From DNA sequence to biological function

Stephen G. Oliver

Genome sequencing is leading to the discovery of new genes at a rate 50–100 times greater than that achieved by classical genetics, but the biological function of almost half of these genes is completely unknown. In order fully to exploit genome sequence data, a systematic approach to the discovery of gene function is required. Possible strategies are discussed here in the context of functional analysis in the yeast *Saccharomyces cerevisiae*, a model eukaryote whose genome sequence will soon be completed.

THE genome sequence of the human pathogen *Haemophilus influenzae* has recently been completed[1]. Significant progress has also been made in sequencing the genomes of several model organisms, including those of the bacteria *Escherichia coli*[2] and *Bacillus subtilis*[3], the budding yeast *Saccharomyces cerevisiae*[4], the nematode *Caenorhabditis elegans*[5], and the higher plant *Arabidopsis thaliana*[6]. In addition, work has begun on the genome of an important crop rice (*Oryza sativa*[7]) and on the human genome itself[8]. The model organisms have been chosen, at least in part, because they have relatively small genomes (Table 1), allowing the sequencing of large chromosomal segments[9] or, in the case of *S. cerevisiae*, entire chromosomes[10–15]. This advantage is shared by rice. By contrast, the human genome of 3,000 megabases (Mb) is being studied mainly by sequencing complementary DNA copies of messenger RNA molecules (so-called ESTs, or expressed sequence tags[16]), although plans to make a high-resolution 'sequence map' have recently been announced[17].

Sequencing chromosomal DNA can produce data unobtainable from the sequence of small genomic fragments or cDNA clones, including information on the relationship between the higher-order structure of chromosomes and such important events as recombination, transposition, replication and gene expression. What has caught the imagination of the general scientific community, however, is the speed with which large numbers of new genes are being revealed by systematic genome sequencing. The European Union Yeast Genome Sequencing Network alone is uncovering three new open reading-frames (ORFs) per day, about 40 times the rate achieved in the past 40 years[18]. It is already clear not only that the gene pool is much richer than we had anticipated, but also that the functions of nearly half the new genes are unknown. Moreover, this is as true of intensively studied organisms like *E. coli*[1] and *S. cerevisiae*[3] as it is of more recent model organisms such as *Arabidopsis*[5]. Biologists are therefore faced with a major challenge in finding out what all these genes do.

Here I consider how this might be achieved for a simple model eukaryote, *S. cerevisiae*, and how such an exercise could assist the analysis of larger genomes.

Yeast as a lead organism for analysis

S. cerevisiae has a small genome (~13 Mb), a large number of chromosomes (sixteen), little repetitive DNA and few introns[5,19], features that have made it an excellent subject for genome sequencing. But it also has many features that commend it as a system with which to pioneer a systematic approach to screening for the function of novel genes. It is unicellular and can grow on chemically defined media, allowing the investigator complete control of its physiology. It can grow in either the haploid or the diploid state, so the effect of cell type or ploidy on gene action may be determined. Finally, and most importantly, excellent tools exist for its genetic manipulation, permitting the precise deletion of genes under investigation[20].

The problem of redundancy

Most of the *S. cerevisiae* genome is transcribed into RNA[21], suggesting that it contains relatively little redundant information. This view was disturbed by an experiment[22] in which the genome was disrupted at random, suggesting that about 70% of it was dispensable for growth on a rich, glucose-containing medium. The completion of the chromosome III sequence allowed a more specific analysis[10]: of 55 ORFs disrupted or deleted, only three were essential, and further analysis of 42 genes revealed a phenotype for only 21. Why does this small genome appear to have such a high level of redundancy?

Duplication of some important genes may provide a selective advantage by insuring against the loss of essential functions by mutation. Such genes may include those required for survival during stress or starvation. The genes encoding Hsp90 heat-shock proteins[23] and *RAS*[24] may be representatives of this class. Exact functional duplicates will be analysed most readily when the sequence of the entire yeast genome is complete and 'synthetic phenotypes', which result from multiple gene knockouts designed on the basis of sequence homology, can be examined.

