the Xp- females was 185.8 months (s.d., 74.9). All subjects were healthy, with no significant neurological disease. Females with an Xp- chromosome were all referred for investigation because of short stature in middle childhood, with one exception who was karyotyped at birth. Neuropsychological test results are presented for subjects with verbal IQs ≥ 65 (three 45,X^m subjects and one 45,X^p subject had verbal IQs that fell out of range). Parents rated 70 normal males and 71 normal females (age range, 6–18 years) on the social-cognition scale. The neuropsychological test battery was used to assess 68 normal males and 91 normal females (age range, 6–25 years). Verbal IQs were in the range 65–151. All normal comparison subjects were recruited from urban and suburban schools (6–18 years) and from hospital staff (18–25 years).

Behavioural and cognitive measures. Initial screening was conducted by postal questionnaires using a well-standardized set of instruments[14–16]. These were completed by parents, teachers and the Turner-syndrome subjects themselves (11 years and over). The social cognition questionnaire (Box 1) was completed by parents only. In a survey of 175 Turner-syndrome subjects for whom we obtained parental ratings on two occasions, a mean of 2.7 years apart, the intraclass correlation coefficient was 0.81 ($P < 0.01$). Scores correlate with the self-rated social problem subscale of the YSR[16] 0.58 ($P < 0.002$), with the teacher rating on the TRF[15] 0.54 ($P < 0.001$), and with the parent-rated CBCL[14] 0.69 ($P < 0.001$). The range of scores was 0–23 in the Turner-syndrome sample and 0–21 in the normal sample (maximum score of 24). The CBCL[14] was completed by 70 parents, the YSR[16] was completed by 40 subjects over 11 years of age, and the TRF[15] was completed by 45 teachers. Clinical significance of social problems was estimated according to clinical *T* scores[14–16]. Measures of cognition included the Wechsler Intelligence Scales for Children (WISC III-UK)[24] and the Wechsler Adult Intelligence Scales–Revised (WAIS-R)[25]. The behavioural inhibition task was the Same–Opposite World subtest from the Test of Everyday Attention for Children[26]. This yields a time measure that ascertains the difference in latency for a subject responding to a series of stimuli on a task of sequential responses, which are named both as they appear and then opposite to their appearance. The subject reads a random series of numbers (1 and 2) saying 'one' to 1, and 'two' to 2. The subjects then repeat the task on a new series, but this time they have to inhibit the prepotent response and instead say 'two' to 1, and 'one' to 2, correcting any errors before proceeding. Test–retest reliability on a sample of 70 normal children gave an intraclass correlation coefficient of 0.62 ($P < 0.001$). The Tower of Hanoi task was based on the procedure described previously[27]. It was scored according to the most complex level of the problem the child could solve reliably. Test–retest reliability gave an intraclass correlation coefficient for the highest level achieved of 0.45 ($P < 0.001$), which is in line with expectations for a test that makes novel demands of this nature[28].

Received 17 February; accepted 1 May 1997.

