

## MicroReview

# Yeast genome sequencing: the power of comparative genomics

Jure Piškur<sup>1,2\*</sup> and Rikke B. Langkjær<sup>1</sup>

<sup>1</sup>BioCentrum-DTU, Building 301, Technical University of Denmark, DK-2800 Kgl. Lyngby, Denmark.

<sup>2</sup>Department of Cell and Organism Biology, Lund University, Lund, Sweden.

### Summary

**For decades, unicellular yeasts have been general models to help understand the eukaryotic cell and also our own biology. Recently, over a dozen yeast genomes have been sequenced, providing the basis to resolve several complex biological questions. Analysis of the novel sequence data has shown that the minimum number of genes from each species that need to be compared to produce a reliable phylogeny is about 20. Yeast has also become an attractive model to study speciation in eukaryotes, especially to understand molecular mechanisms behind the establishment of reproductive isolation. Comparison of closely related species helps in gene annotation and to answer how many genes there really are within the genomes. Analysis of non-coding regions among closely related species has provided an example of how to determine novel gene regulatory sequences, which were previously difficult to analyse because they are short and degenerate and occupy different positions. Comparative genomics helps to understand the origin of yeasts and points out crucial molecular events in yeast evolutionary history, such as whole-genome duplication and horizontal gene transfer(s). In addition, the accumulating sequence data provide the background to use more yeast species in model studies, to combat pathogens and for efficient manipulation of industrial strains.**

### Introduction

Yeasts represent a quite divergent group of fungi that exist

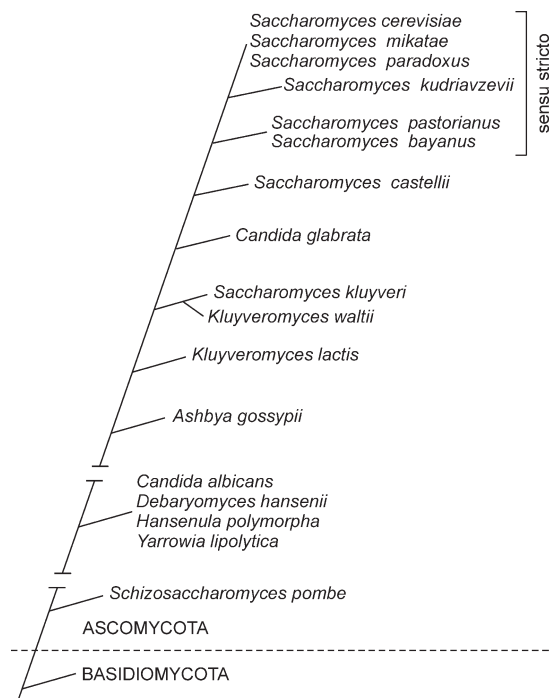
predominantly as unicellular organisms. They include important industrial organisms that have been used for many centuries, pathogens and popular laboratory organisms that serve as general models to understand the eukaryotic cell. For decades, *Saccharomyces cerevisiae*, baker's yeast, has been one of the best characterized organisms from the genetics and physiological point of view. The sequencing of the first eukaryotic chromosome, number III, was begun in the late 1980s by a consortium of yeast researchers and was completed in 1992 (Oliver *et al.*, 1992). This achievement can now be considered as one of the main milestones in molecular biology research and the birth of eukaryote genomics. In 1996, *S. cerevisiae* became the first completely sequenced eukaryote (Goffeau *et al.*, 1996), thus providing the main reference for comparison of any newly sequenced eukaryotic genomes. Recent progress in the sequencing of the genomes of a number of yeast species represents the heart of the exploding field of comparative genomics.

By the time the *S. cerevisiae* genome was almost completed, two other intensive yeast sequencing projects had already been started; the genomes of another important model organism, *Schizosaccharomyces pombe*, and an important human pathogen, *Candida albicans* (Tzung *et al.*, 2001; Wood *et al.*, 2002; <http://www-sequence.stanford.edu/group/candida>). However, the three yeast species are only distantly related to each other and therefore difficult to compare in any detail. In the late 1990s, smaller scale sequencing projects were carried out on rather close *S. cerevisiae* relatives (Ozier-Kalogeropoulos *et al.*, 1998; Langkjær *et al.*, 2000; Cliften *et al.*, 2001), providing a new window to look at various genomic aspects through comparison of very closely related sequences. Furthermore, the Génolevures project analysed in a relatively comprehensive way the genomes of 13 yeasts (Souciet *et al.*, 2000). Just a few months ago, the genomes of several *Saccharomyces* and *Kluyveromyces* species (Cliften *et al.*, 2003; Kellis *et al.*, 2003; 2004), that of the close relative of *Saccharomyces*, *Ashbya gossypii* (Brachat *et al.*, 2003; Dietrich *et al.*, 2004) and that of a related filamentous fungus, *Neurospora crassa* (Galagan *et al.*, 2003), were completely sequenced. In addition, several more or less complete genome sequences

Accepted 15 April, 2004. \*For correspondence at the first address. E-mail: jp@biocentrum.dtu.dk; Tel. (+45) 45 25 25 18; Fax (+45) 45 93 28 09.

that are increasingly relevant for comparative genomics are or will be available soon for the scientific community. These include that of the brewing yeast *Saccharomyces pastorianus/carlsbergensis* (Nakao *et al.*, 2003), the output from the Whitehead Institute sequencing efforts, the Fungal Genome Initiative (2003; <http://www.broad.mit.edu/annotation/fungi/fgi/>) (Fig. 1).

