GENETICS RESEARCH

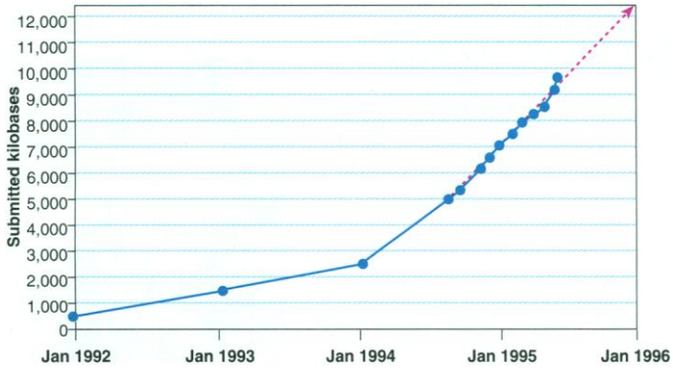
Closing In on the Complete Yeast Genome Sequence

Yeast researchers who gathered in Lisbon, Portugal, this week for their biennial conference on genetics and molecular biology learned that they have a special reason to celebrate. Calculations by a group of researchers at a key meeting held ahead of the main congress predict that the entire genome of the yeast *Saccharomyces cerevisiae* will be sequenced within months—much sooner than anyone expected.

The achievement—the result of a unique international collaborative venture involving dozens of labs, most of them in Europe—will mark a turning point in biology. Several viral genomes have been sequenced, and last month Craig Venter of The Institute for Genomic Research and Hamilton Smith of Johns Hopkins University announced that their team had obtained the complete sequences of the genomes of two bacterial species: *Haemophilus influenzae* and *Mycoplasma genitalium* (*Science*, 2 June, p. 1273). But once the yeast genome sequence is completed, says biochemist Andre Goffeau, who holds a civil service post in the European Union (EU) and coordinates part of the program from the University of Louvain in Belgium, it will provide "the first inventory of all the genes" of one of the higher species known as eukaryotes. "It's a terrific achievement," says Sir Walter Bodmer, director of the Imperial Cancer Research Fund in London. "It will be the first genuinely free-living biological organism to be sequenced."

Eukaryotes, which include humans and other mammals, have much more complex cells than bacteria. The complete catalog of the yeast genome should help researchers identify the genes needed to build and maintain the eukaryotic cellular architecture and understand their functions. "The power of the yeast genome for addressing these questions is enormous," says biochemist Peter Little of Imperial College, London.

In 1989 when the project got under way, researchers thought it would take until the end of the century to sequence the entire yeast genome, which contains a total of 14 million base pairs of DNA distributed among 16 chromosomes. In the early days, that estimate seemed about right. By 1992, only 500 kilobases had been sequenced, and work had not even begun on seven chromosomes, including the very large chromosomes 4 and 12. Project leaders estimated then that the sequence would only be about half completed by 1995. But the project passed a major milestone in 1992: the publication of the sequence of yeast chromosome 3, the first time a full sequence had been determined for a eukaryotic chromosome. After that, the collaboration began to expand, and beginning in 1994, the sequencing effort really took off.



Yeast rising. The plot shows the actual *(blue)* and predicted *(light red)* yeast genome sequences submitted to the database.

SOURCE: P. MORDANT AND A. GOFFEAU/UNIV. OF LOUVAIN

In that year alone, the effort bagged nearly 5000 kilobases of sequence. More than 70% of the sequence is now available either in the public databases or from the Martinsried Institute for Protein Sequences in Germany, the informatics coordination center for the EU project, where checks are carried out before release of the data. Participants at the Lisbon meeting who are carrying out the sequencing project estimate that by the end of the year, the amount sequenced will rise to 95%, including almost all the genes. "Everyone is on target to meet that goal," says Goffeau. "We're confident that the entire project will be finished in the new year. After we have done all the verifications, the full data should be publicly available in a matter of months."

Several factors contributed to the success of the project. One is the compact nature of the yeast genome. Unlike the human genome, which is loaded with DNA that does not code for proteins, yeast DNA is rich in genes. On average, each 2000 base-pair stretch contains a gene encoding a protein made up of more than 100 amino acids. This feature has attracted gene sequencing laboratories with the promise of rapid results.

