Yeast genome evolution in the post-genome era

Cathal Seoighe and Kenneth H Wolfe*

The *Saccharomyces cerevisiae* genome sequence, augmented by new data on gene expression and function, continues to yield new findings about eukaryote genome evolution. Analysis of the duplicate gene pairs formed by whole-genome duplication indicates that selection for increased levels of gene expression was a significant factor determining which genes were retained as duplicates and which were returned to a single-copy state, possibly in addition to selection for novel gene functions. Proteome comparisons between worm and yeast show that genes for core metabolic processes are shared among eukaryotes and unchanging in function, while comparisons between different yeast species identify 'orphan' genes as the most rapidly evolving fraction of the proteome. Natural hybridisation among yeast species is frequent, but its long-term evolutionary significance is unknown.

Addresses
Department of Genetics, University of Dublin, Trinity College, Dublin 2, Ireland
*e-mail: khwolfe@tcd.ie

Current Opinion in Microbiology 1999, **2**:548–554

1369-5274/99/$ – see front matter

Abbreviation
YPD Yeast Proteome Database

Introduction

The complete genome sequence of the yeast *Saccharomyces cerevisiae* [1] has provided an invaluable anchor point for all subsequent studies of the genomes of the yeasts and their evolution. Yeast genome evolution is of interest both because *S. cerevisiae* is a model organism with a three-year head start over other eukaryotes in genomics research, and because of the economic value or pathogenic properties of several yeast species. The yeast family hemiascomycetes has the added advantage of containing a broad range of relatively well studied species, including both close and distant relatives of *S. cerevisiae* [2,3••,4,5]. One of the challenges in this post-genomic era is to use the *S. cerevisiae* sequence efficiently to further our knowledge about these other species. A large-scale sequencing project is currently underway for the pathogenic yeast *Candida albicans* [6–8]. By the end of July 1999 this project had already produced 15 Mb of sequence in fewer than 2,000 contigs. Work is currently under way to at least partially sequence the genomes of *Ashbya gossypii* [9] and many other hemiascomycetes [10]. The recent completion of *Caenorhabditis elegans* sequencing project [11] was also of great significance for the study of the yeast genome [12].

This review discusses recent research into molecular evolution of the *S. cerevisiae* genome, with emphasis on genome duplication and the evolution of gene order along chromosomes.

Genome duplication and gene order evolution

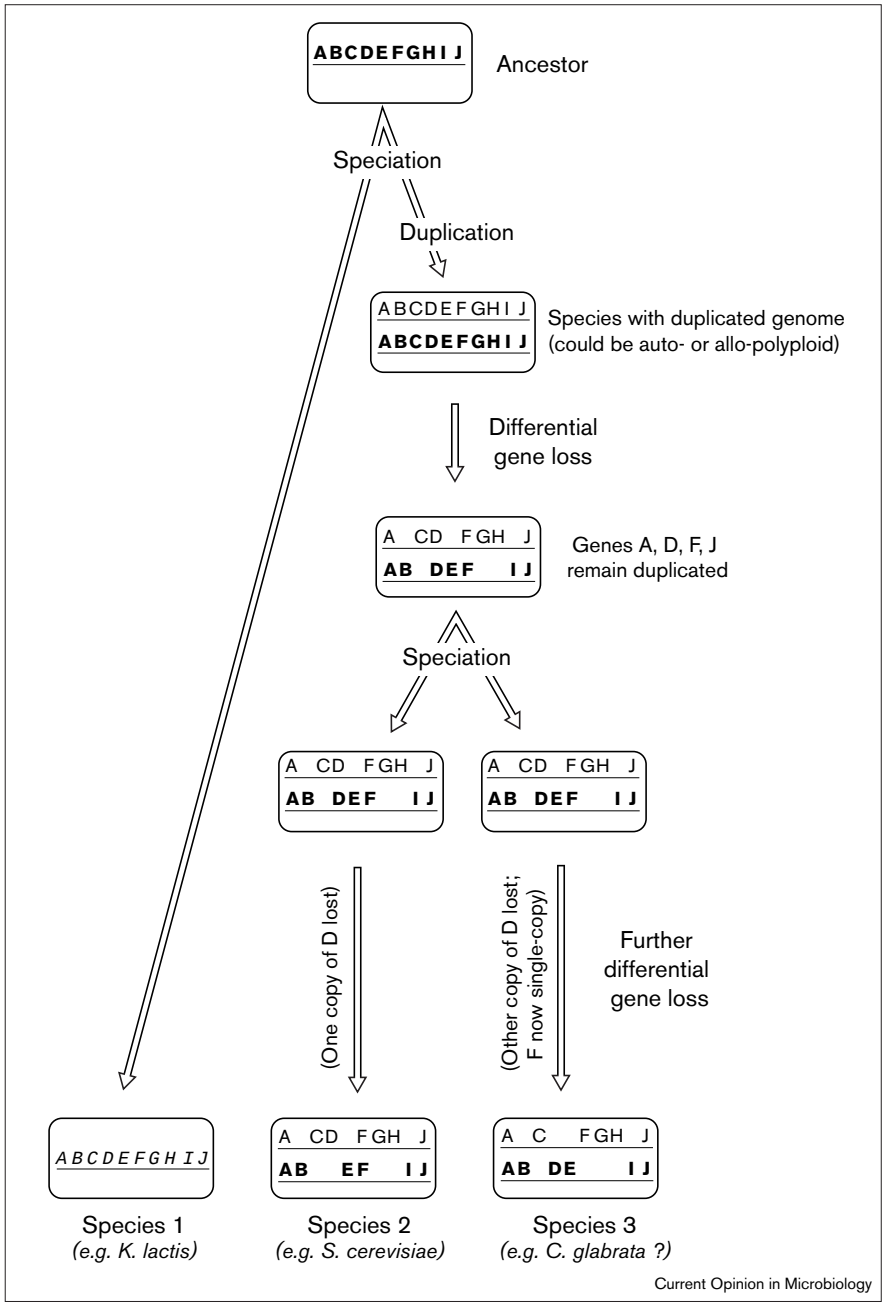
Complete or partial duplication of the genome may have had a profound impact on many organisms' evolution [13]. In addition to altering the karyotype and increasing the number of genes, duplication brings about a reorganisation of local gene order through differential gene loss [14] and may also increase the likelihood that large-scale chromosomal rearrangements will be fixed [15,16]. The large duplicated chromosomal regions found in the genome of *S. cerevisiae* [17–19], as well as limited gene order information from related species [20], suggest that *S. cerevisiae* is a degenerate tetraploid that underwent genome duplication approximately 10^8 years ago [17]. It is not currently possible to test whether this genome duplication occurred within a single species (autotetraploidy) or was the result of hybridisation between two closely related species (allotetraploidy).