Sequencing of yeast genomes tells us directly about the coding potential and regulatory sequences that are necessary for a yeast (and eukaryotic) cell, but also about genes that are specific for a certain species. However, yeasts could also be used as a general model in comparative genomics because of the small genome size(s) and a well-defined evolutionary relationship among different lineages. A whole range of genomes from closely related yeasts is now available for developing a methodological approach to study these relationships in other organisms. In addition, well-characterized genomes provide a background to understand the molecular mechanisms involved in the generation of novel species and to deduce yeast



**Fig. 1.** A relative phylogenetic relationship of selected, extensively sequenced yeasts, based on Kurtzman and Robnett (1998; 2003). Ascomycota, including all the mentioned sequenced yeasts, diverged from Basidiomycota  $\approx$  550 million years ago. The speciation of the 'Saccharomyces complex', including *Saccharomyces* and *Kluveromyces* yeasts, started upon separation from the *A. gossypii* lineage more than 100 million years ago. A detailed taxonomic position of *C. albicans*, *D. hansenii*, *H. polymorpha* and *Y. lipolytica* regarding the 'Saccharomyces complex' is still not completely clear. *S. pastorianus* is actually a hybrid between *S. cerevisiae* and *S. bayanus*. The advantages of yeasts for comparative genomics studies are the small genome size(s) and a relatively well-defined evolutionary relationship among the different species.

evolutionary history. Comparison of closely related sequences is also helpful to deduce the so far 'hidden' sequence motifs that are important, for example, in gene expression. The present sequence data will be useful for breeding of industrial strains and combating yeast pathogens. Finally, newly sequenced yeasts are becoming future model organisms.

### Phylogenetic relationship

The assignment of organisms to different taxa has long been based on morphology and physiology (Kurtzman and Robnett, 1998). The recent application of gene sequence analysis to systematics has revealed that the previous analysis of phylogenetic relationships based only on phenotypic characters was often inconsistent. However, alternative phylogenetic relationships generated from single-gene data sets frequently generate incongruences (Maddison, 1997).

Assuming that the molecular clock for the 18S ribosomal DNA (rDNA) evolution has been constant through time, Ascomycota and Basidiomycota diverged  $\approx$  550 million years ago (Berbee and Taylor, 2001). This figure also assumes that horizontal transfer of genetic material did not occur frequently among sexually isolated yeast lineages. *Candida albicans* and *S. cerevisiae* had the last common ancestor  $\approx$  200 million years ago, and the speciation of *Saccharomyces* and *Kluveromyces* yeasts, upon divergence from the *Ashbya*-like yeast lineages, started  $\approx$  100–150 million years ago (Kurtzman and Robnett, 1998; Berbee and Taylor, 2001). However, the precise timing for separation of the deep branches is still very unclear. The short generation time of yeast, lack of fossils and uneven evolution rate of the analysed genes are the main problems in calibrating the molecular clock. The speciation and phylogenetic relationship of several *Saccharomyces/Kluveromyces* yeasts was studied recently on the basis of the divergence in genes of the rDNA repeat, three single copy nuclear genes and two mitochondrial gene sequences. While single-gene phylogenies were congruent for terminal lineages, the deeper branches on the phylogenetic tree were not clearly resolved. However, combined analysis of the whole data set resulted in a much better resolution (Kurtzman and Robnett, 2003). Similar analysis in other organisms of sets of a limited number of sequences has recently questioned the minimal amount of data that could be sufficient to recover a valid species tree (Soltis *et al.*, 1999). Recent progress in genome sequencing of *Saccharomyces* species (Cliften *et al.*, 2003; Kellis *et al.*, 2003) has presented an unprecedented opportunity to resolve this question.

Rokas *et al.* (2003) have selected 100 genes from eight *Saccharomyces* species for phylogenetic analysis, both singly and by concatenation, and confirmed that data sets

of single or small numbers of genes often gave conflicting topologies. However, consistent results were obtained when different sets of a minimum of 20 genes were analysed (Rokas *et al.*, 2003). This study is now an important guide on how to continue with molecular taxonomy studies and suggests that a partial genome sequencing may be sufficient to resolve phylogenetic relationships within a relatively defined group of organisms. Apparently, only the use of genome-wide data sets from a large number of organisms will help to reconstruct the precise historical relationships in the tree of life. In other words, intensified sequencing must be continued for a while to provide sufficient data sets for highly reliable phylogenetic studies.

### Speciation

A biological species is a group of organisms defined by their inability to mate successfully and produce viable offspring with other species. Barriers that prevent successful mating can be either prezygotic, blocking fertilization, or post-zygotic, and their origin in nature is still poorly understood. Yeasts with their well-characterized genomes now represent ideal models to understand the molecular background underlying speciation.

The 'Saccharomyces complex' has recently been resolved by multigene sequence analysis into 14 clades. The *Saccharomyces sensu stricto* yeasts, including *S. bayanus*, *S. cariocanus*, *S. cerevisiae*, *S. kudriavzevii*, *S. mikatae* and *S. paradoxus*, represent an isolated and well-supported monophyletic group with overall phenotypic similarity (Kurtzman and Robnett, 2003). These species can mate with each other, but interspecific pairings result predominantly in sterile hybrids (Naumov, 1996; Marinoni *et al.*, 1999), and interactions between the nuclear and mitochondrial genome might also be impaired (Sulo *et al.*, 2003). While the post-zygotic barrier already exists among *sensu stricto* species, an efficient prezygotic barrier is still absent and, therefore, this group of yeasts is still in the early stages of species formation. What could be the main driving force behind their speciation in particular and speciation in general?

