Outlook

In "predicting" genes, protein functions, and structures, it is helpful to define our terms precisely and be honest about our achievements. Otherwise, we will continue to be baffled by paradoxical new prediction methods that yield >80% error rates. Gene identification, structure prediction, and functional inference are nontrivial computational tasks, but with the relentless accumulation of sequence data, improvements continue to be made in all areas.

Nature functions by integration, and the adoption of a more holistic view of complex biological systems is an essential next step for bioinformatics. To get the most from genomic data, we need to take account of information on the regulation of gene expression, metabolic pathways, and signaling cascades. Proteins do not work in isolation but are involved in interrelated networks. Unraveling these networks and their interactions will be vital to our understanding of normal and pathologic cell development, and will help us create an integrated mapping between genotype and phenotype.

Genomics-based drug discovery is heavily dependent on accurate functional annotation. Toward this end, bioinformatics will need to deliver highly integrated, interoperable databases (and data "warehouses") that allow the user to reason over disparate data sources and ultimately enable knowledge-based inference and innovation. The more genome annotation is automated, the greater will be the need for collaboration between software developers, annotators, and experimentalists. And the more data we have to handle, the more rigorous we must be in our thinking (and writing) if we are to make sense of the complexities. Sequence-structure-function bioinformatics does not yet yield all the answers, but a future holistic approach should help fuse today's glimmerings of knowledge into a new dawn of understanding.

References and Notes

1. M. G. Reese *et al.*, *Genome Res.* **10**, 483 (2000).
2. K. Hofmann *et al.*, *Nucleic Acids Res.* **27**, 215 (1999).
3. T. K. Attwood *et al.*, *Nucleic Acids Res.* **28**, 225 (2000).
4. A. Bateman *et al.*, *Nucleic Acids Res.* **28**, 263 (2000).
5. J. Henikoff *et al.*, *Nucleic Acids Res.* **28**, 228 (2000).
6. R. Apweiler *et al.*, *Bioinformatics*, in press.
7. W. M. Fitch, *Syst. Zool.* **19**, 99 (1970).
8. G. R. Reeck *et al.*, *Cell* **50**, 667 (1987).
9. F. Jacob, *Science* **196**, 1161 (1977).
10. L. Gold *et al.*, *Curr. Opin. Genet. Dev.* **7**, 848 (1997).
11. M. Ashburner *et al.*, *Nature Genet.* **25**, 25 (2000).
12. B. Rost and S. O'Donoghue, *Comput. Appl. Biosci.* **13**, 345 (1997).
13. A. R. Panchenko *et al.*, *J. Mol. Biol.* **296**, 1391 (2000).
14. M. J. E. Sternberg *et al.*, *Curr. Opin. Struc. Biol.* **9**, 368 (1999).
15. T. I. Zarembinski *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 15189 (1998).
16. K. A. Olszewski *et al.*, *Comput. Chem.* **24**, 499 (2000).
17. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

TECHSIGHTING

SOFTWARE

Conquering by Dividing

The average personal computer spends much less than half a day actually performing useful computations. Many users, concerned about the vulnerability of expensive electronic components to the constant cycling of the power on and off, leave their systems on continuously. It is staggering to imagine the enormous, unused computing resources of several million PCs left running unattended. One popular approach to tapping this computing power is the Search for Extra-Terrestrial Intelligence (SETI) project (*1*), which breaks giant computing problems into pieces that can be solved on personal computers in their spare time.

Popular Power Worker
Popular Power, Inc.
San Francisco, CA
Free
www.popularpower.com

Popular Power, Inc. is a company offering a new twist on this theme. Like SETI, a company computer feeds pieces of large computing problems to networked personal computers via their software program, Popular Power Worker, for idle-time operation. Popular Power's approach differs, however, in providing a variety of computing problems to work on. These include nonprofit projects with no financial incentive to the personal computer owner, as well as commercial jobs that will eventually pay users for tasks performed on their machines.

