View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Hemeroteca Cientifica Catalana

Int Microbiol (2003) 6: 183–190
DOI 10.1007/s10123-003-0132-1

REVIEW ARTICLE

Enrique Herrero · María Angeles de la Torre
Eulogio Valentín

Comparative genomics of yeast species: new insights into their biology

Received: 16 January 2003 / Accepted: 5 June 2003 / Published online: 29 July 2003
© Springer-Verlag and SEM 2003

Abstract The genomes of two hemiascomycetous yeasts (*Saccharomyces cerevisiae* and *Candida albicans*) and one archiascomycete (*Schizosaccharomyces pombe*) have been completely sequenced and the genes have been annotated. In addition, the genomes of 13 more Hemiascomycetes have been partially sequenced. The amount of data thus obtained provides information on the evolutionary relationships between yeast species. In addition, the differential genetic characteristics of the microorganisms explain a number of distinctive biological traits. Gene order conservation is observed between phylogenetically close species and is lost in distantly related species, probably due to rearrangements of short regions of DNA. However, gene function is much more conserved along evolution. Compared to *S. cerevisiae* and *S. pombe*, *C. albicans* has a larger number of specific genes, i.e., genes not found in other organisms, a fact that can account for the biological characteristics of this pathogenic dimorphic yeast which is able to colonize a large variety of environments.

Keywords *Saccharomyces cerevisiae* · *Candida albicans* · *Schizosaccharomyces pombe* · Hemiascomycetes · Comparative genomics

E. Herrero (✉) · M. A. de la Torre
Department of Basic Medical Sciences,
University of Lleida, Faculty of Medicine,
25198 Lleida, Spain
E-mail: enric.herrero@cmb.udl.es
Tel.: +34-973702409
Fax: +34-973702426

E. Valentín
Department of Microbiology and Ecology,
University of Valencia, Faculty of Pharmacy,
46003 Valencia, Spain

Introduction

The increasing number of completely sequenced microbial genomes provides the opportunity to compare different species (or strains of the same species) on a genome scale. Thus, comparative genomics yields fundamental insights into the biological basis of microbial virulence, the metabolic abilities of microorganisms and their ecological adaptations, as well as evolutionary relationships between microorganisms [28, 29]. Also, phylogenetic profiles, i.e., the pattern of co-occurrence of genes across genomes, may reveal the function of proteins that show evolutionary co-occurrence with other proteins of known function [27]. Although comparative genomics cannot substitute for more focused approaches, it may help to discern experimental strategies for the functional characterization of biological processes.

Until now, only three fungal genomes have been completely sequenced and annotated: *Saccharomyces cerevisiae* [13], *Schizosaccharomyces pombe* [38] and *Candida albicans* (http://genomeweb.pasteur.fr/lfrangeu/ca). However, the Génolevures French consortium has partially sequenced 13 Hemyascomycetes yeast species [32]. While the amount of information on yeast whole-genomes is small compared with Bacteria and Archaea, it already can provide insights into the evolution of yeast species and help to explain the genetic basis of biological properties of species such as *C. albicans* that previously remained obscure. In this review, we will introduce the comparative genomics of the above yeast species and give examples of how the plasticity of yeast genomes may be an evolutionary force in the acquisition of new functions.

The genome of *Saccharomyces cerevisiae*

Publication of the sequence of the *S. cerevisiae* genome in 1996 (based on the work of a European Union-USA consortium of more than 100 laboratories) revealed

some interesting facts [13]. A total of 6,275 open reading frames (ORFs) were predicted on the basis of being larger than 99 amino acids and not being included in larger ORFs. Due to their low codon usage, it was estimated that 390 of these ORFs were not translated into proteins, which left a total of 5,855 ''real'' genes. This genetic information is distributed in 16 distinct chromosomes with a total of 13.4 megabases (Mb) (mitochondrial genes are not included in these calculations). Importantly, about one third of the sequenced genes were described as ''orphan'' genes [10], that is, genes without known function, since the sequence of the translation product lacked significant homology to any other protein of *S. cerevisiae* or other organisms with assigned or proposed function. This high proportion of orphan genes was surprising since *S. cerevisiae* had been subjected to extensive research for many years. Later analyses have slightly modified these numbers. At this moment (December 2002), the PEDANT database (http://pedant.gsf.de) assigns 6,449 ORF entries for the *S. cerevisiae* genome (Table 1), 98.3% of which have sequence homologues in other organisms. This large number of known homologues would be the consequence of the amount of prokaryotic and eukaryotic genome sequences recently completed and in the course of completion, which adds a huge amount of sequence information to the databases. However, these sequence data are not accompanied by the same amount of functional data, and on February 2002 1,003 *S. cerevisiae* ORFs were classified as similar to proteins of unknown function, while 516 were non-similar to any other protein (http://mips/gfs.de/yeast/tables). Most of the latter probably are questionable ORFs. In summary, about one sixth of the *S. cerevisiae* are still function orphans.

