ferent approaches. As well as partial overlaps between the regions sequenced by two laboratories, putative frameshift checking, alignment of the sequence with previously published data and a few random resequencing verifications were performed on cosmid subclones. A new method for verifying specific regions of the sequence was developed for this chromosome (G. V. *et al*., manuscript in preparation) and applied to several other chromosomes. The extrapolation from the number of discrepancies observed in the overlaps (102,049 nucleotides, 9.4 %) to the whole sequence suggests that the nucleotide sequence of chromosome VII is 99.974 % accurate. The quality of the coding regions, where frameshifts are quite easy to check, is probably much higher than that of the intergenic regions. Indeed, the quality assessment procedure led to the correction of a total of 90 errors mainly located in the coding regions. A total of 56,344 bp (5.2 %) have been resequenced, and the comparison with the original data makes it possible to estimate that about 120 errors remain in the chromosome VII sequence. Parts of the sequence were published independently[16–30] before assembly of the contig and application of the final quality controls; several other manuscripts are in the press. □

Received 19 July 1996; accepted 11 March 1997.

1. Oliver, S.G. et al. *Nature* 357, 38-46 (1992).
2. Capieaux, E., Ulaszewski, S., Balzi, E. & Goffeau, A. *Yeast* 7, 275–280 (1991).
3. Chen, W. *et al. Yeast* 7, 287–299 (1991).
4. Dujon, B. *et al. Nature* 369, 371–378 (1994).
5. Dujon, B. *Trends Genet*. 12, 263–270 (1996).
6. Klein, P., Kanehisa, M. & Delisi, C. *Biochim. Biophys. Acta* 815, 468–476 (1985).
7. Goffeau, A., Slonimski, P., Nakai, K. & Risler, J.L. *Yeast* 9, 691–702 (1993).
8. Nakai, K. & Kanehisa, M. *Genomics* 14, 897–911 (1992).
9. Woolford, J. L. & Warner, J. R. in *The Molecular and Cellular Biology of the Yeast Saccharomyces* Vol.1 (ed. Jones, E., Pringle, J. & Broach, J.) 597–598 (Cold Spring Harbor Laboratory Press, NY, 1991).
10. Milosavljevic, A. & Jurka, J. *Comput. Adv. Biosci*. 9, 409–411 (1993).
11. Rymond, B. C. *Proc. Natl Acad. Sci. USA* 90, 848–852 (1993).
12. Olson, M. V. in *The Molecular and Cellular Biology of the Yeast Saccharomyces* Vol. 1 (ed. Jones, E., Pringle, J. & Broach, J.) 1–40 (Cold Spring Harbor Laboratory Press, NY, 1991).
13. Tettelin, H. *et al*. in *Methods in Molecular Genetics*, (ed. Adolph, K.W.) (Academic,London, 1995).
14. Thierry, A., Gaillon, L., Galibert, F. & Dujon, B. *Yeast* 11, 121–135 (1995).
15. Louis, E. & Borts, R. *Genetics* 139, 125–136 (1995).
16. Arroyo, J. *et al. Yeast* 11, 587–591 (1995).
17. Bertani, I. *et al. Yeast* 11, 1187–1194 (1995).
18. Coglievina, M. *et al. Yeast* 11, 767–774 (1995).
19. Coissac, E., Maillier, E., Robineau, S. & Netter, P. *Yeast* 12, 1555–1562 (1996).
20. Escribano, V., Eraso, P., Portillo, F. & Mazón, M.J. *Yeast* 12, 887–892 (1996).
21. Guerreiro, P. *et al. Yeast* 11, 1087–1091 (1995).
22. Guerreiro, P. *et al. Yeast* 12, 273–280 (1996).
23. Hansen, M. *et al. Yeast* 12, 1273–1277 (1996).
24. James, C.M., Indge, K.J. & Oliver, S.G. *Yeast* 11, 1413–1419 (1995).
25. Klima, R. *et al. Yeast* 12, 1033–1040 (1996).
26. Rodriguez–Belmonte, E. *et al. Yeast* 12, 145–148 (1996).
27. Skala, J., Nawrocki, A. & Goffeau, A. *Yeast* 11, 1421–1427 (1995).
28. Tizon, B. *et al. Yeast* 12, 1047–1051 (1996).
29. Vandenbol, M., Durand, P., Portetelle, D. & Hilger, F. *Yeast* 11, 1519–1523 (1995).
30. Van der Aart, Q.J.M., Kleine, K. & Steensma, H.Y. *Yeast* 12, 385–390 (1996).