S. cerevisiae has a large number of reasonably close relatives (in genera such as *Kluyveromyces*, *Zygosaccharomyces* and *Torulaspora*, as well as *Saccharomyces*) but their phylogenetic relationships are poorly resolved by ribosomal RNA sequence analyses [3••,4,5,20]. The genome duplication event occurred somewhere within this phylogenetic bush of lineages. Trees drawn from protein sequences show that *C. albicans* and *Kluyveromyces lactis* diverged from the *S. cerevisiae* lineage before genome duplication occurred, but the *Saccharomyces sensu stricto* species (*S. bayanus*, *S. pastorianus*, *S. paradoxus* and *S. cerevisiae*) diverged after it and so share the duplicated genes [17,20]. The placement of *K. lactis* outside the genome-duplicated clade is supported by extensive data showing that the relationship of its gene order to that in *S. cerevisiae* is in line with what is predicted on the basis of genome duplication (Figure 1; [20,21••,22]). The fragmentary evidence currently available from *Candida glabrata* suggests that it diverged from *S. cerevisiae* after the genome duplication occurred, but before the sorting-out process of differential gene loss was complete (Figure 1). If so, the relationship between gene orders (and copy numbers) in these two species may be quite complex. Recent analysis of the available *C. albicans* genome sequence data indicates that gene order comparisons between distantly related species are greatly affected by a high frequency of small-scale inversions involving single genes or a small number of contiguous genes (C Seoighe, KH Wolfe, unpublished data).

Reconstructing the ancestral genome

Gene deletion and chromosomal rearrangement (by reciprocal translocation) after genome duplication are proposed to have given rise to the mosaic of partly identical chromosomal regions (duplicated blocks) in the genome of *S. cerevisiae* that

Figure 1



Schematic representation of gene order evolution following genome duplication (based on Keogh *et al.* [20]). Letters A–J represent genes, with different fonts used to distinguish between homologues. For simplicity, no interchromosomal translocations have been shown. Species 1 diverged from species 2 and 3 before genome duplication occurred. Species 2 and 3 are descended from the same genome duplication event, and have some shared gene losses and some species-specific losses. We hypothesise that the genome of *C. glabrata* is descended from the same genome duplication event as *S. cerevisiae* based on the following observations: *C. glabrata* has 14 chromosomes [47] and two rDNA arrays [48,49]; for two gene pairs in *S. cerevisiae* (*DED1*/*DBP1* and *PDR5*/*PDR15*), *C. glabrata* contains a sequence significantly more closely related to one member of the pair than to the other, as expected if speciation occurred after genome duplication; gene order is generally identical in the two species [49–53], but one example of apparent differential gene loss within a duplicated block has been identified (K Haynes, personal communication).