Those genes with known or reasonable predictable function have been divided into categories in the MIPS database (http://mips/gfs.de/yeast/catalogues) (Table 2). Perhaps the most surprising aspect of this Table 2 is the large number of ORFs assigned to the transcription function category (1,167). Thus, 182 *S. cerevisiae* genes would code for transcription factors of different classes, 52 of them being of the Zn(2)-Cys(6) binuclear cluster domain *trans*-activator class. This is a group of fungal transcription factors that are mainly involved in the regulation of basic metabolic functions such as amino acid or carbohydrate metabolism [34].

Another interesting aspect revealed by the *S. cerevisiae* sequence is gene redundancy [10]. This property is characteristic of all organisms and increases in parallel to genome size; in addition, it can be a source of functional diversity. It was proposed [37] that the present *S. cerevisiae* haploid genome derives from an ancient genome duplication followed by massive gene loss and transpositions, which would be the basis of the existing gene redundancy. However, other studies demonstrate that shorter duplication events in the chromosomes followed by translocation could explain the redundancy [22]. The consequence of redundancy is the existence of paralogous genes coding for protein families in *S. cerevisiae*. A family is defined as groups (of at least two members) of ORF products that show significant sequence homology among them but are not homologous to ORFs not included in the group (for simplicity, we do not distinguish between the concepts of family and partition [6]). With this definition, 2,458 ORFs of *S. cerevisiae* are included in 722 families ranging from two members (457 families) to one family of 108 members. Some large-size families correspond to transporters, while among the families with two members are those corresponding to ribosomal proteins or translation factors [6]. The previous definition of protein families was based exclusively on sequence homology. However, it is now supplemented by the large number of multimember protein classes (based on biochemical function) (http://mips.gsf.de/proj/yeast/catalogues). Thus, the *S. cerevisiae* genome contains genes for 143 ATPases, 22 cyclins and 78 GTP-binding proteins. In contrast, it has a single actin gene.

Function acquisition can also result from new combinations of protein domains, for example, as revealed in the *S. cerevisiae* genome by monothiol glutaredoxins, studied by our group [3]. This is a family of three

Table 1 Summary of the analysis of the genomes of completely sequenced yeast species[a]. Numbers in *parentheses* indicate the percentage of sequences with the indicated characteristics relative to the total number of entries in the database

	Saccharomyces cerevisiae[b]	*Schizosaccharomyces pombe*[b]	*Candida albicans*[c]
Entries in the database	6449	5010	6165
Sequences with homologues to known proteins	6338 (98.3)[b]	4996 (99.7)	5228 (84.8)
Sequences with PROSITE patterns[d]	2353 (36.5)	1833 (36.6)	2102 (34.1)
Sequences with PFAM domains[e]	3190 (49.5)	2749 (54.9)	2919 (47.3)
Sequences belonging to at least one COG[f]	2308 (35.8)	2207 (44.1)	2452 (39.8)
Sequences with superfamily assignments	4086 (63.4)	3089 (61.7)	3456 (56.1)
Sequences with one or more transmembrane regions	3287 (51.0)	1960 (39.1)	2706 (43.9)
Sequences having known or homologous three-dimensional structure	2766 (42.9)	2528 (50.5)	847 (30.0)

[a]From the PEDANT database (http://pedant.gsf.de)
[b]Computational analysis of complete genomic sequence
[c]Experimental and unfinished genomic sequence
[d] PROSITE patterns correspond to short protein sequences with a characterized role
[e]PFAM: Protein Domain Families database
[f]COG: Cluster of Orthologous Groups database

Table 2 Functional categories of the proteins of completely sequenced yeast species[a]

Category	*Saccharomyces cerevisiae*[b]	*Candida albicans*[c]	*Schizosaccharomyces pombe*[b]
Metabolism	1592	2336	1191
Energy	479	925	345
Cell cycle and DNA processing	1075	1619	819
Transcription	1167	1863	942
Protein synthesis	528	807	504
Protein fate	900	1577	765
Cellular transport and transport mechanism	855	1584	673
Cellular communication/ signal transduction mechanism	448	873	414
Cell rescue, defense and virulence	728	1505	509
Regulation of/interaction with cellular environment	350	962	222
Cell fate	855	1588	639
Systemic regulation of/interaction with environment	36	54	27
Development	61	284	45
Transposable elements, viral and plasmid proteins	129	366	19
Control of cellular organization	559	1352	441
Subcellular localization	3188	3764	2477
Protein activity regulation	32	86	27
Protein with binding function or cofactor requirement	12	439	11
Storage protein	7	18	8
Transport facilitation	447	766	323
Classification not yet clear-cut	749	1791	635
Unclassified proteins	3537	3575	1823

