Lawrence Berkeley National Laboratory

Lawrence Berkeley National Laboratory

Title

The US Department of Energy Joint Genome Institute Microbial Genome Program

Permalink

https://escholarship.org/uc/item/7pb1s77q

Author

Lapidus, Alla

Publication Date

2005-02-11

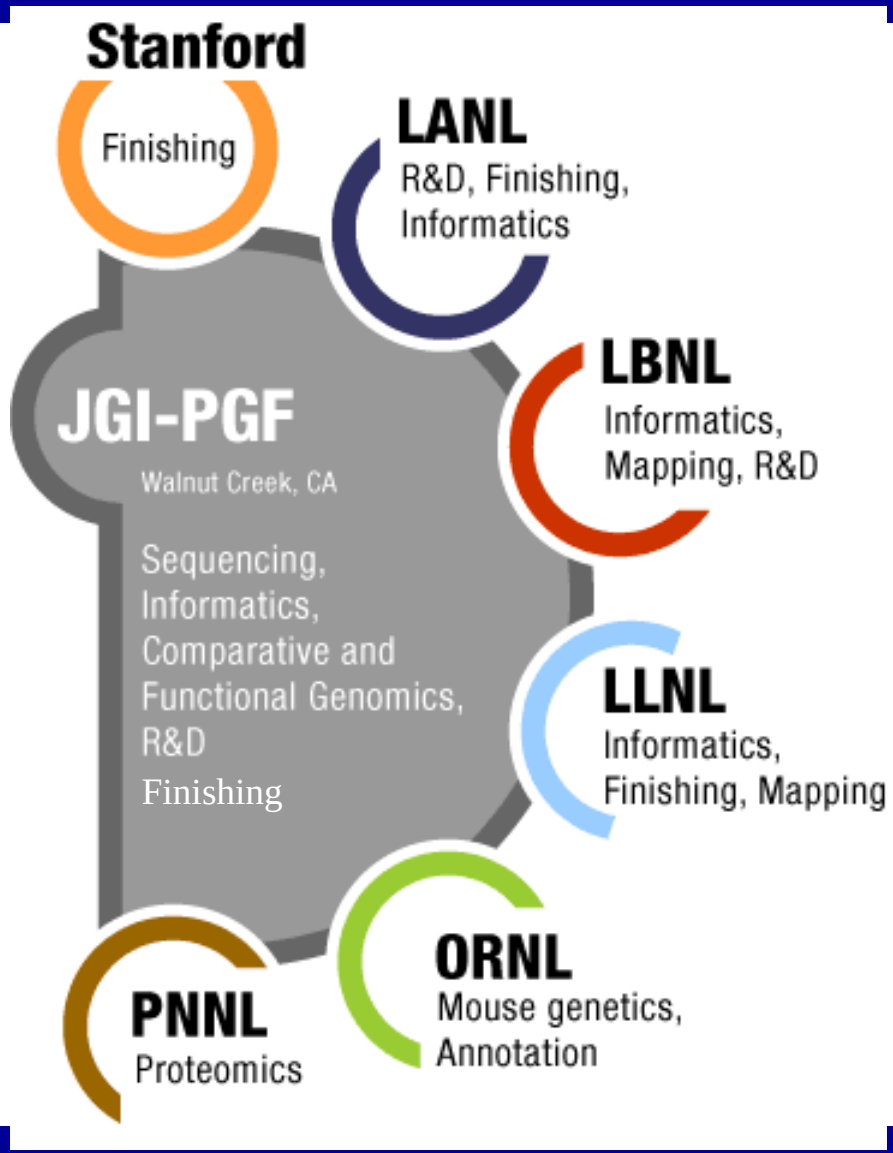
eScholarship.org Powered by the California Digital Library
University of California

DOE JOINT GENOME INSTITUTE
US DEPARTMENT OF ENERGY
OFFICE OF SCIENCE

The US Department of Energy Joint Genome Institute Microbial Genome Program

JGI Partners

The DOE Joint Genome Institute (JGI) is a "virtual institute" that integrates the sequencing and analytical activities of six partner institutions:



Alla Lapidus
February 11, 2005
Marco Island

Address http://www.science.doe.gov/production/ober/microbial.html Go Links

Microbial Genome Program

Divisions
Life Sciences
Medical Sciences
Environmental Sciences
About Us
Contacts
Research Topics
Research Opportunities
Fellowships and Educational Programs
Cool Stuff for All Ages
International Programs
FYI
Directions
DOE

Recommendations for Sequencing Targets (pdf format)

A 1994 spinoff from the Human Genome Project, the Microbial Genome Program's goal has been to completely sequence the genomes of microbes, primarily prokaryotes. However, unlike the human genome, which took years to complete, many microbial genomes can be completely sequenced in weeks or months and, with recent advances in sequencing technologies, even days. Only a few years ago, most scientists could not have imagined having access to the complete genetic sequence of more than a few microbes. As of April 2003, DOE has sequenced the genomes of about 100 microbes, most of them by the Joint Genome Institute. These, in addition to many viruses and higher organisms such as yeast and the roundworm, are available in public databases and are being actively used by academic, medical, and industrial scientists to make comparisons not possible previously. The completed genome of one microbe, *Methanococcus jannaschii*, confirmed the existence of a third major branch of life on earth, the Archaea.

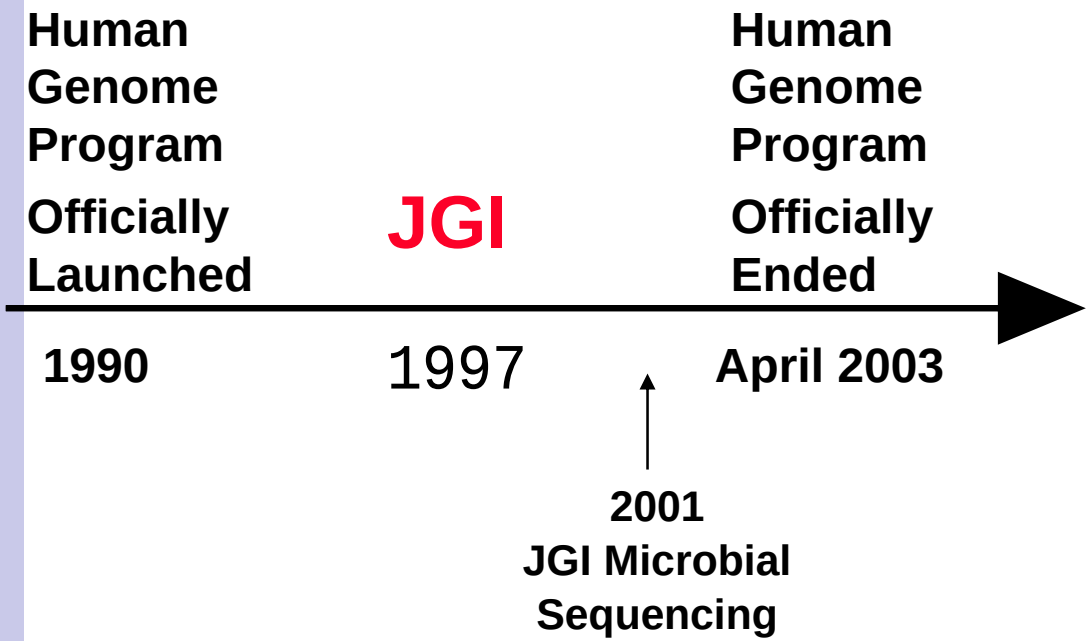
In newly sequenced microbial genomes, about 40% of the predicted open reading frames (potential genes) typically do not have matches in existing databases; this means that even though investigators have high confidence in the existence of these potential genes, their function and biological value are unknown. This exciting finding suggests that a large reservoir of interesting proteins and much fascinating biology remain to be discovered. One of the hardest biological problems, and one in which bioinformatics is beginning to have an impact, is to predict protein structure from primary amino acid sequence. Ultimately, scientists expect to understand the functions of unknown proteins, starting with their gene sequences.

The possible uses of this information are staggering. Microbes, incredibly, make up around 60% of the earth's biomass. They have survived on the planet for over 3.8 billion years and have been found in every conceivable environment, surviving extremes of heat, cold, radiation, pressure, salt, and acid—often where no other forms of life can exist. This rich diversity means that microbes long ago "solved" many problems for which scientists have been actively seeking answers.

Through the study and understanding of a diverse group of microbes, solutions are nearer for DOE mission challenges in environmental cleanup, medicine, agriculture, industrial processes, and energy production and use, to name a few. For example, *M. jannaschii*'s ability to produce methane may have implications for new forms of fuel generation, and *Deinococcus radiodurans* has potential for cleanup of toxic mixed-waste sites containing radionuclides, in addition to heavy metals and organic solvents, because it can survive extremely high levels of radiation and repair its own radiation-damaged DNA. Understanding the genome sequence of *B. anthracis*, which causes anthrax, will promote faster detection methods and new treatments.

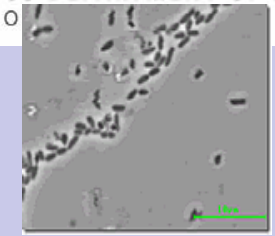
DOE JOINT GENOME INSTITUTE
US DEPARTMENT OF ENERGY
OFFICE OF SCIENCE

JGI Timeline

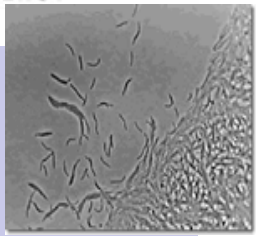


DOE JOINT GENOME INSTITUTE
US DEPARTMENT OF ENERGY

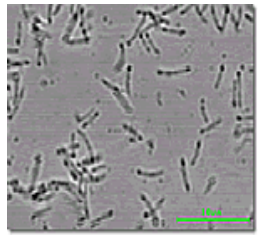
JGI 2001 Microbes (22 projects)



