Duplication of genes and genomes in yeasts

Simon Wong and Kenneth H. Wolfe

Abstract

The molecular evolution of the group of yeasts closely related to *Saccharomyces cerevisiae* has been profoundly affected by an ancient polyploidy event that resulted in duplication of the whole genome. This event occurred in the common ancestor of the *Saccharomyces sensu stricto* and *sensu lato* species, including *Candida glabrata*. Recent progress in genome sequencing has allowed the molecular sorting-out process after genome duplication to be investigated in detail. The loci where both copies of the gene were retained, as opposed to deletion of one copy, appear to be those that have either been subject to selection for high dosage of the gene product, or where functional divergence between the two copies was achieved rapidly.

1 The 'true' yeasts

The kingdom Fungi consists of a vast range of eukaryotic organisms found in diverse environments. Most fungi are composed of hyphae – filamentous, threadlike structures often congregating into systems called mycelia. However, some fungi assume unicellular forms known as yeasts. In addition, some species are dimorphic, incorporating both structural forms in their life cycles depending on environmental conditions. Although many fungal species are unicellular, those in the phylum Ascomycota are often referred to as the 'true' yeasts due to their predominantly unicellular life cycles. This phylum comprises three classes: Archiascomycetes (e.g. *Schizosaccharomyces pombe*), Euascomycetes (e.g. *Neurospora crassa*), and Hemiascomycetes (e.g. *Saccharomyces cerevisiae*). This review focuses on the class Hemiascomycetes, which includes some of the most important yeasts for basic, applied and medical research, features that have made Hemiascomycetes the focus of extensive genomics research.

Saccharomyces cerevisiae (bakers' yeast) is the most renowned and best studied yeast. Its natural capability to produce ethanol by fermentation and carbon dioxide by respiration has been exploited for millennia in the brewing and baking industries. Its economic importance has provided much impetus for basic research into this yeast. Beginning with the work of Winge and Lindegren in the 1930s (reviewed in Mortimer 1993a; 1993b), the ability to perform crosses with *S. cerevisiae* and its tractability in the laboratory made it an attractive research tool

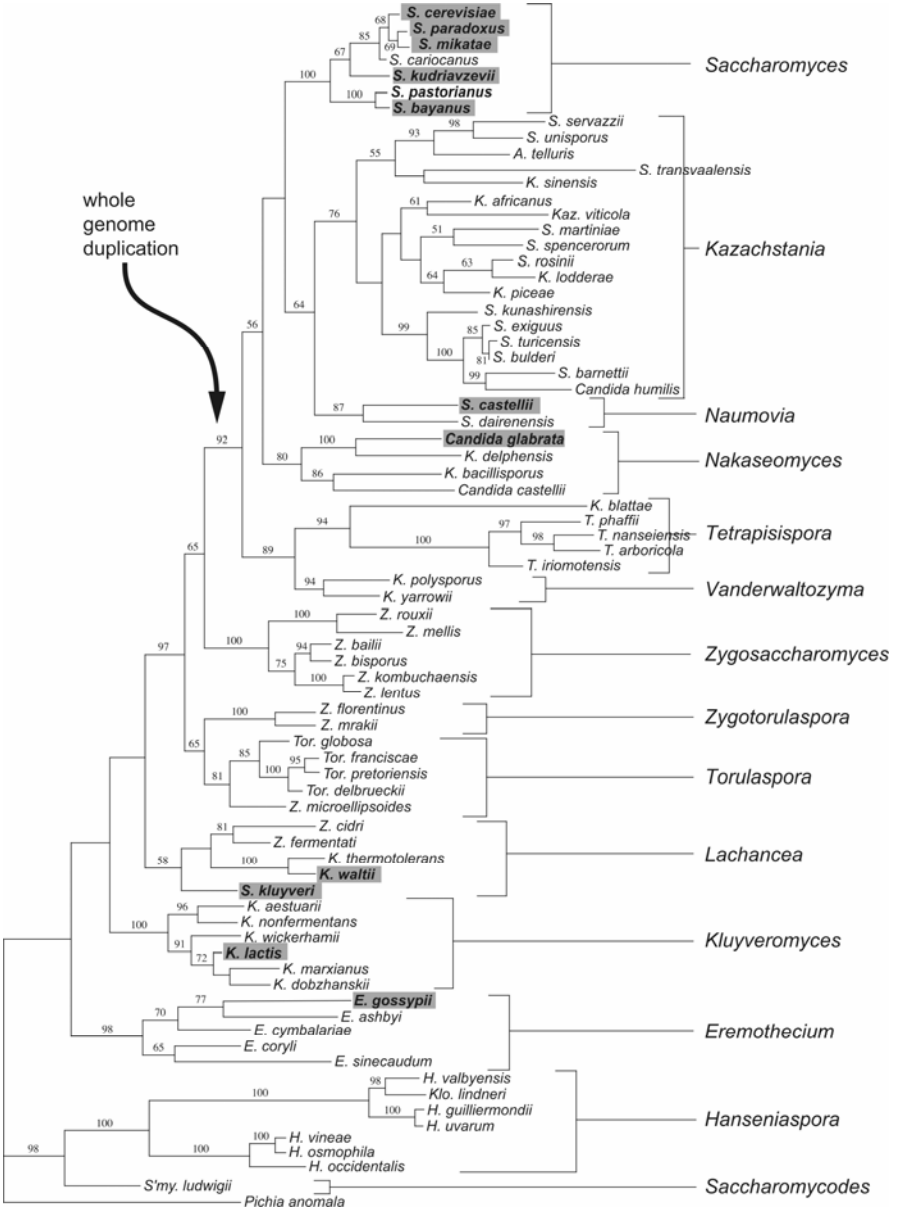
Topics in Current Genetics
P. Sunnerhagen, J. Piškur (Eds.): Comparative Genomics
DOI 10.1007/b105770 / Published online: 7 January 2005

in classical genetics. Combined with modern molecular techniques, it has become one of the best characterized eukaryotic model organisms.

Apart from *S. cerevisiae*, many related yeasts are also widely employed for the production of different wine, beer, and bread (Demain et al. 1998). However, some species can produce other important compounds such as vitamins, citric acid, and lipids. *Candida utilis* is used for the production of animal feed as well as the flavoring substances ethyl acetate and acetaldehyde. Many species of the genera *Candida*, *Debaryomyces*, *Pichia*, and *Yarrowia* can utilize hydrocarbons as sole carbon sources and could potentially be used to clean up oil spills. With recent advances in recombinant DNA technology, a number of yeasts have been developed as host organisms for the production of heterologous protein such as human hormones and enzymes of commercial interest (Gellissen and Hollenberg 1997).

Within the Hemiascomycetes, many species of the genus *Candida* are opportunistic pathogens of humans. They cause a range of diseases and are often associated with immunocompromised patients (Hazen 1995; Calderone 2002). The principal yeast pathogen for human is *Candida albicans*, which is the most common species isolated from bloodstream infections. However, other species such as *C. tropicalis*, *C. dubliniensis*, and *C. glabrata* are emerging concerns as they are less susceptible to some antifungal drugs and their incidence has increased relative to that of *C. albicans*. The medical importance of the *Candida* species has stimulated much research interest and the development of accurate strain detection systems.

