

Update on Bioinformatics

Plant Biology in the Post-Gutenberg Era

Everything You Wanted to Know and More on the World Wide Web

Chris Somerville*, David Flanders, and J. Michael Cherry

Carnegie Institution of Washington, Department of Plant Biology, 290 Panama Street, Stanford, California 94305 (C.S.); and Department of Genetics, Stanford University, Stanford, California 94305 (D.F., J.M.C.)

Once upon a time, not very long ago, students of biology with burning questions about the nature of life used to spend long hours in dusty library stacks searching for knowledge and opinions recorded by their academic ancestors onto random access information retrieval devices made of matted cellulose fibers. Although these "books," as they were called, can still be found in antique stores and some university libraries, their use has declined dramatically since the release of the first user-friendly free World Wide Web (WWW) browser in November of 1993. Publishers were quick to spot the trend and several journals, notably the massive *Journal of Biological Chemistry* (<http://www-jbc.stanford.edu/>), are now available on-line and several others, such as *Plant Physiology*, are partially available on-line (for a list of electronic journals, see <http://www.nih.gov/science/journals/>). It now seems likely that most academic journals will be available via the WWW before the turn of the century, and it seems certain that an ever-increasing number will only be available via the WWW as users and libraries come to grips with using the electronic format.

Although most academics have been selected on the basis of a fondness for paper, an expanding coterie have recognized that there are many advantages to the electronic versions. For instance, it is possible to do a full text keyword search of articles in *Journal of Biological Chemistry*. It is also possible to do forward literature searches. That is, if you are reading an article from a 1995 issue that is cited in a 1996 paper, the 1995 article is annotated with a hypertext link (a highlighted word or phrase that activates a retrieve command when clicked with a mouse) that allows you to move forward to the citing article with the click of a mouse button. Another nice feature is that there need be no limitation on the number of color pictures or film clips associated with an article. The role of video in cell biology research is expanding rapidly and there is every reason to expect that it will soon be routinely included in electronic research publications. It is also possible to append commentaries to articles at any time after they appear. In the

physics journals, where electronic publishing has been routine for many years, the value of an article is established by the number of annotations by readers who can append their comments to an article. Finally, of course, it is possible to connect the journal articles directly to databases via hypertext links so that one can move from the authors' conclusions back to the original data.

The kinds of information available in databases or in some other format via the WWW are expanding daily. Indeed, the main problem with the WWW at the moment is that there is so much information available that it is possible to spend many consecutive days busily hopping from one site to another, finding something interesting at each stop but never arriving at the desired source. Thus, in this short overview we have focused on some of the resources that are available without expense and are likely to be of most value to the readers of *Plant Physiology*. In addition, monthly descriptions of new WWW resources of interest to biologists can be found in "The Spider's Web," a regular column in *Current Biology*.

SEARCH ENGINES

The most powerful approach to finding information is via a keyword search of one of the directory sites such as Yahoo (<http://www.yahoo.com>). Thus, for instance, if you want the latest catalog for a supplier of reagents such as Clontech or need technical advice about a product from Promega or most of the other suppliers of biological reagents, instruments, or books, you need simply enter the word in the space provided and the search engine will find the site. Yahoo also provides hypertext links to other search engines such as Alta Vista (<http://www.altavista.digital.com>) or Lycos (<http://www.lycos.com>), which use slightly different methods for indexing WWW sites. Thus, if a Yahoo search fails to turn up what you need, try at least one other site. Links to Yahoo mirror sites (a duplicate computer providing similar or identical resources, usually at a distant location from the original site) in other countries are available at the universal resource locator (URL) above. Many other search engines are available. For example, using the "Net Search" directory button in Netscape, a

* Corresponding author; e-mail crs@andrew.stanford.edu; fax 1-415-325-6857.

useful tool for finding e-mail addresses and home pages is "WhoWhere?" (<http://www.whowhere.com>). If the number of listings for a common last name is overwhelming, you can limit the search by restricting it to domain names with the "edu" suffix, or with the name of the institution (e.g. "bezerkly.edu"). WhoWhere also searches for U.S. phone numbers. Similar phone and e-mail search facilities are provided by Four11 (<http://www.four11.com>).