Rearrangements within the nuclear genome have been common during yeast evolution. In the *Saccharomyces* complex, chromosome numbers and sizes have been changed (Langkjær *et al.*, 2000), chromosomal translocations have taken place (Fischer *et al.*, 2000), gene order remodelled and differential gene loss has taken place (Blandin *et al.*, 2000). Chromosome translocations found in *Saccharomyces sensu stricto* have been mapped previously and, surprisingly, the number of translocations relative to *S. cerevisiae* does not correlate with the sequence-based phylogeny (Fischer *et al.*, 2000). Therefore, chromosomal translocations in yeast might contribute to the reproductive isolation among *sensu stricto*

species, but are not the only cause of speciation. This hypothesis has been tested recently by Delneri *et al.* (2003) by re-engineering the yeast chromosome sets. Based on the availability of the genome sequences and the development of a method for generating precisely located chromosomal translocations, the authors developed strains of *S. cerevisiae* for studies on reproductive isolation. The corresponding genomes were collinear with those of other *sensu stricto* yeasts, but otherwise the genomic sequences exhibited 100% identity with the wild-type *S. cerevisiae* genome. The imposed collinearity allowed the generation of interspecific hybrids, in crosses among otherwise different species, which could produce viable but aneuploid spores (Delneri *et al.*, 2003). In other words, if the chromosomes were rearranged, the species barrier almost disappeared. Therefore, chromosome rearrangements within developing novel species play an important role during reproductive isolation and contribute to the speciation process. However, other processes might also contribute significantly to speciation.

Another important point in the above work (Delneri *et al.*, 2003) is the existence of a mechanism for chromosome quality control following inter-species crosses (Wolfe, 2003). While hybrid zygotes formed among *sensu stricto* species preserve both parental chromosome sets, hybrid zygotes obtained from crosses among less related *Saccharomyces* species tend to eliminate most of the chromosomes from one of the parents (Marinoni *et al.*, 1999). Similarly, spores produced by *sensu stricto* inter-species hybrids often retain the complete set of chromosomes from one parent and about half the chromosomes from the other (Delneri *et al.*, 2003). Apparently, the presence of the complete genome of only one parent (aneuploidy) provides a way to overcome the meiotic problems caused by translocation differences. New genome sequences from *Saccharomyces* and other related yeasts will now increase the opportunities for further experiments on chromosome stability and species barriers, such as searching for specific genes that are directly involved in these processes.

### Genes and regulatory motifs

Large-scale comparisons of genomes address basic genomics questions, such as the number of functional genes, identification of species-specific genes, distribution of genes among functional families, gene density, preservation of gene order, mechanisms of genome reshuffling, the rate of sequence divergence, etc. In general, the choice of species that mark evolutionary distances for comparative genomics should depend on the aim of the analysis. One should also take into account that some sequences evolve at a faster rate and that some, especially functional ones, evolved more slowly (Frazer *et al.*,

2003). The recent yeast comparative approach focused on both well-conserved and less-preserved sequences by comparing several species with a variable degree of evolutionary relationship. Two benefits of this approach were that it improved the annotation of the already sequenced *S. cerevisiae* genome and predicted novel shorter sequence elements that might be involved in gene expression (Cliften *et al.*, 2003; Kellis *et al.*, 2003). Such sequences are usually difficult to recognize because they are short and highly degenerate and are found at variable positions.

When annotating protein-coding genes in the genome of *S. cerevisiae*, it has generally been accepted to look at open reading frames (ORFs) longer than 100 codons. By accepting only one of two overlapping ORFs, usually the longer one, a large number of spurious ORFs have therefore been eliminated. Several different approaches have been used to analyse the ≈2000 remaining hypothetical ORFs, such as DNA composition and/or homology between closely related species. For several years, the predicted gene count for *S. cerevisiae* exceeded 6000 (Goffeau *et al.*, 1996). However, recent comparative studies (Brachat *et al.*, 2003; Cliften *et al.*, 2003; Kellis *et al.*, 2003) suggest that up to 500 ORFs that were initially annotated as hypothetical are actually spurious. Besides lacking reading frame conservation in related species (i.e. no homologue), most of the spurious ORFs involved uncharacterized ORFs, of which many overlapped with another well-conserved ORF, but had insertions and deletions. In addition, some were small and had atypical codon usage. Thus, *S. cerevisiae* apparently contains only 5700–5800 functional reading frames (Cliften *et al.*, 2003; Kellis *et al.*, 2003). Furthermore, Cebrat's group has estimated the number of protein-coding ORFs to be as low as 5300–5400 (Mackiewicz *et al.*, 2002). They proposed that the majority of the non-coding ORFs have arisen by gene duplication followed by mutations that threw the bulk of the originally duplicated copy out of frame (Mackiewicz *et al.*, 1999). Just now, the complete genome sequences of *A. gossypii* (Dietrich *et al.*, 2004) and *Kluyveromyces waltii* (Kellis *et al.*, 2004) revealed that the number of protein-coding genes, ≈4718 and 5230, respectively, in these two species is notably smaller than the accepted 5700–5800 encoded by the *S. cerevisiae* genome. Comparing the number of genes in *A. gossypii* and *K. waltii* with the fact that almost 500 of the duplicated genes are kept in duplicate in *S. cerevisiae* (Wolfe and Shields, 1997) reveals that there is still some controversy regarding the gene number in *S. cerevisiae*. The recent changes to the *S. cerevisiae* annotation also involved discovery of novel genes, including several shorter genes, novel introns, ORF extensions and fusions of neighbouring ORFs (Brachat *et al.*, 2003; Cliften *et al.*, 2003; Kellis *et al.*, 2003). Some of the changes have been supported

by additional experiments; a few of the novel introns have been verified by 5' rapid amplification of cDNA ends (5' RACE) (Brachat *et al.*, 2003), 84 novel small ORFs (smORFs) have been confirmed by transcription products (Kessler *et al.*, 2003) as well as several errors in the original *S. cerevisiae* genome sequence have been confirmed by resequencing of smaller regions of the genome (Brachat *et al.*, 2003; Kellis *et al.*, 2003). In short, the concise and reliable picture of the *S. cerevisiae* protein-coding potential should soon be available.