[a]For more details see: http://pedant.gsf.de. A protein can be assigned to more than one functional category; therefore, the number of entries for each organism is larger than the total number of coding genes

[b]Computational analysis of complete genomic sequence

[c]Experimental and unfinished genomic sequence

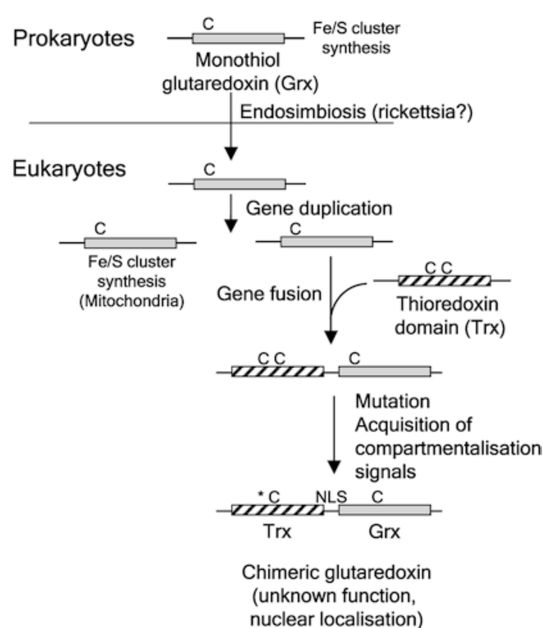

Fig. 1 The hypothetical evolution of monothiol glutaredoxins based on results described in [3] and on sequence comparisons (see text for more details). *Grx* Monothiol glutaredoxin domain, *Trx* thioredoxin domain, *NLS* nuclear localization sequence. The presence of one or two cysteine (*C*) residues in the active site is indicated. The *asterisk* marks the loss of a cysteine from the original thioredoxin active site

proteins, one of which is mitochondrial (Grx5) and is involved in the synthesis of iron–sulfur clusters, while the other two (Grx3 and Grx4) are nuclear and their functions are unknown. Structurally, Grx5 contains a single glutaredoxin domain (Grx), whereas Grx3 and Grx4 contain a modified thioredoxin (Trx) domain (lacking one of the two cysteine residues of the active site of authentic thioredoxins) fused at the N-end of the thioredoxin domain (Fig. 1). This type of hybrid Trx-Grx structure is totally absent in the completed prokaryote sequences, where thioredoxins and glutaredoxins exist as separate molecules, and probably was acquired by an ancestor eukaryote (our unpublished results). As a consequence, in most eukaryotes hybrid Trx-Grx proteins coexist with authentic thioredoxins and glutaredoxins, carrying out separate functions. Figure 1 illustrates our hypothesis on the evolution of monothiol glutaredoxins, based on sequence comparisons. The other two completed yeast genomes (*C. albicans* and *S. pombe*) have a single Trx-Grx molecular species in addition to the single Grx domain species, which indicates that duplication leading to Grx3 and Grx4 is a recent evolutionary event, occurring after separation of the *S. cerevisiae* branch from a common yeast ancestor. Genome plasticity in the form of new protein domain combinations such as that observed in yeast glutaredoxins has probably been a leading force in evolution, as also shown by analysis of the human genome [18].

Genomics of other Hemiascomycetes

The Génolevures project has obtained the partial genome sequences of 13 Hemiascomycetes species (Fig. 2) by using paired sequence reads from both ends

	Ploidy	Chromosome number	% synteny	% homologues
Saccharomyces cerevisiae	n	16	-	-
Saccharomyces bayanus* var. *uvarum	n	16	97.9	**46.5**
Saccharomyces exiguus	2n	14-16	70.7	25.8
Saccharomyces servazii	2n	12	70.2	24.7
Zygosaccharomyces rouxii	2n	7	71.3	**39.1**
Saccharomyces kluyverii	2n	8	53.5	25.1
Kluyveromyces thermotolerans	2n	7	56.5	25.4
Kluyveromyces lactis	n	6	47.3	**41.7**
Kluyveromyces marxianus var. *marxianus*	?	10	49.8	24.9
Pichia angusta	2n	?	18.9	**40.3**
Debaryomyces hansenii var. *Hansenii*	n	?	16.2	20.8
Pichia sorbitophila	2n	7	59.2	**25.6**
Candida tropicalis	2n	12	18.1	18.2
Yarrowia lipolytica	n	6	10.1	**19.7**