A key factor has been the nature of the collaboration itself, particularly the large European contingent funded by the EU's biotechnology program and coordinated by Goffeau. Of the total of 79 laboratories participating in the project worldwide, 74 belong to the EU-sponsored consortium. These small laboratories will sequence 56% of the genome and, with other, larger European laboratories involved in the project but not part of the EU consortium, the total European contribution will be more than 70%.

One key element in the success of the European collaboration was its open access to sequencing laboratories outside the traditional yeast field, many of which were drawn to the project because they thought it would help with work already under way. "The European collaboration has been very successful," says Little. "Because many laboratories in the collaboration have a special interest in the genes encoded within the regions they have sequenced, the data and its interpretation are very good."

A decentralized administration and a payment system firmly tied to results also proved attractive. "At the outset laboratories got 10 kilobases [of DNA] to sequence in a year, and if they did more they got more DNA and more money," says Goffeau. That strategy encouraged laboratories to cooperate to speed up the sequencing. The result was "very successful networks and a noncompetitive atmosphere which we want to build on," says Bernard Dujon of the Institut Pasteur in Paris, who coordinated the sequencing of two yeast chromosomes.

More recently, the project got a big boost when dedicated sequencing laboratories with access to automated technology joined in. Bart Barrell at the Sanger Centre in Cambridge, U.K., and Mark Johnston of the Genome Sequencing Center at Washington University in St. Louis are now running yeast sequencing programs. Only 7% of chromosome 3's sequence was determined by automated techniques, but now almost all the sequences for some chromosomes are coming from the automated labs.

As the sequencing project draws to completion, the participants are planning their next big push: trying to find out what the genes they've identified actually do. Indeed, one of the key surprises to come out of the project has been the discovery that the yeast genome is packed with genes whose functions are completely unknown. Of the genes identified so far, 45% have no obvious homologs among previously identified genes from any organism, which means there are few clues to what they might do. Finding out will be a daunting task.

At a meeting in Paris last October, the Europeans decided that the best way to tackle the problem would be to build on the networks already developed for the sequencing effort. "We considered going back to the old way of working, but there was an over-

whelming majority in favor of developing the networks," says Dujon. The network structure will be used to coordinate activities around sets of genes or techniques and help laboratories to collaborate. Dujon expects an even larger collaboration will be needed: "Biology is not as easy as sequencing, but we want to use the experience of working on the genome," he says.

The EU will decide whether to provide funding for this project at the end of the month. While awaiting the decision, researchers have begun working out which methods to use for full-scale functional analysis of the yeast genes. A pilot study, supported by the EU's biotechnology program and coordinated by Piotr Slonimski at the Centre Génétique Moléculaire at Gif-sur-Yvette, France, is already under way. Its aim is to develop methods to produce mutant strains in which individual yeast genes have been disrupted, as well as better assays for assessing the functional effects of the disruptions. A suggestion is also being floated by Johnston to provide a library of mutants as a research resource. "Having them all in the freezer might stimulate people to look at functions," he says.

The treasure-trove of information being accumulated on the yeast genome is likely to prove valuable to researchers beyond the yeast community, because yeast genes can be used as probes to find related genes in other species including humans. By working back and forth, it should be possible to increase the understanding of gene function by leaps and bounds. "Yeast genes will be a real help for studying human gene functions," says Bodmer.

The biological stakes are high. The discovery of so many new genes has been likened to the characterization of new species which led Darwin to the theory of evolution. "It is impossible to say what might come out of the genome data. We are groping our way to facts we cannot yet imagine," says Stephen Oliver of the University of Manchester Institute of Science and Technology, who coordinated the sequencing of chromosome 3. Johnston, on the other hand, believes that data on the new genes are likely to fall into place along current lines without fundamental surprises.

Whatever yeast genes might ultimately reveal, researchers in Lisbon have reason to be in a buoyant mood. "In a short time it will be hard to realize how we managed without the sequence data. Biology will never be the same again," says Oliver.