Based on comparative genomics, non-coding RNA genes have been found in the analysed genomes, and a number of regulatory sequences that map outside the genes they regulate have been identified. Even if these sequences are non-coding, their relatively strong preservation suggests that they play a functional role. If the evolutionary distance among the compared species is too short, functional and non-functional parts can show a similar high degree of identity, and therefore the regulatory sequences can be 'lost' in the background. This problem was overcome by analysis of *Saccharomyces* genomes with different degrees of evolutionary relationship, which identified a number of novel short sequence motifs that might be recognized by regulatory proteins (Cliften *et al.*, 2003; Kellis *et al.*, 2003). Several of these novel motifs are located immediately upstream of genes coding for proteins belonging to similar functional groups, as well as in front of genes of unknown function. These genes can now be grouped together on the basis of their shared putative regulatory motifs. These relationships provide a new approach to search for functions of orphan genes. One can imagine that this approach might now play a major role in gene expression studies in other eukaryotes. However, a major problem to be overcome is that genomes of higher eukaryotes consist predominantly of non-functional sequences that are difficult to align, and the regulatory motifs might map far away from the genes they control.

## Evolution

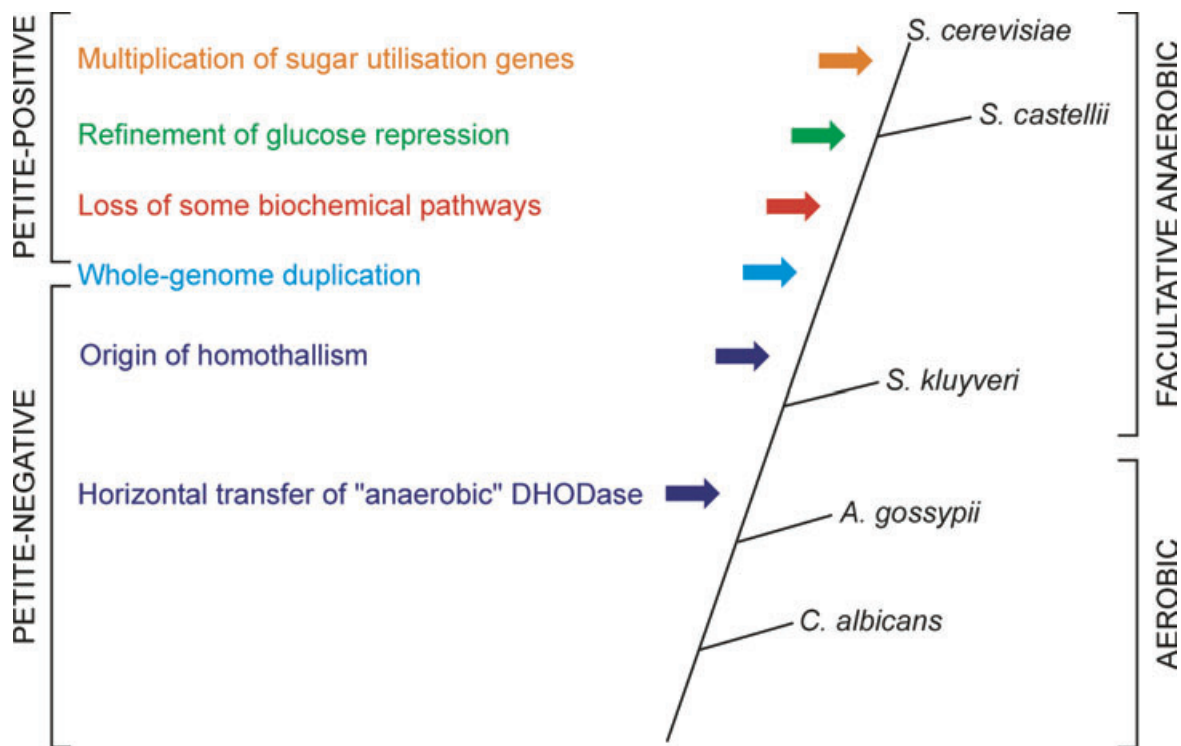
Yeasts belonging to the *Saccharomyces* complex have a number of unique characters not found in other yeast genera (Kurtzman and Fell, 1998). For example, *Saccharomyces sensu stricto* yeasts, as well as other *Saccharomyces* members, primarily degrade hexoses only to the C<sub>3</sub> and C<sub>2</sub> compounds pyruvate and ethanol, even in the presence of oxygen. This phenomenon relies on a 'glucose repression' circuit that represses the respiratory part in the presence of glucose (Johnston, 1999). The occurrence of fermentation under aerobic conditions is sometimes referred to as the Crabtree effect and the yeasts exhibiting it as Crabtree-positive yeasts (Pronk *et al.*, 1996). While a majority of yeasts cannot grow in the

absence of oxygen (aerobic yeasts), a majority of the *Saccharomyces* complex yeasts can also survive without any oxygen (Andreasen and Stier, 1953; Subik *et al.*, 1974; Pronk *et al.*, 1996; Møller *et al.*, 2001). The life cycle of the *Saccharomyces* complex yeasts is also very unique. Sexually reproducing yeasts can undergo mating between either heterothallic lines, which are self-sterile, or homothallic lines, which are self-fertile. In *S. cerevisiae*, homothallism occurs because cells can switch mating type; specifically, a maternal cell can, upon budding, switch its mating type and then mate with its daughter (Strathern and Herskowitz, 1979). The presence of the *HO* gene is responsible for the homothallic behaviour, and *ho* mutants are heterothallic. A site-specific nuclease, encoded by *HO*, represents a central element in the mating type switching (Haber, 1998). In summary, an interesting evolutionary question is when and how did the progenitor of *Saccharomyces* yeasts develop these basic characters and what were the molecular mechanisms operating during this yeast's evolutionary history?

The complete sequence of the *S. cerevisiae* genome, as well as those of additional yeast sequences, has provided a unique tool to study the origin of the modern yeast

traits (Fig. 2). Two molecular mechanisms, whole-genome duplication and horizontal gene transfer, are proposed to play a major role in the evolutionary history of the *Saccharomyces* complex yeasts.

