

# Conserved Processes and Lineage-Specific Proteins in Fungal Cell Wall Evolution<sup>∇†</sup>

Juan E. Coronado,<sup>1,2</sup> Saad Mneimneh,<sup>3</sup> Susan L. Epstein,<sup>3</sup> Wei-Gang Qiu,<sup>1,2</sup> and Peter N. Lipke<sup>1,2,4\*</sup>

Department of Biological Sciences,<sup>1</sup> Center for Gene Structure and Function,<sup>2</sup> and Department of Computer Science,<sup>3</sup> Hunter College, City University of New York, New York, New York 10021, and Department of Biology, Brooklyn College of City University of New York, Brooklyn, New York 11210<sup>4</sup>

Received 9 February 2007/Accepted 3 October 2007

**The cell wall is a defining organelle that differentiates fungi from its sister clades in the opisthokont superkingdom. With a sensitive technique to align low-complexity protein sequences, we have identified 187 cell wall-related proteins in *Saccharomyces cerevisiae* and determined the presence or absence of homologs in 17 other fungal genomes. There were both conserved and lineage-specific cell wall proteins, and the degree of conservation was strongly correlated with protein function. Some functional classes were poorly conserved and lineage specific: adhesins, structural wall glycoprotein components, and unannotated open reading frames. These proteins are primarily those that are constituents of the walls themselves. On the other hand, glycosyl hydrolases and transferases, proteases, lipases, proteins in the glycosyl phosphatidyl-inositol-protein synthesis pathway, and chaperones were strongly conserved. Many of these proteins are also conserved in other eukaryotes and are associated with wall synthesis in plants. This gene conservation, along with known similarities in wall architecture, implies that the basic architecture of fungal walls is ancestral to the divergence of the ascomycetes and basidiomycetes. The contrasting lineage specificity of wall resident proteins implies diversification. Therefore, fungal cell walls consist of rapidly diversifying proteins that are assembled by the products of an ancestral and conserved set of genes.**

This detailed phylogenetic study of homologs to fungal cell wall proteins partially resolves the question of their origin. Cell walls are defining organelles that differentiate fungi from their sister clades in the opisthokonts, including the animals (6, 7, 51, 54). Although fungal cell walls consist of both carbohydrates and proteins, they are not homologous to cell walls in other organisms, such as plants or bacteria. It is, therefore, reasonable to ask whether fungal cell walls evolved independently in multiple lineages or diverged from a common ancestral origin. At one extreme, all or most cell wall components might be conserved in all fungal lineages, so that the same set of orthologous macromolecules would constitute cell walls across all existing fungal species. At the other extreme, few cell wall components might be conserved in structure or function during fungal evolution, so that fungal cell wall components would have lineages that are highly species specific. Fungi consist of the phyla *Ascomycota*, *Basidiomycota* (together called *Dikarya*, or sometimes “higher fungi”), *Glomeromycota*, *Microsporidia*, and two polyphyletic groups, zygomycetes and chytridiomycetes (25). In general, fungi are walled in at least part of the life cycle, and chitin is usually found as a wall component, although it is sometimes minor.

**General features of fungal cell walls.** Fungal walls appear to share a conserved basic design, although chemical and molecular analyses have been rare, except in ascomycetes and basidiomycetes. The major polysaccharides are glucans, chitin or

chitosan (4, 8, 11, 24, 32, 44), and mannans (28, 32, 35, 52). Glycosyl phosphatidyl inositol (GPI)-anchored proteins have also been found whenever sought (12, 13, 58). The best-studied fungal cell walls are from baker’s yeast, *Saccharomyces cerevisiae* (28, 35, 38), and the opportunistic pathogen *Candida albicans* (12). These two share basic design with other ascomycetes:  $\beta$ -linked fibrous polysaccharides cross-linked by glycoproteins (28, 29, 47). The polysaccharides are assembled into fibrils and cross-linked to the Ser-rich and Thr-rich glycoproteins through modified GPI-glycans at the C terminal and/or by ester bonds (17, 28). Disulfide bonds between glycoproteins also contribute to cell wall integrity (15, 48). Many *S. cerevisiae* wall components and assembly mechanisms have been found in the walls of other fungi, and so *S. cerevisiae* appears to be a valid general model (24, 29, 35, 38, 47). The synthetic and secretion pathways that export fungal cell wall components are similar to those of extracellular glycoproteins in other eukaryotes (7, 27, 37, 42). The protein components of *S. cerevisiae* cell walls are structurally and functionally heterogeneous but have several common sequence features that enable their extracellular localization (27). These motifs include N-terminal secretion signal sequences, N- and O-glycosylation motifs, intermolecular disulfide bonding, and often hydrophobic C-terminal GPI addition signals. As in other eukaryotes, membrane translocation occurs in the endoplasmic reticulum (ER), and glycosylation takes place in the ER and Golgi apparatus, as the fungal cell wall components traverse the vesicular secretion pathway (34, 35, 42).

**Molecular evolution and fungal cell walls.** Fungal cell wall evolution has not been systematically investigated by sequence analysis, primarily because standard protein alignment techniques have proved inadequate for many fungal cell wall proteins. Alignment algorithms typically use scoring matrices

\* Corresponding author. Mailing address: Department of Biology, Brooklyn College of City University of New York, Brooklyn, NY 11210. Phone: (718) 951-5000, ext. 1949. Fax: (718) 951-4659. E-mail: plipke@brooklyn.cuny.edu.

† Supplemental material for this article may be found at <http://ec.asm.org/>.

∇ Published ahead of print on 19 October 2007.

based on amino acid composition of globular proteins (1), but fungal cell wall proteins have a very different composition. In *S. cerevisiae* wall proteins, amino acids S, T, A, V, G, and P are present at higher percentages and K, L, N, D, and E are more scarce (9, 64). This composition gives disproportionately large scores to alignments of overrepresented residues, even when they are in nonhomologous positions, a problem called “low-complexity corruption” (50). We devised a technique to minimize low-complexity corruption with scoring matrices adjusted for the frequency of each amino acid in the query sequence (9, 64). Such query-specific matrices lower the score for matches of residues common in the query sequence and so yield alignments free from random pairings of common residues (9, 64).

Query-specific matrices allow identification and alignment of homologous sequences among fungal cell wall proteins (9). We have compared *S. cerevisiae* cell wall proteins to open reading frames (ORFs) in 17 other complete genomes and to the NR database (36). This large collection of ORFs supports the comparison of cell wall-related sequences in diverse fungal species. Our results show that fungal walls are both conserved in origin and diversified in composition.

#### MATERIALS AND METHODS

**Identification of cell wall proteins and their paralogs in *S. cerevisiae*.** We identified a set of 103 *S. cerevisiae* cell wall proteins based on their annotation in the *S. cerevisiae* gene ontology (GO) database (36). This set was the union of the 85 entries GO annotated as “cell wall” with the (overlapping) set of 74 proteins in *S. cerevisiae* known or predicted to have a signal sequence and a GPI attachment signal but no transmembrane domains (2, 13).

