Internet On-Ramp

World Wide Web Resources for Microbiologists

Donald P. Nierlich¹, Dennis Benson², Mary K.B. Berlyn³, Lois Blaine⁴, Peter Karp⁵ and Kenneth Sanderson⁶

¹University of California, Los Angeles, CA, ²National Center for Biotechnology Information, Bethesda, MD, ³Yale University, New Haven, CT, ⁴ATCC, Rockville, MD, ⁵Pangea Systems, Oakland, CA, USA and ⁶University of Calgary, Calgary, AL, Canada

This review is intended for several audiences. For those not yet using the World Wide Web (WWW), it provides an introduction and a roadmap to some important electronic databases and analytical programs. For more familiar users, it describes some Web sites in detail. The multi-faceted focus reflects the diverse resources of the Internet and a subset of particular use to microbiologists. Of equal importance, this article hopes to convey that these sites and others like them are research projects in themselves. Thus, they not only provide easy access to information but also test new ways of organizing and using it. Finally, for those who are leading the way into cyberspace, this article might provoke thought and provide some amusement.

The way that scientists find information is changing rapidly. Visits to the library and calls or letters to colleagues are being replaced by searches of the Internet, e-mail messages and postings on electronic bulletin board systems (BBSs). If you need a mutant strain or a monoclonal cell line, you may be only a few computer-mouse clicks away from locating it, and you may be able to order it with only a little more effort. Along the lines of information resources, microbiologists are widely aware of the GenBank[®] and European Molecular Biology Laboratory (EMBL) nucleic acid and protein sequence databases and the Medline bibliographic database. However, there are many other, very diverse databases, differing in subject matter, internal organization and methods for user searches and information retrieval. In addition, some sites primarily provide analytical services; for example, the MFOLD site performs RNA-secondary structure predictions.

A newer type of information resource, a knowledge base, is represented by EcoCyc[™], a digital compendium of information about *Escherichia coli*. Organized according to metabolic pathways, EcoCyc provides instantaneous cross references to enzymes, substrates, genes and their supporting literature. It is one of a growing number of such resources dealing with either individual organisms, microbes generally (e.g., PUMA) or other information.

As an example of this diversity of resources, a list of selected microbiological databases and analytical programs that are accessible on the Web has been made available as a supplement to this review in the BioResources Section of the BioTechniques' Web site (<http://www.BioTechniques.com>). As explained below, the Web is one part of the Internet. The list is not comprehensive, as there may be several hundred such sites. Moreover, there are databases that are accessible through the Internet but not part of the Web.

Enter the Internet

Many computer databases have been available for some years. The Protein Information Resource (PIR) database of protein sequences and phylogeny was last published in *paper*

format in 1978 (6). The EMBL and GenBank databases were established in the early 1980s. The advantages of computer databases are the ease with which one searches for information, the facility with which additions and revisions are made and the ability to address them with complex questions. For example, many databases can handle Boolean queries, such as "*find keywords HIV and drug-use*". An important additional advantage is that, being electronic, they can be addressed from afar.

The newest developments, in fact, stem from recent developments on the Internet, the electronic network linking academic, commercial, government and private computers (7). First, the wide acceptance of several relatively new protocols (or standards) for the exchange of information means that persons with little technical expertise can retrieve information from the growing multitude of resources. Moreover, this can occur independent of the kind of computer that is used and with little or no expense for special software. These protocols, particularly HyperText Markup Language (HTML), also allow anyone with only moderately more technical knowledge and computer resources to become an information provider as well.

Second, the growth of Internet connections and adoption of these protocols allow one to quickly access global information resources. For example, American and Australian users will in many instances be able to obtain a detailed color graphic from the others' databases in a small fraction of a minute. Fiber-optic and satellite links support rapid communications. Unfortunately, poor connections also occur, often where communications depend on telephone lines, but generally information in text files travels quickly; sometimes it is easier to obtain a reference from across a continent than from a file cabinet across the room. For example, DNA sequences from the GenBank computer, in Bethesda, MD, USA, can be retrieved at the University of California, Los Angeles, in seconds.