–Nigel Williams

IMANISHI-KARI CASE

Marathon Hearing Gets Under Way

This week marked the opening of the latest—and perhaps final—stage of one of the most prominent scientific misconduct cases ever handled by the federal government. A three-person administrative panel began hearing testimony on an appeal from Tufts immunologist Thereza Imanishi-Kari of a finding by the Department of Health and Human Service's (HHS's) Office of Research Integrity (ORI) that she falsified and fabricated data for a 1986 paper in *Cell*, coauthored by Nobel laureate David Baltimore and four other researchers, on immune function in transgenic mice.

Last November ORI concluded after a 2-year investigation that Imanishi-Kari "not only fabricated and falsified critical areas of the reported results, but in denying the original misconduct, she further compounded these violations by fabricating data that she claimed supported her initial findings" (*Science*, 2 December 1994, p. 1468). It has proposed that she be banned from receiving federal funding for 10 years, prohibited from serving on federal advisory panels, and required to submit a full retraction of the *Cell* paper. In a 55-minute opening statement to the panel, ORI counsel Marcus Christ explained why Imanishi-Kari's conduct warrants the harsh sentence requested by the government. "It is not a simple act of falsification and fabrication in a paper," Christ said, "but a continuing pattern of conduct to deceive the government."

Imanishi-Kari's lawyer, Joseph Onek of Crowell and Moring in Washington, D.C., sees another pattern in the case, one distinctly less favorable to ORI. Drawing on his successful defense of National Institutes of Health virologist Robert Gallo and other scientists against charges of scientific misconduct, Onek told the panel in brief opening remarks that "ORI has used the same overheated rhetoric in other cases in which no misconduct was found, and we are confident that no misconduct will be found here."

During a break in the hearing, Onek told *Science* that a successful appeal, in his view, "will bring to an end an ignoble decade of attacks against scientists" by the federal government. ORI Director Lyle Bivens acknowledged that this "is the last of the big cases we've inherited" from ORI's predecessor, the Office of Scientific Integrity. But he offered a different assessment of the outcome. "Whichever way this turns out, I think we've succeeded in showing that we can present a strong case and do a good job investigating scientific misconduct," he said.

The hearing, featuring more than 40 witnesses, is expected to run for 3 weeks this month and then resume for 2 weeks in late August. Witnesses for the defense will include the Massachusetts Institute of Technology's Baltimore and most of the other coauthors of the *Cell* paper. A second Nobelist, Harvard's Walter Gilbert, will appear on behalf of the government.

The first half will focus on the methodology for the experiments, the reliability of various reagents used to determine expression of antibodies, and the scientific significance of the work, as well as philosophical discussions of the ethical responsibilities of scientists. It will delve into Imanishi-Kari's attitude toward conflicting and contradictory data and her lab management practices. The August portion, including testimony from Imanishi-Kari herself, will be devoted to a forensic analysis of her notebooks. Secret Service agents will offer evidence that Imanishi-Kari juggled the scientific record to make her case in the *Cell* paper, while defense experts will argue that the government's forensic analysis is flawed. The testimony is expected to be replete with the arcana of printer ribbons, data tapes, recycled paper rolls, and the impressions left by pens on subsequent pages of a notebook.

KEN HEINEN

My turn. Imanishi-Kari appeals ORI decision.

ORI's witnesses are expected to say that Imanishi-Kari's conduct violated scientific norms and that her alleged attempt to cover up her actions makes her offense particularly flagrant. Witnesses for Imanishi-Kari are expected to defend her scientific and administrative practices as falling within the range of acceptable behavior, noting that other researchers have built on her work.

A ruling is expected by the end of the year. The panel, officially the Research Integrity Adjudications Panel of the HHS Departmental Appeals Board, consists of two lawyers who are members of the board, Cecilia Sparks Ford and Judith Ballard, and an outside virologist, Julius Youngner, who is professor emeritus at the University of Pittsburgh School of Medicine.

The panel has considerable leeway in deciding whether to support or reject ORI's position; its decision on the proposed sanctions will be final in all matters except for the funding ban, which will be made in the form of a recommendation to HHS. If the board upholds ORI's conclusion, Imanishi-Kari's only recourse would be to take her case to federal court.

–Jeffrey Mervis