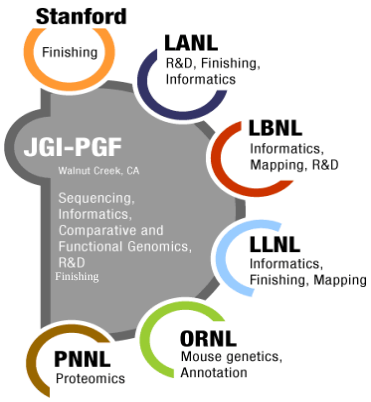
Advancing Science with DNA Sequence

The US DOE Joint Genome Institute Microbial Genome Program

Alla Lapidus, Patrick Chain, Cliff Han, Thomas Brettin, Alex Copeland, Chris Detter, Samuel Pitluck, Tijana Glavinadelrio, Susan Lucas, Kerrie Berry, Miriam Land, Frank Larimer, Nikos Kyrpides, Natalia Ivanova, Jeremy Schmutz, Victor M Markowitz, David Bruce, Paul Gilna, Jim Bristow, Eddy Rubin , Paul Richardson.
DOE Joint Genome Institute

DOE JOINT GENOME INSTITUTE
US DEPARTMENT OF ENERGY
OFFICE OF SCIENCE



JGI is a leader in performing sequences to support the number of U.S. Department of Energy (DOE) Microbial Programs. The Microbial Genome and GTL programs were established to determine the complete genome sequence of a number of microbes selected for their relevance to DOE missions. Community Sequencing Program recently started by JGI (http://www.jgi.doe.gov/CSP/index.html) also includes sequence and detailed analysis of the genomes of the different representatives of the microbial world.

A workflow procedure for all microbial programs has been formalized to process samples from DNA prep through sequencing, assembly, finishing, quality assurance, annotation and analysis. To date, the JGI has sequenced over 300 microbes to draft quality, finished over 60 and is currently working on more than 100 additional microbial projects. Most projects are now targeted for complete finishing.

A completely sequenced, high quality genome is a perfect starting point for the genome annotation (http://img.jgi.doe.gov/v1.1/main.cgi), microarrays, knockout experiments, etc. Despite the fact that draft genomes are useful on their own, a completed genome is, overall, a better product, especially if it will be used to analyze previously unknown and difficult-to-cultivate microbes; for the comparative analysis of clinical isolates, or for the creation of microbial strains overproducing different proteins and amino acids. Knowledge of the completely finished genome will allow scientists to modify specific regions of the genome and, therefore, to affect the expression of the gene being studied. Thus, in order to be able to realize these and many other studies, it is necessary to close most (if not all) of the genomes being sequenced at JGI.

The Integrated Microbial Genomes (IMG) system was created to provide 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. Currently IMG includes 674 genomes from archaea to eukarya sequenced by JGI (134) and other centers.

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-AC02-05CH11231 and Los Alamos National Laboratory under Contract No. W-7405-ENG-36. LBNL-57373 Poster II

Programs:

DOE Microbial Program

DOE GTL Program (GTL)

Community Sequencing Program (CSP)

JGI Internal Program

LSP- Lab Science Program

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

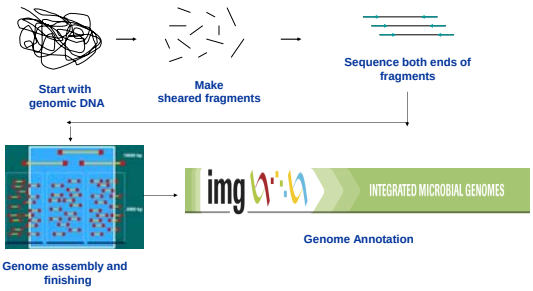
The Community Sequencing Program

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

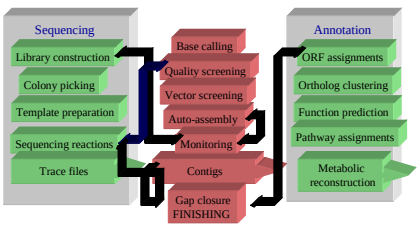
Types of projects:

A wide range of projects.
The most important factor for acceptance is a project's scientific merit.
The deliverables can range from raw sequence traces to well-annotated assembled genomes

