ORIGINS OF VARIATION IN THE FUNGAL CELL SURFACE

Kevin J. Verstrepen, Todd B. Reynolds‡ and Gerald R. Fink**

The increase in hospital-acquired fungal infections has been attributed to the ability of fungi to adhere not only to human tissues, but also to the plastic prostheses and invasive devices that are used to treat disease. These properties are conferred by a family of fungal cell-surface proteins, called adhesins. Adhesins might also have a central role in the formation of fungal biofilms, which are resistant to antimicrobial drugs. The structure of the genes that encode adhesin-family members, and the sequence homology between them, enables genetic reshuffling of domains to form new genes. Coupled with epigenetic changes in gene expression, these genetic rearrangements provide a reservoir of cell-surface molecules with new functions.

Fungal infections pose a mounting clinical problem, especially for the vulnerable population of immunocompromised patients. Although *Candida albicans* is the most common fungal pathogen of humans, *Candida glabrata* is now emerging as another important pathogen[1,2]. Even *Saccharomyces cerevisiae*, which was once thought of as a harmless colonizer of plants and a benign laboratory workhorse, has now been identified as a human pathogen, presumably because it is used in several countries as a biotherapeutic agent to prevent antibiotic-induced diarrhoea[3].

The presence of these fungi in internal organs raises the question of how they gain access to and proliferate within these sites. The ability of *Candida* spp. to gain access to tissues that they do not normally colonize in healthy humans is likely to result from their adhesion to plastic devices such as catheters, prosthetic heart valves, cardiac pacemakers, endotracheal tubes, dentures and cerebrospinal-fluid shunts[4]. This 'piggybacking' theory is supported by the correlation of intravenous catheterization with infections of the blood by *C. albicans* (87%)[5]. The mortality rate that is associated with such blood infections, which is quite high (41%), drops by almost half (to 21%) if the catheter is removed[6]. It has been suggested that the proliferation of fungi on inert surfaces might involve a form of colony growth that is known as biofilm formation, which produces a reservoir of infective cells that have increased antibiotic resistance[4,7].

The immune system recognizes distinctive properties of the cell surface of microorganisms. Many bacteria and protozoans have the ability to switch their cell-surface molecules, a tactic that permits them to elude the immune system and adhere to diverse materials and cells. The immune system also poses similar challenges to fungi, which have characteristic cell-surface molecules that are recognized by dedicated phagocytic cells[8–10].

These considerations raise several questions about fungal adhesion. First, what are the cell-surface molecules that permit the adhesion of fungi to inert surfaces and to mammalian cells? Second, what generates the diversity of these cell-surface molecules? Third, are these adhesion molecules involved in biofilm formation?

The fungal adhesins

The genomes of many fungi contain a family of cell-wall glycoproteins — known as adhesins — that confer unique adhesion properties[11–13]. These molecules are required for the interactions of fungal cells with each other (flocculation and filamentation)[11,13–15], inert surfaces (agar and plastic)[14,16,17] and mammalian tissues[18–20]. They might also be crucial for the formation of biofilms[16,21,22].

In pathogens such as *C. albicans* and *C. glabrata* (in which adhesins are encoded by the *ALS* (agglutinin-like sequence) genes[12] and the *EPA* (epithelial adhesin) genes, respectively[23]), these proteins are responsible for

**Whitehead Institute for Biomedical Research/MIT, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA.*
‡Department of Microbiology, University of Tennessee, Knoxville, Tennessee 37996, USA.
Correspondence to G.R.F.
e-mail: gfink@wi.mit.edu
doi:10.1038/nrmicro927