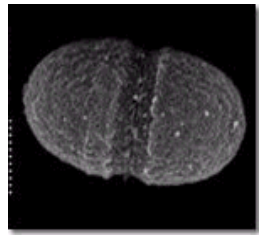
Burkholderia cepacia



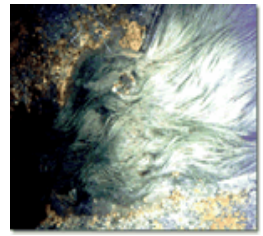
Cytophaga hutchinsonii



Desulfitobacterium halfniense

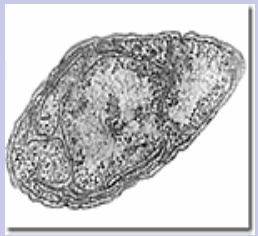


Enterococcus faecium



Ferroplasma acidarmanus

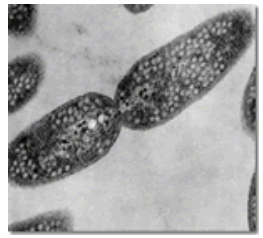
Magnetospirillum magnetotacticum



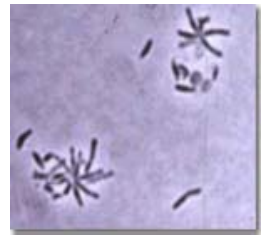
Nitrosomonas europaea

Prochlorococcus marinus

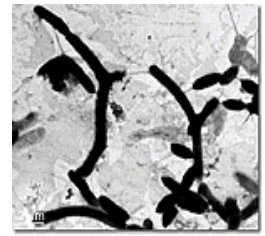
Pseudomonas fluorescens



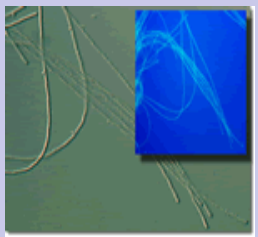
Rhodobacter sphaeroides



Rhodopseudomonas palustris



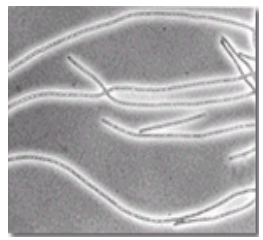
Sphingomonas aromaticivorans



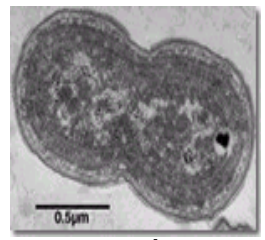
Thermomonospora fusca

Trichodesmium erythraeum

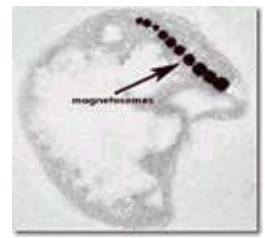
Xylella fastidiosa



Nostoc punctiforme



Marine synechococcus



Magnetococcus MC-1

DOE JOINT GENOME INSTITUTE
US DEPARTMENT OF ENERGY
OFFICE OF SCIENCE

More Microbes …

FY 2002
30 projects

Lactic acid bacteria

Lactobacillus gasseri (Klaenhammer)
Lactobacillus casei (Broadbent/Steele)
Lactobacillus delbrueckii (Steele)
Lactococcus cremoris (Weimer)
Brevibacterium linens (Weimer)
Pediococcus pentosaceus (Broadbent)
Oenoccoccus oeni (Mills)
Leuconostoc mesneteroides (Breidt)
Streptococcus thermophilus (Hutkins)

Bifidobacterium longum (O'Sullivan)

Complex polysaccharide degradation

Clostridium thermocellum (Wu)
Microbulbifer degradans (Weiner)
(complements white rot fungus sequence)

Phototrophic bacteria

Rhodospirillium rubrum (Roberts)
(complements *Rhodopseudomonas palustris* and *Rhodobacter spheroides)*

Toxic waste degradation and microbial ecology

Desulfuromonas acetoxidans (Lovely)
Desulfovibrio desulfuricans
Geobacter metallicreducens (Loveley, Ciufo)
Dechloromonas aromatica
Ralstonia eutropha (Valenzuela)
Azotobacter vinelandi
Trichodesmium erythraeum

Microbes in extreme environments

Psychrobacter (Thomashow)
Exiguobacterium (Thomashow)
Methanococcoides burtonii (Sowers, Cavicchioli)

Infectious diseases of plants and animals

Erlichia chaffeensis (Yu)
Erlichia canis (Yu)
Streptococcus suis (Gottschalk)
Haemophilus somnus (Inzana)
Pseudomonas syringae (Lindow)
Agrobacterium tumefaciens

Anaerobic methane oxidizing consortium "ball of bugs" (DeLong, Monterey Bay)
one (or two?!) reverse methanogenic archaea in core plus sulfur reducing bacterium on surface

DOE JOINT GENOME INSTITUTE
US DEPARTMENT OF ENERGY
OFFICE OF SCIENCE

And More Microbes…

2003

Single Microbes

- Rubrobacter xylanophilus
- Prochlorococcus isolate NATL2A
- Kineococcus radiotolerans sp nov
- Methylobacillus flagellatus, strain KT
- Synechococcus elongates PCC7 942
- Moorella thermoacetica ATCC39073
- Anabaena variabilis ATCC 29413
- Burkholderia complex (genomovar V)
- Crocosphaera watsonii WH8501

Fungus

- Trichoderma reesei - 87.55Mb of Sequence Present
- (Strain RUT-C30, ATCC56765)

Marine Algae

- Emiliania huxleyi strain 1516

Stramenopiles

•Phytophthera ramorum UCD Pr4 – 2.46Mb sequence

•Phytophthora sojae P6497 – 319.72Mb sequence

Microbial Consortia

•Acid mine drainage from site in Iron Mountain

•Chlorochromatium aggregatum

DOE JOINT GENOME INSTITUTE
US DEPARTMENT OF ENERGY
OFFICE OF SCIENCE

2004 DOE Microbial Projects

59 projects

8 species of Chlorobia
Chlorobium limicola, DSMZ 245(T)
Chlorobium phaeobacteroides, MN1
Prosthecochloris spp.
Prosthecochloris aestuarii, SK413/DSMZ 271(t)
Chlorobium vibrioforme f. thiosulfatophilum, DSMZ 265(T)
Chlorobium phaeobacteroides, DSMZ 266(T)
Pelodictyon phaeoclathratiforme, BU-1 (DSMZ 5477(T))
Pelodictyon luteolum, DSMZ 273(T)
Model Syntrophic Consortium:
Syntrophobacter fumaroxidans, MPOB
Syntrophomonas wolfei, Göttingen (DSM 2245B)
Methanospirillum hungateii, JF1
Facultative Metal-reducing Gamma proteobacteria
Shewanella putrifaciens, CN-32
Shewanella sp., PV-4
Shewanella amazonesis
Shewanella baltica, OS1155
Shewanella frigidimarina, NCMB400
*Shewanella denitrificans,OS **217***
Shewanella putrifaciens, 200
five bacteria involved in nitrification
Nirosomonas eutropha C71
Nitrosospira multiformis Surinam
Nitrosomonas oceani
Nitrobacter winogradskyi, Nb-255
Nitrobacter hamburgensis

Single microbes
Synthophobacter fumaroxidans
Synthophus acidotrophicus
Arthrobacter aurescens, TC1
Thermoanaerobacter ethanolicus, X514
Frankia sp., EAN1pec
Frankia sp., CcI3
Anaeromyxobacter dehalogenans, 2CP-C
Nocardioides sp., JS614
Deinococcus geothermalis, DSM11300
Chromohalobacter salexigens, DSM3043
Clostridium beijerincki, NCIMB 8052
Acidobacterium sp., Ellin6076
Clostridium phytofermentans
Arthrobacter sp., FB24
Thiomicrospira crunogena
Thiomicrospira denitrificans
Sphingopyxis alaskensis, RB2256
Alkaliphillus metalliredigenes
Jannaschina sp.CCS1
Roseobacter sp., TM1040
Paracoccus denitrificans, 1222
Thiobacillus denitrificans, ATCC 23644
b-proteobacterium sp., JS666
Eukaryotes
Glomus intraradices
Laccaria bicolor
Pichia stipitis, CBS 6054
Pichia mRNA for cDNA libraries
Communities:
200 BACs from anaerobic bioreactor granules
acid mine drainage community
Picoplankton BACS from HOTS site
Boiling thermal pool

sequencing
quencing Plans and Progress
Statistics
y Sequence Them?
Protocols
Community equencing Program
Genome Portal Site

Community Sequencing Program Sequencing Plans for 2005

Organism	Proposer	Affiliation	Sequence Allocated (MB)	Status
Microbes				
O. algarvensis symbionts	Dubilier	Max Planck Institute of Marine Microbiology	100	In production
Crenarchaeota	Delong	MIT	400	
Marinobacter aquaeolei	Edwards	Woods Hole Oceanographic Institution	176	In production
Staph. Aureus VISA strains	Tomasz	Rockefeller University	50	
Prochlorococcus	Chisholm	MIT	540	
Rhodocyclus-like polyphosphate accum.	Hugenholtz	JGI	150	
Rhodobacter	Kaplan	University of Texas, Houston	90	
Contaminated groundwater	Zhou	ORNL	80	
Lactobacillus reuteri (two strains)	Tannock	University of Otago, Dunedin, NZ	48	
Bacillus cereus (two strains)	Sorokin	INRA, France	80	
Basal Organisms				

In February **2004**, the JGI launched the Community Sequencing Program (CSP).

24 proposals accepted
10 microbial projects

February 25, 2005 – Deadline for receipt of applications for year 2006

Sequencing projects will be chosen based on scientific merit, judged through independent peer review. Criteria for participation in this program, the review process, and interactions between JGI and participants are outlined on the web site:

http://www.jgi.doe.gov/CSP/

DOE JOINT GENOME INSTITUTE
US DEPARTMENT OF ENERGY
OFFICE OF SCIENCE

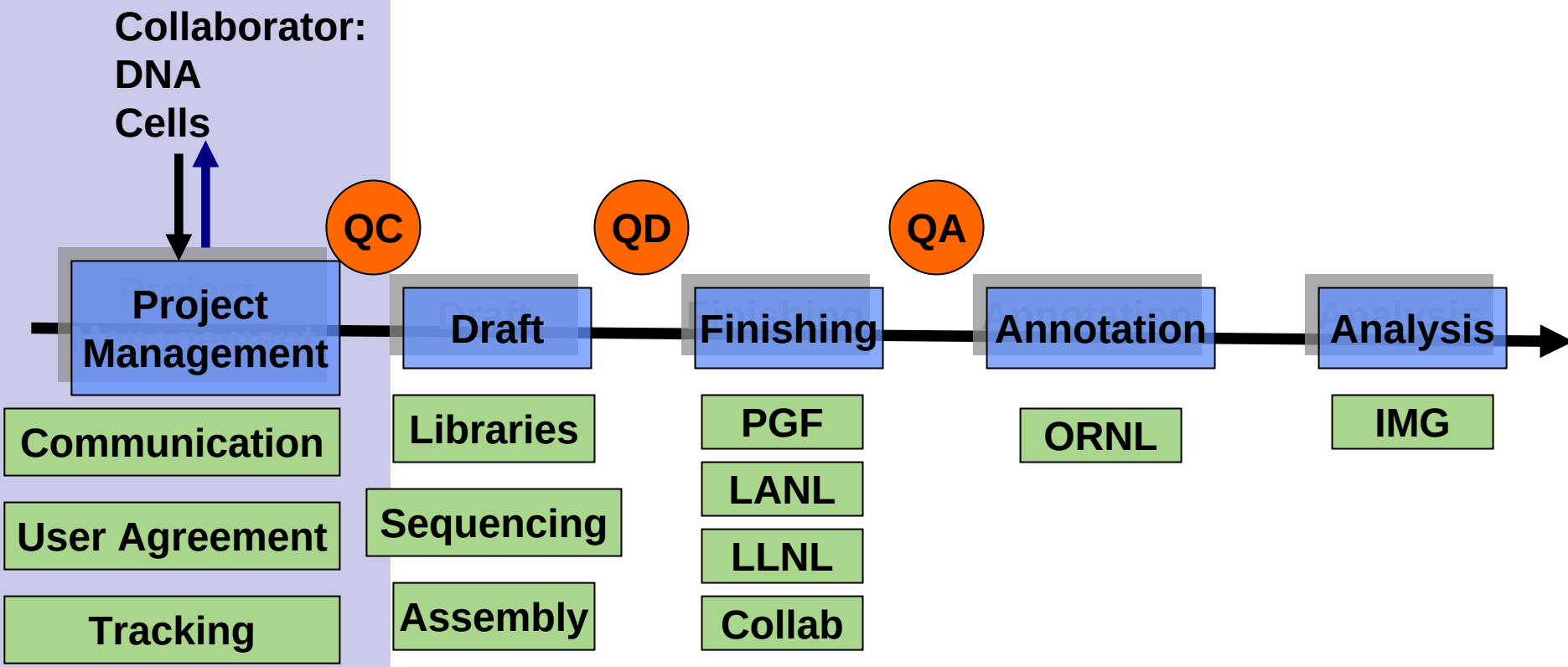
Microbial Programs

DOE Microbial Program
DOE GTL Program (GTL)
Community Sequencing Program (CSP)
JGI Internal Program