Genome Sequencing



Steps in the Sequencing Process



Finished Genomes

2005 DOE Microbe Projects

Rhodopseudomonas palustris HaA2

Acidobacterium Ellen 345

Rhodopseudomonas palustris BisB18

2005 CSP Microbe Projects

Ignicoccus sp. KIN4/I

2005 WFO Microbe Projects

Chlamydia trachomatis Ds2923

Chlamydia trachomatis G9301

2004 DOE Microbe Projects

Pelodictyon luteolum DSM 273

Methanospirillum hungateii, JF1

Nitrosospira multiformis Surinam

Nitrosococcus oceani ATCC 19707(ex Nitrosomonas oceani)

Nitrobacter winogradskyi, Nb-255

Frankia sp., CcI3

Anaeromyxobacter dehalogenans, 2CP-C

Thiomicrospira crunogena XCL-2

Thiomicrospira denitrificans

Jannaschia sp.CCS1

Thiobacillus denitrificans, ATCC 25259

Polaromonas JS666 (ex b-proteobacterium sp., JS666)

Synechococcus sp., CC9902 (Coastal)

Synechococcus sp., CC9605 (Oligotrophic)

Prochlorococcus sp., NATL2A

Chromohalobacter salexigens, DSM3043

Psychrobacter cryohalolentis K5 (ex Psychrobacter cryopegella)

Nitrobacter hamburgensis

2004 GTL

Rhodoferax ferrireducens DSM 15236

Pelobacter carbinolicus DSM 2380

Organsim: 2003 Microbes

Prochlorococcus sp. MIT9312

Methylobacillus flagellatus

Chlorobium chlorochromatii CaD3 (ex Chlorochromatium aggregatum)

Moorella thermoacetica ATCC 39073

Anabaena variabilis

Synechococcus elongatus

Methylobium petroleophilum PM1(ex Rubrivivax gelatinosus PM1)

Burkholderia strain 383(ex R-18194)

2002 Microbes

Ehrlichia canis

Saccharophagus (ex Microbulbifer) degradans 2-40

Ralstonia eutropha JMP134

Dechloromonas aromatica RCB

Pseudomonas syringae B728a

Rhodospirillum rubrum ATCC 11170

Desulfovibrio desulfuricans G20

Geobacter metallireducens GS-15

Psychrobacter sp. 273-4

Methanococcoides burtonii

Haemophilus somnus 129PT

2001 Microbes

Cytophaga hutchinsonii ATCC 33406

Methanosarcina barkeri fusaro

Pseudomonas fluorescens PfO-1

Rhodobacter sphaeroides 2.4.1

Thermobifida Fusca YX

Burkholderia xenovorans LB400 (ex Burkholderia fungorum)

Nostoc punctiforme ATCC 29133

Novosphingobium (ex Sphingomonas) aromaticivurans

Synechococcus WH8102

Rhodopseudomonas palustris CGA009

Nitrosomonas europaea

Prochlorococcus marinus MED4

Prochlorococcus marinus MIT9313

Bacillus anthracis Sterne

Bacillus thuringiensis 9727

Bacillus thuringiensis ALH

Bacillus thuringiensis Zebra killer

Brucella abortus 2308

Francisella philomiragia 2773039

Francisella tularensis OR-960463

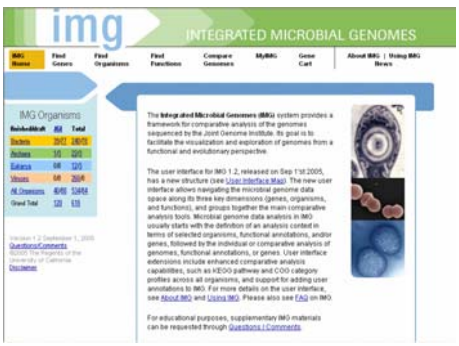
JGI-wide Finishing Standards

- **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.**
- **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.**

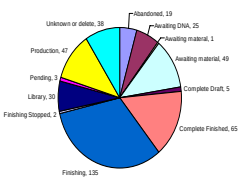
135 Microbial projects in finishing

65 – Microbial genomes finished

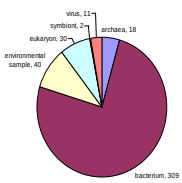
Genome Analysis



Projects by Status



Projects by Taxonomy



finished/draft	JGI	Total
Bacteria	43/87	256/117
Archaea	1/3	23/3
Eukarya	0/0	12/3
Viruses	0/0	260/0
All Organisms	44/90	551/123
Grand Total	134	674

Next IMG release: March 1, 2006