These 103 proteins were then used as queries in searches against the *S. cerevisiae* genome sequence database (SGD). To avoid low-complexity corruption, all searches reported here used BLAST with the query-specific *gtQ* scoring matrices (9). A *gtQ* matrix compensates for overrepresentation of any residue, while it preserves the negative value of the entire matrix. Thus, it is much more highly discriminating but does not sacrifice sensitivity. A transitive closure procedure was conducted with multiple rounds of searches against the *S. cerevisiae* genome, until no new homologous ORFs were identified. In a given round, new ORFs that participated in high-scoring pairs with *e*-values of  $<10^{-5}$  became the query set for the next round, still against the same database. This process continued until no additional proteins with *e*-values below the specified cutoff were obtained (22, 62). This method identified a total of 171 proteins, including the original 103 queries as *S. cerevisiae* cell wall components or their paralogs. To these genes we added 16 other sequences that are not annotated as cell wall in GO in SGD but are known to be associated with wall synthesis and biogenesis (28, 29, 35). These 187 ORFs are listed in Table S1 of the supplemental material and are referred to in the remainder of this paper as *S. cerevisiae* cell wall-related proteins.

**Cell wall-related proteins from *Yarrowia lipolytica*.** We needed an independent set of queries that could be used to test the relationship of homology occurrence and phylogenetic distance (see Fig. 4, below). Therefore, we identified a set of putative cell wall-related proteins from the ascomycete *Yarrowia lipolytica* as query sequences. Because annotation is sparse in species other than *S. cerevisiae*, we used the ORFs predicted to have GPI anchors, which are common in cell wall proteins in *S. cerevisiae*, *C. albicans*, and other fungi (61). *Y. lipolytica* putative cell wall-related sequences were identified as follows: we used the GPI-SOM server to find 237 proteins with predicted GPI signals and secretion signals (19). Of these 237 proteins, the tmHMM prediction server (<http://www.cbs.dtu.dk/services/TMHMM/>) identified 188 proteins without transmembrane domains. The Fungal BIG-PI server ([http://mendel.imp.ac.at/gpi/fungi\\_server.html](http://mendel.imp.ac.at/gpi/fungi_server.html)) was slightly more conservative and identified a subset of 149 (79%) as potential GPI-anchored proteins (18). Like the *S. cerevisiae* cell wall proteins, 115 of the *Y. lipolytica* proteins contained regions with high S and T content: >50 consecutive residues with >30% S and T.

**Genome comparisons.** The genomes were retrieved from their respective databases (see Table S2 in the supplemental material). The 187 *S. cerevisiae* cell wall-related ORFs were used as initial queries in BLAST-*gtQ* searches for sequence similarity in 17 other fungal genomes (listed in Table S2). The 188 GPI-anchored proteins from *Y. lipolytica* were also used as a query set in ho-

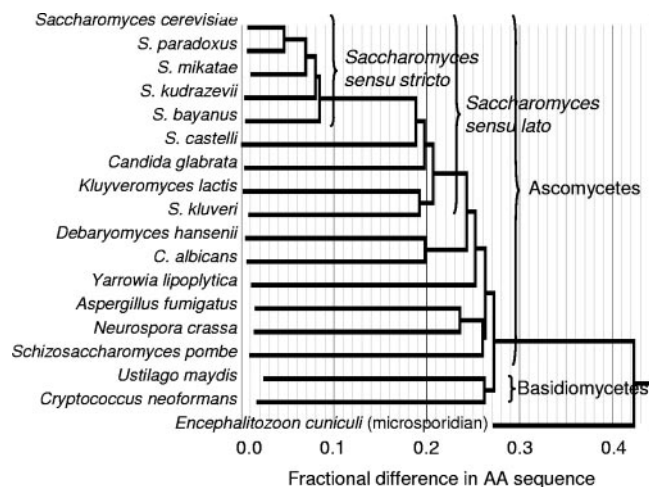


FIG. 1. Neighbor-joining phylogenetic tree based on amino acid substitution rates for all orthologs in 18 fungal genomes.

mology searches of the other fungal genomes, but the results were not analyzed in detail because many of the ORFs have minimal annotations or ones based only on similarity to *S. cerevisiae* sequences.

**Species tree.** The power of any phylogenetic analysis depends on a comparison to a reliable phylogenetic tree for the sampled organisms. Therefore, a tree was generated based on a comparison of amino acid sequences in all orthologous ORFs in all 18 fungal genomes (31). A species tree was produced from a neighbor joining of the amino acid identity distance matrix (31).

**Degree of conservation.** Presence/absence of an ORF on an internal node of the tree was inferred from maximum parsimony, with a postorder tree traversal algorithm (20). The degree of conservation of an ORF (see Fig. 3 and 5, below; see also Table S1 in the supplemental material) was determined as the proportion of the total length of branches on which this ORF is present over the total tree branch length. The branches where an ORF is present are counted only if the ORF is present on both the parent and the child nodes.

#### RESULTS

**Phylogenetics of fungal proteomes.** A species tree was generated based on the average amino acid identity of all orthologs between all pairs of genomes. Such widespread comparisons are more reliable than trees based on 18S rRNA or on other limited sequence comparisons (5, 25). Figure 1 shows the resultant neighbor-joining tree, which was congruent to other phylogenetic analyses of the fungi (25). *S. cerevisiae*, *Saccharomyces paradoxus*, *Saccharomyces bayanus*, *Saccharomyces kudrzewii*, and *Saccharomyces bayanus* form a closely related group and represent *Saccharomyces sensu stricto* (39). The clade *Saccharomyces sensu lato* includes also *Candida glabrata*, *Saccharomyces castellii*, *Kluyveromyces lactis*, and *Saccharomyces kluyveri*. More distant ascomycetes and the basidiomycetes were clearly resolved, and the microsporidian *Encephalitozoon cuniculi* served as an outgroup.

The amino acid distance tree in Fig. 1 shows that major fungal lineages have accumulated extensive sequence differences. These differences imply that the lineages split long ago, with early divergence of the basidiomycetes from the ascomycetes and of the filamentous ascomycetes *Aspergillus fumigatus* and *Neurospora crassa* from the yeasts. The basidiomycetes *Ustilago maydis* and *Cryptococcus neoformans* are themselves as divergent as the budding yeast *S. cerevisiae* and the fission yeast *Schizosaccharomyces pombe* (~48% amino acid differences).

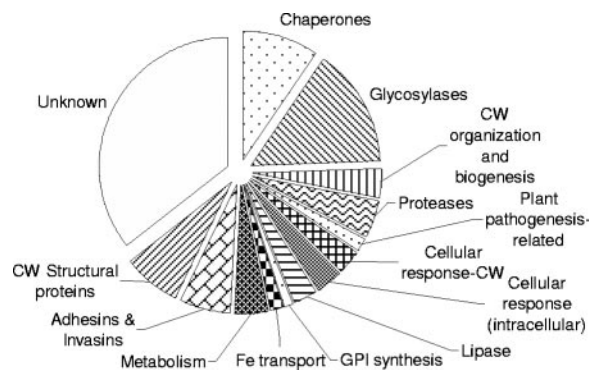


FIG. 2. Cellular function groups of *S. cerevisiae* cell wall-related proteins. The size of each sector is proportional to the number of ORFs in that group. CW, cell wall.

**The functions of *S. cerevisiae* cell wall-related proteins and their paralogs.** We identified a set of 187 cell wall-related proteins as those annotated in the SGD GPI-anchored proteins and their paralogs identified in transitive closure (9). For the 187 cell wall-related proteins, GO annotations report that 108 (63%) are localized either to the cell wall (86 proteins) or to similar cell wall-associated spaces, such as periplasmic (3 proteins) and extracellular (13 proteins). A large set (46 ORFs, 27%) has unknown location or no annotation. Only about 10% of the proteins are reported with intracellular localization, including paralogs of chaperones and glycolytic enzymes. Those two classes of proteins are primarily cytoplasmic, but some proteins in each class are also proposed to be localized in cell walls in *S. cerevisiae* and in *C. albicans* on the bases of analyses of isolated walls and/or immunoassays on intact cells (14, 30, 40, 41, 46).

