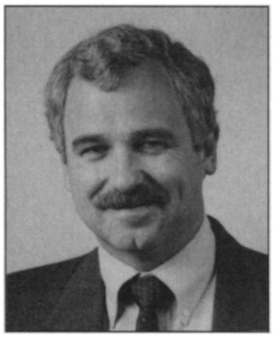
GENOME DATABASES

NIH, DOE Battle for Custody Of DNA Sequence Data

The breakup of the National Institutes of Health (NIH) and the Department of Energy's (DOE) partnership in biological computing is much like a messy divorce, with bystanders on all sides looking on in fascination and waiting to see who gets what. In this case the bystanders are the genome community, and the bureaucratic schism has featured a months-long battle involving not only high-level officials from both agencies but some of molecular biology's leading lights. It's been a "bizarre, interminable, bureaucratic battle," says one of the main protagonists, David Lipman, head of the National Center for Biotechnology Information (NCBI) at NIH. What's most frustrating, say bystanders, is that this dispute could have been resolved far more amicably had not personal animosities escalated so early.

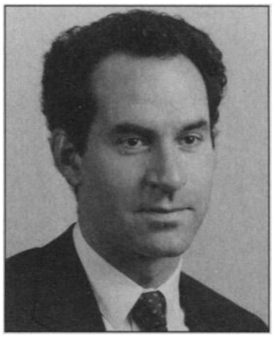

DOE's David Galas

As in many divorces, the split comes only after repeated efforts at reconciliation. But on 14 October the agencies finally decided to separate, ending a 12-year cooperation in the running of GenBank, the nation's main DNA database. The upshot is that there will be two separate but cooperating DNA sequence databases in the United States. Whatever one thinks of divorce, resolution of this messy dispute should come as a relief to researchers, who have been wondering since August where to send their sequence data. Scientists who want to send their data to GenBank should send it to NCBI in Bethesda and not Los Alamos National Laboratory, where the database has been maintained for 20 years. In the meantime, DOE is launching the Genome Sequence DataBase, based on the earlier Los Alamos operation, that will cater to big sequencing labs in addition to the molecular biology community at large.

Views are mixed on just how well this split will serve the community. Robert Robbins of DOE, who is among those pushing for a diversity of approaches to handling the crush of data emerging from the genome project, calls it a "win-win" situation. NIH officials are less than enthralled, however. Throughout this spat, Lipman and others have insisted that there should be just one sequence database in the country. Two separate outfits will only foster confusion, needless duplication of effort, and problems with quality control, they argue. If Los Alamos wants to provide more sophisticated services to the genome community, then why not do so after getting the data from GenBank at NCBI? DOE, however, has stuck to its demand to receive data directly from producers. The final call on whether two DNA sequence databases are a good idea may come from Congress, which may not be ready to cough up funds for them, no matter how diverse their approaches.

The current fracas began several years ago, when NIH announced it would switch management of GenBank from the National Institute of General Medical Sciences (NIGMS) to the newly created NCBI, which is part of the National Library of Medicine (NLM). GenBank is an offshoot of the first DNA sequence database, created by Walter Goad at Los Alamos in the 1970s. It was rechristened in 1982, when NIH and other agencies took over major funding for it from DOE. Although GenBank was largely supported by NIH, the Los Alamos team continued to play a central role in designing and building it.

The transition from NIGMS to NCBI was relatively smooth, despite some misgivings in the genome computing community about Lipman—who is described by even his friends as brash and aggressive, if brilliant. Under an agreement between NIH and DOE, scientists continued to submit sequence data to Los Alamos for processing, while the NCBI team in Bethesda took over data distribution. Then last year Lipman announced that in August his crew would be taking over direct submissions as well, phasing Los Alamos out of its job. Lipman made the decision, he says, because NCBI's budget had been cut and he thought he could save $1 million a year by handling direct submissions at NIH. He also felt he could improve quality control at GenBank, which had had troubles in the past.


NCBI's David Lipman

But to the crew at Los Alamos and at DOE headquarters in Washington, Lipman's move was a hit below the belt. While DOE concedes that GenBank had problems in the 1980s—in particular, failing to keep pace with the onslaught of sequence data—those have been largely remedied. And under David Galas, who from 1990 to 1993 ran the Office of Health and Environmental Research and thus oversaw DOE's genome program, the agency had begun to move more aggressively into the field of biological computing, or "bioinformatics" (also see p. 502).