Acknowledgements. This study is part of the third and final phase of the European Yeast Genome Sequencing Project carried out under the administrative coordination of A. Vassarotti and the Université Catholique de Louvain. We thank P. Mordant for accounting and help; F. Foury for support and advice; G. Gérard for statistical analyses; and our colleagues for help and discussion. We would also like to acknowledge the contribution of S. Kiefer who died in a car accident in 1994. This work was supported by the EU under the BIOTECH II programme and by the Services Fédéraux des Affaires Scientifiques, Techniques et Culturelles; the Pôles d'Attraction Inter-Universitaire; the Région Wallonne; the Fonds pour la Formation à la Recherche dans l'Industrie et l'Agriculture; the Belgian Federal Services for Science Policy; the Research Fund of the Katholieke Universiteit Leuven; the Schweizerisches Bundesamt für Bildung und Wissenschaft; the Fundaçao Calouste Gulbenkian; the International Centre for Genetic Engineering and Biotechnology of Trieste; the Bull HN Information Systems Italy; the Ministerio de Asuntos Exteriores from Spain; the Comisión Interministerial de Ciencia y Tecnología of Spain; and the Groupement de Recherche et d'Etude sur les Génomes.

Correspondence and requests for materials should be addressed to A.G. (e-mail: goffeau@fysa.ucl.ac.be).

The nucleotide sequence of *Saccharomyces cerevisiae* chromosome IX

C. Churcher[1], S. Bowman[1], K. Badcock[1], A. Bankier[2], D. Brown[1], T. Chillingworth[1], R. Connor[1], K. Devlin[1], S. Gentles[1], N. Hamlin[1], D. Harris[1], T. Horsnell[2], S. Hunt[1], K. Jagels[1], M. Jones[1], G. Lye[1], S. Moule[1], C. Odell[1], D. Pearson[1], M. Rajandream[1], P. Rice[1], N. Rowley[2], J. Skelton[1], V. Smith[2], S. Walsh[1], S. Whitehead[1] & B. Barrell[1]

[1]*The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK*
[2]*MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK*

Large-scale systematic sequencing has generally depended on the availability of an ordered library of large-insert bacterial or viral genomic clones for the organism under study. The generation of these large insert libraries, and the location of each clone on a genome map, is a laborious and time-consuming process. In an effort to overcome these problems, several groups have successfully demonstrated the viability of the whole-genome random 'shotgun' method in large-scale sequencing of both viruses and prokaryotes[1–5]. Here we report the sequence of S*accharomyces cerevisiae* chromosome IX, determined in part by a whole-chromosome 'shotgun', and describe the particular difficulties encountered in the random 'shotgun' sequencing of an entire eukaryotic chromosome. Analysis of this sequence shows that chromosome IX contains 221 open reading frames (ORFs), of which approximately 30% have been sequenced previously. This chromosome shows features typical of a small S*accharomyces cerevisiae* chromosome.

The sequence derived for chromosome IX is 439,886 nucleotides in length, and 71.6% codes for proteins or predicted proteins. There are 219 non-overlapping ORFs equal to or greater than 100 amino acids long, and a further two ORFs (YIL060W and YIL059C) that overlap; these are short, and both have a low codon adaptation index (CAI). Although it is unlikely that both are coding, one could not be selected above the other as more likely to encode a protein. A single Ty3-2 retrotransposon containing three ORFs is present on the left arm of chromosome IX (between bases 205,217 and 210,644), leaving 218 *S. cerevisiae*-derived ORFs encoded on this chromosome, of which 116 are on the Crick strand, and 102 (+ 3 transposon ORFs) are on the Watson strand. Of these, 66 (30.3%) have been sequenced previously. A further 68 (31.2%) have some similarity to genes in *S. cerevisiae* and other organisms for which some functional information is available. However, 74 (33.9%) of the predicted genes on this chromosome cannot be assigned even a putative function based on sequence similarity. These can be divided into two groups: those that show no similarity to current database entries (53, 24.3%), and those that are similar to predicted genes of unknown function (21, 9.6%). The remaining 10 (4.6%) are putative pseudogene ORFs.