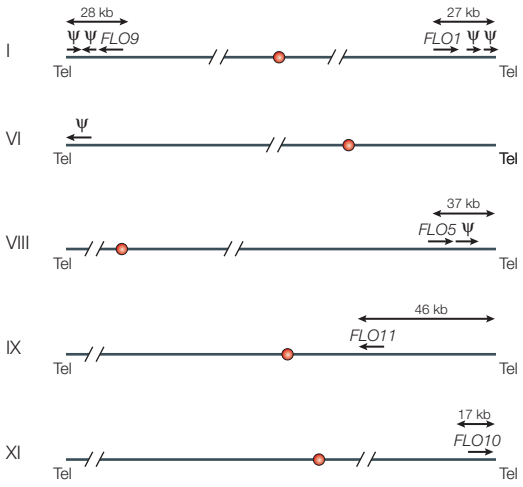

Figure 1 | **Chromosomal localization of adhesin genes and pseudogenes in the *Saccharomyces cerevisiae* laboratory strain S288C.** The numbers on the left indicate the yeast chromosomes on which the *FLO* genes reside; the red dots represent the centromeres. Ψ indicates the sites of *FLO* pseudogenes on the chromosomes. The silent *FLO* genes and pseudogenes are all located within 40 kb of the telomeres (Tel). *FLO11* is neither centromeric nor telomeric. A BLAST search of other sequenced species of the *Saccharomyces* genus[68] shows that these organisms also have orthologues of the *S. cerevisiae FLO* genes.

adhesion to mammalian tissues and plastic. In *S. cerevisiae*, the *FLO* (flocculation) genes confer adhesion to agar, plastic and other yeast cells (flocculation)[11]. The adhesin gene family within each organism consists of many genes, and the protein encoded by each gene has considerable sequence identity with other family members[11–13,24]. Nonetheless, the different adhesin genes confer distinct cell-surface properties and function as a resource for cell-surface variation[11,25,26]. Many members of these fungal gene families are located near TELOMERES (FIG. 1), and this might be important for their expression, genetic interactions and evolution[13,23,27]. Although many other factors might contribute to adhesion, the adhesins seem to have an important role in several fungi, and these glycoproteins are now being studied intensively.

Structure of adhesins

The adhesins of *S. cerevisiae*, *C. albicans* and *C. glabrata* form a superfamily that is united by a common structure. Members of the adhesin family have a modular configuration that consists of three domains (A, B and C; FIG. 2) and an amino-terminal secretory sequence that must be removed as the protein moves through the secretory pathway *en route* to the plasma membrane[28]. The N-terminal domain (A) is thought to provide much of the affinity of these proteins for surfaces[28,29]. The *C. albicans ALS* genes have considerable sequence similarity in the region that encodes the A domain, whereas some of the *S. cerevisiae FLO* genes show less similarity in this region. This domain is followed by a segment of variable length (domain B) that is extremely rich in serine and threonine residues (for example, domain B in *FLO11* contains 60% serine/threonine residues) and contains many tandem repeats. The carboxy-terminal region (domain C), which follows the serine/threonine-rich region, is conserved only among some subgroups within the adhesin gene families (for example, it is conserved among *FLO1*, *FLO5* and *FLO9*).

The adhesins undergo several post-translational modifications. These proteins must move from the endoplasmic reticulum (ER), through the Golgi and pass through the plasma membrane as they exit the cell to reach their final destination in the cell wall[28,30]. A glycosyl phosphatidylinositol (GPI) anchor is added covalently to the C terminus in the ER, and mannose residues are added in the Golgi to the many serine and threonine residues in domain B[30–33]. The presence of many Asn-X-Thr/Ser sequences (where X represents any amino acid) indicates that there might also be considerable *N*-glycosylation. The adhesin proteins are ultimately linked covalently through a GPI remnant to the β-1,6-glucans that protrude from the fungal cell wall[34,35]. Studies of the Epa1 protein of *C. glabrata* indicate that the amino acids that are located around the site of GPI addition are crucial for the localization of Epa1 to the cell wall[36].