The search engines can produce such a flood of information that it can be time consuming to sort through the chaff. Thus, when searching for information about a topic, as opposed to a specific company or institution, it is useful to go directly to a WWW site that has preselected pages of potential interest. A good place to begin, in this respect, is a comprehensive list of biology servers available at <http://golgi.harvard.edu/biopages.html>. Thus, for instance, if you want to know how to find information on *Caenorhabditis elegans* or *Drosophila*, start there. A link to all model organism databases is also available at the Sanger Center (<http://www.sanger.ac.uk/bio/mod.orgs.html#alien>).

BOTANICAL WEB SITES

A good place to begin when looking for information about plants is one of the several web sites that provide hypertext links to plant-related information. The home page of the American Society of Plant Physiologists is a good place to start (<http://www.aspp.org/links.htm>). However, a much more comprehensive site is the AgDB list (<http://www.agnic.org/agdb/erdcalf.html>) compiled by the U.S. National Agriculture Library (NAL; <http://www.nalusda.gov/>). The AdDB site is an excellent attempt to describe all agriculture-related databases, datasets, and information systems. Not all datasets described are available on the Internet, but for those that are, links have been established to the actual data wherever possible. The complete set of databases and sites in AgDB can be browsed or searched by keyword. Recent additions to the list are presented on the *What's New?* page. As noted by the developers of the sites, the list is not comprehensive. In particular, we did not find a connection to any Arabidopsis information, nor to the excellent maize database MaizeDB (<http://www.agron.missouri.edu/>). In spite of these omissions there is a huge number of tremendously useful and unexpected links. In a casual survey of the approximately 500 databases available, we found databases or related information resources devoted to many other plant species.

The site <http://probe.nalusda.gov:8300/plant/index.html> provides searches of databases for Arabidopsis, alfalfa, *Phaseolus* and *Vigna*, *Chlamydomonas*, cool season food legumes, cotton, wheat, barley, rye, maize, millet, fungal pathogens, rice, members of the Solanaceae, sorghum, soybeans, and forest trees. A related site (<http://probe.nalusda.gov:8000/elsewhere/indexbio.html>) provides links to other important species-specific databases such as the Rice Genome Research project (<http://www.staff.or.jp/>), but missed the snapdragon database (<http://www.mpiz-koeln.mpg.de/~stueber/snapdragon/snapdragon.html>).

A useful secondary source is a site provided by the University of Regina that permits keyword searches of hundreds of sites (<http://herb.biol.uregina.ca/liu/bio/botany.shtml>). Another very useful site is the *Botany Virtual Library* maintained by the University of Oklahoma (<http://www.uoknor.edu/cas/botany-micro/www-vl/>). These sites list everything from the Maize Genetics Cooperative Stock Center to the Kiwi Conservation Club, a Royal Forest and Bird Protection Society project for children.

SPECIALIZED BOTANICAL INFORMATION

An important key to using the WWW is to be able to formulate expectations about what might be available so that you can design a useful search. The following databases provide a few interesting examples of the diversity of resources that are now available.

The Poisonous Plant Database (<http://vm.cfsan.fda.gov/~djw/readme.html>) is a set of working files of scientific information about the animal and human toxicology of vascular plants and herbal products of the world. Kew gardens is developing a database of the DNA content of plants (<http://www.rbgekew.org.uk/index.html>). An area of botany with a large amount of database information concerns ethnobotanical information (see <http://www.ars-grin.gov/~ngrlsb/>). This site provides a link to a database (phytochemDB) that provides the chemical composition of many plants or, conversely, allows one to search for plants that contain a particular chemical. A related database is EthnobotDB, which contains 80,000 records of plant uses world-wide (<http://probe.nalusda.gov:8000/related/aboutethnobotdb.html>). Many smaller databases, such as a database of rainforest plants and their uses (<http://rain-tree.com/plants.htm>), can be found as links to the NAL database list.

There is also an excellent WWW page for agriculture and horticulture at http://ipm_www.ncsu.edu/cernag/cern.html, which has connections to international sites for agriculture.