The average length of a chromosome IX ORF is 476 codons, with an average of one ORF every 1,993 base pairs. The largest ORF on chromosome IX is YIL129C, which encodes a hypothetical protein of 2,376 amino acids. The YIL129C protein is similar to another hypothetical protein encoded on *Caenorhabditis elegans* chromosome III (EMBL database, accession numbers CEF21H11, U11279 and ORF F21H11.2) over a region of 2,009 amino acids. In total, 20 chromosome IX ORFs are longer than 1,000 codons. Short *S. cerevisiae* genes with no homology are difficult to detect[6]. On chromosome IX, five ORFs with less than 100 codons have been identified, but future analysis will probably reveal additional short coding regions. Less than 4% of the ORFs on chromosome IX are predicted to be spliced; eight ORFs contain introns. None of the tRNA genes on this chromosome are spliced.

Ten ORFs have been identified as contributing to five putative pseudogenes. These ORFs have very good homology to genes or predicted genes, but are separated from an adjacent ORF with homology to the same protein by internal stop codons or frameshifts. These areas have been sequenced on *S. cerevisiae* genomic DNA, and the frameshifts and stop codons confirmed. However, at least two pseudogene ORFs, YIL168W and YIL167W, probably constitute the single gene *SDL1*, which codes for a serine dehydratase[7]. This gene is not present elsewhere in the *S. cerevisiae* genome and is not essential in *S. cerevisiae*. A second putative pseudogene is highly similar to hexose transporter genes, and may also simply prove to be mutated in AB972 rather than being a true pseudogene. All putative pseudogenes are located near the telomeres of the chromosome.

For this and other *S. cerevisiae* chromosomes, the intergenic distance between two adjacent ORFs varies depending on their orientation with respect to each other. Of the adjacent ORFs on chromosome IX, 95 are arranged in tandem, 54 are divergent and 55 convergent. For ORFs arranged in tandem the intergenic distance averages 472 basepairs; ORFs with divergent promoters average 619 bp; and ORFs with convergent terminators average 421 bp. This is consistent with the greater information content required for transcription initiation and regulation than for transcription termination.

Chromosome IX also contains ten tRNAs, five solo delta elements, one solo sigma element, two sigma elements flanking a transposon, and a single solo tau element.

Reports describing the features of the other small yeast chromosomes suggest that they have used a variety of strategies to achieve a certain minimum length[8–10], and this seems to be the case for chromosome IX. Its

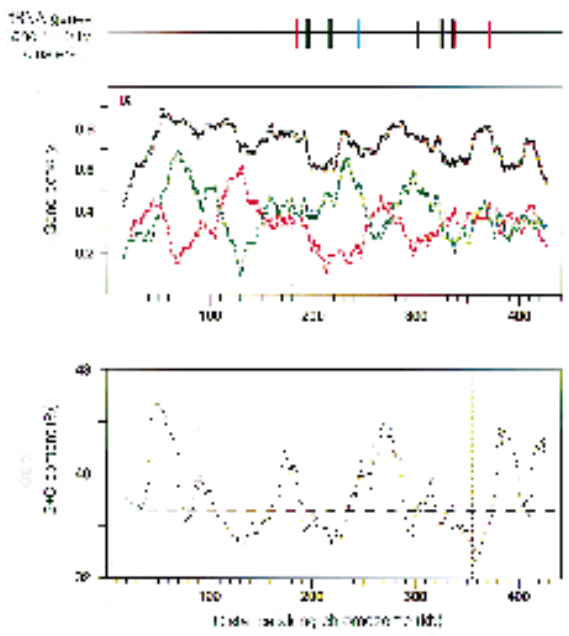

Figure 1 Overall molecular architecture of chromosome IX. The top line indicates positions of tRNA genes, solo long terminal repeat (LTR) or Ty elements (thin vertical lines) or clusters of them (thick vertical lines) along the chromosome map. The panels show variation in gene in gene density (top) and base composition (bottom) along the sequence-based map of chromosome IX (scale in kilobases from left telomere). Vertical broken lines indicate the centromere. Gene density is expressed as the probability for each nucleotide to be part of an ORF, and was calculated using sliding windows of 30 kb (steps of 0.5 kb) for the Watson strand alone (red line), the Crick strand alone (green line), and the sum of both (black line). G+C richness (%) was calculated from the silent positions of codons using a sliding window of 13 consecutive ORFs (horizontal broken line indicates average percentage G+C at silent positions of codons, 35.8%).

