ul style="list-style-type: none">Project TypeSequencing StatusPhylogenetic

1. Register



Register your project information and Metadata in Genomes Online Database

[Register](#)

2. Annotate



Annotate your microbial genome or metagenome with IMG/ER or IMG/MER

[Annotate](#)

3. Publish



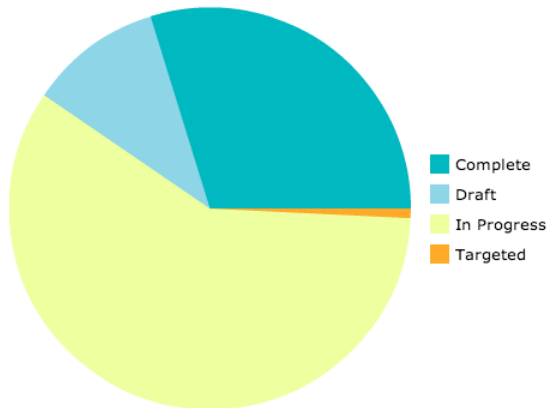
Publish your genome or metagenome in open access standards-supportive journal.

[Publish](#)

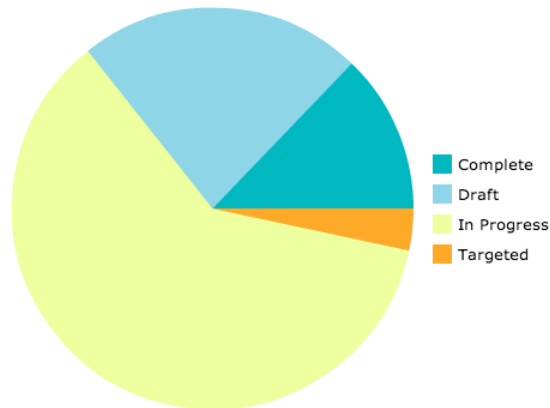
SEQUENCING STATUS DISTRIBUTION

A	ARCHAEA TOTAL: 453	Complete: 267 Permanent Drafts: 39	Draft: 14 In Progress: 77 DNA Received: 71 Awaiting DNA: 0	Targeted: 1
B	BACTERIA TOTAL: 19065	Complete: 6792 Permanent Draft: 1654	Draft: 2944 In Progress: 7863 DNA Received: 99 Awaiting DNA: 0	Targeted: 431
E	EUKARYA TOTAL: 3140	Complete: 1156 Permanent Draft: 34	Draft: 284 In Progress: 1573 DNA Received: 4 Awaiting DNA: 0	Targeted: 6

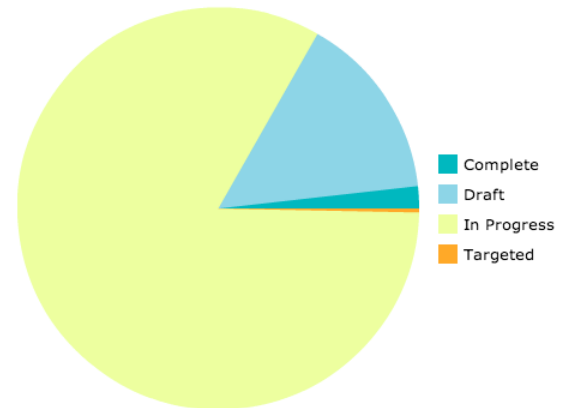
Archaeal Distribution



Bacterial Distribution



Eukaryotic Distribution



Last update: 2013-02-08
Total # of genomes: **22658**

[Home](#)[Genome Map](#)[Genome Earth](#)[Search](#)[News](#)[Statistics](#)[Team](#)[Reference](#)[Contact](#)[Blogger](#)

Statistics

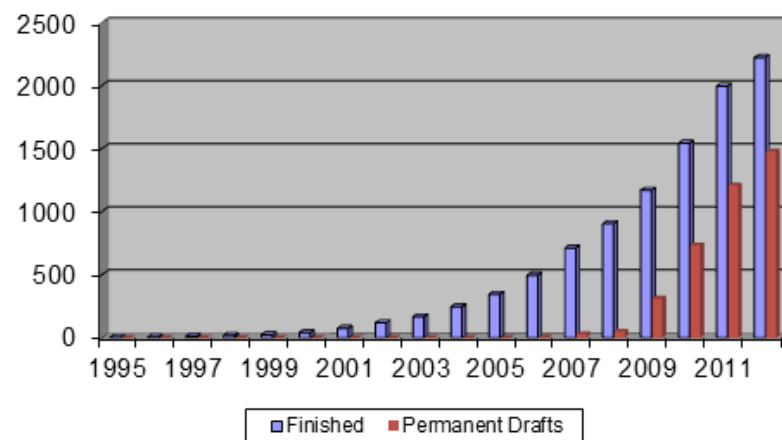
GOLD Indexing Information

1. [Completely Sequenced Genomes](#)
2. [Genome Sequencing Projects on GOLD](#)
3. [Genome Projects according to Phylogenetic Groups](#)
4. [Phylogenetic distribution of Bacterial Genome Projects](#)
5. [Funding Relevance of Bacterial Genome Projects](#)
6. [Major Sequencing Centers](#)
7. [Major Sequencing centers for Bacteria and Archaea](#)

1. Completely Sequenced Genomes

September 2012, 3699 projects

Complete Genome Projects ©
September 2012: 3699 Projects



2. Genome Sequencing Projects on GOLD

October 2011, 10031 projects

Last update: 2013-02-08
Total # of genomes: 22658

[Home](#)

[Genome Map](#)

[Genome Earth](#)

[Search](#)

[News](#)

[Statistics](#)

[Team](#)

[Reference](#)

[Contact](#)