Fig. 2 Summary of the genetic characteristics of the 13 hemiascomycetous species studied in the Génolevures project, compared with those of *S. cerevisiae*. Species sequenced with 0.4× coverage are marked in *bold*; the others were sequenced with 0.2× coverage. *Vertical lines* are drawn from the cladograms shown in [23] and [32], and mark the groups of species resulting from those studies. Percentage of synteny indicates the proportion of gene order conservation between close or neighboring gene pairs when the order in the respective species is compared with that in *S. cerevisiae* [23]. Percentage of homologues indicates the fraction of *S. cerevisiae* proteins that have homologues in the respective species [24]; note that numbers corresponding to species with different sequence coverage are not comparable

of plasmid library clones [32]. This allowed 0.2–0.4× coverage of the entire nuclear genome depending on the species (Fig. 2). The amount of information obtained in this way is sufficient for comparative genomic studies of these species with respect to *S. cerevisiae*, which is also a Hemiascomycete yeast. First, the sequence of the 25S rRNA has allowed the construction of a phylogenetic tree of the 13 species plus *S. cerevisiae* [32]. The clades thus obtained basically agree with the tree derived from 18S rRNA analyses [21] and are also in accordance with relationships based on synteny (gene order conservation between species) [23] or amino acid sequence conservation [24] (Fig. 2). The only important discrepancy is the phylogenetic position of *Pichia sorbitophila*. This species displays large gene synteny with *S. cerevisiae* while the 25S rRNA sequence groups it with *Debaryomyces hansenii* and *Candida tropicalis*, distantly from budding yeast (Fig. 2). A very recent study of Wolfe et al. [39], also based on large rRNA sequences, positions *C. albicans* in the same clade as *C. tropicalis*.

The presence of a large number of gene families is common to all the Hemiascomycetes genomes studied to date [22]. The comprehensive Génolevures study also demonstrated the presence of a large number of Ascomycetes-specific genes, that is, genes not present in organisms different from ascomycetous fungi. At the time of that study, 1,712 genes of these 13 species having an orthologue in *S. cerevisiae* fell into this category [24].

Another interesting conclusion arising from the Génolevures project is the fact that distribution of genes among protein subclasses varies depending on the yeast species [12]. Genes classified as Ascomycetes-specific are those more rapidly evolving, therefore showing a more unequal distribution among species. The gene subclass distribution reflects (and in part explains) the different biology of the species studied, and is detailed in[12]. As an example, a human pathogen such as *C. tropicalis* has over-represented (with respect to *S. cerevisiae*) genes of the following subclasses: nitrogen and sulfur metabolism, glycolysis and gluconeogenesis, tricarboxylic acid pathway, peroxisomal organisation and transport, and β-oxidation of fatty acids. The metabolically versatile species of industrial use, *Yarrowia lipolytica* is over-represented in the following subclasses: nitrogen and sulfur metabolism, transport facilitators, ABC transporters, and peroxisomal biogenesis.

The information on the Hemiascomycete genome sequences has also been employed to study the possible diploidization of an ancestor of present species. Based on gene order conservation, the existence and position of gene pairs, and the number of chromosomes (haploid set, see Fig. 2) in the present hemiascomycetous yeasts, it has been proposed [39] that diploidization (followed by extensive rearrangement of short sequences and gene losses) occurred at a point after separation of *Zygosaccharomyces rouxii* from the two groups *S. cerevisiae*/*Saccharomyces bayanus* and *Saccharomyces exiguus*/*Saccharomyces servazii*. Consequently, the present genome of these four *Saccharomyces* species would be the result of that antique chromosome diploidization. The same would apply to the *Candida glabrata* genome, which has been put (based on the same gene order studies) close to the *Saccharomyces* group and therefore distant from *C. albicans* [39]. These observations point to the phylogenetic heterogeneity of the genus *Candida*.

The genome of *Candida albicans*

The Hemiascomycete *C. albicans* is an opportunistic pathogen causing human fungal infections, mainly in immunocompromised hosts. This species has the ability to grow as a yeast or in a hyphal mode in response to certain environmental conditions (it can be defined therefore as a dimorphic fungus), and is not amenable to standardized functional genomic strategies such as those used for other organisms, e.g. *S. cerevisiae* [5]. *C. albicans* has a nuclear genome of about 16 Mb, approximately 33% larger than that of *S. cerevisiae*, distributed in eight pairs of homologous chromosomes which are numbered from 1 (largest) to 7 (smallest), with the one carrying the ribosomal DNA called R [36]. The complete sequence of the *C. albicans* genome has been determined (http//www-sequence.stanford.edu/group/candida) based on a shotgun strategy carried out at the Stanford Genome Technology Center. The total assembly of the haploid genome comprises 15.5 Mb, corresponding to approximately 5,920 genes annotated by the European Consortium Galar Fungail (http://genolist.pasteur.

fr/CandidaDB). The PEDANT database establishes a somewhat larger number of ORF entries (see Table 1). The availability of the annotated *C. albicans* sequence opens the door for genome functional studies [5].