right telomeric region is gene poor, with only 25.3% of the sequence contributing to ORFs over approximately the last 15 kilobases of the chromosome. Chromosomes I and VI, the two smallest yeast chromosomes, also have a low coding density in their telomeric regions. Ten ORFs on chromosome IX are thought to contribute to five putative pseudogenes, all of which are located in the telomeric regions (four in the left telomeric region and one in the right). Four of these putative pseudogenes have a high degree of similarity to sequences repeated elsewhere in the yeast genome (the remaining putative pseudogene may be a mutated copy of *SDL1*, as discussed previously). Two of these occur in a 21-kb region of the chromosome IX left telomere, which is duplicated almost exactly on the chromosome X left telomere. Comparison of these two pseudogene ORFs with the equivalent regions on chromosome X shows that one region is interrupted by a frameshift on that chromosome, but the second is a single ORF. The putative pseudogene on the right arm of the chromosome is highly similar to a single ORF on chromosome XIII. The centromere[11] of chromosome IX is located towards the right end of the chromosome between bases 355,627 and 355,744.

Sequencing has revealed the definitive chromosomal position for all genes on chromosome IX. The resulting physical map correlates well with the genetic map[12] for this chromosome. Local variations in the ratio of cM to kb occur throughout this chromosome. For example, the region bounded by the genes *REV7* and *HOP1* has a significantly lower ratio of 0.3, indicating that recombination events in this area are less frequent than for the chromosome as a whole.

A cluster of genes on this chromosome occurs within its smaller, right arm between bases 399,775 and 415,615 (YIR023W to YIR032C). Six of the ten ORFs in this region are involved in the allantoin degradation pathway[13,14].

Evidence of several interchromosomal duplications occurs on chromosome IX. As well as the large region common to the left telomeres of chromosomes IX and X already described, a smaller region at the right telomere shows good homology with the telomeres of several other chromosomes. There are also several other internal chromosomal regions with long-range homology to other chromosomes. The largest of these is an area common to chromosomes IX and XIV, occurring at 89,233–186,363 and 478,568–616,076, respectively, and containing 15 homologous ORFs. Smaller interchromosomal duplications have also occurred, including a region at 230,272–258,279 repeated on chromosome V at 273,881–305,178 and a region at 46,201-69,525 repeated on chromosome XI at 617,636–639,600.

A further long-range feature previously observed for *S. cerevisiae* is a variation in the percentage G+C content along the length of its chromosomes[15]. In most cases the G+C composition for the third base of each codon has been analysed, as this is less influenced by biases in amino acid composition. For several chromosomes a periodic variation in G+C content has been observed. A plot of third-position G+C composition for chromosome IX (Fig. 1) shows that, as for other chromosomes, G+C content varies along its length. Consistent with previous analyses, the region around the centromere of the chromosome has low G+C content; the highest content occurs in peaks separated by about 100 kb.

The occurrence of first-, second- and third-codon position G+C composition was analysed for individual ORFs, as well as G+C composition in intragenic regions and the total G+C content of the chromosome (Fig. 2). This was superimposed on a chromosome map showing individual ORFs, as the usual method for plotting G+C composition uses a large window to calculate the average third-position G+C content of several adjacent ORFs, making it difficult to ascertain whether high G+C composition is a general trend in a particular region or derives from a single ORF. This method correctly predicted the peaks in third-position G+C content already described for chromosome III (data not shown).

In regions producing peaks of G+C content several adjacent reading frames show higher than average third-position G+C. Specific areas noted include 1–6, 52–67, 93–97, 175–180, 266–276, 308–313, 380–385 and 421–424 kilobases, with up to six ORFs contributing to each region. The areas of high G+C content are much more pronounced in chromosome III, with many more reading frames over a much longer distance contributing

to the two main peaks on each of the chromosome arms (results not shown).