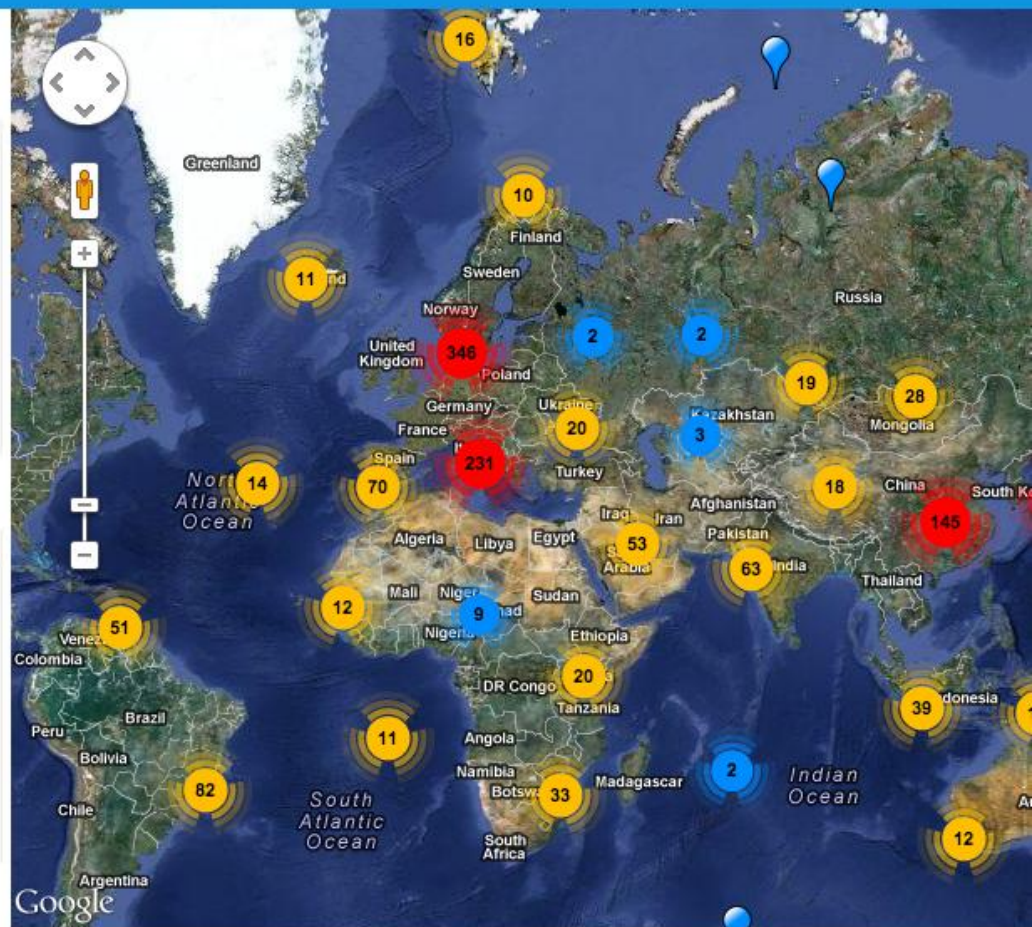
An example of MarkerClusterer v3

Genomes:

[Acaryochloris marina MBIC11017](#)
[Acaryochloris sp CCME 5410](#)
[Acetobacter pasteurianus IFO 3283-12](#)
[Acetobacter pasteurianus IFO 3283-03](#)
[Acetobacter pasteurianus IFO 3283-07](#)
[Acetobacter acetii ATCC 23746](#)
[Acetobacter pasteurianus IFO 3283-01](#)
[Acetobacter pasteurianus IFO 3283-32](#)

Metagenome Samples:

(Acid Mine Drainage (ARMAN) euryarchaeal community from 5-way FS site, at Richmond mine, Iron Mountain, CA, P5-way (FS) sample
0-1 mm depth into microbial mat
02J05
05E01 (contig09038)
05E01-33kb
06d18
1-2mm depth into microbial mat
10-22 mm depth into microbial mat



Contact Us

For Questions regarding the Database or to submit information on Genome Projects,

Please Contact us at:

[Mail at genomesonline.org](mailto:genomesonline.org)

[Search](#)[Advanced Search](#)[Metagenome Search](#)[Metadata Search](#)

SEARCH GOLD

Search by:

METAGENOME STUDY FILTERS

[Get Studies](#)[reset all selections](#)

METAGENOME INFORMATION

Metagenome Information Filter: Study Name

gut

Ecosystem

MATCHES

ALL

Ecosystem Category

MATCHES

ALL

Ecosystem Type

MATCHES

ALL

Ecosystem Subtype

MATCHES

ALL

Specific Ecosystem

MATCHES

ALL

Study Type

MATCHES

ALL

Project Status

MATCHES

ALL

Journal

MATCHES

ALL

[Get Studies](#)[reset all selections](#)

METAGENOME SAMPLE FILTERS

[Get Samples](#)[reset all selections](#)

SAMPLE INFORMATION

Sample Information Filter: Study Name

Ecosystem

MATCHES

ALL

Hands on practice: the Phytozome database

Genome

Genome

[Limits](#) [Advanced](#)



Genome

This resource organizes information on genomes including sequences, maps, ch... annotations.

Using Genome

- [Help](#)
- [Browse by Organism](#)
- [Download / FTP](#)
- [Submit a genome](#)

Genome Tools

- [BLAST the Human Genome](#)
- [Genomic groups BLAST](#)
- [NCBI remap](#)
- [Genome Decoration Page](#)

Custom resources

- [Human Genome](#)
- [Microbes](#)
- [Organelles](#)
- [Plants](#)
- [Viruses](#)

Genome Annotation and Analysis

- [Eukaryotic Genome Annotation](#)
- [Prokaryotic Genome Annotation](#)
- [PASC \(Pairwise Sequence Comparison\)](#)
- [TaxPlot \(3-way Genome Comparison\)](#)

Other Resources

- [Assembly](#)
- [BioProject](#)
- [BioSample](#)
- [Map Viewer](#)
- [Protein Clusters](#)

External Resources

- [GOLD - Genomes On](#)
- [Ensembl Genome Bro](#)
- [Bacteria Genomes at](#)
- [Large-Scale Genome](#)



[Bioenergy Research
at JGI](#)



[Genomic
Technologies](#)



[Fungal Genomics
Program](#)



[Metagenomics
Program](#)



[Microbial Genomics
Program](#)



[Plant Genomics
Program](#)

Characterizing major bioenergy crops and those model plants to understand the complex traits related to biomass yields, efficient deconstruction of cell walls, improved drought and pest resistance, and easier conversion to biofuels.



Announcement

**Genomics &
Environment
8th Annual
User Meeting**

Save the date:
March 26 - 28

REGISTER

Latest News



December 19, 2012

From the stockings decorating mantles to the new outfits in display windows calling to shoppers, cotton is woven into the fabric of the holiday season. For bioenergy researchers, however, fiber composition matters more than color and texture as each cotton strand is composed of more than two dozen coils of cellulose, a target biomass for next-generation biofuels. [» more...](#)

