YEAST VOL. **14:** 443–457 (1998)

Evolution of Gene Order and Chromosome Number in *Saccharomyces*, *Kluyveromyces* and Related Fungi

ROBERT S. KEOGH, CATHAL SEOIGHE AND KENNETH H. WOLFE*

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

Received 11 July 1997; accepted 26 October 1997

The extent to which the order of genes along chromosomes is conserved between *Saccharomyces cerevisiae* and related species was studied by analysing data from DNA sequence databases. As expected, the extent of gene order conservation decreases with increasing evolutionary distance. About 59% of adjacent gene pairs in *Kluyveromyces lactis* or *K. marxianus* are also adjacent in *S. cerevisiae*, and a further 16% of *Kluyveromyces* neighbours can be explained in terms of the inferred ancestral gene order in *Saccharomyces* prior to the occurrence of an ancient whole-genome duplication. Only 13% of *Candida albicans* linkages, and no *Schizosaccharomyces pombe* linkages, are conserved. Analysis of gene order arrangements, chromosome numbers, and ribosomal RNA sequences suggests that genome duplication occurred before the divergence of the four species in *Saccharomyces* sensu stricto (all of which have 16 chromosomes), but after this lineage had diverged from *Saccharomyces kluyveri* and the *Kluyveromyces lactis/marxianus* species assemblage. © 1998 John Wiley & Sons, Ltd.

Yeast 14: 443–457, 1998.

KEY WORDS — evolution; polyploidy; gene duplication; gene order; *LEU2*

INTRODUCTION

The order of genes on a chromosome can be changed during evolution by transposition, translocation, deletion, inversion, or gene duplication, but little is known about the rates at which these processes occur. In mammals and plants many large linkage groups are conserved across species (Copeland *et al*., 1993; Moore *et al*., 1995; Paterson *et al*., 1996) at least at the low level of resolution provided by genetic linkage maps as compared to complete genomic sequences. In eubacteria, virtually no conservation of gene order is seen between *Haemophilus influenzae* and *Escherichia coli* (Mushegian and Koonin, 1996) but there is almost complete conservation between the more closely related species *Mycoplasma genitalium* and *M. pneumoniae* (Himmelreich *et al*., 1997).

*Correspondence to: K. H. Wolfe, Department of Genetics, University of Dublin, Trinity College, Dublin 2, Ireland. Tel. (+353) 1 608 1253; fax (+353) 1 679 8558; e-mail: khwolfe@tcd.ie.
Contract/grant sponsor: Forbairt.
Contract/grant sponsor: European Union.

CCC 0749–503X/98/050443–15 $17.50

Several authors have reported conservation of gene order in ascomycete fungi, particularly between *Saccharomyces cerevisiae* and *Kluyveromyces lactis* (for example, Stark and Milner, 1989; Bergkamp-Steffens *et al*., 1992; Mulder *et al*., 1994; Wesolowski-Louvel and Fukuhara, 1995) or *Ashbya gossypii* (Altmann-Jöhl and Philippsen, 1996). The completion of the yeast genome sequence (Goffeau *et al*., 1997) now makes it possible to relate fragmentary gene order information from many fungi to the *S. cerevisiae* gene map, and so to investigate the extent of gene order conservation in different species. In particular, we have re-analysed EMBL database sequences to look for previously unrecognized homologues of *S. cerevisiae* genes in the regions upstream or downstream of known genes from other fungi, and so gathered additional gene order information.

Interpretation of gene order data is complicated by the presence of many large duplicated chromosomal regions in *S. cerevisiae* (Mewes *et al*., 1997; Philippsen *et al*., 1997; Wolfe and Shields, 1997). We have shown by molecular clock analysis that several of these duplicated regions originated in

the *S. cerevisiae* lineage after it had split off from the lineage leading to *K. lactis*, and proposed that all 55 duplicated chromosomal regions arose simultaneously in a whole-genome duplication making yeast, in effect, a degenerate tetraploid. Some regions of the *K. lactis* genome have gene orders that correspond to an amalgamation of genes from both copies of duplicated regions in *S. cerevisiae* (Wolfe and Shields, 1997), which is consistent with ancient tetraploidy in *S. cerevisiae*.

Figure 1 summarizes our model of yeast gene order evolution through tetraploidy, gene deletion and reciprocal translocation. In this study we have used gene adjacency conservation (the extent to which adjacent genes in one species are also adjacent in another) as a measure of gene order conservation. Gene adjacency is changed to a large extent by gene deletion, and to a lesser extent by reciprocal translocation. In the model (Figure 1) we assumed that genes are deleted one at a time, not as large groups of neighbouring genes. This is approximately true because, if each of the $\sim$320 intervals between the duplicated genes (paralogues) making up the large duplicated regions in yeast is considered separately, there is a strong correlation between the numbers of unique genes in each pair of 'sister' intervals (Coissac *et al.*, 1997).

Reciprocal translocations in a duplicated genome such as *S. cerevisiae* can be divided into two classes which we term 'illegitimate' and 'legitimate', depending on the genomic locations of the recombining sites. This is not the same as the classification of recombinations as illegitimate or legitimate, which depends on whether the recombining sites have local *sequence* similarity (not genomic *location* similarity). Illegitimate translocation involves reciprocal recombination between apparently random sites in two chromosomes. Each illegitimate translocation increases the number of duplicated chromosomal blocks by two, because it breaks up two large blocks into four smaller ones (Figure 1c). The recombination sites need not have any sequence similarity, although in practice repeated DNA sequences of some sort might be involved. The intergenic regions in which illegitimate reciprocal translocations are inferred to have happened during *S. cerevisiae* evolution are now very divergent in sequence and it is impossible to tell whether or not recombination events were guided by local sequence similarity. The important point is that the two recombining sites are *not* at equivalent locations within sister duplicated chromosomal regions. In the sense used here, all reciprocal translocations happening in a species without a duplicated genome (such as *K. lactis*) are illegitimate.

The second class of reciprocal translocations that can occur in a duplicated genome is 'legitimate'. These translocations involve recombination within a pair of paralogous genes derived from genome duplication. They appear to be rare, because the chromosomes of the other species of *Saccharomyces* sensu stricto (*S. paradoxus*, *S. bayanus*, *S. pastorianus*) are generally collinear with those of *S. cerevisiae*. Ryu *et al.* (1996) mapped one legitimate reciprocal translocation of this type, between *S. cerevisiae* and *S. bayanus*, to a point inside duplicated block 3 on *S. cerevisiae* chromosomes II and IV. The total number of legitimate reciprocal translocations in *S. bayanus* is probably less than ten (Ryu *et al.*, 1996), and none have been detected in *S. paradoxus* (Naumov *et al.*, 1992; Hawthorne and Philippsen, 1994). A legitimate reciprocal translocation exchanges the flanking unique markers on each side of the pair of paralogues where the recombination occurred (Figure 1d). This has no effect on the number of blocks in the genome, and the only effect on gene adjacency concerns the genes immediately beside the paralogues. Legitimate reciprocal translocations can probably only occur during a limited period after genome duplication, before sequence divergence between the paralogues becomes too great. Moreover, these events can only be detected if a speciation also occurs during this time period. We have included legitimate reciprocal translocations in Figure 1 because this model is general to any organism undergoing genome duplication, and even though legitimate reciprocal translocations are rare in yeast they may be more frequent in other degenerate polyploid species (Morizot, 1990). We emphasise that legitimate reciprocal translocations, as defined here, can only occur in genome-duplicated organisms.