1. Jacobs, P. A. *et al.* A cytogenetic and molecular reappraisal of a series of patients with Turner's syndrome. *Ann. Hum. Genet.* **54,** 209–223 (1990).
2. Pennington, B. F. *et al.* The neuropsychological phenotype in Turner syndrome. *Cortex* **21,** 391–404 (1985).
3. McCauley, E., Ito, J. & Kay, T. Psychosocial functioning in girls with the Turner syndrome and short stature. *J. Am. Acad. Child Psychiat.* **25,** 105–112 (1986).
4. Damasio, A. R. On some functions of the human prefrontal cortex. *Proc. N. Y. Acad. Sci.* **769,** 241–251 (1995).
5. Barlow, D. P. Gametic imprinting in mammals. *Science* **270,** 1610–1613 (1995).
6. Ballabio, A. & Andria, G. Deletions and translocations involving the distal short arm of the human X chromosome: review and hypotheses. *Hum. Mol. Genet.* **1,** 221–227 (1995).
7. Lyon, M. F. Gene action in the X-chromosome of the mouse (*Mus musculus* L). *Nature* **190,** 372–373 (1961).
8. Bailey, A., Philips, W. & Rutter, M. Autism: towards an integration of clinical, genetic, neuropsychological and neurobiological perspectives. *J. Child Psychol. Psychiat.* **37,** 89–126 (1996).
9. Ledbetter, D. H. & Engel, E. Uniparental disomy in humans: development of an imprinting map and its implications for prenatal diagnosis. *Hum. Mol. Genet.* **4,** 1757–1764 (1995).
10. Zuccotti, M. & Monk, M. Methylation of the mouse *Xist* gene in sperm and eggs correlates with imprinted *Xist* expression and paternal X-inactivation. *Nature Genet.* **9,** 316–320 (1995).
11. McCauley, E., Kay, T., Ito, J. & Trader, R. The Turner syndrome: cognitive deficits, affective discrimination and behaviour problems. *Child Dev.* **58,** 464–473 (1987).
12. Saenger, P. Clinical Review 48: The current status of diagnosis and therapeutic intervention in Turner's syndrome. *J. Clin. Endocrinol. Metabol.* **77,** 297–301 (1993).
13. Skuse, D., Percy, E. L. & Stevenson, J. in *Growth, Stature, and Adaptation. Behavioral, Social, and Cognitive Aspects of Growth Delay* (eds Stabler, B. & Underwood, L.) 151–164 (UCP, Chapel Hill, 1994).
14. Achenbach, T. M. *Manual for the Child Behavior Checklist/4-18 and 1991 Profile* (Department of Psychiatry, University of Vermont, Burlington, VT, 1991).
15. Achenbach, T. M. *Manual for the Teacher's Report Form and 1991 Profile* (Department of Psychiatry, University of Vermont, Burlington, VT, 1991).
16. Achenbach, T. M. *Manual for the Youth Self-Report Form and 1991 Profile* (Department of Psychiatry, University of Vermont, Burlington, VT, 1991).
17. Eagley, A. H. The science and politics of comparing men and women. *Am. Psychol.* **50,** 145–158 (1995).
18. World Health Organization *The ICD-10 Classification of Mental and Behavioural Disorders: Clinical Descriptions and Diagnostic Guidelines* (World Health Organizaiton, Geneva, 1992).
19. Pennington, B. F. & Ozonoff, S. Executive functions and developmental psychopathology. *J. Child Psychol. Psychiat.* **37,** 51–87 (1996).
20. Bjorklund, D. F. & Kipp, K. Parental investment theory and gender differences in the evolution of inhibition mechanisms. *Psychol. Bull.* **120,** 163–188 (1996).
21. Disteche, C. M. Escape from X inactivation in human and mouse. *Trends Genet.* **11,** 17–22 (1995).
22. Wolff, D. J., Miller, A. P., Van Dyke, D. L., Schwartz, S. & Willard, H. F. Molecular definition of breakpoints associated with human Xq isochromosomes: implications for mechanism of formation. *Am. J. Hum. Genet.* **58,** 154–160 (1996).
23. Hassold, T., Pettay, D., Robinson, A. & Uchida, I. Molecular studies of parental origin and mosaicism in 45,X conceptuses. *Hum. Genet.* **89,** 647–652 (1992).
24. Wechsler, D. *Wechsler Intelligence Scale for Children* 3rd UK edn (Psychological Corporation, London, 1992).
25. Wechsler, D. *Wechsler Adult Intelligence Scales-Revised* (Psychological Corporation, New York, 1986).
26. Borys, S. V., Spitz, H. H. & Dorans, B. A. Tower of Hanoi performance of retarded young adults and nonretarded children as a function of solution length and goal state. *J. Exp. Child Psychol.* **33,** 87–110 (1982).
27. Manly, T., Robertson, I. H. & Anderson, V. *The Test of Everyday Attention for Children (TEACh)* (Thames Valley Test Company, Bury St Edmunds, in the press).
28. Rabbitt, P. M. A. in *Methodologies of Frontal and Executive Function* (ed. Rabbitt, P. M. A.) (Psychology Press, Hove, in the press).
29. Temple, C. M. & Carney, R. A. Patterns of spatial functioning in Turner's syndrome. *Cortex* **31,** 109–118 (1995).

Acknowledgements. We thank E. Percy, S. Cave, A. O'Herlihy, R. South, J. Smith, M. Power and D. Robinson for assistance; M. Pembrey for comments and discussion; many paediatric consultants for assisting with the recruitment of patients, the schools who participated, and all of the subjects of our investigation and their families for their time. This research was supported by the Wellcome Trust and the Child Growth Foundation. Compilation of the national register of Turner syndrome was supported by the British Society for Paediatric Endocrinology and by Pharmacia.

Correspondence and requests for material should be addressed to D.H.S. (e-mail: dskuse@ich.ucl.ac.uk).

Molecular evidence for an ancient duplication of the entire yeast genome

Kenneth H. Wolfe & Denis C. Shields

Department of Genetics, University of Dublin, Trinity College, Dublin 2, Ireland

Gene duplication is an important source of evolutionary novelty[1,2]. Most duplications are of just a single gene, but Ohno[1] proposed that whole-genome duplication (polyploidy) is an important evolutionary mechanism. Many duplicate genes have been found in *Saccharomyces cerevisiae*, and these often seem to be phenotypically redundant[3–7]. Here we show that the arrangement of duplicated genes in the *S. cerevisiae* genome is consistent with Ohno's hypothesis. We propose a model in which this species is a degenerate tetraploid resulting from a whole-genome duplication that occurred after the divergence of *Saccharomyces* from *Kluyveromyces*. Only a small fraction of the genes were subsequently retained in duplicate (most were deleted), and gene order was rearranged by many reciprocal translocations between chromosomes. Protein pairs derived from this duplication event make up 13% of all yeast proteins, and include pairs of transcription factors, protein kinases, myosins, cyclins and pheromones. Tetraploidy may have facilitated the evolution of anaerobic fermentation in *Saccharomyces*.

We searched systematically for duplicated regions[6,7] in the complete yeast genome[8] by using BLASTP[9] amino-acid sequence similarity searches of all yeast proteins against one another, and plotted the results on dot matrices. Duplicate regions are visible as a diagonal series of 'hits' with conserved gene orientation. In the example shown in Fig. 1, three separate diagonals indicate three distinct regional duplications between chromosomes X and XI. Within each region the homologues are interspersed with genes that are not now duplicated. We propose that this is the result of random deletion of individual duplicated genes from one or other chromosome subsequent to the initial duplication of the whole region.

