Yeast
Yeast (in press)
Published online in Wiley InterScience
(www.interscience.wiley.com) **DOI:** 10.1002/yea.1515

Review

Yeast genome evolution — the origin of the species

Devin R. Scannell[1,2], Geraldine Butler[3] and Kenneth H. Wolfe[1]*

[1] *Smurfit Institute of Genetics, Trinity College, Dublin 2, Ireland*
[2] *Lawrence Berkeley National Laboratory, 1 Cyclotron Road, MS 84R0171, Berkeley, CA 94720, USA*
[3] *UCD School of Biomolecular and Biomedical Science, Conway Institute, University College Dublin, Belfield, Dublin 4, Ireland*

**Correspondence to:*
Kenneth H. Wolfe, Smurfit Institute of Genetics, Trinity College, Dublin 2, Ireland.
E-mail: khwolfe@tcd.ie

Received: 8 February 2007
Accepted: 24 May 2007

Abstract

With almost 20 genomes sequenced from unicellular ascomycetes (Saccharomycotina), and the prospect of many more in the pipeline, we review the patterns and processes of yeast genome evolution. A central core of about 4000 genes is shared by all the sequenced yeast genomes. Gains of genes by horizontal gene transfer seem to be very rare. Gene losses are more frequent, and losses of whole sets of genes in some pathways in some species can be understood in terms of species-specific differences in biology. The wholesale loss of redundant copies of duplicated genes after whole-genome duplication in the ancestor of one clade of yeasts is likely to have caused the emergence of many reproductively isolated lineages of yeasts at that time, but other processes are responsible for species barriers that arose more recently among close relatives of *Saccharomyces cerevisiae*.

Keywords: comparative genomics; phylogenetics; gene loss; gene gain; gene duplication; speciation

Contents

Phylogenetic position of the Saccharomycotina among ascomycetes

The number of extant fungal species is thought to be in the millions, although only about 80 000 have been described [1]. The known species are typically divided into five phyla, Ascomycota, Basidiomycota, Glomeromycota, Zygomycota and Chytridomycota, although some of these may not be monophyletic [2]. The largest phylum, Ascomycota, is defined by the production of a specialized structure, the ascus, that surrounds the spores formed during meiosis [2]. The Ascomycota diverged from the Basidiomycota in the region of 741–1195 million years ago (95% confidence interval from Ref. 3), and the monophyly of the Ascomycota is well supported by molecular data [4,5]. There are two main subphyla within the Ascomycota: the Pezizomycotina (which includes hyphal fungi such as *Neurospora crassa*) and the Saccharomycotina (which includes yeasts such as *Candida albicans* and *Saccharomyces cerevisiae* and are sometimes called the hemiascomycetes). These two subphyla diverged about 798–1166 million years ago [3], shortly after the split from Basidiomycota. Recent molecular evidence showing that *Schizosaccharomyces pombe* is an outgroup to both of these taxa has prompted the proposal of a third ascomycete class [6], the Archiascomycotina (Figure 1). Additional molecular sequence data have supported the novel classification [2,4,5], indicating that unicellular yeasts have evolved from multicellular hyphal progenitors more than once.

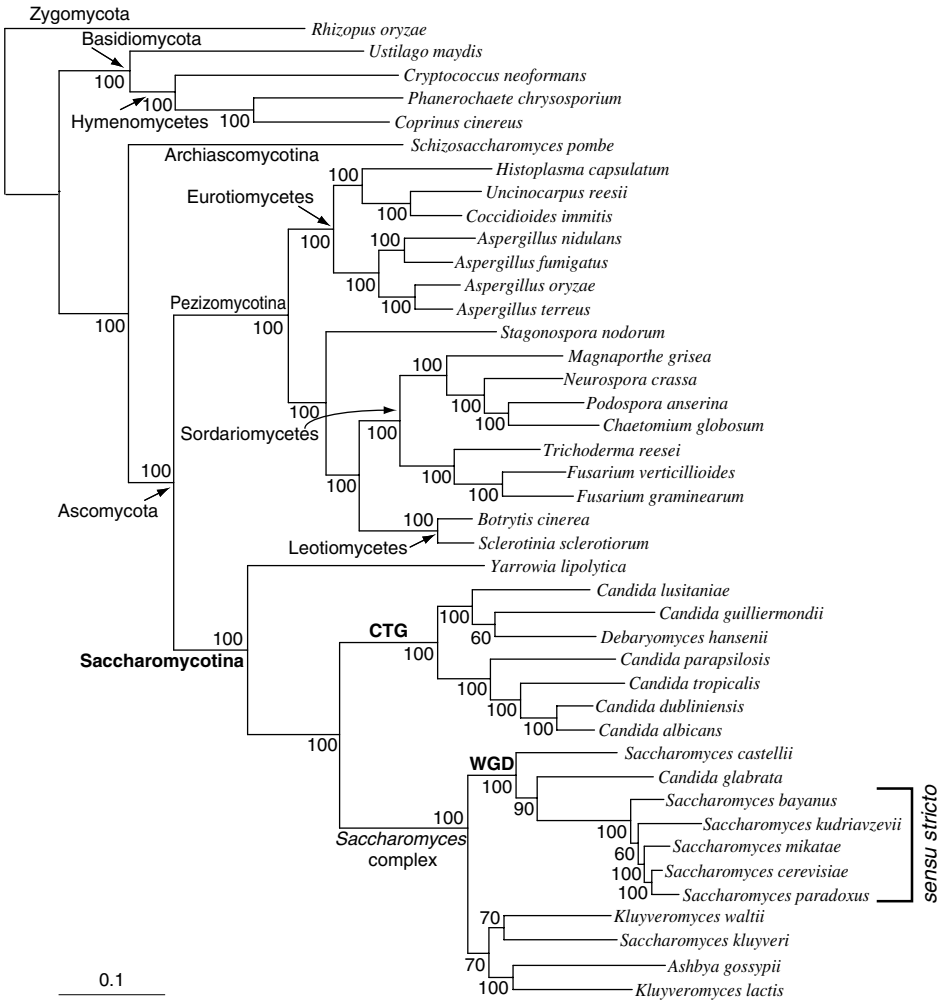

Figure 1. Phylogenetic relationships among the sequenced fungal genomes [5]. The tree is a maximum-likelihood phylogeny reconstructed using the concatenated sequences of 153 genes that are universally present in the 42 genomes shown. Bootstrap percentages are shown for all nodes. Major clades are named, including the '*Saccharomyces* complex' [9], the group of species that share the whole-genome duplication (WGD), those with the variant genetic code (CTG) and the *Saccharomyces sensu stricto* group. Modified from Ref. 5

More than 1000 species of Saccharomycotina have been described. Their centre of greatest biodiversity appears to be in insect guts, and it is likely that thousands more species remain to be discovered from this habitat [7,8]. All of the Saccharomycotina genomes sequenced thus far fall into three clusters (Figure 1). The first cluster is comprised primarily of species from the genera *Saccharomyces* and *Kluyveromyces* (Figure 1), and is generally referred to as the '*Saccharomyces* complex' [9]. The second cluster consists of species that translate CTG codons as serine rather than leucine, a reassignment that occurred more than 170 million years ago [10,11]. This cluster includes *Candida* species and yeasts such as *Debaryomyces hansenii* and *Lodderomyces elongisporus* [5]. The sole sequenced member of the third cluster is *Yarrowia lipolytica* [12].

