

Free and public expression

After a slow start, progress towards developing public repositories for gene expression data is poised to accelerate. For the many biologists working with DNA microarrays, that should be welcome news.

In the era of functional genomics, DNA microarrays have emerged as uniquely powerful tools for studying gene expression. They allow biologists to take samples of tissue and study the activity of thousands of genes at a time. Such experiments churn out reams of data, of which only a concise summary appears in published papers.

Given some of the potential pitfalls of microarray research (see page 860), it is important for other researchers to have access to the underlying data. But even for the most carefully conducted studies, anyone wanting to build from previous work would be foolish not to scrutinize their peers' data before designing their own experiments.

It is for this reason that *Nature* requires microarray data relating to papers we publish to be made freely available. But at present, the system for doing so is far from perfect: data are generally posted on authors' own websites; there is little standardization, and no firm guarantee that data will remain available in perpetuity.

It would be much better for the data to be submitted to public repositories along the lines of the GenBank sequence database. But developing a robust and useful database for gene expression data has proved a more formidable challenge.

DNA sequences are linear strings of data, and require relatively little ancillary information — such as the species involved, and whether the sequence comes from nuclear DNA, or was derived from messenger RNA. Gene expression data, on the other hand, are richer and much more varied. Readouts from DNA microarrays also make sense only in the light of a suite of supporting information defining the type of chip used, the nature of the tissue samples involved, the precise procedures used to hybridize the RNA in the samples to the DNA spots on the chip, and so on.

These challenges may explain why public repositories for gene expression data have been slow to get off the ground. Now, however, several embryonic databases exist, notably the Gene Expression

Omnibus (GEO) at the National Center for Biotechnology Information in Bethesda, Maryland (<http://www.ncbi.nlm.nih.gov/geo>), and ArrayExpress, run by the European Bioinformatics Institute at Hinxton near Cambridge (<http://www.ebi.ac.uk/arrayexpress>). A new consortium is also proposing to construct a database for gene expression in cancer and other diseases (see page 855).

So far, existing databases have attracted few submissions. But that may soon change, as bioinformaticians are close to agreeing a standard mark-up language for formatting gene expression data. There have been two competing standards, MAML and GEML. Proponents of the two are now working to adopt a unified standard, incorporating aspects of both. By mid-June, this is expected to be adopted by an expert panel known as the Object Management Group.

With a single format for gene expression data, databases should be able to 'talk' to one another and exchange data. The existence of a standard language should also spur development of software tools to query the databases, and to manage and display gene expression data.

Issues remain to be resolved, however. Perhaps the most important is agreement on the suite of supporting information to be included in the databases to allow entries to be readily interpreted. The international Microarray Gene Expression Database Group (<http://www.mged.org>) has applied considerable thought to this question, developing guidelines called MIAME (for Minimal Information About a Microarray Experiment). But they are not universally accepted, and at present GEO seems to be less concerned than ArrayExpress about data being MIAME-compliant.

Nature and its sister journals will be watching developments closely, to determine the point at which it may become appropriate to require authors to submit microarray data to a public database or databases. In the meantime, we welcome the views of the diverse community of biologists interested in gene expression. ■

Science without a tinge of green

With no explanation, the first budget proposed by George W. Bush proposes drastic reductions in environmental research.

The Bush administration has tried to portray its 2002 budget proposal as a case study in moderation. But when it gets down to the small print, the proposal is anything but moderate, slashing programmes that are not to the administration's liking.

Environmental science is one of the victims. Research that deals with air and water pollution, or with biodiversity, would be cut to the bone under the proposal. Practically no agency that supports work in toxicology, hydrology, oceanography, atmospheric science or whole-organism biology is left untouched.

The numbers are, by the usual standards of such things, quite remarkable. The research programmes of the Environmental Protection Agency, for example, are reduced by more than 20%, from \$919 million to \$707 million; biological and environmental science at the Department of Energy falls by 8%; and the United States Geological Survey's water-resources division is cut by almost a quarter.

The National Science Foundation may have a level budget for

research, but this means that the agency's ambitious plans to expand its environmental-science programmes will be stillborn. The science office at the National Oceanic and Atmospheric Administration, meanwhile, loses 10% of its funding. Even the Smithsonian Institution, sensing that the new administration is not going to be a generous benefactor for environmental research, has seen fit to close an important conservation centre (see *Nature* 410, 727; 2001).

There has been no strategic explanation for any of this. One is left to conclude that the Bush administration — which has not yet named appointees to its most important scientific positions — simply does not buy the case for the scientific study of environmental questions.

Yet an unusually wide consensus in the United States holds that good science is the only reasonable basis for environmental stewardship. A majority of members of both houses of Congress claim to believe that they should act now to reverse Bush's rash attack on environmental science. ■