was described by our laboratory [17] and recently updated [22]. The pattern of blocks was assessed to see whether their arrangement in the ancestral genome prior to duplication could be inferred using the principle of maximum parsimony [23]. The result was that there are many alternative and equally parsimonious series of reciprocal translocations, suggesting many indistinguishable alternatives for the order of the blocks before genome duplication. Furthermore, many regions in the *S. cerevisiae* genome cannot currently be paired up because they lack a sufficient number of genes that have been retained in duplicate since genome duplication. This missing information means that the minimum number of reciprocal translocation steps required to rearrange the current map of duplicated blocks into a symmetrical pattern (i.e. to reconstruct the order of blocks in the pre-duplication genome) is less than the actual number of translocations that took place after duplication (estimated to be approximately 70–100 [23]). The problem of estimating the minimal number of steps has been tackled both computationally and analytically [23,24•]. Although an interesting problem, it is, perhaps, biologically uninformative.

By Southern hybridisation of probes to pulsed-field gels, Ryu *et al.* [25] found that chromosomes II and IV of

Table 1

Fraction of *S. cerevisiae* proteins in different functional categories that have been retained in duplicate since genome duplication*.

Protein category	Number of proteins in category	Percent retained in duplicate	χ^2†
All proteins	5792	12.9	
Essential proteins	731	2.7	59
Nonessential proteins	2255	16.6	24
YPD functional categories			
Cyclins	22	54.5	30
Protein phosphatases	40	32.5	12
Heat-shock proteins	32	31.3	8
Protein kinases	123	29.3	26
GTPase-activating proteins	19	26.3	
Glucose metabolism	223	26.0	30
Guanine nucleotide exchange factors	23	21.7	
GTPases	55	18.2	
Amino acid metabolism	189	12.7	
Transcription factors	261	12.3	
tRNA synthetases	42	11.9	
ABC cassette proteins	30	10.0	
Proteases (non-proteasomal)	72	9.7	
Ubiquitin-conjugating proteins	24	8.3	
Proteasome subunits	34	2.9	
Serine-rich proteins	10	0.0	
AAA ATPase domain proteins	16	0.0	
Ribosomal proteins	209	39.2	112
Mitochondrial ribosomal proteins	44	0.0	6
Cytosolic ribosomal proteins	165	50.3	179

*Adapted from [30]. †χ^2 values with one degree of freedom are shown if significant at the 5% level.

S. cerevisiae were rearranged in its sibling *S. bayanus*, and proposed that chromosomal rearrangement could have caused their speciation. Further mapping identified the rearrangement as a reciprocal translocation between the two copies of the gene *RPL2*, which are part of a large duplicated block on these chromosomes. Phylogenetic analysis of the noncoding sequences upstream and downstream of *RPL2A* and *RPL2B* in the two species demonstrated elegantly that a recombination had occurred within the gene [16]. The authors suggested that duplicated loci are frequently involved in chromosomal rearrangements. Reciprocal translocations between duplicate loci do not disrupt the map of duplicated blocks and, consequently, are not included in our estimate of the number of reciprocal translocations since duplication [23].

Tracking proteome evolution and finding families for orphans

In a report coinciding with the publication of the *C. elegans* genome sequence, Chervitz *et al.* [26] compared the complete predicted sets of proteins in *C. elegans* and *S. cerevisiae*. In general, proteins carrying out core biological processes were found to be conserved in sequence and function, whereas proteins that are substantially different in the two species are associated with organism-specific functions or pathways [26]. This type of comparative proteomics approach will become increasingly important as the sequences of other fungal genomes reach completion. Ultimately, the differences in biology among species must be explicable in terms of differences in their genomes.