An initial striking observation on the *C. albicans* genome is that the proportion of ORF products non-homologous to known proteins is larger than that in *S. cerevisiae* or *S. pombe* (see Table 1), that is, the genome of *C. albicans* seems to contain a significantly high proportion of genes specific to this organism, which could be related to its versatile way of living. In accordance with this, a relatively larger number of proteins is assigned to functional categories related to interactions with the environment and morphogenesis in *C. albicans* than in the other two yeasts, compared with proteins involved in metabolism, proliferation or subcellular compartmentalization (see Table 2). The larger size of the *C. albicans* genome can be explained in part by the greater number of retrotransposon families. The *C. albicans* genome also contains families of proteases, lipases, and cell wall proteins with a large number of members, which are lacking in *S. cerevisiae* (http://genolist.pasteur.fr/CandidaDB).

Initial studies based on small ribosomal RNA sequences proposed an evolutionary proximity between *C. albicans* and *S. cerevisiae*, with the divergence estimated having occurred about 140 million years ago [17]. More recently, studies based on a large set of protein sequences have estimated that *S. cerevisiae* and *C. albicans* diverged about 841 million years ago [16]. During this period of independent evolution, inversions of small segments of DNA, less than ten genes long, were the major cause of gene rearrangement and the significant loss of synteny between both species [30].

Although both species are evolutionarily divergent, many *C. albicans* proteins have orthologues in *S. cerevisiae*. Given their differences in virulence and habitat, genomic comparisons between these fungi are likely to illuminate aspects of the unique cell biology of both microorganisms. Comparative genomics has recently been used to examine the possibility of a sexual life cycle in *C. albicans* [35]. This species has no known sexual stage and diverse studies have indicated that naturally occurring populations are predominantly clonal [14]. Genomic analysis showed that *C. albicans* possesses many homologues of genes that are found in sexual pathways in other eukaryotes, including *S. cerevisiae*, which indicates that *C. albicans* may have the genetic capacity to carry out a sexual cycle in nature [35]. However, some of the genes known to be involved in meiosis in *S. cerevisiae* are lacking in *C. albicans*, whereas other meiotic gene homologues present in filamentous fungi, such as *Aspergillus nidulans* and *Neurospora crassa*, are found in the *C. albicans* genome. This species appears to contain homologues of *S. cerevisiae* genes involved in two processes, glucose repression and nitrogen metabolism (*MIG1, GAT1, UME6, RAS, SNF1, MCK1, and CAT8*), related to the meiotic programs.

The availability of the *C. albicans* genome sequence has made it possible to construct DNA microarrays for the simultaneous analysis of expression levels of large sets of genes, such as those initially developed for *S. cerevisiae*. A *C. albicans* array has recently been used to identify new cellular targets of several transcriptional regulators that play a central role in the control of metabolism and yeast-hypha morphogenesis [25]. Microarrays have also been used to study the response of *C. albicans* to antifungals, identifying possible targets in the fungus to itroconazol treatment [9]. It has also been found that the evolution of drug resistance is accompanied by changes in gene expression that persist in the absence of the drug, the new patterns of gene expression being constitutive [8]. Transcript profiling of *C. albicans* during the yeast to hypha transition has identified several genes related to virulence factors that are modulated by the addition of serum at the 37 °C growth temperature [26]. In summary, the availability of the *C. albicans* genome sequence will allow an analysis of the dynamic and complex interactions between this fungus and its host, providing further information regarding the development of novel methods to control the infection and to identify the determinants required for *C. albicans* growth and proliferation in humans.

The genome of *Schizosaccharomyces pombe*

S. pombe (fission yeast) is an Archiascomycete distant in evolution from Hemiascomycetes [21]. Recent calculations [16] based on a large number of amino acid sequences indicate that it separated from Ascomycetes more than 1,100 million years ago. The fact that *S. pombe* is a single-celled eukaryote and is nearly as easy to manipulate as *S. cerevisiae* has also made it useful as a model to study molecular mechanisms in eukaryotic cells. Its genome was the sixth eukaryotic genome to be sequenced (http://www.sanger.ac.uk/Projects/S.pombe) [38].