Many yeast genes may be required to deal with challenges that are never encountered in the laboratory, but which are common in the rotting fig[25] or the brewery[26]. In the laboratory, any of several closely related genes might be able to substitute for one or all of the others, but there may nevertheless be specific physiological challenges in which only one of the set can supply the exact function required for continued growth or survival. Yeast proteases furnish one example[27]. A more specific example comes from chromosomes III and XIV, which apparently evolved from a common ancestor[28]. Each has a citrate synthase gene at an equivalent position, but *CIT2* on chromosome III encodes the cytoplasmic enzyme, whereas *CIT1* on chromosome XIV encodes the mitochondrial enzyme (an apparent example of evolution through gene duplication).

It may also be that once the constraints on genome size imposed by the single, circular bacterial chromosome are released, organisms can afford to use multiple genes to perform identical (or nearly identical) functions in different contexts. Even in complex bacteria such as *Streptomyces* (which has a large, linear

TABLE 1 Sizes of genomes in current systematic sequencing projects

Organism	Genome size (Mb)
Bacillus subtilis	4.2
Escherichia coli	4.7
Saccharomyces cerevisiae	13.5
Caenorhabditis elegans	100
Arabidopsis thaliana	100
Oryza sativa	400
Homo sapiens	3,000

chromosome[29]), multiple genes encoding enzymes of essentially identical biochemical function are used in secondary metabolism[28] and differentiation[30].

What physiological challenges need be analysed to reveal the functions of novel genes uncovered by sequencing? Obvious stresses such as heat, cold, osmotic shocks or starvation can be examined, but vast numbers of such conditions may be relevant to the natural history of yeast, many of them unknown. Another example from chromosome III illustrates the point. Although YCR32w (6,501 base pairs) is the longest ORF on the chromosome, its complete deletion had no obvious effect, and the mutant was exposed to many physiological challenges without results before it was discovered that it died when grown at low pH on glucose and challenged with acetic acid[31]. As YCR32w encodes a membrane protein, its product is probably an acetic acid exit pump. But can we reproduce this exercise for 3,000 more genes of unknown function once the yeast genome sequence is complete? A more systematic approach must be found.

If much redundancy is more apparent than real, the apparent rarity of gene deletions with a discernible phenotype may reflect a lack of adequate functional tests. There may thus be whole classes of functions to be found. But how can we uncover them? It is hard to believe that many undefined tracts of metabolism remain (although secondary metabolism in organisms other than antibiotic-producing bacteria is admittedly much neglected). New genes are therefore more likely to be involved in as-yet unstudied levels of regulation that may not be constitutively employed, such as the integration of different areas of metabolism, the temporal control of activity, or apoptosis. In order to reveal such functions, new methods of analysis will be needed; I now consider the strengths and weaknesses of several systematic approaches.

Analysis

Although the sequence of an ORF is itself a major resource that enables function to be inferred from the similarity that its putative protein product shows to other protein sequences in the data libraries, it is important to realize the limitations of such analyses. Significant matches of a novel ORF to another sequence may be in any of four classes. (1) A match that predicts both the biochemical and physiological function of the novel gene (such as that between ORF YCR24c from yeast chromosome III and *E. coli* Asn-tRNA synthetase). (2) A match that defines the biochemical function of a gene product without revealing its physiological function (such as the five new protein kinase genes found on chromosome III, whose biochemical function is clear—they phosphorylate proteins—but whose particular physiological function in yeast remains obscure). (3) A match to a gene from another organism whose function is unknown in that organism (such as that between ORF YCR63w from yeast chromosome III, protein G10 from *Xenopus*, and novel genes from *C. elegans* and humans, none of whose function is known). Such matches are becoming increasingly common as data accumulates from the different genomic sequencing projects. Although they illustrate the potential of systematic sequencing projects, they demonstrate that this potential may only be realized by an equally systematic experimental approach to functional analysis. (4) A match to a gene of known function that merely reveals that our understanding of that function is very superficial, such as that between YCL17c on yeast chromosome III and the NifS protein of nitrogen-fixing bacteria. In this instance, although similar genes were found in *Lactobacillus*, *Bacillus* and *E. coli*, neither yeast nor any of these other bacteria fix molecular nitrogen. Further studies in *Azotobacter*[32] were necessary to show that NifS is responsible for inserting sulphur into the metal–sulphur centre of metalloenzymes, using pyridoxal phosphate as a cofactor. Whereas it may play an analogous role in *Bacillus subtilis*[31], both sequence alignments[33,34] and some experimental data[35] indicate that NifS proteins are pyridoxal-phosphate-dependent aminotransferases. Thus the discovery of similar genes in unexpected places has stimulated experiments that uncovered the true role of a protein whose function was previously thought to be known. Similar matches may exist that do not seem incongruous to us and so do not immediately reveal the superficial nature of our knowledge.