Goal: ***to provide the scientific community access to high throughput sequencing and to operate as a Genomic Infrastructure for American Science***

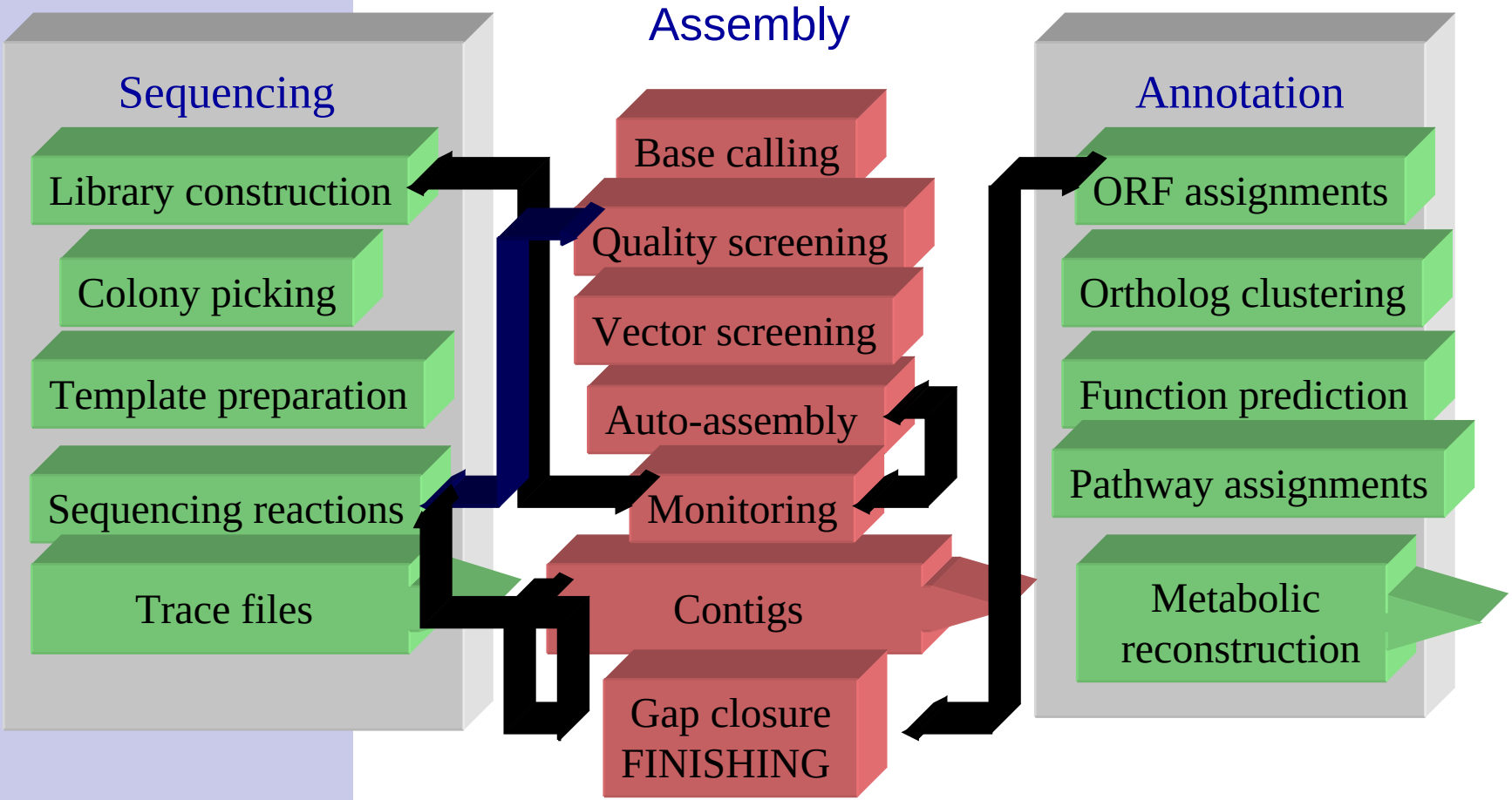
DOE JOINT GENOME INSTITUTE
US DEPARTMENT OF ENERGY
OFFICE OF SCIENCE

Program Structure



DOE JOINT GENOME INSTITUTE
US DEPARTMENT OF ENERGY
OFFICE OF SCIENCE

A typical genome sequencing project



Guidelines for Collaborators supplying Microbial DNA to the JGI for sequencing.

DNA sent to the JGI for library construction:

A. Required information for samples:

Name, email, and address of the PI of the Collaborating Laboratory.
Name, email, and address of the person who isolated the DNA.
Name, email and phone number of the person who packs and ships the material.
Itemized contents of the shipment.
Bacterial strain being shipped and culture conditions.
NCBI name and ATCC number of strain (if applicable).
Estimated genome size, and how this was determined. The preferred method is Pulsed Field Gel Electrophoresis.
Please include as much information as is known about the genome, including G+C content, and any available sequence information such as 16s, known repeat sequences, or any sample sequences or contigs available.
Also, please include the names and accession numbers of any sequenced microbial strains that share significant homology with the strain being sequenced at the JGI.

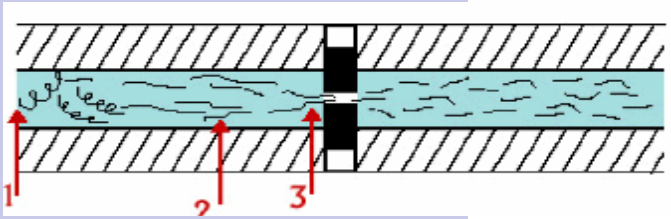
B. Web site information:

In addition, the JGI requests the following information which will be used to describe on going sequencing projects at the JGI.
Photograph-preferably EM with scale and credit.
Basic facts including why it is important and motivation behind sequencing.
Names, email, and addresses of collaborators to be listed on web site.
Any links or other resources about the microbe that can be listed on our site.

Library Construction:

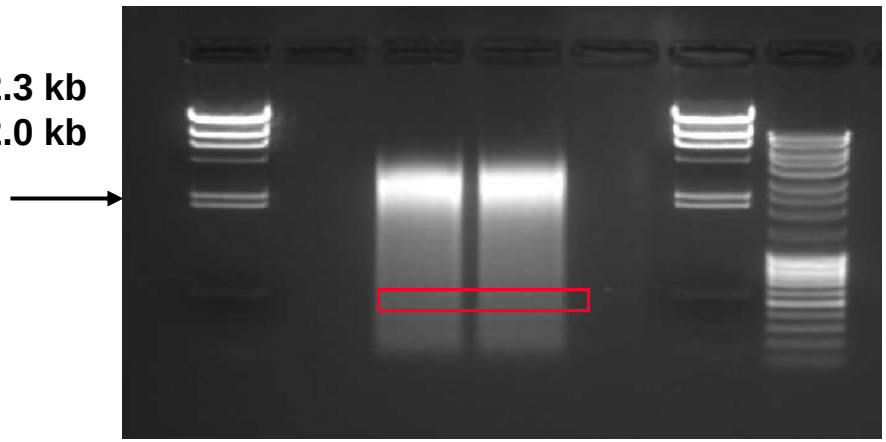
Multiple size insert libraries for each organism and sequence them to a specific depth.

- **4x Sequence of 2-4kbs – Small Insert**
- **4x Sequence of 8-10kbs – Medium Insert**
- **15x Clone coverage of Fosmid Ends**



GeneMachines Hydrashear

Sheared Genomic DNA



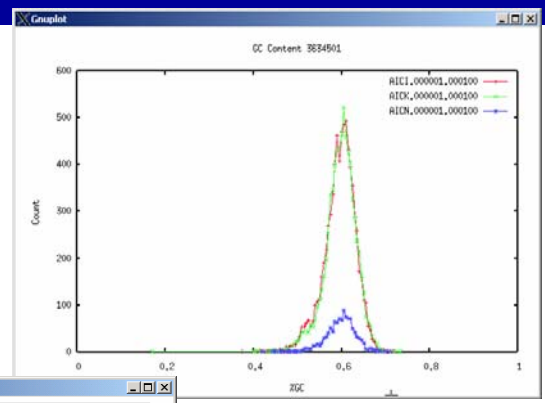
DOE JOINT GENOME INSTITUTE
US DEPARTMENT OF ENERGY
OFFICE OF SCIENCE

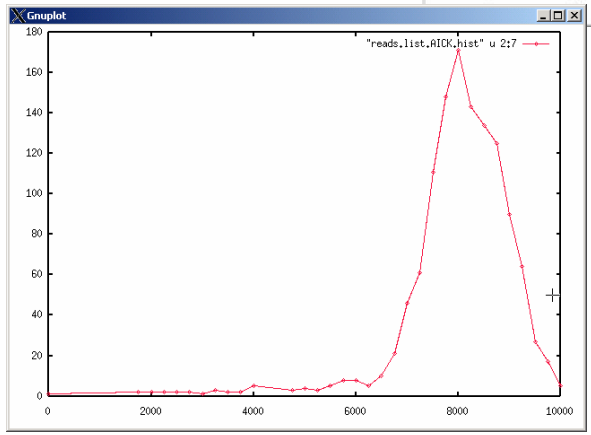
Library and Production QC

10 Plate QC

Project: 3634501
Organism: Roseobacter sp. TM1040
Lineage: cellular organisms; Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales; Rhodobacteraceae; Roseobacter
Vector: pUC, pMCL200
Shearing Operator: CC
Date: Jan 12, 2004