S. cerevisiae and its close relatives have an unusually large number of chromosomes as compared to other yeasts, despite having similar genome sizes (de Jonge *et al.*, 1986; Sor and Fukuhara, 1989). The discovery that the duplicated regions in *S. cerevisiae* include three pairs of centromeres (*CEN2*/*CEN4*; *CEN8*/*CEN11*; *CEN3*/*CEN14*), and that two of these can be related to two of the six *K. lactis* centromeres (Figure 2a), prompted us to re-examine data on chromosome numbers and genome sizes in ascomycetes in the

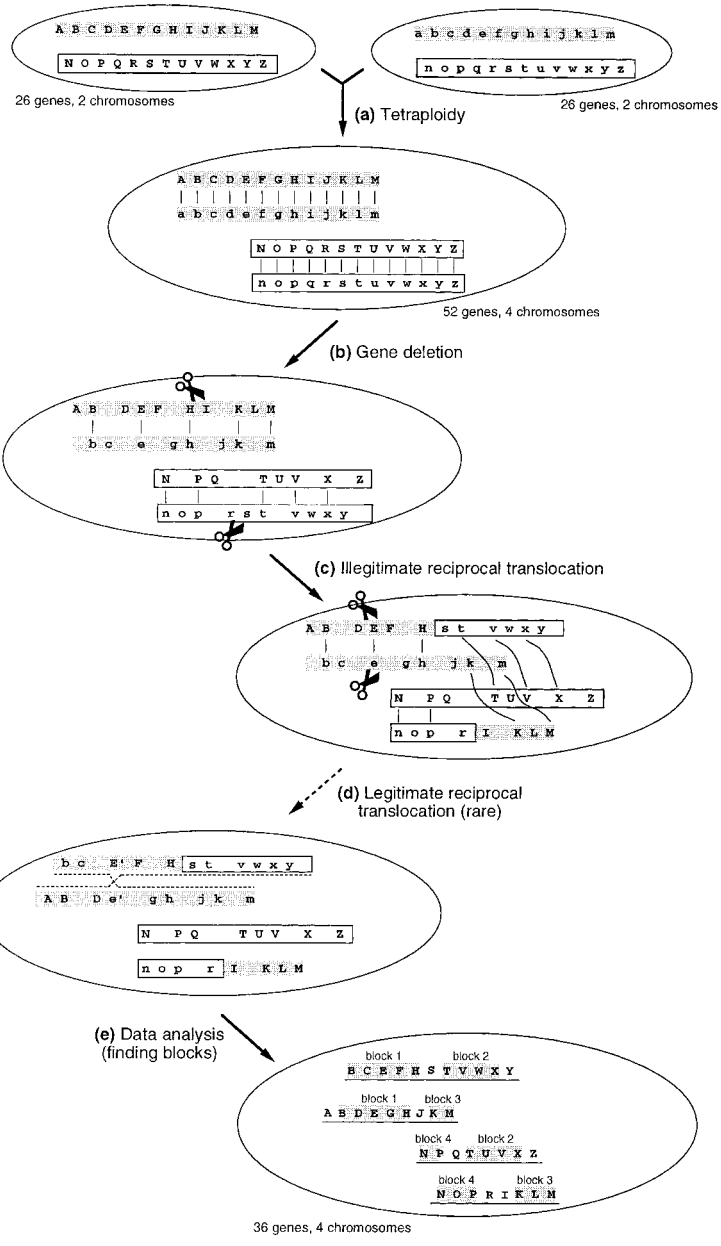

Figure 1. Model of gene order evolution in a duplicated genome such as yeast. A schematic genome is shown with two chromosomes (one grey, one boxed) and 26 genes (letters A–Z). Upper- and lower-case lettering is used to distinguish between the two original sets of chromosomes giving rise to the tetraploid. Vertical lines connect orthologous genes. Stage (c) shows the state of the genome after gene deletion and a single 'illegitimate' reciprocal translocation. Stage (d) illustrates the effect of a rare 'legitimate' reciprocal translocation (involving recombination within a pair of paralogues), such as happened in *S. bayanus* (Ryu *et al.*, 1996). This produces two new hybrid genes (designated E′ and e′) and new combinations of unique genes within blocks (for example, placing gene C near gene F in block 1). Stage (e) shows how this genome would be interpreted using a block-finding method (only upper-case lettering is used because, in practice, it is not possible to determine the origin of each gene in a pair but only to recognize that they are duplicates).

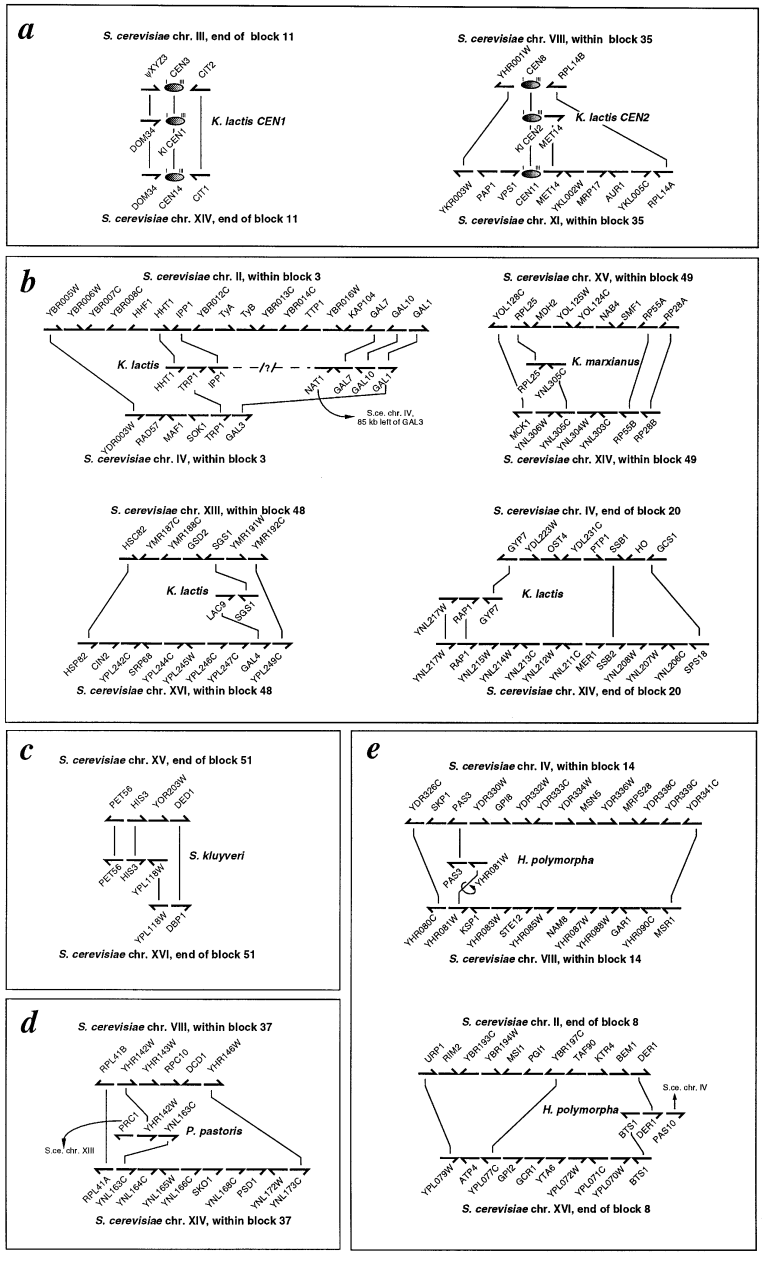

Figure 2. Gene order relationships between some ascomycete species and duplicated regions in *S. cerevisiae*. Arrows indicate the direction of transcription of genes and are not to scale. Vertical lines connect orthologous genes. (a) Relationship between two *K. lactis* centromeres (Heus *et al.*, 1993) and two pairs of *S. cerevisiae* centromeres. Shaded ovals denote centromeres with the relative positions of the CDE I and CDE III elements indicated. ψXYZ3 is a DOM34-related pseudogene on yeast chromosome III (Lalo *et al.*, 1993). Other panels show *S. cerevisiae* relationships to regions from: (b) *Kluyveromyces* species (Webster and Dickson, 1988; Stark and Milner, 1989; Bergkamp-Steffens *et al.*, 1992; Larson *et al.*, 1994; this study); (c) *Saccharomyces kluyveri* (Weinstock and Strathern, 1993); (d) *Pichia pastoris* (Ohi *et al.*, 1996); (e) *Hansenula polymorpha* (Nuttley *et al.*, 1995; Baerends *et al.*, 1996).

light of new ribosomal RNA-based phylogenies for these species (Cai *et al*., 1996; James *et al*., 1997).

MATERIALS AND METHODS

Approximately 1000 sequences from hemiascomycete species other than *S. cerevisiae* and *Schizosaccharomyces pombe* were taken from the EMBL database (release 50 and subsequent daily updates until September 1997). These were searched using BLASTX against the database of all 5790 *S. cerevisiae* proteins used in Wolfe and Shields (1997; http://acer. gen. tcd. ie/~khwolfe/yeast). The significance of matches was assessed by eye, to permit inclusion of some very short but highly conserved sequence matches that occurred at the ends of database sequences. The *S. pombe* dataset comprised 1·577 megabases (55 cosmids in 27 contigs) from chromosome I sequenced at the Sanger Centre, and a consistent significance threshold (BLASTP high score ≥ 200) was used in the analysis of this data. In analyses of the extents of linkage conservation between species, genes that do not have homologues in *S. cerevisiae* were treated as if they were non-existent.

New gene order data was obtained from *K. lactis* by single-pass sequencing of subclones adjacent to previously cloned genes. We identified a *K. lactis* homologue of *SGS1* by sequencing from a *Hin*dIII site 2 kb downstream of *LAC9* in pJ431 (Salmeron and Johnston, 1986), and a homologue of *YML050W* by sequencing from an *Xho*I site 1 kb downstream of *GAL80* in pKLGAL80 (Zenke *et al*., 1993).