Of course it is nice (and cheap) to think that the computer is the simplest route to function, but experiments remain essential. However, we cannot afford the time or money to perform all tests on all unknown genes, so a hierarchical taxonomy of gene function is needed that will allow us to perform only the tests relevant for genes of a given class (as in the 'keys' that are used as an aid to identifying species in classical taxonomy).

Yeast molecular genetics has recently been very successful in widening the net of gene interactions, identifying many new genes in the process. Two methods predominate: the isolation of successive extragenic suppressor mutations, and the creation of 'synthetic' phenotypes by combining two or more non-allelic mutations. The former can be applied to the current problem by cloning novel genes into multicopy vectors and examining their ability to suppress null mutations in known genes. The examination of all possible pairwise combinations is not really feasible, however, and a set of known genes whose suppression is likely to be particularly informative must therefore be chosen. The synthetic phenotype approach may also be helpful, but here too a battery of phenotypic tests will be required, as well as some hierarchical structure to guide their selective application, if it is not to become unwieldy. Similarity searches may make it obvious which combinations of deletions to analyse, but criteria other than sequence similarity may also be needed.

One way of classifying newly discovered proteins is to determine in which cellular compartment they are found. Burns *et al.*[36] have used mini-Tn3::*lac* Z to generate *lac* Z fusions with yeast ORFs in *E. coli*, producing fusion proteins that can be located by immunocytochemistry on transfer back into yeast. Of 2,373 fusions examined, 32% gave no staining, 58% gave general cytoplasmic staining, and 10% gave localized staining (Table 2). A parallel screen identified 39 fusions whose expression was specifically induced by meiosis.

Bolotin-Fukuhara is pursuing a complementary approach, in which a mini-Mu transposon is again used to make *lac* Z fusions[37]. In this case, however, the *lac* Z fusion library is expressed in a pair of yeast hosts that differ only in the presence or absence of a particular transcriptional activator, allowing genes under the activator's control to be identified by their altered colour in the two strains. Although a similar approach was previously used to isolate co-regulated *E. coli* genes[38], this is probably the first application of the method to a eukaryote. A pilot experiment[37] identified 26 fusions regulated by the CCAAT-box-binding protein, Hap2p (refs 39, 40). This approach will be particularly helpful for new transcriptional activators identified by sequence analysis.

Classification

Virtually all the ORFs predicted by the chromosome III sequence are expressed by haploid cells growing on a rich glucose-based medium, at least as mRNAs[41,42]. Nevertheless, expression levels are likely to vary greatly under different conditions, and a rapid procedure for northern analysis is required. This might entail hybridizing antisense oligonucleotide probes to a set of cDNA filters derived from cells (*MAT***a**, *MAT*α and *MAT***a**/*MAT*α diploid) grown under different conditions. A second, more direct approach is to establish a set of yeast ESTs[11]. Initial experiments have shown that the ESTs obtained from yeast grown on a minimal medium include many more genes of unknown function that those obtained from cells grown on rich medium[43]. Careful cost-benefit analysis will be required to choose between the two approaches. Whichever is selected, it will be important to choose a small set of informative physiological conditions so as to simplify the grouping of genes according to function.