Library: AICI, AICK
Insert Size: 3kb,8kb





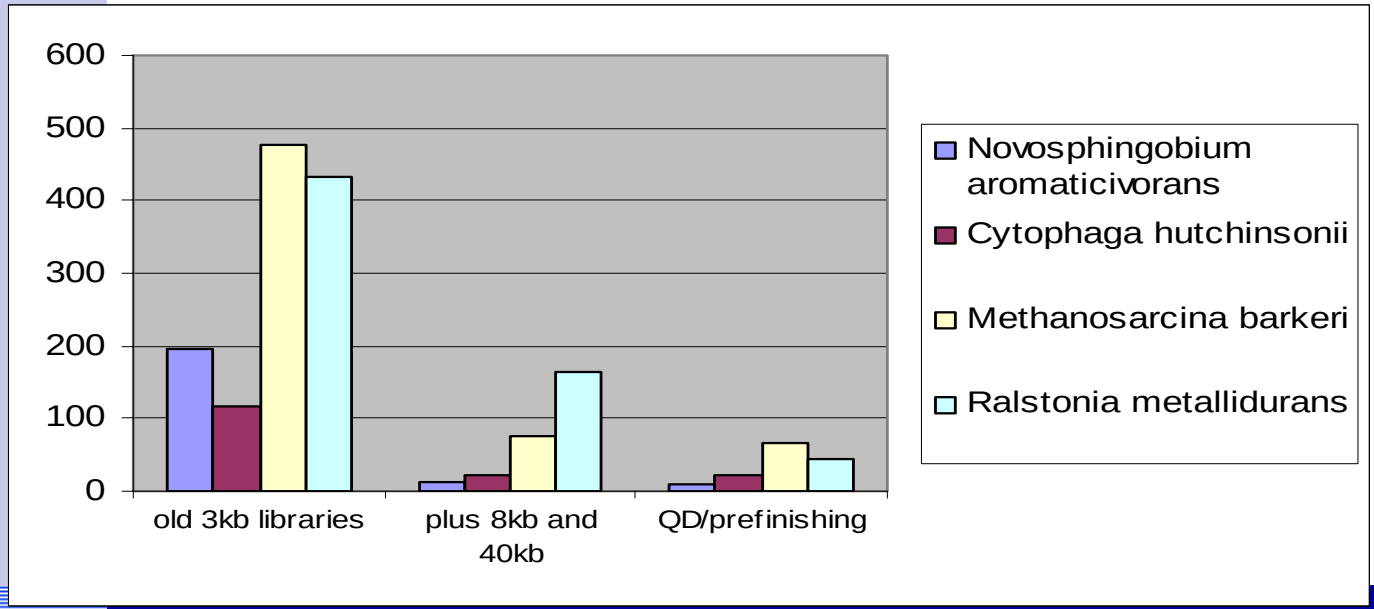
- Contamination Check
  - -Known JGI contaminants
  - -Vector
  - -GC content
  - -Correct microbe
- Library QC
  - -Read distribution
  - -Insert size distribution
  - -Compare to ideal assembly

Assembled: 3941181 (trimmed)
Phrap: 3489815
Current Depth Estimate : 4.180703 +/- 0.856594
About half the reads are in 105 contigs containing at least 37 reads each
N 50 analytical
N50 (analytic): About half the reads will be in 130 contigs containing at least 38 reads each (3.9 MB)
N50 (analytic): About half the reads will be in 756 contigs containing at least 7 reads each (9.0 MB)

DOE JOINT GENOME INSTITUTE
US DEPARTMENT OF ENERGY
OFFICE OF SCIENCE

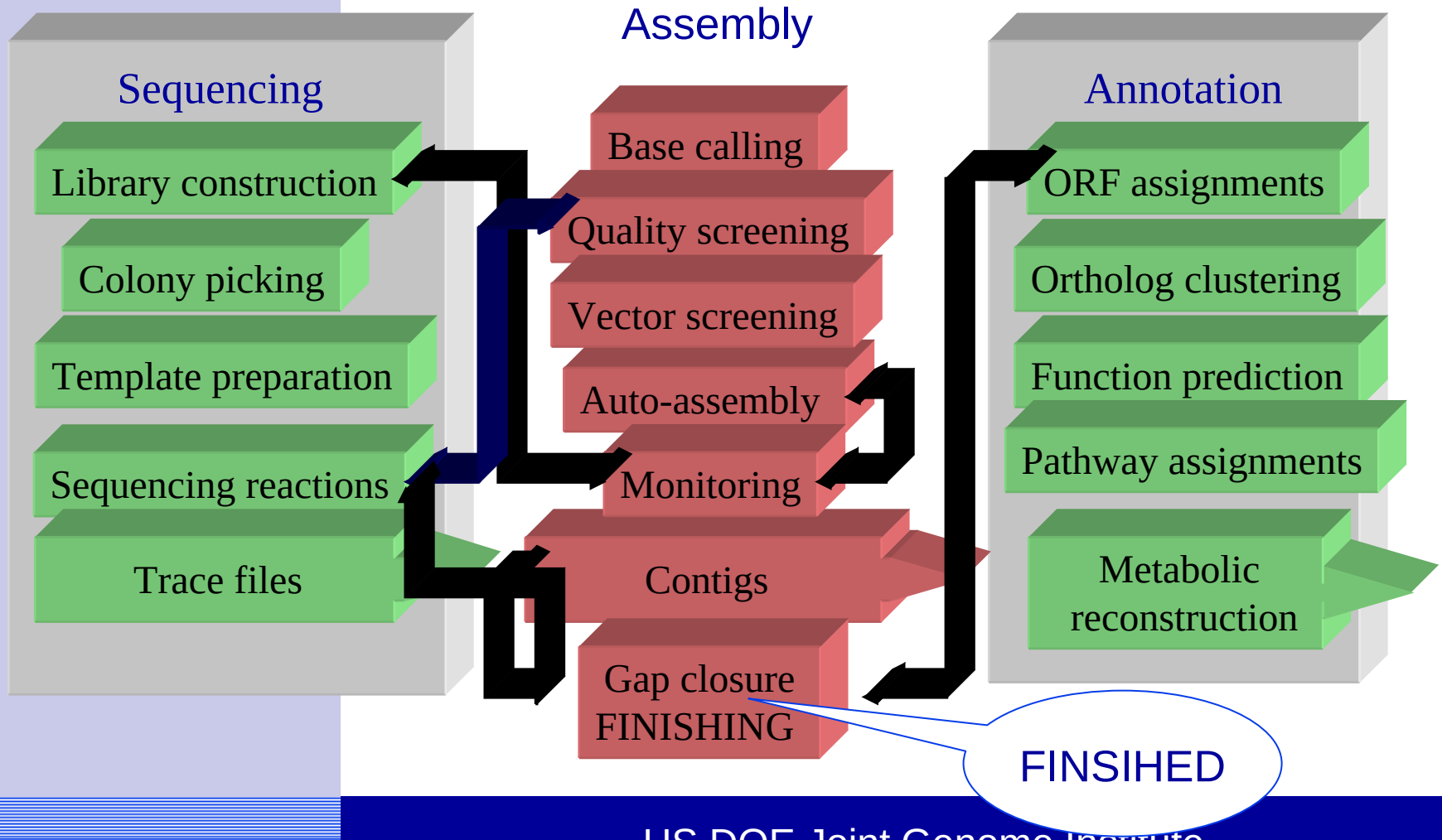
Effects of Improvements on Genome Quality

	old 3kb libraries		plus 8kb and 40kb		QD/prefinishing	
	Major Contigs	Genome size (MB)	Major Contigs	Genome size (MB)	Major Contigs	Genome size (MB)
Novosphingobium aromaticivorans	**197**	4.17	**13**	4.21	**9**	4.215
Cytophaga hutchinsonii	**118**	4.36	**23**	4.41	**22**	4.41
Methanosarcina barkeri	**478**	3.88	**77**	4.83	**67**	4.84
Ralstonia metallidurans	**432**	NA	**165**	6.83	**45**	6.83



DOE JOINT GENOME INSTITUTE
US DEPARTMENT OF ENERGY
OFFICE OF SCIENCE

A typical genome sequencing project



DOE JOINT GENOME INSTITUTE
US DEPARTMENT OF ENERGY
OFFICE OF SCIENCE

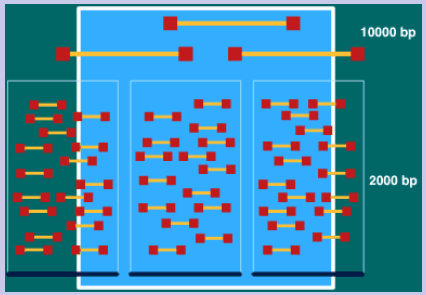
Finishing Standards

Guideline and finishing checklist for all JGI microbes:

- All low quality areas (<Q30) should be reviewed and re sequenced.
- Final error rate should be < 0.2 per 10 Kb.
- No single clone coverage, i.e. minimum of 2X depth everywhere.
- Manually inspect and quantify single stranded regions.
- Check all high quality discrepancies.
- Final sequence should have a base at every position (no strings of xxxx anywhere).
- Verify all repeats (paired ends and PCR if necessary).
- Make sure to check ends of final contigs (chromosomes, plasmids)
- Using Assembly Viewer and phrapViewer tools check correctness of final assembly. Confirm questionable areas with PCR.