Yeast strain designations in different culture collections were interconverted using the World-Wide Web-accessible catalogues of the Centraalbureau voor Schimmelcultures, Netherlands (CBS; http://www. cbs. knaw. nl), the American Type Culture Collection (ATCC; http://www. atcc. org) and Teikyo University Institute of Medical Microbiology (TIMM; http://timm. main. teikyo-u. ac. jp).

RESULTS AND DISCUSSION

Extent of gene order conservation

We used the BLASTX program (Altschul *et al*., 1990) to search every sequence from hemiascomycete fungi (excluding *S. cerevisiae* and *S. pombe*) in the EMBL database against a library of all protein sequences encoded by the *S. cerevisiae* genome (Goffeau *et al*., 1997). BLASTX compares the conceptual six-frame translations of a DNA query sequence against a protein sequence library and so will find matches even if the query sequence is not annotated or contains frameshifts. These searches identified 147 hemiascomycete sequences that contain two adjacent genes (or fragments of genes), both of which have homologues in *S. cerevisiae* (Tables 1 and 2). A similar analysis of data from the *S. pombe* genome project identified 625 pairs of adjacent *S. pombe* genes with *S. cerevisiae* homologues. The adjacent pairs from other species were then compared to the maps of *S. cerevisiae* genes and duplicated chromosomal regions (Wolfe and Shields, 1997). Four possible categories of gene order conservation were recognized (Table 1), depending on whether transcriptional orientation was conserved, and on whether the *S. cerevisiae* genes were adjacent on the same chromosome or were on 'sister' copies of a duplicated chromosomal block (Figure 2).

This information was placed in a phylogenetic context using an approximate tree of 18S ribosomal RNA sequences (Figure 3). The extent of linkage conservation falls off with increasing evolutionary distance from *S. cerevisiae*. At the extremes, gene order is completely conserved in the three other species of *Saccharomyces* sensu stricto (*S. paradoxus*, *S. bayanus* and *S. pastorianus*), whereas there is no linkage conservation at all in *S. pombe*. In *K. lactis* and *K. marxianus* 74–83% of adjacent pairs can be explained in terms of the *S. cerevisiae* map after allowance is made for block duplications and inversions in *S. cerevisiae*. The conservation values are lower for the more distantly related species *Candida albicans* (13%), *C. maltosa* (33%) and *Hansenula polymorpha* (18%), as well as for *Pichia pastoris* (20%), which could not be shown in Figure 3 because its 18S rRNA has not been sequenced but which is expected to lie among these deep branches. Gene order data from other species are scarce but are generally consistent with phylogenetic position (Figure 3). The complete lack of adjacency conservation in the large sample of *S. pombe* genes serves as a control experiment to show that the levels of conservation in other ascomycetes are significant even though they are low.

Table 1. Extent of gene order conservation between *S. cerevisiae* and other ascomycetes.

Species	Adjacent pairs known	Adjacent in *S. cerevisiae*		Conserved between blocks[a]		Not conserved	Total conservation (%)[b]
		Same orientation	Inverted	Same orientation	Inverted		
Ashbya gossypii	4	4	0	0	0	0	100
Candida albicans	31	1	2	1	0	27	13
Candida glabrata	2	2	0	0	0	0	100
Candida guilliermondii	1	0	0	0	0	1	0
Candida maltosa	6	1	1	0	0	4	33
Candida parapsilosis	1	0	0	0	0	1	0
Candida tropicalis	2	0	0	0	0	2	0
Candida utilis[c]	3	1	0	1	0	1[d]	66
Hanseniaspora uvarum[c]	1	0	0	0	0	1	0
Hansenula anomala[c]	1	0	0	0	0	1	0
Hansenula polymorpha	11	0	0	1	1	9	18
Kluyveromyces lactis	31	18	0	4	1	8[d]	74
Kluyveromyces marxianus	6	4	0	1	0	1	83
Pichia pastoris[c]	5	0	0	1	0	4	20
Schwannomyces occidentalis[c]	2	0	0	0	0	2	0
Yamadazyma ohmeri[c]	1	0	1	0	0	0	100
Yarrowia lipolytica	5	1	0	0	0	4	20
Zygosaccharomyces rouxii	1	1	0	0	0	0	100
Saccharomyces kluyveri	4	2	0	1	0	1	75
Saccharomyces carlsbergensis[e]	19	19	0	0	0	0	100
Saccharomyces monacensis[e]	1	1	0	0	0	0	100
Saccharomyces pastorianus[e]	1	1	0	0	0	0	100
Saccharomyces paradoxus	6	6	0	0	0	0	100
Saccharomyces bayanus[f]	1	1	0	0	0	0	100
Saccharomyces uvarum[f]	1	1	0	0	0	0	100
Schizosaccharomyces pombe	625	0	0	0	0	625	0

[a]See Figures 2 and 4.
[b]Sum of all four categories of conservation.
[c]Species not shown in Figure 3 (full-length rRNA sequence not available).
[d]Includes one pair that are almost adjacent in *S. cerevisiae* (see Table 2).
[e]*S. carlsbergensis* is probably an allotetraploid hybrid between *S. monacensis* and *S. cerevisiae* (Hansen and Kielland-Brandt, 1994). Taxonomically, *S. carlsbergensis* and *S. monacensis* are regarded as synonyms of *S. pastorianus* (Barnett, 1992), but *MET2* gene sequences from *S. monacensis* (CBS 1503, the type strain) and the lager chromosome of *S. carlsbergensis* (CBS 1513, type strain) are identical and different from that of *S. pastorianus* (CBS 1538, type strain) (Hansen and Kielland-Brandt, 1994).
[f]*S. uvarum* (type strain: CBS 395) is regarded as a synonym of *S. bayanus* (type strain: CBS 380) (Barnett, 1992) and their *MET2* sequences are identical (Hansen and Kielland-Brandt, 1994).

Comparison of Kluyveromyces *results to theoretical predictions*

Analysis of *Kluyveromyces* data shows that 22 out of 37 adjacent gene pairs (59%) are also adjacent in *S. cerevisiae*, and six out of 37 (16%) are conserved between duplicated blocks (Table 1). Is this consistent with the hypothesis of whole genome duplication in *S. cerevisiae*?

To predict these two quantities, which we term 'adjacency conservation' and 'block conservation', we need to take account of three factors: (i) the incompleteness of the map of duplicated regions in the yeast genome; (ii) the break-up of adjacencies caused by reciprocal translocations; and (iii) the presence of duplicated genes in *S. cerevisiae* which will increase the number of apparent conserved adjacencies. Assuming random single-gene deletions, the predicted extent of adjacency conservation is $P_{adj}=t\{1-0{\cdot}5(1-2d)^2\}$, and of block conservation is $P_{block}=t\{b0{\cdot}5(1-2d)^2\}$, where d is the proportion of original genes retained in

Table 2. Ascomycete EMBL database entries containing two or more genes with *S. cerevisiae* homologues.[a]

Accession numbers[b]	Genes[c]	Linkage conservation
Ashbya gossypii:		
A29820	TEF2→ MUD1→	conserved on II
X91046 and ref. 1	RSC6→ THR4→ ←CTR86 ←PWP2	conserved on III
Candida albicans:		
U13193	STE6→ ← UBA1	conserved on XI
U58133	RAD16→ LYS2→	inverted on II
AF000120/AF000121	PET8→ NFS1→ LEU2→	III/XIV, beside block 11 (see Figure 4)
D83180/D83181	CEG1→ ←FRE1	not conserved
L04305	ERG7→ ←YCR010C	not conserved
L04943	ENO1→ ←YLR231C	not conserved
L08824	FMS1→ YPL225W→	not conserved
L25759	OYE2→ ←YHR052W	not conserved
M29935	TEF1→ ←YPL247C	not conserved
M94160	←PEP8 CDC25→ ←HTB1	not conserved
M94674	←IAH1 MAL32→ ←YNL321W	not conserved
S65451/J04230	HSP12→ TMP1→	not conserved
U09781	PTR2→ ←YPL009C	not conserved
U37371/X78466	CCT8→ ←TRP1 ←YJL029C	not conserved
U67193	ERG11→THR1→	not conserved[d]
U72980	STE7→ ←TAF61	not conserved
X52420/X96850/X88804	CHS2→ SRM1→ ←POL3	not conserved
X53823	YJL084C→ YBR008C→	not conserved
X62496	YLR326W→ ←YDR357C	not conserved
X74952	FAS1→ ←YLR410W	not conserved
X76689	CAN1→ ←HAL2	not conserved
X78968	DFR1→ ←YJL054W	not conserved
X81025	RAD14→ HSP82→	not conserved
Y10377	TOP2→ ←SDH1	not conserved
Z25870	CDC10→ RAD27→	not conserved
Z54197	UBI4→ YHL030W→	not conserved
Candida glabrata:		
M69146	ACE1→ ←YGL164C	conserved on VII
X97320 and ref. 2	SEC14→ ←NAM7	conserved on XIII
Candida maltosa:		
D29759	←GAL10 GAL1→	conserved on II
X05459/X72939	NFS1→ LEU2→	inverted on III (see Figure 4)
D12717	FRE2→ ERG11→	not conserved
D12718	DIT2→ ←YPL135W	not conserved
M58322	ADE1→ ←YHR031C	not conserved
X17310	HIS5→ YMR188C→	not conserved
Candida parapsilosis:		
X99635	URA3→ ←YIL006W	not conserved
Candida tropicalis:		
M23673	ERG11→ THR1→	not conserved[d]
X54875	VMA2→ ←YJL200C	not conserved
Candida utilis:		
D67040	←RPL41B YHR142W→	conserved on VIII
M16014	LEU2→ ←RLP7	III/XIV, beside block 11 (see Figure 4)
D14851/D32213	ERG10→ ←SHA3	almost conserved; S. ce. XVI has ERG10→ YPL027W→ ←SHA3
Hansenula anomala:		
X16051	CYB2→ ←SDS22	not conserved
Hansenula polymorpha:		
U37763	YHR081W→ PAS3→	IV/VIII, block 14, inverted (see Figure 2e)