Duplications are believed to play one of the major roles in biological evolution by providing extra genetic material, which can subsequently be remodelled into 'novel' gene products (Ohno, 1970). Eukaryote genomes generally show a high degree of redundancy. In several organisms, including the fruit fly *Drosophila melanogaster*, the nematode worm *Caenorhabditis elegans* and *C. albicans*, most duplicated genes are dispersed, whereas in *S. cerevisiae*, as well as in humans and the plant *Arabidopsis thaliana*, the duplicated genes coincide with large segmental duplications (Wolfe and Shields, 1997). Apparently, several independent groups of organisms have used whole-genome duplications as a source of novel genes. Recent analyses using different yeast genome sequences showed that almost the entire genome lies in duplicated sister regions, clearly demonstrating that the entire genome became duplicated at some point, followed by rearrangements and gene loss (Wong *et al.*, 2002). Analysis of duplicated gene sequence sets from different



**Fig. 2.** The origin of several modern yeast traits. The whole-genome duplication (Wolfe and Shields, 1997; Kellis *et al.*, 2004) and horizontal transfer of genetic material (Butler *et al.*, 2004; Gojković *et al.*, 2004) provided new genes, which became the background for the development of facultative anaerobic lifestyle (Pronk *et al.*, 1996; Møller *et al.*, 2001), homothallism (Butler *et al.*, 2004) and efficient glucose repression circuit (Johnston, 1999). The precise origin of refined glucose repression (before or after the split between the *S. castellii* and *S. cerevisiae* lineages) is still unclear. Loss of single biochemical pathways and the corresponding genes (such as the pyrimidine catabolic pathway; Gojkovic *et al.*, 2000) could have taken place independently and at several time points. Comparative genomics now helps to place these events at different branching points of the yeast phylogenetic tree and estimates the relative timing of these events (for further details, see the text).

'*Saccharomyces complex*' yeasts has shown that, upon duplication, each duplicated gene pair has been modified independently through specialization, differentiation or deletion of a single copy and through the evolution of the corresponding yeast lineages (Langkjær *et al.*, 2003). As mentioned previously, *A. gossypii* and *K. waltii*, divergence of which from *S. cerevisiae* precedes the duplication event, have been sequenced. These yeasts have now been used to deduce the ancient gene order and the timing of the whole-genome duplication event (Dietrich *et al.*, 2004; Kellis *et al.*, 2004). Upon duplication and massive gene loss,  $\approx 10\%$  of the duplicated genes have been preserved, and often one member of each duplicated pair has evolved rapidly into a 'novel' gene with a derived function. Apart from the whole-genome duplication, several other genes have been duplicated independently. These relatively recent duplications usually map to the telomeric regions, which apparently act as 'nests' for newly duplicated sequences and often encompass genes involved in sugar uptake and metabolism, such as the *SUC* and *MEL* genes (Carlson *et al.*, 1985; Greig and Travisano, 2004). Further understanding of telomere dynamics and its role in the birth of genes is likely to represent an important research focus in the coming years.

The segmental duplication, which occurred  $\approx 100$ – $150$  million years ago (Fig. 2), apparently provided new genes, which were the basis for major remodelling of metabolism, including the development of an efficient glucose repression pathway and oxygen independence, in *Saccharomyces complex* yeasts. For example, some of the duplicated genes have 'remodelled' their expression to become dependent on the presence/absence of oxygen and glucose (Kwast *et al.*, 2002). This event might have coincided with the radiation and subsequent specialization of the modern fruit-bearing plants,  $\approx 100$ – $200$  million years ago and, consequently, the larger available quantity of free mono- and oligosaccharides in nature. The developing yeast lineages apparently had a competitive advantage: they could grow fast with or without oxygen and they preferentially produced ethanol, which was toxic for their bacterial competitors. Similar yeast–bacterial interactions and competition through removal of oxygen and production of ethanol can be found nowadays in grape-wine ecosystems (Fleet, 2003).

Horizontal transfer of bacterial genes contributed to the fine tuning of oxygen independence and the origin of efficient homothallism. One of the crucial requirements for facultative anaerobiosis in yeast is the independence of the *de novo* pyrimidine biosynthesis, more precisely the fourth enzymic activity catalysed by dihydroorotate dehydrogenase (DHODase), from the active respiratory chain (Nagy *et al.*, 1992). While a majority of yeasts has a mitochondrial DHODase, *S. cerevisiae* has a cytoplasmic

enzyme, independent of the presence of oxygen. Apparently, the gene for this enzyme, which is closely related to the bacterial *Lactococcus lactis* gene coding for DHODase A, was adopted before the *S. cerevisiae* and *S. kluyveri* lineages separated (Fig. 2; Gojković *et al.*, 2004) and promoted gradual independence from the presence of oxygen. Homothallism, as known in the modern *Saccharomyces sensu stricto* yeasts, apparently originated in a similar way, through acquisition of an intein-like sequence, a selfish mobile element, which can get inserted in frame into host genes (Keeling and Roger, 1995). After separation of the *S. cerevisiae* and *S. kluyveri* lineages, an intein invaded the yeast *VMA1* gene from an unknown source and subsequently duplicated and gave rise to *HO* (Fig. 2; Butler *et al.*, 2004). Efficient homothallic switching as seen in *S. cerevisiae* is the background for a predominantly diploid lifestyle found among modern *Saccharomyces* yeasts. The *HO* gene product promotes the mating-type switch already after the first division of the haploid cells originating from meiotic spores. Subsequently, the mother and daughter cells mate and generate a diploid cell lineage (Haber, 1998). Diploidization has increased the level of robustness of the yeast genome and promoted the evolution of a repair system based on efficient homologous recombination. However, several aspects of the evolution of homothallism and related processes are still poorly understood.