In the whole genome, 55 duplicate regions were identified

containing 376 pairs of homologous genes (Fig. 2). Amino-acid sequence identity between the pairs ranges from 24% to 100%, with a mean of 63%. The criteria used to define a duplicate region were: (1) BLASTP high scores of ⩾200 for each gene pair (these have an associated significance of $P = 10^{-18}$ or less); (2) at least three pairs of homologues with intergenic distances of ⩽50 kilobases (kb) on each chromosome; and (3) conservation of gene order and orientation (with allowance for small inversions within some blocks). The extent of duplicated regions is significantly in excess of what would be expected by chance if homologous genes were distributed at random, as measured by two statistical tests (see Methods). The duplicated regions are on average 55 kb long and contain a mean of 6.9 duplicate gene pairs. Together they span 50% of the genome. This is a minimal estimate because we did not consider possible duplicated regions that contain just one or two gene pairs. In the 55 duplicate regions, 25% of the genes (743 of 2,905) are duplicated, and again this is a minimal estimate because some less well conserved duplicate genes (BLASTP < 200) may be present within the blocks. Many regions also contain duplicate tRNA genes at equivalent locations. The 376 pairs of duplicate proteins account for only 12% of all the yeast sequence pairs having a BLASTP score above 200 (after excluding yeast retrotransposon (Ty) element sequences); most of the others are hits among members of large families such as sugar permeases, protein kinases, the AAA superfamily and the proteins encoded by subtelomeric repeats. However, these 376 pairs located in duplicate regions account for 42% of all 'simple' duplicate gene pairs in yeast (that is, pairs of genes that are one another's only significant BLASTP hit).

How did these 55 duplicated regions arise? They were formed either by many successive independent duplications (each involving dozens of kilobases), or simultaneously by a single duplication of the entire genome (tetraploidy) followed by reciprocal translocations between chromosomes to produce a mosaic of duplicated blocks. A polyploid origin for the yeast genome was first proposed in 1987 by Smith[5].

We favour the tetraploidy and translocation model for two reasons. First, for 50 of the 55 duplicate regions, the orientation of the entire block with respect to the centromere is the same in the two copies (Fig. 2), which is significantly non-random ($\chi_1^2 = 18$). Block orientations are expected to be conserved if the blocks were formed by reciprocal translocations among duplicate chromosomes, whereas if each block was made by an independent duplication its orientation should be random. Moreover, even the gene pairs outside the identifiable duplicated blocks (Fig. 2) show a bias towards conservation of transcriptional orientation with respect to the centromere. For 'simple' gene pairs (defined above) in which both genes lie outside the 55 blocks, most (117 out of 172 pairs) show conserved transcriptional orientation, whereas no bias of orientation is seen in pairs in which one of the genes lies in a block but the other does not. This suggests that our method may have overlooked numerous smaller duplicated blocks. For example, the large unassigned region on the right arm of chromosome X (Fig. 2) becomes paired with unassigned parts of chromosomes XVI, XIII, IV and VIII if the criterion for defining a block is reduced from three to two pairs of duplicated genes.

Second, based on a Poisson distribution of block sizes, 55 successive duplications would be expected to result in about seven triplicated regions (that is, duplicates of duplicates), but we observe none (or at most one; see Methods). This difference is highly significant. Together, these two statistical tests support a model of tetraploidy in which the regional duplications are relics of a whole-genome duplication.

The mosaic pattern of duplicated segments leads us to propose that *S. cerevisiae* is an ancient tetraploid, similar to maize[10] and perhaps vertebrates[11]. Our model for this is that two ancestral diploid yeast cells, each containing about 5,000 genes, fused to form a tetraploid. Depending on the species relationship of the two cells, this could have been either autotetraploidy or allotetraploidy. This species then became diploid (underwent a decay of sequence identity), and most (about 85%) of the duplicate copies were deleted, leaving the current species with a haploid/diploid life cycle and 5,800 genes that include many duplicates. The original chromosome-sized duplications were then broken up into smaller blocks by reciprocal translocations. About half of the 800 duplicate gene pairs are shown in Fig. 2, and the remainder are presumed to lie in regions that have been fragmented too severely by translocation, deletion or transposition to be detectable by the methods used here. Autotetraploidy sometimes occurs in yeast by fusion between spontaneously arising diploids that are homozygous at the *MAT* locus[12], and recent allotetraploidy has been proposed for *S. carlsbergensis*[13]. Reciprocal translocation, which we propose to be the mechanism causing fragmentation of the duplicated blocks, has been observed in comparisons of chromosomes between *S. cerevisiae* and *S. bayanus*[14]. The arrangement of some sets of blocks, such as the juxtaposition of both copies of blocks 50 and 14 beside both copies of blocks 23 and 37 (Fig. 2), also implicates reciprocal translocation. This model of yeast genome evolution could be tested by comparison of gene order in related genera such as *Kluyveromyces* and *Candida*. A similar model in which most, but not all, of the original chromosomes were duplicated (aneuploidy) seems less likely but cannot be ruled out.

The closest relatives of yeast for which a substantial sequence data set is available are species in the genus *Kluyveromyces*[15]. Comparison of the gene sequences and gene order of *S. cerevisiae* and