Table 2. Ascomycete EMBL database entries containing two or more genes with *S. cerevisiae* homologues.[a]

Accession numbers[b]	Genes[c]	Linkage conservation
U22930	←BTS1 DER1→ PAS10→	BTS1-DER1 is conserved on II/XVI near block 8 (see Figure 2e)
A06214	YOR388C→ PCL1→	not conserved
A11156	←YOR155C TRP3→	not conserved
A11168/X02424	GUK1→ ←RVS161 TKL1→	not conserved
U00889	LEU2→ YOL105C→	not conserved (see Figure 4)
U40996	YLR364W→ YOR374W→	not conserved
X58862	←YFR021W YML070W→	not conserved
Z46868	SPR1→ YKL034W→	not conserved
Hanseniaspora uvarum:		
U13635	PDC1→ ←YFL017C	not conserved
Kluyveromyces lactis:		
L25779	←RFT1 HAP3→	conserved on II
U65983 and ref. 3	TKL2→ LYS2→	conserved on II
U04714	ERD1→ ←YDR412W	conserved on IV
X76027	APA2→ QCR7→	conserved on IV
U48701	CDC68→ ←CHC1	conserved on VII
L05777/L05772	←RPL32 RPL30A→	conserved on VII
X76026	ERG20→ QCR8→	conserved on X
X74292	←YLR181C SWI6→	conserved on XII
Z21512 and this study	GAL80→ YML050W→	conserved on XIII
AJ001358	URA5→ ←SEC65	conserved on XIII
A26615	RPL41A→ YNL161W→	conserved on XIV
M68870	GAL11→ YOL049W→	conserved on XV
A36834	YOL119C→ RP28A→	conserved on XV
X07039	←GAL1 GAL10→ GAL7→ ←NAT1	GAL genes are conserved on II (see Figure 2b)
X70373/X07038 and ref. 4	←ZWF1 ←YNL240C KEX2→ YTP1→	conserved on XIV but S. ce. has LAP3 between YNL240C and KEX2
X14230	HHT1→ TRP1→ ←IPP1	II/IV, in block 3 (see Figure 2b)
X65545	RLP7→ LEU2→	III/XIV, near block 11 (see Figure 4)
X73629	YNL217W→ RAP1→ ←GYP7	IV/XIV, near block 20 (see Figure 2b)
M15210 and this study	GAL4→ ←SGS1	XIII/XVI, in block 48 (see Figure 2b)
A27712/X17654	←LAG2 PGK1→	not conserved
U19586	←KIN28 MRF1→	not conserved
U72486	MET17→ ←YLL015W	not conserved
X52871 and ref. 5	←GAP1 ADH1→	not conserved
X76028	←CTF18 CBF1→	not conserved
Z17316	GLO1→ PFK2→	not conserved
Kluyveromyces marxianus:		
D10580	←YHR142W RPL41B→	conserved on VIII
S53438/S53436/S53434	CRY2→ ←RPS24A RPL46→	conserved on X
X69583	RED1→ RPS33B→	conserved on XII
S53422	RPL25→ ←YNL305C	XV/XVI, in block 49 (see Figure 2b)
X57202	YHL040C→ SUC2→	not conserved
Pichia guilliermondii:		
Z74991/Z49093	TOP2→ RIB1→	not conserved
Pichia guilliermondii:		
Z74991/Z49093	TOP2→ RIB1→	not conserved
Pichia pastoris:		
X87987	←YNL163C ←YHR142W PRC1→	YNL163C-YHR142W is conserved in block 37 (see Figure 2d)
U58140	RRN3→ YMR026C→	not conserved
U69170	HIS3→ TTP1→	not conserved
X96945	PAS5→ VPS15→	not conserved

Table 2. Ascomycete EMBL database entries containing two or more genes with *S. cerevisiae* homologues.[a]

Accession numbers[b]	Genes[c]	Linkage conservation
Schwanniomyces occidentalis:		
S38381	YIL172C→ PTC2→	not conserved
U23210	ADE2→ ←YBL028C	not conserved
Yamadazyma ohmeri:		
Z35101	NFS1→ LEU2→	inverted on III (see Figure 4)
Yarrowia lipolytica:		
Z22571/Z22570	URA5→ ←SEC65	conserved on XIII
X99537/X99538	←YGL054C SEC62→	not conserved
M17741	GYP7→ PRB1→	not conserved
M91598	PGK1→ ←YDR541C	not conserved
X69988	POT1→ ←HSP42	not conserved
Zygosaccharomyces rouxii:		
D00134	TDH2→ ←YJR008W	conserved
Saccharomyces kluyveri:		
M82964	CDC25→ IMH1→	conserved on XII
Z14125	←PET56 HIS3→ ←YPL118W	XV/XVI, beside block 51 (see Figure 2c)
X56042	←COX17 CYR1→	not conserved
Saccharomyces carlsbergensis[e]:		
'lager' chromosomes:		
Z86109	YCL010C thru CEN3[f]	conserved on III
U13062	BIK1→ HIS4→ YCL031C→	conserved on III
L26504	MET10→ ←SMC2	conserved on VI
cerevisiae-like chromosomes:		
K01752/K01609	←GAL1 GAL10→ GAL7→	conserved on II
M12601/M27823	←AGT1 ←YGR291C MAL12→	conserved on VII
X01100	RP28B→ RP55B→ ←YNL303W ←YNL304W	conserved on XIV
Saccharomyces monacensis:		
Y08688	←ORM1 ACB1→	conserved on VII
Saccharomyces pastorianus (strain KBY001):		
D86480	YGR178C→ ATF2→	conserved on VII
Saccharomyces paradoxus (S. douglasii):		
X73886	DED81→ ARG4→ ←YSC83	conserved on VIII
X94370	YHL037C→ CBP2→ ←YHL039W	conserved on VIII
X12864	←RHC18 NAM2→ ←YLR381W	conserved on XII
Saccharomyces bayanus:		
D12534	ACT1→ ←YFL040W	conserved on VI
Saccharomyces uvarum:		
X07976	SUV3→ ERG10→	conserved on XVI

[a]For simplicity, we have listed only one *S. cerevisiae* 'homologue' for each gene. For some sequences there were two or more equally similar *S. cerevisiae* genes and we examined all possibilities for linkage conservation. For some genes, such as transketolase (*TKL1*; Flechner *et al*., 1996; Schenk *et al*., 1997), sequences from different ascomycetes may not be true orthologues.
[b]Where multiple accession numbers are listed the sequences overlap. Additional references: 1, Altmann-Jöhl and Philippsen (1996); 2, Dundon and Islam (1997) and W. Dundon, personal communication; 3, Jacoby and Heinisch (1997); 4, Wesolowski-Louvel and Fukuhara (1995); 5, Shuster (1990).
[c]Only *S. cerevisiae* gene names are listed. Arrows indicate direction of transcription.
[d]The *ERG11→ THR1→* linkage is conserved betwen *C. albicans* and *C. tropicalis* but not *S. cerevisiae*.
[e]*S. carlsbergensis* sequences are described as either '*cerevisiae*-like' or 'lager' based on degree of similarity to the *S. cerevisiae* genome sequence in intergenic spacer regions. 'Lager' sequences are probably derived from *S. monacensis* (Hansen and Kielland-Brandt, 1994).
[f]Ten genes (Andersen and Nilsson-Tillgren, 1997). The *DOM34* homologue *XYZ3* (Lalo *et al.,* 1993; see Figure 2a) is intact on this *S. carlsbergensis* chromosome but not in *S. cerevisiae*. Homologues of *S. cerevisiae CWH36* and *YCL006C* are not intact on this *S. carlsbergensis* chromosome and are either pseudogenes in *S. carlsbergensis* or spurious ORFs in *S. cerevisiae*.