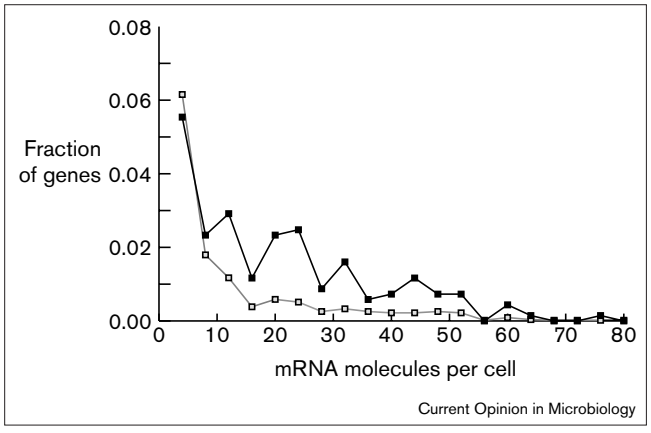
Ozier-Kalogeropoulos *et al.* [21••] sequenced 658 random genomic tags from *K. lactis* and identified 296 genes, more than tripling the number of identified *K. lactis* genes. Genes were identified by sequence similarity, with the majority (all but five of the 296) having homologues in *S. cerevisiae*. Of particular interest was the discovery of *K. lactis* orthologues for 59 so-called orphan genes from *S. cerevisiae*. Orphans are genes having no previously identified orthologues or paralogues, in any species [27]. The results of Ozier-Kalogeropoulos *et al.* [21••] show that these genes are yeast-specific rather than truly unique. Interestingly, the sequence similarity between *K. lactis* and *S. cerevisiae* was lower for orphans than for non-orphans (i.e. genes having homologues in other species). This result also held for comparisons between *C. albicans* and *S. cerevisiae* [21••]. It is not clear whether these genes appear as orphans precisely because they are more rapidly evolving, or whether taxon-specific genes evolve more rapidly. Ozier-Kalogeropoulos *et al.* [21••] measured similarity in terms of BLASTX score, but this question deserves more detailed study using complete gene sequences and molecular evolutionary methods.

Why keep duplicated genes?

After gene duplication one member of a gene pair may accumulate deleterious mutations and be lost, or both copies of the gene may be retained. There are two likely evolutionary reasons for retaining both copies: selection for increased levels of expression, or divergence of gene function. Functional divergence can be produced through complementary degeneration [28], where each daughter gene retains only a subset of the functions of the parent, or (perhaps more rarely) if one daughter acquires a new function. Degenerative tetraploidy provides an opportunity to study the evolution of many duplicated pairs of genes, which were all formed simultaneously.

We estimate that approximately 8% of the genes in the pre-duplication *Saccharomyces* genome were retained in duplicate [23], so that duplicate pairs formed by polyploidy account for approximately 16% of the current *S. cerevisiae* gene set. We have identified 12.9% of *S. cerevisiae*'s genes as polyploidy-derived duplicates [22], so most of the pairs formed by this event have already been found. The remainder lie in regions of the genome that were heavily fragmented by rearrangements. Compared to the average for the genome (12.9%), genes classified as essential are significantly under-duplicated, and nonessential genes are significantly over-duplicated (2.7% and 16.6%, respectively; Table 1). This illustrates the

Figure 2



Over-representation of highly expressed genes among duplicates. The fraction of genes having a particular expression level is shown, considering all genes in the genome (grey line) or just those genes that have been retained in duplicate after genome duplication (black line). Expression data is from Holstege *et al.* [32••] and was grouped into bins of genes expressed at 0–4, 4–8, 8–12 (etc.) mRNA molecules per cell. The large bin representing the lowest expression category (0–4 mRNA molecules per cell) is not shown and contained 87% of all genes and 76% of the duplicated genes.

Table 2

Gene pairs differing by the existence of an intron.

Gene containing intron	Gene lacking intron
SNC1	*SNC2*
ECM33	*YDR055W*
YHR097C	*YDR348C*
MRK1	*MDS1*
RPL22A	*YFL034C-B*
COX5B	*COX5A*
RPL36A	*YPL249C-A*
YML056C	*YLR432W*
RPL21B	*RPL21A*
*ANC1**	*SAS5*
SEC14†	*YKL091C*

*An outgroup is not available for this pair. †This pair is probably not a part of the genome duplication (see text).

apparent genetic redundancy of many duplicated genes, although Thatcher *et al.* [29] reported that yeast genes that were previously classified as nonessential may in fact make a small contribution to evolutionary fitness. It is apparent from the functional classifications of duplicated proteins and from the excess of duplicated genes classified as nonessential (Table 1) that genes that were retained in duplicate continue to perform closely related functions. Of the 280 duplicated pairs for which the Yeast Proteome Database (YPD) [30] lists a functional category for both proteins, the categories are different for only 26 pairs and most of these differences do not appear significant when examined more closely. Taken together these observations suggest that in many cases duplicated genes were retained to improve the efficiency with which existing functions were carried out.

The genes that have been retained in duplicate in *S. cerevisiae* are also not distributed evenly among YPD functional categories (Table 1), indicating some non-randomness or predetermination of the fates of duplicated genes. Some functional categories are over-duplicated, including cyclins and much of the signal transduction apparatus (protein kinases and phosphatases, GTPases, GTPase-activating proteins and guanine nucleotide exchange factors, but not transcription factors). Many cytosolic ribosomal protein genes, but no mitochondrial ribosomal protein genes, are duplicated. For cytosolic ribosomal proteins, 50.3% of the genes are mapped to duplicated chromosomal regions and can be attributed to genome duplication. Many of the remainder are also duplicated [31] but do not form part of larger paired regions.