Fig. 1 (overleaf). Phylogeny of hemiascomycetes in the '*Saccharomyces* complex', redrawn from Kurtzman (2003). Species whose genomes have been extensively sequenced (> 3x coverage) are highlighted (Goffeau et al. 1996; Cliften et al. 2003; Kellis et al. 2003; Dietrich et al. 2004; Dujon et al. 2004; Kellis et al. 2004). *Eremothecium gossypii* is synonymous with *Ashbya gossypii*. The tree is based on parsimony analysis of six genes. Numbers on internal branches are bootstrap percentages, and branches where no number is shown recurred in <50% of bootstrap replicates. Names on the right are new genus names proposed by Kurtzman (2003). The probable phylogenetic position of the whole-genome duplication (WGD) event is shown by an arrow. The positioning of the WGD after the divergence between *Zygosaccharomyces* and the upper clades is based on the presence of 7 chromosomes in the type strain of *Z. rouxii* (Torok et al. 1993; Sychrova et al. 2000) and the extensive colinearity of gene order between *Z. rouxii* and outgroup species such as *K. waltii* or *K. lactis*, contrasting with its 1:2 relationship to *S. cerevisiae* (Wong et al. 2002; J. Gordon and K. H. Wolfe, unpublished results). The positioning of the WGD before the divergence of *Tetrapisispora* and *Vanderwaltozyma* from the upper clades is less certain and is inferred from the presence of about 20 chromosomes in *K. yarrowii* and about 13 in *K. polysporus* (Belloch et al. 1998).



2 Taxonomy and phylogeny

A consistent and well established phylogenetic relationship is fundamental to infer evolutionary events within a group of species. Traditionally, yeast taxonomy has

been based on phenotypic and metabolic features often prone to ambiguity. This is especially true in closely related species such as the hemiascomycete yeasts. To address this issue, various sequence-based reconstructions of the phylogeny of the ascomycetes have been carried out. Some of the more comprehensive studies involved the use of 18S rRNA sequences (Cai et al. 1996; James et al. 1997; Keogh et al. 1998), partial 26S rRNA sequences (Kurtzman and Robnett 1998), the complete rDNA repeating unit (Wong et al. 2003), and the mitochondrial cytochrome-c oxidase II (*cox2*) gene (Belloch et al. 2000). Although the phylogenetic trees obtained from these studies are reasonably congruent, significant progress was achieved by a recent phylogenetic analysis that combined sequence data from multiple loci and included almost all the known species in the group of Hemiascomycetes called the 'Saccharomyces complex' (Kurtzman and Robnett 2003). Their analysis placed 75 species associated with the genera *Saccharomyces* and *Kluyveromyces* into 14 clades. Kurtzman (2003) subsequently used this phylogeny as the basis for proposing a reorganization of the taxonomy of this group of species (Fig. 1).

Species in the current genus *Saccharomyces* can be divided into three groups (Fig. 1). The *Saccharomyces sensu stricto* species, including *S. cerevisiae*, are in a homogeneous group. The phylogenetic relationships within the *sensu stricto* were recently re-examined using a 106-gene data set from whole genome sequences, which produced a tree slightly different from that in Figure 1 (*S. paradoxus* clustered with *S. cerevisiae* instead of with *S. mikatae*; Rokas et al. (2003); see also Phillips et al. (2004) and Holland et al. (2004)).

The *sensu stricto* group is phylogenetically distinct from the *Saccharomyces sensu lato* species, which form a heterogeneous group that is not monophyletic with respect to other species in the genera *Kluyveromyces* and *Candida*. Many of the *Saccharomyces sensu lato* species have been placed into a new genus, *Kazachstania*, by Kurtzman (2003) (Fig. 1). The third *Saccharomyces* group consists of just one species, *S. kluyveri*, which is phylogenetically distant from the *sensu stricto* and *sensu lato* groups. Physiologically, *S. kluyveri* is quite dissimilar to the other *Saccharomyces* yeasts. Most notably, it is unable to form true petite mutants (Moller et al. 2001). It can utilize pyrimidines and purines as sole sources of nitrogen (Gojkovic et al. 1998). Cytogenetic analysis have revealed that its karyotype reveals a lack of small chromosomes and it contains roughly half the number of chromosomes found in other *Saccharomyces* species (Petersen et al. 1999). The latter, along with subsequent studies, have established that it diverged from other yeasts before the whole genome duplication event leading to other *Saccharomyces* lineages (Wolfe and Shields 1997; Wong et al. 2002). Taken together, the placement of *S. kluyveri* in the genus *Saccharomyces* seems questionable and Kurtzman (2003) placed it in a new genus, *Lachancea*.

Kurtzman and Robnett (2003) highlighted the disparity between the well-defined phylogenetic clades in their analysis and the way in which species were grouped into genera under current systematic treatments (summarized in Kurtzman and Fell 1998). Many well-known genera, including *Saccharomyces*, *Kluyveromyces*, *Zygosaccharomyces*, and *Candida* were polyphyletic. This led Kurtzman (2003) to propose sweeping revisions to the taxonomy of this group of

species, whereby the 14 well-supported clades in the multi-gene phylogenetic tree (Kurtzman and Robnett 2003) became 14 genera that are probably monophyletic. Rapid progress in genomics has resulted in almost-complete genome sequences becoming available from representatives of six of these 14 clades (Fig. 1). However, even though each of the 14 clades seems reasonably robust, there is still some doubt about the branching order of these clades relative to one other. Many of the internal branches along the 'spine' of the tree in Figure 1 have low bootstrap confidence, and analyses based on complete genome sequences tend to arrange *S. kluyveri*, *K. waltii*, *K. lactis*, and *E. gossypii* into one or two monophyletic groups as opposed to the three separate lineages represented by *Lachancea*, *Kluyveromyces*, and *Eremothecium* in Figure 1 (Hittinger et al. 2004; J. Gordon, D. Scannell, K. Byrne, S. Wong and K. H. Wolfe, unpublished results).

3 Yeast genome sequencing projects

In 1995, *Haemophilus influenzae* became the first free-living organism to have its genome completely sequenced (Fleischmann et al. 1995). While the genome of this bacterium is only around 1.8 Mb in size, it heralded the genomics era where the full complement of genes of an organism can be systematically identified and analyzed. At the same time, researchers worldwide were busy sequencing the appreciably larger genomes of various eukaryotes. In 1996, *S. cerevisiae* became the first eukaryotic genome to be completely sequenced (Goffeau et al. 1996). The sequence was determined by a large consortium of laboratories over several years, beginning with chromosome III in 1992 (Oliver et al. 1992). It consists of 16 chromosomes that add up to approximately 14 Mb, much smaller than those of other model eukaryotes such as *Arabidopsis thaliana* (125 Mb) or *Drosophila melanogaster* (137 Mb). Yet, it still poses a substantial challenge for researchers trying to decipher its contents. The number of protein-coding genes was originally estimated to be in the region of 6,200 but has since been modified to a more conservative 5,500 to 5,700 (Wood et al. 2001; Kellis et al. 2003). Unlike multicellular organisms, the genome of *S. cerevisiae* is very compact with around 70% of the total sequence coding for genes. In addition, only around 4% of genes contain introns, greatly assisting the annotation process. Repetitive elements in the genome are represented by the yeast retrotransposons, the Ty elements, which occur in about 50 copies often associated with tRNA genes (Hani and Feldmann 1998). But these make up a relatively small proportion of the genome compared with multicellular eukaryotes such as human, where over 50% of the genome can be classified as repetitive DNA (Baltimore 2001). Hence, the compact nature of the *S. cerevisiae* genome permits useful comparative genomics studies to be carried out using relatively small amounts of sequence data from other similar yeasts.

Subsequently, a number of fungal genomes have been completely sequenced. They include the genomes of *Schizosaccharomyces pombe* (in class Archiascomycetes; Wood et al. 2002) and *Neurospora crassa* (in class Euascomycetes; Galagan et al. 2003). These species are so distantly related to *S. cerevisiae* that, although

some interesting comparisons can be made in terms of their proteome contents and organism-specific biology, there is almost no conservation of gene order along chromosomes between these species and *S. cerevisiae* and many genes do not fall into simple one-to-one orthology relationships between these genomes.

In 2003, extensive genome sequence data became available for several other *Saccharomyces sensu stricto* species. *S. mikatae* and *S. bayanus* were independently sequenced to 7x coverage by Kellis et al. (2003) and to 3x coverage by Cliften et al. (2003); Kellis et al. also sequenced *S. paradoxus* (7x), and Cliften et al. also sequenced *S. kudriavzevii* (3x). The close relationship between these yeasts and *S. cerevisiae* means that their genomes are almost identical in organization, with few chromosomal rearrangements disrupting syntenic regions. However, it proved to be extremely useful in the identification of rapidly evolving regulatory elements.

