GENETIC SEQUENCING

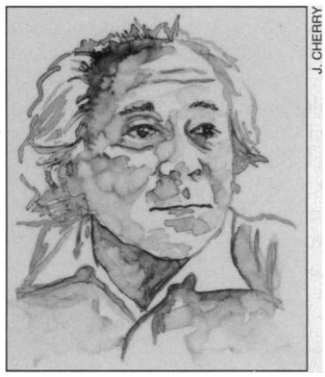
Yeast Chromosome III Reveals A Wealth of Unknown Genes

This week, for the first time ever, molecular biologists will be able to peruse the complete DNA sequence of an entire eukaryote chromosome: the 315,357 base pairs of chromosome III of the yeast *Saccharomyces cervisiae*. They are likely to get a big surprise. The chromosome—the largest continuous stretch of DNA ever sequenced—is packed with genes whose functions are completely unknown. "All of a sudden we learn that there is a whole class of genes, more than half of the genome, of which we are totally ignorant," says biochemist Andre Goffeau of the University of Louvain in Belgium. Goffeau, who also holds an appointment as a civil servant in the European Community (EC), organized a massive EC-funded project that worked out the sequence.


J. CHERRY

Andre Goffeau

Publication of this sequence has been eagerly awaited. EC researchers have been working on the yeast genome since 1989 and are set to keep going for the rest of the century in an attempt to sequence most, if not all, of the yeast's 16 chromosomes (*Science*, 24 April, p. 462). The complete sequence of chromosome III was assembled at the Martinsried Institute for Protein Sequences in Germany from data supplied by 35 labs and was published yesterday by *Nature* in a paper with 147 coauthors.


J. CHERRY

Piotr Slonimski

The sequence makes for rich reading. Indeed, yeast was targeted for the EC's massive sequencing effort because there is no junk DNA to wade through—its genome, about 14 million base pairs, is almost totally devoid of pseudogenes and repetitive sequences—and yeast genes contain hardly any introns. "Almost every base pair sequenced contains information," says Goffeau. And, "in all likelihood, the [yeast] genome represents a basic gene repertoire for eukaryotes. A major proportion of this repertoire will probably turn up in other higher organisms as well," says Goffeau.

The sequence contains 182 open reading frames, all potentially coding for proteins of 100 amino acids or more. "It comes really as a surprise that such a large fraction of the open reading frames fail to show significant homology to the 35,000 or so gene sequences deposited in the databases so far," says Goffeau. Stephen Oliver, DNA coordinator for the project and professor at the University of Manchester Institute for Science and Technology's Biotechnology Center, is confident that the open reading frames really do code for unknown functional proteins. He points out that in 1990, Japanese yeast researchers Katsumi Isono and Akikazu Yoshikawa of Kobe University estimated, through a Northern analysis experiment, that chromosome III specified 160 messenger RNAs. "That was a conservative estimate," comments Oliver, "and anyway, not far off the 182 protein coding genes that we have found. So probably, the bulk of the newly discovered genes are normally transcribed. They may, however, have quite subtle functions."

Piotr Slonimski, project coordinator for functional analysis and director of the Centre de Génétique Moleculaire at Gif-sur Yvette, France, has been probing those functions. He has found already that some of the new genes function only under unusual conditions. One example: He tried deleting a 6.5-kilobase open reading frame, the longest in the entire chromosome, and found it had no effect on the yeast under normal conditions. But the gene turned out to be essential to resist killing by acetic acid at low pH. "At pH 4.5 we discerned an effect and at pH 4.0 the absence of the gene proved lethal," says Slonimski.

Searching for the function of the new genes is going to be a time-consuming business—far tougher than the original sequencing. "With genes of totally unknown function, it is of course difficult to decide what conditions to look for. One can test for obvious characteristics like morphology, temperature sensitivity, and sporulation behavior, but the list can never be exhaustive," says Slonimski. To try to slow the growth of the gap between the rate at which new genes are sequenced and the rate at which their functions are discovered, Slonimski is developing strategies for systematic analysis of phenotypic function. "We are thinking of developing kits in which dozens and dozens of disrupted genes can be tested at once under standardized sets of conditions," he says.

Even some of the genes that had been seen before sprang some surprises. Among the 15 genes showing significant homology to known nonyeast genes was one similar to part of the nitrogen fixation operon of bacteria. Although yeast does not fix nitrogen, the gene proves essential for yeast growth. Another intriguing find is an open reading frame homologous to a gene in tobacco that is turned on by the plant growth hormone auxin. "Its function in tobacco is a complete mystery," says Goffeau, "but in principle it is now possible to unravel the function of the yeast homolog by means of disruption studies."

The yeast sequence is also going to get a warm welcome from geneticists. One longstanding problem that can be tackled is why there are recombination "hot-spots." Comparison of sequence data with genetic map data could help reveal the factors that determine recombination frequency.

While researchers sift through this pile of new data, the yeast sequencers are busy generating even more. Work on chromosome II and XI is well under way and Goffeau expects that they will be completed next year. By 1995, the EC project, along with independent international efforts, will have plowed through about 8 million base pairs and nine chromosomes. Goffeau expects that five more chromosomes could be easily completed "before the end of 1997." That would leave the massive chromosomes IV and XII, but Slonimski believes even these giants may be attempted fairly soon. "We are looking for someone who is willing to take responsibility for the coordination," he says.


J. CHERRY

Stephen Oliver

After completing the first chromosome, EC researchers believe that European science has benefited. "It may sound strange," says Oliver, "but up till now, many of us had not been paying that much attention to what was going on in Europe. Instead, we tended to look to what was happening in the United States." Now, "as a result of the chromosome III project, we have an effective 'yeast grapevine' in Europe," he says.

–Felix Eijgenraam

Felix Eijgenraam is a reporter for NRC Handelsblad *in Rotterdam.*