

Human genome databases at the crossroads

- Learning to cope with avalanche of data
- Will biologists accept rapid release?

Washington

As the Human Genome Project gets fully up to speed, genome researchers face a problem: information will soon be accumulating so fast that the prompt sharing of gene-mapping data threatens to become a bottleneck for the planned 15-year genome initiative. Data-handling experts and software engineers who design new computer databases say that action is needed now to handle the avalanche of data.

These concerns will be raised at a meeting later this month of the directors of the National Institutes of Health (NIH) genome centres. But the solution to the problem may be hampered by a clash of visions between the biologists working on the genome project and the computer scientists designing ways to disseminate rapidly the gene-mapping data. The computer scientists' proposals are forcing the biologists to determine their position on a fundamental, but still unresolved, issue: how quickly are they willing to let their data be made publicly available?

The main repository for human gene mapping data is now the Genome Database, run from Johns Hopkins University in Baltimore, Maryland. It contains a 'consensus map', in which all the information has been carefully screened, to iron out any inconsistencies in data submitted by different researchers.

But as the volume of raw data coming into Hopkins increases, the delay between submission and the appearance of screened data in the database will increase. "People are going to have to give up on consensus data if they are to have rapid access to gene-mapping information," says Nat Goodman, an information scientist at the Whitehead Institute of the Massachusetts Institute of Technology.

Within two years, Goodman predicts, the consensus process will be overwhelmed. This makes it essential, he says, that a new system of intermediate databases be developed to act as clearing houses for unscreened data.

Many information scientists agree. Peter Pearson, who runs the Genome Database at Hopkins, thinks that these intermediate databases should be the responsibility of the individual genome centres set up by NIH and the Department of Energy. Each centre is the hub for work on one or several individual chromosomes, and is supposed to act as a resource for investigators working on these chromosomes at other laboratories.

Although few of the main centres have

made significant progress on the informatics front, database experts are hard at work at these centres, and if all goes as planned, should have intermediate databases ready within a couple of years.

That will still leave the problem of communication between the different databases, which may have different formats. Researchers expect that it will soon be possible to design a 'mediator program' allowing researchers to communicate with each of the major databases via the NSFNET computer network.

In the meantime, Goodman and Thomas Marr from the Cold Spring Harbor Laboratory would like to set up a new central database to handle mapping data before it is screened for inclusion in the Genome Database. This could be set up in a year or so, for as little as \$100,000, Goodman says.

Goodman and Marr's vision, however, may get a cold shoulder from some researchers. Elbert Branscomb, of the Lawrence Livermore National Laboratory, recalls the stunned reactions of biologists at a seminar last year, when he explained the potential for rapid dissemination of data from Livermore's chromosome-19 database. The attitude of some researchers was, he says: "If you do that, we won't collaborate with you any more."

Besides an understandable reluctance to share data with directly competing research groups without at least some delay, many biologists are worried about the potential embarrassment caused by the general distribution of gene-mapping data that have not been fully checked for accuracy.

Leroy Hood, of the California Institute of Technology, believes that controls on data access should be sensitive to the confidence limits placed by researchers on the accuracy of their own data. When data are ascribed a confidence level of only 80 per cent, for example, Hood believes that they should be distributed only to biologists for whom the information is directly relevant in their own research.

Working out an acceptable data access policy for the next generation of genome databases will, however, require a greater degree of communication between the 'two cultures' of scientists working on the genome project than has so far been the case. At most genome project meetings, Hood says, "informaticists speak to informaticists and biologists speak to biologists." **Peter Aldhous**

US okays Antarctic pact

Washington

A potential rift among the Antarctic Treaty nations has been averted, now that the US Administration has agreed to sign a protocol to protect the continent's environment. Last month, the US delegation to a meeting in Madrid irritated other delegates by refusing to agree to a version of the protocol that had been revised specifically to meet US concerns. The controversy was over a provision prohibiting mining or oil exploration in the continent for at least 50 years.

The version now accepted by the United States allows the Antarctic nations to amend the mining ban after 50 years. If three-quarters of the current 26 parties to the treaty support a relaxation of the ban, but a minority of nations refuse to ratify the change, any country can withdraw from the protocol after another two years.

P.A.

Hubble fix on course

Washington

National Aeronautics and Space Administration (NASA) officials begin negotiations this week with contractors hoping to build the COSTAR optical correction device for the Hubble Space Telescope. COSTAR will place a series of lenses in front of three of the instruments at the base of the telescope, in the hope of correcting the spherical aberration of Hubble's primary mirror.

To accommodate the \$20-million COSTAR, one instrument — the High Speed Photometer — must be removed from the base of the telescope. NASA hopes to have completed most of the interesting scientific research with this instrument by the time the fix is attempted, during a 1993 space shuttle flight.

P.A.

Stanford unrepentant

Washington

In a move that seems likely to prompt further outrage in Congress, Stanford University has told federal officials that the university should be reimbursed for indirect research costs to 1991-92 at the rate of 76 per cent of the value of its research grants. This is 2 per cent more than the rate that sparked a congressional investigation earlier this year and precipitated the current indirect costs imbroglio.

"We know the political winds are shifting [and] . . . expect that this rate will be revised," says Byers, but he maintains that the starting point for negotiations must be Stanford's estimate of its true research costs. In calculating the overhead rate, Stanford has not taken into account the 26 per cent limit on administrative costs proposed recently by the White House Office of Management and Budget. Stanford was given a temporary cut from 74 to 55 per cent earlier this year.

P.A.