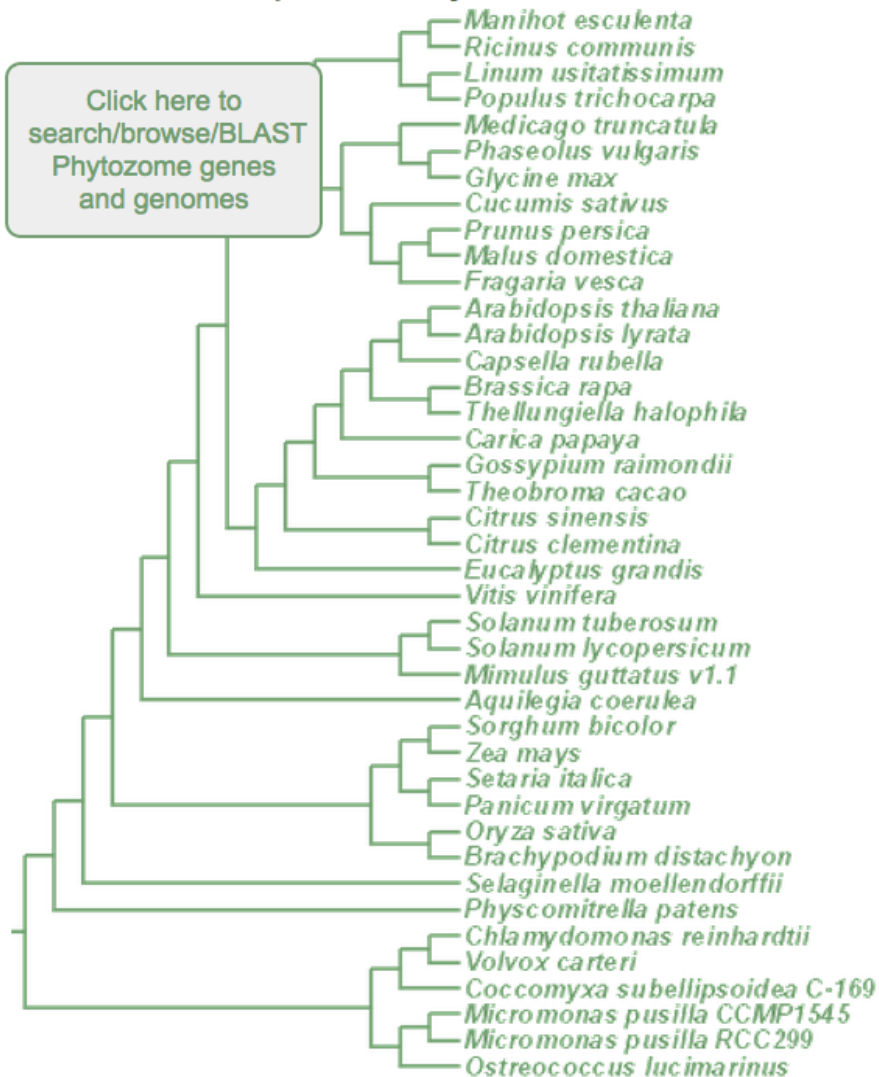
DOE JGI 2013 User Meeting

10 Year Strategic Plan

A major objective of the JGI is to couple the

Species in Phytozome v9.0

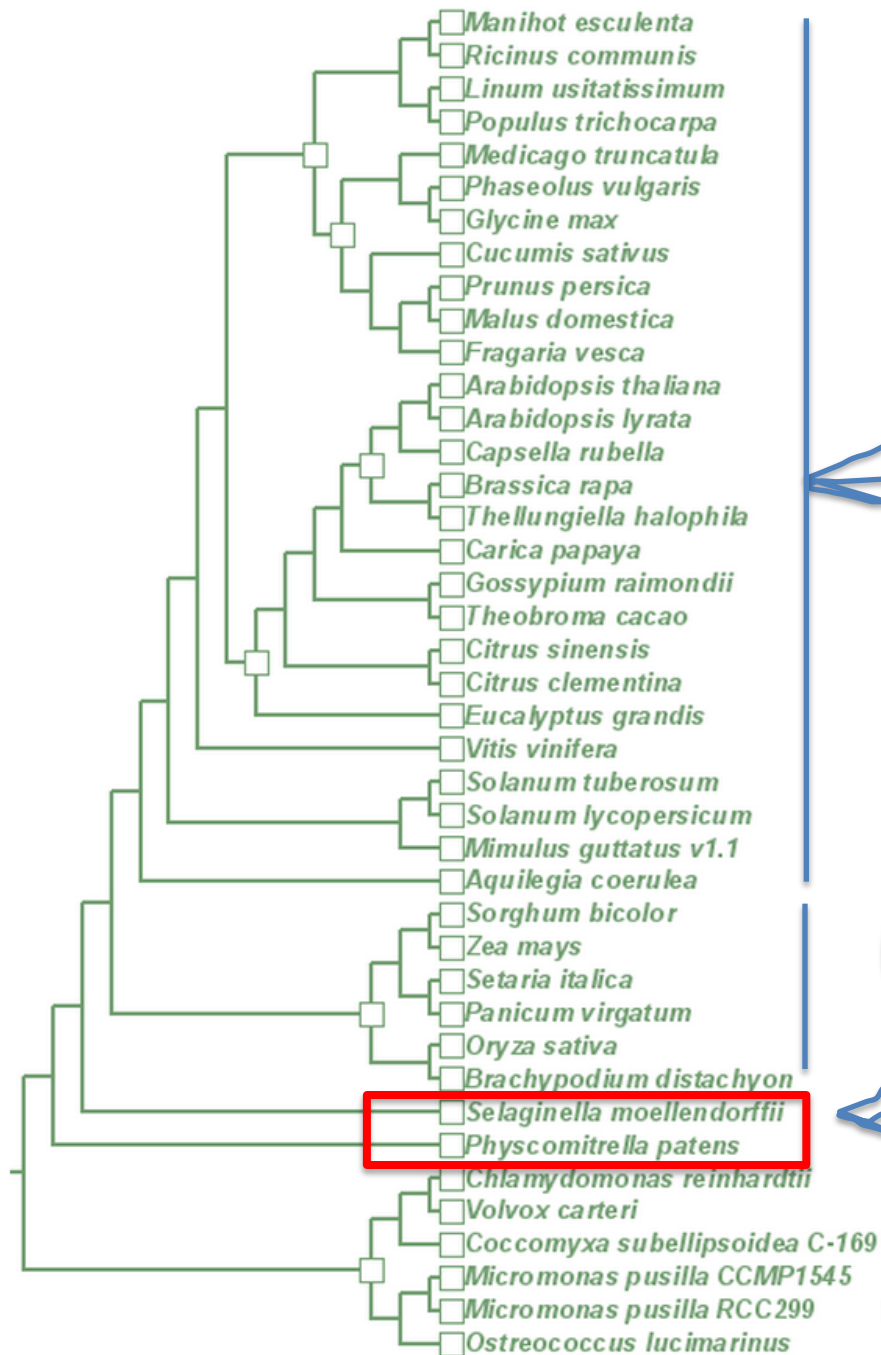
Click here to
search/browse/BLAST
Phytozome genes
and genomes



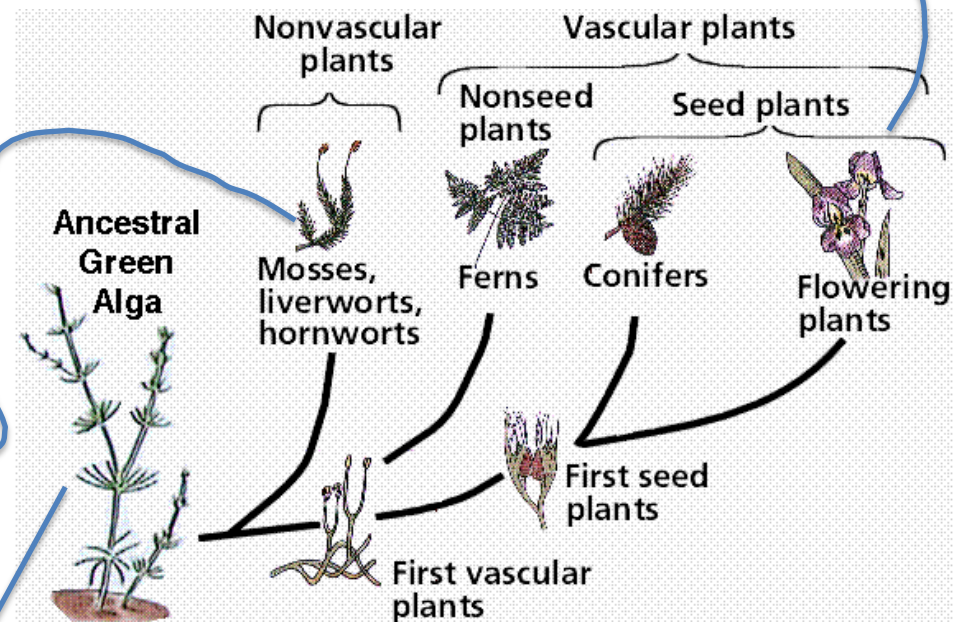
Phytozome is a joint project of the Department of Energy's Joint Genome Institute and the Center for Integrative Genomics to **facilitate comparative genomic studies amongst green plants**. Families of orthologous and paralogous genes that represent the modern descendants of ancestral gene sets are constructed at key phylogenetic nodes. These families allow easy access to clade specific orthology/paralogy relationships as well as clade specific genes and gene expansions. As of release v9.0, Phytozome provides access to **forty-one sequenced and annotated green plant genomes** which have been clustered into gene families at six evolutionarily significant nodes. Where possible, each gene has been **annotated with PFAM, KOG, KEGG, and PANTHER assignments**, and publicly available annotations from RefSeq, UniProt, TAIR, JGI are hyper-linked and searchable.