As expected, the G+C content in intergenic regions is, on average, lower than that for coding DNA. However, it is not uniformly low, and local areas of high G+C content can be observed in some non-coding areas. The largest of these regions have been examined in detail for potential coding sequence. Although any ORFs present in these areas could conceivably be coding, they are all short, and standard methods for predicting coding sequences suggest that this is unlikely. A BLASTX search does not show any significant similarity to other *S. cerevisiae* ORFs, indicating that these areas are not remnants of ancient pseudogenes. Other possibilities have not yet been fully investigated.

Very few genes show high G+C content in position 2 of their codons. However, YIL169C, a putative glycoprotein, and YIR019C (*STA1*), a glucoamylase gene, show high second-position G+C content. Both genes code for proteins with a biased amino-acid composition. YIR019C is surrounded by regions low in G+C, but coincides with a high G+C peak on a plot of overall G+C content (results not shown), showing that a single gene can cause a peak on these plots. The role of these differences in base composition found in all yeast chromosomes has not been determined, although high G+C content might promote DNA replication or recombination.

The chromosome IX sequencing project served to highlight the difficulties involved in a whole-chromosome 'shotgun' project. This approach is described in detail in the Methods section. The purity of the starting chromosomal DNA is critical to avoid redundant sequencing effort and to minimize problems in assembly. Contamination of the chromosome IX preparation with DNA from other chromosomes resulted in a large number of single reads in the database that were not from that chromosome, but which increased the complexity of sequence assembly, slowing down both assembly and subsequent manipulation of contigs. Re-running the PFG-purified chromosome on a second pulsed-field gel has been shown to improve significantly the purity of the DNA preparation[16]. Data from the whole-chromosome 'shotgun' helped to fill the gaps in the cosmid and lambda libraries, but on their own were difficult to manage. The cosmid clones provided a framework on which to build data from the whole-chromosome shotgun. □

Methods

The chromosome IX sequencing project was initiated using a lambda clone library generated by M. Olsen and L. Riles[17]. This was an inefficient approach because of short insert sizes (20 kb) and relatively large overlaps (about 5 kb) between lambda clones. Few cosmids were available at the time, so the approach of randomly 'shotgun' sequencing the entire chromosome was attempted.

Yeast genomic DNA was electrophoresed on pulsed-field gels[18]. To increase the purity of the chromosomal preparation, yeast plugs were excised early in the run after chromosome IX had entered the gel, reducing the background DNA level. Degradation of the material of higher molecular weight occurs with prolonged running of the gel and is the likely cause of contamination of material of lower

