nature genetics
volume 14 no. 1 september 1996

Yeast ferments genome-wide functional analyses

Andre Goffeau

Imagine a time when identification of the gene relevant to a particular phenotype can be made simply by sequencing as little as 20 base pairs, and cloning of that same gene requires only proper primer design or, even better, just simply ordering the appropriate plasmid. Imagine beginning functional studies without having to do any of the brute-force, rudimentary work of selection, identification and confirmation, and with a preliminary assessment of the basic phenotype already in hand. Well, those days are already here for the yeast community. As was evident at the recent Yeast Genetics Society meeting*, the completion of the *Saccharomyces cerevisiae* genome sequence in April of this year marked a profound change in the way science will be done in yeast and also heralds the future for research in other organisms.

The timely completion of the *S. cerevisiae* genome sequence was due to the laudable collaborative effort of groups in Europe and in the United States. Andre Goffeau (Universite de Louvain) spearheaded the European Union Yeast Genome Sequencing Network, which launched the sequencing effort and made the largest contribution to the completed sequence. Mark Johnston (Washington University, St. Louis) guided much of the US effort, providing a competitive edge that aided in the earlier-than-expected release of the sequence. Later this year, *Nature* plans to publish a special issue on the yeast genome, and *Nature* and *Science*, in what can only be described as a reflection of the collaborative spirit of the yeast community, both plan to make a CD-ROM version of the genome sequence available to their subscribers.

Although differences of opinion once existed concerning organisation of the yeast genome project and release of information, such disagreements inevitably dissolved into friendly rivalry — exemplified by Peter Philippsen (University of Basel) sliding a case of French champagne to the front of the stage during his talk while announcing that David Botstein (Stanford University) had won a bet between them on the ultimate release of the complete sequence. (Botstein and Philippsen ended up sharing a bottle in mutual celebration of the genome project's overall success.) All disagreements aside, the concerted effort of the yeast community has clearly paid off, and more than one yeast geneticist commented wryly on the lack of completion so far of the much smaller *Escherichia coli* genome sequence.

The effect of the completed sequence is already echoing in ongoing studies, for example in the satisfying ring of finality in the voices of researchers announcing with complete confidence the *exact* number of gene homologues that exist for

*1996 Yeast Genetics and Molecular Biology Meeting. Yeast Genetics and Molecular Biology Society. Aug. 6–11, 1996. Madison WI, USA.

Peter Phillippsen

David Botstein

their gene of interest. Additionally, new experiments are already being designed that could not be contemplated without the genome data, and ambitious plans on how to utilize this new resource to define gene function have taken root.

Most interestingly, the cohesive approach so successful for yeast genome sequencing is continuing for functional analysis. Europe has already established the European Functional Analysis Network (EUROFAN), a rigorously defined organisation with a standard set of goals for every laboratory involved[1]. Similarly, the United States is orchestrating a joint effort for functional studies, but this will be much more loosely organised[2]: M. Johnston envisages every yeast group receiving a complete set of deletion strains (a set of 6,000 yeast strains, each with a single open reading frame (ORF) deleted[a]) and then independently performing their own brand of analysis on these strains. Johnston is currently writing a grant to fund deletion strain production and is encouraging everyone to send him e-mail[b] describing their favourite phenotypic test for analysing yeast gene knockouts. Both European and US groups generally agree that an organized effort will provide the quickest insight into gene function and are working together to generate a standard set of deletion strains. All materials and data produced will be made freely available.

Genome-wide functional analyses are already underway. Transposon-based analyses include a method described by D. Botstein that utilizes transposon Ty1: random insertion allows researchers to create PCR-fragment 'genomic footprints'; changes in PCR band patterns can then pinpoint global gene expression changes under various growth conditions. Using another transposon, Tn3, M. Snyder (Yale) has generated *lacZ*-fusion and epitope-tagged proteins to localize all proteins involved in vegetative growth, sporulation or pheromone regulation. A more systematic approach for tagging every gene by essentially 'bar-coding' each deletion strain was described by D. Shoemaker (Stanford): a unique 20mer sequence tag added during generation of each deletion strain will allow PCR analysis of a mix of deletions strains (potentially all 6,000 at once!) Amplification of the tags, which are then hybridized to a thumbnail-sized DNA chip containing an ordered grid of all the tags, quickly and accurately identifies those strains that can grow under a variety of conditions. (These 20mer tags will be used to create the deletion strains to be distributed to the community). Additional genomic analyses include (i) fluorescent hybridization studies of individually amplified ORFs on a glass-slide grid, giving a genome-wide snapshot of gene expression changes under different growth conditions and in different yeast strains (D. Lashkari, Stanford), or allowing quick, accurate mapping of whole genome recombination events (J. McKusker, Duke University), and (ii) protein analyses, such as 2D gel identification of every yeast protein (J. Garrels, Proteome) and the development of a protein-interaction database utilizing the yeast two-hybrid system (S. Field, University of Washington, Seattle, for yeast proteins; M. Vidal, Harvard University, for mammalian proteins).

Overall, genome-based analyses provide primarily functional classification rather than detailed understanding of each gene. More specific, individual testing will be required for comprehensive understanding of these genes' roles. F. Sherman (University of Rochester) commented that the yeast genome sequence is not unlike a Sigma catalogue: at one time, researchers manufactured or isolated every chemical every enzyme they needed, without the luxury of a company catalogue. The genome sequence is just such a catalogue, providing the informational material used by every genetics laboratory. The availability of this vast array of information and the development of genome-wide tools will allow researchers to study increasingly complex genetic patterns as a web is spun containing information defining gene and protein expression and interaction. The complete yeast genome sequence is not only altering yeast research. Studies in this organism — described by B. Magasanik as more like a small cow than a large bacterium — are laying the groundwork for studies in higher eukaryotes whose own genome projects are yet to be completed.

[a]6,000 open reading frames (ORFs) have been identified in *S. cerevisiae*. However, these include only those ORFs containing 100 or more codons. Some *S. cerevisiae* genes of fewer codon number do exist, but more directed methods are required to assess the presence of genes of this size. The cutoff of 100 or more codons, however, is predicted to identify most genes in *S. cerevisiae*.

[b]E-mail addresses can be found by doing a colleague search through SGD (Saccharomyces Genome Database) on the world wide web: http://genome-www.stanford.edu/Saccharomyces/.

1. Oliver, S. *Trends Genet.* **12**, 241-242 (1996).
2. Johnston, M. *Trends Genet.* **12**, 242-243 (1996).