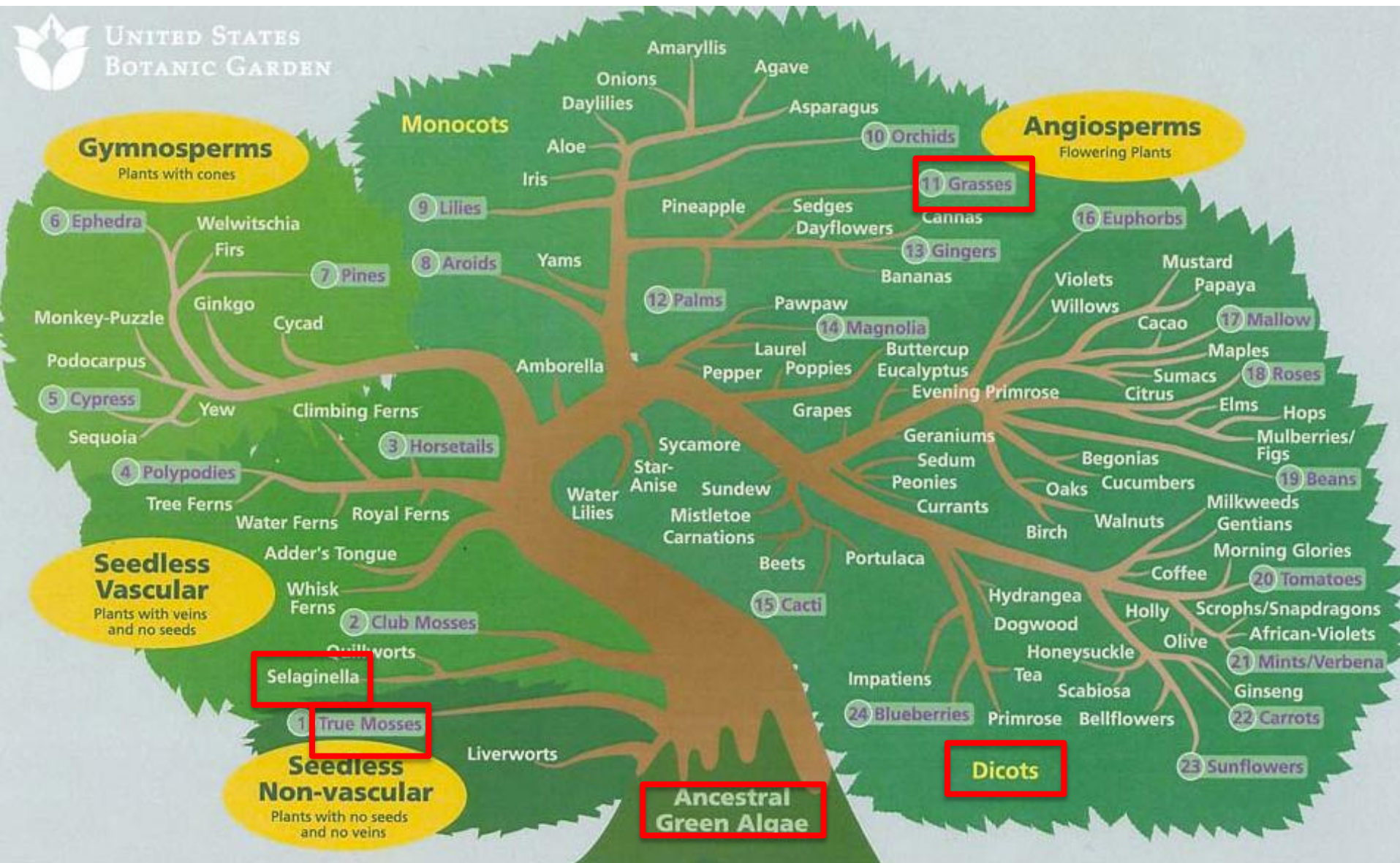
1. Choose a species or ancestral node by clicking a ☐ on the tree

2. Choose a tool:



Keyword search	expand
BLAST search	expand
BLAT search	expand
Genome browser	Launch
Info page	Get info
Bulk data	Get data







2. Choose a tool:

Keyword search

submit

Target: Grass gene families

Search term: ☒ add trailing wildcard

Search type: ☒ Symbols/Identifiers/Define
☐ Ontologies

Family members: ☐ Restrict species

Summary ?

Analyze Results / Get Data

37 genes found

SEARCH CRITERIA [revise your search](#)

Search term: **cellulose**

Search target: **Arabidopsis thaliana columbia genes**

Member filtering: none

Search type: **Symbols/Identifiers/Defines**

1 - 37 of 37 genes

Arabidopsis thaliana columbia genes retrieved

<input type="checkbox"/>	MEMBER COUNT	DESCRIPTION	<i>A. thaliana</i>
<input type="checkbox"/>	1	Cellulose-synthase-like C4	1
<input type="checkbox"/>	1	Cellulose-synthase-like C5	1
<input type="checkbox"/>	1	Cellulose-synthase-like C6	1
<input type="checkbox"/>	1	Cellulose-synthase-like C12	1
<input type="checkbox"/>	1	Cellulose synthase family protein	1
<input type="checkbox"/>	1	Cellulose synthase family protein	1
<input type="checkbox"/>	1	Cellulose synthase family protein	1

Look for CesA1

Arabidopsis thaliana gene AT4G32410 : cellulose synthase 1

About this gene ?

