up with was an *SMCX* sequence, which encoded nine of the 11 amino acids. At that point, the *SMCY* gene was not in the database, but when Engelhard called Bishop about *SMCX*, he learned about the other gene and in 24 hours had the *SMCY* sequence in hand to compare. "It was a perfect match," Engelhard says.

The Simpson-Mitchell team submitted its paper to *Nature*, while Engelhard, Goulmy, and their colleagues sent theirs to *Science*. The manuscript editors at the two journals agreed to publish the papers simultaneously. But it was not to be. One of Bishop's colleagues had done some of the *SMCY* sequencing while at Promega. Thinking it was OK, that colleague and Bishop supplied Engelhard with the entire gene. But as Bishop now says ruefully, "Apparently, I should not have done that." Promega complained and demanded time to file a patent on the gene, thus delaying publication of the *Science* paper.

Meanwhile, onlookers are wondering whether H-Y warrants such a legal frenzy. "The clinical significance [of H-Y] is not so clear," says Harald von Boehmer, an immunologist at the Basel Institute for Immunology in Switzerland and the Institut Necker, INSERM, in Paris. "Nobody really knows how high the risk [of rejection] is with this particular antigen." Attempts to use the peptide to develop anti-rejection therapies could also be complicated, because Simpson has shown there is more than one H-Y antigen. There are even hints that *SMCY* may not be the only gene encoding them.

A great many other questions also remain to be answered about the *SMCY* gene. One concerns its normal function. "Nature didn't worry about skin grafts. The true biological role of *SMCY* can't be in transplant rejection," Bishop quips. Nor is it, as was once proposed, the sex-determining gene, even though it seems to exist in many different species, showing that it has been conserved during evolution.

What else it might do remains unclear, although one clue comes from the structure of the SMCY protein. It resembles that of other proteins known to be transcription factors that turn genes on or off. But at this point, no one has direct proof that H-Y is a transcription factor, never mind what genes it might control.

All in all, the H-Y antigen mystery is beginning to take on the features of a Russian doll: Removing the first layer only reveals the deeper mysteries within, each a little more intricate. But solving the mysteries will be well worth the trouble, von Boehmer predicts: "It must have some important function; otherwise it wouldn't be so conserved."

–Elizabeth Pennisi

Elizabeth Pennisi is a science writer based in Takoma Park, Maryland.

MICROBIOLOGY

Call to Desegregate Microbial Databases

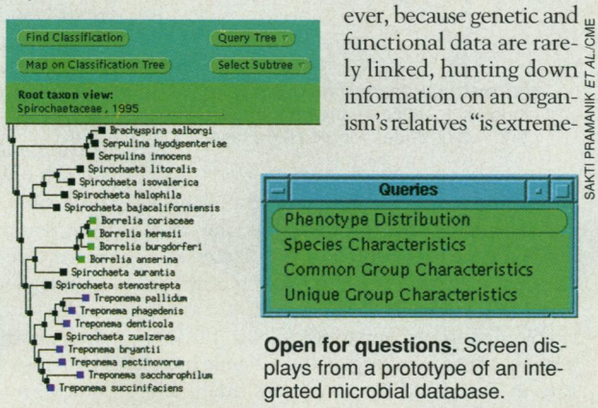
Biological databases can be powerful tools for comparing organisms or pooling different kinds of information about a single species. But when complementary information about an organism is scattered in different databases, they can also become a powerful source of frustration. That's the problem facing microbiologists, who have built a vast array of separate databases on the ecology, biochemistry, physiology, and genetic makeup of microbes—with no easy way to extract data from several sites at once. But relief is on the way: Two weeks ago, at the International Symposium on Microbial Ecology in Santos, Brazil, researchers announced an international effort to end this fragmentation of their field by linking their databases into an integrated web that will be known as the Federated Microbiology Database.

A modest version of this web, connecting three different databases, is already under construction at the National Science Foundation's Center for Microbial Ecology (CME) at Michigan State University in East Lansing. But the larger scheme announced at the Brazil meeting would enable microbiologists to trace both functional and genetic traits through a vast number of microbes without having to consult many databases individually. "The amount of time-saving will be so great," says microbiologist Gary Olsen of the University of Illinois, Urbana-Champaign, "that we'll be able to pose questions we just haven't been able to before."

A major goal of the effort, led by the CME and initiated by its director, Jim Tiedje, is bridging a gap that reflects the field's history. Like botanists, microbiologists originally classified organisms by visible characteristics such as shape and color; later, they took a whole array of physiological characteristics into account as well. Since the 1970s, genetic technologies have also made it possible to study and classify microbes at the genetic level. The result, however, has been two distinct sets of databases.

The federated database project should end this separation, allowing biologists to "understand and explore taxonomic characteristics in terms of an organism's phylogeny," says CME's Larry Forney. As a result, it should help microbiologists explore microbial evolution and deepen their understanding of individual species.

For example, when a researcher isolates a new pathogen, its genetic relatives can provide clues to the new pathogen's behavior and how to control it. At the moment, however, because genetic and functional data are rarely linked, hunting down information on an organism's relatives "is extremely tedious and time-consuming," says Olsen. When the federated database is in place, researchers will be able to automate this kind of complex data search using specialized software designed to interrogate all the relevant databases and report back.



SAKTI PRAMANIK *ET AL.*/CME

Open for questions. Screen displays from a prototype of an integrated microbial database.

Such software tools exist already, says Forney. A more difficult part of the project, he says, will be persuading database managers to open their data collections, many of which are incomplete or proprietary. The first step toward that goal came in early August, when representatives of the world's leading microbial databases met at the CME, agreed on the need to link their databases, and laid plans to link three of the most important ones: the Ribosomal Database (a University of Illinois project that contains sequences of ribosomal RNA genes), a fatty acid methyl ester profile database (a set of data on cellular fatty acids), and the Phenotypic Database in *Bergey's Manuals* (a collection of taxonomic data based at Michigan State). This three-way linkage should be completed within a year, with work being coordinated by the CME.

Expanding that core into the federation proposed in Brazil will take a good deal more effort and expense, says Forney—much of it to complete the data sets in the individual databases and prepare them for open access. The organizers expect that much of the funding will come from the U.S. and other governments, but they are also hoping to attract corporate and foundation sponsorship. If they succeed, the world of microbes may start to look a little less fragmented.

–Margaret Wertheim

Margaret Wertheim's book Pythagoras' Trousers—*a history of the relationship between physics and religion—has just been published by Times Books.*