Many of the over-duplicated functional categories (Table 1) include very highly expressed genes, such as heat shock, glucose metabolism, and cytosolic ribosomal proteins. The correlation between the expression level of a gene and its likelihood of being retained after whole-genome duplication was explored further using whole-genome transcription data from Holstege *et al.* [32••]. The tendency to retain high-expression genes in duplicate (or alternatively, the tendency for each copy of a duplicate gene to be highly expressed) is not confined to the highest categories of gene expression but extends down to expression levels of about 10 mRNA molecules per cell (Figure 2). Thus it appears that increased gene expression (and consequent rapid growth) was a significant concern in the sorting-out of which genes were retained and which were lost. It must, however, be noted from Figure 2 that a majority of duplicated genes have expression levels below 10 molecules per cell, and that selection for diversification of gene function may have been important for these genes. It will be of interest to see whether the criteria for sorting-out were the same in other lineages such as *C. glabrata* (Figure 1).

Intron losses

The set of gene pairs retained in duplicate includes 49 pairs in which at least one gene contains an intron. Of these, 11 pairs are missing the intron in one copy (Table 2). By comparison to the available nucleotide sequences from *C. albicans*, we conclude that in almost all cases the intron was present in the ancestral gene, so that one intron was lost in *S. cerevisiae* after the genome duplication. The single exception to this is the gene pair *SEC14*/*YKL091C*. In this case the intron, present in *S. cerevisiae SEC14* [33], is missing from *YKL091C* and all the available orthologues of *SEC14* (*K. lactis*, *C. glabrata*, *C. albicans*). Further analysis suggests that this pair of genes should not have been attributed to the genome duplication, despite their similar sequences and paired genomic locations [17], because in phylogenetic trees the hemiascomycete *SEC14* sequences cluster together with *YKL091C* as an outgroup. Nonetheless, the intron must

have been gained in *S. cerevisiae SEC14* after its divergence from *C. glabrata* and other ascomycetes. The above data provide an idea of the rate at which introns are lost in *S. cerevisiae* (10 introns lost out of 96 in ~10^8 years, ignoring possible parallel loss).

Isochores laid to rest?

Chromosome III of *S. cerevisiae*, the first chromosome to be sequenced, showed two striking peaks of G + C content [34]. This variation in G + C content was particularly evident in the third coding position (GC3s) and seemed to be analogous to the isochores that had been described in mammals [35]. As complete sequences emerged for more chromosomes, some appeared to contain similar peaks of GC3s content, whereas others did not contain any significant peaks. A reassessment of the variation of GC3s along all 16 yeast chromosomes [36] revealed clusters of genes with similar GC3s content on most of the chromosomes. However, no periodic GC3s content variation could be detected in any chromosome, with the possible exception of chromosome III [36]. On a more local scale it was found that the GC3s content of neighbouring genes is weakly correlated on every chromosome. Li *et al.* [37] also examined compositional heterogeneity along yeast chromosomes, but used raw nucleotide data without distinguishing between coding and noncoding DNA, or separating codon positions. In agreement with Bradnam *et al.* [36], this work also indicated some heterogeneity along yeast chromosomes.

Non-linear genome evolution

The *sensu stricto* yeast *S. pastorianus* has long been recognised as a hybrid between *S. cerevisiae* and *S. bayanus* (or an *S. bayanus*-like species such as *S. monacensis*). Further analysis reveals that both *S. cerevisiae*-type and *S. bayanus*-type chromosomes are present in *S. pastorianus* indicating that this species is an allotetraploid [38,39,40••]. Karyotypic analysis continues to reveal natural hybrids among *Saccharomyces* species including, most recently, *Saccharomyces* sp. CID1 (CBS 8614) which appears to contain chromosomes from two parent species and mitochondria from a third [41,42••].

Despite the frequency of natural yeast hybrids, polyploidy appears to have been extremely rare in the history of *S. cerevisiae*. Our analysis of the genome suggests that tetraploidy occurred only once in approximately 10^8 years [17]. This appears to indicate that the probability of allotetraploids such as *S. pastorianus* surviving as a species is small. It is possible that hybrids frequently cause horizontal transfer of small numbers of genes, but that fixation of complete allotetraploidy is a much rarer event.

Conclusions

The availability of the *S. cerevisiae* genome sequence has sparked interest in the organisation and evolution of many other ascomycete genomes. The growth of comparative genomics and proteomics as a research area indicates that we are beginning to recognise that the reductionist approach, which has been a hugely powerful tool in molecular biology, has its limitations. What is of interest now are the differences between organisms in terms of proteomes and physiology; the similarities are taken for granted.