VIRUS DATABASE

Virtually everything known about viruses, including pictures, taxonomy, and nucleotide sequences, is available from an excellent database maintained by Cornelia Büchen-Osmond and Adrian Gibbs at the Australian National University (<http://life.anu.edu.au/viruses/virus.html>). The database contains information on most species of viruses known to infect plants; not only those whose virions have been described, but also those such as umbraviruses that have no virion protein genes of their own and use the virion proteins of their symbiotic helper viruses instead. Data are available on host range; transmission and control; geographical distribution; physical, chemical, and genomic properties; taxonomy and relationships; and selected literature references. The database also includes accession numbers of the genomic sequences of viruses and of satellite RNAs, and provides links to the National Center

for Biotechnology Information (NCBI) taxonomy database to facilitate searches for these and for more current accessions. These pages also contain generic-level summaries of data for viruses that are definitive or tentative members of genera or "groups." This is truly one-stop shopping for information on viruses.

TAXONOMY

The most ambitious database for taxonomy is *The Tree of Life* project, which is being developed by a consortium of investigators with complementary expertise in systematics (<http://phylogeny.arizona.edu/tree/phylogeny.html>). This database is accessible as a dendrogram in which many of the entries are hypertext links to deeper levels. *The Tree of Life* is a project designed to provide information about the phylogenetic relationships and characteristics of organisms, to illustrate the diversity and unity of living organisms, and to link biological information available on the Internet in the form of a phylogenetic navigator. Of the more than 1070 pages currently in *The Tree of Life*, housed on 11 computers, a few (for example, the basal pages on beetles, jumping spiders, terrestrial vertebrates, amniotes, and sac fungi) are nearing completion, whereas many are just beginning.

To browse the Tree, you should first enter at a chosen page. You can then wander up and down the branches to other pages in the Tree. You can go up toward the tips of the Tree by touching on the names of organisms on the phylogeny or classification shown on a particular page. You can also go back down the Tree toward its root, either by touching on the *Down* or *Deep* buttons on the toolbar at the top of a page, or by touching on the root of the phylogeny shown on a particular page. (More information about navigation is given on the help page.)

The Missouri Botanical Garden provides access to its nomenclatural database system through w³TROPICOS (<http://www.mobot.org/MOBOT/database.html>). This database provides access to the data collected by staff, students, and collaborators over the last 15 years. Information is available for over 750,000 scientific plant names. The records frequently have links to other associated names, types, synonymy, and bibliographic references. This information was collected for a wide variety of projects, curatorial tasks, and collaborative endeavors; consequently, many individual records are not complete or accurate. The most complete information in the database is for taxa covered by floristic or taxonomic projects underway at the Missouri Botanical Garden.

METABOLISM

At least some biologists do not have every metabolic pathway committed to memory and may occasionally find themselves searching through old college texts trying to figure out how some intermediate is made. A broadly useful first approach to this problem is to log on to the graphically oriented database of metabolic maps from the Kyoto Encyclopedia of Genes and Genomes (http://www.genome.ad.jp/kegg/metabolism_map.html). An example of the organization of information in this database is shown in Figure 1. Each of the enclosed boxes is a hypertext link that leads to detailed information about the enzyme or compound in question. Similarly, the more focused EcoCyc database of metabolic pathways in *Escherichia coli* is very useful (<http://www.ai.sri.com/ecocyc/server.html>). One simply enters the name of a compound and the pathway appears with the intermediate names, enzymes, and gene symbols as hypertext links to the structures and reference