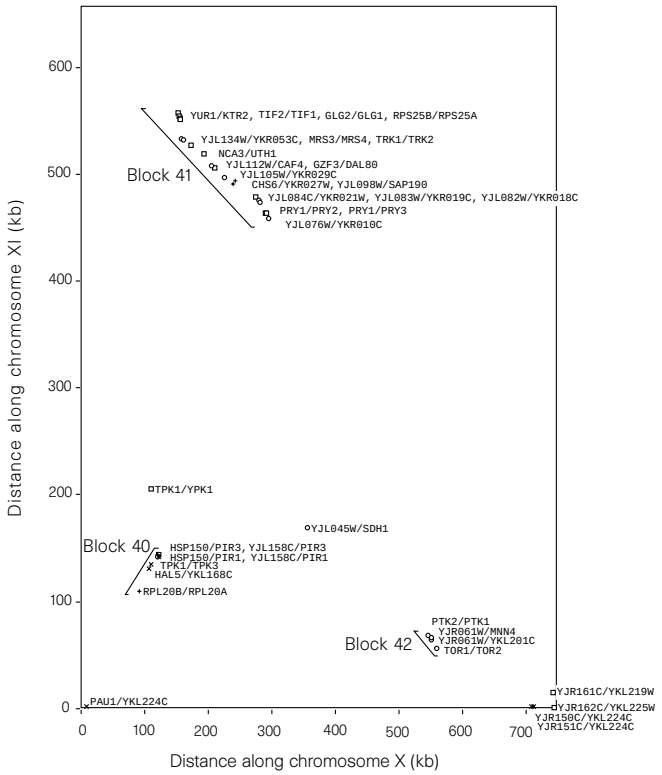
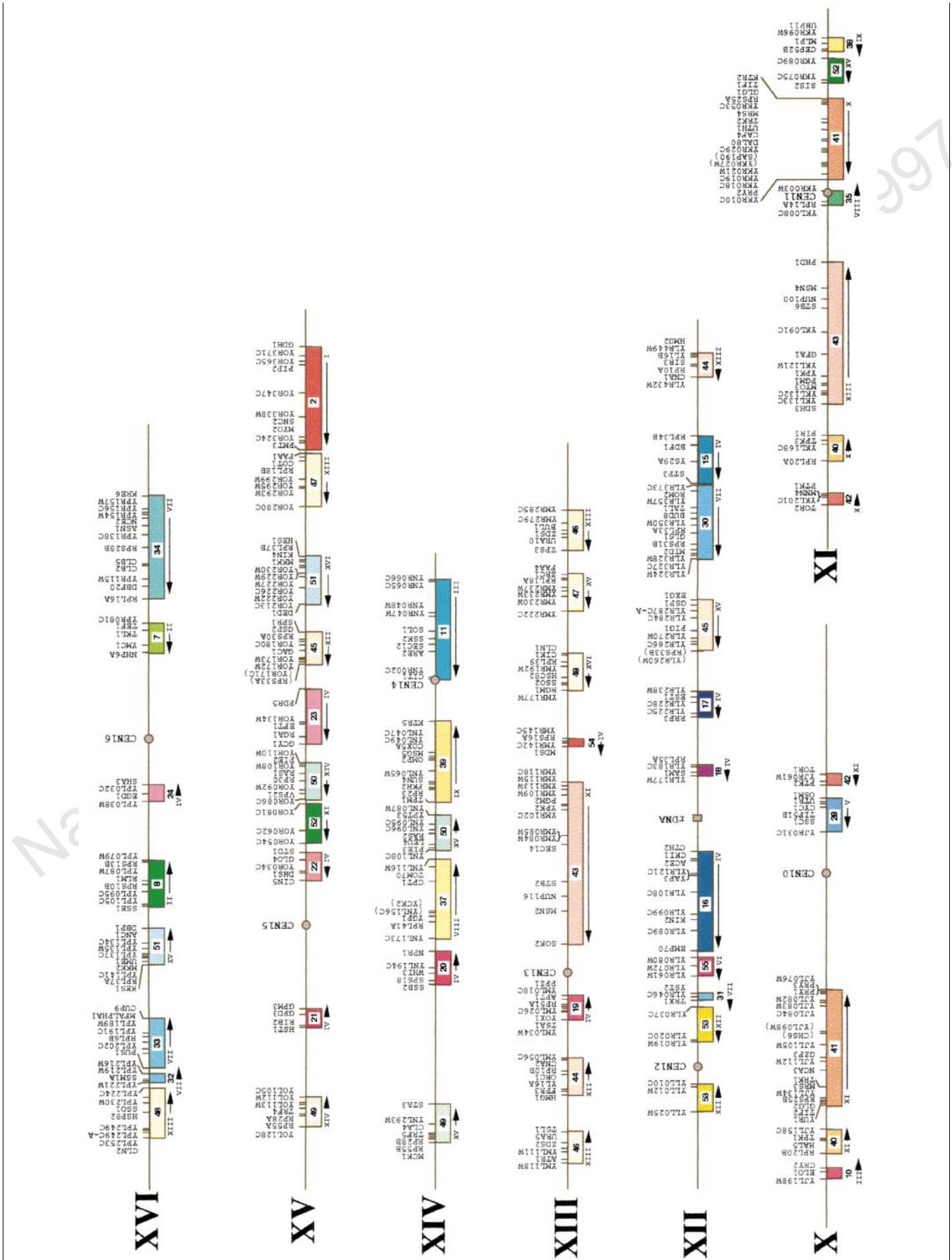


Figure 1 Locations of similar genes on chromosomes X and XI. All pairs of proteins with BLASTP scores ⩾200 are plotted at the position of their genes on the two chromosomes. Ty elements have been omitted. Symbols indicate gene orientations: +, W (Watson strand; left-to-right transcription) on both chromosomes; ×, C (Crick strand; right-to-left transcription) on both; squares, C on chromosome X but W on chromosome XI; circles, W on chromosome X but C on chromosome XI. Three diagonals, corresponding to duplicated regions shown in Fig. 2, are marked.