Is yeast a good model for genome evolution in other eukaryotes? Many of the phenomena discussed here also occur in vertebrates, including isochores [35], gene order change [43], genome duplication [44], horizontal gene transfer [45], and taxon-specific genes [46]. But because the quality of information available about the yeast genome far exceeds that for any vertebrate, it remains to be seen whether the findings from yeast have exact parallels in vertebrates.

Acknowledgements

We thank Jure Piskur for discussion and Avril Coghlan for help with gene expression data.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Goffeau A, Aert R, Agostini-Carbone ML, Ahmed A, Aigle M, Alberghina L, Allen E, Alt-Mörbe J, André B, Andrews S *et al.*: **The Yeast Genome Directory.** *Nature* 1997, **387 (suppl)**:5-105.

2. Kurtzman CP, Fell JW (Eds): *The Yeasts, a Taxonomic Study*, edn 4. Amsterdam: Elsevier; 1998.

3. •• Kurtzman CP, Robnett CJ: **Identification and phylogeny of ascomycetous yeasts from analysis of nuclear large subunit (26S) ribosomal DNA partial sequences.** *Antonie Van Leeuwenhoek* 1998, **73**:331-371.

In a *tour de force* of systematics, Kurtzman and Robnett sequenced the variable D1/D2 domain of 26S ribosomal RNA from the type strains of virtually every ascomycete yeast species (about 500 species). Although the sequenced region is rather short (600 nucleotides) and the phylogenetic trees often have poor resolution on key branches, this work will provide a firm molecular footing for the future taxonomy of ascomycetes.

4. Cai J, Roberts IN, Collins MD: **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* 1996, **46**:542-549.

5. James SA, Cai J, Roberts IN, Collins MD: **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* 1997, **47**:453-460.

6. Sequencing of *Candida albicans* at Stanford's DNA Sequencing and Technology Center on World Wide Web URL: http://www-sequence.stanford.edu/group/candida/

7. Sanger Centre *Candida albicans* Pilot Sequencing Project on World Wide Web URL: http://www.sanger.ac.uk/Projects/C_albicans/

8. Tait E, Simon MC, King S, Brown AJ, Gow NA, Shaw DJ: **A *Candida albicans* genome project: cosmid contigs, physical mapping, and gene isolation.** *Fungal Genet Biol* 1997, **21**:308-314.

9. Dietrich FS, Voegeli S, Gaffney T, Mohr C, Rebischung C, Wing R, Choi S, Goff S, Philippsen P: **Gene map of chromosome I of *Ashbya gossypii*.** *Curr Genet* 1999, **35**:233.

10. Genoscope STS Sequencing from 13 Hemiascomycete Yeasts on World Wide Web URL: http://www.genoscope.cns.fr/externe/English/Projets/Projet_AR/AR. html

11. *C. elegans Sequencing Consortium*: **Genome sequence of the nematode *C. elegans*: a platform for investigating biology.** *Science* 1998, **282**:2012-2018.

12. Chervitz SA, Aravind L, Sherlock G, Ball CA, Koonin EV, Dwight SS, Harris MA, Dolinski K, Mohr S, Smith T *et al.*: **Comparison of the complete protein sets of worm and yeast: orthology and divergence.** *Science* 1998, **282**:2022-2028.

13. Ohno S: *Evolution by Gene Duplication*. London: George Allen and Unwin; 1970.

14. Lundin LG: **Evolution of the vertebrate genome as reflected in paralogous chromosomal regions in man and the house mouse.** *Genomics* 1993, **16**:1-19.

15. Ahn S, Tanksley SD: **Comparative linkage maps of the rice and maize genomes.** *Proc Natl Acad Sci* 1993, **90**:7980-7984.

16. Ryu SL, Murooka Y, Kaneko Y: **Reciprocal translocation at duplicated *RPL2* loci might cause speciation of *Saccharomyces bayanus* and *Saccharomyces cerevisiae*.** *Curr Genet* 1998, **33**:345-351.

17. Wolfe KH, Shields DC: **Molecular evidence for an ancient duplication of the entire yeast genome.** *Nature* 1997, **387**:708-713.

18. Mewes HW, Albermann K, Bähr M, Frishman D, Gleissner A, Hani J, Heumann K, Kleine K, Maierl A, Oliver SG *et al.*: **Overview of the yeast genome.** *Nature* 1997, **387 (suppl)**:7-65.

19. Coissac E, Maillier E, Netter P: **A comparative study of duplications in bacteria and eukaryotes: the importance of telomeres.** *Mol Biol Evol* 1997, **14**:1062-1074.

20. Keogh RS, Seoighe C, Wolfe KH: **Evolution of gene order and chromosome number in *Saccharomyces*, *Kluyveromyces* and related fungi.** *Yeast* 1998, **14**:443-457.