Methionine metabolism

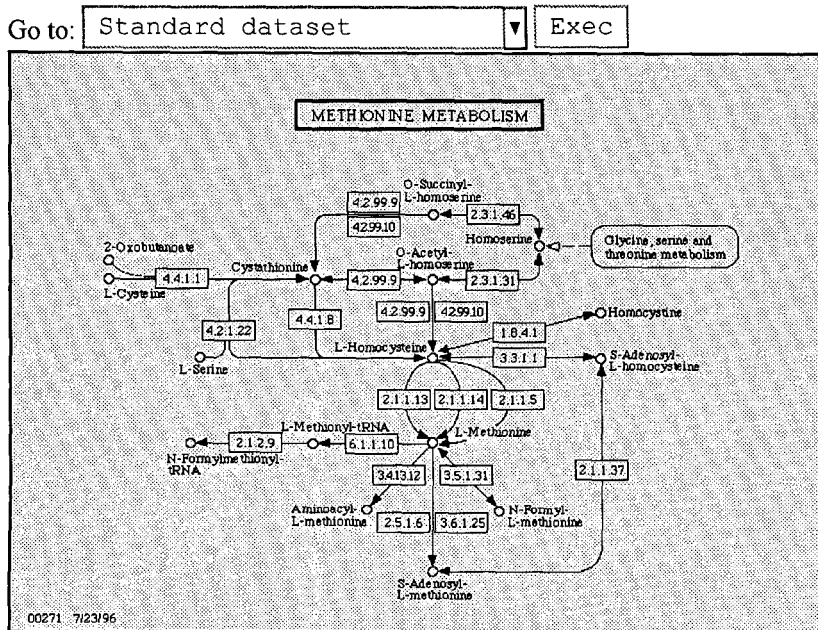


Figure 1. An example of the information on metabolic pathways available at the Kyoto Encyclopedia of Genes and Genomes (http://www.genome.ad.jp/kegg/metabolism_map.html). Each box is an active hypertext link to detailed information about the corresponding enzyme.

literature. These are some of the most impressive sites on the WWW and are valuable tools for teaching and research.

A database called EMP (Enzymes and Metabolic Pathways) claims to be by far the largest and most comprehensive collection of enzymatic and metabolic data in either electronic or printed form. It contains 16,400 enzymatic records, each with more than 300 different fields with properties for each enzyme, covering all aspects of enzymology (<http://www.biobase.com/emp/>). More than 10,000 scientific publications have been completely encoded and converted into a fully queryable database (as opposed to an electronic transcript of literature, which cannot be queried). This part of EMP is commercial (U.S. \$400 for one year's subscription; \$2000 for commercial users). The creators of EMP estimate that it has taken approximately 150 person-years to develop the database. A freely available subset of the EMP database is a file of more than 1,800 metabolic pathway diagrams that are available for downloading or viewing on-line (<http://www.cme.msu.edu/WIT/>).

Other sites of interest are the Klotho database of three-dimensional structures of small molecules (<http://www.ibc.wustl.edu/klotho/>) and a site that points to the latest information in chemical nomenclature (<http://www.chem.qmw.ac.uk/iupac.html/>).

ARBOHYDRATES

A relatively large amount of information is available via the WWW for those interested in carbohydrates. The optimal starting place is the CARBYHD WWW site, which provides links to several dozen sites with information in this area (<http://www.public.iastate.edu/~pedro/carbyhd/carbyhd.html>). For example, there are links to pages that offer information or analysis concerning the nomenclature and stereochemistry of carbohydrates; sugar chemistry and chirality/stereochemistry; graphic tools for modular representation of carbohydrates, proteoglycans, and glycosaminoglycans; sugar structures; cellulose morphing (films); conformation and shape of molecules; sugarbase (a $^1\text{H-NMR}$ and $^{13}\text{C-NMR}$ database for carbohydrate structures); glycosyltransferases; lectins; and many other topics.

The Complex Carbohydrate Structure Database (CCSD) is a database containing complex carbohydrate structures and associated text that must be downloaded to a local computer. The database has a flat file format (i.e. one record contains a single structure with its associated text and citation). The computer program to access the database, CarbBank, is available at <http://bssv01.lanccs.ac.uk/gig/pages/gag/carbbank.htm>.

CCRC-Net (Complex Carbohydrate Research Center Neural Networks) is an automated chemical object recognition system developed at the University of Georgia Complex Carbohydrate Research Center by Faramarz Valafar and Homayoun Valafar (<http://www.crcr.uga.edu/>). Its search engine utilizes a neural network-based pattern-matching mechanism. Current functionality of this software includes identification of GC-MS spectra of partially

methylated alditol acetates and $^1\text{H-NMR}$ spectra of xyloglucan oligosaccharides. Work is underway to incorporate $^1\text{H-NMR}$ spectra recognition engines for various other libraries. There are similar plans for data on infrared, ultraviolet, and other methods of chemical analysis.

Glycoscience network has a membership fee of \$50, but excellent resources (<http://bellatrix.pcl.ox.ac.uk/TGN/>). A database of carbohydrate pathways is available at (<http://www.mcs.anl.gov/home/compbio/pathways/CAR.MPW/CarbChoices.html>).