DOE JOINT GENOME INSTITUTE
US DEPARTMENT OF ENERGY
OFFICE OF SCIENCE

Assembly, Finishing and Annotation

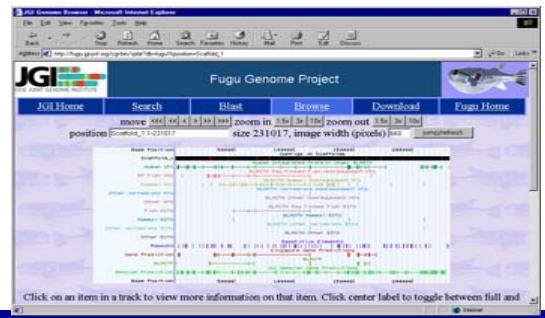


Draft Assembly
All libraries sequenced to completion, data assembled and verified

Draft assembly and annotation available to the collaborator

Project transfers to Microbial Finishing Groups for complete finishing

Annotation



Provide genome and tools to community

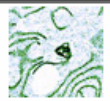
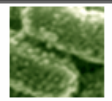
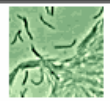
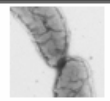
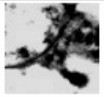
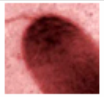
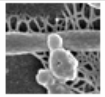
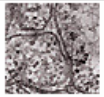
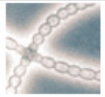
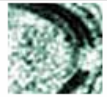
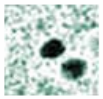
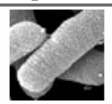
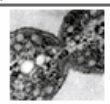
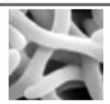
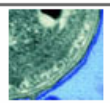
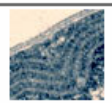
Search Favorites Media Back
Address http://genome.jgi-psf.org/microbial/index.html

http://genome.jgi-psf.org/microbial/index.html

DOE JOINT GENOME INSTITUTE
US DEPARTMENT OF ENERGY
OFFICE OF SCIENCE

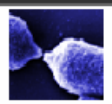
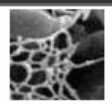
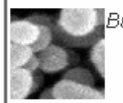
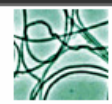
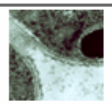
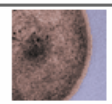
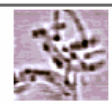
Microbial Genomics

Finished Genomes

 *Anabaena variabilis*	*Burkholderia xenovorans*	 *Chlorochromium aggregatum*	 *Cytophaga hutchinsonii*	 *Desulfovibrio desulfuricans*	*Ehrlichia canis*
 *Haemophilus somnus*	 *Methylobium petroleophilum*	 *Microbulbifer degradans*	 *Nitrosomonas europaea*	 *Nostoc punctiforme*	 *Prochlorococcus marinus MED4*
Prochlorococcus sp. MIT9312	 *Prochlorococcus marinus MIT9313*	 *Pseudomonas syringae*	*Psychrobacter*	 *Rhodobacter sphaeroides*	*Rhodopseudomonas palustris*
 *Rhodospirillum rubrum*	 *Synechococcus WH8102*	 *Synechococcus elongatus*	*Thiobacillus denitrificans*		

31 genomes finished

Draft Genomes

 *Azotobacter vinelandii*	 *Bifidobacterium longum*	*Brevibacterium linens*	 *Burkholderia vietnamensis*	*Burkholderia strain 383*	 *Chloroflexus aurantiacus*
Chromohalobacter	*Clostridium*	 *Crocosphaera*	*Dechloromonas*	 *Deinococcus*	 *Desulfitobacterium*

- >100 genomes drafted
- 66 posted

Internet

File Edit View Favorites Tools Help

Search Favorites Media Back Links

Address http://durian.jgi-psf.org/img_w_v10a/main.cgi Go

JGI DOE JOINT GENOME INSTITUTE US DEPARTMENT OF ENERGY OFFICE OF SCIENCE

Search for Organisms (Examples): GO

img INTEGRATED MICROBIAL GENOMES

IMG Home | Gene Search | Organism Browser | Phylogenetic Profiler | Organism Stats | Gene Cart | Preferences

About IMG | Data Sources
Related Links

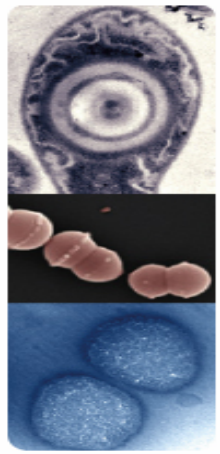
IMG Organisms

finished/draft	JGI	Total
Bacteria	10/8	168/8
Archaea	0/1	19/1
Eukaryota	0/0	9/0
All Organisms	10/9	196/9

Version 1.0alpha, January 3, 2005
For questions or comments, contact *eszeto@lbl.gov*
©2005 The Regents of the University of California
Disclaimer

The **Integrated Microbial Genomes (IMG)** system provides a framework for comparative analysis of the genomes sequenced by the Joint Genome Institute. Its goal is to facilitate the visualization and exploration of genomes from a functional and evolutionary perspective. There are two main approaches to using the system: genes and genomes. Genes can be found by keyword or BLAST query via the Gene Search and then analyzed in the context of their chromosomal, metabolic, and phylogenetic neighborhoods. Genomes of multiple organisms can be selected with the Organism Browser. You can then compare them through the Phylogenetic Profiler in order to identify common or unique properties. The first release of IMG includes 19 JGI genomes together with most publicly available complete microbial genomes.

Note: cookies should be enabled in your browser. A fast internet connection also helps.



Internet

Total genomes in IMG v1.0 - 296: 195 bacterial, 20 archaea, 9 eukaryote
101 bacterial genomes are from JGI

DOE JOINT GENOME INSTITUTE
US DEPARTMENT OF ENERGY
OFFICE OF SCIENCE

Background: JGI Microbial Genome Data

img

- **Sequenced at JGI (PGF) - 101**
- **Finishing by JGI partners and collaborators - 31**
- **Annotated automatically at ORNL**

  ORF calling: Glimmer, Critica, Generation (multi evidence)

  \+ COG, Pfam, EC, SwissProt, KEGG, InterPro, NCBI NR

- **Public Repositories**

  JGI microbes eventually go into RefSeq, EBI-GR, CMR, etc.

- **Released via JGI individual portals**

  Supports individual genome data search (BLAST), download, annotation viewer (ORNL).

IMG Data Exploration: Taxon View

DOE JOINT GENOME INSTITUTE

img INTEGRATED MICROBIAL GENOMES

All Organisms selected

IMG Home | Gene Search | Organism Browser | Phylogenetic Profiler | Organism Stats | Gene Cart | Preferences

About IMG | Data Sources | Related Links

Organism Browser

[View As Phylogenetic Tree]

Save Selections | Select All | Clear All

hint: Selections do not take effect until you save them.

Select	Organism Name
☑	Acinetobacter sp. (strain ADP1)
☑	Aeropyrum pernix (strain K1)
☑	Agrobacterium tumefaciens (strain C58 / ATCC 33970 / ATCC 33970 [Cereon])
☑	Agrobacterium tumefaciens (strain C58 / ATCC 33970 / ATCC 33970 [Dupont])
☑	Anabaena sp. (strain PCC 7120)
☑	Aquifex aeolicus (strain VF5)
☑	Archaeoglobus fulgidus (strain DSM 4304 / VC-16 / ATCC 49558)
☑	Bacillus anthracis (strain Ames / isolate 0581)
☑	Bacillus anthracis (strain Ames / isolate Porton)
☑	Bacillus anthracis (strain Sterne) (JGI)
☑	Bacillus cereus (strain ATCC 10987)
☑	Bacillus cereus (strain ATCC 14579 / DSM 31)
☑	Bacillus halodurans (strain C-125 / JCM 9153)
☑	Bacillus licheniformis (strain DSM 13 / ATCC 14580 [Novozymes])

img INTEGRATED MICROBIAL GENOMES

All Organisms selected

IMG Home | Gene Search | Organism Browser | Phylogenetic Profiler | Organism Stats | Gene Cart | Preferences

About IMG | Data Sources | Related Links

Organism Detail

Taxon Name	Acinetobacter sp. (strain ADP1)
Taxon ID	62977
Lineage	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Moraxellaceae; Acinetobacter; calcoaceticus
Sequencing Status	Finished
Sequencing Center	Genoscope
Funding Agency	
Finishing Group	
Comment	

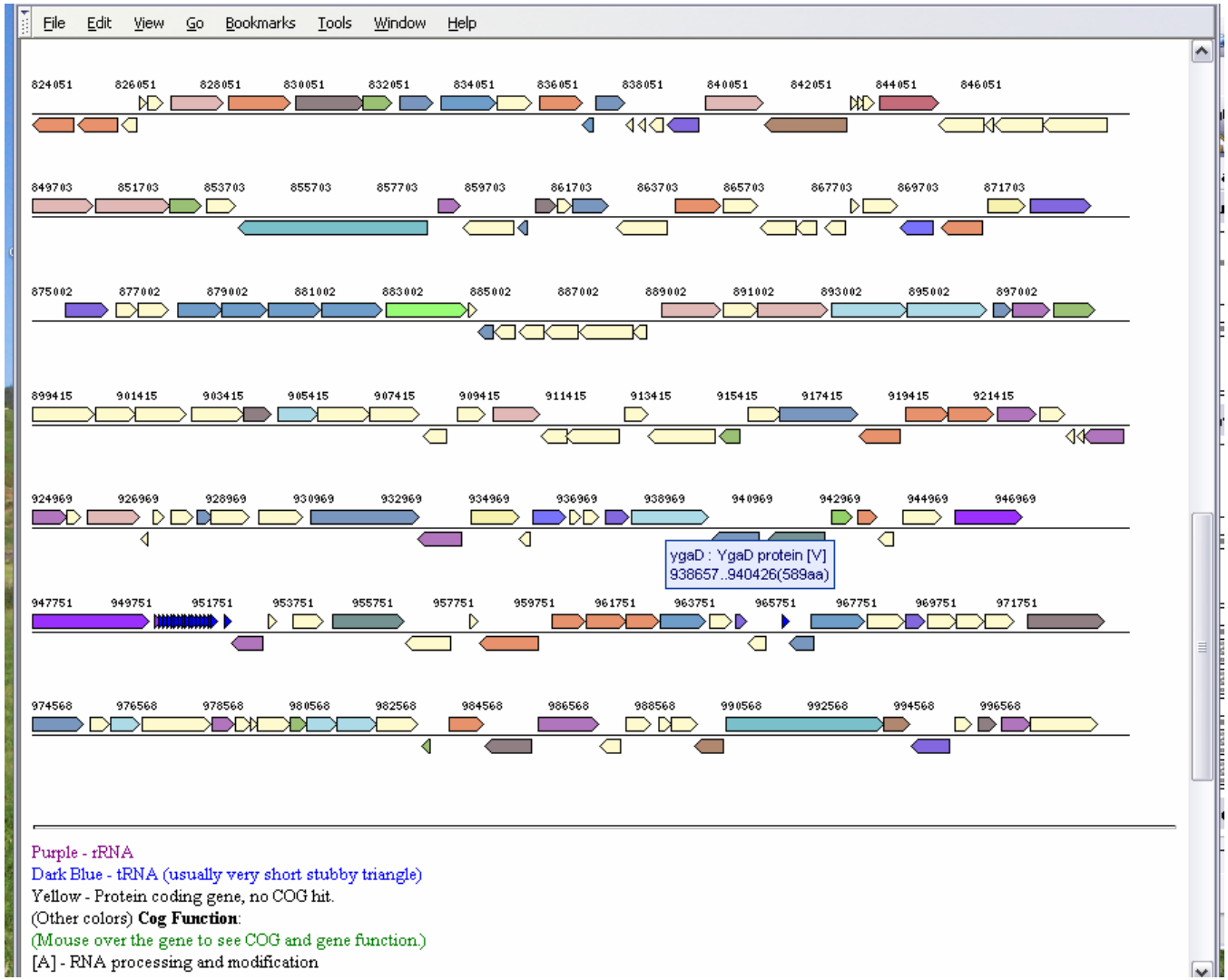
Statistics

	Number	% of Total
DNA, total number of bases	3598621	100.00%
DNA coding number of bases	3197442	88.85%
DNA G+C number of bases	1454907	40.43%
DNA scaffolds	1	100.00%
Genes total number	3422	100.00%
Protein Coding Genes	3325	97.17%
RNA Genes	97	2.83%
Genes with function prediction	2087	60.99%
Genes without function prediction	1238	36.18%
Genes w/o function with similarity	1107	32.35%
Genes w/o function w/o similarity	131	3.83%
Genes connected to KEGG pathways	752	21.98%
Genes not connected to KEGG pathways	2573	75.19%
Genes in Ortholog Clusters	2927	85.53%
Genes in Paralog Clusters	922	26.94%
Genes in COGs	1788	52.25%
Genes in Pfam	1521	44.45%
Genes in InterPro	2531	73.96%
Pfam Clusters	1091	15.04%
Clusters, orthologous groups	2796	3.41%