A related approach is based on two-dimensional protein gels,

TABLE 2 Screen for cellular location of protein products of yeast genes

Site of staining	No. of genes	No. of previously known genes	No. of sequenced genes not previously known	No. of unsequenced genes not previously known
Nucleus	98	11	3	7
Nucleolus	5			1
Endoplasmic reticulum/ nuclear rim	2		1	1
Punctate	143	3	34	
Cell periphery	10		1	1
Bud tip	2			1
Prospore wall	1			1

Previously known genes refer to database matches to the few genes for which partial sequence data is available where that matched was to a named yeast gene defined by 'classical' (function-first) genetic analysis. Not previously known genes refer to genes which have not been discovered by the classical route and whose sequence either is, or is not, currently available in the Genbank database. (This interpretation of the current situation was compiled from data supplied by Petra Ross-MacDonald.)

rather than northern blots. Although it is technically more complex, time-consuming and expensive than the northern approach, it can exploit deletion and amplification mutants, and can be combined with cell fractionation to demonstrate subcellular location (although cell fractionation with yeast is neither as efficient nor as discriminatory as the immunoctyochemical approach described here[44]). Most importantly, however, such analysis can reveal the set of proteins dependent on the expression of a given gene or that phosphorylated by a particular kinase.

Systematic phenotype screening

The phenotypic effect of deleting individual novel genes from yeast chromosome III has been examined in a qualitative but systematic manner. Tests were carried out on agar plates containing a large number of different growth media, sometimes incorporating specific metabolic inhibitors (K. Reiger *et al.*, manuscript submitted). Efficient methods of generating the required deletion mutants are now being developed[45,46], and it is hoped eventually to include deletions in all novel genes. Such a collection will be a resource of enormous importance.

Top-down metabolic control analysis

An approach that may be ideally suited to uncovering functional hierarchies rests on the principles of metabolic control (or 'flux') analysis[47,48]. The two versions of this approach can be described as 'bottom-up' and 'top-down'. The former seeks to determine the contribution made by each enzyme in a metabolic pathway to the flux of carbon through that pathway. The analysis may be extended to enzymes more and more remote from the pathway being examined until all relevant enzymes have been identified and their contribution quantified. However, this approach is not applicable to the current problem, as the pathway affected by the gene under study is, by definition, unknown. The 'top-down' approach[49], by contrast, is admirably suited to the problem. Here, it is the flux through large metabolic domains that is measured. Smaller and smaller domains can then be examined, until the control of specific pathways or steps is elucidated. Top-down control analysis, therefore, has exactly the type of hierarchical structure necessary to facilitate the search for gene function, and the concepts involved can readily be applied to spheres of biological activity other than metabolism.

Conclusions and prospects

Systematic genome sequencing presents biology with enormous opportunities, and some dangers. The dangers include the mistaken belief that analysis *in silicio* will by itself be sufficient to reveal the function of all novel genes uncovered. An equally wrong-headed alternative is to regard the genome sequences as vast reference works whose purpose is simply to save researchers the bother of sequencing the genes on which they alight in the course of their investigations. Seductive though this view is, there may be good reasons why classical genetics has not discovered these genes, and we should not be so arrogant as to imagine that there are no new functions left to discover. Many genes of unknown function are under-represented in cDNA libraries (R. Waterston, personal communication), indicating that they are only expressed transiently or at very low levels, perhaps because they play a regulatory role. Even if they should eventually be revealed by current approaches, this is an enormous waste of the hard-won resource the sequences represent, which could instead be used to transform the pace of biological research.

Although we do not understand the function of about half the genes in most organisms, we have a good understanding of the function of the other half in organisms such as yeast and *E. coli*. Like the sequences themselves, this is an enormous resource which should be exploited in any systematic scheme for functional analysis. The known genes may be used to optimize the tests employed in our function searches, and double-blind trials can be carried out to ask whether new schemes can correctly classify the functions of known genes. Moreover, metabolic control analysis suggests that the overexpression of particular genes of known function will permit the construction of a set of strains that will be sensitized to the detection of the effects of novel genes in specific domains of biological activity.

S. cerevisiae has a special role to play in this process. The function of yeast's 7,000 or so genes, once known, will provide a working description of a eukaryotic cell. It will also provide a unique tool with which to search for drugs to treat not only infectious diseases but also major killers such as cancer[50].