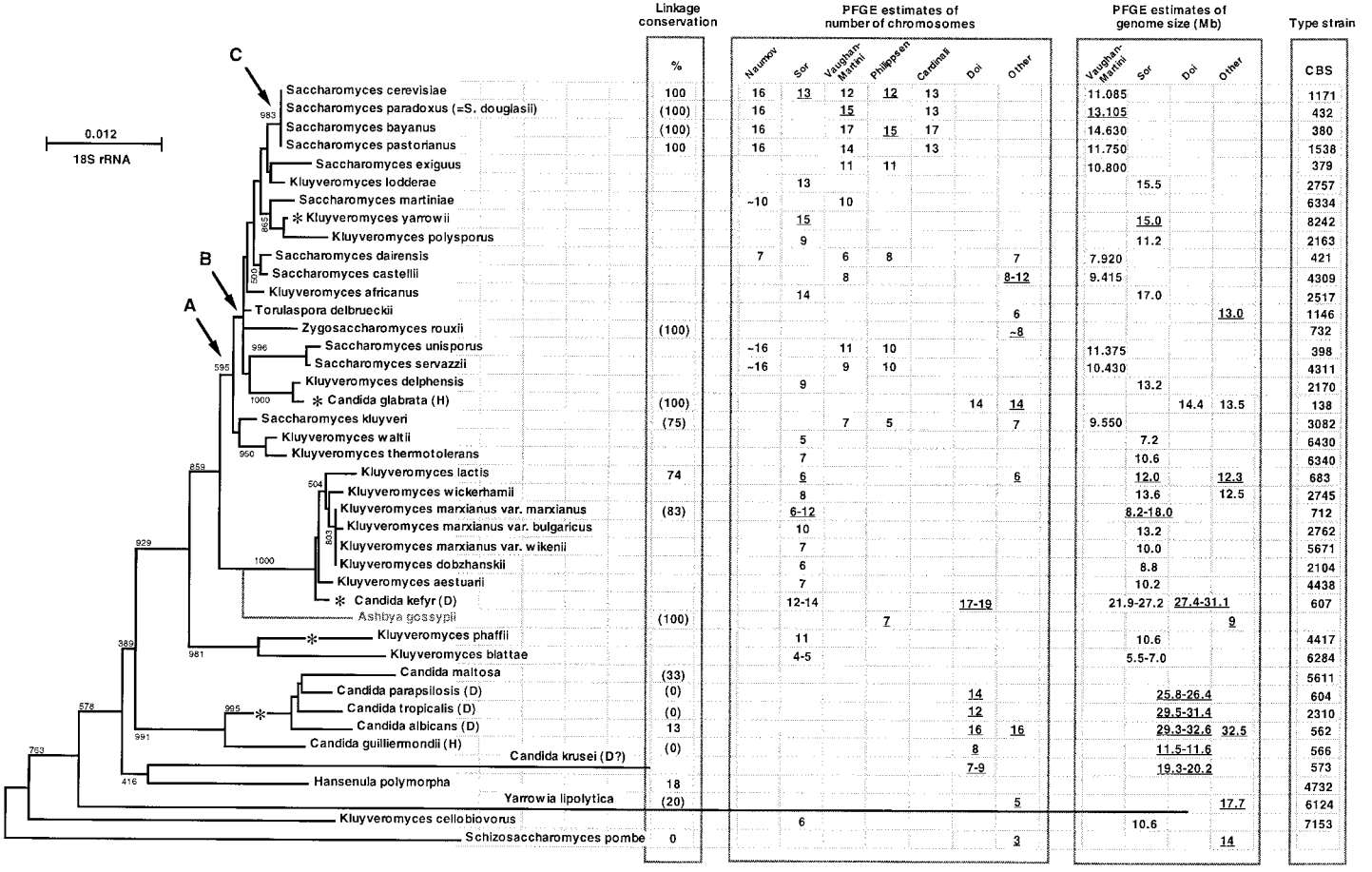

Figure 3. 18S ribosomal RNA phylogeny of some ascomycetes and summary of published data on their genomes. The phylogenetic tree is essentially the same as those published by Cai *et al.* (1996) and James *et al.* (1997) and was produced by the neighbour-joining method from a ClustalW alignment (Thompson *et al.*, 1994) of near-full-length sequences. Bootstrap values (1000 replicates) that are not shown were below 500. Points A, B and C are discussed in the text. Asterisks indicate places where major changes in chromosome number may have occurred. H and D after *Candida* species names indicate their designation as either haploid or diploid by Doi *et al.* (1992; *Candida krusei* was described as 'probably diploid') . The 'Linkage conservation' panel refers to the 'Total Conservation' column in Table 1; values in parentheses are based on fewer than 10 linked pairs. The 'Number of chromosomes' and 'Genome size' panels summarize estimates from pulsed-field gel electrophoresis (PFGE) experiments by several laboratories. Tildes in estimates of chromosome number indicate cases where no explicit statement was made in the text of the cited reference. The type strains of each species are named in the rightmost panel. These strains were used for most of the PFGE analyses (except where indicated by underlining), and rRNA sequencing (all except *C. glabrata* and *Ashbya gossypii*). The placement of *A. gossypii* is an estimate based on its position in a separate tree drawn from the 800 bases of 18S rRNA sequence that are available for this species (Messner *et al.*, 1995), but is consistent with the results of Prillinger *et al.* (1997) for its close relative *Holleya sinecauda*. Major references: Jäger and Philippsen (1989); Sor and Fukuhara (1989); Doi *et al.* (1992); Naumov *et al.* (1992; 1995); Vaughan-Martini *et al.* (1993); Cardinali and Martini (1994). Other references for chromosome number: P. Philippsen, personal communication (*A. gossypii*); Vaughan-Martini and Barcaccia (1996; *S. dairensis*, *S. castellii*); Oda and Tonomura (1995; *T. delbrueckii*, *Z. rouxii*); Kaufmann and Merz (1989; *C. glabrata*); Weinstock and Strathern (1993; *S. kluyveri*); Heus *et al.* (1993; *K. lactis*); Chu *et al.* (1993; *C. albicans*). Other references for genome size: Maleszka and Clark-Walker (1993; *T. delbrueckii*, *C. glabrata*, *K. lactis*, *K. wickerhamii*); P. Philippsen, personal communication (*A. gossypii*); Chu *et al.* (1993; *C. albicans*); Vernis *et al.* (1997; *Y. lipolytica*); Hoheisel *et al.* (1993; *S. pombe*).

duplicate after tetraploidy; b is the fraction of the genome covered by the map of duplicated blocks; and t is the probability that two genes that were originally adjacent have not been separated by a reciprocal translocation.

Under the genome-duplication hypothesis every region of the yeast genome should be paired with a 'sister' region, but so far we have only been able to map 50% of the genome into duplicated blocks (Wolfe and Shields, 1997). The other half of the genome is assumed to contain many additional small, fragmented blocks, as well as undiscovered end fragments of the known blocks. We have estimated elsewhere (C.S. and K.H.W., in preparation) that the combined fraction of the genome occupied by blocks that have been at least partially discovered is $b=0{\cdot}68$, that $d=0{\cdot}08$, and that about 85 reciprocal translocations occurred within the yeast genome after its duplication. We estimated previously that the age of the whole-genome duplication was 0·71 times the age of the divergence between *S. cerevisiae* and *K. lactis* (Wolfe and Shields, 1997), so assuming a molecular clock for translocations this suggests that approximately 240 translocations have occurred between *S. cerevisiae* and *K. lactis*. Each translocation disrupts two adjacencies (Figure 1c), so 480 breakpoints among $\sim$5400 original genes yields a value of $t=0{\cdot}91$. Substituting these values into the formulae above gives $P_{adj}=0{\cdot}59$ and $P_{block}=0{\cdot}22$, which are reasonably close to the observed values. There are many uncertainties and approximations in these calculations, but they indicate that the observed extent of linkage conservation in *K. lactis* is consistent with the genome-duplication hypothesis.