The Internet supports a broad variety of uses, each one reflecting the existence of specific protocols for establishing communications (7). Electronic mail is the traditional use. Usenets or newsgroups, the approximately 10 000 electronic bulletin boards that support discussion among persons sharing specialized interests and particularly about one hundred BioSci newsgroups, are another important resource. The file transfer protocol (FTP) and Telnet are two other, long-established means by which one can transfer information to/from remote computers. But "long-established" is rare in this realm. For example, Gopher, an information search/retrieval protocol that is useful for text-data exchange, which was dominant 6 years ago, has been largely replaced by the Web. The Web, developed at the European Center for Nuclear Research (CERN), depends on two protocols: HyperText Transfer Protocol (HTTP) and HTML. The first describes how computers communicate over a patchwork electronic network, and the second describes how documents are written and displayed. The Web, which is rapidly becoming the de facto standard because of its simplicity of use and power, is in essence a subdomain of the Internet that "speaks" these protocols.

Surfing the Web

Underlying the ease of use of the Web is multimedia technology. To most, multimedia means information conveyed by several means: text, graphics, audio and video. In the realm of informatics, however, it also means that electronic pointers, called hypertext links, can be embedded in a document, associated with a word, on-screen push button or icon, so that

Internet On-Ramp

when one clicks the mouse on them, “something” happens. In this context, this something could be to display a graph (play music or run a video clip), hop to another point in a document or database, or retrieve information from another computer elsewhere on the Internet. Thus, with clicks of a mouse one “surfs” the Internet. For example, when looking at the pathway of tryptophan biosynthesis in EcoCyc, with two clicks one can see a structural representation of the enzyme tryptophan synthetase.

In the last example, this is accomplished by following a link across the Internet from the EcoCyc server in California to the Protein Database (PDB) server in New York. Similarly, one might find on a Web site an index of other sites of interest and then, by clicking on a listing, display information from that site without knowing anything of the computer commands or Internet addresses that are involved. The consequence of these tools is that researchers can often find what they want by simply browsing. If this is not successful, there are powerful “search engines” and global indexes available. Searching is also aided by the customary design of “home pages”, often artful indexes to what is available on a Web site.

In the following sections, there are short descriptions of some widely used resources plus a few new ones. The uniform resource locators (URLs) are the Internet addresses of the resources. Note that the list of URLs provided is not meant to be comprehensive. However, a visit to a few of the “list servers” provided, including one maintained on the American Society for Microbiology’s (ASM) Web page (<http://www.asmusa.org/others.htm>), should provide a more complete view of what is available. List servers are sites providing an index to other sites. Like other on-line “hypertext” lists, once you have it displayed, you can go to the sites listed directly with a mouse click.

GenBank

GenBank (URL: <http://www.ncbi.nlm.nih.gov/>) is a database of all publicly available nucleotide and protein sequences, including supporting bibliographic and biological information (2). Established by the U.S. National Institutes of Health (NIH) in 1982 at the Los Alamos National Laboratory, since 1992, GenBank has been based at the National Center for Biotechnology Information (NCBI), a division of the National Library of Medicine, located on the NIH campus. Currently, GenBank contains close to three million sequence entries, and the number is rising exponentially. Although human sequences predominate, more than 40 000 organisms are represented. Through collaborations with the EMBL Data Library and the DNA Databank of Japan (DDBJ), data are exchanged daily, ensuring that each of the three resources maintains a comprehensive set of all known, public sequence information.

In addition to GenBank’s sequence databases, there are a number of specialized databases. Some of these are maintained by NCBI’s staff, others are collected there for distribution. The expressed sequence tags database (dbEST) includes several hundred thousand tags, including homology assignments, and is maintained at NCBI. NCBI’s computers are also host to the Online Mendelian Inheritance of Man (OMIM) database, a comprehensive catalog of human genes and diseases.

The databases can be searched directly from the NCBI home page (see above URL). Entrez allows one to retrieve sequences using sequence-associated identifiers, such as the names of genes or keywords and bibliographic citations. The

Basic Local Alignment Search Tool (BLAST) provides sequence-similarity searching. The BLAST family of search programs are fast and provide rigorous statistical methods for judging the significance of matches. Detailed instructions on the use of the programs can also be found on NCBI’s Web pages. One new feature of the BLAST programs’ output is a graphical view of aligned sequences in addition to a residue-by-residue display of the matches.