Analysis of the yeast genome is likely to contribute in several ways to the human genome project[51]. It will assist in identifying the function of newly discovered human genes, both by sequence comparisons, and by complemetation of defined genetic lesions in yeast with human cDNA clones. This process might be facilitated by the construction of a 'minimalist yeast', a strain in which all apparently redundant gene sets are reduced to the minimal set permitting growth on complex media in the laboratory. The remaining genes may then be successively replaced by human or viral sequences. As the viability of such a strain would then be absolutely dependent on the heterologous gene, it would form a powerful tool for screening antitumour and antiviral drugs.

The analysis of the yeast genome will also assist studies of proteins implicated in human heritable diseases. Many such proteins have yeast homologues[52,53] (Table 3) and the study of their physiological role and their interactions with other yeast proteins will facilitate understanding of the corresponding disease. Many important medical conditions in humans, whether actual diseases (such as early-onset diabetes[54]) or a predisposition to them (such as colon cancer[55] or heart disease[56,57]) are controlled by multiple genes[58], and uncovering the complete set of genes involved in a particular condition is a difficult and lengthy process[54–58]. Recognition of homology between a yeast protein and a human protein implicated in a particular disease may thus provide an important key to improvements in diagnosis or therapy. *In vivo* approaches to identify interactions between

TABLE 3 Sequence similarity between positionally cloned human genes and *S. cerevisiae* ORFs discovered by systematic sequencing

Human disease	MIM no.	Human gene	GenBank no. for human cDNA	BLASTX P-value	Yeast gene	GenBank no. for yeast DNA
Lowe's syndrome	309000	*OCRL*	M88162	1.2e-47	*YIL002c*	Z47047
Breast cancer, early onset	600185	*BRCA2*	U43746	4.3e-4	*YER033c*	U18796
Neurofibromatosis, type 2	101000	*NF2*	L11353	4.3c-4	*N2231*	X85811
Kallmann's syndrome	308700	*KAL*	M97252	1.0e-3	*YKL083w*	Z28082
Tuberous sclerosis	191090	*TSC*	X75621	1.5e-3	*ORF00954*	X83121
Marfan's syndrome	154700	*FBN1*	L13923	1.9e-2	*YCR89w*	X59720
Huntington's disease	143100	*HD*	L12392	3.4e-2	*D4411*	X82086
Long Q–T syndrome, type 1	192500	*KVLQT1*	U40990	3.1e-1	*YBR235w*	Z36104
Fragile-X syndrome	309550	*FMR1*	S65791	4.1e-1	*UND407*	U43491
Emery–Dreifuss muscular dystrophy	310300	*STA*	X82434	5.7e-1	*YBL046W*	Z35807
Norrie's disease	310600	*NDP*	X65882	9.4e-1	*YIL037c*	Z47047

BLASTX searches were performed using each positionally cloned human cDNA listed against a non-redundant database of *S. cerevisiae* sequences (NRSC) maintained at the *Saccharomyces* Genomic Information Resource. For the most statistically significant yeast protein match for each human query, the BLAST (v1.4) *P*-value for the match is given together with the GenBank accession number for the corresponding *S. cerevisiae* genomic sequence. Data shown represent a subset of a more complete analysis (D. E. Bassett Jr which is accessible on the World-Wide Web (http://www.ncbi.nlm.nih.gov/Bassett/Yeast/PosiCloneSceNew.html). Further details of this analysis and its results are given in Scientific Correspondence by D. E. Bassett *et al.* in this issue.

genes or their products (such as the identification of extragenic suppressors, control-of-expression studies[38], and the use of the two-hybrid system[59]) may also reveal other proteins whose human homologues are involved in the disease.

It is clear that the functions of many novel genes discovered by DNA sequencing will be uncovered and set in their correct physiological context over the next few years. This process will be most important for human genes, but the search for function in the genes of *S. cerevisiae* will undoubtedly be an essential navigation aid for what is to follow. □

Stephen Oliver is in the Department of Biochemistry and Applied Molecular Biology, UMIST, PO Box 88, Manchester M60 1QD, UK.