Comprehensive sequence information has also become available in the past year from more distantly related hemiascomycetes, as summarized in Figure 1 (Cliften et al. 2003; Dietrich et al. 2004; Dujon et al. 2004; Kellis et al. 2004). The result is that we now have sequences from two species in (or close to) the *sensu lato* group (*S. castellii* and *Candida glabrata*), and four species that are somewhat more distantly related (*K. waltii*, *S. kluyveri*, *K. lactis*, and *E. gossypii*). In addition to the species highlighted in Figure 1, which covers only the '*Saccharomyces* complex' (Kurtzman and Robnett 2003), the genome sequences of several other more distantly related hemiascomycetes are known: the genome sequences of *Candida albicans*, *Debaryomyces hansenii*, and *Yarrowia lipolytica* sequences are public (Dujon et al. 2004; Jones et al. 2004), and the *Pichia angusta (Hansenula polymorpha*) sequence is available under restricted terms (Ramezani-Rad et al. 2003). Lastly, more limited amounts of random sequence information from the genomes of several other yeasts were produced by the Génolevures project (Souciet et al. 2000).

4 The origin of new genes

There are four possible ways for a new gene to emerge during evolution: (i) duplication of an existing gene, (ii) combination of parts of different genes to create a mosaic gene, (iii) *de novo* generation of a gene from non-coding DNA, and (iv) horizontal transfer of a gene from another species. While examples of all four routes have been documented (reviewed in Wolfe and Li 2003), by far the most common way to create new genes is by gene duplication.

In his classic book *Evolution by Gene Duplication*, Ohno (1970) proposed that biodiversity evolved in big leaps by the creation – through duplication – of novel, redundant genetic raw material. Some three decades later, this mechanism of genome evolution is universally accepted. In Ohno's view, the original copy of a gene retained the original function (a sort of backup mechanism) while the extra copy was free to vary in sequence. Under this hypothesis, a newly formed copy of a gene faces one of two possible alternative outcomes: either it is lost from the ge-

nome due to the accumulation of deleterious mutations (nonfunctionalization), or else it is preserved in the genome by virtue of acquiring a novel role that is selectively advantageous (neofunctionalization). Since deleterious mutations occur more frequently than beneficial ones, it was expected that most new gene duplicates would quickly pick up an inactivating mutation that would turn them into pseudogenes, eventually becoming deleted from the genome, but occasionally the extra copy of a gene would survive because it acquired a sequence change that conferred a beneficial new function. Ohno's model predicts that, in cases where a duplicated gene has survived, the rate of sequence change in the new copy of the gene will have been faster than in the original copy. The problem, of course, is that it is usually not possible to know which member of a pair of paralogous sequences is the 'original' gene and which is the 'copy'. In fact, the distinction is meaningless for some types of duplication (e.g. polyploidy) and only makes sense in some very specific circumstances where it is possible to tell which copy is derived from which (e.g. in the case of retrotransposed mammalian genes that have lost introns).

In the decades following Ohno's work, it has become apparent that all genomes contain many large gene families, which indicates that gene duplication has been a major force in organismal evolution. However, the ubiquity of gene duplication has led to a problem: there have been so many duplications that it is hard to see how they can all have involved the gain of novel gene functions. A solution to this problem was proposed by Lynch and Force, who suggested that subfunctionalization could provide a mechanism of gene preservation in the immediate aftermath of gene duplication (Force et al. 1999; Lynch and Force 2000). Subfunctionalization is a process whereby a gene with multiple functions (e.g. a gene whose expression is induced under several distinct conditions) becomes duplicated, and random inactivation of some of the functions in each of the daughter copies results in selection against loss of either of the daughters from the genome. Subfunctionalization is, thus, a mechanism whereby mutations that are not adaptive (i.e. most mutations) can lead to the preservation of both copies of a duplicated gene because the daughters both perform subsets of the parent's suite of functions. Later on, it is possible (but not essential) that further mutations could result in the gain of a new function (neofunctionalization) by one of the daughters.

For yeast, however, subfunctionalization is not expected to be an effective mechanism of duplicate gene preservation (Lynch and Force 2000). This is because the population sizes of yeast species are very large. For subfunctionalization to happen, loss-of-subfunction alleles must become fixed by genetic drift at the daughter loci. But when the effective population size exceeds $10^6 – 10^7$ individuals, which is almost certainly true for yeast species, the very long time required for a neutral loss-of-subfunction allele to drift to fixation in the population means that the allele is very likely to acquire a second, inactivating, mutation in transit before it can become fixed (Lynch and Force 2000). This means that loss-of-subfunction alleles will not drift to fixation, so subfunctionalization will not occur in yeast.

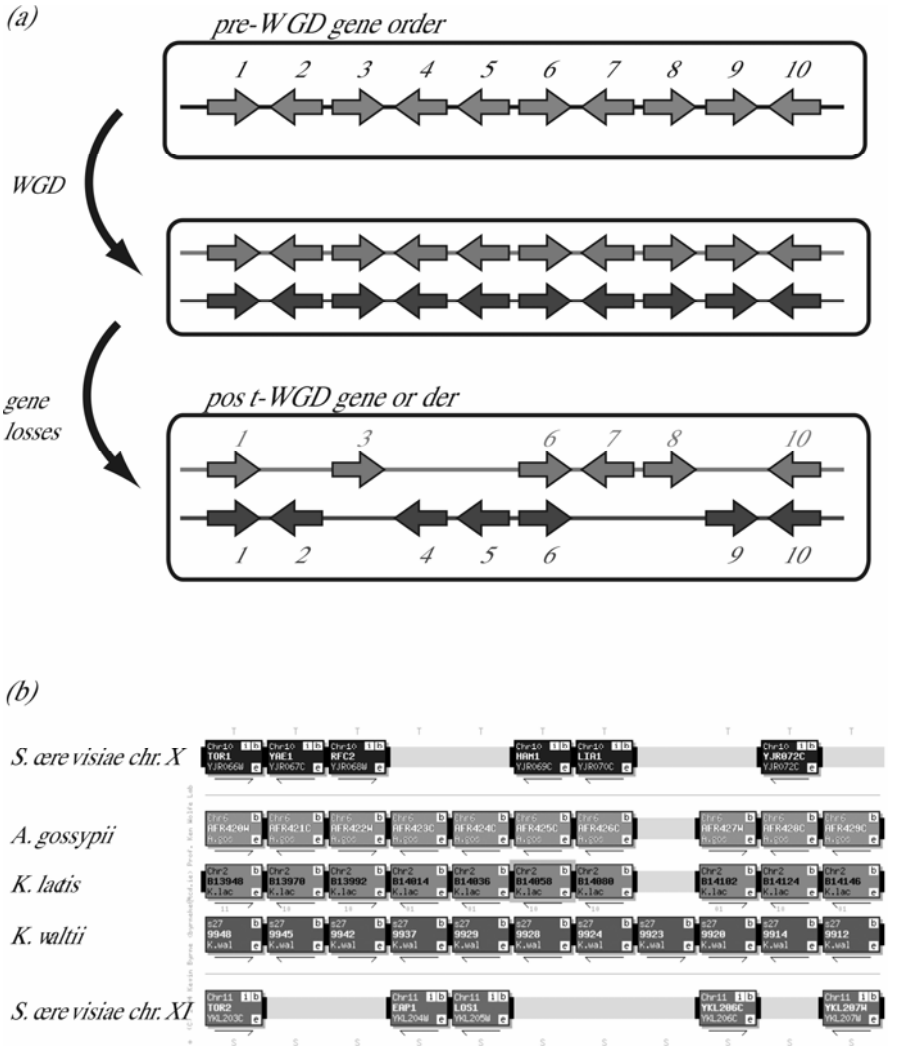
So, have most of the duplicated genes in the yeast genome therefore been retained because they have novel functions? Before tackling this question in Section

8, we will review one of the sources of duplicated genes in the *Saccharomyces* complex of species.