The primary search tool for NCBI’s “text” databases is Entrez, an integrated system that accesses sequence databases, related Medline references and collections of genomic data and three-dimensional (3-D) structures of macromolecules. With Entrez, GenBank records can be searched by accession numbers, keywords or bibliographic citations. The nucleic acid sequences, protein sequences, structures and bibliographic information are also linked to provide easy traversal among the four types of data. Entrez also features a “neighboring” facility that provides for rapid browsing of groups of related sequence or bibliographic records.

Entrez undergoes continuing enhancements, and NCBI is seeking feedback from its users. Entrez now displays schematics of chromosomes and genomes from the Genomes Division of GenBank. With more than 100 entries, the complete sequence of the yeast genome and several bacterial genomes are included, as well as genomic segments of other bacteria and several higher eukaryotes.

The NCBI home page provides a variety of other services. BankIt is a new sequence submission tool that uses a simple forms-based approach for formatting and submitting GenBank entries. Accession numbers are assigned within 24 h, and all entries undergo extensive review.

The NCBI also offers a number of alternative ways of searching its databases (2). For example, e-mail access allows the retrieval of sequences, and e-mail BLAST provides for sequence-similarity searching. If one sends e-mail with the word “help” (only) as the body of the message to either query@ncbi.nlm.nih.gov or blast@ncbi.nlm.nih.gov, instructions will be sent by return e-mail. The NCBI databases can also be obtained through direct downloading by FTP. For further information about any NCBI service, contact: info@ncbi.nlm.nih.gov or phone (301) 406-2475.

EcoCyc: Encyclopedia of *E. coli* Genes and Metabolism

The EcoCyc project (URL: <http://ecocyc.pangeasystems.com/ecocyc/>) has compiled a large knowledge base (KB) describing the genes and intermediary metabolism of *E. coli*, documented with selected literature citations (9,10). The KB describes each pathway of *E. coli* metabolism, and the enzyme that carries out each reaction, including its cofactors, activators, inhibitors and the subunit structure of the enzyme. When known, the genes encoding the subunits of an enzyme are listed along with the map position of the gene on the *E. coli* chromosome. In addition, the KB describes every chemical compound involved in each step, listing synonyms for the compound name, the molecular weight of the compound and, in many cases, its chemical structure. Currently, the KB describes more than 4000 genes, 700 enzymes, 900 reactions in 130 pathways and 1200 chemical compounds. EcoCyc contains links to several other databases, including GenBank, SWISS-PROT, the Brookhaven Protein Data Bank (PDB), PROSITE and Medline.

The EcoCyc KB is designed for a number of uses. It is an electronic reference source for scientists who work with *E. coli* and related microorganisms. The KB can be viewed as

Internet On-Ramp

an electronic review article because it contains hundreds of citations to the primary literature. Utilizing a graphical user interface, the user can visualize the layout of genes on the chromosome or the biochemical features of an individual reaction or of a complete pathway (9). The navigation capabilities allow one to move from a display of an enzyme to a display of a reaction that the enzyme catalyzes or of the gene(s) that encodes the enzyme. It also supports a variety of queries, such as generating a display of the map positions of all genes that code for particular enzymes.

As well as being a reference source, EcoCyc enables complex computations related to metabolism, such as design of novel biochemical pathways for biotechnology, studies of the evolution of pathways and simulation of metabolic reactions. In this view, EcoCyc is an *in silico* model that can be probed and analyzed computationally. The EcoCyc KB could also form the foundation for a computer-based instruction system for biochemistry. The longer-term goals of the project are to integrate additional types of data into EcoCyc, such as descriptions of protein products other than enzymes and mechanisms of genetic and metabolic regulation.

The computer software that underlies EcoCyc is not specific to *E. coli*; it can be applied to other organisms. Currently, the program (HinCyc) is being used to predict the metabolic pathways of *H. influenzae* based on its genomic sequence. The demands of EcoCyc development are also driving a computer science project that aims at a new generation of information management software that supports the collaborative development of large biological knowledge bases. EcoCyc can be queried through the WWW at the URL above. Subscriptions for its use are available free to academic users. EcoCyc can also be run on SUN[®] workstations. The EcoCyc Web pages provide links to all publications produced by the EcoCyc project, including the EcoCyc User's Guide, installation instructions for SUN computers and detailed documentation of the EcoCyc schema. Questions about the EcoCyc project can be addressed to (ecocyc@panbio.com).