21. •• Ozier-Kalogeropoulos O, Malpertuy A, Boyer J, Tekaia F, Dujon B: **Random exploration of the *Kluyveromyces lactis* genome and comparison with that of *Saccharomyces cerevisiae*.** *Nucleic Acids Res* 1998, **26**:5511-5524.

The *K. lactis* genome was surveyed by random sequencing. The most informative data came from sequencing both ends of plasmid clones with 2–3 kb inserts, because this provided information about gene order as well as gene content.

22. Seoighe C, Wolfe KH: **Updated map of duplicated regions in the yeast genome.** *Gene* 1999, in press.

23. Seoighe C, Wolfe KH: **Extent of genomic rearrangement after genome duplication in yeast.** *Proc Natl Acad Sci USA* 1998, **95**:4447-4452.

24. • El-Mabrouk N, Bryant D, Sankoff D: **Reconstructing the pre doubling genome.** In *RECOMB '99: Proceedings of the Third International Conference on Computational Molecular Biology:* 1999, 11–14 April: Lyon, France. Edited by Istrail S, Pevzner P, Waterman M. New York: ACM Press; 1999:154-163.

The authors, in a rigorous treatment of the problem of estimating the 'pre-doubling' organisation of the yeast genome, calculated that an absolute minimum of 38 reciprocal translocations must have occurred to scamble gene order after genome duplication. They also suggested an algorithm for constructing the pre-doubling genome, and proposed one such ancestor. Also of interest here, however, would be a description of the class of equivalent solutions for the pre-doubling genome. Common features of the class of most parsimonious solutions, if they exist, could suggest something about the real pre-doubling genome.

25. Ryu SL, Murooka Y, Kaneko Y: **Genomic reorganization between two sibling yeast species, *Saccharomyces bayanus* and *Saccharomyces cerevisiae*.** *Yeast* 1996, **12**:757-764.

26. Chervitz SA, Aravind L, Sherlock G, Ball CA, Koonin EV, Dwight SS, Harris MA, Dolinski K, Mohr S, Smith T *et al.*: **Comparison of the complete protein sets of worm and yeast: orthology and divergence.** *Science* 1998, **282**:2022-2028.

27. Dujon B: **The yeast genome project: what did we learn?** *Trends Genet* 1996, **12**:263-270.

28. Force A, Lynch M, Pickett FB, Amores A, Yan Y, Postlethwait J: **Preservation of duplicate genes by complementary degenerative mutations.** *Genetics* 1999, **151**:1531-1545.

29. Thatcher JW, Shaw JM, Dickinson WJ: **Marginal fitness contributions of nonessential genes in yeast.** *Proc Natl Acad Sci* 1998, **95**:253-257.

30. Hodges PE, McKee AH, Davis BP, Payne WE, Garrels JI: **The Yeast Proteome Database (YPD): a model for the organization and presentation of genome-wide functional data.** *Nucleic Acids Res* 1999, **27**:69-73.

31. Planta RJ, Mager WH: **The list of cytoplasmic ribosomal proteins of *Saccharomyces cerevisiae*.** *Yeast* 1998, **14**:471-477.

32. •• Holstege FC, Jennings EG, Wyrick JJ, Lee TI, Hengartner CJ, Green MR, Golub TR, Lander ES, Young RA: **Dissecting the regulatory circuitry of a eukaryotic genome.** *Cell* 1998, **95**:717-728.

Although it was not the primary aim of the study, this paper is one of the best examples so far of the power of microarray technology to study gene expression on a global scale. The authors estimated the number of mRNA molecules per cell, and mRNA half-life, for every gene in the *S. cerevisiae* genome under conditions of growth to mid-log phase in YPD media, using Affymetrix high-density oligonucleotide arrays. Included in the wealth of data from this experiment is the finding that under these conditions, the expression levels of cytosolic and mitochondrial ribosomal proteins average 29 and 1.4 mRNA molecules per cell, respectively.

33. Bankaitis VA, Malehorn DE, Emr SD, Greene R: **The *Saccharomyces cerevisiae SEC14* gene encodes a cytosolic factor that is required for transport of secretory proteins from the yeast Golgi complex.** *J Cell Biol* 1989, **108**:1271-1281.

34. Sharp PM, Lloyd AT: **Regional base composition variation along yeast chromosome III: evolution of chromosome primary structure.** *Nucleic Acids Res* 1993, **21**:179-183.

35. Bernardi G, Olofsson B, Filipski J, Zerial M, Salinas J, Cuny G, Meunier-Rotival M, Rodier F: **The mosaic genome of warm-blooded vertebrates.** *Science* 1985, **228**:953-958.