In this review we focus on the yeasts in the '*Saccharomyces* complex', the most densely sequenced of the three Saccharomycotina clusters. We begin by reviewing a number of studies that illustrate the key mechanisms underlying gene gain and gene loss in yeasts and discuss the whole-genome duplication (WGD) that occurred in a common ancestor of several yeast species. In the second part we

DOI: 10.1002/yea

review the literature on speciation in yeasts and describe recent work implicating gene duplication and the subsequent loss of redundant gene copies as a key mechanism of speciation. In particular, we focus on work from our laboratory showing that loss of duplicate gene pairs after the WGD may have led to a burst of speciation after the WGD event.

Gene losses and gains

The *S. cerevisiae* and *Ashbya gossypii* genomes share 4281 genes that are syntenic and orthologous [13]. This set of genes forms a common core that is largely shared by all the sequenced genomes of Saccharomycotina species. A particular species can deviate from this core by losing some genes that it does not need, or by gaining new genes, either by gene duplication or by horizontal transfer. How do these gene losses and gains become established in a species?

Gene loss

Even though many *S. cerevisiae* genes reveal no phenotype when deleted, it is unlikely that any genes in yeast genomes are truly redundant. If any such genes existed, there would be no advantage to maintaining them and hence no selection against losing them, so we would expect them to have been mutated and lost from genomes long ago. Genes with no obvious phenotype may nevertheless make a small contribution to growth rate, or may only be needed under certain conditions that were not studied in the laboratory [14,15].

Given that truly redundant genes are unlikely to exist and assuming that a gene is initially present in all individuals in a population, there are three circumstances that could allow it to become lost: (a) if the selection pressure that caused the gene to be maintained disappears [16]; (b) if a new selection pressure emerges that causes the gene to be maladaptive [17]; or (c) if another gene is present that can complement the loss of the original gene [18]. In population genetics terms these three situations can be described, respectively, as the removal of purifying selection, as gene loss favoured by positive selection, and as a selectively neutral event. Examples of all three situations are known, as described below.

Seven genes in the *GAL* pathway are absent or pseudogenes in the genome of *S. kudriavzevii*, despite being present in all other studied *Saccharomyces sensu stricto* species [16]. These genes function to sense, import and metabolize the sugar galactose. Their loss has been attributed to the removal of a selective pressure, and is presumed to have occurred because *S. kudriavzevii* does not encounter galactose in its natural environment. Although it is hard to exclude the possibility that their loss was beneficial in some way, the fact that the *GAL* genes were also lost independently in several other yeast lineages that occupy very different ecological niches argues against the possibility that the *GAL* genes were maladaptive in the specific environment (rotting leaves [16]) preferred by *S. kudriavzevii*.

In contrast, the loss of five genes in the *BNA* pathway in *Candida glabrata* is likely to have occurred under strong positive selection, because this pathway plays an important role in virulence [19]. The *BNA* pathway is responsible for the synthesis of nicotinic acid and allows *S. cerevisiae* and other yeasts to replenish their pool of NAD^+ if it is depleted by the transcriptional repressor Sir2. *C. glabrata*, however, is entirely dependent on external sources of nicotinic acid. When this is not available, genes that are usually repressed by Sir2 become expressed. Notably, the human urinary tract is very low in nicotinic acid, and expression of the Sir2-regulated *EPA* adhesin genes is increased in this environment [19].

Lastly, the loss of the *a*2 gene from the ancestral mating-type (*MAT*) locus in yeasts appears to be an example of gene loss due to redundancy [20,21]. In *C. albicans a*2 is required to activate **a**-specific genes in **a** cells, but in *S. cerevisiae* these genes are expressed by default in **a** cells and are instead repressed by $\alpha 2$ in α cells. By examining how **a**-specific genes are regulated in yeasts that are more closely related to *S. cerevisiae* than to *C. albicans*, Tsong *et al*. [21] reconstructed the evolutionary steps that brought the **a**-specific genes from positive control in an ancestor similar to *C. albicans* to negative control in *S. cerevisiae*, and showed that an intermediate stage is likely to have involved redundant control by both systems. Thus, loss of the *a*2 gene was possible because, although there was strong purifying selection for appropriate expression of **a**-specific genes, compensatory changes arose that could complement the loss. As

DOI: 10.1002/yea

in the case of the *GAL* pathway, it is hard to exclude the possibility that the change was favoured by selection for some unknown reason, but these three examples serve to illustrate the possible conditions under which gene loss may occur.

Gene gain

Although gene gain might at first glance appear to be just gene loss in reverse, there are some important differences. Foremost among these is that, irrespective of the selection pressures operating on the new gene, it may originate by a variety of processes, including gene duplication, horizontal gene transfer (HGT) and *de novo* emergence of an open reading frame. Of these, gene duplication is by far the most important but a number of noteworthy cases of HGT into yeasts have also been described. The best-documented example is *URA1*, coding for an enzyme (dihydro-orotate dehydrogenase) in the uracil biosynthesis pathway, which was gained in the Saccharomycotina lineage by horizontal transfer from a bacterium resembling *Lactococcus lactis* [22,23]. *URA1* initially coexisted with the distantly related native gene *URA9*, and eventually displaced *URA9* in some lineages, including *S. cerevisiae*. The enzymes encoded by the two genes use different electron acceptors: the Ura9 enzyme requires a functioning mitochondrial electron transport chain, whereas the Ura1 enzyme is active even in the absence of respiration. Thus, the gain of *URA1* by horizontal transfer likely contributed to the increased ability of yeasts in the *S. cerevisiae* lineage to grow without oxygen [22]. Hall *et al*. [23] searched systematically for genes that were transferred from bacteria into the *S. cerevisiae* or *A. gossypii* genome and found only 11 such genes, including *URA1*. Remarkably, eight of the 11 transferred genes are located in subtelomeric regions of the yeast genomes, which perhaps suggests that subtelomeric regions are a preferred site for the uptake of novel DNA. Horizontal transfer of an adenyl deaminase (ADA) gene from *Burkholderia* bacteria into the wine spoilage yeast *Dekkera bruxellensis* (or from an unknown third species into both *Dekkera* and *Burkholderia*) has also been reported recently [24].

As noted above, the key mechanism of gene gain in yeast and probably throughout eukaryotes is gene duplication [25]. It now appears that duplicates must be preserved by one of two process: (a) neofunctionalization, in which one duplicate evolves a useful new function while the other performs the ancestral function; [26] or (b) subfunctionalization, in which the duplicates partition ancestral functions between them so that both duplicates are required for full fitness. [27] Although computational analyses have tended to provide support primarily for neofunctionalization, [28,29] experimental analyses have provided support for both models. For instance, Thomson *et al*. [30] used ancestral sequence reconstruction and chemical synthesis to 'resurrect' the common ancestor of Adh1 and Adh2, homologous alcohol dehydrogenases that favour opposing chemical reactions. They showed that the ancestral Adh enzyme favoured the forward reaction converting acetaldehyde to ethanol, thus behaving in the same manner as Adh1. This strongly suggests that the *ADH2* gene was preserved because it conferred a novel beneficial function (converting ethanol to acetaldehyde), although it must be pointed out that Adh1 and Adh2 differ by 24 (of 348) amino acids, and thus the possibility cannot be excluded that *ADH2* was originally preserved for a different reason and only lately has gained its novel behaviour.