***E. coli* Genetic Stock Center**

The *E. coli* Genetic Stock Center (CGSC) (URL: <http://cgsc.biology.yale.edu>) is a repository for *E. coli* K-12 strains and information. The information documents the genotypes and pedigrees of the strains, the map positions of genes, phenotypic traits affected in mutant strains and the protein or RNA product encoded by the genes, the sources of strains and supporting literature. Over 7000 genetic derivatives of the original K-12 isolate are available. The Stock Center also serves as a registry for gene names and allele numbers and, for the past two decades, its curators have published editions of a comprehensive linkage map for the organism (1,3,5). The initial information used to create the CGSC database originated from the notebooks, index cards and published genetic maps of Barbara Bachmann during the course of 25 years as curator. These provided the initial core of information for the electronic database and guided design efforts. The on-line version is a relational database with several user interfaces and with links to other databases for related information (4). The information is augmented by updated genetic-map segments in the format used for the published linkage maps and a link to the latest published map figure and gene tables (3). The external links provide access to information in databases such as GenBank, EC_Enzyme, SWISS-PROT, Medline and SoyBase MetabolicDB, in addition to organismal databases, such as Maize Genome Database (GDB), Fly-

base, SacchDB and the human GDB. Because CGSC uses its core database for both the daily updates by staff and for users' on-line access, responses to queries are based on the most recent information about any item. The database can be reached on the Web at the above URL.

Searches begin on a query form, tailored for specifying properties or other characteristics of strains, genes or mutations of interest. For example, querying for "recB" or "endonuclease V" on the Gene (Site) form will retrieve the record describing gene position, polypeptide encoded, synonyms, direction of transcription, descriptive comments, references and a list of alleles of that gene. The hypertext link to the allele leads to a mutation record that includes links to the descriptions of strains carrying that mutation. Other links from gene records lead to the GenBank sequence record, Medline abstracts for the references and gene product links to an EC database and SWISS-PROT. Very diverse and specific searches can be carried out; for example, to find: (i) all of the strains with transposons within a certain region of the chromosome, (ii) F-primes carrying specific mutations or (iii) isogenic pairs of strains differing at a single, specific site. Requests for strains, questions about strains, mutations, etc., and questions about use of the database can be sent to mary.berlyn@yale.edu or by mail, phone or fax to the Stock Center (phone: 203-432-9997 or fax: 203-432-3854). The staff welcomes additions, corrections and clarifications communicated by users.

***Salmonella* Genetic Stock Centre**

The *Salmonella* Genetic Stock Centre (SGSC) (URL: <http://www.acs.ucalgary.ca/~kesander/>) maintains and distributes genetic stocks internationally. The largest numbers (several thousand) are mutants of *S. typhimurium* LT2, but there are also sets composed of wild-type strains of the species/serovars of *Salmonella* (the SARA, SARB and SARC sets) and Tn10 insertion mutants of *S. typhi*, *S. enteritidis* and *S. paratyphi* A, B and C. In addition to databases of its collections and the strain kits it makes available, the SGSC also provides an updated genetic map of *S. typhimurium* and a DNA sequence database. It also offers useful stock-collection software.

The *Salmonella* genetic map and databases are accessible at the URL address above. Included here are several files taken from Edition 8 of the *S. typhimurium* genetic map (11,12). They include the text, a list of over 1000 genes, a list of genes that have been sequenced and their GenBank accession numbers and figures showing map locations. Updates are posted periodically.

On-line queries of the strains databases can be made. Most of the database files (8) also can be downloaded from the Web site and searched locally. One can also download the database program used by the SGSC to manage its strain collection (filename SGSC-FM-SIMS). It contains no strain records, and users can make modifications to suit their personal information management requirements; FileMaker Pro[®] 2, a Macintosh[®] application, is required for the program to operate. Also available is StySeq1, a nonredundant *S. typhimurium* DNA sequence database modeled after the EcoSeq collection of *E. coli* DNA sequences (12). The sequence of the chromosome of *S. typhimurium* LT2 is being determined in collaboration with the Genome Sequencing Center, St. Louis, MO, USA (<http://genome.wustl.edu/gsc/>).