The whole genome of *S. pombe* is only slightly larger in size (13.8 Mb) than that of *S. cerevisiae*. However, there exist substantial differences between budding and fission yeast: (1) All the genetic information in *S. pombe* is contained in three chromosomes. (2) Whereas only 5% of the *S. cerevisiae* genes have predicted introns, in *S. pombe* 4,730 introns have been characterized and confirmed in 43% of the genes. The proportion of the genome that codifies for proteins in fission yeast is 60.2% (lower than in budding yeast), which corresponds to an estimated 4,940 ORFs. These data clearly indicate that in *S. cerevisiae* gene density is more compact. (3) A comparison between the proteins that have been characterized in both fission and budding yeast showed that two thirds of *S. pombe* proteins have homologues in *S. cerevisiae*. Further analyses (Blast Crust, ftp://ncbi.nlm.nih.gov/blast/documents/README.bcl) indicated that in *S. pombe* there are fewer gene duplications

than in *S. cerevisiae*. This may account in part for the smaller genome size. (4) Transposable elements exist in the *S. pombe* genome. However, their proportion is the lowest compared to other sequenced eukaryotic organisms such as *S. cerevisiae* [20], *Arabidopsis thaliana* [33], *Drosophila melanogaster* [1], and humans [15]. (5) The presence of tRNA coding genes in the regions adjacent to centromeres is characteristic of the *S. pombe* genome. These centromeres are notably longer than in *S. cerevisiae*. Wood et al. [38] interpreted this information by assuming that each yeast needs different cell structures in order to carry out correct mitosis and/or meiosis.

On the other hand, with the exception of the above mentioned transposable elements, the proportion of functional categories for proteins is essentially similar in budding and fission yeast (see Table 2). This is remarkable considering that both genomes evolved divergently over such a long period of time that differences between *S. cerevisiae* and *S. pombe* are equivalent to divergences with their human homologues [16].

Comparison between protein sequences of *S. pombe* and *S. cerevisiae* with human proteins involved in several diseases has revealed interesting results [38]. There are a considerable number of proteins with a high level of conserved homology between the three organisms. Some of them are involved in cancer (xeroderma pigmentosum). In addition, two homologues to genes linked to human diseases have been characterized in *S. pombe* that are absent in *S. cerevisiae*. One of the genes is related to cancer, and the other to cell metabolism.

A comparative view of the yeast genomes

The amount of information gained after sequencing the genomes of *S. cerevisiae*, *S. pombe* and *C.albicans* places the basis for further studies not only on the function of specific proteins (or protein families) but also on the evolution from prokaryotic to eukaryotic cell organization. Other fungal genomes are at different stages in the process of being sequenced (*Candida glabrata*, *Criptococcus neoformans*, *Aspergillus fumigatus*) and they will add information to this integrative view of the evolution of life, and of fungi in particular. It is remarkable that eukaryotic cellular organization does not necessarily require more genes than are present in prokaryotic organization. In fact, the two bacteria that have the largest sequenced genomes up to this moment, *Streptomyces coelicolor* [4] and *Mesorhizobium loti* [19], contain more genes than in each of the three yeast species sequenced (7,825 and 6,752 potentially coding genes, respectively). Therefore, it is not the number of genes but the characteristics of these genes that differentiate both types of cellular organization. Wood et al. [38] gave some clues in this respect by describing several functional categories of genes that are present in all eukaryotic organisms and absent in all prokaryotes. Besides the expected genes responsible for cytoskeletal proteins or chromatin organization, there are genes for cell cycle regulation and damage control, and kinases/phosphatases for rapid control of protein activities. It seems that, in addition to structural differences, it is through the regulation of cell proliferation and protein activity that eukaryotes have gained versatility in their responses to variable environments. Combination in a single eukaryotic molecule of independent protein domains existing as such in the prokaryotic world added the potential for regulation of responses by eukaryotic cells.

In contrast, when analyzing differences between unicellular (*S. cerevisiae* and *S. pombe*) and multicellular eukaryotes (humans, *D. melanogaster*, *A. thaliana*) regarding gene conservation, only three genes were found to be specific to multicellular organisms [38]. They code for a transcription factor, an RNA-binding protein, and a selenium-binding protein. The authors of this study concluded that transition from unicellularity to multicellularity is probably less complex in terms of new genes than the transition from prokaryotic to eukaryotic cell. This hypothesis has received some criticism from other authors [11], who argued that in those studies two important aspects were not sufficiently considered: (1) the fact that plants and animals, both multicellular, probably evolved independently, and (2) the lapse of time since unicellular organisms evolved to multicellular organisms compared to the much longer time that was required to change from prokaryotic to eukaryotic cell. The availability in the near future of sequences of other prokaryotic and eukaryotic species will provide a more accurate evolutionary interpretation.