36. Bradnam KR, Seoighe C, Sharp PM, Wolfe KH: **G+C content variation along and among *Saccharomyces cerevisiae* chromosomes.** *Mol Biol Evol* 1999, **16**:666-675.

37. Li W, Stolovitzky G, Bernaola-Galvan P, Oliver JL: **Compositional heterogeneity within, and uniformity between, DNA sequences of yeast chromosomes.** *Genome Res* 1998, **8**:916-928.

38. Kielland-Brandt M, Nielsson-Tillgren T, Gjermansen C, Holmberg S, Pedersen MB: **Genetics of Brewing Yeasts.** In *The Yeasts*. Edited by Wheals AE, Rose AH, Harrison JS. London: Academic Press; 1995:223-254.

39. Tamai Y, Momma T, Yoshimoto H, Kaneko Y: **Co-existence of two types of chromosome in the bottom fermenting yeast *Saccharomyces pastorianus*.** *Yeast* 1998, **14**:923-933.

40. •• Petersen RF, Marinoni G, Nielsen ML, Piskur J: **Molecular approaches for analyzing diversity and phylogeny among yeast species.** In *Dimorphism in Human Pathogenic and Apathogenic Yeasts*. Edited by Ernst JF, Schmidt A. Basel: Karger; 1999.

Viable hybrid yeast strains containing most chromosomes from both parents were constructed between pairs of *sensu stricto* species, although spore viability is low. In hybridisation between *sensu stricto* and *sensu lato* species, the hybrids lose all or most of one set of parental chromosomes during mitosis. Petersen *et al.* suggest that some genes may be retained in two copies either by limited aneuploidy or by integration into the chromosomes of the other parent.

41. Masneuf I, Hansen J, Groth C, Piskur J, Dubourdieu D: **New hybrids between *Saccharomyces* sensu stricto yeast species found among wine and cider production strains.** *Appl Environ Microbiol* 1998, **64**:3887-3892.

42. •• Groth C, Hansen J, Piskur J: **A natural chimeric yeast containing genetic material from three species.** *Int J Syst Bacteriol* 1999, in press.

Sequence analysis of a yeast isolated from cider (CID1) shows that its nuclear genome contains sequences derived from both *S. cerevisiae* and *S. bayanus*. Its mitochondrial DNA is derived from a third species, *Saccharomyces* sp. IFO 1802. Exchange of genetic material between *Saccharomyces* species may be more frequent than previously supposed.

43. Gilley J, Fried M: **Extensive gene order differences within regions of conserved synteny between the *Fugu* and human genomes: implications for chromosomal evolution and the cloning of disease genes.** *Hum Mol Genet* 1999, **8**:1313-1320.

44. Amores A, Force A, Yan YL, Joly L, Amemiya C, Fritz A, Ho RK, Langeland J, Prince V, Wang YL *et al.*: **Zebrafish hox clusters and vertebrate genome evolution.** *Science* 1998, **282**:1711-1714.

45. Kordis D, Gubensek F: **Unusual horizontal transfer of a long interspersed nuclear element between distant vertebrate classes.** *Proc Natl Acad Sci USA* 1998, **95**:10704-10709.

46. Lawn RM, Schwartz K, Patthy L: **Convergent evolution of apolipoprotein(a) in primates and hedgehog.** *Proc Natl Acad Sci USA* 1997, **94**:11992-11997.

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

48. Maleszka R, Clark-Walker GD: **Yeasts have a four-fold variation in ribosomal DNA copy number.** *Yeast* 1993, **9**:53-58.

49. Miyazaki H, Miyazaki Y, Geber A, Parkinson T, Hitchcock C, Falconer DJ, Ward DJ, Marsden K, Bennett JE: **Fluconazole resistance associated with drug efflux and increased transcription of a drug transporter gene, *PDH1*, in *Candida glabrata*.** *Antimicrob Agents Chemother* 1998, **42**:1695-1701.

50. Zhou PB, Thiele DJ: **Isolation of a metal-activated transcription factor gene from *Candida glabrata* by complementation in *Saccharomyces cerevisiae*.** *Proc Natl Acad Sci USA* 1991, **88**:6112-6116.

51. Dundon W, Islam K: **Nucleotide sequence of the gene coding for SEC14p in *Candida (Torulopsis) glabrata*.** *Gene* 1997, **193**:115-118.

52. Nagahashi S, Lussier M, Bussey H: **Isolation of *Candida glabrata* homologs of the *Saccharomyces cerevisiae KRE9* and *KNH1* genes and their involvement in cell wall beta-1,6-glucan synthesis.** *J Bacteriol* 1998, **180**:5020-5029.

53. Cormack BP, Falkow S: **Efficient homologous and illegitimate recombination in the opportunistic yeast pathogen *Candida glabrata*.** *Genetics* 1999, **151**:979-987.