

A permanent education

As data analysis shifts towards protein expression and structure determination, demands for standardization and ease of access are leaving bioinformaticians scrambling to keep their skills up to date, says Helen Gavaghan.

CSK



David Searls: ambitious scientists will find exciting work in bioinformatics.

Andy Brass (right) co-founded the popular bioinformatics course at Manchester.

Anyone who has taken apart a comparatively simple machine, such as a vacuum cleaner, in an effort to figure out why it won't work — or how it is supposed to work — knows just how difficult it is to fit all the bits back together. There is usually some disobliging piece left over that turns out to be crucial.

So pity the biologists. Having energetically deconstructed organisms ranging from fruitflies to humans into strings of bases, they are now reassembling them in a way meant to elucidate life, death, disease and evolution at every scale from organelle, through to cells, organs and organisms.

Enter bioinformatics, the merger of computing and biology, and a subject that David Searls, head of bioinformatics training at pharmaceuticals company GlaxoSmithKline and one of the founders of the field, likes to call the midwife of the post-genomic era. Ambitious scientists in this area who can cross disciplines will have exciting work for a long time to come, Searls says.

But getting and keeping jobs in bioinformatics means gaining an ever-shifting skill set. Practitioners

must, of course, learn how to analyse genomes comparatively as more organisms are sequenced. And now bioinformaticians must also tackle microarray data. In the future, as even more data from ever more kinds of sources enter repositories, bioinformaticians will need to work within standards for everything from terminology to software.

Although a clutch of scientists such as Searls first saw the promise of combining biology with computing in the late 1980s, biologists, even five or six years ago, did not appreciate just how central to their discipline computers would become.

Now there is no doubt, and there is a clamouring at the gates of universities offering bioinformatics courses. In Britain, the University of Manchester's masters course in bioinformatics is routinely oversubscribed. And this year, when it added a distance-learning masters in the subject, it did not even need to advertise for students.

In just a decade, bioinformatics has evolved from a curiosity into a fully fledged academic discipline and industrial activity. It now has all the usual accoutrements, including dedicated journals, societies, international meetings, textbooks, divisions in industry, dedicated chairs and academic courses (at post- not undergraduate level).

FRIENDLY EXPRESSIONS

Bioinformatics is also a discipline on the move, which means there is a constant need to update skills. "Until a year or so ago, we were sequence obsessed," says Alan Robinson, team leader of the industrial group at the European Bioinformatics Institute (EBI) in Hinxton, near Cambridge, UK. Acutely conscious of the need to broaden his knowledge, Robinson is taking a sabbatical in Michael Ashburner's laboratory at the University of Cambridge (until this year Ashburner also headed bioinformatics research at the EBI). Robinson plans to use the sabbatical to brush up on his 'omics' (such as



UNIV. MANCHESTER

proteomics and functional genomics) and to study the potential of computer models of biological processes.

The technology driving many of the developments in the 'omics' world is microarrays. Andy Brass, who helped to establish bioinformatics as a discipline at the University of Manchester, says that in the future it will be as routine to put expression data into public databases as it currently is to deposit sequences.

Already, the EBI has made gene expression central to its mission. At the beginning of October, Janet Thornton, a professor of protein-structure analysis at University College London, took up a five-year post as head of the EBI with responsibility for research. Not surprisingly, gene expression and training in the latest bioinformatics techniques are high on her agenda.

This month, too, the EBI began accepting expression data and ran a workshop on the analysis and informatics of DNA-array gene-expression data. "When the institute offered this workshop, 170 people applied for 25 places," Robinson says.

SHIFT TO STANDARDS

In the long term, the dominant issue will be standards. This could encompass tasks as basic as developing or consulting tables of synonyms to ensure everyone is calling genes and proteins by the same name, or developing open-source, standard chunks of mark-up code, or establishing standard interfaces so that scientists can write their own programs confident that they will work on different databases.

One even more ambitious international standardization programme is the genome ontology

collaboration. This seeks to assign common codes to genes and proteins of a particular function, cellular location or biological role, thus enabling searches across different databases. As a result, a particular apoptosis regulator, say, in the database of one model organism, could be located in another.

Although the thought of developing standards for search tools, database access or terminology, or of learning what these standards are, may not set the pulse racing, this is the way of the future. Without standards, there is the risk that crucial biological data will be left languishing in some inaccessible database — and a biologist who dismantled a piece of biological machinery may never reassemble it into working order.

Helen Gavaghan is a freelance writer based in West Yorkshire.



Janet Thornton: bioinformatics training is top of her priority list.

Web links

European Bioinformatics Institute ♦ www.ebi.ac.uk
 Gene Ontology Consortium ♦ www.geneontology.org
 Bioinformatics resource ♦ www.hgmp.mrc.ac.uk/CCP11/index.jsp
 Bioperl Project ♦ www.bioperl.org
 BioJava Project ♦ www.biojava.org
 Java tutorial ♦ java.sun.com/docs/books/tutorial
 University of Manchester bioinformatics ♦ bioinf.man.ac.uk
 UCL biochemistry group ♦ www.biochem.ucl.ac.uk/bsm
 Nature's genome gateway ♦ www.nature.com/genomics

Training options abound

Oversubscription of bioinformatics courses is not unusual. So a fair degree of self-reliance, including scouring the Internet for tutorials in bioinformatics — or its constituent parts of biology and aspects of computing — is essential for anyone wanting to learn about or keep up to date with the subject.

Fortunately, there are many

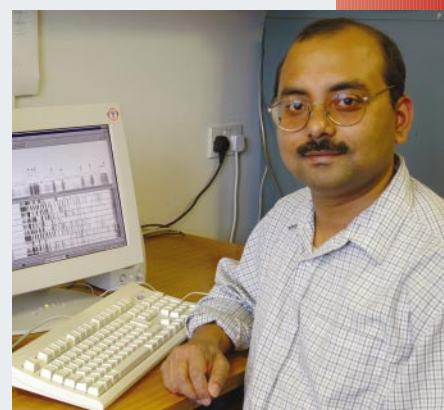
excellent websites for willing autodidacts, says Alan Robinson, team leader of the industrial group at the European Bioinformatics Institute (EBI) in Hinxton, near Cambridge, UK. But non-biologists should consider going along to their friendly local biology laboratory, he says.

David Searls, head of bioinformatics training at GlaxoSmithKline, adds that they should develop their skills using real data, and not work in an 'ivory tower' of idealized problems.

Biologists and non-biologists alike would benefit from participating in ventures such as Bioperl and BioJava, says Robinson, which aim to develop open sources of software for bioinformatics tasks. Apart from enhancing skills, and learning about which solutions work in the real world, participants become

known to one another and the networking helps with job searches.

Nevertheless, some formal training is still helpful. For Debashis Rana, now working as a project scientist on *Brassica* at the John Innes Centre in Norwich, UK, studying bioinformatics using distance-learning modules for the masters course at the University of Manchester is proving invaluable. His biology background is strong — he studied agricultural science as an undergraduate at the College of Agriculture in Vidya-Bharati, India, and then made deep forays into genetics during his PhD and postdoctoral work in the United States. Now he is conscious of the need to figure out how to store, manage and explore efficiently the data that four different institutes in Britain are producing on *Brassica*.



Experience pays: Debashis Rana

One thing is certain, whether people choose a full-time course, distance learning, on-the-job problem solving or web exploration, bioinformatics is moving fast, and those who don't move with the flow and update their skills will miss out on the biological harvest of genomics.

H.G.



Data demystified: Alan Robinson teaching.