The distribution of molecular function for these 187 proteins (as annotated by the SGD) includes several major classes (Fig. 2). Sixty-one proteins (33%) have no known molecular function; 33 (18%) are glycosyl transferases, hydrolases, or transglycosylases (collectively called “carbohydrate-active enzymes” in the CAZY database [http://www.cazy.org], but abbreviated as “glycosylases” in subsequent discussions) (10); 16 (9%) are involved in unfolded protein binding (called “chaperones,” subsequently); 13 (7%) are structural constituents of the cell wall; and 8 (4%) have protease activity. The GO annotation for the biological process most often associated with the 187 *S. cerevisiae* cell wall-related proteins was “cell wall organization and biogenesis.” Forty-eight proteins, including chaperones and glycosylases, were annotated this way.

**Occurrence of cell wall protein homologs in fungi.** To study the conservation of cell wall proteins across the fungi, the 187 *S. cerevisiae* cell wall-related proteins were used as queries in BLAST-gtQ searches against 17 other fungal genomes. Sequences that aligned with *e*-values of  $<10^{-10}$  were counted as homologs. A homology was designated as an orthology only if the members of the high-scoring pair were reciprocal best hits and at least 30% of the aligned amino acids were identities across at least 70% of the length of the shorter member of the pair (31).

Figure 3 illustrates the occurrence of homologs and orthologs in two functional classes of cell wall proteins. Among

the adhesins and invasins shown at the top in Fig. 3, those participating in mating were conserved within the closely related *Saccharomyces sensu stricto* species. For the *FLO* family of mannose-binding flocculins and invasins, homologs were also found in *Saccharomyces sensu lato* species and *Debaryomyces hansenii*, but not in the more distant yeasts or the filamentous ascomycetes. In contrast, the glycosylases at the bottom of the figure show a different conservation pattern. For many members of this large functional class, there were homologs in each of the ascomycete and basidiomycete genomes. Thus, these two functional classes of genes show different patterns of sequence conservation. Table S1 in the supplemental material lists the 187 *S. cerevisiae* cell wall-related proteins.

The occurrence pattern for most genes followed the expected general pattern: there were homologs in the species most closely related to *S. cerevisiae*, and the probability of recognizable homology decreased as phylogenetic distance increased. Figure 4 illustrates this trend; it shows decreasing homolog identification with increasing phylogenetic distance. To confirm that this trend was not specific to *S. cerevisiae* and its cell wall-related proteins, we carried out a similar search and analysis for homologs of 188 potential GPI-anchored wall proteins from *Y. lipolytica* (a set of similar size). Most of the other species were similarly distant and had homologs to about 35% of the *Y. lipolytica* proteins. The corresponding points in Fig. 4 cluster near the occurrence-distance curve for *S. cerevisiae*, except for the *Y. lipolytica*/*E. cuniculi* comparison, which identified homologs to only 10 of the *Y. lipolytica* proteins.

**Degree of conservation is related to molecular function.** The origins of cell walls can be understood in terms of the roles of the conserved and nonconserved cell wall ORFs. For each function of the 187 ORFs in the *S. cerevisiae* wall (Fig. 2), we determined the degree of conservation, as detailed in Materials and Methods. The different functional classes differed significantly in their degrees of conservation (Fig. 5). Several functional classes were poorly conserved, with degree of conservation values below 0.4 (on the right in Fig. 5). These protein classes included a small set of Fe transport-related proteins, the uncharacterized ORFs, cell wall structural components, and adhesins. The uncharacterized ORFs include many with known location in the cell wall but no known phenotype in gene deletions (see Table S1 in the supplemental material). Thus, these uncharacterized ORFs may encode proteins without unique or essential molecular activities or those replaceable with other gene products (17, 61). In contrast, components of the GPI synthesis pathway, lipases, proteases, metabolic enzymes, glycosylases, and chaperones were strongly conserved and were present in most tested fungi. In summary, a general observation is that the biosynthetic capability for cell walls is well conserved but noncatalytic wall components, including the adhesins, and structural and putative structural proteins are not.

**Cell wall-related proteins occur across the eukaryotes.** Comparisons of the *S. cerevisiae* cell wall-related proteins against the GenBank NR database identified homologs in 137 ascomycetes species other than the genomes used in the initial fungal comparisons. Homologs were also found in 30 basidiomycete species, 7 zygomycete species, and 2 chytridiomycetes species. The GPI biosynthetic enzymes, metabolic enzymes, and chaperones were most likely to have homologs (Fig. 6).