Focusing our attention on unicellular fungi, recent sequence-based genome analyses have basically confirmed the previously proposed [21] phylogenetic relationships between *S. cerevisiae*, *S. pombe, C. albicans*, and other Hemiascomycete species. However, some new views have been introduced. For instance, although the existence of gene families is common to all the species, the possible duplication of the haploid genome of a *S. cerevisiae* ancestor has added complexity to gene families in this yeast and in close species. As this duplication would have occurred recently, it would not yet have allowed the functional divergence of sequence-homologous members of these families. Another consideration is that the view that *S. pombe* is significantly closer to humans than *S. cerevisiae* can no longer be sustained, since in both yeasts the proportion of genes homologous to those of metazoans (and to humans in particular) is similar [38]. This is in agreement with recently proposed phylogenetic trees based on protein sequences [2,16]. Finally, the existence of Ascomycetes-specific genes is another aspect to consider. Many of these genes code for proteins important for cell wall structure and biogenesis. It is notable that the *S. cerevisiae* genome employs a large set of genes for cell-wall-related functions [7,31].

Another important aspect is that genomic studies of yeasts are currently being carried out with single isolates of each species that are not necessarily representative of

the whole species. In the future, genomic studies will have to extend to different isolates from the same species, in the same way as is already occurring with bacterial strains. Comparative genomics using the information thus obtained will be the basis for further advances in our knowledge of the biology and diversity of yeasts. The pioneering and seminal studies of Professor Herman Phaff will no doubt be in the minds of the researchers involved in this work.