5 Whole genome duplication

One of the most dramatic ways to increase the gene repertoire of an organism involves the duplication of the entire genome (polyploidization). Genomic data has provided evidence of ancient polyploidization events in many species that are now genetically diploid – a situation referred to as paleopolyploidy. Paleopolyploid species include plants such as *Arabidopsis* and the cereals (Blanc et al. 2003; Paterson et al. 2004), ray-finned fishes such as the zebrafish and *Fugu* (Taylor et al. 2001; Vandepoele et al. 2004), tetrapods such as frogs of the genus *Xenopus* (Hughes and Hughes 1993), and a large clade of yeasts in the *Saccharomyces* complex (Fig. 1).

Fig. 2 (overleaf). (a). Illustration of our model of gene order evolution following whole-genome duplication (WGD). The box at the top shows a hypothetical region of chromosome containing ten genes numbered 1–10. After WGD, the whole region is briefly present in two copies. However, many genes subsequently return to single-copy state because there is no evolutionary advantage to maintaining both copies. In this example, only genes 1, 6 and 10 remain duplicated. However, the arrangement of these three homolog pairs in the post-WGD species (bottom) would be sufficient to allow the sister regions to be detected using that genome sequence alone. Also, the order of genes in sister regions in post-WGD species have well-defined relationships to the gene order that existed in the pre-WGD genome (top), which will also be similar to the gene order seen in any species that diverged from the WGD lineage before the WGD occurred. Based on Keogh et al. (1998). Figure 2 (b). An example of gene order relationships between parts of two sister regions in *S. cerevisiae* (from chromosomes X and XI), and the homologous single chromosome regions from *A. gossypii*, *K. lactis* and *K. waltii*. In this representation, each rectangle represents a gene and homologs are arranged as vertical columns. Arrows below the rectangles show transcriptional orientation. Gray lines connect adjacent genes but do not indicate the actual gene spacing on the chromosome. In this example, the *S. cerevisiae* genes *TOR1* and *TOR2* are the only pair of ohnologs in the region and there is a single *TOR1*/*TOR2* ortholog in the other species. Nine other genes have all returned to a single-copy state following WGD. Apart from the post-WGD gene losses in *S. cerevisiae* and the presence of a gene (9923) in *K. waltii* that has no ortholog in the other species, there have been no other rearrangements of the region in any species. This image is a screenshot from a Yeast Gene Order Browser (YGOB) currently under development in our laboratory (K. Byrne and K. H. Wolfe, unpublished).



Ohno (1970) envisaged that whole genome duplication (WGD) provides a simple mechanism to generate vast numbers of duplicated genes. His name is so often associated with this process that gene duplicates produced by polyploidization are sometimes referred to as 'ohnologs' (Wolfe 2001). There are compelling advantages for polyploidy in evolution. In a polyploid species every gene is copied, including all the necessary regulatory elements. The relative stoichiometric proportions of all the gene products are also preserved in this process, minimizing potential damaging dosage effects caused by gene copy number imbalance (Papp et al. 2003). Polyploidy also generates fully redundant biochemical pathways, allowing freedom for radical biochemical innovation that can lead to major evolutionary transitions. Duplicating all the components of a pathway would be impos-

sible by other means unless all the genes are physically close. Therefore, polyploidization is able to provide great genetic flexibility without some of the problems associated with smaller scale duplication events.

In 1997, we proposed that *S. cerevisiae* is a paleopolyploid species derived from an ancestor whose genome duplicated in a single event roughly 10^8 years ago (Wolfe and Shields 1997). The hypothesis was that subsequent chromosomal translocation and gene loss events have shaped the *S. cerevisiae* genome into its current form. By assessing the locations of duplicated genes in the *S. cerevisiae* genome, several pieces of supporting evidence for this model were found, including (i) that approximately half of the genome could be paired into sister regions where a series of genes on one chromosome had a series of homologs on another chromosome; (ii) that the large sister regions did not overlap with one another; and (iii) that the overall orientation of duplicated regions, with respect to centromeres and telomeres, had largely remained the same. That a whole genome duplication (WGD) has occurred during the evolution of *S. cerevisiae* was confirmed in 2004 through the sequencing of the genomes of several species that separated from the *S. cerevisiae* lineage prior to the WGD (Dietrich et al. 2004; Dujon et al. 2004; Kellis et al. 2004; Fig. 2).

6 Intraspecific detection of genome duplication

Genomic data from a particular species provides two ways to uncover a past polyploidization event, using information from that species alone. Both methods are dependent on the presence of ohnologs. The map-based approach involves the matching up of chromosomes, or parts of chromosomes, that can be linked by homologs located in each sister region. In the example cartooned in Figure 2a, genes 1, 6, and 10 occur as duplicated pairs in the current genome and would allow identification of the whole region between genes 1 and 10 as a duplicated 'block'. Wolfe and Shields (1997) identified 55 such duplicated blocks in *S. cerevisiae*, and subsequent work verified 52 of these with a further 32 possible paired regions (Seoighe and Wolfe 1999a). Such fragmented blocks are proposed to have arisen by WGD with subsequent reciprocal translocation between chromosomes. This view is supported by the fact that nearly all pairs of sister regions are interchromosomal, as confirmed by independent analyses (Friedman and Hughes 2001; Cavalcanti et al. 2003). Physically, the identified duplicated blocks covered about 50% of the genome (compared to a theoretical expectation of 100% for a complete genome duplication), highlighting the limitation of intraspecific block detection methods.

Extensive loss of gene duplicates has occurred in *S. cerevisiae*. It was estimated that 16% of the total gene set are ohnologs, meaning that only 8% of duplicates were preserved from the pre-polyploid ancestor. Therefore, the low number of ohnologs retained in yeast is likely to cause the sizes of sister regions found by map-based approaches to be underestimated, and may even cause some regions to lie undetected in the case of small sister regions where every pair of duplicated genes

has been reduced to single-copy (Seoighe and Wolfe 1998; Kellis et al. 2004). An additional factor in yeast is the estimated 10^8 years of sequence divergence, which means that some ohnologs may not be identified using standard sequence similarity searches, further diminishing the coverage of sister regions.

Despite these shortcomings, there is convincing evidence from the *S. cerevisiae* genome itself that this species is a paleopolyploid. Llorente et al. (2000b) proposed a contradictory model where the duplicated blocks were produced by independent segmental duplications occurring at different times. The key to the resolution of the two hypotheses lies on the organization of sister regions. Under the WGD model, blocks produced by a single genome duplication event cannot overlap with each other, because the blocks are the surviving fragments of structures that were originally whole duplicated chromosomes (Wolfe and Shields 1997; Keogh et al. 1998). Under the alternative segmental duplication model, parts of chromosomes can be duplicated multiple times, creating significant overlaps between sister regions. The data clearly reveal that the majority of blocks do not overlap. Despite this, recent experimental evidence has shown that segmental duplications of large sections of chromosome can be formed in *S. cerevisiae* during artificial evolution experiments (Koszul et al. 2003), but for unknown reasons segmental duplications do not seem to have contributed significantly to the actual evolutionary history of the *S. cerevisiae* genome.

An alternative to map-based approaches to study paleopolyploidy is to use phylogenetic tree-based approaches to re-trace the origin of duplicated genes. The expectation is that ohnologs produced by a single round of genome duplication should be the same age. Friedman and Hughes (2001) tested this by estimating when, in evolutionary time, pairs of *S. cerevisiae* ohnologs diverged (i.e. the time they were duplicated). They identified 28 blocks containing genes that seemed to have duplicated simultaneously and 11 blocks containing varying amounts of relatively young duplicated genes. However, most of these recent duplicates are located within subtelomeric regions (regions near the telomeres of chromosomes) which can be subject to gene homogenizing effects (Wolfe 2001).