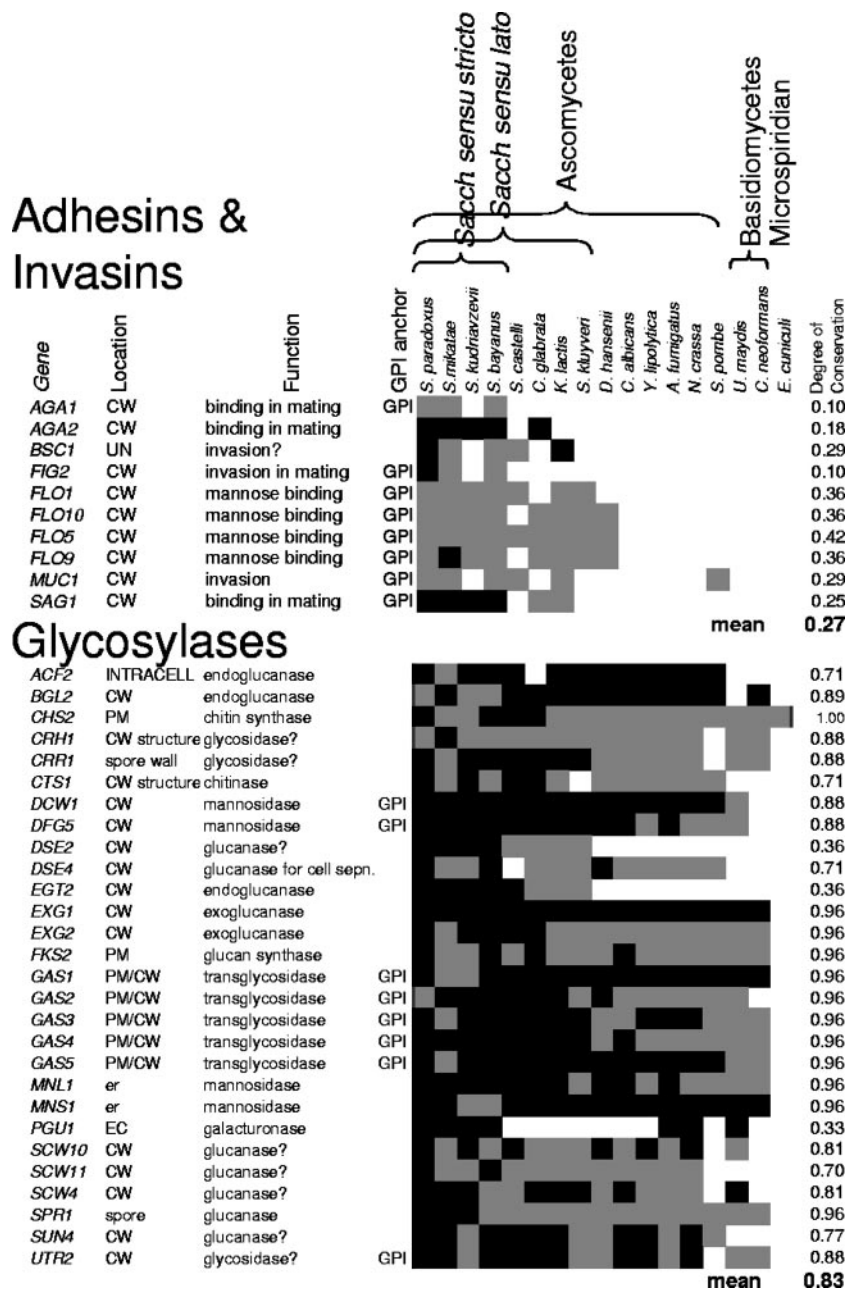


FIG. 3. Presence of homologs and orthologs for two functional groups of wall-related genes in 17 fungi. The query sequences were *S. cerevisiae* cell wall-related proteins in the adhesin and glycosylase functional groups. The presence of one or more homologs with  $e$  of  $<10^{-10}$  is shown as a gray square, and the presence of an ortholog is shown as a black square (31). On the right is the total conservation score for each gene, with the group mean conservation score at the bottom of the group. Cellular locations and abbreviations are as follows: CW, cell wall; PM, plasma membrane; ER, endoplasmic reticulum; UN, unknown. The column labeled GPI anchor shows those proteins predicted to have GPI anchors.

Searches from the 187 *S. cerevisiae* cell wall-related proteins in the NCBI nonredundant protein database identified similar sequences in organisms outside of the fungi as well (Fig. 6). Forty of the 187 proteins had homologs in other kingdoms. These 40 ORFs were in the functional groups most conserved in the fungi, including glycosylases, GPI synthetic proteins, aspartyl proteases, a group of related proteins important for plant pathogenesis, and two cytoplasmic groups with known or suspected cell wall involvement: chaperones and metabolic

enzymes (14, 30, 40, 41, 46). Because these kinds of genes are present in many eukaryote kingdoms, they are likely to be ancestral to all eukaryotes, including fungi.

### DISCUSSION

**What are the origins of fungal cell walls?** The cell surface is at the front line of species adaptation. While the cell wall needs to maintain a basic functionality like other organelles, it is

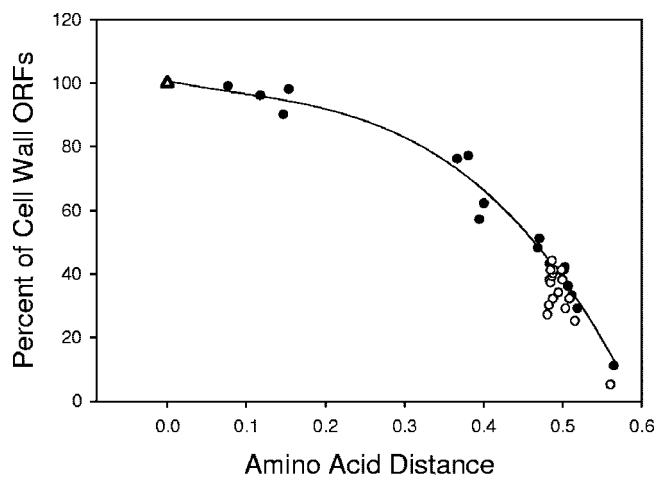


FIG. 4. Relationship between homolog identification and phylogenetic distance. For each tested genome, the fraction of query sequences that had a homolog was plotted against the phylogenetic distance derived from Fig. 1. The query protein sets were all *S. cerevisiae* cell wall-related proteins (●) or *Y. lipolytica* GPI-anchored proteins (○). The top, left-most point (△) corresponds to the searches of *S. cerevisiae* queries against the *S. cerevisiae* genome and to searches of the *Y. lipolytica* queries against the *Y. lipolytica* genome.

expected to evolve more quickly because it is directly exposed to biotic and abiotic selective forces. Cell surface structures are a defining feature of some of the deepest evolutionary lineages, including bacteria, archaea, fungi, plants, and animals. Indeed, innovations in cell surface structures may be responsible for the success and persistence of major evolutionary lineages (5). One way to test for a causal relation between the evolution of the cell wall and the cell as a whole is to study the evolutionary conservation of cell wall macromolecular components. For example, a key evolutionary innovation in a lineage should be conserved and shared by all descendants, without repeated gain and loss.

Cavalier-Smith has argued that fungal cell walls are derived from ancestral chitinous walls of desiccation-resistant cysts in many protist clades (7, 8). Certainly, chitinous cysts and spores are present in other eukaryote kingdoms, including *Amoebozoa* (such as *Entamoeba* [11]), *Chromalveolata* (*Phytophthora* [43, 44]), and *Excavata* (*Giardia* [57]). In addition, *S. cerevisiae* walls share extracellular disulfide cross-links and GPI-anchored cell adhesion molecules with *Amoebozoa*, perhaps the earliest-diverged eukaryotes (Fig. 6) (7, 26, 55, 59, 63). Cavalier-Smith's hypothesis motivated our use of genomics to identify genes that are key to wall biogenesis and structure and to test which are conserved and which are subject to adaptive selection.

The deep roots for the ascomycete-basidiomycete split show large amino acid distances among all homologs in the fungi (25) (Fig. 1). This observation implies that cell wall genes could also have diverged broadly. Therefore, the identification of homologs in some cell wall proteins among distantly related organisms implies greater-than-average sequence conservation. In other words, anciently diverged fungal clades should retain sequence similarity only in the most conserved of the ancestral cell wall genes. The sequence comparisons summarized in this study demonstrate precisely that.

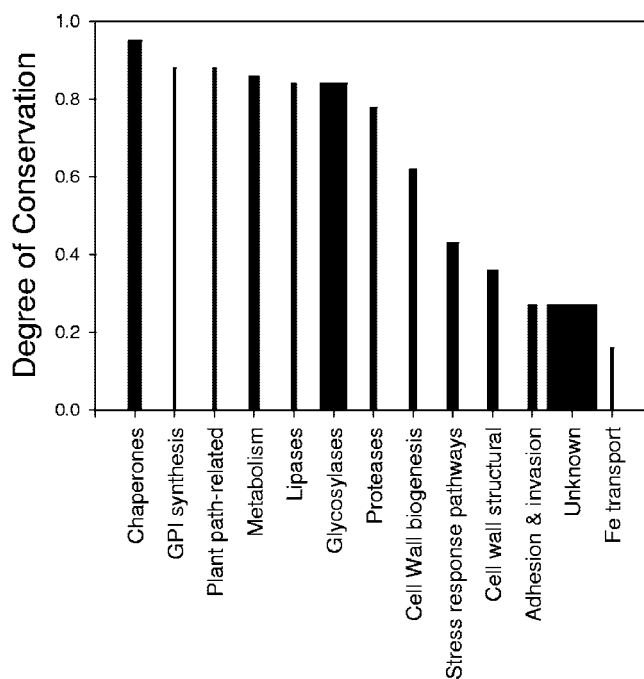


FIG. 5. Degrees of conservation for 13 functional groups of cell wall proteins. The width of each bar is proportional to the number of ORFs in that group. The degree of conservation was calculated as described in Materials and Methods.