If it is difficult to prove that a duplicate pair has been preserved by neofunctionalization because the behaviour of evolutionary intermediates may not be the same as that of modern sequences, it is no less difficult to demonstrate that subfunctionalization has occurred. This is because even when it is clear that each of a pair of duplicates can perform only a subset of functions that a single-copy homologue (or a reconstructed common ancestor) can perform, it is all but impossible to exclude the possibility that one of the duplicates also has an unknown novel function that was not possessed by the progenitor sequence and that this is the ultimate reason for preservation. Nevertheless, van Hoof [31] has provided compelling evidence that four pairs of duplicate genes preserved in *S. cerevisiae* since the WGD have been preserved by subfunctionalization. Although all four pairs were chosen on the basis that they were likely candidates for preservation by neofunctionalization, *SIR3* and *ORC1* are of particular interest because both are well-studied genes with few obvious functional similarities. *SIR3* is a non-essential gene involved in silencing at the *HML and HMR* loci and at subtelomeres,

DOI: 10.1002/yea

whereas *ORC1* is an essential subunit of the origin recognition complex (ORC) that initiates DNA replication. Van Hoof [31] showed that the single-copy *SIR3/ORC1* homologue in *S. kluyveri* can complement deletions of both *SIR3* and *ORC1* in *S. cerevisiae*, even though the two *S. cerevisiae* deletions cannot complement one another. This result implies that neither *SIR3* nor *ORC1* in *S. cerevisiae* performs an essential function that was not present in their shared ancestor, and that they have partitioned (subfunctionalized) the ancestral functions between them.

These examples lend themselves to two conclusions. First, as was the case with gene loss, gene gain seems to be possible both with and without positive Darwinian selection. Although *a*2 is required for mating in *C. albicans*, it was lost from an ancestor of the '*Saccharomyces* complex' species without any obvious compensating benefit. Similarly, the replacement of the single-copy *SIR3/ORC1* gene that existed prior to the WGD (and was probably also required for mating) by a pair of paralogues in *S. cerevisiae* also occurred without conferring any apparent advantage. In contrast, adaptive rationales can be ascribed to the loss of the genes in the *BNA* pathway and to the gain of *ADH2*. Second, it suggests that the resolution of duplicate gene pairs after the WGD that occurred in the '*Saccharomyces* complex' involved a variety of different processes. In particular, although most gene loss is likely to have been neutral because of the presence of a second copy, the preservation in duplicate of hundreds of previously single-copy genes may have involved both neofunctionalization and subfunctionalization.

The yeast whole-genome duplication

The major phylogenetic division within the *Saccharomyces* complex is between those yeasts whose common ancestor underwent a whole-genome duplication (WGD [12,13,28,32]) and those that diverged prior to this event (Figure 1). The genome of modern *S. cerevisiae* is dominated by the changes wrought by the WGD and the subsequent return of the genome to a haploid/diploid inheritance cycle [33]. The most obvious structural change is the doubling of the number of chromosomes relative to the ancestral 'pre-WGD' yeast, which can be tracked by identifying pairs of duplicated centromeres [28,34]. The genome transiently increased its number of genes from about 5000 to 10 000, and then lost one member of most of the pairs to leave the present-day set. *S. cerevisiae* currently has about 5500 protein-coding genes, of which 1102 form 551 duplicated 'ohnolog' pairs [35].

The loss of one member of most of the previously duplicated gene pairs and the resulting interleaving of single-copy genes between sister regions (Figure 2) means that around half of all neighbouring gene relationships were altered in the aftermath of the WGD. It has recently been shown that following a WGD event in the plant *Arabidopsis*, the pattern of duplicate loss between sister regions

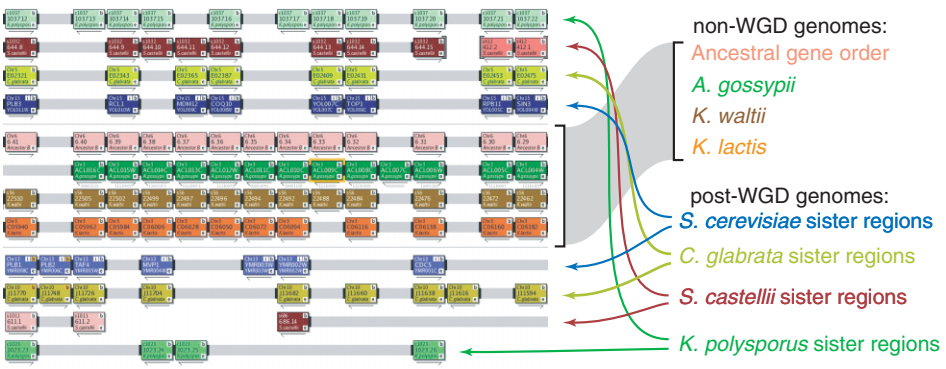

Figure 2. Screenshot from the Yeast Gene Order Browser (for details, see ref. 35). Each rectangle represents a gene and homologous genes are arranged in columns. Genes from the two sister regions in each post-WGD species appear 'interleaved' in single genomic regions from non-WGD species. In addition to data from ref. 35, gene order in the *Kluyveromyces polysporus* genome and the inferred 'ancestral' gene order that existed immediately prior to the WGD event are shown (D. R. Scannell, J. L. Gordon and K. H. Wolfe, unpublished)

DOI: 10.1002/yea

was not random but resulted in the production of 'gene-rich' and 'gene-poor' regions [36]. Although there is no evidence that a similar process occurred after the WGD in yeast, the orientation bias of neighbouring genes has been changed, resulting in a reduction of the excess of convergently and divergently transcribed neighbouring gene pairs (as opposed to parallel orientations) seen in non-WGD yeasts [37]. This appears to have had an effect on the correlation in expression of neighbouring genes [37] and raises the obvious question of whether the chromosomal clustering of co-expressed genes [38] and genes involved in the same biological process [39] in *S. cerevisiae* was affected by the reorganization of neighbouring gene relationships after the WGD.

Ancient WGDs (paleopolyploidizations) have been analysed at the genome-sequence level in all four eukaryotic kingdoms: plants, [40–44] animals, [45–47] fungi [13,28,32] and protists [48]. Among these WGD events, the fraction of the original genes that survived in duplicate varies from approximately 10% to 50%, but this is largely a function of the amount of time since polyploidization. Nevertheless, it is notable that similar functional classes of genes have been retained in duplicate after many of these events. For instance, cytosolic ribosomal protein genes have been retained in duplicate in both plants [49] and fungi [50]. Similarly, transcription factors and/or kinases ('regulatory' genes) were preferentially retained in duplicate after the WGDs in yeast, [50] plants [51] and animals. [52] In addition, it has been shown that duplicates derived from a first WGD event have a significantly increased chance of being re-retained after subsequent WGD events, [48,53] and that the types of genes that are retained in duplicate after WGD typically do not give rise to duplicates by other mechanisms [51]. Because the characteristics of cytosolic ribosomal protein genes and 'regulatory' genes are very different, it is likely that more than one explanation will be required to account for these observations. In the former case, it has been proposed that ribosomal protein genes are retained for increased dosage [50] and that WGD is the primary mode of duplication by which this can happen, because a duplication of only some of the set of ribosomal protein genes would lead to dosage imbalance and a dominant negative phenotype [54]. In reference to our comment in the previous section ('Gene gain') that duplicated genes are preserved by either subfunctionalization or neofunctionalization, we note that the increased gene dosage conferred by a second duplicate can be regarded as a form of quantitative neofunctionalization. Although no qualitatively new function has evolved, the presence of a duplicate confers an advantage relative to the ancestral genotype.