STRUCTURAL BIOLOGY

Not long ago a workstation and a working knowledge of UNIX or other languages was required to gain access to the rapidly expanding databases of structural information. This has recently changed with the development of several new tools that permit anyone with an up-to-date WWW browser to visualize three-dimensional molecules. An excellent place to begin is at the Brookhaven database, where one can scan the list of available structures (4607 proteins, 361 nucleic acids, 12 carbohydrates) and visualize them with the click of a mouse (<http://pdb.pdb.bnl.gov/cgi-bin/pdbmain>). Simply click on the name of the molecule you would like to see and the structure comes up. You can rotate the molecule in three-space by simply holding down the button while dragging the mouse. This should be a very useful tool for teaching and is aesthetically pleasing. Ribbon images of proteins for downloading are also available at <ftp://pdb.pdb.bnl.gov/images/GIF/>.

A somewhat different format for viewing is available via the VRML (Virtual Reality Modeling Language Browsers), which is also available from the Brookhaven home page (<http://pdb.pdb.bnl.gov/>). Simply click on the VRML hypertext link to bring up a list of available structures, then click on the name of the molecules of interest. A set of buttons will appear at the bottom of the screen that can be used to manipulate the image in three-space.

EVERYTHING ARABIDOPSIS

Because of the large number of plant biologists using Arabidopsis as an experimental organism, there are a relatively large number of information resources available and the list is growing. A beginning point for all searches for Arabidopsis information is the *Arabidopsis thaliana* Database (AtDB) maintained by J.M.C. and D.F. (<http://genome-www.stanford.edu/Arabidopsis/>). The home page provides links to all other Arabidopsis resources, including individual lab home pages, such as David Meinke's page, where researchers register the symbols for new genetic loci in Arabidopsis (<http://mutant.lse.okstate.edu/meinke.html>). In addition to being a launch site, AtDB is developing a new Informix (Illustra)-based database that represents the latest technology in database design. A particularly strong feature of this database is that it is capable of constructing graphical representations directly from the raw mapping data in the database. Thus,

the maps are continually updated as information is added to the database. This feature will become particularly crucial as the full genome sequencing progresses.

AtDB also provides links to all forms of genetic map data, YAC and BAC contigs, all forms of molecular mapping markers, lists of conferences and meetings, and addresses of plant biologists. AtDB provides access to all previous messages from arab-gen, the BioSci electronic Arabidopsis discussion group, and pages to perform BLAST or FASTA searches on Arabidopsis-only sequences. It also provides links to major databases such as the Arabidopsis Information Management System (AIMS), which is used to manage all information related to the Arabidopsis Biological Resource Center at Ohio State University (<http://aims.cps.msu.edu/aims/>), and the database used at the Nottingham Arabidopsis Stock Centre (<http://nasc.life.nott.ac.uk/home.html>). Both stock center databases allow searching of the extensive collections of accessions and clones, and direct on-line ordering. An important feature of these databases is that they permit a query of whether a particular stock has been ordered previously. For example, it is possible to determine how many other labs have ordered a particular EST clone. The Nottingham Centre also provides the full text of all the abstracts presented at the latest international Arabidopsis meeting.

Important EST databases include dbEST at NCBI (<http://www.ncbi.nlm.nih.gov/>) and that provided by Ernie Retzel and colleagues at the University of Minnesota (<http://www.cbc.umn.edu/ResearchProjects/Arabidopsis/index.html>). This database provides an extensive analysis of plant ESTs from all plant species. Another important EST database is the TIGR database of EST contigs in which the overlapping EST clones have been grouped into consensus sequences (<http://www.tigr.org/tdb/at/at.html>). Finally, a new dimension in plant biology research has just begun with the initiation of the full sequencing of the Arabidopsis genome. The five research groups currently involved will post news of their progress to the WWW on their home pages, which are listed in AtDB under the heading *International Sequencing Efforts*. Each of the groups list the BAC or YAC clones that are currently being prepared for sequencing. Thus, if you know that your favorite gene is on one of the BAC or YAC clones being sequenced, it may be prudent to engage in chromosome resting until the sequence becomes available.

Finally, the Arabidopsis community has an electronic journal called *Weeds World* that contains short research articles of particular interest to the Arabidopsis community. This journal and other services such as Lehle Seeds, a commercial supplier of Arabidopsis seeds and research materials, can be accessed from links at AtDB and other sites.