Bioinformatics

After consulting with numerous experts, DOE had begun articulating a vision of what the genome databases of the future should be like—and Lipman, they decided, was clearly not leading GenBank in that direction. Specifically, explains Chris Fields of The Institute for Genomic Research, a big sequencing outfit near Washington, D.C., genome researchers must be able to ask complex questions if they are going to be able to interpret the voluminous mass of sequence data. And that, in turn, depends on a distributed system of linked databases that can be directly queried by the investigator. Lipman, though he embraces much of this vision of the future, had other priorities to tend to first. What's more, he says, his mission is to serve the broad molecular biology community, which means that every specialized request cannot be met.

To Galas and DOE's supporters, who include Tom Marr at Cold Spring Harbor Laboratory and Leroy Hood at the University of Washington, Lipman's move to take over direct submissions looked like an effort to exert monolithic control and to shut out Los Alamos. Galas asked Lipman to delay his plans for GenBank, but Lipman did not. By all accounts, discussions quickly degenerated, with accusations of bad faith flying back and forth by e-mail. Finally Galas and his deputy, John Wooley, decided they had no option but to forge ahead with DOE's plans, with or without Lipman's support. So in August, just as NCBI began processing direct submissions for GenBank, Galas sent out a message on Internet explaining that DOE would be expanding its informatics efforts at Los Alamos and asking researchers to continue sending their sequences to that address.

Not only did Galas' announcement enrage Lipman and higher-ups at NIH, it confounded researchers, who did not know where to send their data. NIH genome head Francis Collins, fearing that the agencies were headed toward duplicative databases, began trying to broker a compromise. He supported Lipman's plan to phase out direct submissions to Los Alamos while the lab expanded its own research efforts. Collins also pushed for the creation of a joint

NIH-DOE advisory committee that would keep Lipman attuned to the needs of the genome community.

Lipman and Wooley, who took over after Galas left for Darwin Molecular Corp. in Seattle, have spent the past several weeks trying to craft a joint statement that would clarify the situation, only to have those efforts disintegrate into yet more bickering. Lipman's boss, Donald Lindberg, head of NLM, complained: "Both Collins and I are sick and tired of problems with DOE. We don't seem to be making any headway."

As *Science* went to press, DOE was still hoping for an interagency agreement and on 13 October sent NCBI yet another proposal. One sticking point is that DOE wants Lipman to endorse its new database—something he and his bosses have so far declined to do. At this stage, NIH seems disinclined to pursue any joint agreement. Having failed to find a political solution, says Lipman, the issue now is to figure out technically how best to interact with the new player.

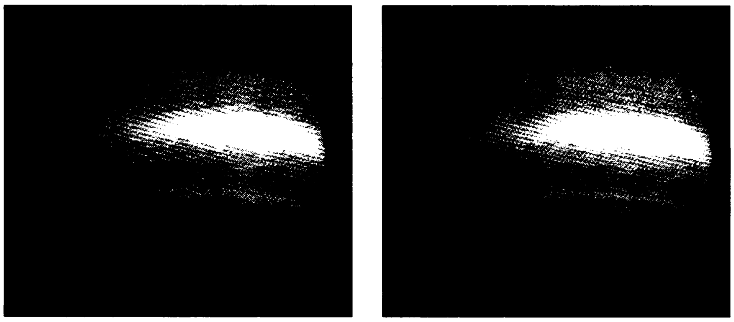
Both Lipman and Wooley vow they will share data. And each is intent on creating the better product. Lipman promises a cleaned up version, called GenBank Select, in which redundancies and errors are reduced and all coding DNA sequences have proteins associated with them. DOE will continue on its path toward a "federated information infrastructure" that will "enable data to flow electronically from producers to databases to users." Meanwhile, both agencies are vying for the affections of the few big sequencing labs—DOE can boast Hood's, while NCBI is claiming Robert Waterston at Washington University. As in any divorce, it will be some time before the community can decide who came out ahead.