No plausible explanation has yet been offered for the preferential retention of kinases and transcription factors in duplicate after WGD, although a number of possibilities can be considered. First, there is some evidence that genes in these functional classes have more complex promoters [55,56] and thus may be particularly good candidates for preservation by subfunctionalization. Second, kinases and transcription factors often have many substrates [57] or targets, [58] respectively, and can thus be considered to be under a high level of pleiotropic constraint. Because target phosphorylation sites or *cis*-regulatory elements are likely to be heterogeneous (i.e. all deviating from the consensus in slightly different ways), partial loss-of-function mutations in each member of a pair of duplicates may result in each having high affinity for only a subset of the ancestral targets, and thus a requirement for both copies to be retained. This is reminiscent of both coding region subfunctionalization [59] and quantitative subfunctionalization [60]. Third, it is possible that the simultaneous duplication of multiple regulatory genes prevents dysregulation. Indeed, it is notable that kinases and transcription factors are among the functional classes most likely to produce a deleterious phenotype when overexpressed in isolation [61]. Finally, studies in both plants [49] and yeasts [62] have suggested that duplicated pathways may become independently expressed following WGD. It is possible that regulatory genes are only recruited when new pathways require regulation.

There are some convincing examples of neofunctionalization associated with the WGD in yeast, the most striking of which involve adaptation of duplicated gene copies to growth in low-oxygen environments. Among the Saccharomycotina, yeasts are classified as either Crabtree-positive or Crabtree-negative. In Crabtree-positive species (such as *S. cerevisiae*) glucose is metabolized through fermentation in preference to respiration, even when oxygen levels are high. *S. cerevisiae* is also among the

DOI: 10.1002/yea

few yeasts that can grow in the complete absence of oxygen, and can be induced to generate respiratory-deficient mitochondrial mutants ('petites') [63]. A recent study by Merico *et al*. [64] showed that these phenotypes are generally present in the species that are descended from the WGD event, whereas the majority of the non-WGD species show a reduced Crabtree effect and a reduced ability to grow in the absence of oxygen. Many of the gene pairs that have been retained in duplicate by *S. cerevisiae* are involved in carbohydrate metabolism, [50] and the members of many of these pairs are differentially expressed in response to either oxygen [65] or glucose [66] availability. The evidence suggests that the last common ancestor of the species with duplicated genomes was adapted to living in a low-oxygen environment, and had the capacity to ferment. The expansion of lineages descended from the WGD may have coincided with the increased availability of fruit sugars at the end of the Cretaceous [32,67]. Further adaptations to the 'fermentative lifestyle' occurred specifically in an ancestor of the *Saccharomyces sensu stricto* species, with duplications of *ADH* and *PDC* genes, among others, happening more recently than the WGD [30].

Origins of new species

Saccharomyces sensu stricto yeast species are generally accepted to be distinct on the basis of low viability of spores produced by hybridization. Whereas mating between members of the same *S. cerevisiae* strain produces spores with viabilities of close to 100% and spores produced by mating between *S. cerevisiae* strains often show viabilities of ~80%, [68] mating between *S. cerevisiae* and *S. paradoxus* or other *Saccharomyces sensu stricto* species typically result in <1% of spores being viable (for references, see ref. 69). The determinants of reproductive barriers among yeast species have been investigated intensely over the last few years. Three are reviewed here: chromosomal rearrangements, sequence divergence acted on by the mismatch repair system, and a modified Dobzhansky–Muller mechanism related to reciprocal gene loss. The third of these mechanisms provides a link between the loss of duplicated genes, particularly after WGD, and the emergence of reproductively isolated lineages.

Chromosomal speciation

Chromosomal rearrangements are hypothesized to lead to hybrid inviability by inducing the formation of multivalents at meiosis. Multivalents are prone to mis-segregation and can result in the production of aneuploid spores with decreased fitness. The reduction of fitness may be due either to spores being deficient for essential genes or to the increased likelihood of mis-segregation in future meioses. Both retrospective and interventionist approaches have been employed to estimate the contribution of chromosomal rearrangements to hybrid viability between *S. cerevisiae* and other *sensu stricto* yeasts.

Fischer *et al*. [70] used a combination of electrophoresis and PCR to identify karyotype changes in *sensu stricto* yeasts relative to *S. cerevisiae*. They detected no rearrangements in *S. paradoxus* or *S. kudriavzevii* relative to *S. cerevisiae*, but four in *S. cariocanus* and *S. bayanus* and two in *S. mikatae*. These observations are inconsistent with the known levels of spore viability among these species. For instance, if each rearrangement reduces spore viability by 50%, then the expected spore viability in a cross between *S. cariocanus* and *S. paradoxus* is 6.25%, but the observed viability is only one-tenth of this. Additional factors must therefore contribute and Fischer *et al*. concluded that chromosomal rearrangements were not a prerequisite for speciation.

Nevertheless, the possibility remained that rearrangements contribute quantitatively to reproductive isolation, or that they may reinforce species barriers after they have arisen by another mechanism. To address this issue, Delneri *et al*. [71] used the *Cre-lox* inducible recombination system to engineer strains of *S. cerevisiae* that are colinear to one of two strains of *S. mikatae*. One of these *S. mikatae* strains differs from wild-type *S. cerevisiae* (but not the engineered strain Sct1) by a single rearrangement, and the other differs from wild-type *S. cerevisiae* (but not the engineered strain Sct1/2) by two rearrangements. In subsequent crosses between these engineered strains and wild-type *S. cerevisiae*, spore viabilities of 60% and 25% were obtained with Sct1 and Sct1/2, respectively. These percentages are close to what is expected under the assumption of 50% loss of viability per rearrangement noted above, and suggests that mis-segregation contributes to spore death. In addition, interspecific crosses between Sct1 and the

DOI: 10.1002/yea

S. mikatae strain with which it is co-linear, resulted in 20–30% spore viability in four of ten crosses. These data clearly support the view that chromosomal rearrangements at least have the potential to contribute to species barriers in yeast, but the failure to restore full viability indicates that other mechanisms must also be invoked. Indeed, it was noticed that all of the viable spores were aneuploid, with some having up to 25 chromosomes. It is therefore possible that these extra chromosomes are masking recessive Dobzhansky–Muller incompatibilities (discussed below) that might otherwise reduce viability.

Sequence divergence acted on by the mismatch repair system

In contrast to the chromosomal rearrangement model of speciation, there is unambiguous evidence that sequence differences between homologous chromosomes can interfere with recombination and lead to non-productive meioses between diverged yeast species [72]. Moreover, there is evidence that this interference is mediated by the mismatch repair system and that it results in spore inviability by two separate mechanisms, non-disjunction at meiosis I [72] and mismatch-stimulated chromosome loss [73]. Both of these mechanisms result in potentially lethal aneuploidy. Indeed, the most attractive aspect of this model is that it predicts the existence of the widespread aneuploidy that has arisen during (and confounded) attempts to study other possible mechanisms of speciation.