Sequences

Protein Homologs

Gene Ancestry

Get Data


Info:

Locus name AT4G32410

Transcript name AT4G32410.1

Aliases RSW1
CESA1
AtCESA1

Description cellulose synthase 1

Links to external DBs  **tair**

Functional annotations for this locus:

Pfam:03552  Cellulose synthase

Panther:13301:SF21

Panther:13301 X-BOX TRANSCRIPTION FACTOR-RELATED

KEGGORTH:10999 cellulose synthase A [EC:2.4.1.12]

GO:0030244 cellulose biosynthesis

GO:0016757 transferase activity, transferring glycosyl groups



9 more annotations



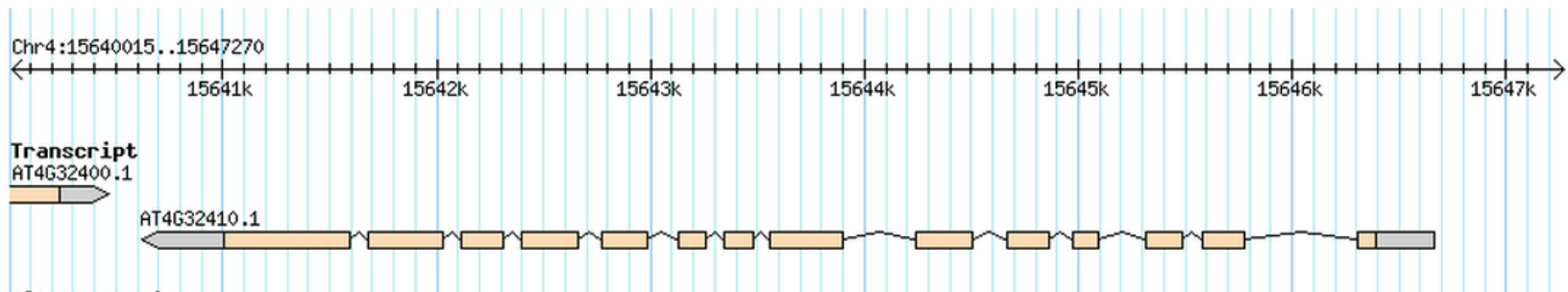
Protein domain view:

1

1081

Genomic view:

Location: Chr4: 15640620 - 15646665 :: [View in Gbrowse environment](#)





2. Choose a tool:

Keyword search	expand
BLAST search	expand
BLAT search	expand
Genome browser	Launch
Info page	Get info
Bulk data	Get data

Physcomitrella patens (Moss)



Browse
Genome



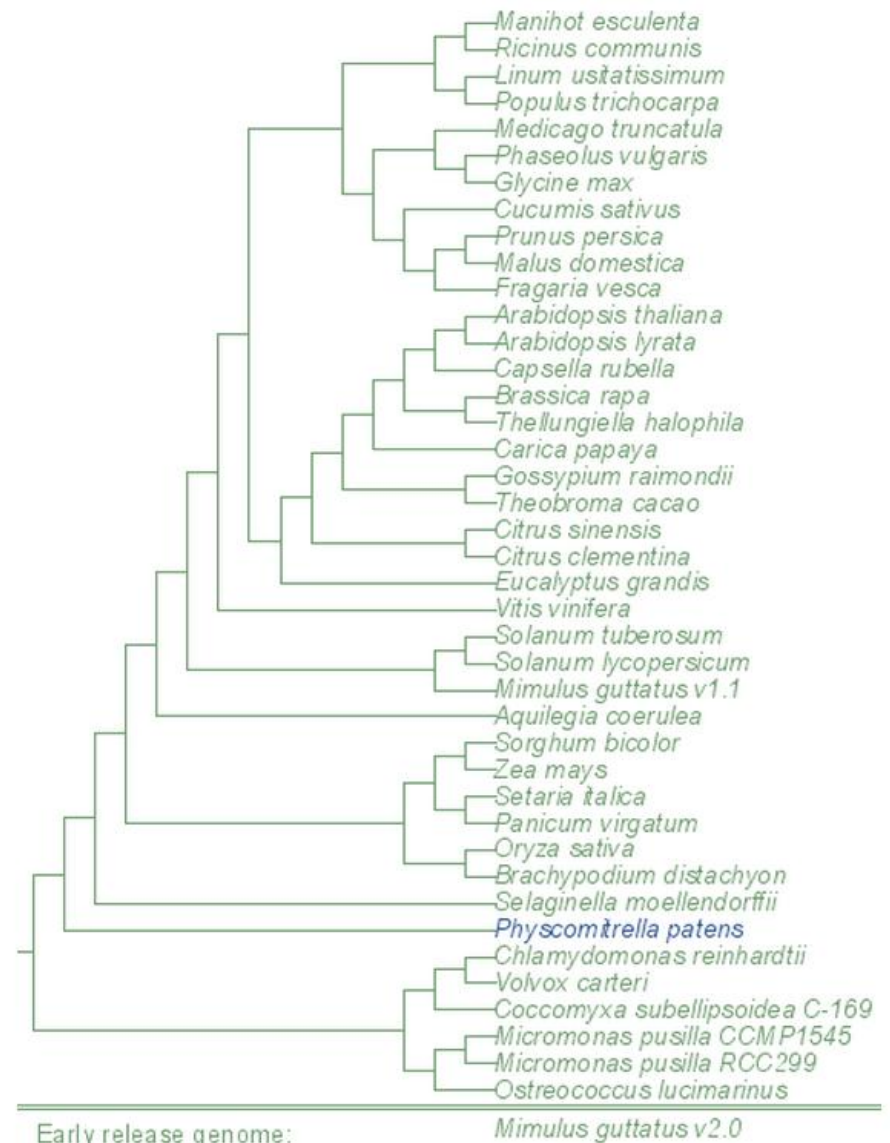
BLAST
Genome



Download
data

About the genome:

- Overview
- Statistics



Panicum virgatum Data Usage Policy

As a public service, the raw sequence data from this genome sequencing project are being made available by the Department of Energy Joint Genome Institute (JGI) before scientific publication. 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 collaborators and the JGI will publish their results without concerns about potential preemption by other groups that did not participate in the effort.

By accessing these data, **you agree not to publish** any articles containing analyses of genes or genomic data on a whole genome or chromosome scale prior to publication by JGI and its collaborators of its comprehensive genome analysis

Reserved analyses include the identification of complete (whole genome) sets of genomic features such as genes, gene families, regulatory elements, repeat structures, GC content, etc., and whole-genome comparisons of regions of evolutionary conservation or change. Interested parties are encouraged to contact the the principal collaborator and JGI to discuss such possibilities

The data may be freely downloaded and used by all who respect the restrictions in the previous paragraphs. While our in training period status, the assembly and raw sequence reads should not be redistributed or repackaged without permission from the JGI. 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." Once moved to unreserved status, the data are freely available for any subsequent use.

OK, proceed to data

lftp ftp://ftp.jgi-psf.org/pub/compgen/phytozome/v9.0/Pvirgatum/

```
yyin@glu:~$ lftp ftp://ftp.jgi-psf.org/pub/compgen/phytozome/v9.0/Pvirgatum/  
cd ok, cwd=/pub/compgen/phytozome/v9.0/Pvirgatum  
lftp ftp.jgi-psf.org:/pub/compgen/phytozome/v9.0/Pvirgatum> ls  
-rw-rw-r-x      1 24799      124          3055 Dec 13 11:00 Pvirgatum_202_readme.txt  
drwxrwxr-x      2 24799      124          482 Dec 13 11:00 annotation  
drwxrwxr-x      2 24799      124          133 Dec 13 11:00 assembly  
drwxrwxr-x      2 32295      124          113 Feb 07 14:09 gbrowse  
drwxrwxr-x      2 24799      124           0 Dec 13 11:00 related_files  
lftp ftp.jgi-psf.org:/pub/compgen/phytozome/v9.0/Pvirgatum> █
```

File Help

Arabidopsis thaliana: 6.046 kbp from Chr4:15,640,468..15,646,513

Browser Select Tracks Snapshots Custom Tracks Preferences

Search

Landmark or Region:

Chr4:15,640,468..15,646,513 Search

Examples: Chr3:5550001..5600000.