Scaffolds and Contigs

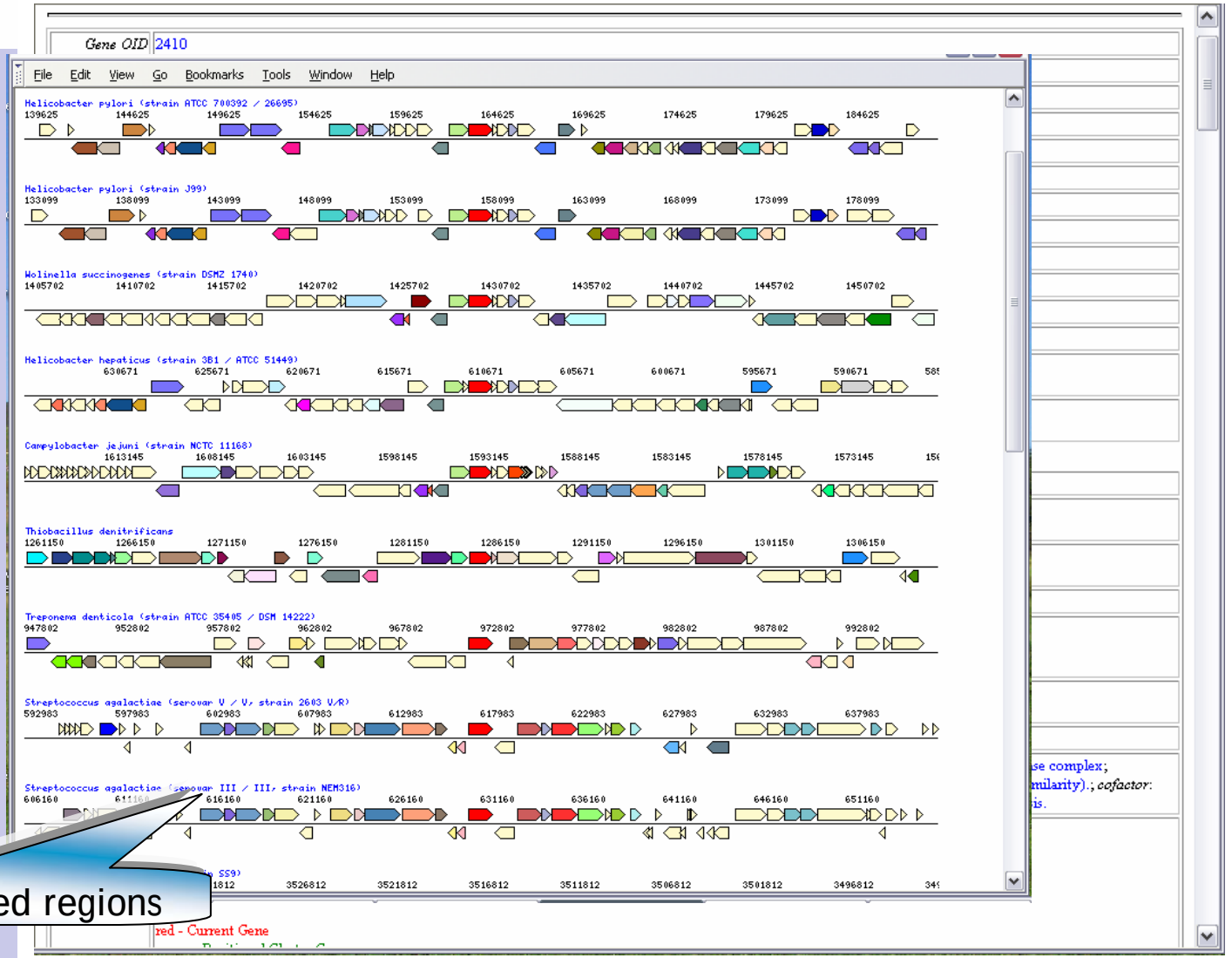
DOE JOINT GENOME INSTITUTE
US DEPARTMENT OF ENERGY
OFFICE OF SCIENCE

IMG Data Exploration: Chromosome View



DOE JOINT GENOME INSTITUTE
US DEPARTMENT OF ENERGY
OFFICE OF SCIENCE

IMG Data Exploration: Gene Page



Other taxon conserved regions

Sequence Similarities

Orthologs (Bi-directional best hits of each organism against each other organism)

Taxonomic Domains(D): B = bacteria, A = archaea, E = eukarylota.

Select	Ortholog	Name	Percent Identity	Subject Alignment	Subject Length	E-value	Bit Score	Cons. Region Score	(D)	Taxon Name
☐	248710	Enolase	97.65		426aa	0.0e+00	821.2	9	B	Helicobacter pylori (strain J99)
☐	2366830	Enolase	71.23		425aa	0.0e+00	612.1	5	B	Helicobacter hepaticus (strain 3B1 / ATCC 51449)
☐	5040780	PUTATIVE ENOLASE	71.39		422aa	0.0e+00	608.6	5	B	Wolinella succinogenes (strain DSMZ 1740)
☐	3032250	Enolase	63.27		414aa	0.0e+00	517.7	3	B	Campylobacter jejuni (strain NCTC 11168)
☐	3761080	Enolase	61.02		429aa	0.0e+00	500.4	0	B	Bacillus halodurans (strain C-125 / JCM 9153)
☐	2573440	Enolase	61.17		422aa	0.0e+00	497.7	0	B	Thermus thermophilus (strain HB27 / ATCC BAA-163 / DSM 7039)
☐	991030	Enolase	57.65		429aa	0.0e+00	496.5	0	B	Thermoanaerobacter tengcongensis (strain MB4 / JCM 11007)
☐	25850	Enolase	59.71		427aa	0.0e+00	490	0	B	Thermotoga maritima (strain ATCC 43589 / DSM 3109 / MSB8)
☐	200202130	Enolase	58.8		429aa	0.0e+00	489.6	0	B	Magnetococcus sp. MC-1
☐	200085380	Enolase	57.65		430aa	0.0e+00	487.6	0	B	Exiguobacterium
☐	1804620	Enolase	57.92		431aa	0.0e+00	486.5	0	B	Clostridium tetani (strain Massachusetts / E88)
☐	2439700	Enolase	57.41		428aa	0.0e+00	486.5	0	B	Geobacter sulfurreducens (strain ATCC 51573 / PCA)
☐	118500	Enolase	57.08		426aa	0.0e+00	483.4	0	B	Aquifex aeolicus (strain VF5)
☐	200007290	Enolase	57.08		422aa	0.0e+00	483.4	0	B	Deinococcus geothermalis, DSM11300
☐	4163040	Enolase	57.48		429aa	0.0e+00	481.9	0	B	Oceanobacillus iheyensis (strain DSM 14371 / JCM 11309 / HTE831)
☐	200370250	Enolase	57.88		429aa	0.0e+00	481.9	0	B	Ralstonia metallidurans
☐	3445080	Enolase	57.65		427aa	0.0e+00	481.1	0	B	Ralstonia solanacearum (strain GMI1000)
☐	283750	Enolase	58.01		428aa	0.0e+00	480.7	0	B	Neisseria meningitidis (serogroup B / B, strain MC58)

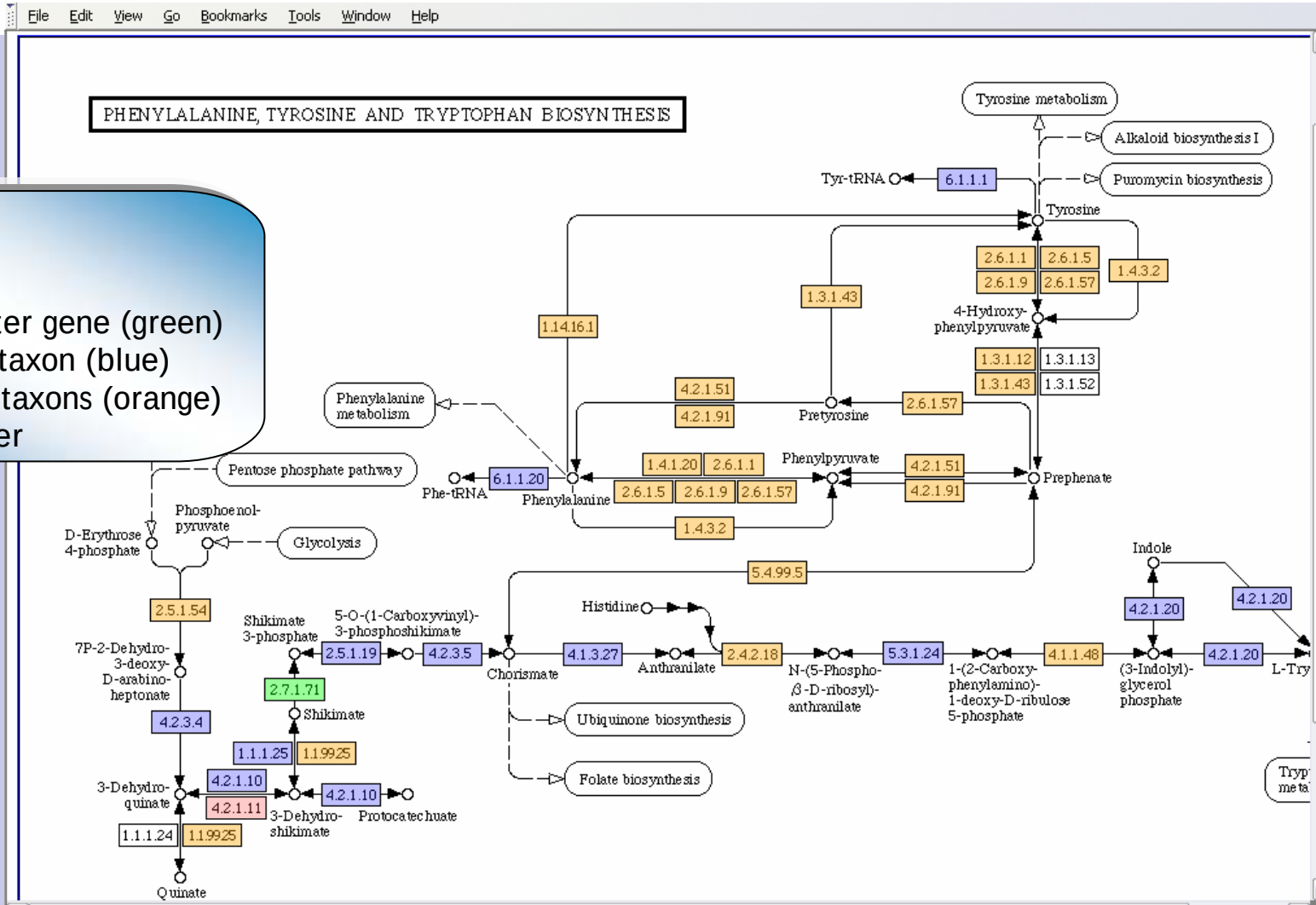
[Show Unidirectional Hit Sequence Similarities]