SEQUENCE DATABASES

The analysis of nucleotide and protein sequences is impossible without computer programs. The simplest query

is to recover the sequence. This is accomplished by a search of the NCBI at <http://www.ncbi.nlm.nih.gov/>. This is the mother of all biological databases and, in the eyes of many researchers, is considered to be the single most useful activity of the Federal Government. One of the most useful features is the ENTREZ suite, which allows retrieval of citations and records in the area of molecular biology from the NCBI databases. These include the molecular biology subset of MEDLINE, the National Library of Medicine's database of biomedical articles, proteins from several databases around the world, and nucleotides from GenBank and other sources. A very helpful primer on how to use ENTREZ is available at <http://www3.ncbi.nlm.nih.gov/Entrez/entrezhelp.html#GettingStarted>.

A rich collection of sequence analysis tools has been developed during the past 20 years and there is now a relatively standard approach to the analysis of a new nucleotide or protein sequence. A well organized and fast site for quick access to the most useful sequence analysis tools such as PROSITE, BLOCKS, BEAUTY (Worley et al., 1995), BLAST, and BLITZ, is available at the Human Genome Center, Baylor College of Medicine (<http://kiwi.imgen.bcm.tmc.edu:8088/search-launcher/launcher.html>). The BCM Search Launcher is an attempt to organize molecular biology-related search and analysis services available on the WWW by function by providing a single point of entry for related searches (e.g. a single page for launching protein sequence searches using standard parameters). All of the analyses are conducted by simply pasting a raw sequence into open boxes and clicking on a few parameter choices. Simple analyses produce an immediate response. Computer-intensive applications return a result by e-mail. Although it is biased toward human applications, there are plant-specific gene finder functions under the "specific species" buttons.

Although a complete description of the utility of the various programs is beyond the scope of this article, a few are of such great utility that they deserve special mention. PROSITE is a tool for assigning probable function to uncharacterized proteins translated from genomic or cDNA sequences. It consists of a database of biologically significant sites, patterns, and profiles that help to identify reliably to which known family of protein (if any) a new sequence belongs (<http://expasy.hcuge.ch/sprot/prosite.html>). If your favorite protein does not have obvious homology to a known protein, the logical next step is to explore the power of the BLOCKS database to reveal homology to domains of known function (<http://www.blocks.fhrc.org/blocks>). Blocks are multiply aligned, ungapped segments corresponding to the most highly conserved regions of proteins. The ProDom database is a comprehensive collection of homologous domains detected in the SWISS-PROT database by the DOMAINER algorithm. It has been devised to assist with the analysis of the domain arrangement of proteins. This database has recently been supplanted by PFAM, which performs similar functions (<http://genome.wustl.edu/Pfam/>).

In addition to functions for sequence comparisons, there is a small suite of functions for finding things such as mem-

brane spanning domains (see also http://www.embl-heidelberg.de/tmap/tmap_sin.html and http://ulrec3.unil.ch/software/TMPRED_form.html). Useful programs available at other sites include PSORT (Nakai and Kanehisa, 1992), which finds targeting signals (<http://psort.nibb.ac.jp/>).

Sequence database searching is an art, and a summary of the steps that should be followed in using these powerful tools is beyond the scope of this article. However, a good way to learn the art is to use the excellent Sequence Retrieval System (SRS) tutorial at the Sanger center (<http://www.sanger.ac.uk/srstutor-main.html>). SRS indexes a wide range of molecular biology databases at many sites around the world and is particularly good at cross-referencing databases, including the literature abstracts on MedLine. This tutorial aims to take you through the first stages of learning to use SRS through the WWW. It shows you some of the most useful search options and how to get what you really want from the databases, while protecting you from some of the confusion that the vast range of options can induce.

TOOLBOX FOR MOLECULAR BIOLOGY

The single most useful site for general tools for molecular biology is Pedro's Research Tools, a huge and unruly list of sites that probably lists everything available on the WWW (http://www.public.iastate.edu/~pedro/research_tools.html). Pedro's collection is mirrored in the U.S., Germany, Switzerland, and Japan, and access is fast. The problem with Pedro is that he is indiscriminate—there is tremendous redundancy. Thus, for most forms of sequence analysis one should go to the BCM Search Launcher described above. Pedro is mercifully organized into five main sections: Molecular Biology Search and Analysis; Bibliographic, Text, and WWW Searches; Guides, Tutorials, and Help Tools; Journals: Bio/Chemical Journals and Newsletters; and a List of Pedro's Biological Resources. There are lots of useful tools there, such as a simple program for selection of primers (<http://alces.med.umn.edu/webprimers.html>).