Data Source

Arabidopsis thaliana

Annotate VISTA Plot: A. coerulea v.1.0

Configure...

Go

Save Snapshot

Load Snapshot

Scroll/Zoom:

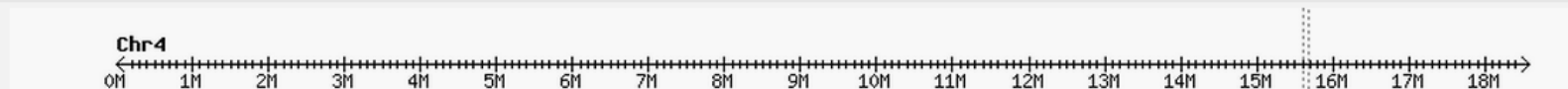


Show 6.046 kbp



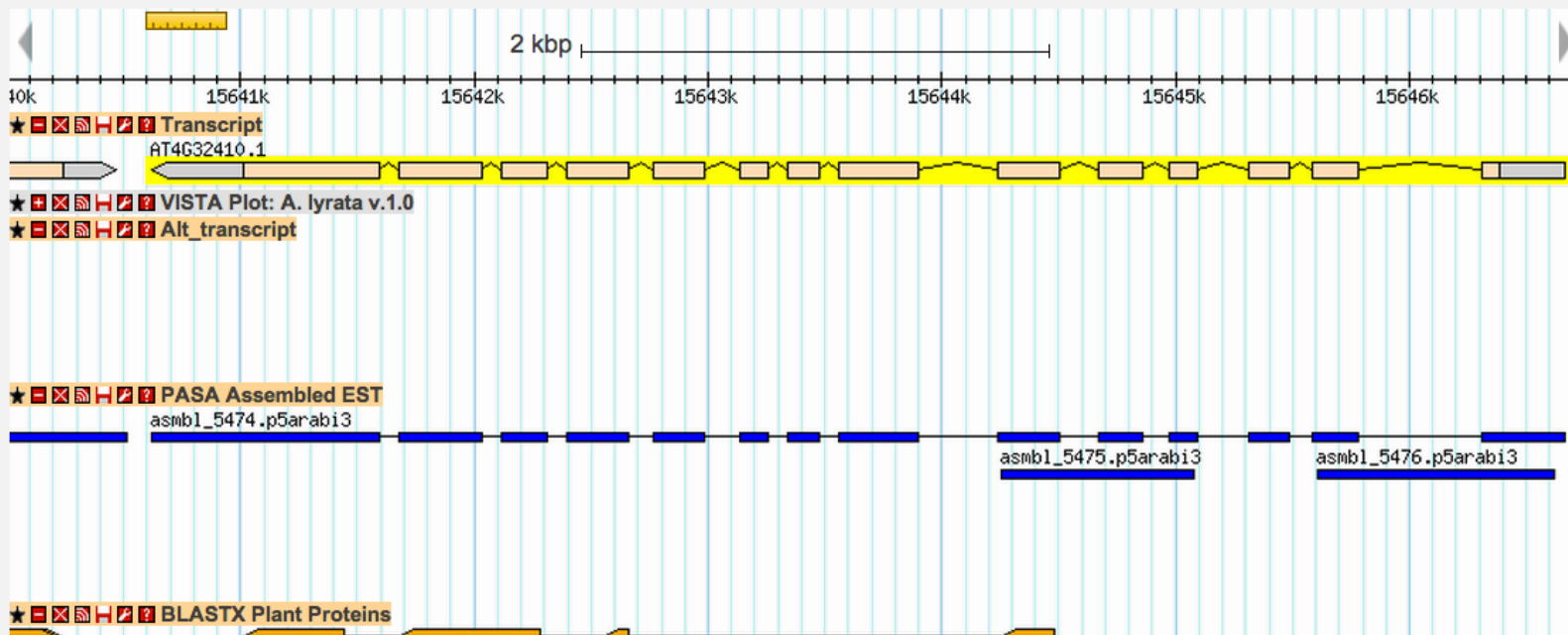
Flip

Overview



Region

Details



Hands on practice:
the microbial genome/metagenome
database (IMG/M)



DOE Joint Genome Institute

Enabling Advances in Bioenergy & Environmental Research



[Bioenergy Research at JGI](#)



[Genomic Technologies](#)



[Fungal Genomics Program](#)



[Metagenomics Program](#)



[Microbial Genomics Program](#)



[Plant Genomics Program](#)

A significant portion of the DOE JGI's projects are related to bioenergy and focus on three areas: developing plant feedstocks; using microbes to break down cellulose in plant cell walls; and fermenting sugars into biofuels.



Announc

Genom
Enviro
8th An
User M

Save th
March 2

REGI

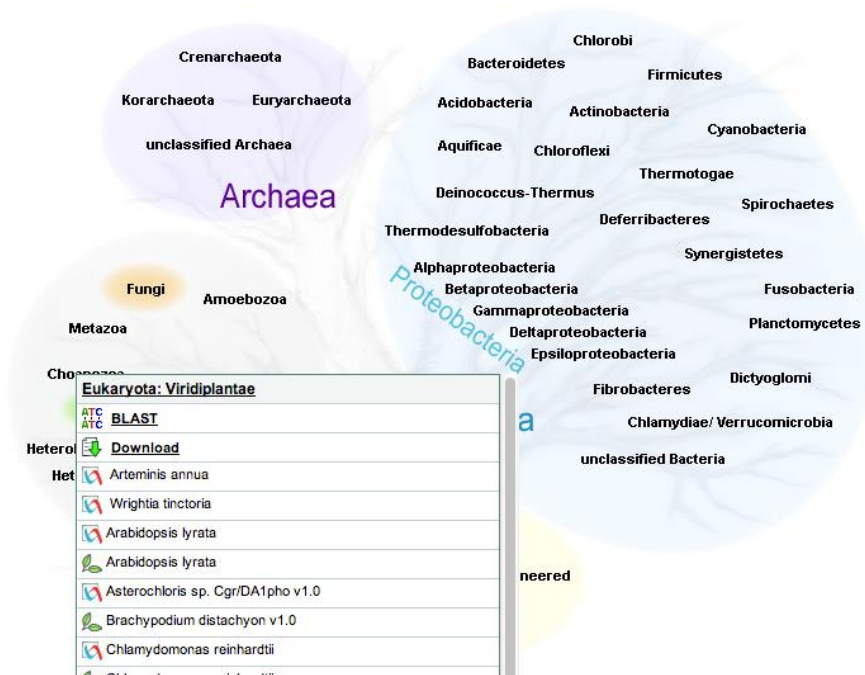
Latest News

December 19, 2012

From the stockings decorating mantles to the new outfits in display windows calling to

Please help us to improve the JGI Genome Portal. Your feedback is very important to us. [Click here](#) to take the annual Portal survey.