It is important to note that although a single genome duplication event has undoubtedly taken place in the lineage leading to *S. cerevisiae*, there are many other groups of paralogous genes in *S. cerevisiae* that were not formed by this event. The ancestral organism that underwent WGD was itself a complex eukaryote with numerous gene families – many of which expanded further via the WGD. With the availability of complete genome sequences, there is now good evidence that genes are often duplicated as tandem repeats, creating locally clustered multigene families. This is exemplified by the *SUC*, *MAL*, and *MEL* gene families in *S. cerevisiae* (Carlson et al. 1985; Michels et al. 1992; Turakainen et al. 1994). Interestingly, these tandem duplicates tend to be located in subtelomeric regions. It is known that recombinational exchanges, a process that can generate tandem repeats, are relatively frequent near chromosome ends compared to the rest of the genome (Pryde and Louis 1997). There is extensive population variation in the repertoire of some subtelomeric genes even between different yeast strains, and these genes often play adaptive roles, such as the utilization of different carbon sources or resistance to stresses (e.g. Maciaszczyk et al. 2004; Nomura and Takagi 2004). As

well as tandem repeats, some other gene families have copies with highly similar sequences at dispersed locations around the genome, but these were not formed by the WGD, for example the pyruvate decarboxylase (*PDC1*/ *PDC5*/ *PDC6*/ *THI3*) gene family (Moller et al. 2004).

7 Interspecific detection of genome duplication

Evidence of genome duplication can be obscured by events such as extensive gene loss, chromosomal rearrangements and independent gene duplications. As described above, the ohnolog approach to infer polyploidization is useful but it is limited in its ability to detect small sister regions in a genome. This can be greatly supplemented with genomics data from related species. Sister regions in *S. cerevisiae* are interspersed with 'singletons' – genes that were duplicated *en bloc* but have subsequently returned to single-copy (in Fig. 2a, genes 2, 3, 4, 5, 7, 8, and 9 are singletons). They have little informative value in intraspecific comparative mapping because only ohnologs can be considered. However, singletons can be brought into play using genomics data from an outgroup species that diverged before polyploidization. Immediately after genome duplication, every ancestral chromosomal region corresponds to a pair of duplicated blocks in the polyploid genome. In terms of gene order, it follows that every pair of neighboring genes is also duplicated. Due to the stochastic nature of gene loss after diploidization, a pair of previously adjacent genes may end up as singletons residing on different chromosomes, although still within the same duplicated block (*e.g.* genes 2 and 3 in Fig. 2a). Without nearby ohnologs to act as anchors, the pairing of the region would have been impossible to detect intraspecifically. Yet, the gene adjacency relationship is readily preserved in the genome of a species that diverged from the *S. cerevisiae* lineage before the WGD occurred (*e.g.* a species with the 'pre-WGD' gene order shown at the top of Fig. 2a). Therefore, ancestral gene order information can be invaluable in providing the missing links between sister regions.

Several early studies used fragmentary gene order information from other species to study the origins of sister regions in *S. cerevisiae* (Keogh et al. 1998; Ozier-Kalogeropoulos et al. 1998; Seoighe and Wolfe 1999a; Ladrière et al. 2000; Langkjaer et al. 2000; Llorente et al. 2000b; Wong et al. 2002). These studies suggested that species such as *K. lactis* had 'pre-WGD' gene orders similar to what is inferred to have existed in an ancestor of *S. cerevisiae* before the genome duplication happened. These findings have now been comprehensively confirmed through the sequencing of the complete genomes of three pre-WGD species: *E. gossypii* (Dietrich et al. 2004), *K. waltii* (Kellis et al. 2004), and *K. lactis* (Dujon et al. 2004), each of which shows a 1:2 gene order relationship to sister regions in the *S. cerevisiae* genome (Fig. 2b).

In contrast, 'post-WGD' species (e.g. *S. bayanus*, *S. mikatae*, *S. paradoxus*) showed extensive gene order conservation with *S. cerevisiae* (Keogh et al. 1998; Fischer et al. 2001). Most of the disruptions in synteny in these species have been attributed to genome rearrangements such as translocations after the WGD event.

However, some genomic regions in post-WGD species can also display an apparent pre-WGD organization (Langkjaer et al. 2000; Llorente et al. 2000b; Fischer et al. 2001). This can be explained by species divergence after WGD but before the process of gene loss is complete, resulting in differential gene loss between sister regions (Seoighe and Wolfe 1999b).

Phylogenetic analysis of genes in related species provides another way of detecting polyploidization. Ohnologs retained by a paleopolyploid genome are predicted to be present as singletons in species that diverged before genome duplication. Unless the complete genome sequence of an outgroup species is available, straightforward gene counting to investigate this 2:1 relationship, as attempted by Llorente et al. (2000a), may not be reliable due to the confounding influence of multigene families. In order to obtain a clearer picture, phylogenetic methods are required. A pair of ohnologs in one species is expected to be more closely related to one another than to their ortholog in a species that diverged before the duplication event. This is represented by an A(BC) topology in a phylogenetic tree, where A corresponds to the gene in the outgroup (pre-WGD) species and the ohnologs in the post-WGD species are denoted by B and C. Other possible topologies, C(AB) and B(AC), can reflect shared older gene duplication events followed by gene loss (i.e. misidentification of putative ohnologs), or rapid sequence divergence of one gene, causing aberrant phylogenetic tree reconstruction. Under perfect circumstances, trees drawn only from *S. cerevisiae* ohnologs and their orthologs in a pre-WGD species should all assume the A(BC) topology. Furthermore, the timing of the duplication event (the coalescence date) should be uniform among different ohnolog pairs.

Due to the lack of appropriate outgroup sequences, Wolfe and Shields (1997) obtained coalescent dates for only 12 pairs of duplicated genes and concluded that the genome duplication event had occurred on the order of magnitude of 10^8 years ago. This date is consistent with the results from a later analysis using a larger ohnolog data set with *C. albicans* genes acting as outgroups (Pal et al. 2001). However, some relatively young ohnolog pairs were found in both studies. There are several possible reasons why two pairs of genes that in fact duplicated simultaneously might appear to be different ages. The age can be underestimated if gene conversion has acted to homogenize the sequences at any time after their initial duplication. Aberrantly old date estimates can result if the pre-WGD genome contained a pair of tandemly duplicated genes that were already different in sequence, and each sister region in the post-WGD species retained one of these paralogs (Smith et al. 1999). In the context of a genome doubling process, a set of ohnologs may have apparent variable duplication dates due to the asynchronous nature of the diploidization process, as may have happened in maize (Gaut and Doebley 1997; Wolfe 2001).

Apart from irregular coalescent dates, there is also another difficulty in using of tree-based methods to place the WGD event on the phylogeny of hemiascomycetes. Trees constructed from *S. cerevisiae* ohnologs and their putative pre-polyploidization orthologs do not always conform to the A(BC) topology. For example, the citrate synthase genes of *S. cerevisiae* (*ScCIT1* and *ScCIT2*) are ohnologs based on their genomic locations, but phylogenetic analysis grouped

ScCIT1 with *SkCIT1* from *S. kluyveri* to the exclusion of *ScCIT2*, even though extensive gene order information indicates that *S. kluyveri* is a pre-WGD species (Langkjaer et al. 2000). More recently, Langkjaer et al. (2003) analyzed the phylogenetic relationship of 38 *S. cerevisiae* ohnolog pairs and their orthologs in five other yeasts. Surprisingly, significant proportions of orthologs from *S. kluyveri* and *K. lactis* (58% and 28% respectively) grouped with one member of their corresponding ohnolog pair. The authors arrived at the conclusion that the WGD event pre-dated the speciation of *S. kluyveri* and *K. lactis* from *S. cerevisiae*, and that different ohnolog pairs diverged in sequence independently at different times. The complete genome sequence data now available from the *S. kluyveri* and *K. lactis* genomes makes this hypothesis untenable, however, because these species are clearly pre-WGD (they only have one locus orthologous to each ohnolog pair in *S. cerevisiae*). We suspect that the phylogenetic trees reported by Langkjaer et al., including the *CIT* tree, have been affected by long branch attraction, an artifact of phylogenetic methodology that causes erroneous tree topologies and can arise if sequences have very unequal evolutionary rates (M. A. Fares and K. H. Wolfe, unpublished results).