–Leslie Roberts

PLANETARY SCIENCE

Jupiter Hits May Be Palpable After All

When astronomers first predicted earlier this year that a shattered comet would smack into Jupiter next July, they thought it might produce a spectacle rivaling the Earth impact that may have wiped out the dinosaurs. But when they started getting better views of Comet Shoemaker-Levy 9, estimates of the size of the largest fragments dwindled from as much as 10 kilometers in diameter to less than 1 kilometer, and many researchers began to fear a dud (*Science*, 30 July p. 552). Now, however, prospects are brightening again: Rather than dwindling out of sight, the impacts look like they will be both observable and scientifically productive.



Ripple effect. A computer simulation traces pressure variations in Jupiter's atmosphere 5 and 10 hours after the impact of a 2-kilometer comet fragment.

HARRINGTON *ET AL.*

Continued analysis of Space Telescope images and of slight differences in the fragments' orbits suggest that the largest of them are probably at least a couple of kilometers in diameter. "I think there is pretty good evidence the largest ones are not below 2 to 2.5 kilometers," says Harold Weaver of the Space Telescope Science Institute. Objects that large would put on a nice show for any observer, whether using a spacecraft or a ground-based telescope. And researchers are realizing that even if the impacts themselves don't generate the expected fireworks, they will probably trigger detectable events—from the damping of Jovian radio emissions to ripples in the Jovian atmosphere—with scientific payoffs.

The earlier diminished expectations for Shoemaker-Levy came from analysis of images made last July by the Hubble Space Telescope. Weaver and others electronically peeled away obscuring dust and debris in search of the solid fragments, but they couldn't convince themselves that the bright core of each fragment was anything but a particularly dense blob of dust. But with more time to reflect, Weaver now feels that the single, exceptionally bright pixel at the center of the brightest dust coma is an actual comet fragment measuring at least 2.5 kilometers across. If so, and if Shuttle astronauts can correct Hubble's myopic vision during the repair mission this December, the sharper images from the repaired telescope should put Weaver's size estimate on a solid footing.

The length of the train of Shoemaker-Levy fragments also suggests that they will pack a wallop. The pieces formed and began spreading apart in July of last year, when the intact comet passed within 37,000 kilometers of Jupiter's cloud tops, too close to survive the strain imposed by the massive planet's gravity. After the breakup, each of the 21 or more pieces followed a slightly different path depending on its initial distance from Jupiter, which could vary by no more than the diameter of the intact comet. The current length of the fragment train thus gives a clue to the size of the parent body—and hence of its 21 fragments.

Researchers James Scotti and Jay Melosh of the University of Arizona have modeled the breakup of Shoemaker-Levy under different assumptions of its original size and compared the resulting fragment trains with observations. The best fit was with a parent comet 2.3 kilometers in diameter, which would yield fragments averaging 0.74 kilometer in size. A group at the Jet Propulsion Laboratory—Zdenek Sekanina, Donald Yeomans, and Paul Chodas—considers this smallish estimate to be a lower bound because it does not include some factors that could make for a larger parent comet, such as possible breakup before closest approach to Jupiter. Taking such factors into account, the JPL group envisions an intact comet of perhaps 10 kilometers, which would yield 20 or so major fragments ranging from 5 kilometers down to sub-kilometer size.

Even if the smaller estimates are on target, the impacts should still make a scientific splash. Joseph Harrington, Timothy Dowling, and their colleagues at the Massachusetts Institute of Technology have run a model simulating the waves each impact should trigger in Jupiter's bottomless sea of gas. Even a 1-kilometer impacter should create waves strong enough for observers to track at visual or infrared wavelengths, says Dowling. Knowing the speed at which waves move through Jupiter's atmosphere—"arguably the most fundamental unknown quantity in Jovian atmospheric dynamics"—would provide insight into its structure. An accurate number, he adds, could let researchers "throw away half the models of the Great Red Spot."

Whatever the size of the fragments themselves, the dust and debris sweeping along with them may provide a valuable probe of another region of Jupiter—its magnetosphere, the vast envelope of charged particles trapped by the planet's magnetic field. Space physicists Alexander Dessler and Bill R. Sandel of the University of Arizona think the comet's debris and the gas it gives off should play havoc with trapped plasma and radiation, and with the radio waves they emit. By temporarily quenching the trapped radiation belt, for example, the cometary debris may reveal details of magnetospheric behavior to earthbound observers. One way or another, it seems, the show next July should be worth tuning in.

–Richard A. Kerr