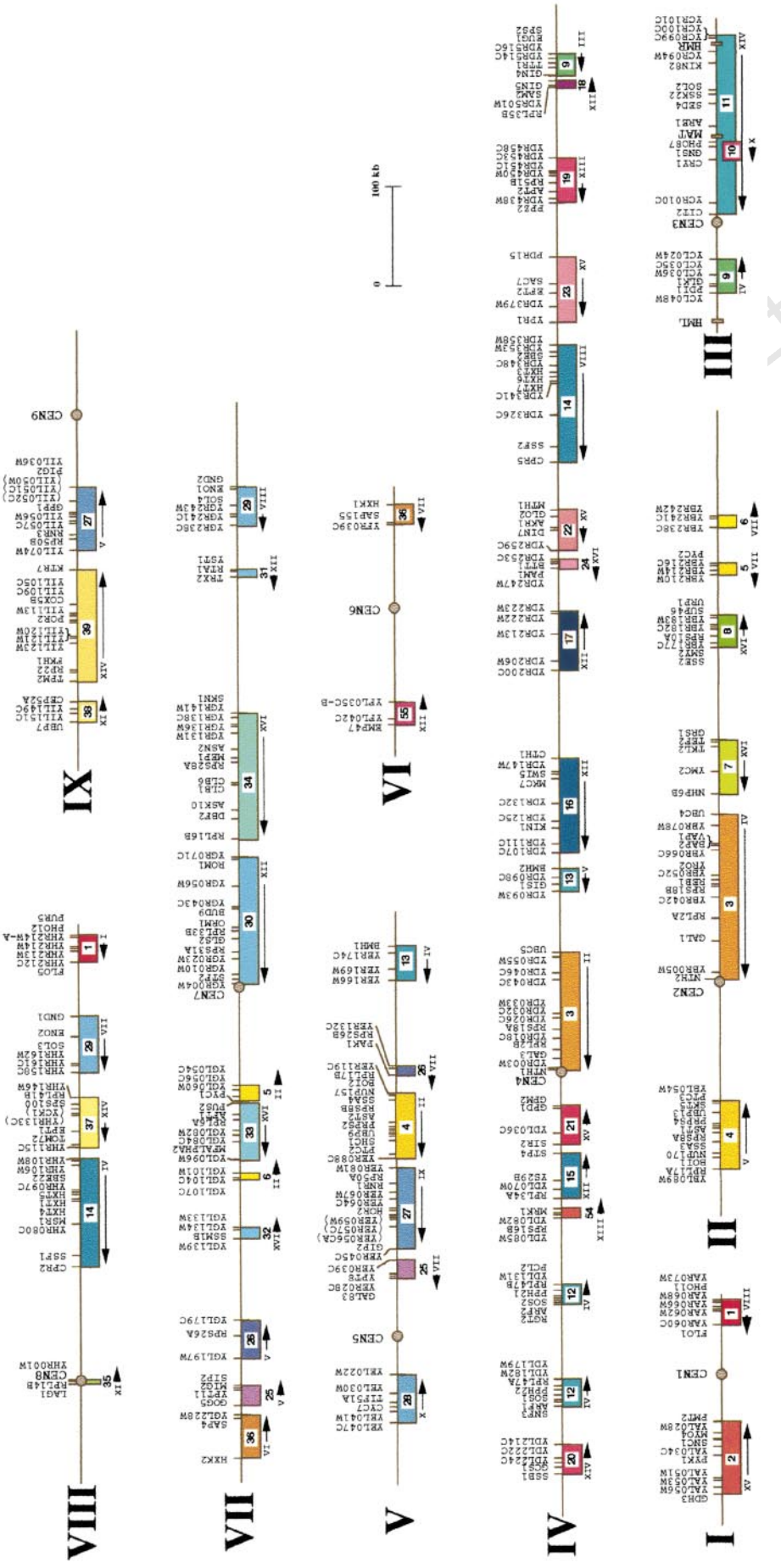
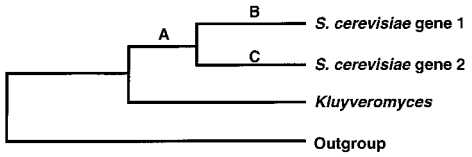


Figure 2 Locations of 55 duplicated chromosomal regions. Only duplicated genes are named (YPD database names). Regions are numbered and coloured arbitrarily. Arrows show the relative orientations of the two copies of a block; orientations are conserved with respect to the centromere for all except blocks 6, 8, 17, 18 and 33. Roman numerals below blocks indicate the chromosomal location of the other copy. Parentheses indicate genes with an order and orientation that is not as expected (that is, probable cases of postduplication inversions). Braces group two ORFs in one copy of a block that correspond to one ORF in the other. CEN indicates the position of the centromere on that chromosome. Details of some additional putative gene pairs that are located within or beside the 55 regions but have lower BLASTP scores can be obtained from http://acer.gen.tcd.ie/~khwolfe/yeast.



Block	Gene 1 / Gene 2	Bootstrap support (%)	Age	90% Confidence interval	Amino-acid sites	Branch length A	B	C	Outgroup
3	*NTH2 / NTH1*	100	0.70	(0.59 - 0.82)	714	0.052	0.131	0.115	*Candida albicans*
3	*GAL1 / GAL3*	99	0.55	(0.41 - 0.76)	468	0.100	0.115	0.130	*Homo sapiens*
37	*YCK1 / YCK2*	97	0.48		343	0.036	0.046	0.024	*Sch. pombe*
36	*HXK1 / HXK2*	94	0.77	(0.62 - 0.95)	484	0.037	0.147	0.107	*Sch. pombe*
45	*EXG1 / SPR1*	90	0.80	(0.67 - 0.97)	418	0.051	0.166	0.270	*Candida albicans*
10	*CRY1 / CRY2*	89	0.40		137	0.015	0.014	0.007	*Homo sapiens*
21	*SIR2 / HST1*	45	0.86	(0.66 - 1.00)	402	0.028	0.167	0.182	*Caenorhabditis elegans*
28	*CYC7 / CYC1*	34	0.82		106	0.017	0.110	0.057	*Sch. pombe*
30	*YGR043C / TAL1*	24	1.00		332	0.000	0.231	0.149	*Homo sapiens*
Genes with apparent gene conversion									
8	*RPS10A / RPS10B*	100	0.04		233	0.042	0.000	0.004	*Sch. pombe*
20	*SSB1 / SSB2*	100	0.06	(0.01 - 0.11)	613	0.049	0.002	0.004	*Candida albicans*
37	*RPL41B / RPL41A*	100	0.00		105	0.054	0.000	0.000	*Candida tropicalis*