GENOME DATABASES

A good place to begin with any query about plant genomics is the USDA list of sites for plant genome databases (<http://probe.nalusda.gov:8300/plant/index.html>). This location provides ready links to all the major plant genome projects. There are several complementary resources for Arabidopsis, which are described separately above.

The complete genomes of four bacteria and the yeast *Saccharomyces cerevisiae* have recently been sequenced. Although the nucleotide sequence of the genomes is available from GenBank, the quality of the annotation associated with GenBank entries leaves much to be desired. Thus, specialized databases are available for each of these.

The first three genomes to be completed were reported by Craig Venter and colleagues at the Institute for Genomic

Research (TIGR). The TIGR Database (http://www.tigr.org/tigr_home/index.html) is a collection of molecular biology databases containing the annotated genome sequences of *Haemophilus influenzae*, *Methanococcus jannaschii*, and *Mycoplasma genitalium*, as well as extensive information on human and Arabidopsis cDNAs.

The complete 3.57-Mb genome of the cyanobacterium *Synechocystis* sp. strain PCC6803 has recently been determined by Satoshi Tabata and colleagues at the Kazusa DNA Research Institute. It is available from their database, Cyanobase (<http://www.kazusa.or.jp/cyano/cyano.html>). Cyanobase provides an easy way of accessing the sequence and all-inclusive annotation data through image maps, keyword searches, and the gene category list.

The *Saccharomyces* Genome Database (<http://genome-www.stanford.edu/Saccharomyces/>) contains a variety of genomic and biological information about the yeast *S. cerevisiae*.

VECTORS

There are now thousands of plasmids circulating, but all too frequently the scraps of paper with crucial information about the plasmids are stuck in someone else's notebook. VectorDB contains annotations and sequence information for many vectors commonly used in molecular biology (<http://www.atcg.com/vectordb/>). Information for more than 2600 vectors is available with search facilities. Vectors for which sequences are also in GenBank have direct links to that database via NCBI's Entrez browser!

SEEDS, STRAINS, AND CELL LINES

A common problem in certain areas of research is locating seeds of a plant species that has been reported to have some interesting or useful trait. Seeds of some economically useful plants are available from the USDA via the GRIN database, which contains approximately 35,000 accessions (<http://www.ars-grin.gov/npgs/tax/taxecon.html>). The database is very useful but not particularly friendly (e.g. searches are case-sensitive). Other useful sources are companies that specialize in herbs, ornamentals, wildflowers, or ethnobotanicals. One company claims to be able to provide access to more than 240,000 species of plants (<http://www.ethnobotany.com/>). The number is a bit surprising, since it approximates the estimated number of angiosperms. A generally useful web page is provided by Gardeners net (<http://trine.com/GardenNet/CatalogCenter/catshow.htm>).

Bacterial strains, fungi, plasmids, and animal cell lines (e.g. hybridomas) can also be obtained on-line from the American Type Culture Collection (ATCC) (<http://www.atcc.org/>). Additional hybridoma lines, including some that secrete antibodies against plant proteins, are available in other collections (e.g. <http://wcdm.nig.ac.jp/wcdm/HDB.html>). The World Data Centre for Microorganisms provides a comprehensive directory of culture collections, databases on microbes and cell lines, and the

gateway to biodiversity, molecular biology, and genome projects (<http://wdcm.nig.ac.jp/>). Accessions from a huge collection of *E. coli* strains can be obtained without cost from the *E. coli* stock center at Yale University (<http://cgsc.biology.yale.edu/>).

PATENTS

With over 35 million patents published worldwide and a million new ones published each year, patents are an immense source of technical and commercial information. Until recently, this source of information has generally been unavailable to practicing biologists. An important new development is that the U.S. Patent Office has made title and abstracts of recently issued patents available online (<http://www.uspto.gov/>). If you find something of interest by searching abstracts, the full text can be ordered online for \$3 and is mailed immediately. For an additional fee it is possible to receive the patent by fax or express mail. The patent office site also has links to many other patent offices around the world. Unfortunately, none of them offer comparable service.

The USDA offers a database of biotechnology patents that provides complete text of all patents since 1994 (http://www.nal.usda.gov/bic/Biotech_Patents/). IBM has recently mounted a site that provides access to the last 26 years of U.S. patents (<http://patent.womplex.ibm.com/index.html>).