To obtain further information or to make requests from the collection, contact *Salmonella* Genetic Stock Centre, Depart-

Internet On-Ramp

ment of Biological Sciences, University of Calgary, Calgary, Alberta, Canada T2N 1N4; phone: (403) 220-6792 or 220-3572; Fax: (403) 289-9311; e-mail: kesander@ucalgary.ca.

ATCC Databases on the WWW

The American Type Culture Collection (ATCC; Rockville, MD, USA), the largest and most diverse service culture collection in the world, provides access to its microbiological and cell-line data through the WWW (URL: <http://www.atcc.org/>). Together its databases cover information on over 50 000 strains of algae, protozoa, bacteria, phages, fungi, yeasts, viruses, plant tissues, cell lines and hybridomas. The ATCC is also the official repository for U.S. patent strains of microorganisms and has recombinant DNAs available for distribution, including over one million human cDNA clones that were recently added. Finally, it provides information on strains, media and storage of organisms, and offers broad microbiological services on a contract basis. Strains can also be ordered on line. Catalogs of the strain collections can also be obtained in printed format or on CD-ROM.

Each of the major ATCC collections provides its own searchable database. Records contain data on name of the strain, growth conditions and media, strain history, special applications and references. ATCC's largest and most comprehensive database is the recombinant DNA materials database, covering information on thousands of clones, hosts, libraries, kits, vectors and oligonucleotides.

Searches of all of the databases are made using a standardized query form. Searches allow recovery of entries not only by entry name or number, but also by other information, such as the mutations an organism carries or the name of the person who deposited the strain in the collection. In addition, queries can be made more precise using Boolean operators. Field-based searches have been implemented for the ATCC databases. With this retrieval system, one can search on name of recombinant material, organism and tissue, gene name, chromosome number, insert information, restriction sites, vectors, genotype, primers, application of materials, depositors' names and references as appropriate to the subset of the data being queried.

Data in the recombinant material database is linked to GenBank/EMBL/DBJ records whenever possible; that is, a user can click on a GenBank number to retrieve the record corresponding to an ATCC clone. Links to GDB are also in process. Most of the other ATCC databases are not yet linked to external microbiological data resources; however, work in that direction is underway, with the viral databases having just been completed.

An Information Cornucopia?

Although the sites described above are important examples of the resources available on the Web, because of differences in needs, a given user can use none or only a few of them. Individuals differ in specialized interests, in the form that they would like to obtain information and in physical location. In regards to location, it is instructive to know how information travels over the Internet. The "host" breaks a message into packets and sends each out tagged with an address. The packets, independently, then travel from node to node across the net, and once at their destination, they are reunited. When not excessively burdened, sizable messages can travel quickly. Burgeoning use, however, threatens the speed of the Internet in spite of its capacity. Users, therefore, may benefit in choosing resources based on accessibility. Thus, while

North American users will likely obtain sequences from GenBank, Asians might use the DBJ and Europeans use the EMBL databases. Similarly, many important sites are "mirrored" at several global locations to circumvent the bottlenecks. For this reason, the EMBL databases that normally reside on the European Bioinformatics Institute (EBI) server near Cambridge, England, UK can be found on hosts in most European nations and in Israel. The French site, INFOBIO-GEN, provides the EMBL databases and other resources, and a uniform user interface, the Sequence Retrieval System (SRS), by which to access them. Also in this way, two comprehensive lists of microbiology and biochemistry resources, *All the Virology Sites on the Web* (<http://www.tulane.edu/~dmsander/garryfavweb.html>) and *Pedro's BioMolecular Research Tools* (http://www.public.iastate.edu/~pedro/research_tools.html), are each mirrored at several locations.

Thus, both new and experienced users will surf the Web to find resources that meet their needs and are accessible. New sites are frequently added, and old sites are often in constant revision, so there are always discoveries to be made. In addition, several publications review Web sites; besides *BioTechniques*, these include *ASM News*, *Nature* and *Trends in Biochemical Sciences*. Also, at two-year intervals, *Nucleic Acids Research* dedicates an issue to databases and related programs (the most recent is: Vol. 26, No. 1, January, 1998).

In spite of the great resources offered by the Web, it does have its limitations. The most immediate relate to costs, including: (i) the costs of increasing use and expansion will have to be paid, and it isn't clear yet how charges will be levied and collected; (ii) while one can order bacterial strains and technical books, along with many general goods and services, a secure system of electronic payment is just being worked out; and (iii) while every month there are new on-line research journals and other copyrighted materials, means for recouping their costs and obtaining payment are still in flux.