Figure 3 Phylogenetic analysis of *S. cerevisiae* gene pairs and their *K. lactis* or *K. marxianus* homologues. Bootstrap support for the indicated branching order, and estimated age of the *S. cerevisiae* duplication relative to the *Saccharomyces/Kluyveromyces* speciation, are shown. *Sch.*, *Schizosaccharomyces*.

Kluyveromyces suggests that these lineages diverged before the proposed genome duplication in *S. cerevisiae* (Fig. 3). This implies that the *Kluyveromyces* homologues of all of the genes named in Fig. 2 should be single-copy, and we did not find any examples to contradict this in the literature. *K. lactis* has a genome of 12 million base pairs, slightly smaller than that of *S. cerevisiae*, but has only six chromosomes. Two of its centromeres[16] are orthologous to pairs of *S. cerevisiae* centromeres that are located in duplicated blocks 11 and 35.

Phylogenetic trees were drawn from protein sequences of 12 *S. cerevisiae* duplicate pairs with homologues in *Kluyveromyces* and an outgroup (Fig. 3). In nine of these there is strong bootstrap support ($\geq$89%) for a branching order that places the two *S. cerevisiae* sequences together; in the others there is no strong support for any order. We estimated the ages of the duplications in *S. cerevisiae* relative to the speciation between the two yeast species. Three gene pairs (*SSB1/SSB2* and two ribosomal proteins) yield very young ages, probably because they have been involved in recent gene conversions within *S. cerevisiae* (Fig. 3). Of the five sequences for which there are sufficient data to calculate a confidence interval by bootstrapping, and excluding *SSB1/SSB2*, the mean relative age of duplication is 0.74 (s.d. 0.12). The date of the divergence between *Saccharomyces* and *Kluyveromyces* is not known with any certainty, but assuming a constant molecular clock for 18S ribosomal RNA[15] and an animal/fungal divergence time of 10^9 years, we estimate it to be roughly 1.5×10^8 years ago, which places the genome duplication roughly 10^8 years ago.

Our model of massive gene deletion in the wake of genome duplication predicts that some groups of genes that are adjacent in *Kluyveromyces* should have homologues in *S. cerevisiae* that are not themselves duplicated, but are located within different copies of duplicated blocks. This is the case with the genes *GAL4* (*LAC9*) and *SGS1*. In *S. cerevisiae* these genes are single-copy and are located on chromosomes XVI and XIII, respectively, in equivalent intervals within block 48 (between the pairs *HSP82/HSC82* and *YPL249C/YMR192W*; Fig. 2). In *K. lactis GAL4* (*LAC9*) and *SGS1* are neighbours (R. S. Keogh and K.H.W., unpublished data). This suggests that *GAL4* and *SGS1* were originally adjacent in *Saccharomyces*, but after the genome duplication one copy of each gene was deleted. Similar relationships are seen between the *K. lactis* gene cluster *HHT1–TRP1–IPP1* and block 3; between *K. lactis* *YNL217W–RAP1–GYP7* and block 20; and between *K. marxianus* *RPL25–YNL305C* and block 49 (refs 17–19). In each of these cases the transcriptional orientation of all genes has been conserved between the two species. These results could be explained either by our model of degenerate tetraploidy or by multiple independent regional duplications, but in the latter case all four duplications must have occurred in *Saccharomyces* after the speciation.

For most of the 376 gene pairs in *S. cerevisiae*, the function of both genes is not known. Only a few pairs have functions that have clearly diverged, notably the genes encoding mitochondrial and peroxisomal isozymes of citrate synthase (*CIT1/CIT2*)[6], the *RAS1/RAS2* genes[20], the genes *ACE2/SWI5* which encode transcription factors[21], *TOR1/TOR2* which encode phosphatidylinositol kinases[22], and *MYO2/MYO4* which encode myosins[23]. The differences between other pairs seem less important, for example the myosin genes *MYO3/MYO5*, and several pairs of genes that encode cyclins (including *CLN1/CLN2*, *CLB1/CLB2* and *CLB5/CLB6*), nucleoporins (*NUP100/NUP116* and *NUP157/NUP170*), the α1 and α2 mating pheromones, and 18 pairs of protein kinases. Nevertheless, according to our model these genes have been retained in duplicate for about 10^8 years. Before the genome duplication their separate functions must either have been embodied in a single protein[2], or one of the functions did not exist (or, less likely, one of the functions was performed by a different gene that was supplanted). This implies that the physiology of the ancestral yeast may have been quite different from that seen today, and was perhaps more similar to that of *Kluyveromyces*. The most striking physiological difference between *Saccharomyces* and other yeasts is its ability to ferment sugars vigorously under anaerobic conditions, producing ethanol. The proposed genome duplication may have been instrumental in its evolutionary adaptation to anaerobic growth; for example, the duplicate genes include several pairs that are regulated differently under aerobic and anaerobic conditions (*CYC1* and *CYC7*; *COX5A* and *COX5B*), as well as several genes encoding sugar transporters. It may not be a coincidence that the estimated date of the genome duplication corresponds to the time when angiosperms (and their fruit) became abundant in the earth's flora[24]. □

Methods

Data. Yeast proteome lists from the Yeast Protein Database (YPD), the *Saccharomyces* Genome Database (SGD) and Martinsried Institute for Protein

Sequences (MIPS) databases were reconciled. Ty elements and dubious open reading frames (ORFs) were excluded. The data set (5,790 proteins) and search results can be viewed at the URL http://acer.gen.tcd.ie/~khwolfe/yeast. Repetitive regions within proteins were masked using the SEG filter in BLAST.