**Prevalence of cell wall-related sequences in fungi.** The evolutionary distance between two fungal species is correlated with the divergence of their cell wall proteins (Fig. 4). The *Saccharomyces sensu stricto* species have homologs for at least 166 of the 187 *S. cerevisiae* cell wall-related ORFs, with an amino acid identity of at least 86%. The number of conserved wall genes decreases in more distant fungal clades. *Saccharomyces sensu lato* yeasts have homologs for at least 70% of the *S. cerevisiae* cell wall-related proteins; filamentous ascomycetes had fewer. Among the homologs not found in filamentous ascomycetes were genes encoding extracellular proteins, including cell wall structural proteins, adhesins, and extracellular members of the GO classification “cell wall organization and biosynthesis.” Also not conserved between yeast and filamentous ascomycetes were many of the proteins involved in the cell wall stress response pathway.

The two basidiomycete genomes in our study had homologs for 68 of the *S. cerevisiae* putative cell wall proteins (36%). The intracellular components were mostly conserved, except for the cell wall stress response pathway components. Among the cell wall-localized proteins, all but six of the glycosylases were conserved, as were lipases and proteases. Twelve GPI-anchored proteins from *S. cerevisiae* had basidiomycete homologs. The large number of homologs to the *Y. lipolytica* GPI proteins supports this concept of conservation of GPI-anchored proteins, as previously noted in genomic analyses of several fungi (12, 13, 29). Thus, basidiomycetes share with *S. cerevisiae* intracellular and extracellular activities that are critical for biosynthesis and assembly of fungal cell walls.

In the search of the nonredundant NCBI database, almost every major fungal phylum contained homologs of some *S.*

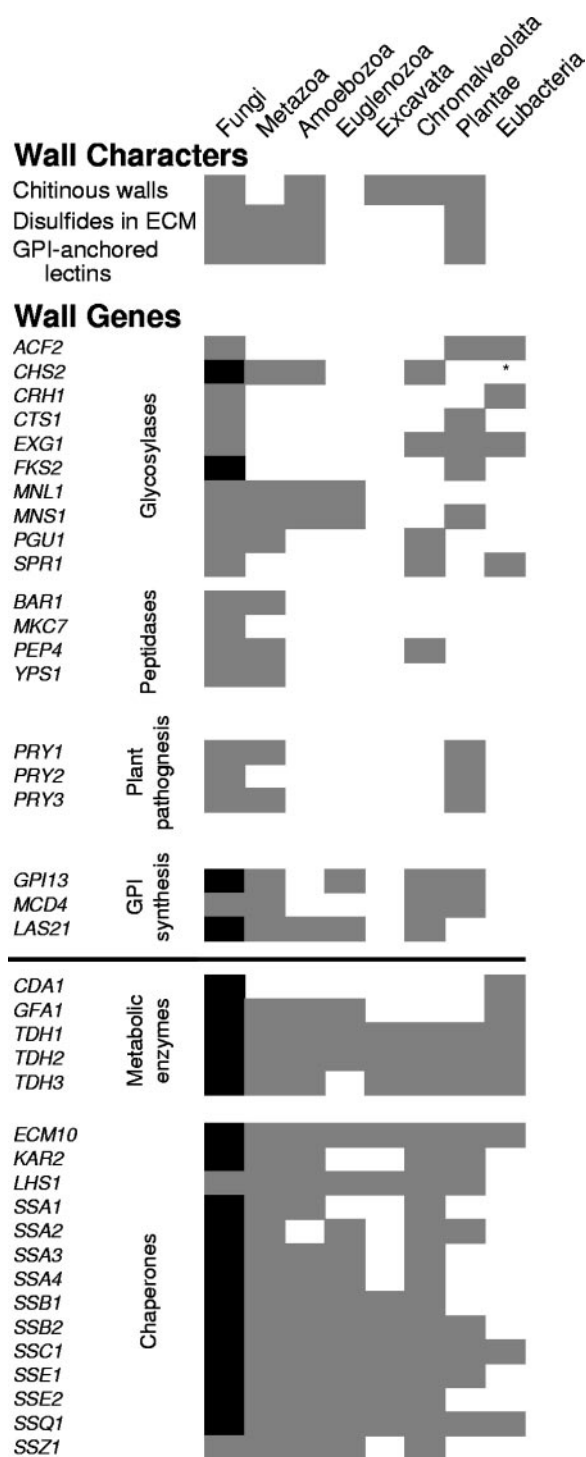


FIG. 6. Occurrence of wall characters and homologs of *S. cerevisiae* cell wall-related ORFs in other kingdoms and domains. (Top) Conserved characters in walls are shown by gray shading where known to occur. Note that an absence of shading is not definitive, because it may represent a lack of data. ECM, extracellular matrix. (Bottom) Each *S. cerevisiae* cell wall-related ORF was the query sequence for a BLAST-gtQ search through the NR database. Homologs (gray squares) were characterized by kingdom and domain. Black squares indicate ORFs present in each of the 18 fungal genomes searched and in all fungal phyla except *Glomeromycota*. There was a single bacterial *CHS2* homolog, and its uniqueness and *e*-value,  $<10^{-18}$ , implies that it is derived from horizontal gene transfer. A horizontal line sets off the two classes of proteins that are primarily intracellular.

*cerevisiae* cell wall proteins (Fig. 6). (The single exception, *Glomeromycota*, is probably due to the small number of sequences in GenBank.) Thus, homologs of some of the *S. cerevisiae* cell wall-related proteins are likely to be present throughout the fungi.

**Differences in divergence of functional classes of cell wall-related proteins.** A major finding of this analysis is that the degree of conservation among cell wall proteins is strongly correlated with the cellular roles of those proteins (Fig. 5). Many of the proteins that reside in *S. cerevisiae* cell walls have diverged to the extent that homologs were not recognized in organisms outside of the *Saccharomyces* sensu lato group. The least-conserved include the three Fit iron transport proteins and the largest class, homologs of unannotated *S. cerevisiae* cell wall ORFs. Homologs of unannotated sequences were often found only in the sensu stricto group (Fig. 5; see also Table S1 in the supplemental material). These poorly conserved proteins probably have important (45) but noncatalytic roles in walls.

Among the adhesins, most have homologs in the *Saccharomyces* sensu lato group but not in more-distant ascomycetes (Fig. 3; see also Table S1 in the supplemental material). There is anecdotal evidence that adhesins are poorly conserved, because they are subject to sexual isolation and to diversifying selection for adaptation to different environments (16, 33, 56). The structural cell wall proteins (in the *CWP*, *TIR*, *PAU*, and *DAN* families) are also poorly conserved. This observation may reflect the idea that this class is rich in proteins with very-low-complexity compositions. Within each sequence, only short segments are under selective pressure to retain sequence: the secretion and GPI signals and the relatively short glycosylation and transglycosylation sites (9, 16, 28). The majority of each sequence has low complexity and would have evolved faster than other parts of the genome (49).

Two protein classes show intermediate levels of conservation: the cell wall stress response pathways and cell wall biogenesis pathways. Their intermediate mean conservation scores result from their composition, a mixture of poorly conserved and highly conserved proteins. In general, those proteins resident in the walls are poorly conserved. Within these two classes, the plasma membrane and intracellular proteins are a mixture of conserved and nonconserved sequences (see Table S1 in the supplemental material).

**Strongly conserved proteins.** In contrast, the strongly conserved functional classes are key biosynthetic and processing enzymes that must be useful in wall biosynthesis and assembly. These include the glycosylases, GPI synthetic enzymes, proteases, and lipases. These proteins include many orthologs that may function in homologous roles in wall biogenesis across the fungi. Among the most highly conserved are three sets of partially redundant glycosylases: the *Chs* chitin synthases, the *Gas* transglycosidases, and the *Fks*  $\beta$ -1,3-glucan synthases. Other highly conserved components were peptidases, phosphatases, lipases, and enzymes of the *N*-acetylglucosamine synthesis pathway (see Table S1 in the supplemental material).