### Practical perspectives

Yeasts have been widely used for millennia as cell factories. Some of the oldest products are alcoholic beverages, such as beer and wine, and bread. Later on, the processed food products were joined by vitamins, organic acids, lipids and recently also heterologous proteins, such as insulin, growth hormone, vaccines, etc. (Rose and Harrison, 1993). Several yeasts are also pathogenic for humans, domesticated animals and crops (Kurtzman and Fell, 1998). Already, the first three genomic sequences, from *S. cerevisiae*, *S. pombe* and *C. albicans*, were important for the understanding of the yeast applied potential in general. The accumulating sequence data now provide an additional tool to study a number of yeasts, either directly by availability of the genomic sequence of the yeast in question or indirectly using a comparative approach. In addition, the availability of the sequence data from so far rarely studied yeasts has, because of their interesting traits, already turned these into novel model organisms.

Lager-brewing yeast strains are natural hybrids between *S. cerevisiae* and another *Saccharomyces sensu stricto* species. Owing to the fact that more than half the genome has been unknown and because of the complicated genetic structure, application of genetic and molecular biology techniques has been challenging

(Nilsson-Tillgren *et al.*, 1981; Kielland-Brandt *et al.*, 1995). Recently, one of the lager strains has been sequenced at the Suntory Breweries providing a new tool to optimize and manipulate the behaviour of these yeasts (Nakao *et al.*, 2003). In addition, the complex allopolyploid nature and co-existence of homeologous and hybrid chromosomes will continue to be a challenge for chromosome structure and recombination studies.

Even though *A. gossypii* is a close relative of the *Saccharomyces* yeasts (Fig. 1), it is very different with respect to growth form and habitat. *A. gossypii* is a filamentous fungus growing as a multinucleated branched hyphal network and therefore an emerging model organism to study filamentous growth. In addition, this yeast is a cotton and citrus pathogen. The recently released genome sequence from the Basel group (Brachat *et al.*, 2003; Dietrich *et al.*, 2004) will undoubtedly have an enormous impact on our understanding of the evolutionary history of yeasts, as well as the switch between unicellular and hyphal growth.

The Génolevures consortium has recently completed the genome sequences of three industrially important yeasts and the second most common human pathogen causing candidiasis, *Candida glabrata* (Sherman *et al.*, 2004; <http://cbi.labri.fr/Genolevures>). *Kluyveromyces lactis* has already for decades been a popular laboratory organism and a successful production strain for heterologous proteins such as chymosin, an active agent of the cheese rennet. *Debaryomyces hansenii* is an osmotolerant yeast that can assimilate hydrocarbons, and so can *Yarrowia lipolytica*. *C. glabrata* is becoming an important organism for understanding the switch between sexual and asexual reproduction among yeasts (Wong *et al.*, 2003). The genome sequence of *C. glabrata* will now help to identify species-specific genes and thereby help to develop species-specific drugs for this increasingly frequent pathogen. Ramezani-Rad *et al.* (2003) have recently sequenced the genome of another industrial yeast species, the methylotrophic *Hansenula polymorpha*, also frequently used as a model for peroxysome studies. While *Y. lipolytica* and *H. polymorpha* were studied intensively for their physiology in the past, the genome sequence now opens further possibilities for gene modification in these yeasts and the generation of novel mutant strains with increased production potential.

Two other recently sequenced yeasts, *S. castellii* and *S. kluyveri* (Cliften *et al.*, 2003), have just started to be used as model organisms for some specialized traits. The homogeneous telomeric repeats have made *S. castellii* an emerging model for studying the maintenance of telomeric DNA (Wahlin *et al.*, 2003). *S. kluyveri*, which is a facultative anaerobe but petite-negative yeast, is used in comparison with *S. cerevisiae* to understand glucose repression (Møller *et al.*, 2001) and to study the degradation of nucleic acids precursors (Gojkovic *et al.*, 2000).

The *S. castellii* and *S. kluyveri* total genome sequences will provide an additional platform for their increasing popularity among yeast researchers. In particular, *S. castellii* and its carbon metabolism (Fig. 2) would deserve to be studied in the near future.

In conclusion, yeast has for millennia been one of man's favourite organisms, providing wine, beer and bread. During the last several decades, yeast has increasingly become a central model organism to understand different biological aspects of the eukaryotic cell. Now yeasts are at the cutting edge of comparative genomics, providing a unique model and opportunity for the development of new tools to understand other eukaryotic and prokaryotic genomes. However, one of the crucial technical problems to be solved in the future is the availability of the data and how to organize the yeast genome sequences into a single database or at least databases that can successfully communicate with each other.

### Acknowledgements

The authors would like to thank Morten Kielland-Brandt for reading and commenting on an early version of this manuscript. Several laboratories have kindly provided their results before publication, and apologies to all of those whose contribution to the yeast sequencing database has not been mentioned in this paper.

### References

- Andreasen, A.A., and Stier, T.J.B. (1953) Anaerobic nutrition of *Saccharomyces cerevisiae*. I. Ergosterol requirement for the growth in a defined medium. *J Cell Comp Physiol* **41**: 23–36.
- Berbee, M.L., and Taylor, J.W. (2001) Systematics and evolution. In *The Mycota VII B*. Mclaughlin, D.J., Mclaughlin, E.G., and Lemke, P.A. (eds). Berlin: Springer, pp. 229–245.
- Blandin, P., Durrens, P., Tekaiia, F., Aigle, M., Bolotin-Fukuhara, M., Bon, E., *et al.* (2000) Genomic exploration of the hemiascomycetous yeasts. 4. The genome of *Saccharomyces cerevisiae* revisited. *FEBS Lett* **487**: 31–36.
- Brachat, S., Dietrich, F.S., Voegeli, S., Zhang, Z., Stuart, L., Lerch, A., *et al.* (2003) Reinvestigation of the *Saccharomyces cerevisiae* genome annotation by comparison to the genome of a related fungus: *Ashbya gossypii*. *Genome Biol* **4**: R45.
- Butler, G., Kenny, C., Fagan, A., Kurischko, C., Gaillardin, C., and Wolfe, K.H. (2004) Evolution of the *MAT* locus and its Ho endonuclease in yeast species. *Proc Natl Acad Sci USA* **101**: 1632–1637.
- Carlson, M., Celenza, J.L., and Eng, F.J. (1985) Evolution of the dispersed *SUC* gene family of *Saccharomyces* by rearrangements of chromosome telomeres. *Mol Cell Biol* **5**: 2894–2902.
- Cliften, P.F., Hillier, L.W., Fulton, L., Graves, T., Miner, T., Gish, W.R., *et al.* (2001) Surveying *Saccharomyces* genomes to identify functional elements by comparative DNA sequence analysis. *Genome Res* **11**: 1175–1186.