To test the hypothesis that sequence divergence detected by the mismatch repair system can lead to aberrant meioses, Hunter *et al* [72]. crossed strains of *S. cerevisiae* and *S. paradoxus* and then measured the rates of both recombination and aneuploidy in the resulting gametes. This was done using wild-type and mismatch repair-deficient (*pms1* null and *msh2* null) strains of *S. cerevisiae*. Comparisons between crosses performed using the wild-type and mutant strains showed that recombination, non-disjunction and viability changed in concert. For instance, both the spore viability and the rate of recombination seen when wild-type *S. cerevisiae* was crossed to wild-type *S. paradoxus* was approximately 1% of that seen in intraspecific crosses. In contrast, when *msh2* null *S. cerevisiae* was crossed with wild-type *S. paradoxus*, recombination and viability both rose to ~10%. In addition, non-disjunction was significantly lower when an *msh2* null strain was crossed to *S. paradoxus* than when a wild-type strain was used. These data support the view that, when diverged sequences pair at meiosis but fail to recombine (due to the mismatch repair system), non-disjunction may occur and lead to inviable aneuploid spores. Subsequent work clarified the mechanism by which this occurs. Chambers *et al*. [73] showed that asci that contain two viable spores tend to be disomic, consistent with meiosis I non-disjunction, but asci with three viable spores typically contain no disomes and one recombinant spore. They proposed that the unpaired recombinant genotype arises because, although the sequences of *S. cerevisiae* and *S. paradoxus* are similar enough that one successful strand invasion may occur, the probability of the reciprocal strand invasion occurring is negligible. Hence, one recombinant chromosome is formed and the other is aborted.

Two lines of evidence suggest that sequence divergence acted on by the mismatch repair system may be sufficient to account for reproductive isolation among *sensu stricto* yeast species. First, Greig *et al*. [74] used the same assays described above to assess the impact of between-strain sequence differences on reproductive isolation in *S. cerevisiae* and *S. paradoxus* and found in both cases that it could account for at least 50% of the variation: spore viability and recombination were both increased in a *msh2* null background. Second, Liti *et al*. have shown that, once chromosomal rearrangements are taken into account, there is a linear relationship between the level of sequence divergence and the level of spore inviability [75]. This is consistent with a causal relationship and, in the absence of any significant evidence that genic incompatibilities play a role in species barriers among *sensu stricto* yeasts, suggests that sequence divergence may be a sufficient explanation.

Dominant and recessive Dobzhansky–Muller incompatibilities

An alternative to the chromosomal basis for hybrid infertility is the existence of Dobzhansky–Muller incompatibilities between epistatically interacting genes. This model posits that after an ancestral species splits to create two daughter lineages, incompatible changes can arise in alternative members of a pair of loci that interact and are co-adapted [76]. Thus, in one lineage one of the genes

DOI: 10.1002/yea

diverges from its ancestral sequence, whereas in the second lineage the other gene diverges from its ancestral sequence. These changes are neutral (or possibly beneficial), provided that the other locus has not changed in sequence, but if the diverged versions of both genes are brought together in a hybrid they will interact in such a way as to reduce fitness. It is important to note that the incompatibility can be either dominant or recessive. In the former case, the presence of the two diverged genes will reduce fitness irrespective of what other genes are present. In the latter case, however, the existence of an incompatibility can be masked by the presence of an ancestral type sequence at both loci (e.g. in an F_1 hybrid).

To test the possibility that *dominant* Dobzhansky–Muller incompatibilities might play a role in reproductive isolation between *sensu stricto* yeast lineages, Greig *et al*. [68] repeated the test originally performed by Dobzhansky in *Drosophila [77]*. Dobzhansky had observed that in infertile *D. pseudoobscura* hybrids, homologous chromosomes failed to pair at meiosis, thus arresting spermatogenesis. In order to distinguish between the possibility that the chromosomes could not pair because their sequences were too divergent and the possibility that genetic incompatibilities between the two parents had prevented successful meiosis, Dobzhansky examined the pairing of tetraploid spermatocytes. Because tetraploidy is achieved by duplication of the homologous chromosomes that are present in diploids, failure to pair cannot be due to the lack of a homologous partner. When Dobzhansky performed this test using tetraploid spermatocytes, he observed that the hybrids were still infertile and concluded that sterility was due to genetic factors. Strikingly, when repeated using *sensu stricto* yeast species, precisely the opposite result was obtained [68].

Greig *et al*. [68] first created pseudo-haploids of several yeast species by deleting a single copy of the *MAT* locus from non-hybrid diploids. They then made interspecific crosses between *S. cerevisiae* pseudo-haploids and pseudo-haploids from the other *sensu stricto* species. In each case, the spore viability of the hybrid was ~90% compared to <1% for true hybrid diploids. Indeed, the spore viability of the hybrids obtained by crossing pseudo-haploids was not significantly different from that obtained in intraspecific crosses of normal haploids. These data indicate comprehensively that hybrid infertility among these yeast species is not due to dominant Dobzhansky–Muller incompatibilities. If dominant interactions between loci were responsible, increasing the number of copies of each gene present would not be able to rescue the infertile phenotype.

That *recessive* Dobzhansky–Muller incompatibilities do not play a role in speciation of *sensu stricto* yeasts is suggested by the fact that *S. cerevisiae* chromosome III can be replaced by *S. paradoxus* chromosome III without any loss of viability in the haploid [73]. This indicates that, although the chromosomes are ~15% divergent at the DNA level and ~10% divergent at the protein sequence level, [78] all the functional elements on chromosome III are conserved between these two species. Moreover, because the *S. paradoxus* chromosome III is present in an otherwise completely *S. cerevisiae* background, no recessive Dobzhansky–Muller incompatibilities can exist between loci on *S. paradoxus* chromosome III and other loci in the genome. Liti *et al*. [75] mention (without evidence) that most chromosomes in *S. cerevisiae* can similarly be replaced individually by their *S. paradoxus* counterparts. If this is true, it strongly suggests that Dobzhansky–Muller incompatibilities play little part in *sensu stricto* yeast speciation. Moreover, because the *S. paradoxus* and *S. cerevisiae* genomes are co-linear, [70,79] it suggests that sequence divergence acted on by the mismatch repair system is the primary mechanism of speciation in these yeasts.

There is, however, some indirect evidence that recessive Dobzhansky–Muller incompatibilities exist in yeast species, based on interspecific crosses. Whereas dominant epistatic interactions can be revealed by crossing haploids from two parental species and examining the fertility of the F_1 generation, recessive incompatibilities can only be revealed by examining F_2 or successive generations in which regions of the genome may be homozygous at the locus of interest. To investigate the fertility of an F_2 generation, Greig *et al*. exploited the fact that most F_1 hybrid diploids are fertile at a low level (typically <1%) and collected 80 gametes from a large cross [69]. They then allowed these to auto-diploidize to obtain a homozygous F_2 generation. Interestingly, the F_2 hybrids fulfilled the two main requirements for a new species, high fertility (~80%) and isolation from the ancestral population (back-cross hybrid

DOI: 10.1002/yea

fertility ~7%). Nevertheless, the reason for the ~20% decrease in fertility relative to the pure parental strain is unclear. As the authors point out, chromosomal incompatibilities cannot explain the difference, since the F_2 hybrids were produced by auto-diploidization and must therefore be able to pair chromosomes at meiosis. In addition, the authors argue that aneuploidy is not the explanation, although they show — as was also observed for the hybrids obtained by crossing *S. mikatae* to artificially co-linear *S. cerevisiae* strains [71] — that the F_2 hybrids are highly aneuploid. By this process of exclusion, Greig *et al*. [69] concluded that the decreased fertility must be attributable to recessive Dobzhansky–Muller incompatibilities. However, given the results of the chromosome complementation experiments cited above, [73] direct evidence for a role in reproductive isolation will be required to establish their relevance.