The current version of the Popular Power Worker runs only on Windows and Linux systems and is officially in pre-release form. The preliminary status of the software is readily apparent; numerous bugs, frequent crashes, and difficulties in installation plague the program currently. If information at the company Web site is accurate, personal computer owners interested in Popular Power's computing model may find dealing with the problems of the early release worth their while. Users of the pre-release software are promised priority of access to commercial computing jobs after the official version is released. Popular Power Worker can be downloaded for free from the company's Web site, and it installs as a screen saver, which starts the program running when it becomes active. Future versions of the program for Macintosh and Solaris systems are planned.

The benefits of the Popular Power scheme for distributed computing tasks do not accrue solely to the user whose computer is used. The flexible nature of Popular Power's design provides access for businesses, scientists, and anyone with massive computing projects to computing power that is potentially far greater than they would gain from a fixed piece of hardware. Personal computer users might be able to select which commercial job to run through Popular Power Worker depending on the return offered by the originating contractor. A key to the success of the computing model is likely to be the price Popular Power demands for acting as the interface between the computing project creators and the personal computer users.

In summary, the current version of Popular Power Worker is still in the testing phase and users may find the software unstable. Tech-savvy personal computer enthusiasts are best suited to test the current pre-release product. The remaining users are advised to wait at least for the official release of the software.

—**KEVIN AHERN**

Department of Biochemistry and Biophysics, Oregon State University, Corvallis, OR 97331, USA. E-mail: ahernk@ucs.orst.edu

References

1. J. Kaiser, *Science* **282**, 839 (1998).

Tech.Sight is published in the third issue of each month. Contributing editor: Kevin Ahern, Department of Biochemistry and Biophysics, Oregon State University. Send your comments by e-mail to techsight@aaas.org

---

TECHSIGHTING

SOFTWARE

Eyes on the Skies

The orbital space above Earth contains an astonishing collection of man-made satellites. Tracking all of these objects is no small task. Liftoff is a NASA Web site that provides several software tools to locate, track, and identify Earth-orbiting satellites. At the Web site, three programs are available: J-Pass (identifies satellites passing overhead); J-Track (allows one to track orbiting objects); and J-Track 3D (allows one to view satellites orbiting Earth from a perspective far away in space). Each of these platform-independent applications is written in Java and is accessible from both Internet Explorer and Netscape

J-Track, J-Track 3D, and J-Pass
NASA
Free
http://liftoff.msfc.nasa.gov/realtime/JTrack/Spacecraft.html

Communicator browsers (version 4.0 or newer). Users load each software package within their browsers by clicking on the name of the program on a palette at the site. This organizational scheme is a bit confusing because the palette also contains links to other spacecraft-specific information, such as the locations of MIR, the Space Station, and the Shuttle. (When the Shuttle was not flying, clicking on its name gave an error message about a broken link, instead of simply saying the shuttle was on the ground.)

Of the three software products, J-Track 3D is the most impressive. Providing the user a viewpoint beyond the position of geosynchronous orbiting satellites (22,000 miles), the program shows a hemisphere of Earth with the satellites around it as points of light (see the figure). Users can either click on a point to learn more about it (such as the satellite's orbit, ground tracking, launch, or data collecting information) or select a satellite by name from the Satellite menu option. Simple controls allow one to rotate the planet in any direction. Satellite positions are updated at user-specified intervals of real time as short as 0.25 seconds.

Our world. View of Earth and man-made satellites in J-Track 3D.

J-Track illustrates the track of satellites superimposed on a two-dimensional (2D) map of the earth, similar to plots one might see on NASA television. Though the information overlaps with that available in J-Track 3D, the advantages of J-Track are that users can specify which spacecraft to follow instead of seeing all of them and that the 2D map of J-Track is easier to view than the globe of J-Track 3D.

