

# Browsing small genomes on the WWW

John C. Robbins

Sequence data from genome projects dedicated to unraveling the genetic codes of literally dozens of microorganisms are available on the Web. Unfortunately, finding these resources can be a daunting and frustrating task.

Without a doubt, the best place to initiate one's search for genomic information, microbial or not, is the National Center for Biotechnology Information (NCBI) web site (<http://www.ncbi.nlm.nih.gov>). Here, a full menu of services is offered, including search engines that allow queries of GenBank via many different avenues. The Entrez search system can aid in finding sequences from any organism in the nucleotide and protein databases by searching numerous nonsequence search fields.

Alternatively, one may find genes encoded by a particular organism by climbing through a virtual phylogenetic tree using the

*John C. Robbins has research interests in microbial genomics and may be contacted at [jcr15@cornell.edu](mailto:jcr15@cornell.edu); <http://genomics.simplenet.com>.*

Taxonomy Browser or to retrieve all sequences in GenBank ascribed to any organism using Batch Entrez. By submitting a protein or DNA sequence to NCBI's BLAST service, one may find related sequences in GenBank. All BLAST and Entrez search results have copious links to related information, including the PubMed literature citation database. DNA sequence data from a genome project may also be scanned for open reading frames by using the ORF Finder feature. Before using NCBI's site, though, it may be instructive to peruse the Entrez help guide online at <http://www.ncbi.nlm.nih.gov/Entrez/entrezhelp.html>.

Some other starting-point sites are also worth mentioning. If studying yeast is your wont, then two URLs devoted to *Saccharomyces cerevisiae* are definitely required reading: The *Saccharomyces* Genome Database (<http://genome-www.stanford.edu/Saccharomyces/>) offers combined physical and genetic maps and systematic functional analysis of the yeast genome; and extensively annotated sequences and interactive tools for genome analysis are available at The Yeast

Genome Project site (<http://www.mips-biochem.mpg.de/yeast/>). The Microbial Database (MDB) at The Institute for Genomic Research (TIGR) (<http://www.tigr.org/mdb/mdb.html>) provides signposts to microbial sequencing projects, including a trio of their own. Overall, you will find a very well-organized set of databases and search tools for the analysis of the *Haemophilus influenzae*, *Mycoplasma genitalium*, and *Methanococcus jannaschii* genomes. For further information about fully and partially sequenced genomes, as well as nascent projects, one may access small genome indices on the WWW at MAGPIE (<http://www.mcs.anl.gov/home/gaasterl/genomes.html>) and the NCGR's Genome Sequence Database (<http://www.ncgr.org/gsd/>).

The sites listed below should also serve as an enticing entree into the world of microbial genomics. Links to these, along with other related sites pertaining to microbial genomics, are also available at <http://genomics.simplenet.com>.

## A sampler of product and service sites related to unicellular genome projects

|  |  |   |
|--|--|---|
| <b>General information</b>                           |  |   |
| NCBI Genomes Database                                | <a href="http://www.ncbi.nlm.nih.gov/Entrez/Genome/org.html">http://www.ncbi.nlm.nih.gov/Entrez/Genome/org.html</a>  | The NCBI provides browsable maps of small genomes and genomic sequences, par excellence.  |
| <b>Sites related to specific organisms</b>           |  |   |
| The <i>E. coli</i> Genome Project                    | <a href="http://www.genetics.wisc.edu:80/index.html">http://www.genetics.wisc.edu:80/index.html</a>  | The <i>E. coli</i> strain K-12 genome at the University of Wisconsin.   |
| CyanoBase  | <a href="http://www.kazusa.or.jp/cyano/cyano.html">http://www.kazusa.or.jp/cyano/cyano.html</a>  | Browsable genome maps of the <i>Synechocystis</i> sp. genome and the sibling proteome project, Cyano 2Dbase at the Kazusa DNA Research Institute.   |
| The <i>Mycoplasma pneumoniae</i> Genome Project      | <a href="http://mail.zmbh.uni-heidelberg.de/M_pneumoniae/">http://mail.zmbh.uni-heidelberg.de/M_pneumoniae/</a>  | Analysis of the genome of <i>M. pneumoniae</i> , the causative agent of atypical pneumonia, at the University of Heidelberg.  |
| <b>Sites related to metabolism reconstruction</b>    |  |   |
| EcoCyc and HinCyc                                    | <a href="http://www.ai.sri.com/ecocyc/ecocyc.html">http://www.ai.sri.com/ecocyc/ecocyc.html</a> =20<br><a href="http://www.ai.sri.com/ecocyc/hincyc.html">http://www.ai.sri.com/ecocyc/hincyc.html</a> | Fully integrated encyclopedias of the genes and metabolic pathways of <i>E. coli</i> and <i>Haemophilus influenzae</i> .  |
| WIT  | <a href="http://www.cme.msu.edu/WIT/">http://www.cme.msu.edu/WIT/</a>  | Microbe metabolism reconstruction, with fully interactive modeling.   |
| <i>Saccharomyces</i> Genome Deletion Project         | <a href="http://sequence-www.stanford.edu/group/yeast_deletion_project/deletion.html">http://sequence-www.stanford.edu/group/yeast_deletion_project/deletion.html</a>                                  | Functional genomics of yeast.   |
| <b>Ongoing genome projects</b>                       |  |   |
| The <i>Mycobacterium tuberculosis</i> Genome Project | <a href="http://www.sanger.ac.uk/Projects/M_tuberculosis/">http://www.sanger.ac.uk/Projects/M_tuberculosis/</a>  | A listing of cosmid contigs cloned and sequenced at the Institut Pasteur and Sanger Centre, respectively.   |
| Advanced Center for Genome Technology (ACGT)         | <a href="http://dna1.chem.uoknor.edu/">http://dna1.chem.uoknor.edu/</a>  | The <i>Streptococcus pyogenes</i> , <i>Neisseria gonorrhoeae</i> and <i>Actinobacillus actinomycetemcomitans</i> Genome Projects at the University of Oklahoma. BLAST searching of raw data is available. |