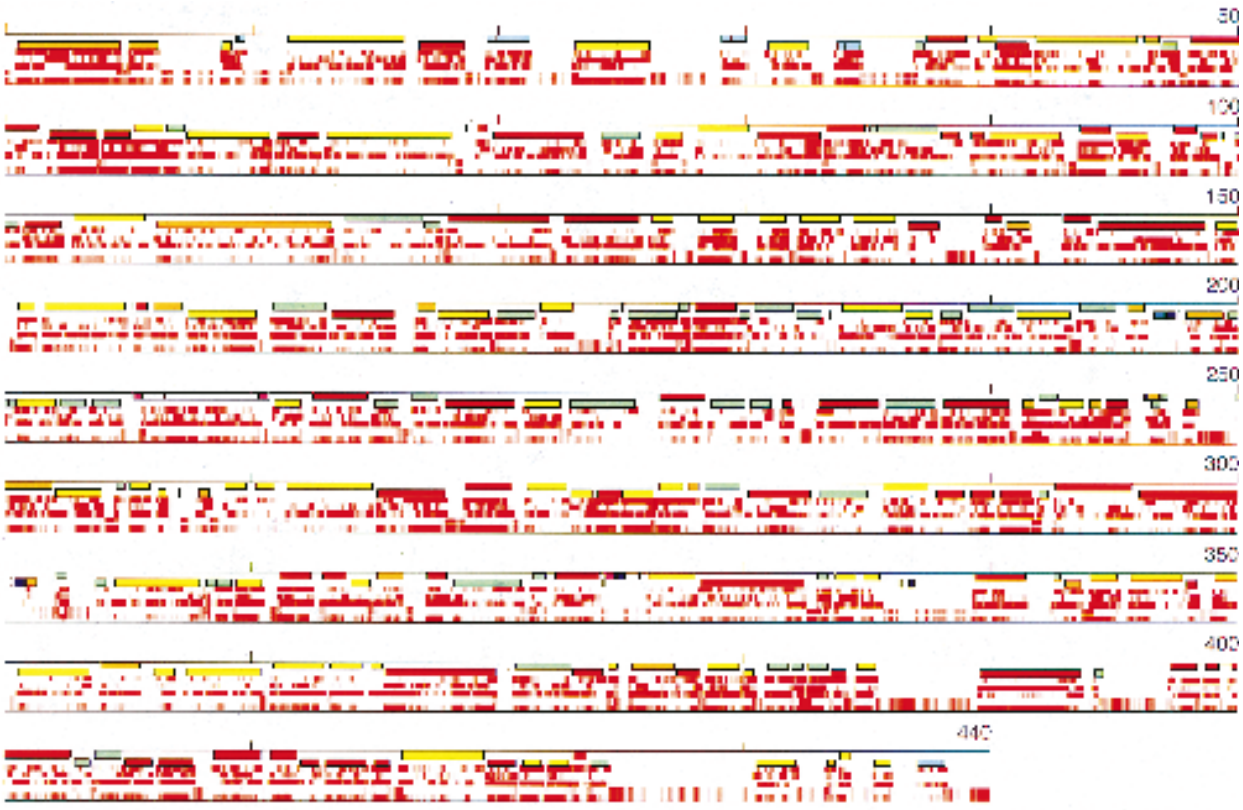

Figure 2 G+C composition of chromosome IX (drawn to scale). The top graduated line represents the chromosome split into 50-kb segments, with ORFs indicated below this as coloured boxes. ORFs located on the Watson strand are shown above those on the Crick strand. ORFs encoding previously identified genes are shown in red, those with similarities to known genes in yellow, those with similarities to hypothetical proteins in orange, and those with no significant similarities in green; pseudogene ORFs are shown in blue. tRNA genes are shown as white boxes, as are transposon-derived ORFs, with LTRs shown in dark blue (delta), turquoise (tau) and pink (sigma). Below this, variations in G+C composition (calculated using a sliding window of 200 bases) are shown as bars, with gradations of red varying from 35% to 45% (P.R., unpublished data); areas lower than 35% G+C are white, and those over 45% are red. Five bars of G+C variation are shown: the lowest bar shows total G+C content; the second shows G+C content in intergenic regions alone; and the G+C composition in each of the three bases of each codon are shown above this, with the central of the five bars representing first-position G+C, the next representing second-position G+C, and the top bar showing third-position G+C content.

molecular weight. The chromosome IX band was excised under long-wave ultraviolet transillumination to minimize DNA damage. Chromosomal DNA was purified by melting and phenol extraction, sonicated and end repaired. Two libraries were prepared: fragments 1.4–2 kb in length were cloned into M13mp18, and fragments 6–9 kb long were cloned into the phagemid vector pBS. Over 10,000 independent M13 clones were sequenced and assembled into a database using the program XBAP[19]. Sequencing strategy and methods used for sequence assembly are as described[19–25]. The lambda-clone consensus sequences were also entered into this database, which contained several thousand contigs at this stage, most of which contained a single gel read. We concluded that the chromosome IX DNA preparation was approximately 30% contaminated with DNA from other chromosomes, and that this was the source of most single-read contigs. This contamination, together with repetitive sequences in the database, caused many problems with the data assembly.

To overcome this problem, all single-read contigs were removed from the working copy of the chromosome IX database and collected in a secondary database. As further data was generated, the secondary database was periodically rescreened, and single reads were re-entered if they found matches in the primary database. This reduced the number of reads in the primary database to approximately 7,000, which represented coverage of the chromosome five times over. The database still contained several hundred contigs. At this stage a cosmid library covering most of chromosome IX became available[17]. The chromosomal 'shotgun' data were 'seeded' with reads from cosmid clones selected to give coverage over regions not previously sequenced by lambda clones. This approach also allowed the chromosome to be split up into manageable sections to solve double-stranding and compression problems. A minimal 'shotgun' of 300–500 reads was performed on each cosmid clone. Data from these cosmids were entered into the chromosome IX 'shotgun' database to contiguate the entire chromosome, and into separate cosmid databases for ease of handling, together with overlapping reads from the whole-chromosome shotgun. Each cosmid-sized project was contiguated, double stranded and all compressions were resolved.