1. Fleischmann, R. D. *et al. Science* **269,** 538–540 (1995).
2. Blattner, F., Daniels, D. L., Burland, V. D., Plunkett, G. & Chang, S. in *The Chromosome* (eds Heslop-Harrison, J. S. & Flavell, R. B.) 43–59 (βios, Oxford, 1993).
3. Devine, K. *Trends Biotechn.* **13,** 210–216 (1995).
4. Oliver, S. G., James, C. M., Gent, M. E. & Indge, K. J. in *The Chromosome* (eds Heslop-Harrison, J. S. & Flavell, R. B.) 233–248 (βios, Oxford, 1993).
5. Sulston, J. *et al. Nature* **356,** 37–41 (1992).
6. Schmidt, R. & Dean, C. *Bioessays* **15,** 63–69 (1993).
7. Havukkala, I., Ichimura, H., Nagamura, Y. & Sasaki, T. *J. Biotech.* **41,** 139–148 (1995).
8. Olson, M. V. *Proc. natn. Acad. Sci. U.S.A.* **90,** 4338–4344 (1993).
9. Wilson, R. *et al. Nature* **368,** 32–38 (1994).
10. Oliver, S. G. *et al. Nature* **357,** 38–46 (1992).
11. Dujon, B. *et al. Nature* **369,** 371–378 (1994).
12. Johnston, M. *et al. Science* **265,** 2077–2082 (1994).
13. Feldmann, H. *et al. EMBO J.* **13,** 5795–5809 (1994).
14. Bussey, H. *et al. Proc. natn. Acad. Sci. U.S.A.* **92,** 3809–3813 (1995).
15. Murkamai, Y. *et al. Nature Genet.* **10,** 261–268 (1995).
16. Adams, M. D. *et al. Science* **252,** 1651–1656 (1991).
17. Waterston, R. & Sulston, J. *Nature* **376,** 111 (1995).
19. Olson, M. V. in *The Molecular and Cellular Biology of the Yeast* Saccharomyces cerevisiae (eds Broach, J. R., Pringle, J. R. & Jones, E. W.) 1–39 (Cold Spring Harbor Laboratory, New York, 1991).
20. Rothstein, R. J. *Meth. Enzym.* **101,** 202–211 (1983).
21. Kaback, D. B., Angerer, L. M. & Davidson, N. *Nucleic Acids Res.* **6,** 2499–2517 (1979).
22. Goebl, M. E. & Petes, T. D. *Cell* **46,** 983–992 (1986).
23. Craig, E. A., Gambill, B. D. & Nelson, R. J. *Microbiol. Rev.* **57,** 402–413 (1993).
24. Tatchell, K., Robinson, L. C. & Breitenbach, M. *Proc. natn. Acad. Sci. U.S.A.* **82,** 3785–3789 (1985).
25. Mortimer, R. K. & Johnston, J. R. *Genetics* **113,** 35–43 (1989).
26. Hammond, J. R. M. in *The Yeasts* Vol. 5 (eds Rose, A. H. & Harrison, J. S.) 7–67 (Academic, London, 1993).
27. Rendueles, P. S. & Wolf, D. H. *FEMS Microbiol. Rev.* **54,** 17–46 (1988).
28. Lalo, D., Stettler, S., Mariotte, S., Slonimski, P. P. & Thuriaux, P. *Compt. Rend. Acad. Sci.* (III) **316,** 137–143 (1993).
29. Lin, Y. S., Kieser, H. M., Hopwood, D. A. & Chen, C. W. *Molec. Microbiol.* **10,** 923–933 (1993).
30. Chadwick, D. J. & Whelan, J. (eds) *Secondary Metabolites: Their Function and Evolution* (CIBA Foundation Symposium 171), (Wiley, Chichester, 1992).
31. Chater, K. F. in *Regulation of Prokaryotic Development* (eds Smith, I. Slepecky, R. A. & Setlow, P.) 277–299 (Am. Soc. Microbiol., Washington, 1989).
31. Jia, Y. thesis, Univ. Pierre et Marie Curie, Paris (1993).
32. Zheng, L., White, R. H., Cash, V. L., Jack, R. F. & Dean, D. R. *Proc. natn. Acad. Sci. U.S.A.* **90,** 2754–2758 (1958).
33. Sun, D. & Setlow, P. *J. Bact.* **175,** 1423–1432 (1993).