- Cliften, P., Sudarsanam, P., Desikan, A., Fulton, L., Fulton, B., Majors, J., *et al.* (2003) Finding functional features in *Saccharomyces* genomes by phylogenetic footprinting. *Science* **301**: 71–76.
- Delneri, D., Colson, I., Grammenoudi, S., Roberts, I.N., Louis, E.J., and Oliver, S.G. (2003) Engineering evolution to study speciation in yeasts. *Nature* **422**: 68–72.
- Dietrich, F.S., Voegeli, S., Brachat, S., Lerch, A., Gates, K., Steiner, S., *et al.* (2004) The *Ashbya gossypii* genome as a tool for mapping the ancient *Saccharomyces cerevisiae* genome. *Science* **304**: 304–307.
- Fischer, G., James, S.A., Roberts, I.N., Oliver, S.G., and Louis, E.J. (2000) Chromosomal evolution in *Saccharomyces*. *Nature* **405**: 451–454.
- Fleet, G.H. (2003) Yeast interactions and wine flavour. *Int J Food Microbiol* **86**: 11–22.
- Frazer, K.A., Elnitski, L., Church, D.M., Dubchak, I., and Hardison, R.C. (2003) Cross-species sequence comparisons: a review of methods and available resources. *Genome Res* **13**: 1–12.
- Fungal Genome Initiative (2003) *A White Paper for Fungal Comparative Genomics*, 10 June, 2003. [WWW document] URL <http://www.broad.mit.edu/annotation/fungi/fgi/>.
- Galagan, J.E., Calvo, S.E., Borkovich, K.A., Selker, E.U., Read, N.D., Jaffe, D., *et al.* (2003) The genome sequence of the filamentous fungus *Neurospora crassa*. *Nature* **422**: 859–868.
- Goffeau, A., Barrell, B.G., Bussey, H., Davis, R.W., Dujon, B., Feldmann, H., *et al.* (1996) Life with 6000 genes. *Science* **274**: 563–567.
- Gojkovic, Z., Jahnke, K., Schnackerz, K.D., and Piškur, J. (2000) *PYD2* encodes 5,6-dihydropyrimidine amidohydrolyase, which participates in a novel fungal catabolic pathway. *J Mol Biol* **295**: 1073–1087.
- Gojković, Z., Knecht, W., Zameitat, E., Warneboldt, J., Coutelis, J.-B., Pynyaha, Y., *et al.* (2004) Horizontal gene transfer promoted the evolution of the yeast ability to propagate under anaerobic conditions. *Mol Genet Genomics* published online 11 March, 2004 [DOI: 10.1007/s00438-004-0995-7].
- Greig, D., and Travisano, M. (2004) The prisoner's dilemma and polymorphism in yeast *SUC* genes. *Proc R Soc London B Suppl* **271**: S25–S26.
- Haber, J.E. (1998) Mating-type gene switching in *Saccharomyces cerevisiae*. *Annu Rev Genet* **32**: 561–599.
- Johnston, M. (1999) Feasting, fasting and fermenting. Glucose sensing in yeast and other cells. *Trends Genet* **15**: 29–33.
- Keeling, P.J., and Roger, A.J. (1995) The selfish pursuit of sex. *Nature* **375**: 283.
- Kellis, M., Patterson, N., Endrizzi, M., Birren, B., and Lander, E.S. (2003) Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* **423**: 241–254.
- Kellis, M., Birren, B.W., and Lander, E.S. (2004) Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* **428**: 617–624.
- Kessler, M.M., Zeng, Q., Hogan, S., Cook, R., Morales, A.J., and Cottarel, G. (2003) Systematic discovery of new genes in the *Saccharomyces cerevisiae* genome. *Genome Res* **13**: 264–271.
- Kielland-Brandt, M.C., Nilsson-Tillgren, T., Gjermansen, C., Holmberg, S., and Pedersen, M.B. (1995) Genetics of brewing yeasts. In *The Yeasts*, Vol. 6, 2nd edn. Wheals, A.E., Rose, A.H., and Harrison, J.S. (eds). London: Academic Press, pp. 223–254.
- Kurtzman, C.P., and Fell, J.W. (eds) (1998) *The Yeasts. A Taxonomic Study*, 4th edn. Amsterdam: Elsevier.
- Kurtzman, C.P., and Robnett, C.J. (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, C.P., and Robnett, C.J. (2003) Phylogenetic relationships among yeasts of the '*Saccharomyces complex*' determined from multigene sequence analyses. *FEMS Yeast Res* **3**: 417–432.
- Kwast, K.E., Lai, L.-C., Menda, N., James, D.T., III, Aref, S., and Burke, P.V. (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.
- Langkjær, R.B., Nielsen, M.L., Daugaard, P.R., Liu, W., and Piškur, J. (2000) Yeast chromosomes have been significantly reshaped during their evolutionary history. *J Mol Biol* **304**: 271–288.
- Langkjær, R.B., Cliften, P.F., Johnston, M., and Piškur, J. (2003) Yeast genome duplication was followed by asynchronous differentiation of duplicated genes. *Nature* **421**: 848–852.
- Mackiewicz, P., Kowalczyk, M., Gierlik, A., Dudek, M.R., and Cebrat, S. (1999) Origin and properties of non-coding ORFs in the yeast genome. *Nucleic Acids Res* **27**: 3503–3509.
- Mackiewicz, P., Kowalczyk, M., Mackiewicz, D., Nowicka, A., Dudkiewicz, M., Laszkiewicz, A., *et al.* (2002) How many protein-coding genes are there in the *Saccharomyces cerevisiae* genome? *Yeast* **19**: 619–629.
- Maddison, W.P. (1997) Gene trees and species trees. *Syst Biol* **46**: 523–536.
- Marinoni, G., Manuel, M., Petersen, R.F., Hvidtfeldt, J., Sulo, P., and Piškur, J. (1999) Horizontal transfer of genetic material among *Saccharomyces* yeasts. *J Bacteriol* **181**: 6488–6496.
- Møller, K., Olsson, L., and Piškur, J. (2001) Ability for anaerobic growth is not sufficient for development of the petite phenotype in *Saccharomyces kluyveri*. *J Bacteriol* **183**: 2485–2489.
- Nagy, M., Lacroute, F., and Thomas, D. (1992) Divergent evolution of pyrimidine biosynthesis between anaerobic and aerobic yeasts. *Proc Natl Acad Sci USA* **89**: 8966–8970.
- Nakao, Y., Kodama, Y., Nakamura, N., Ito, T., Hattori, M., Shiba, T., and Ashikari, T. (2003) Whole genome sequence of a lager brewing yeast. *Proceedings of the 29th European Brewery Convention Congress*, Dublin, pp. 524–530.
- Naumov, G.I. (1996) Genetic identification of biological species in the *Saccharomyces sensu stricto* complex. *J Ind Microbiol* **17**: 295–302.
- Nilsson-Tillgren, T., Gjermansen, C., Kielland-Brandt, M.C., Petersen, J.G.L., and Holmberg, S. (1981) Genetic differences between *Saccharomyces carlsbergensis* and *S. cer-*