It is not surprising that chaperone sequences and triose phosphate dehydrogenases (in the metabolic enzyme class) are strongly conserved, because these are genes whose sequences are highly conserved across the biome (3, 21, 60). Their primary locations are intracellular. Although their role as cell wall

proteins is less well known and somewhat controversial, it is well documented in two ascomycetous yeasts, *S. cerevisiae* and *C. albicans* (14, 30, 40, 41, 46). Given their controversial or auxiliary role in cell walls and their dual roles as wall resident and key metabolic activities, these classes could be excluded from our analysis (“chaperones” and “metabolism” in Fig. 4 and 6). That exclusion would in fact strengthen our key observation, that wall biosynthetic genes generally have conserved sequences and wall resident proteins do not.

**Modes of sequence divergence in cell wall components.** In contrast to the homology of cell wall biosynthetic pathways, the actual composition of cell walls is lineage specific. This dichotomy is similar to the distinction in genomics between conserved “housekeeping” genes and a faster-evolving set of “accessory” genes (23). Two evolutionary mechanisms could give rise to such diversity and lineage specificity in the “accessory-like” genes: either rapid sequence divergence or frequent loss and substitution of the proteins. In the first evolutionary scenario, a homologous core set of cell wall proteins exists, but they are conserved only in protein structure and function, not in sequence. For instance, the greatest mean amino acid distance between genomes of *Saccharomyces sensu stricto* species is 16% (Fig. 1). The *S. cerevisiae* wall proteins generally have homologs within the *sensu stricto* group, and so the sequences conserved in this clade must have been conserved within the timescale corresponding to this difference. Rapid evolution would make homology unrecognizable in more distant groups. This result is consistent with the observation that many of these sequences are rich in low-complexity sequences, which are evolutionarily less constrained (49). In such a case, proteins resident in fungal cell walls may have a single origin but have diverged among different fungal lineages, through either neutral or adaptive divergence.

In the second evolutionary scenario, fungal cell wall architecture is conserved as a whole, with frequent additions and deletions of specific macromolecular components during lineage diversification. In this case, fungal cell walls are more analogous to a cell organelle that has evolved independently among lineages with few underlying homologous components. Functional features of fungal cell walls have been maintained by natural selection, similar to the multiple emergences of fins among different vertebrate groups. These fins show functional homology and anatomic convergence despite their multiple origins. Possible fungal examples might be the relocation of novel proteins to the wall following acquisition of secretion signals and GPI anchor sequences or the internal repeats that include Gln residues to be esterified to wall glucans (17). Either of these additions could happen by recombination or by insertion of foreign DNA (53, 56).

Whether the cause is a rapid sequence divergence or frequent loss and substitution, it appears beyond dispute that fungal cell wall components evolve faster than do core metabolic protein sequences (Fig. 5). Fast evolution in a large number of cell wall-related ORFs may be driven by adaptive divergence. Unlike other cell organelles, cell walls in fungi are in contact with the external environment, play a direct role in cell adaptation, and must have evolved as highly specialized and dynamic structures for colonization, host immune system evasion, signal transduction, transport, and structural maintenance. Fungi have continuously put these structures to the test

through natural selection, and the selection continues. Cell wall proteins that are harmful or neutral to the cell can be mutated or removed from the genome. As a result, the organism loses a maladaptive factor and gains efficiency in its growth and replication rates. Given that natural selection favors improved fitness and that the organism can gain two benefits from a single such event, the likelihood of selection for such mutations and deletions increases.

**Conservation of cell wall biogenesis.** Several classes of fungal cell wall-related proteins are common to eukaryotes: glycosylases, proteases, proteins needed for GPI synthesis, and the poorly characterized Pry proteins, as well as chaperones. The presence of these classes throughout the fungi and in several eukaryote kingdoms (Fig. 6) implies that they are ancestral. These genes represent a conserved set of metabolic activities that were coordinated in the formation of cell walls. Many of the conserved glycosylases have plant homologs with annotations that show roles in plant cell wall biogenesis. These proteins include glucosidases Dse4, Spr1, and Exg1, mannosidases involved in glycoprotein biogenesis (Mns1 and Mnl1), and the Fks  $\beta$ -1,3-glucan synthase subunits. The roles in plant wall biogenesis include synthesis and processing of callose ( $\beta$ -1,3-glucan), cellulose, and glycoproteins.

The enzymes of chitin metabolism are extremely highly conserved. Two enzymes in the biosynthetic pathway for *N*-acetylglucosamine (Gfa1 and YMR84w) are conserved in ascomycetes and basidiomycetes, as are chitin deacetylases (Cda1 and Cda2) (see Table S1 in the supplemental material). Fungal chitin synthases (Chs) are homologous to those in other fungi, metazoa, amoebozoa, and chromalveolata (Fig. 6). This high degree of conservation is consistent with an ancestral and continuing role of chitin in cell walls of many eukaryotes (4, 7, 43, 44, 55, 57).

Given the conservation of wall biosynthetic sequences, the broad distribution of specific wall glycoconjugates, and the similarity in roles of plant and fungal glycosylases, it is reasonable to infer that a common ancestor to fungi and other eukaryotes was an organism with an extensive carbohydrate chemistry repertoire. The commonalities identified here suggest that each eukaryotic lineage has conserved the mechanism to generate diverse extracellular structures but has altered the products of the mechanism in their ancestor to form cell walls (in plants) or some other extracellular matrices (in metazoans). This conserved mechanism uses GPI synthesis and processing enzymes, carbohydrate-processing enzymes, and chaperones to enable the localization and proper folding of the proteins (3, 21, 40, 41).

In summary, genes that encode the proteins that make up fungal cell walls have evolved so fast that their homology is often not recognized, even within the *Saccharomyces sensu lato* group. This rapid divergence appears to be driven by the great diversity of strong selective pressures on cell interactions, including mating, colony and biofilm formation, pathogenesis, and immune escape. On the other hand, there is a conserved core of sequences that are involved in wall biogenesis throughout the fungi and in other eukaryote kingdoms. The conserved metabolic capabilities in Fig. 6 are apparently ancestral. Their conservation implies that the ability to organize a wall may predate the divergence of the plants from the opisthokonts, including the fungi.

## ACKNOWLEDGMENTS

We thank Eugene Koonin for his comments.

This work was supported by NIH/NIGMS SCORE grants S06 GM 060654 and S06 GM 076168 and NIH/RCMI grant RR 030307. J.E.C. was supported in part by a fellowship from NSF MAGNET-STEM.