Finally, it must be realized that the World Wide Web is just one stage in the democratization of the Internet. First used as a military resource, then as a tool of academic researches, it is now becoming the province of school children in rich countries and a sought after prize in the less affluent. This is largely the result of three developments: inexpensive computers, global high-speed communications and easy-to-use software. But the Web's underlying protocols also have technical limitations. Perhaps the most important is the question-and-answer way in which the user's computer, the "client" and the "server" exchange information. But new protocols are already being introduced. Currently, Java™ allows the client to obtain software tools (applets) with data from the server. These tools then can be used for incremental manipulation of the information, e.g., calculations, viewing different aspects of a genetic map or rotating a molecular model in space. However, in the longer term, the Web is likely to be just an evolutionary stage in the development of global information resources and communications.

NOTES

GenBank is maintained by the NCBI, David Lipman, Director. Dennis Benson heads the Information Resources Branch. GenBank is a registered trademark of the U.S. Department of Health and Human Services. EcoCyc is a collaborative project between Peter D. Karp of Pangea Systems and Monica Riley of the Marine Biological Laboratory. Kenneth Rudd, formerly of the National Library of

Internet On-Ramp

Medicine, has contributed much of the genomic data. The current director of the CGSC is Mary Berlyn. The database design and data model were by Berlyn and Stanley Letovsky of Letovsky Associates (letovsky@gdb.org); the software development was the work of Letovsky. The curator of SGSC is Kenneth E. Sanderson. The assistance of Andrew Hessel, Neil Mattatall and P. Ronald MacLachlan in developing and/or maintaining the databases is gratefully acknowledged. StySeq1 was developed by Kenneth E. Rudd in collaboration with Andrew Hessel. The SGSC is supported by an Infrastructure grant from the Natural Sciences and Engineering Research Council of Canada. The director of bioinformatics at ATCC is Lois Blaine.

REFERENCES

1. **Bachmann, B.J.** 1990. Linkage map of *Escherichia coli* K-12, Edition 7, p. 807-876. In F.C. Neidhardt, J.L. Ingraham, K.B. Low, B. Magasanik, M. Schaechter and H.E. Umbarger (Eds.), *Escherichia coli* and *Salmonella typhimurium*, Cellular and Molecular Biology. ASM Press, Washington, D.C.
2. **Benson, D.A., M. Boguski, D.J. Lipman and J. Ostell.** 1994. GenBank. *Nucleic Acids Res.* 22:3441-3444.
3. **Berlyn, M.K.B.** 1998. Linkage Map of *Escherichia coli* K-12, Edition 10: the traditional map. *Microbiol. Mol. Biol. Rev.* 62:814-984.
4. **Berlyn, M.B. and S. Letovsky.** 1992. Genome-related datasets within the *E. coli* Genetic Stock Center Database. *Nucleic Acids Res.* 20:6143-6151.
5. **Berlyn, M.K.B., K.B. Low and K.E. Rudd.** 1996. Linkage map of *Escherichia coli* K-12, Edition 9, p. 2489-2495. In F.C. Neidhardt, R. Curtiss III, J.L. Ingraham, E.C.C. Lin, K.B. Low, B. Magasanik, W.S. Reznikoff, M. Riley, M. Schaechter and H.E. Umbarger (Eds.), *Escherichia coli* and *Salmonella*: Cellular and Molecular Biology, 2nd ed. ASM Press, Washington, D.C.
6. **Dayhoff, M.O., R.M. Schwartz and B.C. Orcutt.** 1978. A model for evolutionary change, p. 345-358. In M.O. Dayhoff (Ed.), *Atlas of Protein Sequence and Structure*, Vol. 5, Suppl. 3. National Biomedical Research Foundation, Washington, D.C.
7. **Gilster, P.** 1995. *The New Internet Navigator*. John Wiley & Sons, New York.
8. **Hessel, A. and K.E. Sanderson.** 1996. Computer system for maintaining strains at the *Salmonella* Genetic Stock Centre, Chapter 135. In F.C. Neidhardt, R. Curtiss III, J.L. Ingraham, E.C.C. Lin, K.B. Low, B. Magasanik, W.S. Reznikoff, M. Riley, M. Schaechter and H.E. Umbarger (Eds.), *Escherichia coli* and *Salmonella*: Cellular and Molecular Biology, 2nd ed. ASM Press, Washington, D.C.
9. **Karp, P. and S. Paley.** 1996. Integrated access to metabolic and genomic data. *J. Comp. Biol.* 3:191-212.
10. **Karp, P., M. Riley, S. Paley and A. Pellegrini-Toole.** 1998. EcoCyc: electronic encyclopedia of *E. coli* genes and metabolism. *Nucleic Acids Res.* 26:50-53.
11. **Sanderson, K.E., A. Hessel, S.-L. Liu and K.E. Rudd.** 1996. Genetic map of *Salmonella typhimurium*, Edition VIII, Chapter 110. In F.C. Neidhardt, R. Curtiss III, J.L. Ingraham, E.C.C. Lin, K.B. Low, B. Magasanik, W.S. Reznikoff, M. Riley, M. Schaechter and H.E. Umbarger (Eds.), *Escherichia coli* and *Salmonella*: Cellular and Molecular Biology, 2nd ed. ASM Press, Washington, D.C.
12. **Sanderson, K.E., A. Hessel and K.E. Rudd.** 1995. Genetic map of *Salmonella typhimurium*, Edition VIII. *Microbiol. Rev.* 59:241-303.