Inversions at LEU2

Table 1 includes a few examples where linkage of a pair of adjacent genes has been conserved between *S. cerevisiae* and another species, but the transcriptional orientation of one of the genes has been inverted. The relationship between *LEU2* and its neighbours is interesting because data are available from a range of species (Figure 4). We interpret Figure 4 to mean that the four genes *LEU2*, *NFS1*, *PET8* and *RLP7* were all adjacent in an ascomycete ancestor. Genome duplication and subsequent deletions in *S. cerevisiae* left *LEU2* and *NFS1* on chromosome III, but *PET8* and *RLP7* on chromosome XIV; this may be an extension of duplicated block 11 (Wolfe and Shields, 1997; see also Lalo *et al.*, 1993), which lies to the right of these genes. However, the orientation of *LEU2* in *S. cerevisiae* and *C. utilis* is opposite to that in other species (see also Sharp and Wolfe, 1993), and no simple explanation for the current gene arrangements is apparent. One possible (but convoluted) explanation of the data in Figure 4 is that the ancestral gene order was $\leftarrow$*LEU2* $\leftarrow$*NFS1* $\leftarrow$*PET8* $\leftarrow$*RLP7*, with an inversion of *LEU2* in *S. cerevisiae*, an independent multigene inversion in *C. utilis* spanning the three-gene cluster *LEU2-NFS1-PET8* bringing *LEU2* and *RLP7* into their present tail-to-tail arrangement, and a transposition of *NFS1-PET8* to elsewhere in the *K. lactis* genome.

Evolution of chromosome number and genome size

We tried to examine the evolution of chromosome number and genome size in ascomycetes by combining the 18S ribosomal RNA phylogeny with published pulsed-field gel electrophoresis (PFGE) profiles for the same species (Figure 3). The PFGE technique tends to underestimate the number of chromosomes because bands may co-migrate on gels, but there is a qualitative difference between *Saccharomyces* sensu stricto and other yeasts in terms of the presence of many small chromosomes of <500 kb (de Jonge *et al.*, 1986; Johnston and Mortimer, 1986; Vaughan-Martini *et al.*, 1993). There is also considerable variation among laboratories in PFGE results (Figure 3), so apparent differences between species are probably only reliable if the data come from a single laboratory.

Much of Figure 3 is inconclusive as regards chromosome number evolution. This is caused by poor resolution (low bootstrap values) in the phylogenetic tree, as well as possible inaccuracies in the PFGE data and/or possible aneuploidy, as seen in industrial and clinical strains of yeast (Johnston *et al.*, 1989; Hadfield *et al.*, 1995; Clemons *et al.*, 1997). However, as pointed out by others (de Jonge *et al.*, 1986; Johnston and Mortimer, 1986; Sor and Fukuhara, 1989), most of the species that lie on the deeper branches (below point A in Figure 3), including the *C. albicans* group and the *K. lactis*/*K. marxianus* group, have haploid chromosome numbers of between six and eight, which implies an approximate doubling in *Saccharomyces* sensu stricto. A parsimonious explanation of the data in Figure 3 alone is that chromosome number increased from 6–8 to 16

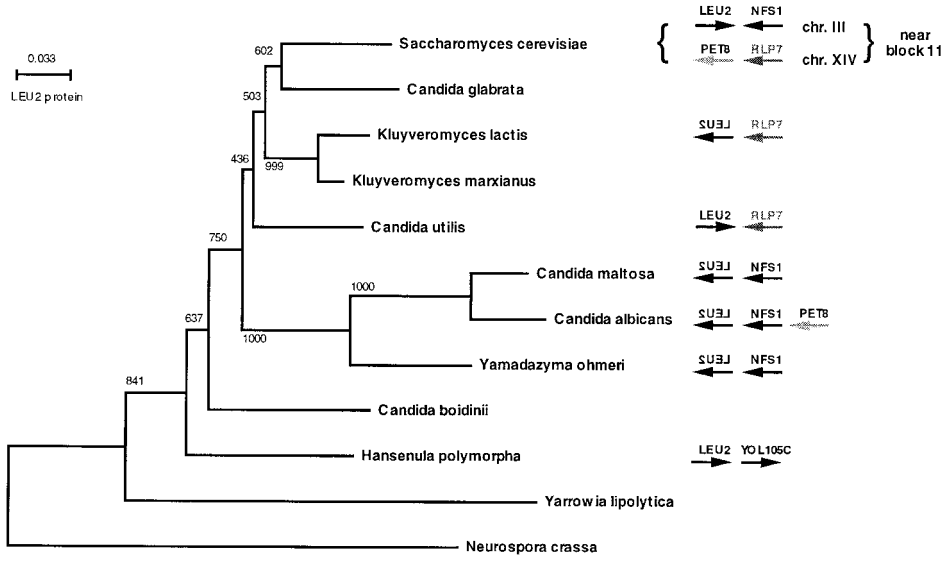

Figure 4. Evolution of gene order and orientation near *LEU2*. Arrows indicate directions of gene transcription and are not to scale. The phylogenetic tree was drawn by the neighbour-joining method from a ClustalW alignment (Thompson *et al.*, 1994) of *LEU2* protein sequences. Bootstrap values from 1000 replicates are shown. There is no information about genes neighbouring *LEU2* in some species. References (from top to bottom of the tree): Goffeau *et al.* (1997); Kitada (1997); Zhang *et al.* (1992); Bergkamp *et al.* (1991); Hamasawa *et al.* (1987); Becher *et al.* (1994); Plant and Poulter (1997); Piredda and Gaillardin (1994); Sakai and Tani (1992); Agaphonov *et al.* (1994); Davidow *et al.* (1987); Li *et al.* (1993).

somewhere on the *S. cerevisiae* lineage between points A and C.

The arrangement of one set of adjacent genes in *Saccharomyces kluyveri* (*PET56-HIS3-YPL118W*; Figure 2c; Weinstock and Strathern, 1993) indicates that genome duplication in *S. cerevisiae* occurred after *S. kluyveri* and *S. cerevisiae* diverged. This is consistent with *S. kluyveri* having only seven chromosomes (eight in one strain; Vaughan-Martini *et al.*, 1993; Weinstock and Strathern, 1993), and places the whole-genome duplication somewhere between points B and C (Figure 3).

Figure 3 suggests that several other major changes in ploidy may have occurred during ascomycete evolution. The clearest example is the comparison of *K. blattae* to its close relative *K. phaffii*, which contains approximately twice as much DNA and twice as many chromosomes (Sor and Fukuhara, 1989). A ploidy change may also have occurred between *K. delphensis* and its close relative *C. glabrata* (genetically haploid; Whelan, 1987; Doi *et al.*, 1992), which have similar genome sizes but 9 and 14 PFGE bands, respectively. Other apparent substantial changes are marked by asterisks in Figure 3. Sor and Fukuhara (1989) reported a wide range of genome sizes and chromosome numbers in *K. marxianus* var. *marxianus*, and some strains (such as CBS 1553 with 12 chromosomes and 18·0 megabases) may be tetraploid with respect to others.

Many asexual ascomycete species such as *C. albicans* appear to be permanently stuck in a diploid state. Given sufficient time, an asexual diploid genome would be expected to undergo 'haploidization' (Ohno, 1970) as its alleles diverge in sequence from one another, or allele deletions occur. In *C. albicans* alleles are highly similar in sequence (Miyasaki *et al.*, 1994), but some divergence is apparent in terms of the sizes of allelic chromosomes (Chu *et al.*, 1993). Whether asexual lineages can persist for long times has been questioned (Berbee and Taylor, 1993) but if they can, haploidization will cause gene order changes as in Figure 1. It is possible that repeated cycles of long periods of asexuality followed by sexual exchanges could result in multiple successive genome duplications followed by downsizing, with consequent turnover of the gene order during each cycle.

ACKNOWLEDGEMENTS

The relationship shown in Figure 2d was first noted by Bobby Baum (Yeast Update 4.6, March 1996). We thank P. Philippsen for *Ashbya* information, and two anonymous reviewers for helpful comments. This study was supported by Forbairt and the European Union biotechnology programme.

NOTE ADDED IN PROOF

Recent EMBL database updates include three further examples of pairs of adjacent genes in *K. lactis* that are distributed on sister blocks in *S. cerevisiae*, similar to those in Figure 2b. These are accession numbers U93209 (ARG8→ ←KRE1, corresponding to block 49 on yeast chromosomes XIV and XV), AF023920 (←YDR101C PDA1→, block 13 on chromosomes IV and V), and AF022776 (UBP2→ YDR372C→, block 23 on chromosomes IV and XV).

REFERENCES

Agaphonov, M. O., Poznyakovski, A. I., Bogdanova, A. I. and Ter-Avanesyan, M. D. (1994). Isolation and characterization of the *LEU2* gene of *Hansenula polymorpha*. *Yeast* **10,** 509–513.