## REFERENCES

- Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**:3389–3402.
- Bendtsen, J. D., H. Nielsen, G. von Heijne, and S. Brunak. 2004. Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* **340**:783–795.
- Brodsky, J. L., and G. Chiosis. 2006. Hsp70 molecular chaperones: emerging roles in human disease and identification of small molecule modulators. *Curr. Top. Med. Chem.* **6**:1215–1225.
- Cabib, E. 2004. The septation apparatus, a chitin-requiring machine in budding yeast. *Arch. Biochem. Biophys.* **426**:201–207.
- Cavalier-Smith, T. 2006. Cell evolution and Earth history: stasis and revolution. *Philos. Trans. R Soc. Lond. B* **361**:969–1006.
- Cavalier-Smith, T. 2004. Only six kingdoms of life. *Proc. Biol. Sci.* **271**:1251–1262.
- Cavalier-Smith, T. 2002. The phagotrophic origin of eukaryotes and phylogenetic classification of Protozoa. *Int. J. Syst. Evol. Microbiol.* **52**:297–354.
- Cavalier-Smith, T. 2001. What are Fungi?, p. 1–37. *In* D. J. McLaughlin, E. C. McLaughlin, and P. A. Lemke (ed.), *The Mycota*. Springer, Berlin, Germany.
- Coronado, J. E., O. Attie, S. L. Epstein, W. G. Qiu, and P. N. Lipke. 2006. Composition-modified matrices improve identification of homologs of *Saccharomyces cerevisiae* low-complexity glycoproteins. *Eukaryot. Cell* **5**:628–637.
- Coutinho, P. M., E. Deleury, G. J. Davies, and B. Henrissat. 2003. An evolving hierarchical family classification for glycosyltransferases. *J. Mol. Biol.* **328**:307–317.
- Das, S., K. Van Dellen, D. Bulik, P. Magnelli, J. Cui, J. Head, P. W. Robbins, and J. Samuelsen. 2006. The cyst wall of *Entamoeba invadens* contains chitosan (deacetylated chitin). *Mol. Biochem. Parasitol.* **148**:86–92.
- De Groot, P. W., A. D. de Boer, J. Cunningham, H. L. Dekker, L. de Jong, K. J. Hellingwerf, C. de Koster, and F. M. Klis. 2004. Proteomic analysis of *Candida albicans* cell walls reveals covalently bound carbohydrate-active enzymes and adhesins. *Eukaryot. Cell* **3**:955–965.
- De Groot, P. W., K. J. Hellingwerf, and F. M. Klis. 2003. Genome-wide identification of fungal GPI proteins. *Yeast* **20**:781–796.
- Delgado, M. L., J. E. O'Connor, I. Azorin, J. Renau-Piqueras, M. L. Gil, and D. Gozalbo. 2001. The glyceraldehyde-3-phosphate dehydrogenase polypeptides encoded by the *Saccharomyces cerevisiae* *TDH1*, *TDH2* and *TDH3* genes are also cell wall proteins. *Microbiology* **147**:411–417.
- de Nobel, J. G., F. M. Klis, J. Priem, T. Munnik, and H. van den Ende. 1990. The glucanase-soluble mannoproteins limit cell wall porosity in *Saccharomyces cerevisiae*. *Yeast* **6**:491–499.
- Dranginis, A. M., J. M. Rauceo, J. E. Coronado, and P. N. Lipke. 2007. A biochemical guide to yeast adhesins: glycoproteins for social and antisocial occasions. *Microbiol. Mol. Biol. Rev.* **71**:282–294.
- Ecker, M., R. Deutzmann, L. Lehle, V. Mrsa, and W. Tanner. 2006. Pir proteins of *Saccharomyces cerevisiae* are attached to beta-1,3-glucan by a new protein-carbohydrate linkage. *J. Biol. Chem.* **281**:11523–11529.
- Eisenhaber, B., G. Schneider, M. Wildpaner, and F. Eisenhaber. 2004. A sensitive predictor for potential GPI lipid modification sites in fungal protein sequences and its application to genome-wide studies for *Aspergillus nidulans*, *Candida albicans*, *Neurospora crassa*, *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*. *J. Mol. Biol.* **337**:243–253.
- Fankhauser, N., and P. Maeser. 2006. KohGPI: identification of GPI-anchor signals by a Kohonen self organizing map. <http://gpi.unibe.ch/>.
- Fitch, W. 1971. Toward defining the course of evolution: minimum change for a specified tree topology. *Syst. Zool.* **20**:406–416.
- Fu, X., W. Jiao, and Z. Chang. 2006. Phylogenetic and biochemical studies reveal a potential evolutionary origin of small heat shock proteins of animals from bacterial class A. *J. Mol. Evol.* **62**:257–266.
- Gerstein, M. 1998. Measurement of the effectiveness of transitive sequence comparison, through a third “intermediate” sequence. *Bioinformatics* **14**:707–714.
- Hecker, M. 2003. A proteomic view of cell physiology of *Bacillus subtilis*: bringing the genome sequence to life. *Adv. Biochem. Eng. Biotechnol.* **83**:57–92.
- Hochstenbach, F., F. M. Klis, H. van den Ende, E. van Donselaar, P. J. Peters, and R. D. Klausner. 1998. Identification of a putative alpha-glucan synthase essential for cell wall construction and morphogenesis in fission yeast. *Proc. Natl. Acad. Sci. USA* **95**:9161–9166.
- James, T. Y., F. Kauff, C. L. Schoch, P. B. Matheny, V. Hofstetter, C. J. Cox, G. Celio, C. Gueidan, E. Fraker, J. Miadlikowska, H. T. Lumbsch, A. Rauhut, V. Reeb, A. E. Arnold, A. Amtoft, J. E. Stajich, K. Hosaka, G. H. Sung, D. Johnson, B. O'Rourke, M. Crockett, M. Binder, J. M. Curtis, J. C. Slot, Z. Wang, A. W. Wilson, A. Schussler, J. E. Longcore, K. O'Donnell, S. Mozley-Standridge, D. Porter, P. M. Letcher, M. J. Powell, J. W. Taylor, M. M. White, G. W. Griffith, D. R. Davies, R. A. Humber, J. B. Morton, J. Sugiyama, A. Y. Rossman, J. D. Rogers, D. H. Pfister, D. Hewitt, K. Hansen, S. Hambleton, R. A. Shoemaker, J. Kohlmeyer, B. Volkman-Kohlmeier, R. A. Spotts, M. Serdani, P. W. Crous, K. W. Hughes, K. Matsuura, E. Langer, G. Langer, W. A. Untereiner, R. Lucking, B. Budel, D. M. Geiser, A. Aptroot, P. Diederich, I. Schmitt, M. Schultz, R. Yahr, D. S. Hibbett, F. Lutzoni, D. J. McLaughlin, J. W. Spatafora, and R. Vilgalys. 2006. Reconstructing the early evolution of Fungi using a six-gene phylogeny. *Nature* **443**:818–822.
- Jose-Estanyol, M., F. X. Gomis-Ruth, and P. Puigdomenech. 2004. The eight-cysteine motif, a versatile structure in plant proteins. *Plant Physiol. Biochem.* **42**:355–365.
- Kapteyn, J. C., H. Van Den Ende, and F. M. Klis. 1999. The contribution of cell wall proteins to the organization of the yeast cell wall. *Biochim. Biophys. Acta* **1426**:373–383.
- Klis, F. M., A. Boorsma, and P. W. De Groot. 2006. Cell wall construction in *Saccharomyces cerevisiae*. *Yeast* **23**:185–202.
- Klis, F. M., P. de Groot, and K. Hellingwerf. 2001. Molecular organization of the cell wall of *Candida albicans*. *Med. Mycol.* **39**(Suppl. 1):1–8.
- Klotz, S. A., M. L. Pendrak, and R. C. Hein. 2001. Antibodies to  $\alpha_5\beta_1$  and  $\alpha_6\beta_3$  integrins react with *Candida albicans* alcohol dehydrogenase. *Microbiology* **147**:3159–3164.
- Konstantinidis, K. T., and J. M. Tiedje. 2004. Trends between gene content and genome size in prokaryotic species with larger genomes. *Proc. Natl. Acad. Sci. USA* **101**:3160–3165.
- Leal, J. A., B. Gomez-Miranda, A. Prieto, and M. Bernabe. 1993. Differences in cell wall polysaccharides of several species of *Eupenicillium*. *FEMS Microbiol. Lett.* **108**:341–345.
- Lee, J. H., S. Waffenschmidt, L. Small, and U. Goodenough. 2007. Between-species analysis of short-repeat modules in cell wall and sex-related hydroxyproline-rich glycoproteins of *Chlamydomonas*. *Plant Physiol.* **144**:1813–1826.
- Lehle, L., S. Strahl, and W. Tanner. 2006. Protein glycosylation, conserved from yeast to man: a model organism helps elucidate congenital human diseases. *Angew. Chem. (Int. Ed. Engl.)* **45**:6802–6818.
- Lesage, G., and H. Bussey. 2006. Cell wall assembly in *Saccharomyces cerevisiae*. *Microbiol. Mol. Biol. Rev.* **70**:317–343.
- Liolios, K., N. Tavernarakis, P. Hugenholz, and N. C. Kyrpides. 2006. The Genomes On Line database (GOLD) v. 2: a monitor of genome projects worldwide. *Nucleic Acids Res.* **34**:D332–D334.
- Lipke, P. N., and J. Kurjan. 1992. Sexual agglutination in budding yeasts: structure, function, and regulation of adhesion glycoproteins. *Microbiol. Rev.* **56**:180–194.
- Lipke, P. N., and R. Ovalle. 1998. Cell wall architecture in yeast: new structure and new challenges. *J. Bacteriol.* **180**:3735–3740.
- Liti, G., and E. J. Louis. 2005. Yeast evolution and comparative genomics. *Annu. Rev. Microbiol.* **59**:135–153.
- Lopez-Ribot, J. L., H. M. Allouh, B. J. Masten, and W. L. Chaffin. 1996. Evidence for presence in the cell wall of *Candida albicans* of a protein related to the Hsp70 family. *Infect. Immun.* **64**:3333–3340.
- Lopez-Ribot, J. L., and W. L. Chaffin. 1996. Members of the Hsp70 family of proteins in the cell wall of *Saccharomyces cerevisiae*. *J. Bacteriol.* **178**:4724–4726.
- Lu, C. F., J. Kurjan, and P. N. Lipke. 1994. A pathway for cell wall anchorage of *Saccharomyces cerevisiae* alpha-agglutinin. *Mol. Cell. Biol.* **14**:4825–4833.
- Maia, J. C. 1994. Hexosamine and cell wall biogenesis in the aquatic fungus *Blastocladiella emersonii*. *FASEB J.* **8**:848–853.
- Mort-Bontemps, M., L. Gay, and M. Fevre. 1997. CHS2, a chitin synthase gene from the oomycete *Saprolegnia monoica*. *Microbiology* **143**:2009–2020.
- Mrsa, V., M. Ecker, S. Strahl-Bolsinger, M. Nitz, L. Lehle, and W. Tanner. 1999. Deletion of new covalently linked cell wall glycoproteins alters the electrophoretic mobility of phosphorylated wall components of *Saccharomyces cerevisiae*. *J. Bacteriol.* **181**:3076–3086.
- Ngondi-Ekome, J., F. Thiebault, J. M. Strub, A. Van Dorselaer, R. Bonaly, C. Contino-Pepin, M. Wathier, B. Pucci, and J. Coulon. 2003. Study on agglutinating factors from flocculent *Saccharomyces cerevisiae* strains. *Biochimie* **85**:133–143.
- Nickerson, W. J. 1974. Chemical composition of cell walls and membranes of yeasts. *Ann. N. Y. Acad. Sci.* **235**:105–108.
- Nickerson, W. J., and G. Falcone. 1956. Enzymatic reduction of disulfide bonds in cell wall protein of baker's yeast. *Science* **124**:318–319.
- Romov, P. A., F. Li, P. N. Lipke, S. L. Epstein, and W. G. Qiu. 2006. Comparative genomics reveals long, evolutionarily conserved, low-complexity islands in yeast proteins. *J. Mol. Evol.* **63**:415–425.
- Schaffer, A. A., L. Aravind, T. L. Madden, S. Shavirin, J. L. Spouge, Y. I. Wolf, E. V. Koonin, and S. F. Altschul. 2001. Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. *Nucleic Acids Res.* **29**:2994–3005.
- Simpson, A. G., and A. J. Roger. 2004. The real “kingdoms” of eukaryotes. *Curr. Biol.* **14**:R693–R696.