Address correspondence to:

Dr. Donald Nierlich
Department of Microbiology
Immunology and Molecular Genetics, UCLA
Los Angeles, CA 90095-1489, USA
Internet: donn@microbio.lifesci.ucla.edu

Some Useful Web Sites that Provide Indexes to Other Resources.

BioInformatics (Johns Hopkins)	http://www.gdb.org/
Biology and Molecular Biology (Harvard)	http://golgi.harvard.edu/
Biology (U.S. Geological Survey)	http://info.er.usgs.gov/network/science/biology/index.html/
Biology Meta-Index (Madison)	http://www.genetics.wisc.edu/html/bio.html
Bio-wURLd (Biochemistry, Bioinformatics, Molecular Biology)	http://www.ebi.ac.uk/htbin/bwurl.pl/
BioMolecular Research Tools (Pedro's)	http://www.public.iastate.edu/~pedro/research_tools.html
Carbohydrate (Pedro's)	http://www.public.iastate.edu/~pedro/carbhyd/carbhyd.html
Genome Research (TIGR)	http://www.tigr.org/
Human Genome Project, Sites/Resources (Oak Ridge)	http://www.ornl.gov/TechResources/Human_Genome/genetics.html#hgpsites/
Immunology: WWW Virtual Library (Harvard)	http://golgi.harvard.edu/biopages/immuno.html/
Microbial Information Network (China)	http://sun.im.ac.cn/
Microbiology Links (American Type Culture Collection)	http://www.atcc.org/links.html
Microbiology Links and Resources (BioTech, U. Indiana)	http://biotech.chem.indiana.edu/lib/scitools.html
Microbiology, Immunology and Molecular Genetics - this list (UCLA)	http://www.lifesci.ucla.edu/mmg/www/
Microbiology, Immunology and Virology Server (U. Adelaide)	http://www.microbiology.adelaide.edu.au/general/microwww.htm
The Microbiology Network	http://microbiol.org/
Microbiology Resources (American Society for Microbiology)	http://www.asmsa.org/others.htm
Molecular Biology (Pasteur Inst.)	http://www.pasteur.fr/other/biology/english/rubriques-uk.html
Molecular Biology and Biochemistry (ExPASy, U. Geneva)	http://expasy.hcuge.ch/
Mycology Resources (Cornell)	http://muse.bio.cornell.edu/~fungi/
Vector Sequence Database (Queens U., Canada)	http://vectordb.atcg.com/
Virus Databases (Australian National U.)	http://life.anu.edu.au:80/.viruses/virus.html
Virology (All the Virology Sites on the Web; Tulane U.)	http://www.tulane.edu/~dmsander/garryfavweb.html
WFCC World Data Center for Microorganisms (Japan)	http://wdcm.nig.ac.jp/

Links to these and other sites are provided at <http://www.BioTechniques.com>

Note, URLs are case-sensitive; that is, capital and lower-case letters must be entered correctly.