Three regions of the chromosome remained unrepresented in either cosmid or lambda libraries: the left and right telomeres, and a region near the centre of the chromosome flanked by lambda clones 6569 and 3299. The right telomere was sequenced by primer walking using a plasmid clone[26]. The left telomere was finished using data from the whole-chromosome 'shotgun' and some primer walking from polymerase chain reaction (PCR) products generated from PFGE-purified chromosome IX DNA. The gap near the centre of the chromosome was filled using data from the whole-chromosome 'shotgun' and by sequencing a 1 kb fragment generated by PCR from genomic DNA. The gap between the lambda clones 6569 and 3299 was approximately 7 kb.

Received 24 July 1996; accepted 16 January 1997.

1. Davison, A. J. *DNA Sequence* 1, 389-394 (1991).
2. Telford, E.A.R. *et al. Virology* 189, 304-316 (1992).
3. Rawlinson, W.D. *et al. J. Virol* 70, 8833–8849 (1996).
4. Fleischmann, R. D. *et al. Science* 269, 496–512 (1995).
5. Fraser, C. M. *et al. Science* 270, 397–403 (1995).
6. Termier, M. & Kalogeropoulos, A. *Yeast* 12, 369–384 (1996).
7. Seufert, W. *Nucleic Acids Res.* 18, 3653 (1990).
8. Oliver, S. G. *et al. Nature* 357, 38–46 (1992).
9. Bussey, H. *et al. Proc. Natl Acad. Sci. USA* 92, 3809–3813 (1995).
10. Murakami, Y. *et al. Nature Genet.* 10, 261–268 (1995).
11. Fitzgerald-Hayes, M. *Yeast* 3, 187–200 (1987).
12. Mortimer, R. K. *et al.* http://genome.www.stanford.edu/saccdb/edition12.html (1995).
13. Cooper, T. G., Gorski , M. & Turoscy, V. *Genetics* 92, 383–396 (1979).
14. Turoscy, V., Chisholm, G. & Cooper, T. G. *Genetics* 108, 827–831 (1984).
15. Sharp, P. M. & Lloyd, A. T. *Nucleic Acids Res.* 21, 179–183 (1993).
16. Vaudin, M. *et al. Nucleic Acids Res.* 23, 670–674 (1995).
17. Riles, L. *et al. Genetics* 134, 81–150 (1993).
18. Schwartz, D. C. & Cantor, C. R. *Cell* 37, 67–75 (1984).
19. Dear, S. & Staden, R. *Nucleic Acids Res.* 19, 3907–3911 (1991).
20. Craxton, M. *Methods: A Companion to Methods in Enzymology* Vol. 3 (ed. Roe, B.) 20–26 (Academic, San Diego, 1991).
21. Halloran, N., Du, Z. & Wilson, R. K. in *Methods in Molecular Biology* Vol. 10, *DNA Sequencing: Laboratory Protocols* (eds Griffin, H. G. & Griffin, A. M.) 297–316 (Humana, Clifton, NJ, 1992).
22. Smith, V. *et al. Methods Enzymol.* 218, 173–187 (1993).
23. Hawkins, T. *DNA Sequence* 3, 65–69 (1992).
24. Staden, R. *Methods Mol. Biol.* 25, 27–36 (1994).
25. Staden, R. *Methods Mol. Biol.* 25, 37–67 (1994).
26. Louis, E. *Biochemica* 3, 25–26 (1995).

Acknowledgements. We thank A. Fraser for cosmid DNA preparation; the subcloning group for lambda and cosmid library preparation; the staff in the gel pouring and media kitchens for their help; the staff in the computer support and software development groups and R. Staden for software support; L. Riles and M. Olsen for lambda clones and cosmids; E. Louis for telomeric plasmids; B. Dujon for providing Fig. 1, which was prepared using unpublished software developed by C. Marck in collaboration with B. Dujon; and our colleagues for proof reading and criticism of this manuscript. This work was funded by the Medical Research Council and the Wellcome Trust.

Correspondence and requests for materials should be addressed to B.B. (e-mail: barrell@sanger.ac.uk). Clone accession numbers and other information can be found on http://www.sanger.ac.uk/yeast/home.html.