52. Southard, S. B., C. A. Specht, C. Mishra, J. Chen-Weiner, and P. W. Robbins. 1999. Molecular analysis of the *Candida albicans* homolog of *Saccharomyces cerevisiae* MNN9, required for glycosylation of cell wall mannoproteins. *J. Bacteriol.* **181**:7439–7448.
53. Sprague, G. F., Jr. 1991. Genetic exchange between kingdoms. *Curr. Opin. Genet. Dev.* **1**:530–533.
54. Steenkamp, E. T., J. Wright, and S. L. Baldauf. 2006. The protistan origins of animals and fungi. *Mol. Biol. Evol.* **23**:93–106.
55. Van Dellen, K., S. K. Ghosh, P. W. Robbins, B. Loftus, and J. Samuelson. 2002. *Entamoeba histolytica* lectins contain unique 6-Cys or 8-Cys chitin-binding domains. *Infect. Immun.* **70**:3259–3263.
56. Verstrepen, K. J., A. Jansen, F. Lewitter, and G. R. Fink. 2005. Intragenic tandem repeats generate functional variability. *Nat. Genet.* **37**:986–990.
57. Ward, H. D., J. Alroy, B. I. Lev, G. T. Keusch, and M. E. Pereira. 1988. Biology of *Giardia lamblia*. Detection of N-acetyl-D-glucosamine as the only surface saccharide moiety and identification of two distinct subsets of trophozoites by lectin binding. *J. Exp. Med.* **167**:73–88.
58. Weig, M., L. Jansch, U. Gross, C. G. De Koster, F. M. Klis, and P. W. De Groot. 2004. Systematic identification *in silico* of covalently bound cell wall proteins and analysis of protein-polysaccharide linkages of the human pathogen *Candida glabrata*. *Microbiology* **150**:3129–3144.
59. West, C. M. 2003. Comparative analysis of spore coat formation, structure, and function in *Dictyostelium*. *Int. Rev. Cytol.* **222**:237–293.
60. Yarbrough, P. O., M. A. Hayden, L. A. Dunn, P. S. Vermersch, M. R. Klass, and R. M. Hecht. 1987. The glyceraldehyde-3-phosphate dehydrogenase gene family in the nematode, *Caenorhabditis elegans*: isolation and characterization of one of the genes. *Biochim. Biophys. Acta* **908**:21–33.
61. Yin, Q. Y., P. W. de Groot, H. L. Dekker, L. de Jong, F. M. Klis, and C. G. de Koster. 2005. Comprehensive proteomic analysis of *Saccharomyces cerevisiae* cell walls: identification of proteins covalently attached via glycosylphosphatidylinositol remnants or mild alkali-sensitive linkages. *J. Biol. Chem.* **280**:20894–20901.
62. Yona, G., N. Linial, and M. Linial. 2000. ProtoMap: automatic classification of protein sequences and hierarchy of protein families. *Nucleic Acids Res.* **28**:49–55.
63. Yoshida, M., N. Sakuragi, K. Kondo, and E. Tanesaka. 2006. Cleavage with phospholipase of the lipid anchor in the cell adhesion molecule, csA, from *Dictyostelium discoideum*. *Comp. Biochem. Physiol. B* **143**:138–144.
64. Yu, Y. K., and S. F. Altschul. 2005. The construction of amino acid substitution matrices for the comparison of proteins with nonstandard compositions. *Bioinformatics* **21**:902–911.