On Tuesday, Feb 12, JGI SysOps and NERSC will be conducting quarterly computer system maintenance tasks. This may impact the availability and/or performance of this site during that time. This site will most likely be unavailable for most of that day.



New to Genome Portal?
[Explore](#)

Genome Releases

- [Fungal Releases](#)
- [Metagenomics Releases](#)
- [Microbial Releases](#)
- [Plant Releases](#)

[Integrated Microbial Genomes \(IMG\)](#) - a comprehensive resource for comparative genomics and annotation of all publicly available genomes.

[Mycocosm](#) - the Fungal Genomics Resource - access to the annotated fungal genomes and analysis.

[Phytozome](#) - a comparative hub for plant genomes and gene family data and analysis.



IMG 4 Data Management

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 bacterial, archaeal, eukaryotic, and phage genomes, as well as engineered, environmental and host associated metagenome samples. **RefSeq** is the main source of publicly available genome sequence datasets. Genes in IMG are characterized using several functional resources, including **COG**, **KOG**, **KEGG** (release 63.0, July 2012), **PFAM** (version 26.0, Nov 2011), **TIGRfam** (release 12.0, Feb 2012), **MetaCyc** (release 16.1, July 2012), **Gene Ontology** (June 2012), and **Interpro** (Apr 2012).

About **30%** of the **genomes** and **75%** of the **metagenome samples** in IMG were **sequenced at DOE's Joint Genome Institute**.

IMG ER and IMG/M ER have about **2,500** registered **users** from **61 countries** across **North America** (57%), **Europe** (20%), **Asia** (13%), **South America** (5%), **Oceania** (4.5%), and **Africa** (.5%).

[Register as an IMG User](#)

Microbial Genome Analysis Systems



[IMG](#)

IMG provides users with analysis tools ([IMG UI Map](#)) for examining publicly available genomes in IMG.



[IMG/ER](#)

IMG/ER (Expert Review) provides users with analysis tools ([IMG UI Map](#)) for examining and revising the annotations for their private (password protected) genomes in the context of publicly available genomes in IMG.



[IMG/EDU IMG/ACT](#)

IMG Education (IMG/EDU) provides support for teaching microbial genome analysis courses.

Metagenome Analysis Systems



[IMG/M](#)

IMG/M provides users with analysis tools ([IMG/M UI Map](#)) for examining publicly available metagenome samples and genomes in IMG.




[IMG/MER](#)

IMG/M ER (Expert Review) provides users with analysis tools ([IMG/M UI Map](#)) for examining and revising the annotations of their private (password protected) metagenome samples and genomes in the context of publicly available genomes and metagenome samples in IMG.



[IMG/HMP Metagenomes](#)

IMG/M HMP (Human Microbiome Project) provides users with analysis tools for examining the annotations of HMP specific genomes and metagenome samples in the context of all publicly available genomes and metagenome samples in IMG.

 On Feb. 12, 2013 All IMG systems will be down for general maintenance. Sorry for the inconvenience.

IMG/M Content

Datasets

[Bacteria](#) 4878

[Archaea](#) 184

[Eukarya](#) 187

[Plasmids](#) 1184

[Viruses](#) 2809

[Genome Fragments](#) 654

[Metagenome](#) 1237

[Total Datasets](#) 11133

[GEBA](#) 238

Last updated: [2013-01-17](#)

**IMG 4.0 is dedicated to the
memory of our colleague,
Iain Anderson**

[Metagenome Projects Map](#)
[System Requirements](#)
[About IMG](#)
[FAQ](#)

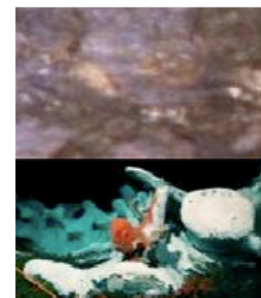


Hands on
training available
at the

[Microbial Genomics &
Metagenomics Workshop](#)

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 isolate and single cell genomes, and a rich set of publicly available metagenome samples.

For details, see [IMG Release Notes](#) (Dec. 12, 2012), in particular the workspace and background computation capabilities available to IMG registered users.



[IMG/M Statistics](#)

[IMG/M ER Account Request](#)

IMG/M ([Nucleic Acids Research, Vol 40, 2012](#)) provides users with tools ([IMG/M UI Map](#)) for analyzing their private (password protected access) metagenome samples in the context of all public (free access) genome and metagenome samples in IMG.

IMG/M contains 1237 public metagenome samples distributed as follows:

Engineered	36	Environmental	370	Host-associated	831
Bioremediation	3	Air	2	Arthropoda	23
Biotransformation	5	Aquatic	305	Birds	5
Lab enrichment	4	Terrestrial	63	Human	753
Solid waste	11			Mammals	17
Wastewater	13			Mollusca	8
				Plants	22
				Porifera	3

On Feb. 12, 2013 All IMG systems will be down for general maintenance. Sorry for the inconvenience.

Status	Bacteria	Archaea	Eukaryota	Plasmids	Viruses	Genome Fragments	Microbiome	Total
Finished	2052	150	36	1184	2809	654	11	6896
Draft	2215	14	151	0	0	0	1221	3601
Permanent Draft	640	20	0	0	0	0	5	665
Total	4907	184	187	1184	2809	654	1237	11162

DNA	Bacteria	Archaea	Eukaryota	Plasmids	Viruses	Genome Fragments	Microbiome
DNA, no. of bases	18,752,807,313	455,435,633	38,339,677,249	33,515,725	87,874,492	22,728,454	438,230,359,861
DNA, no. of coding bases	16,172,027,014	391,589,615	10,580,493,706	25,452,415	77,225,608	19,679,430	58,975,823,848
DNA, G+C no. of bases	9,696,405,163	225,602,414	14,621,930,290	16,202,763	39,337,990	14,530,743	217,640,299,436
DNA, Scaffolds	508,874	3,249	1,110,164	1,185	4,286	654	2,526,564,010
Total	44,621,748,364	1,072,630,911	63,543,211,409	75,172,088	204,442,376	56,939,281	717,373,047,155

On Feb. 12, 2013 All IMG systems will be down for general maintenance. Sorry for the inconvenience.