The nucleotide sequence of *Saccharomyces cerevisiae* chromosome XII

M. Johnston[1], L. Hillier[1], L. Riles[1], other members of the Genome Sequencing Center[1], K. Albermann[2], B. André[3], W. Ansorge[4], V. Benes[4], M. Brückner[5], H. Delius[6], E. Dubois[7], A. Düsterhöft[8], K.-D. Entian[9], M. Floeth[8], A. Goffeau[10], U. Hebling[6], K. Heumann[2], D. Heuss-Neitzel[8], H. Hilbert[8], F. Hilger[11], K. Kleine[2], P. Kötter.[9], E. J. Louis[12], F. Messenguy[13], H. W. Mewes[2], T. Miosga[14], D. Möstl[8], S. Müller-Auer[2], U. Nentwich[4], B. Obermaier[15], E. Piravandi[15], T. M. Pohl[16], D. Portetelle[11], B. Purnelle[10], S. Rechmann[4], M. Rieger[5], M. Rinke[15], M. Rose[9], M. Scharfe[17], B. Scherens[18], P. Scholler[19], C. Schwager[4], S. Schwarz[19], A. P. Underwood[12], L. A. Urrestarazu[3], M. Vandenbol[11], P. Verhasselt[20], F. Vierendeels[13], M Voet[20], G. Volckaert[20], H. Voss[4], R. Wambutt[17], E. Wedler[17], H. Wedler[17], F. K. Zimmermann[14], A. Zollner[2], J. Hani[2] & J. D. Hoheisel[20]

[1] The Genome Sequencing Center, Department of Genetics, Washington University School of Medicine, 630 S. Euclid Avenue, St. Louis, Missouri 63110, USA
[2] Martinsrieder Institut für Protein Sequenzen, Max-Planck-Institut für Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, Germany
[3] Laboratoire de Physiologie Cellulaire et de Génétique des Levures, Université Libre de Bruxelles - Campus Plaine CP 244 Boulevard du Triomphe, B-1050 Bruxelles, Belgium
[4] EMBL, Meyerhofstrasse 1, D-69117 Heidelberg, Germany
[5] Genotype, Klingenstrasse 35, D-69434 Hirschhorn and Angelhofweg 39, D-69259 Wilhelmsfeld, Germany
[6] DNA Sequencing, Deutsches Krebsforschungszentrum, Im Neuenheimer Feld 506, D-69120 Heidelberg, Germany
[7] Laboratoire de Microbiologie de l' Université Libre de Bruxelles, B-1070 Brussels, Belgium
[8] QIAGEN GmbH, Max-Volmer-Strasse 4, D-40724 Hilden, Germany
[9] Johann Wolfgang Goethe-Universität Frankfurt, Institut für Mikrobiologie, Marie-Curie-Strasse 9; Geb. N250, D-60439 Frankfurt/Main, Germany
[10] Unité de Biochimie Physiologique, Faculté des Sciences Agronomiques, Université Catholique de Louvain, Place Croix du Sud, 2-20, B-1348 Louvain-la-Neuve, Belgium.
[11] Faculté Universitaire des Sciences Agronomiques, Unité de Microbiologie, 6, avenue Maréchal Juin, 5030 Gembloux, Belgium
[12] Department of Yeast Genetics, Institute of Molecular Medicine, John Radcliffe Hospital, Oxford OX3 9DU, UK
[13] Research Institute of the CERIA-COOVI, B-1070 Brussels, Belgium
[14] Institut für Mikrobiologie und Genetik, Schnittspahnstrasse 10, D-64287 Darmstadt, Germany
[15] MediGene GmbH, Lochhamer Strasse 11, 82152 Martinsried, Germany
[16] GATC Gesellschaft für Analyse-Technik und Consulting mbH, Fritz-Arnold-Strasse 23, 78467 Konstanz, Germany
[17] AGON GmbH, Glienicker Weg 185, D-12489 Berlin, Germany
[18] Laboratorium voor Erfelijkheidsleer en Microbiologie van de Vrije Universiteit Brussel Vlaams Interuniversitair Instituut voor Biothechnologie, Departement Microbiologie, Avenue E. Gryson 1, B-1070 Brussels, Belgium
[19] Katholieke Universiteit Leuven, Laboratory of Gene Technology, Willem de Croylaan 42, B-3001 Leuven, Belgium
[20] Molecular Genetic Genome Analysis, Deutsches Krebsforschungszentrum, Im Neuenheimer Feld 506, D-69120 Heidelberg, Germany