8 Genes lost, genes kept

The occurrence of WGD during the evolution of *S. cerevisiae* enables us to study the evolutionary fates of a large sample of genes (i.e. every gene in the genome) that were all duplicated simultaneously. Some of these genes survived in two copies, whereas many others went back to being single copies. Studying the functions of these sets of genes provides some answers to the question of how duplicated genes can survive in a species where subfunctionalization is impossible. There seem to have been two major mechanisms by which duplicates formed by WGD survived: selection for increased dosage, and neofunctionalization.

For some types of gene, the presence of additional copies in the genome can confer a selective advantage even without any divergence in the function of the loci. This can occur through 'dosage' effects if a cell gains a competitive advantage simply by merit of having higher quantities of the protein or RNA encoded by the gene. This concept is familiar from examples such as the tandem amplification of metallothionein genes in response to high concentrations of copper (Fogel and Welch 1982), or the correlation between the numbers of copies of tandem repeats of the rDNA array and cell division rate (Rustchenko et al. 1993). Selection for increased dosage is the likely reason why almost every gene for cytosolic ribosomal proteins has been retained in duplicate following the WGD in an ancestor of *S. cerevisiae*; for most of these highly expressed gene pairs there has been little or no divergence in the sequence of the two copies and they are probably being homogenized by gene conversion. Similarly, selection for increased dosage probably underlay the retention, after WGD, of duplicated genes for chaperones such as *SSB1*/*SSB2* and *HSP82*/*HSC82*.

For other pairs of genes, neofunctionalization is the probable reason why both copies have been retained in the genome. One of the clearest examples of apparent *en masse* neofunctionalization of genes duplicated by WGD is in the establishment of a set of gene isoforms specialized for growth under highly anaerobic conditions. Well studied examples of aerobic/anaerobic gene pairs include *CYC1*/*CYC7*, and *COX5A*/*COX5B*. Microarray experiments identified ten ohnolog pairs that display alternate expression profiles under aerobic or hypoxic conditions and suggested that these are only the tip of the iceberg: one-quarter of *S. cerevisiae* ohnolog pairs have at least one member that shows differential expression depending on oxygen levels (Kwast et al. 2002). The group of species that are descended from the WGD (Fig. 1) also show other evidence of adaptation towards specialization for rapid anaerobic growth: all are likely petite-positive (meaning that they can dispense with their mitochondrial if grown on a fermentable carbon source; Piskur 2001), and their genomes are depleted of genes coding for oxygen-requiring peroxisomal oxidase enzymes (S. Wong and K. H. Wolfe, unpublished results). Another example of neofunctionalization after WGD is the formation of a specialized myosin heavy chain (Myo4) that is involved specifically in setting up the asymmetry between mother and daughter cells, while its ohnolog (Myo2) is not involved in this process and continues to carry out the more usual functions of myosin (Bohl et al. 2000). A further example of neofunctionalization is the evolution of Gal3 into an inducer of galactose catabolism, whereas its ohnolog Gal1 retains enzymatic activity as a galactokinase (Platt et al. 2000).

It is still unclear what fraction of the duplicated genes retained in *S. cerevisiae* after WGD were retained for dosage reasons, and what fraction underwent neofunctionalization. In fact, the two processes are not mutually exclusive, and some gene pairs that were originally retained for dosage reasons may subsequently have undergone functional divergence. Kellis et al. (2004) searched for examples of ohnolog pairs where one copy shows evidence of significantly accelerated evolution, as expected under Ohno's model, and found evidence of acceleration in 76 out of 457 ohnolog pairs (17%). Although there are many possible causes of such an acceleration, this result suggests that neofunctionalization may have occurred in many of the retained pairs. Kellis et al. pointed to several examples where the faster-evolving member of the pair also seemed to be the one with the more 'derived' function. Their result contrasts with an earlier study of ohnologs in tetraploid *Xenopus* (an organism with a much lower population size, making it much more likely that subfunctionalization will be a major factor in the retention of duplicate frog genes): Hughes and Hughes (1993) did not find any evidence of sequence acceleration in either copy of the *Xenopus* gene pairs.

Can any generalizations be made about which genes are retained after a WGD and which become single-copy again? Genome duplication provides a unique opportunity to compare the fates of duplicated genes in different functional categories because, unlike the case for individual gene duplications in a genome, all the ohnologs are the same age so those that have survived in duplicate have survived for the same length of time. One of the most striking early results about the WGD in *S. cerevisiae* was that almost all the genes for cytosolic ribosomal proteins were retained in duplicate, and genes coding for protein kinases and other signal trans-

duction components were also significantly over-represented among the ohnologs (Seoighe and Wolfe 1999b). Was this an accident, or was this outcome somehow inevitable? An indication that some types of genes might have higher probabilities of survival after a polyploidy has recently come from analyses of the genome of *Arabidopsis thaliana*. This plant underwent several successive polyploidizations during its evolution, the most recent of which was about 24-40 million years ago and so is considerably younger than the yeast WGD. For the most recent WGD in *Arabidopsis*, signal transduction (*i.e.* protein kinases and protein phosphatases) is among the categories of gene function that are over-represented among the retained genes, just like in yeast (Blanc and Wolfe 2004; Seoighe and Gehring 2004). Ribosomal proteins are also over-represented. Furthermore, the same types of gene tend to have been retained in duplicate after each round of WGD in *Arabidopsis* – that is, genes that were retained in duplicate after the earlier rounds of duplication are more likely also to have been retained in duplicate in the recent WGD (Seoighe and Gehring 2004). This suggests that there is a degree of inevitability to the sorting-out process after a WGD. Diversifying a signal transduction pathway by retaining duplicate genes for many of its components would be a powerful way to increase the regulatory complexity of an organism following a WGD. Interestingly, though, the end players in signal transduction cascades – transcription factors – are over-represented among the ohnologs in *Arabidopsis* but not in *Saccharomyces* (Seoighe and Wolfe 1999b; Blanc and Wolfe 2004; Seoighe and Gehring 2004).

However, yeasts and plants are so distantly related that perhaps one should not read too much into these apparently convergent results. It would be preferable to make comparisons about the outcomes of WGDs in groups of more closely-related species, as is now becoming possible in yeast species. It is particularly notable that *C. glabrata*, which is a descendant of the same WGD event as in *S. cerevisiae* (Fig. 1), does not retain two copies of most cytosolic ribosomal protein genes (Dujon et al. 2004). This suggests that species-specific factors can also strongly affect the outcome of a WGD. Further investigation of these types of questions should lead to a better understanding of the (r)evolutionary effects that the WGD had on yeast biology.

References

Baltimore D (2001) Our genome unveiled. Nature 409:814-816

Belloch C, Barrio E, Garcia MD, Querol A (1998) Inter- and intraspecific chromosome pattern variation in the yeast genus *Kluyveromyces*. Yeast 14:1341-1354

Belloch C, Querol A, Garcia MD, Barrio E (2000) Phylogeny of the genus *Kluyveromyces* inferred from the mitochondrial cytochrome-c oxidase II gene. Int J Syst Evol Microbiol 50:405-416

Blanc G, Hokamp K, Wolfe KH (2003) A recent polyploidy superimposed on older large-scale duplications in the *Arabidopsis* genome. Genome Res 13:137-144

Blanc G, Wolfe KH (2004) Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. Plant Cell 16:1679-1691

Bohl F, Kruse C, Frank A, Ferring D, Jansen RP (2000) She2p, a novel RNA-binding protein tethers *ASH1* mRNA to the Myo4p myosin motor via She3p. EMBO J 19:5514-5524

Cai J, Roberts IN, Collins MD (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

Calderone RA (2002) *Candida* and Candidiasis. ASM Press, Washington D.C.