Although the evidence for a contribution of Dobzhansky–Muller incompatibility to reproductive isolation among *sensu stricto* species is equivocal, it should be noted that abundant epistasis has been detected in genome-wide scans for expression quantitative trait loci (QTLs) [80] and that negative fitness consequences have been demonstrated for certain pairs of alleles from different *S. cerevisiae* strains [81]. For instance, haploids with an *MLH1* allele from strain S288C (*cMLH1*) and a *PMS1* allele from strain SK1 (*kPMS1*) were shown to accumulate mutations at approximately 100 times the rate of any other combination of alleles (*cMLH1–cPMS1, kMLH1–kPMS1, kMLH1–cPMS1*). This defect was observed in both genetic backgrounds and shown to result in a significant reduction in the number of complete tetrads over the course of ~100 generations, consistent with a fitness cost [81]. Thus, although the *cMLH1–kPMS1* interaction results in neither inviability nor sterility of spores produced by crossing S288C and SK1, it indicates that incompatibilities exist between genotypes of different strains and that other more severe incompatibilities may also be segregating.

Modified Dobzhansky–Muller mechanism

In spite of the popularity of the Dobzhansky–Muller model, [76] only a handful of 'speciation genes' have been identified, and the two members of a pair of epistatically interacting loci have been identified in only a single case [82]. One possible explanation for why speciation genes have been so elusive — even in taxa such as *Drosophila*, where evidence supports the existence of Dobzhansky–Muller incompatibilities [77] — is that another mechanism, which behaves similarly to Dobzhansky–Muller incompatibility in genetic crosses but does not involve co-adapted gene pairs, also exists. One such mechanism was suggested by Werth and Windham, [83] based on studies of polyploid plants, and was subsequently recognized by Lynch and Force as a special case of Dobzhansky–Muller incompatibility [84]. Werth and Windham [83] proposed that reciprocal loss of different members of a duplicated gene pair in two lineages can lead to reduced hybrid fitness, because gametes produced by a hybrid may receive a null copy of the previously duplicated gene from each of the parental genotypes (Figure 3). In this scenario, a pair of null genes takes the place of the pair of epistatically interacting protein-coding genes in the classical Dobzhansky–Muller model. Loss of fitness arises because the hybrid gamete (or spore) is deficient for a required function, rather than because of an incompatibility *per se*. Nevertheless, the expected results in genetic crosses are the same as in the case of classical recessive Dobzhansky–Muller incompatibility; assuming the previously duplicated gene is essential and that the surviving copies reside on different chromosomes in the parental lineages, one-quarter of hybrid gametes will be inviable (Figure 3).

Despite their similarities, the modified Dobzhansky–Muller mechanism differs significantly from the classical model in terms of the underlying mutations. Whereas classical Dobzhansky–Muller incompatibility invokes co-adapted alleles segregating at pairs of loci and is often thought to arise as a consequence of adaptive substitutions, [76] the modified Dobzhansky–Muller model relies on gene duplication and subsequent null (inactivating) mutations. Two considerations arise from this. First, because the modified mechanism does not rely on adaptive mutations and the rates of both gene duplication [25] and null mutations are high in eukaryotes, it may be a frequent and hence comparatively important mechanism by which reproductive isolation is created [85]. Second, because large numbers of genes are duplicated and subsequently lost following WGD events, the modified

DOI: 10.1002/yea

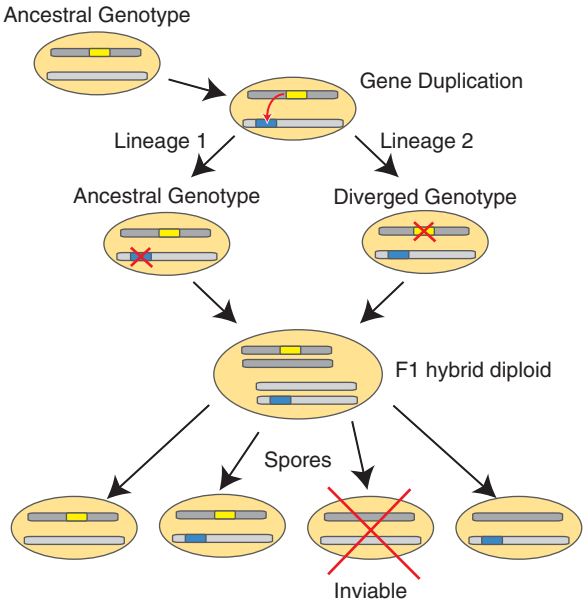

Figure 3. Cartoon illustration of the modified Dobzhansky–Muller mechanism of reproductive isolation resulting from reciprocal loss of duplicate gene copies. In this example, the blue gene was formed by duplication of the yellow gene. Two lineages containing the duplicate gene pair then separate. Lineage 1 later loses the blue gene, restoring the ancestral genotype. Lineage 2 loses the yellow gene and retains the blue one. If lineages 1 and 2 subsequently meet and hybridize, 1/4 of the spores produced by their hybrid will lack both the yellow and the blue genes and will be inviable if the gene product is essential. If the same process occurs at several duplicated gene pairs, the net spore viability will be approximately $(3/4)^n$, where n is the number of duplicated essential genes that were lost reciprocally in the two lineages [83,84]. For 50 such genes, only one spore per 1.7 million is expected to be viable

Dobzhansky–Muller model may provide a mechanism by which species radiations can occur. Werth and Windham calculated that for hybrids between a pair of lineages that diverged soon after a polyploidy event, when 70% of the ancestral genome is still duplicated and assuming only 500 essential genes in the genome, less than 0.5% of gametes are expected to be viable [83]. As more loci become single-copy and more realistic numbers of essential genes are considered, the probability that hybrids could produce viable gametes declines rapidly and the number of mutually reproductively isolated lineages that can emerge rises sharply.

We examined the hypothesis that reciprocal gene loss after a WGD can lead to the emergence of multiple daughter lineages by comparing the genomes of three yeasts, *S. cerevisiae, S. castellii* and *C. glabrata*, that diverged from after the WGD in their common ancestor [86]. We used the Yeast Gene Order Browser (Figure 2 [35]) to trace the fates of ancestrally duplicated genes among lineages and showed that reciprocal gene loss had occurred at hundreds of ancestral loci between all pairwise combinations of species. Consistent with the expectation outlined above, we estimated that the probability of producing a fertile hybrid spore following a mating between *S. castellii* and *S. cerevisiae* was at most 6×10^{-9}, thus confirming that the level of reciprocal gene loss is more than sufficient to account for reproductive isolation among these species [86]. Moreover, reciprocal gene loss occurred at the same time as speciation. By inferring the number of genes that were still duplicated at internal nodes in the phylogenetic tree, we found that the WGD was followed by a period of rapid and widespread gene loss and that the majority of reciprocal gene loss events occurred contemporaneously with the divergence of the lineages represented by *S. cerevisiae, S. castellii* and *C. glabrata.*

Our results and the evidence that sequence divergence operated on by the mismatch repair system [75] can both affect spore inviability suggest that at least two mechanisms contribute to speciation among yeasts in the Saccharomycotina. Immediately following the WGD, the loss of large numbers of duplicate genes from the genomes of incipient yeast species resulted in the emergence of several major lineages (corresponding to Clades 1–6 in Kurtzman's phylogenetic tree [9]). As the rate of gene loss slowed, however, reciprocal gene loss and the modified Dobzhansky–Muller mechanism contributed progressively less to the establishment of new reproductive barriers. Sequence divergence operated on by the mismatch repair system appears to be the principal isolating mechanism among modern *sensu stricto* species.