Statistical analysis. Chi-square tests (data not shown) indicate that duplicated genes in yeast are distributed in a highly non-random manner with regard to both the order in which homologous genes occur on pairs of chromosomes and the transcriptional orientations of those genes. A simultaneous origin of duplicate regions, as opposed to 55 independent duplications, is supported by a chi-square test on block orientations and by the lack of triplicated regions. The Poisson expectation if blocks were duplicated sequentially is for approximately 40 duplicated blocks, and 7 blocks that are replicated more than once (mainly triplicated). There is only one possible candidate for a triplicated region: the genes *YDR474C*, *YDR492W* and *GNP1* on chromosome IV and *YOR019W*, *YOL002C* and *SCM2* on chromosome XV meet our criteria for a duplicated chromosomal region; this is not shown in Fig. 2 because this area of chromosome IV overlaps with blocks 18 and 9, which have a higher density of homologues than the proposed chromosome IV/XV block. The three-gene match between chromosomes IV and XV is probably spurious, but even if this is counted as a triplication the departure from Poisson expectations is significant ($P = 0.001$).

Phylogenetic analysis. Protein sequences were aligned using default settings in ClustalW with manual editing to remove regions whose alignment was not clear. Branch lengths were estimated with correction for multiple hits[25]. The mean age of duplication was estimated as $(B/(A + B) + C/(A + C))/2$, where A, B and C correspond to the lengths of branches A, B and C shown in Fig. 3. Confidence intervals were estimated by bootstrap analyses for genes where there were >10 inferred substitutions on branch A. One gene pair, *ORC1/SIR3*, was omitted because one of the yeast genes appeared more similar to its human homologue than to its duplicate.

Received 17 December 1996; accepted 20 April 1997.

1. Ohno, S. *Evolution by Gene Duplication* (George Allen and Unwin, London, 1970).
2. Hughes, A. L. The evolution of functionally novel proteins after gene duplication. *Proc. R. Soc. Lond. B* **256,** 119–124 (1994).
3. Kaback, D. B. Yeast genome structure. In *The Yeasts* Vol. 6 (eds Wheals, A. E., Rose, A. H. & Harrison, J. S.) 179–222 (Academic, London, 1995).
4. Olson, M. V. in *The Molecular and Cellular Biology of the Yeast Saccharomyces* Vol. 1 (eds Broach, J. R., Pringle, J. R. & Jones, E. W.) 1–40 (Cold Spring Harbor Laboratory Press, NY, 1991).
5. Smith, M. M. Molecular evolution of the *Saccharomyces cerevisiae* histone gene loci. *J. Mol. Evol.* **24,** 252–259 (1987).
6. Lalo, D., Stettler, S., Mariotte, S., Slonimski, P. P. & Thuriaux, P. Une duplication fossile entre les régions centromériques de deux chromosomes chez la levure. *C.R. Acad. Sci.* **316,** 367–373 (1993).
7. Melnick, L. & Sherman, F. The gene clusters ARC and COR on chromosomes 5 and 10, respectively, of *Saccharomyces cerevisiae* share a common ancestry. *J. Mol. Biol.* **233,** 372–388 (1993).
8. Goffeau, A. *et al*. Life with 6000 genes. *Science* **274,** 546–567 (1996).
9. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215,** 403–410 (1990).
10. Ahn, S. & Tanksley, S. D. Comparative linkage maps of the rice and maize genomes. *Proc. Natl Acad. Sci. USA* **90,** 7980–7984 (1993).
11. Spring, J. Vertebrate evolution by interspecific hybridisation–are we polyploid? *FEBS Lett.* **400,** 2–8 (1997).
12. Roman, H. & Sands, S. M. Heterogeneity of clones of *Saccharomyces* derived from haploid ascospores. *Proc. Natl Acad. Sci. USA* **39,** 171–179 (1953).
13. Kielland-Brandt, M. C., Nilsson-Tillgren, T., Gjermansen, C., Holmberg, S. & Pedersen, M. B. Genetics of brewing yeasts. In *The Yeasts* Vol. 6 (eds Wheals, A. E., Rose, A. H. & Harrison, J. S.) 223–254 (Academic, London, 1995).
14. Ryu, S.-L., Murooka, Y. & Kaneko, Y. Genomic reorganization between two sibling yeast species, *Saccharomyces bayanus* and *Saccharomyces cerevisiae*. *Yeast* **12,** 757–764 (1996).
15. Hendriks, L. *et al*. Phylogenetic relationships among ascomycetes and ascomycete-like yeasts as deduced from small subunit ribosomal subunit RNA sequences. *Syst. Appl. Microbiol.* **15,** 98–104 (1992).
16. Heus, J. J., Zonneveld, B. J. M., Steensma, H. Y. & van den Berg, J. A. The consensus sequence of *Kluyveromyces lactis* centromeres shows homology to functional centromeric DNA from *Saccharomyces cerevisiae*. *Mol. Gen. Genet.* **236,** 355–362 (1993).
17. Stark, M. J. R. & Milner, J. S. Cloning and analysis of the *Kluyveromyces lactis TRP1* gene: a chromosomal locus flanked by genes encoding inorganic pyrophosphatase and histone H3. *Yeast* **5,** 35–50 (1989).
18. Larson, G. P., Castanotto, D., Rossi, J. J. & Malafa, M. P. Isolation and functional analysis of a *Kluyveromyces lactis RAP1* homologue. *Gene* **150,** 35–41 (1994).
19. Bergkamp-Steffens, G. K., Hoekstra, R. & Planta, R. J. Structural and putative regulatory sequences of *Kluyveromyces* ribosomal protein genes. *Yeast* **8,** 903–922 (1992).
20. Hurwitz, N., Segal, M., Marbach, I. & Levitzki, A. Differential activation of yeast adenylyl cyclase by Ras1 and Ras2 depends on the conserved N terminus. *Proc. Natl Acad. Sci. USA* **92,** 11009–11013 (1995).
21. Dohrmann, P. R. *et al*. Parallel pathways of gene regulation: homologous regulators SWI5 and ACE2 differentially control transcription of *HO* and chitinase. *Genes Dev.* **6,** 93–104 (1992).
22. Schmidt, A., Kunz, J. & Hall, M. N. TOR2 is required for organization of the actin cytoskeleton in yeast. *Proc. Natl Acad. Sci. USA* **93,** 13780–13785 (1996).
23. Jansen, R. P., Dowzer, C., Michaelis, C., Galova, M. & Nasmyth, K. Mother cell-specific *HO* expression in budding yeast depends on the unconventional myosin Myo4p and other cytoplasmic proteins. *Cell* **84,** 687–697 (1996).
24. Friis, E. M., Chaloner, W. G. & Crane, P. R. (eds) *The Origins of Angiosperms and their Biological Consequences* (Cambridge Univ. Press, 1987).
25. Fitch, W. M. & Margoliash, E. Construction of phylogenetic trees. *Science* **155,** 279–284 (1967).