Altmann-Jöhl, R. and Philippsen, P. (1996). *AgTHR4*, a new selection marker for transformation of the filamentous fungus *Ashbya gossypii*, maps in a four-gene cluster that is conserved between *A. gossypii* and *Saccharomyces cerevisiae*. *Mol. Gen. Genet.* **250,** 69–80.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215,** 403–410.

Andersen, T. H. and Nilsson-Tillgren, T. (1997). A fungal minisatellite. Nature 386, 771.

Baerends, R. J., Rasmussen, S. W., Hilbrands, R. E., et al. (1996). *The Hansenula polymorpha PER9* gene encodes a peroxisomal membrane protein essential for peroxisome assembly and integrity. *J. Biol. Chem.* **271,** 8887–8894.

Barnett, J. A. (1992). A taxonomy of the genus *Saccharomyces* Meyen ex Reess: a short review for non-taxonomists. *Yeast* **8,** 1–23.

Becher, D., Schulze, S., Kasuske, A., Schulze, H., Samsonova, I. A. and Oliver, S. G. (1994). Molecular analysis of a *leu2*-mutant of *Candida maltosa* demonstrates the presence of multiple alleles. *Curr. Genet.* **26,** 208–216.

Berbee, M. L. and Taylor, J. W. (1993). Ascomycete relationships: dating the origin of asexual lineages with 18S ribosomal RNA gene sequence data. In Reynolds, D. R. and Taylor, J. W. (Eds), *The Fungal Holomorph: Mitotic, Meiotic and Pleomorphic Speciation in Fungal Systematics*. CAB International, Wallingford, UK, pp. 67–78.

Bergkamp, R. J., Geerse, R. H., Verbakel, J. M., Musters, W. and Planta, R. J. (1991). Cloning and disruption of the *LEU2* gene of *Kluyveromyces marxianus* CBS 6556. *Yeast* **7,** 963–970.

Bergkamp-Steffens, G. K., Hoekstra, R. and Planta, R. J. (1992). Structural and putative regulatory sequences of *Kluyveromyces* ribosomal protein genes. *Yeast* **8,** 903–922.

Cai, J., Roberts, I. N. and Collins, M. D. (1996). Phylogenetic relationships among members of the ascomycetous yeast genera *Brettanomyces*, *Debaryomyces*, *Dekkera*, and *Kluyveromyces* deduced by small-subunit rRNA gene sequences. *Int. J. Syst. Bacteriol.* **46,** 542–549.

Cardinali, G. and Martini, A. (1994). Electrophoretic karyotypes of authentic strains of the sensu stricto group of the genus *Saccharomyces*. *Int. J. Syst. Bacteriol.* **44,** 791–797.

Chu, W.-S., Magee, B. B. and Magee, P. T. (1993). Construction of a *Sfi*I macrorestriction map of the *Candida albicans* genome. *J. Bacteriol.* **175,** 6637–6651.

Clemons, K. V., Park, P., McCusker, J. H., McCullough, M. J., Davis, R. W. and Stevens, D. A. (1997). Application of DNA typing methods and genetic analysis to epidemiology and taxonomy of *Saccharomyces* isolates. *J. Clin. Microbiol.* **35,** 1822–1828.

Coissac, E., Maillier, E. and Netter, P. (1997). A comparative study of duplications in bacteria and eukaryotes: the importance of telomeres. *Mol. Biol. Evol.*, in press.

Copeland, N. G., Jenkins, N. A., Gilbert, D. J., *et al.* (1993). A genetic linkage map of the mouse: current applications and future prospects. *Science* **262,** 57–66.

Davidow, L. S., Kaczmarek, F. S., DeZeeuw, J. R., *et al.* (1987). The *Yarrowia lipolytica LEU2* gene. *Curr. Genet.* **11,** 377–383.

de Jonge, P., de Jongh, F. C. M., Meijers, R., Steensma, H. Y. and Scheffers, W. A. (1986). Orthogonal-field-alternation gel electrophoresis banding patterns of DNA from yeasts. *Yeast* **2,** 193–204.

Doi, M., Homma, M., Chindamporn, A. and Tanaka, K. (1992). Estimation of chromosome number and size by pulsed-field gel electrophoresis (PFGE) in medically important *Candida* species. *J. Gen. Microbiol.* **138,** 2243–2251.

Dundon, W. and Islam, K. (1997). Nucleotide sequence of the gene coding for SEC14p in *Candida (torulopsis) glabrata*. *Gene* **193,** 115–118.

Flechner, A., Dressen, U., Westhoff, P., Henze, K., Schnarrenberger, C. and Martin, W. (1996). Molecular characterization of transketolase (EC 2.2.1.1) active in the Calvin cycle of spinach chloroplasts. *Plant Mol. Biol.* **32,** 475–484.

Goffeau, A., *et al*. (1997). The Yeast Genome Directory. *Nature* **387**(Suppl.), 5–105.

Hadfield, C., Harikrishna, J. A. and Wilson, J. A. (1995). Determination of chromosome copy numbers in *Saccharomyces cerevisiae* strains via integrative probe and blot hybridization techniques. *Curr. Genet.* **27,** 217–228.

Hamasawa, K., Kobayashi, Y., Harada, S., Yoda, K., Yamasaki, M. and Tamura, G. (1987). Molecular cloning and nucleotide sequence of the 3-isopropylmalate dehydrogenase gene of *Candida utilis*. *J. Gen. Microbiol.* **133,** 1089–1097.

Hansen, J. and Kielland-Brandt, M. C. (1994). *Saccharomyces carlsbergensis* contains two functional *MET2* alleles similar to homologues from *S. cerevisiae* and *S. monacensis*. *Gene* **140,** 33–40.

Hawthorne, D. and Philippsen, P. (1994). Genetic and molecular analysis of hybrids in the genus *Saccharomyces* involving *S. cerevisiae*, *S. uvarum* and a new species, *S. douglasii*. *Yeast* **10,** 1285–1296.

Heus, J. J., Zonneveld, B. J., Steensma, H. Y. and van den Berg, J. A. (1993). The consensus sequence of *Kluyveromyces lactis* centromeres shows homology to functional centromeric DNA from *Saccharomyces cerevisiae*. *Mol. Gen. Genet.* **236,** 355–362.

Himmelreich, R., Plagens, H., Hilbert, H., Reiner, B. and Herrmann, R. (1997). Comparative analysis of the genomes of the bacteria *Mycoplasma pneumoniae* and *Mycoplasma genitalium*. *Nucl. Acids Res.* **25,** 701–712.

Hoheisel, J. D., Maier, E., Mott, R., *et al*. (1993). High resolution cosmid and P1 maps spanning the 14 Mb genome of the fission yeast *S. pombe*. *Cell* **73,** 109–120.

Jacoby, J. J. and Heinisch, J. J. (1997). Analysis of a transketolase gene from *Kluyveromyces lactis* reveals that the yeast enzymes are more related to transketolases of prokaryotic origins than to those of higher eukaryotes. *Curr. Genet.* **31,** 15–21.

Jäger, D. and Philippsen, P. (1989). Many yeast chromosomes lack the telomere-specific Y′ sequence. *Mol. Cell. Biol.* **9,** 5754–5757.

James, S. A., Cai, J., Roberts, I. N. and Collins, M. D. (1997). A phylogenetic analysis of the genus *Saccharomyces* based on 18S rRNA gene sequences: description of *Saccharomyces kunashirensis* sp. nov. and *Saccharomyces martiniae* sp. nov. *Int. J. Syst. Bacteriol.* **47,** 453–460.

Johnston, J. R., Curran, L., Contopoulou, R. C. and Mortimer, R. K. (1989). Electrophoretic karyotyping of commercial brewing and distilling strains of *Saccharomyces* and of other yeasts. *Yeast* **5,** S255–S259.

Johnston, J. R. and Mortimer, R. K. (1986). Electrophoretic karyotyping of laboratory and commercial strains of *Saccharomyces* and other yeasts. *Int. J. Syst. Bacteriol.* **36,** 569–572.

Kaufmann, C. S. and Merz, W. G. (1989). Electrophoretic karyotypes of *Torulopsis glabrata*. *J. Clin. Microbiol.* **27,** 2165–2168.

Kitada, K. (1997). Host-vector systems in *Candida glabrata*. GenBank/EMBL/DDBJ database accession number U90626.

Lalo, D., Stettler, S., Mariotte, S., Slonimski, P. P. and Thuriaux, P. (1993). Une duplication fossile entre les régions centromériques de deux chromosomes chez la levure. *C. R. Acad. Sci. Paris* **316,** 367–373.

Larson, G. P., Castanotto, D., Rossi, J. J. and Malafa, M. P. (1994). Isolation and functional analysis of a *Kluyveromyces lactis RAP1* homologue. *Gene* **150,** 35–41.