34. Mehta, P. K. & Christen, P. *Eur. J. Biochem.* **211,** 373–376 (1993).
35. Ouzunis, C. & Sander, C. *FEBS Lett.* **322,** 159–164 (1993).
37. Leong-Morgenthaler, P., Oliver, S. G., Hottinger, H. & Söll, D. *Biochimie* **76,** 45–49 (1994).
36. Burns, N. *et al. Genes Dev.* **8,** 1087–1105 (1994).
37. Dang, V.-D., Valens, M., Bolotin-Fukuhara, M. & Daignan-Fornier, B. *Yeast* **10,** 1273–1283 (1994).
38. Casadaban, M. J. & Cohen, S. N. *Proc. natn. Acad. Sci. U.S.A.* **76,** 4530–4533 (1979).
39. Olesen, J., Hahn, S. & Guarente, L. *Cell* **51,** 953–961 (1987).
40. De Winde, J. H. & Grivell, L. A. *Progr. nucleic Acids Res.* **46,** 51–91 (1992).
41. Yoshikawa, A. & Isono, K. *Yeast* **6,** 383–401 (1990).
42. Tanaka, S. & Isono, K. *Nucleic Acids Res.* **21,** 1149–1153 (1992).
43. Weinstock, K. G., Kirkness, E. F., Lee, N. H., Earle-Hughes, J. A. & Venter, J. C. *Curr. Opin. Biotech.* **5,** 599–603 (1994).
44. Kreudtzfeldt, C. & Witt, W. in Saccharomyces (eds Tuite, M. F. & Oliver, S. G.) Biotech. Handbooks Vol. **4,**, 5–58 (Plenum, New York, 1991).
47. Rieger, K., Orlowska, G., Kaniak, A., Aljinovic, G. & Slonimski, P. P. *Yeast* (in the press).
45. Baudin, A., Ozier-Kalogeropoulos, O., Denouel, A., Lacroute, F. & Cullin, C. *Nucleic Acids Res.* **21,** 3329–3330 (1993).
46. Wach, A., Brachat, A., Pöhlmann, R. & Philippsen, P. *Yeast* **10,** 1793–1808 (1995).
47. Kacser, H. & Burns, J. A. *Symp. Soc. exp. Biol.* **32,** 65–104 (1973).
48. Heinrich, R. & Rapoport, T. A. *Eur. J. Biochem.* **42,** 89–95 (1974).
49. Quant, P. A. *Trends biochem. Sci.* **18,** 26–30 (1993).
50. Dulbecco, R. *Gene* **135,** 259–260 (1993).
51. Short, N. *Nature* **377** (suppl), 1 (1995).
52. Tugendreich, S., Bassett, D. E. Jr, McKusick, V. A., Boguski, M. S. & Hieter, P. *Human molec. Genet.* **3,** 1509–1517 (1994).
53. Bassett, D. E. Jr, Boguski, M. S. & Hieter, P. *Nature* **379,** 589–590 (1996).
54. Davies, J. L. *et al. Nature* **371,** 130–136 (1994).
55. Bodmer, W., Bishop, T. & Karran, P. *Nature Genet.* **6,** 217–219 (1994).
56. Ward, R. in *Hypertension: Pathophysiology, Diagnosis and Management* (eds Laragh, J. H. & Brenner, B. M.) 81–100 (Raven, New York, 1990).
57. Schwartz, K. *Nature Genet.* **8,** 110–111 (1994).
58. Lander, E. S. & Schork, N. J. *Science* **265,** 2037–2048 (1994).
59. Chien, C. T., Bartel, P. L., Sternglanz, R. & Fields, S. *Proc. natn. Acad. Sci. U.S.A.* **88,** 9578–9582 (1991).

ACKNOWLEDGEMENTS. Work on yeast genome sequencing and analysis in my own laboratory is supported by the European Commission, BBSRC, and by Pfizer Central Research. I would like to thank J. Fincham for asking the right questions; H. Boucherie, B. Dujon, S. Fey, P. Hieter, P. Mose Larsen and M. Snyder for critically reading the manuscript; K. Indge for his help with computer analyses; and D. Bassett, A. Goffeau, P. Slonimski and P. Ross-MacDonald for their free communication of data.