Acknowledgements. We thank our colleagues in the yeast genome project; J. I. Garrels for use of the YPD database, which was central to this study; G. Butler, A. T. Lloyd, L. Skrabanek, C. Seoighe, B. Baum and R. Rothstein for comments; and S. Kossida, M. Lewis and R. Keogh for initial work on this project. Yeast genome sequencing in our laboratory was supported by the European Union. *In silico* analysis is supported by the European Union and Forbairt (to K.H.W.) and the Wellcome Trust (to D.C.S.).

Correspondence and requests for materials should be addressed to K.H.W. (e-mail: khwolfe@tcd.ie). Additional details of the dataset and results are available on the World-Wide Web at URL http://acer.gen.tcd.ie/~khwolfe/yeast.

A dendritic-cell-derived C–C chemokine that preferentially attracts naive T cells

Gosse J. Adema*, Franca Hartgers*, Riet Verstraten*, Edwin de Vries*, Gill Marland*, Satish Menon†, Jessica Foster†, Yuming Xu†, Pete Nooyen‡, Terrill McClanahan†, Kevin B. Bacon† & Carl G. Figdor*

** Departments of Tumour Immunology and ‡ Pathology, University Hospital Nijmegen St Radboud, Philips van Leydenlaan 25, 6525 EX Nijmegen, The Netherlands*

† Departments of Immunology and Molecular Biology, DNAX Research Institute, Palo Alto, California 94304, USA

Dendritic cells form a system of highly efficient antigen-presenting cells. After capturing antigen in the periphery, they migrate to lymphoid organs where they present the antigen to T cells[1,2]. Their seemingly unique ability to interact with and sensitize naive T cells gives dendritic cells a central role in the initiation of immune responses and allows them to be used in therapeutic strategies against cancer, viral infection and other diseases. How they interact preferentially with naive rather than activated T lymphocytes is still poorly understood. Chemokines direct the transport of white blood cells in immune surveillance[3,4]. Here we report the identification and characterization of a C-C chemokine (DC-CK1) that is specifically expressed by human dendritic cells at high levels. Tissue distribution analysis demonstrates that dendritic cells present in germinal centres and T-cell areas of secondary lymphoid organs express this chemokine. We show that DC-CK1, in contrast to RANTES, MIP-1α and interleukin-8, preferentially attracts naive T cells ($CD45RA^+$). The specific expression of DC-CK1 by dendritic cells at the site of initiation of an immune response, combined with its chemotactic activity for naive T cells, suggests that DC-CK1 has an important rule in the induction of immune responses.

Dendritic cells are key regulators in immune responses, capable of priming naive T cells. Their potent antigen-presenting capacity can be explained in part by their unique life cycle and their high expression of major histocompatibility complex (MHC) class I and II molecules as well as co-stimulatory molecules[1]. Detailed molecular analysis of dendritic cell function has been hampered, however, by the low numbers of dendritic cells present in blood mononuclear cells. The mechanism by which dendritic cells interact with or activate resting naive T cells to initiate an immune response is not fully understood. One possibility is that secreted cytokines or chemokines preferentially attract or activate naive rather than activated T cells. We generated sufficient numbers of dendritic cells *in vitro*[5,6] to prepare a panel of dendritic-cell cDNA libraries, which allowed us to analyse dendritic cells at the molecular level.