J-Pass is a tool for anyone interested in watching satellites overhead. It provides users with a list of satellites with time plots and sky maps of when and where each one can be viewed. Users define their location on Earth by specifying longitude and latitude or, in the United States, by using their ZIP code. This clever feature works well and simplifies user interaction with the program.

On the downside, the results of having the roots of each program in Java are interfaces that are not exactly cutting edge. Menus appear in different places depending on whether one uses Netscape or Internet Explorer. Each program contained several bugs that were annoying. Ideally, visitors to the site should have a network connection that is faster than a 56K modem to avoid slow downloads, although each program can be used on a modem, if one is patient.

—KEVIN AHERN

Department of Biochemistry and Biophysics, Oregon State University, Corvallis, OR 97331, USA. E-mail: ahernk@ucs.orst.edu

TECHSIGHTING

SOFTWARE

Genomics Made Easy

This is an era in which molecular biologists can often obtain gene sequences of interest by searching databases rather than by screening libraries of cloned DNA fragments. Thus, instead of wandering around labs with tubes of radioactive probes, scientists are often found at their desks, peering hopefully at their computer screens. As whole genomes find their way into the databases, the probability is ever greater that a sequence of interest is only a BLAST search away.

Although the power of the Internet makes enormous amounts of sequence data readily accessible, molecular biologists still need tools to help them with routine sequence analysis tasks such as restriction mapping, identifying protein coding regions or special sequence motifs, and carrying out sequence similarity searches. There are, of course, numerous shareware programs on the Web, but most of these will perform a single type of analysis (e.g., identifying restriction enzyme recognition sites) and each has a different user interface. This variability can be frustrating if one needs to carry out a variety of analyses fairly often. A convenient if more expensive alternative is to buy a comprehensive sequence analysis program designed to perform the most commonly used analyses with a single, user-friendly interface. One such software package is called MacVector, the latest release of which provides the molecular biologist with a wide variety of tools for analyzing nucleic acid and protein sequences. MacVector is designed to be used on Macintosh computers, and has a well-deserved reputation for being swift, powerful, and amazingly easy to use. A short, but by no means complete, list of nucleic acid analyses available includes restriction mapping, open reading frame searches to identify protein coding regions within a sequence, base composition and codon usage analyses, nucleic acid subsequence motif searches, primer design for PCR (polymerase chain reaction), multiple sequence alignments, and translation of open reading frames to the corresponding amino acid sequences. Protein sequences may be analyzed to determine (among other things) the molecular weight and pI (isoelectric pH) of the protein of interest and to obtain profiles of hydrophilicity, antigenicity, and secondary structure along the length of the protein. MacVector will also retrieve protein or nucleic acid sequences of interest from databases by conducting an Internet search, which is trivial to set up as well as extremely fast. Another useful feature is the ability to perform BLAST searches, where a query sequence is compared with sequences in databases accessible through the Internet, without leaving the program. Such features offer an improvement over version 6.5 [see the review of version 6.5 in (*1*)].

MacVector 7.0
Genetics Computer Group

Campbell, CA
$2950
800-876-9994
www.gcg.com

New in version 7.0 is the ability to generate true phylogenetic trees with the use of distance matrix methods. Other notable features in version 7.0 include additional algorithms for identifying coding regions, and an improved multiple sequence alignment editor. BLAST searches done through MacVector are now carried out with NCBI's enhanced BLAST 2.0 programs, which permit gapped alignments. A number of small, but thoughtful, changes in the interface make the program easier to use than ever. As with earlier versions, this one is a winner. One complaint, however, concerns the Keyserver used in the network-licensed version of the program. Though the MacVector software itself is easy to use, the network installation of it is overly complicated.

—INDIRA RAJAGOPAL

Department of Biochemistry and Biophysics, Oregon State University, Corvallis, OR 97331, USA. E-mail: rajagopi@ucs.orst.edu

References

1. I. Rajagopal, *Science* **283**, 652 (1999).