References

1. Ashburner M, Misra S, Roote J, Lewis SE, Blazej R, Davis T, et al (1999) An exploration of the sequence of a 2.9-Mb region of the genome of *Drosophila melanogaster*: the *Adh* region. Genetics 153:179–219
2. Baldauf SL, Roger AJ, Wenk-Siefert I, Doolittle WF (2000) A kingdom-level phylogeny of eukaryotes based on combined protein data. Science 290:972–977
3. Bellí G, Polaina J, Tamarit J, de la Torre MA, Rodríguez-Manzaneque MT, Ros J, Herrero E (2002) Structure-function analysis of yeast Grx5 monothiol glutaredoxin defines essential amino acids for the function of the protein. J Biol Chem 277:37590–37596
4. Bentley SD, Chater KF, Cerdeño-Tárraga AM, Challis GL, Thomson NR, James KD, et al (2002) Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). Nature 417:141–147
5. Berman J, Sudbery PE (2002) *Candida albicans*: a molecular revolution built on lessons from budding yeast. Nat Rev Genet 3:918–930
6. Blandin G, Durrens P, Tekaia 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
7. Caro LH, Tettelin H, Vossen JH, Ram AF, van den Ende H, Klis FM (1997) *In silico* identification of glycosyl-phosphatidylinositol-anchored plasma membrane and cell wall proteins of *Saccharomyces cerevisiae*. Yeast 13:1477–1489
8. Cowen LE, Nantel A, Whiteway MS, Thomas DY, Tessier DC, Kohn LM, Anderson JB (2002) Population genomics of drug resistance in *Candida albicans*. Proc Natl Acad Sci USA 99:9284–9289
9. De Backer MD, Ilyna T, Ma XJ, Vandoninck S, Luyten HLMW, van den Bossche H (2001) Genomic profiling of the response of *Candida albicans* to itroconazol treatment using a DNA microarray. Antimicrob Agents Chemother 45:1660–1670
10. Dujon B (1996) The yeast genome project: what did we learn? Trends Genet 12:263–270
11. Eisen JA (2002) Brouhaha over the other yeast. Nature 415:845–848
12. Gaillardin C, Duchateau-Nguyen G, Tekaia F, Llorente B, Casaregola S, Toffano-Nioche C, et al (2000) Genomic exploration of the Hemiascomycetous yeasts: 21. Comparative functional classification of genes. FEBS Lett 487:134–149
13. Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, et al (1996) Life with 6000 genes. Science 274:563–567
14. Gräser Y, Volovsek M, Arrington J, Schönian G, Presber W, Mitchell TG, Vilgalys R (1996) Molecular markers reveal that population structure of the human pathogen *Candida albicans* exhibits both clonality and recombination. Proc Natl Acad Sci USA 93:12473–12477
15. Gu Z, Wang H, Nekrutenko A, Li WH (2000) Densities, length proportions, and other distributional features of repetitive sequences in the human genome estimated from 430 megabases of genomic sequence. Gene 259:81–88
16. Heckman DS, Geiser DM, Eidell BR, Stauffer RL, Kardos NL, Hedges SB (2001) Molecular evidence for the early colonization of land by fungi and plants. Science 293:1129–1133
17. Hendriks L, Goris A, Van de Peer Y, Neefs JM, Vancanneyt M, Kersters K, Hennebert GL, de Wachter R (1991) Phylogenetic analysis of five medically important *Candida* species as deduced on the basis of small ribosomal subunit RNA sequences. J Gen Microbiol 137:1223–1230
18. International Human Genome Consortium (2001) Initial sequencing and analysis of the human genome. Nature 409:860–921
19. Kaneko T, Nakamura Y, Sato S, Asamizu E, Kato T, Sasamoto S, et al (2000) Complete genome structure of the nitrogen-fixing symbiotic bacterium *Mesorhizobium loti*. DNA Res 7:331–338
20. Kim JM, Vanguri S, Boeke JD, Gabriel A, Voytas DF (1998) Transposable elements and genome organization: a comprehensive survey of retrotransposons revealed by the complete *Saccharomyces cerevisiae* genome sequence. Genome Res 8:464–478
21. Kurtzman CP, Blanz PA (2000) Ribosomal RNA/DNA sequence comparisons for assessing phyogenetic relationships. In: Kurtzman CP, Fell JW (eds) The yeasts: a taxonomic study, 4th edn. Elsevier, Amsterdam, pp 69–74
22. Llorente B, Durrens P, Malpertuy A, Aigle M, Artiguenave F. Blandin G, et al (2000) Genomic exploration of the Hemiascomycetous yeasts: 20. Evolution of gene redundancy compared to *Saccharomyces cerevisiae*. FEBS Lett 487:122–133
23. Llorente B, Malpertuy A, Neuvéglise C, de Montigny J, Aigle M, Artiguenave F, et al (2000) Genomic exploration of the Hemiascomycetous yeasts: 18. Comparative analysis of chromosome maps and synteny with *Saccharomyces cerevisiae*. FEBS Lett 487:101–112
24. Malpertuy A, Tekaia F, Casarégola S, Aigle M, Artiguenave F, Blandin G, et al (2000) Genomic exploration of the Hemiascomycetous yeasts: 19. Ascomycetes-specific genes. FEBS Lett 487:113–121
25. Murad AMA, d́Enfert C, Gallardin C, Tournu H, Tekaia F, Talibi D et al (2001) Transcript profiling in *Candida albicans* reveals new cellular functions for the transcriptional repressors CaTup1, CaMig1 and CaNrg1. Mol Microbiol 42:981–993
26. Nantel A, Dignard D, Bachewich C, Harcus D, Marcil A, Bouin AP, et al (2002) Transcription profiling of *Candida albicans* cells undergoing the yeast-to-hyphal transition. Mol Biol Cell 13:3452–3465
27. Pellegrini M, Marcotte EM, Thompson Mj, Eisenberg D, Yeates TP (1999) Assigning protein functions by comparative genome analysis; protein phylogenetic profiles. Proc Natl Acad USA 96:42885–42888
28. Ross Fitzgerald J, Musser JM (2001) Evolutionary genomics of pathogenic bacteria. Trends Microbiol 9:547–553
29. Schoolnik GK (2002) Functional and comparative genomics of pathogenic bacteria. Curr Opin Microbiol 5:20–26
30. Seoighe C, Federspiel N, Jones T, Hansen N, Bivolarovic V, Surzycki R, et al (2000) Prevalence of small inversions in yeast order evolution. Proc Natl Acad Sci USA 97:14433–14437
31. Smits GJ, Kapteyn JC, van den Ende H, Klis FM (1999) Cell wall dynamics in yeast. Curr Opin Microbiol 2:348–352
32. Souciet JL, Aigle M, Artiguenave F, Blandin G, Bolotin-Fukuhara M, Bon E, et al (2000) Genomic exploration of the Hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies. FEBS Lett 487:3–12
33. The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408:796–815
34. Todd RB, Andrianopoulos A (1997) Evolution of a fungal regulatory gene family: the Zn(II)2Cys6 binuclear cluster DNA binding motif. Fungal Genet Biol 21:388–405
35. Tzung KW, Willims RM, 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

36. Wickes B, Staudinger J, Magee BB, Kwon-Chung KJ, Magee PT, Scherer S (1991) Physical and genetic mapping of *Candida albicans*: genes previously assigned to chromosome I map on chromosome R, the rDNA-containing linkage group. Infect Immun 59:2480–2484
37. Wolfe KH, Shields DC (1997) Molecular evidence for an ancient duplication of the entire yeast genome. Nature 387:708–713
38. Wood V, Gwilliam R, Rajandream MA, Lyne M, Stewart A, Sgouros J, et al (2002) The genome sequence of *Schizosaccharomyces pombe*. Nature 415:871–879
39. Wong S, Butler G, Wolfe KH (2002) Gene order evolution and paleopolyploidy in hemiascomycete yeasts. Proc Natl Acad Sci USA 99:9272–9277