Carlson M, Celenza JL, Eng FJ (1985) Evolution of the dispersed *SUC* gene family of *Saccharomyces* by rearrangements of chromosome telomeres. Mol Cell Biol 5:2894-2902

Cavalcanti AR, Ferreira R, Gu Z, Li WH (2003) Patterns of gene duplication in *Saccharomyces cerevisiae* and *Caenorhabditis elegans*. J Mol Evol 56:28-37

Cliften P, Sudarsanam P, Desikan A, Fulton L, Fulton B, Majors J, Waterston R, Cohen BA, Johnston M (2003) Finding functional features in *Saccharomyces* genomes by phylogenetic footprinting. Science 301:71-76

Demain AL, Phaff HJ, Kurtzman CP (1998) The industrial and agricultural significance of yeasts. In: Fell JW (ed) The Yeasts, A Taxonomic Study. Elsevier, Amsterdam

Dietrich FS, Voegeli S, Brachat S, Lerch A, Gates K, Steiner S, Mohr C, Pohlmann R, Luedi P, Choi S, Wing RA, Flavier A, Gaffney TD, Philippsen P (2004) The *Ashbya gossypii* genome as a tool for mapping the ancient *Saccharomyces cerevisiae* genome. Science 304:304-307

Dujon B, Sherman D, Fischer G, Durrens P, Casaregola S, Lafontaine I, de Montigny J, Marck C, Neuvéglise C, Talla E, Goffard N, Frangeul L, Aigle M, Anthouard V, Babour A, Barbe V, Barnay S, Blanchin S, Beckerich J-M, Beyne E, et al. (2004) Genome evolution in yeasts. Nature 430:35-44

Fischer G, Neuvéglise C, Durrens P, Gaillardin C, Dujon B (2001) Evolution of gene order in the genomes of two related yeast species. Genome Res 11:2009-2019

Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, Bult CJ, Tomb JF, Dougherty BA, Merrick JM, et al. (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae Rd*. Science 269:496-512

Fogel S, Welch JW (1982) Tandem gene amplification mediates copper resistance in yeast. Proc Natl Acad Sci USA 79:5342-5346

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

Friedman R, Hughes AL (2001) Gene duplication and the structure of eukaryotic genomes. Genome Res 11:373-381

Galagan JE, Calvo SE, Borkovich KA, Selker EU, Read ND, Jaffe D, FitzHugh W, Ma LJ, Smirnov S, Purcell S, Rehman B, Elkins T, Engels R, Wang S, Nielsen CB, Butler J, Endrizzi M, Qui D, Ianakiev P, Bell-Pedersen D, et al. (2003) The genome sequence of the filamentous fungus *Neurospora crassa*. Nature 422:859-868

Gaut BS, Doebley JF (1997) DNA sequence evidence for the segmental allotetraploid origin of maize. Proc Natl Acad Sci USA 94:6809-6814

Gellissen G, Hollenberg CP (1997) Application of yeasts in gene expression studies: a comparison of *Saccharomyces cerevisiae*, *Hansenula polymorpha* and *Kluyveromyces lactis* - a review. Gene 190:87-97

Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M, Louis EJ, Mewes HW, Murakami Y, Philippsen P, Tettelin H, Oliver SG (1996) Life with 6000 genes. Science 274:546, 563-567

Gojkovic Z, Paracchini S, Piskur J (1998) A new model organism for studying the catabolism of pyrimidines and purines. Adv Exp Med Biol 431:475-479

Hani J, Feldmann H (1998) tRNA genes and retroelements in the yeast genome. Nucleic Acids Res 26:689-696

Hazen KC (1995) New and emerging yeast pathogens. Clin Microbiol Rev 8:462-478

Hittinger CT, Rokas A, Carroll SB (2004) Parallel inactivation of multiple GAL pathway genes and ecological diversification in yeasts. Proc Natl Acad Sci USA 101:14144-14149

Holland BR, Huber KT, Moulton V, Lockhart PJ (2004) Using consensus networks to visualize contradictory evidence for species phylogeny. Mol Biol Evol 21:1459-1461

Hughes MK, Hughes AL (1993) Evolution of duplicate genes in a tetraploid animal, *Xenopus laevis*. Mol Biol Evol 10:1360-1369

James SA, Cai J, Roberts IN, Collins MD (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

Jones T, Federspiel NA, Chibana H, Dungan J, Kalman S, Magee BB, Newport G, Thorstenson YR, Agabian N, Magee PT, Davis RW, Scherer S (2004) The diploid genome sequence of *Candida albicans*. Proc Natl Acad Sci USA 101:7329-7334

Kellis M, Patterson N, Endrizzi M, Birren B, Lander ES (2003) Sequencing and comparison of yeast species to identify genes and regulatory elements. Nature 423:241-254

Kellis M, Birren BW, Lander ES (2004) Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. Nature 428:617-624

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

Koszul R, Caburet S, Dujon B, Fischer G (2003) Eucaryotic genome evolution through the spontaneous duplication of large chromosomal segments. EMBO J 23:234-243

Kurtzman CP, Fell JW (1998) The Yeasts, a taxonomic study. Elsevier, Amsterdam

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

Kurtzman CP (2003) Phylogenetic circumscription of *Saccharomyces*, *Kluyveromyces* and other members of the *Saccharomycetaceae*, and the proposal of the new genera *Lachancea*, *Nakaseomyces*, *Naumovia*, *Vanderwaltozyma* and *Zygotorulaspora*. FEMS Yeast Res 4:233-245

Kurtzman CP, Robnett CJ (2003) Phylogenetic relationships among yeasts of the '*Saccharomyces* complex' determined from multigene sequence analyses. FEMS Yeast Res 3:417-432

Kwast KE, Lai LC, Menda N, James DT 3rd, Aref S, Burke PV (2002) Genomic analyses of anaerobically induced genes in *Saccharomyces cerevisiae*: functional roles of Rox1 and other factors in mediating the anoxic response. J Bacteriol 184:250-265

Ladrière JM, Georis I, Guerineau M, Vandenhaute J (2000) *Kluyveromyces marxianus* exhibits an ancestral *Saccharomyces cerevisiae* genome organization downstream of *ADH2*. Gene 255:83-91

Langkjaer RB, Nielsen ML, Daugaard PR, Liu W, Piskur J (2000) Yeast chromosomes have been significantly reshaped during their evolutionary history. J Mol Biol 304:271-288

Langkjaer RB, Cliften PF, Johnston M, Piskur J (2003) Yeast genome duplication was followed by asynchronous differentiation of duplicated genes. Nature 421:848-852

Llorente B, Durrens P, Malpertuy A, Aigle M, Artiguenave F, Blandin G, Bolotin-Fukuhara M, Bon E, Brottier P, Casaregola S, Dujon B, de Montigny J, Lepingle A, Neuveglise C, Ozier-Kalogeropoulos O, Potier S, Saurin W, Tekaia F, Toffano-Nioche C, Wesolowski-Louvel M, et al. (2000a) Genomic Exploration of the Hemiascomycetous Yeasts: 20. Evolution of gene redundancy compared to *Saccharomyces cerevisiae*. FEBS Lett 487:122-133

Llorente B, Malpertuy A, Neuveglise C, de Montigny J, Aigle M, Artiguenave F, Blandin G, Bolotin-Fukuhara M, Bon E, Brottier P, Casaregola S, Durrens P, Gaillardin C, Lepingle A, Ozier-Kalogeropoulos O, Potier S, Saurin W, Tekaia F, Toffano-Nioche C, Wesolowski-Louvel M, et al. (2000b) Genomic Exploration of the Hemiascomycetous Yeasts: 18. Comparative analysis of chromosome maps and synteny with *Saccharomyces cerevisiae*. FEBS Lett 487:101-112