Add Selections To Gene Cart | Select All | Clear All Selections

IMG Data Exploration: KEGG Map Viewer

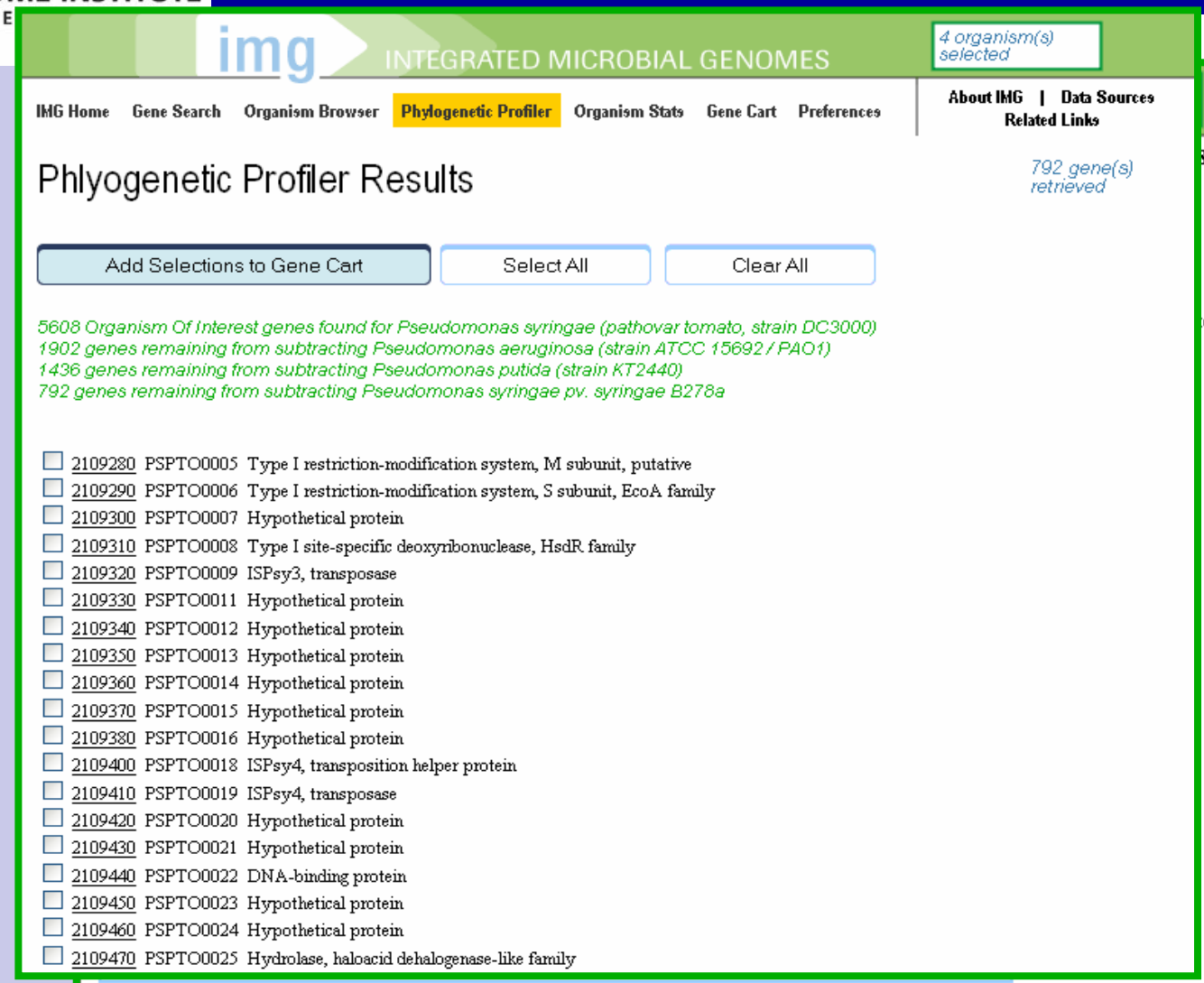
DOE JOINT GENOME INSTITUTE
US DEPARTMENT OF ENERGY
OFFICE OF SCIENCE

Genes:
- current (red)
- positional cluster gene (green)
- genes in same taxon (blue)
- genes in other taxons (orange) via EC number



DOE JOINT GENOME INSTITUTE
US DEPARTMENT OF E
OFFICE OF SCIENCE

IMG Data Exploration: Profile Search



DOE JOINT GENOME INSTITUTE
US DEPARTMENT OF ENERGY
OFFICE OF SCIENCE

IMG Data Exploration: Gene Search

Gene Search

Keyword:

Filters: Gene Name

Examples

Go Reset

BLASTp

Find matches in selected organisms (all IMG organisms if none are selected).
Protein vs. protein, paste amino acid sequence here:

E-value: 1e-5

Database: All IMG Genes - One large Database

Go Reset

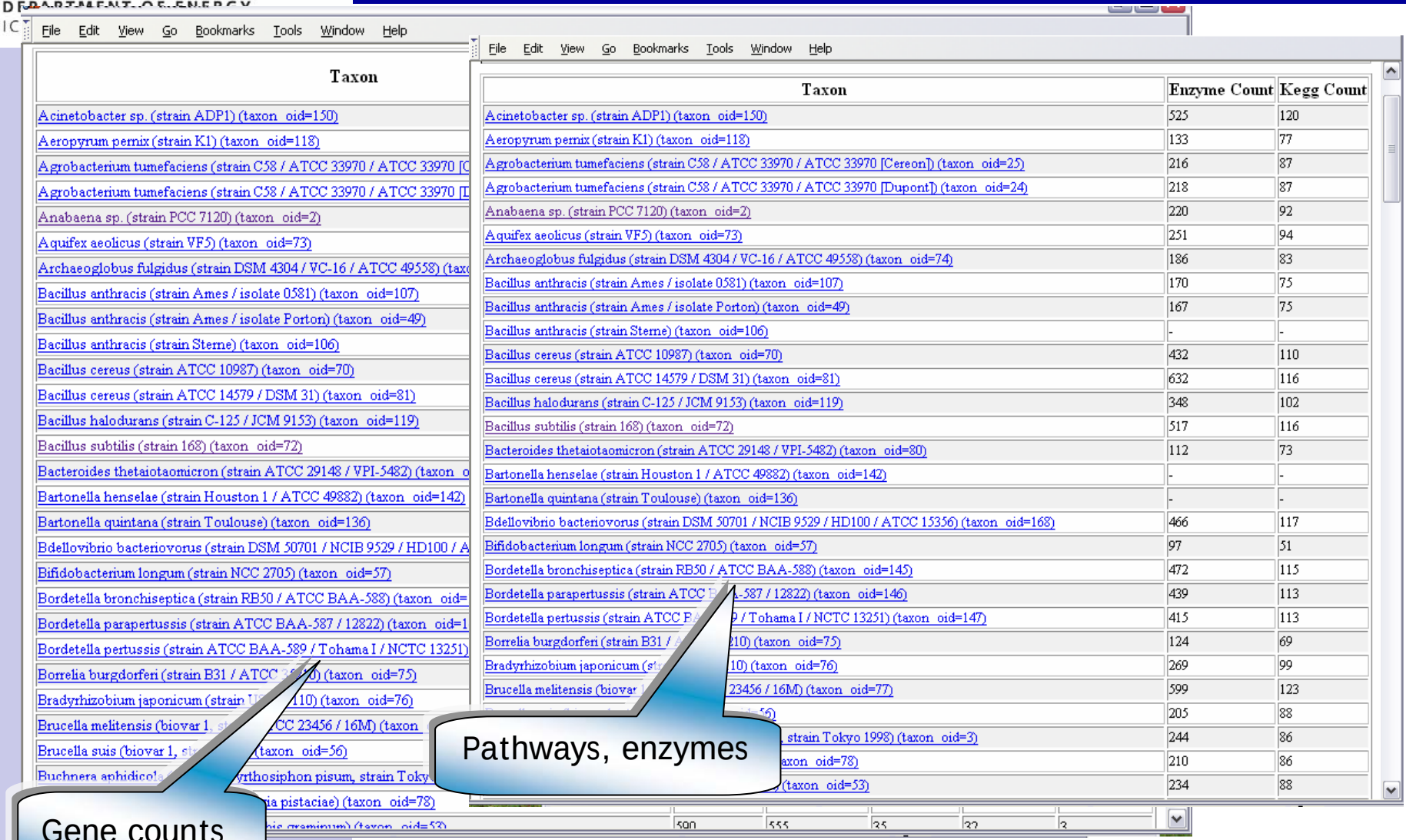
hint: *– BLAST may be slow for large protein queries.*
– The optimum E-value will depend on the size of the BLAST database you select. (For a larger database, use a larger E-value cutoff.)