Li, Q., Jarai, G., Yaghmai, B. and Marzluf, G. A. (1993). The *leu-1* gene of *Neurospora crassa*: nucleotide and deduced amino acid sequence comparisons. *Gene* **136,** 301–305.

Maleszka, R. and Clark-Walker, G. D. (1993). Yeasts have a four-fold variation in ribosomal DNA copy number. *Yeast* **9,** 53–58.

Messner, R., Prillinger, H., Ibl, M. and Himmler, G. (1995). Sequences of ribosomal genes and internal transcribed spacers move three plant parasitic fungi, *Eremothecium ashbyi*, *Ashbya gossypii*, and *Nematospora coryli*, towards *Saccharomyces cerevisiae*. *J. Gen. Appl. Microbiol.* **41,** 31–42.

Mewes, H. W., Albermann, K., Bähr, M., *et al*. (1997). Overview of the yeast genome. *Nature* **387**(Suppl.), 7–65.

Miyasaki, S. H., White, T. C. and Agabian, N. (1994). A fourth secreted aspartyl proteinase gene (*SAP4*) and a *CARE2* repetitive element are located upstream of the *SAP1* gene in *Candida albicans*. *J. Bacteriol.* **176,** 1702–1710.

Moore, G., Devos, K. M., Wang, Z. and Gale, M. D. (1995). Cereal genome evolution: Grasses, line up and form a circle. *Curr. Biol.* **5,** 737–739.

Morizot, D. C. (1990). Use of fish gene maps to predict ancestral vertebrate genome organization. In Ogita, Z.-I. and Markert, C. L. (Eds), *Isozymes: Structure, Function, and Use in Biology and Medicine*. Wiley-Liss, New York, pp. 207–234.

Mulder, W., Scholten, I., van Roon, H. and Grivell, L. (1994). Isolation and characterisation of the linked genes *APA2* and *QCR7*, coding for Ap4A phosphorylase II and the 14 kDa subunit VII of the mitochondrial bc1-complex in the yeast *Kluyveromyces lactis*. *Biochim. Biophys. Acta* **1219,** 719–723.

Mushegian, A. R. and Koonin, E. V. (1996). Gene order is not conserved in bacterial evolution. *Trends Genet.* **12,** 289–290.

Naumov, G. I., Naumova, E. S. and Korhola, M. (1995). Karyotypic relationships among species of *Saccharomyces* sensu lato: *S. castellii, S. dairensis, S. unisporus* and *S. servazzii*. *Syst. Appl. Microbiol.* **18,** 103–108.

Naumov, G. I., Naumova, E. S., Lantto, R. A., Louis, E. J. and Korhola, M. (1992). Genetic homology between *Saccharomyces cerevisiae* and its sibling species *S. paradoxus* and *S. bayanus*: electrophoretic karyotypes. *Yeast* **8,** 599–612.

Nuttley, W. M., Szilard, R. K., Smith, J. J., Veenhuis, M. and Rachubinski, R. A. (1995). The *PAH2* gene is required for peroxisome assembly in the methylotrophic yeast *Hansenula polymorpha* and encodes a member of the tetratricopeptide repeat family of proteins. *Gene* **160,** 33–39.

Oda, Y. and Tonomura, K. (1995). Electrophoretic karyotyping of the yeast genus *Torulaspora. Lett. Appl. Microbiol.* **21,** 190–193.

Ohi, H., Ohtani, W., Okazaki, N., Furuhata, N. and Ohmura, T. (1996). Cloning and characterization of the *Pichia pastoris PRC1* gene encoding carboxypeptidase Y. *Yeast* **12,** 31–40.

Ohno, S. (1970). *Evolution by Gene Duplication*. George Allen and Unwin, London.

Paterson, A. H., Lan, T. H., Reischmann, K. P., *et al*. (1996). Toward a unified genetic map of higher plants, transcending the monocot-dicot divergence. *Nature Genet.* **14,** 380–382.

Philippsen, P., *et al*. (1997). The nucleotide sequence of *Saccharomyces cerevisiae* chromosome XIV and its evolutionary implications. *Nature* **387**(Suppl.), 93–98.

Piredda, S. and Gaillardin, C. (1994). Development of a transformation system for the yeast *Yamadazyma (Pichia) ohmeri. Yeast* **10,** 1601–1612.

Plant, E. P. and Poulter, R. T. M. (1997). Disruption of the *LEU2* gene of *Candida albicans*. GenBank/EMBL/DDBJ database accession number AF000121.

Prillinger, H., Schweigkofler, W., Breitenbach, M., *et al*. (1997). Phytopathogenic filamentous (Ashbya, Eremothecium) and dimorphic fungi (Holleya, Nematospora) with needle-shaped ascospores as new members within the Saccharomycetaceae. *Yeast* **13,** 945–960.

Ryu, S. L., Murooka, Y. and Kaneko, Y. (1996). Genomic reorganization between two sibling yeast species, *Saccharomyces bayanus* and *Saccharomyces cerevisiae. Yeast* **12,** 757–764.

Sakai, Y. and Tani, Y. (1992). Directed mutagenesis in an asporogenous methylotrophic yeast: cloning, sequencing, and one-step gene disruption of the 3-isopropylmalate dehydrogenase gene (*LEU2*) of *Candida boidinii* to derive doubly auxotrophic marker strains. *J. Bacteriol.* **174,** 5988–5993.

Salmeron, J. M., Jr, and Johnston, S. A. (1986). Analysis of the *Kluyveromyces lactis* positive regulatory gene *LAC9* reveals functional homology to, but sequence divergence from, the *Saccharomyces cerevisiae GAL4* gene. *Nucl. Acids Res.* **14,** 7767–7781.

Schenk, G., Layfield, R., Candy, J. M., Duggleby, R. G. and Nixon, P. F. (1997). Molecular evolutionary analysis of the thiamine-diphosphate-dependent enzyme, transketolase. *J. Mol. Evol.* **44,** 552–572.

Sharp, P. M. and Wolfe, K. H. (1993). Identification of a *Candida utilis* gene encoding ribosomal protein L7: evidence for two divergent subclasses of the eukaryotic ribosomal protein L7 family. *Nucl. Acids Res.* **21,** 3577.

Shuster, J. R. (1990). *Kluyveromyces lactis* glyceraldehyde-3-phosphate dehydrogenase and alcohol dehydrogenase-1 genes are linked and divergently transcribed. *Nucl. Acids Res.* **18,** 4271.

Sor, F. and Fukuhara, H. (1989). Analysis of chromosomal DNA patterns of the genus *Kluyveromyces. Yeast* **5,** 1–10.

Stark, M. J. and Milner, J. S. (1989). 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.

Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucl. Acids Res.* **22,** 4673–4680.

Vaughan-Martini, A. and Barcaccia, S. (1996). A reconsideration of species related to *Saccharomyces dairensis* (Naganishi). *Int. J. Syst. Bacteriol.* **46,** 313–317.

Vaughan-Martini, A., Martini, A. and Cardinali, G. (1993). Electrophoretic karyotyping as a taxonomic tool in the genus *Saccharomyces. Antonie van Leeuwenhoek* **63,** 145–156.

Vernis, L., Abbas, A., Chasles, M., *et al*. (1997). An origin of replication and a centromere are both needed to establish a replicative plasmid in the yeast *Yarrowia lipolytica. Mol. Cell. Biol.* **17,** 1995–2004.

Webster, T. D. and Dickson, R. C. (1988). Nucleotide sequence of the galactose gene cluster of *Kluyveromyces lactis. Nucl. Acids Res.* **16,** 8192–8194.

Weinstock, K. G. and Strathern, J. N. (1993). Molecular genetics in *Saccharomyces kluyveri*: the *HIS3* homolog and its use as a selectable marker gene in *S. kluyveri* and *Saccharomyces cerevisiae. Yeast* **9,** 351–361.

Wesolowski-Louvel, M. and Fukuhara, H. (1995). A map of the *Kluyveromyces lactis* genome. *Yeast* **11,** 211–218.

Whelan, W. L. (1987). The genetics of medically important fungi. *Crit. Rev. Microbiol.* **14,** 99–170.

Wolfe, K. H. and Shields, D. C. (1997). Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* **387,** 708–713.

Zenke, F. T., Zachariae, W., Lunkes, A. and Breunig, K. D. (1993). Gal80 proteins of *Kluyveromyces lactis* and *Saccharomyces cerevisiae* are highly conserved but contribute differently to glucose repression of the galactose regulon. *Mol. Cell. Biol.* **13,** 7566–7576.

Zhang, Y. P., Chen, X. J., Li, Y. Y. and Fukuhara, H. (1992). *LEU2* gene homolog in *Kluyveromyces lactis. Yeast* **8,** 801–804.