- visiae*. Analysis of chromosome III by single chromosome transfer. *Carlsberg Res Commun* **46**: 65–76.
- Ohno, S. (1970) *Evolution by Gene Duplication*. London: George Allen & Unwin.
- Oliver, S.G., van der Aart, Q.J., Agostoni-Carbone, M.L., Aigle, M., Alberghina, L., Alexandraki, D., *et al.* (1992) The complete DNA sequence of yeast chromosome III. *Nature* **357**: 38–46.
- Ozier-Kalogeropoulos, O., Malpertuy, A., Boyer, J., Tekaiia, F., and Dujon, B. (1998) Random exploration of the *Kluyveromyces lactis* genome and comparison with that of *Saccharomyces cerevisiae*. *Nucleic Acids Res* **26**: 5511–5524.
- Pronk, J.T., Steensma, H.Y., and van Dijken, J.P. (1996) Pyruvate metabolism in *Saccharomyces cerevisiae*. *Yeast* **12**: 1607–1633.
- Ramezani-Rad, M., Hollenberg, C.P., Lauber, J., Wedler, H., Griess, E., Wagner, C., *et al.* (2003) The *Hansenula polymorpha* (strain CBS4732) genome sequencing and analysis. *FEMS Yeast Res* **4**: 207–215.
- Rokas, A., Williams, B.L., King, N., and Carroll, S.B. (2003) Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* **425**: 798–804.
- Rose, A.H., and Harrison, A.H. (eds) (1993) *The Yeasts*, 2nd edn, Vol. 5, *Yeast Technology*. London: Academic Press.
- Sherman, D., Durrens, P., Beyne, E., Nikolski, M., Souciet, J.L., for the Génolevures Consortium (2004) Génolevures: comparative genomics and molecular evolution of hemiascomycetous yeasts. *Nucleic Acids Res* **32**: D315–D318.
- Soltis, P.S., Soltis, D.E., and Chase, M.W. (1999) Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology. *Nature* **402**: 402–404.
- Souciet, J., Aigle, M., Artiguenave, F., Blandin, G., Bolotin-Fukuhara, M., Bon, E., *et al.* (2000) Génolevures: Genomics exploration of the hemiascomycetous yeasts. *FEBS Lett* **487**: 1–149.
- Strathern, J.N., and Herskowitz, I. (1979) Asymmetry and directionality in production of new cell types during clonal growth: the switching pattern of homothallic yeast. *Cell* **17**: 371–381.
- Subik, J., Kolarov, J., and Kovac, L. (1974) Anaerobic growth and formation of respiration-deficient mutants of various species of yeasts. *FEBS Lett* **45**: 263–266.
- Sulo, P., Špírek, M., Šoltéssová, A., Marinoni, G., and Piškur, J. (2003) The efficiency of functional mitochondrial replacement in *Saccharomyces* species has directional character. *FEMS Yeast Res* **4**: 97–104.
- Tzung, K.W., Williams, R.M., Scherer, S., Federspiel, N., Jones, T., Hansen, N., *et al.* (2001) Genomic evidence for a complete sexual cycle in *Candida albicans*. *Proc Natl Acad Sci USA* **98**: 3249–3253.
- Wahlin, J., Rosen, M., and Cohn, M. (2003) DNA binding and telomere length regulation of yeast RAP1 homologues. *J Mol Biol* **332**: 821–833.
- Wolfe, K. (2003) Speciation reversal. *Nature* **422**: 25–26.
- Wolfe, K.H., and Shields, D.C. (1997) Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* **387**: 708–713.
- Wong, S., Butler, G., and Wolfe, K.H. (2002) Gene order evolution and paleopolyploidy in hemiascomycete yeasts. *Proc Natl Acad Sci USA* **99**: 9272–9277.
- Wong, S., Fares, M.A., Zimmermann, W., Butler, G., and Wolfe, K.H. (2003) Evidence from comparative genomics for a complete sexual cycle in the ‘asexual’ pathogenic yeast *Candida glabrata*. *Genome Biol* **4**: R10.
- Wood, V., Gwilliam, R., Rajandream, M.A., Lyne, M., Lyne, R., Stewart, A., *et al.* (2002) The genome sequence of *Schizosaccharomyces pombe*. *Nature* **415**: 871–880.