Conclusion

Yeast comparative genomics has proved to be a powerful system for addressing questions in genome evolution. The existence of a highly stable core of genes in almost all sequenced yeast species and the remarkable conservation of synteny in the '*Saccharomyces* complex' have been of particular

DOI: 10.1002/yea

importance. They have permitted detailed studies of both gene gain (by duplication and horizontal transfer), gene loss (under a variety of circumstances) and the combined process of gene gain and loss that occurs with WGD. Recent studies have greatly clarified the process of speciation among *sensu stricto* yeasts, as well as implicating WGD and reciprocal loss of previously duplicated genes in the emergence of several major lineages.

Acknowledgements

Research in the Wolfe and Butler laboratories is supported by Science Foundation Ireland. We thank David Fitzpatrick for drawing Figure 1.

References

1. Hedges SB. 2002. The origin and evolution of model organisms. *Nat Rev Genet* **3**: 838–849.
2. James TY, Kauff F, Schoch CL, *et al*. 2006. Reconstructing the early evolution of fungi using a six-gene phylogeny. *Nature* **443**: 818–822.
3. Hedges SB, Blair JE, Venturi ML, Shoe JL. 2004. A molecular timescale of eukaryote evolution and the rise of complex multicellular life. *BMC Evol Biol* **4**: 2.
4. Galagan JE, Henn MR, Ma LJ, Cuomo CA, Birren B. 2005. Genomics of the fungal kingdom: insights into eukaryotic biology. *Genome Res* **15**: 1620–1631.
5. Fitzpatrick DA, Logue ME, Stajich JE, Butler G. 2006. A fungal phylogeny based on 42 complete genomes derived from supertree and combined gene analysis. *BMC Evol Biol* **6**: 99.
6. Heckman DS, Geiser DM, Eidell BR, *et al*. 2001. Molecular evidence for the early colonization of land by fungi and plants. *Science* **293**: 1129–1133.
7. Suh S-O, McHugh JV, Pollock DD, Blackwell MD. 2005. The beetle gut: a hyperdiverse source of novel yeasts. *Mycol Res* **109**: 261–265.
8. Boekhout T. 2005. Gut feeling for yeasts. *Nature* **434**: 449–451.
9. Kurtzman CP, Robnett CJ. 2003. Phylogenetic relationships among yeasts of the '*Saccharomyces* complex' determined from multigene sequence analyses. *FEMS Yeast Res* **3**: 417–432.
10. Massey SE, Moura G, Beltrao P, *et al*. 2003. Comparative evolutionary genomics unveils the molecular mechanism of reassignment of the CTG codon in *Candida* spp. *Genome Res* **13**: 544–557.
11. Miranda I, Silva R, Santos MA. 2006. Evolution of the genetic code in yeasts. *Yeast* **23**: 203–213.
12. Dujon B, Sherman D, Fischer G, *et al*. 2004. Genome evolution in yeasts. *Nature* **430**: 35–44.
13. Dietrich FS, Voegeli S, Brachat S, *et al*. 2004. The *Ashbya gossypii* genome as a tool for mapping the ancient *Saccharomyces cerevisiae* genome. *Science* **304**: 304–307.
14. Thatcher JW, Shaw JM, Dickinson WJ. 1998. Marginal fitness contributions of nonessential genes in yeast. *Proc Natl Acad Sci USA* **95**: 253–257.
15. Delneri D, Gardner DC, Oliver SG. 1999. Analysis of the seven-member *AAD* gene set demonstrates that genetic redundancy in yeast may be more apparent than real. *Genetics* **153**: 1591–1600.
16. 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.
17. Olson MV. 1999. When less is more: gene loss as an engine of evolutionary change. *Am J Hum Genet* **64**: 18–23.
18. Morett E, Korbel JO, Rajan E, *et al*. 2003. Systematic discovery of analogous enzymes in thiamine biosynthesis. *Nat Biotechnol* **21**: 790–795.
19. Domergue R, Castano I, De Las Penas A, *et al*. 2005. Nicotinic acid limitation regulates silencing of *Candida* adhesins during UTI. *Science* **308**: 866–870.
20. Tsong AE, Miller MG, Raisner RM, Johnson AD. 2003. Evolution of a combinatorial transcriptional circuit: a case study in yeasts. *Cell* **115**: 389–399.
21. Tsong AE, Tuch BB, Li H, Johnson AD. 2006. Evolution of alternative transcriptional circuits with identical logic. *Nature* **443**: 415–420.
22. Gojkovic Z, Knecht W, Zameitat E, *et al*. 2004. Horizontal gene transfer promoted evolution of the ability to propagate under anaerobic conditions in yeasts. *Mol Genet Genom* **271**: 387–393.
23. Hall C, Brachat S, Dietrich FS. 2005. Contribution of horizontal gene transfer to the evolution of *Saccharomyces cerevisiae*. *Eukaryot Cell* **4**: 1102–1115.
24. Woolfit M, Rozpedowska E, Piskur J, Wolfe KH. 2007. Genome survey sequencing of the wine spoilage yeast *Dekkera (Brettanomyces) bruxellensis*. *Eukaryot Cell* **6**: 721–733.
25. Lynch M, Conery JS. 2000. The evolutionary fate and consequences of duplicate genes. *Science* **290**: 1151–1155.
26. Ohno S. 1970. *Evolution by Gene Duplication*. George Allen and Unwin: London.
27. Force A, Lynch M, Pickett FB, *et al*. 1999. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151**: 1531–1545.
28. Kellis M, Birren BW, Lander ES. 2004. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* **428**: 617–624.
29. Byrne KP, Wolfe KH. 2007. Consistent patterns of rate asymmetry and gene loss indicate widespread neofunctionalization of yeast genes after whole-genome duplication. *Genetics* **175**: 1341–1350 (PMID 17194778).
30. Thomson JM, Gaucher EA, Burgan MF, *et al*. 2005. Resurrecting ancestral alcohol dehydrogenases from yeast. *Nat Genet* **37**: 630–635.
31. van Hoof A. 2005. Conserved functions of yeast genes support the duplication, degeneration and complementation model for gene duplication. *Genetics* **171**: 1455–1461.
32. Wolfe KH, Shields DC. 1997. Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* **387**: 708–713.
33. Wolfe KH. 2001. Yesterday's polyploids and the mystery of diploidization. *Nat Rev Genet* **2**: 333–341.
34. Wolfe KH. 2006. Comparative genomics and genome evolution in yeasts. *Phil Trans R Soc Lond B Biol Sci* **361**: 403–412.