Genome Browser

Add Selected to Genome Cart

Select All

Clear All

View Phylogenetically

Filter column: Domain Filter text Apply

Export

Page 1 of 13 << first < prev 1 2 3 4 5 6 7 8 9 10 next > last >>

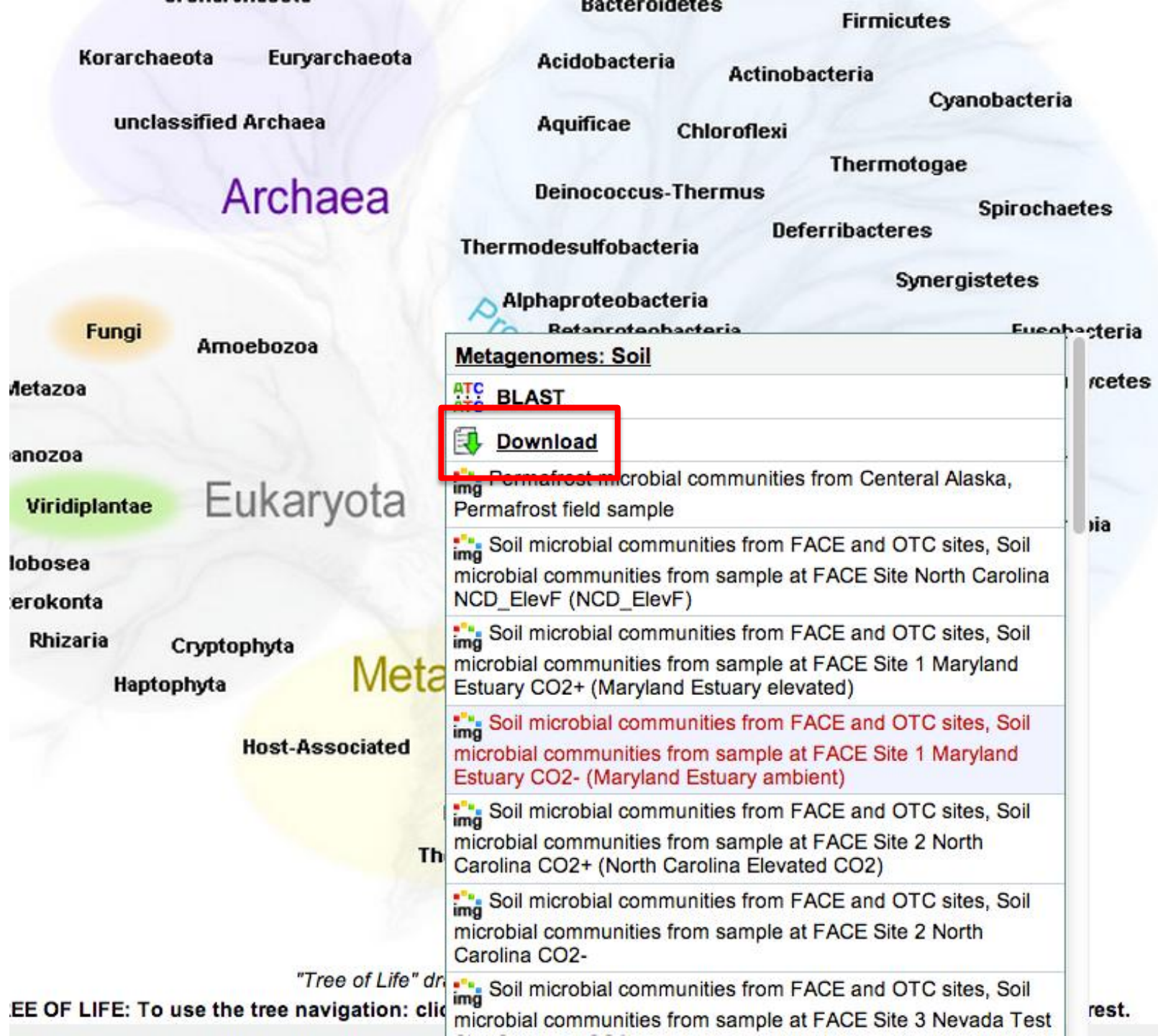
100

Column Selector

Select Page


Deselect Page

Select	Domain	Status	Proposal Name	Genome Name / Sample Name	Sequencing Center	Phylum	Genome Size	Gene Count	GC %
<input type="checkbox"/>	*	D	Human Microbiome Production Phase community	Human Supragingival plaque microbiome from visit number 2 of subject 764143897	Washington Univ	Host-associated	111531229	425536	.46843
<input type="checkbox"/>	*	D	Human Microbiome Production Phase community	Human Stool microbiome from visit number 2 of subject 159247771	Broad Institute	Host-associated	82506363	238942	.4681
<input type="checkbox"/>	*	D	Environmental microbial communities from Soda lakes and Soda Solonchak soils	Soda Lake 5KL_010_BRINE (Soda Lake 5KL_010_BRINE, July 2012 Assem)	DOE Joint Genome Institute	Environmental	318457321	700745	1.04435
<input type="checkbox"/>	*	D	Human Microbiome Production Phase community	Human Tongue dorsum microbiome from visit number 2 of subject 159571453	Broad Institute	Host-associated	115317748	390682	.41907
<input type="checkbox"/>	*	D	Human Microbiome Production Phase	Human Stool microbiome from visit number 1 of subject	Washington Univ	Host-associated	106116893	408960	.43445



"Tree of Life" dr
 EE OF LIFE: To use the tree navigation: clic

rest.

 On Tuesday, Feb 12, JGI SysOps and NERSC will be conducting quarterly computer system maintenance tasks. This may impact the availability and/or performance of this site during that time. This site will most likely be unavailable for most of that day.

Data Release Policy

As a public service, the raw sequence data from the genome sequencing projects is being made available by the Department of Energy (DOE) and JGI to balance the imperative of DOE and JGI that the data from its sequencing projects be made available to the scientific community. JGI will publish their results without concerns about potential preemption by other groups that did not participate in the effort.


These pre-publication data are preliminary and may contain errors. The goal of our policy is that early release should enable the community to perform analyses of genes or genomic data on a whole genome or chromosome scale prior to publication by JGI and its collaborators. The publication of the whole genome analysis on organism by organism basis and reflected on their corresponding portal web-pages, not compete directly with planned publications (e.g. reserved analyses) of the JGI and collaborators. A principal collaborator is listed on the JGI website about their intentions and any potential collaboration.

Reserved analyses include the identification of complete (whole genome) sets of genomic features such as genes, gene families, comparisons of regions of evolutionary conservation or change. Manually annotated genes within the Genome Portal are also reserved. Analyses not planned by the JGI and its collaborators may also be undertaken following an agreement to that effect. Interested parties are encouraged to contact JGI for more information.

If these data are used for publication, the following acknowledgment should be included: "These sequence data were produced by JGI in collaboration with the user community." We request that you notify us upon publication.

The data may be freely downloaded and used by all who respect the restrictions in the previous paragraphs. While still in waiting period, the data may be repackaged without permission from the JGI. Any redistribution of the data during this period should carry this notice: "The Joint Genome Institute (JGI) does not assume any legal liability or responsibility for any purpose for which the data are used." Once moved to unreserved status, the data may be used for any purpose without restriction.

☒ Do not show Data Usage Policy for **environmental-terrestrial-soil** next time

 On Tuesday, Feb 12, JGI SysOps and NERSC will be conducting quarterly computer system maintenance tasks. This may impact the availability and/or performance of this site during that time. This site will most likely be unavailable for most of that day.

Download Selected Files

ExpandAll

CollapseAll

Rescan

☒ Organize by file type.
 ☐ Organize by genome.

- ☒ **All Downloads**
 - ☐ Download All (in one zip file)
 - ☐ Assembly
 - ☐ Feature
 - ☐ Functional
 - ☐ Additional Files
- ☒ **Raw Data**
 - ☐ Additional Files

Next lecture: Galaxy for
sequence analysis