DOE JOINT GENOME INSTITUTE

US DEPARTMENT OF ENERGY

OFFICE

IMG Data Exploration: Statistics



File Edit View Go Bookmarks Tools Window Help

Taxon
Acinetobacter sp. (strain ADP1) (taxon_oid=150)
Aeropyrum pernix (strain K1) (taxon_oid=118)
Agrobacterium tumefaciens (strain C58 / ATCC 33970 / ATCC 33970 [C
Agrobacterium tumefaciens (strain C58 / ATCC 33970 / ATCC 33970 [D
Anabaena sp. (strain PCC 7120) (taxon_oid=2)
Aquifex aeolicus (strain VF5) (taxon_oid=73)
Archaeoglobus fulgidus (strain DSM 4304 / VC-16 / ATCC 49558) (taxo
Bacillus anthracis (strain Ames / isolate 0581) (taxon_oid=107)
Bacillus anthracis (strain Ames / isolate Porton) (taxon_oid=49)
Bacillus anthracis (strain Sterne) (taxon_oid=106)
Bacillus cereus (strain ATCC 10987) (taxon_oid=70)
Bacillus cereus (strain ATCC 14579 / DSM 31) (taxon_oid=81)
Bacillus halodurans (strain C-125 / JCM 9153) (taxon_oid=119)
Bacillus subtilis (strain 168) (taxon_oid=72)
Bacteroides thetaiotaomicron (strain ATCC 29148 / VPI-5482) (taxon_o
Bartonella henselae (strain Houston 1 / ATCC 49882) (taxon_oid=142)
Bartonella quintana (strain Toulouse) (taxon_oid=136)
Bdellovibrio bacteriovorus (strain DSM 50701 / NCIB 9529 / HD100 / A
Bifidobacterium longum (strain NCC 2705) (taxon_oid=57)
Bordetella bronchiseptica (strain RB50 / ATCC BAA-588) (taxon_oid=
Bordetella parapertussis (strain ATCC BAA-587 / 12822) (taxon_oid=1
Bordetella pertussis (strain ATCC BAA-589 / Tohama I / NCTC 13251)
Borrelia burgdorferi (strain B31 / ATCC [illegible]) (taxon_oid=75)
Bradyrhizobium japonicum (strain [illegible]) (taxon_oid=76)
Brucella melitensis (biovar 1, [illegible]) (taxon
Brucella suis (biovar 1, [illegible]) (taxon_oid=56)
Buchnera aphidicol[illegible] [illegible]) (taxon_oid=78)
[illegible] (taxon_oid=53)

File Edit View Go Bookmarks Tools Window Help

Taxon	Enzyme Count	Kegg Count
Acinetobacter sp. (strain ADP1) (taxon_oid=150)	525	120
Aeropyrum pernix (strain K1) (taxon_oid=118)	133	77
Agrobacterium tumefaciens (strain C58 / ATCC 33970 / ATCC 33970 [Cereon]) (taxon_oid=25)	216	87
Agrobacterium tumefaciens (strain C58 / ATCC 33970 / ATCC 33970 [Dupont]) (taxon_oid=24)	218	87
Anabaena sp. (strain PCC 7120) (taxon_oid=2)	220	92
Aquifex aeolicus (strain VF5) (taxon_oid=73)	251	94
Archaeoglobus fulgidus (strain DSM 4304 / VC-16 / ATCC 49558) (taxon_oid=74)	186	83
Bacillus anthracis (strain Ames / isolate 0581) (taxon_oid=107)	170	75
Bacillus anthracis (strain Ames / isolate Porton) (taxon_oid=49)	167	75
Bacillus anthracis (strain Sterne) (taxon_oid=106)	-	-
Bacillus cereus (strain ATCC 10987) (taxon_oid=70)	432	110
Bacillus cereus (strain ATCC 14579 / DSM 31) (taxon_oid=81)	632	116
Bacillus halodurans (strain C-125 / JCM 9153) (taxon_oid=119)	348	102
Bacillus subtilis (strain 168) (taxon_oid=72)	517	116
Bacteroides thetaiotaomicron (strain ATCC 29148 / VPI-5482) (taxon_oid=80)	112	73
Bartonella henselae (strain Houston 1 / ATCC 49882) (taxon_oid=142)	-	-
Bartonella quintana (strain Toulouse) (taxon_oid=136)	-	-
Bdellovibrio bacteriovorus (strain DSM 50701 / NCIB 9529 / HD100 / ATCC 15356) (taxon_oid=168)	466	117
Bifidobacterium longum (strain NCC 2705) (taxon_oid=57)	97	51
Bordetella bronchiseptica (strain RB50 / ATCC BAA-588) (taxon_oid=145)	472	115
Bordetella parapertussis (strain ATCC [illegible]-587 / 12822) (taxon_oid=146)	439	113
Bordetella pertussis (strain ATCC [illegible] / Tohama I / NCTC 13251) (taxon_oid=147)	415	113
Borrelia burgdorferi (strain B31 / [illegible]) (taxon_oid=75)	124	69
Bradyrhizobium japonicum (st[illegible]) (taxon_oid=76)	269	99
Brucella melitensis (biovar [illegible] 23456 / 16M) (taxon_oid=77)	599	123
[illegible]	205	88
[illegible], strain Tokyo 1998) (taxon_oid=3)	244	86
[illegible]axon_oid=78)	210	86
[illegible] (taxon_oid=53)	234	88

Pathways, enzymes

Gene counts

DOE JOINT GENOME INSTITUTE
US DEPARTMENT OF ENERGY
OFFICE OF SCIENCE

General Goals

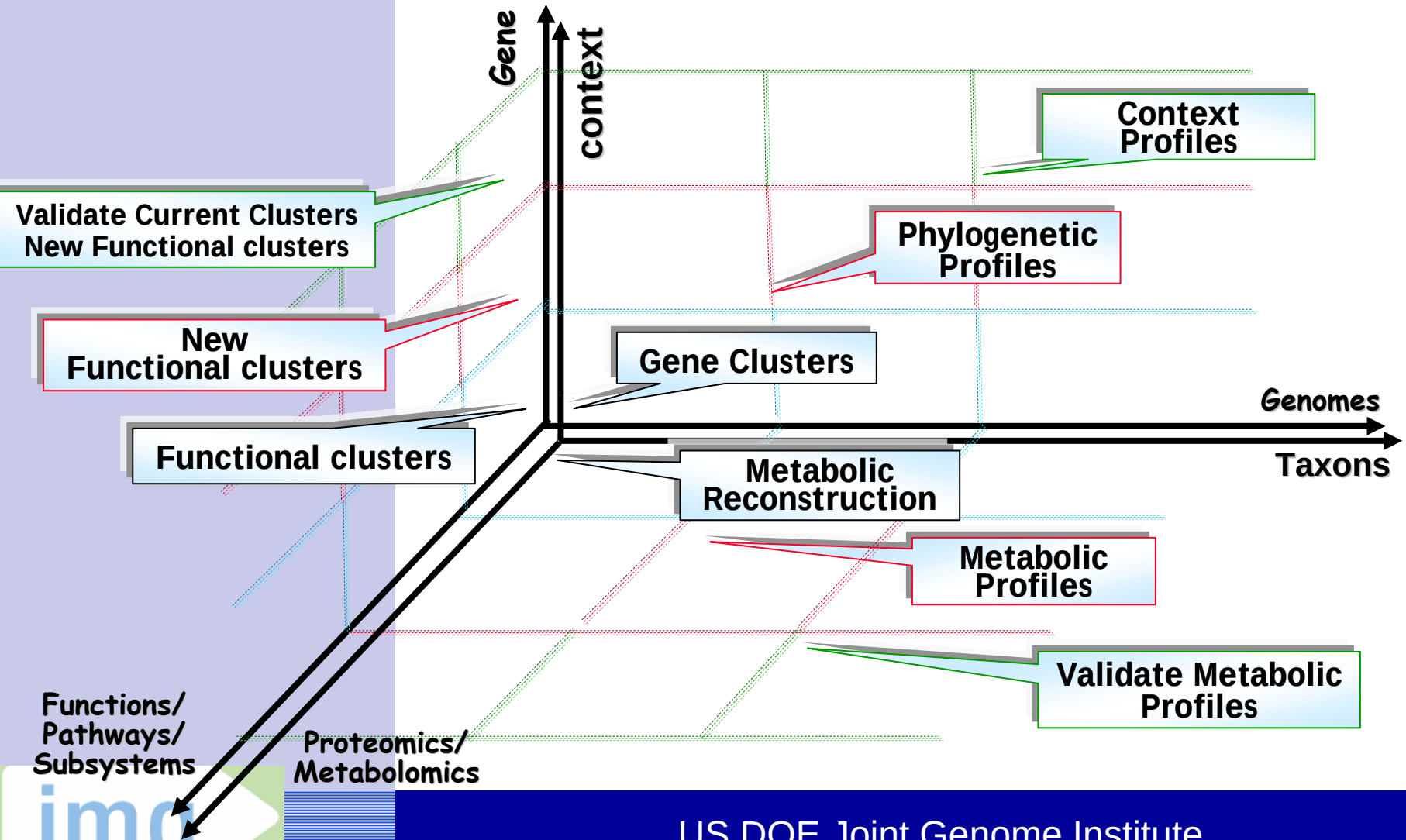
You cannot analyze one genome in isolation

- **Provide a data management system that will:**
  - **support the analysis of all genomes**
  - **support community annotation of genomic sequences**
  - **support modeling cellular networks**
  - **become a platform for understanding genomes**
- **Provide a data management system that** enables scientists to explore microbial community genomic data in the context of relevant environmental, geographical, geochemical and phylogenetic data

DOE JOINT GENOME INSTITUTE
US DEPARTMENT OF ENERGY
OFFICE OF SCIENCE

Genome Analysis Data Types

Levels and Dimensions



The US Department of Energy Joint Genome Institute Microbial Genome Program

DOE JOINT GENOME INSTITUTE
US DEPARTMENT OF ENERGY
OFFICE OF SCIENCE

Program director Paul Richardson

Production (PGF)
Susan Lucas
Tijana Glavina
Miranda Harmon
Nancy Hammon
Sanjay Israni
Dan Baker
Chris Daum
Marty Pollard
Simon Roberts
Alex Copeland,
Kerrie Barry
Jenny White
Chris Detter
Mariana Anaya
Corey Chinn
Eileen Dalin
Doug Smith
Hope Tice

Microbial Finishing

Alla Lapidus (PGF)
Eugene Goltsman
Genevieve DiBartolo
Michele Martinez
James Thiel
Steve Lowry
Stephan Trong

Patrick Chain (LLNL)
Stephanie Malfatti
Lisa Vergez
Maria Shin

Paul Gilna (LANL)
Cliff Han
Thomas Brettin
Roxanne Tapia
Olga Chertkov
David Bruce
Lynne Goodwin
Liz Saunders
Chris Munk
John J. Fawcett

Eddy Rubin
Jim Bristow

ORNL(annotation)
Frank Larimer
Miriam Land

Stanford (QC)
Jeremy Schmutz
Jane Grimwood

Webportal/Ftp (PGF)
Sam Pitluck
Linda Peters
Leila Hornick

IMG/annotation (PGF)
Nikos Kyrpides
Victor Markowitz
Natalia Ivanova
Thanos Lykidis
Phil Hugenholtz
Anu Padki
Amber Shao
Xueling Zhao
Greg Werner
Annette Greiner
Kristin Taylor
David Goodstein
Dan Rokshar
Ernest Szeto
Krishna Palaniappan
Frank Korzeniewski

DOE JOINT GENOME INSTITUTE
US DEPARTMENT OF ENERGY
OFFICE OF SCIENCE

This work was performed under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program and by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48, Lawrence Berkeley National Laboratory under contract No. DE-AC03-76SF00098 and Los Alamos National Laboratory under contract No. W-7405-ENG-36.