35. Byrne KP, Wolfe KH. 2005. The Yeast Gene Order Browser: combining curated homology and syntenic context reveals gene fate in polyploid species. *Genome Res* **15**: 1456–1461.
36. Thomas BC, Pedersen B, Freeling M. 2006. Following tetraploidy in an *Arabidopsis* ancestor, genes were removed preferentially from one homeolog leaving clusters enriched in dose-sensitive genes. *Genome Res* **16**: 934–946.
37. Byrnes JK, Morris GP, Li WH. 2006. Reorganization of adjacent gene relationships in yeast genomes by whole-genome duplication and gene deletion. *Mol Biol Evol* **23**: 1136–1143.
38. Cohen BA, Mitra RD, Hughes JD, Church GM. 2000. A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. *Nat Genet* **26**: 183–186.
39. Lee JM, Sonnhammer EL. 2003. Genomic gene clustering analysis of pathways in eukaryotes. *Genome Res* **13**: 875–882.
40. Simillion C, Vandepoele K, Van Montagu MC, Zabeau M, Van de Peer Y. 2002. The hidden duplication past of *Arabidopsis thaliana*. *Proc Natl Acad Sci USA* **99**: 13627–13632.
41. Blanc G, Wolfe KH. 2004. Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. *Plant Cell* **16**: 1667–1678.
42. 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.
43. Yu J, Wang J, Lin W, *et al*. 2005. The genomes of *Oryza sativa*: a history of duplications. *PLoS Biol* **3**: e38.
44. Tuskan GA, Difazio S, Jansson S, *et al*. 2006. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* **313**: 1596–1604.
45. McLysaght A, Hokamp K, Wolfe KH. 2002. Extensive genomic duplication during early chordate evolution. *Nat Genet* **31**: 200–204.
46. Dehal P, Boore JL. 2005. Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biol* **3**: e314.
47. Jaillon O, Aury JM, Brunet F, *et al*. 2004. Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. *Nature* **431**: 946–957.
48. Aury JM, Jaillon O, Duret L, *et al*. 2006. Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. *Nature* **444**: 171–178.
49. Blanc G, Wolfe KH. 2004. Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. *Plant Cell* **16**: 1679–1691.
50. Seoighe C, Wolfe KH. 1999. Yeast genome evolution in the post-genome era. *Curr Opin Microbiol* **2**: 548–554.
51. Maere S, De Bodt S, Raes J, *et al*. 2005. Modeling gene and genome duplications in eukaryotes. *Proc Natl Acad Sci USA* **102**: 5454–5459.
52. Blomme T, Vandepoele K, De Bodt S, *et al*. 2006. The gain and loss of genes during 600 million years of vertebrate evolution. *Genome Biol* **7**: R43.
53. Seoighe C, Gehring C. 2004. Genome duplication led to highly selective expansion of the *Arabidopsis thaliana* proteome. *Trends Genet* **20**: 461–464.
54. Papp B, Pal C, Hurst LD. 2003. Dosage sensitivity and the evolution of gene families in yeast. *Nature* **424**: 194–197.
55. Nelson CE, Hersh BM, Carroll SB. 2004. The regulatory content of intergenic DNA shapes genome architecture. *Genome Biol* **5**: R25.
56. Iwama H, Gojobori T. 2004. Highly conserved upstream sequences for transcription factor genes and implications for the regulatory network. *Proc Natl Acad Sci USA* **101**: 17156–17161.
57. Ptacek J, Devgan G, Michaud G, *et al*. 2005. Global analysis of protein phosphorylation in yeast. *Nature* **438**: 679–684.
58. Harbison CT, Gordon DB, Lee TI, *et al*. 2004. Transcriptional regulatory code of a eukaryotic genome. *Nature* **431**: 99–104.
59. Dermitzakis ET, Clark AG. 2001. Differential selection after duplication in mammalian developmental genes. *Mol Biol Evol* **18**: 557–562.
60. Lynch M, Force A. 2000. The probability of duplicate gene preservation by subfunctionalization. *Genetics* **154**: 459–473.
61. Sopko R, Huang D, Preston N, *et al*. 2006. Mapping pathways and phenotypes by systematic gene overexpression. *Mol Cell* **21**: 319–330.
62. Conant GC, Wolfe KH. 2006. Functional partitioning of yeast co-expression networks after genome duplication. *PLoS Biol* **4**: e109.
63. Snoek IS, Steensma HY. 2007. Factors involved in anaerobic growth of *Saccharomyces cerevisiae*. *Yeast* **24**: 1–10.
64. Merico A, Sulo P, Piskur J, Compagno C. 2007. Fermentative lifestyle in yeasts belonging to the *Saccharomyces* complex. *FEBS J* **274**: 976–989.
65. Kwast KE, Lai LC, Menda N, *et al*. 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.
66. DeRisi JL, Iyer VR, Brown PO. 1997. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278**: 680–686.
67. Piskur J. 2001. Origin of the duplicated regions in the yeast genomes. *Trends Genet* **17**: 302–303.
68. Greig D, Borts RH, Louis EJ, Travisano M. 2002. Epistasis and hybrid sterility in *Saccharomyces*. *Proc R Soc Lond B Biol Sci* **269**: 1167–1171.
69. Greig D, Louis EJ, Borts RH, Travisano M. 2002. Hybrid speciation in experimental populations of yeast. *Science* **298**: 1773–1775.
70. Fischer G, James SA, Roberts IN, Oliver SG, Louis EJ. 2000. Chromosomal evolution in *Saccharomyces*. *Nature* **405**: 451–454.
71. Delneri D, Colson I, Grammenoudi S, *et al*. 2003. Engineering evolution to study speciation in yeasts. *Nature* **422**: 68–72.
72. Hunter N, Chambers SR, Louis EJ, Borts RH. 1996. The mismatch repair system contributes to meiotic sterility in an interspecific yeast hybrid. *EMBO J* **15**: 1726–1733.
73. Chambers SR, Hunter N, Louis EJ, Borts RH. 1996. The mismatch repair system reduces meiotic homeologous recombination and stimulates recombination-dependent chromosome loss. *Mol Cell Biol* **16**: 6110–6120.
74. Greig D, Travisano M, Louis EJ, Borts RH. 2003. A role for the mismatch repair system during incipient speciation in *Saccharomyces*. *J Evol Biol* **16**: 429–437.
75. Liti G, Barton DB, Louis EJ. 2006. Sequence diversity, reproductive isolation and species concepts in *Saccharomyces*. *Genetics* **174**: 839–850.

76. Coyne JA, Orr HA. 2004. *Speciation*. Sinauer: Sunderland, MA.
77. Dobzhansky T. 1933. On the sterility of the interracial hybrids in *Drosophila pseudoobscura*. *Proc Natl Acad Sci USA* **19**: 397–403.
78. Cliften PF, Hillier LW, Fulton L, *et al*. 2001. Surveying *Saccharomyces* genomes to identify functional elements by comparative DNA sequence analysis. *Genome Res* **11**: 1175–1186.
79. 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.
80. Brem RB, Storey JD, Whittle J, Kruglyak L. 2005. Genetic interactions between polymorphisms that affect gene expression in yeast. *Nature* **436**: 701–703.
81. Heck JA, Argueso JL, Gemici Z, *et al*. 2006. Negative epistasis between natural variants of the *Saccharomyces cerevisiae MLH1* and *PMS1* genes results in a defect in mismatch repair. *Proc Natl Acad Sci USA* **103**: 3256–3261.
82. Wu CI, Ting CT. 2004. Genes and speciation. *Nat Rev Genet* **5**: 114–122.
83. Werth CR, Windham MD. 1991. A model for divergent, allopatric speciation of polyploid pteridophytes resulting from silencing of duplicate-gene expression. *Am Nat* **137**: 515–526.
84. Lynch M, Force AG. 2000. The origin of interspecies genomic incompatibility via gene duplication. *Am Nat* **156**: 590–605.
85. Lynch M. 2004. Gene duplication and evolution. In *Evolution: from Molecules to Ecosystems*, Moya A, Font E (eds). Oxford University Press: Oxford; 33–47.
86. Scannell DR, Byrne KP, Gordon JL, Wong S, Wolfe KH. 2006. Multiple rounds of speciation associated with reciprocal gene loss in polyploid yeasts. *Nature* **440**: 341–345.

DOI: 10.1002/yea