INTERACTIVE MODELS OF PLANT GROWTH

An important new direction in understanding complex processes such as plant morphogenesis or global climate change is the development of spatially resolved simulations. As an example, plant morphogenesis and growth can be modeled three-dimensionally using L-systems at the Centre for Tropical Pest Management in Brisbane, Australia. "Virtual plants" are computer simulations of the structural dynamics of individual plants in three-space (<http://www.ctpm.uq.edu.au/Programs/IPI/ipivp.html>). A virtual plant is generated from a model that captures the "rules" for the formation of new plant parts and for changes in size and shape of existing parts. In addition to structural development and growth, the rules may represent processes such as wilting, compensation for damage, etiolation, tropisms, physical degeneration, senescence, and abscission. They may be restricted to the dynamics observed in a constant environment or may incorporate conditional responses triggered by environmental variations. A virtual plant may simulate a whole plant or parts of a shoot or root system. It may represent a single real plant, the morphogenetic rules of which have been derived from measurements, the average of several real plants, or a hypothetical plant. Multiple virtual plants may be generated simultaneously to explore three-dimensional spatial phenomena involving members of a stand. This site provides an interesting list of related WWW sites relating to plant morphogenesis and architecture.

RESOURCES FOR TEACHING

Many of the resources described here should be useful supplements to traditional course materials. A glimpse of the textbook of the future is available as an experimental "hypertextbook" for an introductory biology class (Biology 7.01) at Massachusetts Institute of Technology (<http://esg-www.mit.edu:8001/esgbio/>). The hypertextbook, which is designed to supplement the Biology 7.01 course materials, is organized in a very straightforward manner and contains a variety of components designed to be used independently or together to facilitate understanding of introductory biology. Each chapter has a directory page and includes a self-quiz and practice problems at the end. There is a searchable index that allows one to type in a word or topic, and the hypertextbook will return a list of places in the handbook that one can read about that topic.

Another example of WWW-supported undergraduate teaching is the "WWW Arabidopsis Plant Physiology Manual" from Jonathon Monroe at James Madison University (<http://www.jmu.edu/biology/pplab/ppjml.html>). Tom Warner and Les Hickok also describe the use of the fern *Ceratopteris* as a teaching tool and provide a detailed experimental protocol for a laboratory experiment (<http://www.bio.utk.edu/cfern/>). The goal of the experiment is to observe development of the gametophytic generation of a fern.

Many useful aids for teaching biology at the high school level are provided by Access Excellence (<http://www.gene.com/ae/>), the web site for a national program sponsored by Genentech, Inc. In the U.K., *Science and Plants for Schools* (SAPS) (<http://nasc.nott.ac.uk:8100/sapshome.html>) aims to interest young people in plant science and molecular biology. SAPS is a member of PHYTOS, an international network that links plant science and biotechnology education programs in Australia, the U.S., and the U.K.

MISCELLANEOUS

A convenient list of all federal government sites can be found at <http://www.law.vill.edu/Fed-Agency/fedwebloc.html#toc>. A database of institutional sources of fellowships and grants is available at <http://cos.gdb.org/repos/fund/fund.html>. The full text of 1000 books and reports published by the U.S. National Academy of Sciences is available at <http://www.nap.edu>.

CONCLUDING REMARKS

The resources that became available when every individual with a computer and Internet access could mount their own WWW page are overwhelming. People who otherwise would not have had the financial resources to contribute are now able to devote their talents and enthusiasms to organizing some aspect of human knowledge, and to finding its connection to other aspects of knowledge. In moving from paper to electronic publishing and databases, the post-Gutenberg era promises to be a time of explosive

growth in access to knowledge, comparable to that which resulted from the invention of movable type more than 500 years ago.

The URLs in this article are available as a list of hyper-text links in the Arabidopsis database A+DB.

Received December 18, 1996; accepted December 21, 1996.
Copyright Clearance Center: 0032-0889/97/113/1015/07.

LITERATURE CITED

- Nakai K, Kanehisa M** (1992) A knowledge base for predicting protein localization sites in eukaryotic cells. *Genomics* **14**: 897-911
- Worley KC, Wiese BA, Smith RF** (1995) BEAUTY: an enhanced BLAST-based search tool that integrates multiple biological information resources into sequence similarity search results. *Genome Res* **5**: 173-184