Lynch M, Force A (2000) The probability of duplicate gene preservation by subfunctionalization. Genetics 154:459-473

Maciaszczyk E, Wysocki R, Golik P, Lazowska J, Ulaszewski S (2004) Arsenical resistance genes in *Saccharomyces douglasii* and other yeast species undergo rapid evolution involving genomic rearrangements and duplications. FEMS Yeast Res 4:821-832

Michels CA, Read E, Nat K, Charron MJ (1992) The telomere-associated *MAL3* locus of *Saccharomyces* is a tandem array of repeated genes. Yeast 8:655-665

Moller K, Olsson L, Piskur J (2001) Ability for anaerobic growth is not sufficient for development of the petite phenotype in *Saccharomyces kluyveri*. J Bacteriol 183:2485-2489

Moller K, Langkjaer RB, Nielsen J, Piskur J, Olsson L (2004) Pyruvate decarboxylases from the petite-negative yeast *Saccharomyces kluyveri*. Mol Genet Genomics 270:558-568

Mortimer RK (1993a) Øjvind Winge: Founder of yeast genetics. In: Hall MN, Linder P (eds) The Early Days of Yeast Genetics. Cold Spring Harbor Laboratory Press, New York, p 3-16

Mortimer RK (1993b) Carl C. Lindegren: Iconoclastic father of *Neurospora* and yeast genetics. In: Hall MN, Linder P (eds) The Early Days of Yeast Genetics. Cold Spring Harbor Laboratory Press, New York, p 17-38

Nomura M, Takagi H (2004) Role of the yeast acetyltransferase Mpr1 in oxidative stress: Regulation of oxygen reactive species caused by a toxic proline catabolism intermediate. Proc Natl Acad Sci USA 101:12616-12621

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

Oliver SG, van der Aart QJ, Agostoni-Carbone ML, Aigle M, Alberghina L, Alexandraki D, Antoine G, Anwar R, Ballesta JP, Benit P, et al. (1992) The complete DNA sequence of yeast chromosome III. Nature 357:38-46

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

Pal C, Papp B, Hurst LD (2001) Highly expressed genes in yeast evolve slowly. Genetics 158:927-931

Papp B, Pal C, Hurst LD (2003) Dosage sensitivity and the evolution of gene families in yeast. Nature 424:194-197

Paterson AH, Bowers JE, Chapman BA (2004) Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. Proc Natl Acad Sci USA 101:9903-9908

Petersen RF, Nilsson-Tillgren T, Piskur J (1999) Karyotypes of *Saccharomyces sensu lato* species. Int J Syst Bacteriol 49:1925-1931

Phillips MJ, Delsuc F, Penny D (2004) Genome-scale phylogeny and the detection of systematic biases. Mol Biol Evol 21:1455-1458

Piskur J (2001) Origin of the duplicated regions in the yeast genomes. Trends Genet 17:302-303

Platt A, Ross HC, Hankin S, Reece RJ (2000) The insertion of two amino acids into a transcriptional inducer converts it into a galactokinase. Proc Natl Acad Sci USA 97:3154-3159

Pryde FE, Louis EJ (1997) *Saccharomyces cerevisiae* telomeres. A review. Biochemistry (Mosc) 62:1232-1241

Ramezani-Rad M, Hollenberg CP, Lauber J, Wedler H, Griess E, Wagner C, Albermann K, Hani J, Piontek M, Dahlems U, Gellissen G (2003) The *Hansenula polymorpha* (strain CBS4732) genome sequencing and analysis. FEMS Yeast Res 4:207-215

Rokas A, Williams BL, King N, Carroll SB (2003) Genome-scale approaches to resolving incongruence in molecular phylogenies. Nature 425:798-804

Rustchenko EP, Curran TM, Sherman F (1993) Variations in the number of ribosomal DNA units in morphological mutants and normal strains of Candida albicans and in normal strains of *Saccharomyces cerevisiae*. J Bacteriol 175:7189-7199

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

Seoighe C, Wolfe KH (1999a) Updated map of duplicated regions in the yeast genome. Gene 238:253-261

Seoighe C, Wolfe KH (1999b) Yeast genome evolution in the post-genome era. Curr Opin Microbiol 2:548-554

Seoighe C, Gehring C (2004) Genome duplication led to highly selective expansion of the Arabidopsis thaliana proteome. Trends Genet 20:461-464

Smith NGC, Knight R, Hurst LD (1999) Vertebrate genome evolution: a slow shuffle or a big bang? Bioessays 21:697-703

Souciet J, Aigle M, Artiguenave F, Blandin G, Bolotin-Fukuhara M, Bon E, Brottier P, Casaregola S, de Montigny J, Dujon B, Durrens P, Gaillardin C, Lepingle A, Llorente B, Malpertuy A, Neuveglise C, Ozier-Kalogeropoulos O, Potier S, Saurin W, Tekaia F, et al. (2000) Genomic exploration of the hemiascomycetous Yeasts: 1. A set of yeast species for molecular evolution studies. FEBS Lett 487:3-12

Sychrova H, Braun V, Potier S, Souciet JL (2000) Organization of specific genomic regions of *Zygosaccharomyces rouxii* and *Pichia sorbitophila*: comparison with *Saccharomyces cerevisiae*. Yeast 16:1377-1385

Taylor JS, Van de Peer Y, Braasch I, Meyer A (2001) Comparative genomics provides evidence for an ancient genome duplication event in fish. Philos Trans R Soc Lond B Biol Sci 356:1661-1679

Torok T, Rockhold D, King AD, Jr. (1993) Use of electrophoretic karyotyping and DNA-DNA hybridization in yeast identification. Int J Food Microbiol 19:63-80

Turakainen H, Kristo P, Korhola M (1994) Consideration of the evolution of the *Saccharomyces cerevisiae MEL* gene family on the basis of the nucleotide sequences of the genes and their flanking regions. Yeast 10:1559-1568

Vandepoele K, De Vos W, Taylor JS, Meyer A, Van De Peer Y (2004) Major events in the genome evolution of vertebrates: Paranome age and size differ considerably between ray-finned fishes and land vertebrates. Proc Natl Acad Sci USA 101:1638-1643

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

Wolfe KH (2001) Yesterday's polyploids and the mystery of diploidization. Nat Rev Genet 2:333-341

Wolfe KH, Li W-H (2003) Molecular evolution meets the genomics revolution. Nature Genet 33 Suppl:255-265

Wong S, Butler G, Wolfe KH (2002) Gene order evolution and paleopolyploidy in hemiascomycete yeasts. Proc Natl Acad Sci USA 99:9272-9277

Wong S, Fares MA, Zimmermann W, Butler G, Wolfe KH (2003) Evidence from comparative genomics for a complete sexual cycle in the "asexual" pathogenic yeast *Candida glabrata*. Genome Biol 4:R10

Wood V, Rutherford KM, Ivens A, Rajandream M-A, Barrell B (2001) A re-annotation of the *Saccharomyces cerevisiae* genome. Compar Funct Genomics 2:143-154

Wood V, Gwilliam R, Rajandream MA, Lyne M, Lyne R, Stewart A, Sgouros J, Peat N, Hayles J, Baker S, Basham D, Bowman S, Brooks K, Brown D, Brown S, Chillingworth T, Churcher C, Collins M, Connor R, Cronin A, et al. (2002) The genome sequence of *Schizosaccharomyces pombe*. Nature 415:871-880

Wolfe, Kenneth H.
Department of Genetics, Smurfit Institute, University of Dublin, Trinity College, Dublin 2, Ireland
khwolfe@tcd.ie

Wong, Simon
Department of Genetics, Smurfit Institute, University of Dublin, Trinity College, Dublin 2, Ireland