The structural and functional analysis of the adhesin multigene families is most straightforward in *S. cerevisiae*, which has a highly tractable genetic system and for which a complete genome sequence has been assembled. These features make it possible to analyse the functions of each member of a multigene family. The genome sequence shows that there are five unlinked *FLO* genes in the adhesin family (*FLO1*, *FLO5*, *FLO9*, *FLO10* and *FLO11*), four of which (*FLO1*, *FLO5*, *FLO9* and *FLO10*) are located adjacent to telomeres (FIG. 1). The fact that only *FLO11* is expressed in laboratory strains of *S. cerevisiae* — whereas *FLO1*, *FLO5*, *FLO9* and *FLO10* are transcriptionally silent[37] — simplifies the functional analysis of this family, as all adhesion-specific phenotypes in these strains are dependent on a functional *FLO11* gene[11,14].

The SUBTELOMERIC *FLO* genes are transcriptionally silent, but their functions can be assessed by replacing the endogenous promoter for each gene with the inducible *GAL1* promoter[11]. The expression of these *GAL1–FLO* fusions was determined for the silent genes

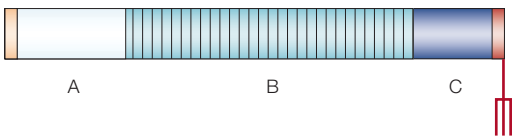

Figure 2 | **Domain structure of adhesins.** Adhesins comprise three domains — A, B and C — which are preceded by an amino-terminal signal sequence. The N-terminal domain (A) is thought to confer adhesion. The central domain (B) contains a serine/threonine-rich region that is encoded by many repeated nucleotide sequences. The carboxy-terminal domain (C) contains a site for the covalent attachment of a glycosyl phosphatidylinositol anchor (shown in red).

TELOMERES
The physical ends of linear chromosomes. They are associated with specialized nucleoprotein complexes that are required for the protection, replication and stabilization of the chromosome ends. In most organisms, telomeres contain many tandemly repeated DNA sequences called 'terminal repeats'.

SUBTELOMERIC REGIONS
DNA sequences close to telomeres. Genes in these regions are often found in multiple copies on different chromosomes and might be subjected to common regulatory mechanisms as a consequence of their proximity to the telomeres.

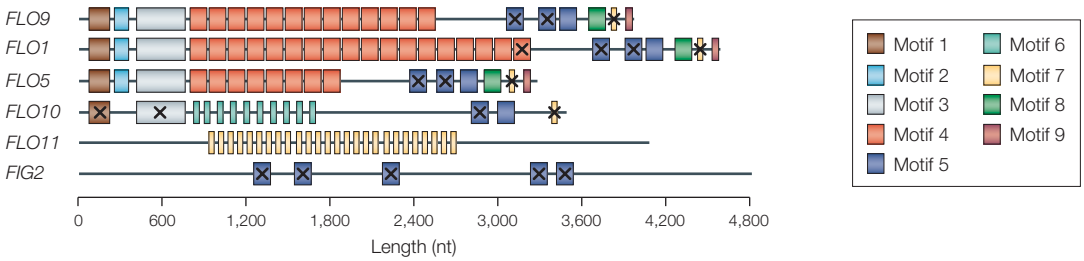

Figure 3 | **Repeated nucleotide motifs in the *FLO* genes.** The five *FLO* genes and a gene with a related structure (*FIG2*), but which has only limited sequence similarity with the *FLO* genes, are shown. The boxes show highly conserved DNA-sequence motifs. The nucleotide motifs were established by first identifying conserved amino-acid sequences within each protein using the Meme[69], BLAST[70], Etandem[71] and ClustalX[72] programs, and then identifying the corresponding DNA sequences. The analysis required the designation of a repeat size that is large enough to represent a good target for recombination (>30 nucleotides). The numbers below the sequences show the distance in nucleotides from the translational start signal. Boxes marked with an 'X' represent repeated amino-acid motifs that are not conserved in the DNA sequence (see FIG. 5 for a detailed analysis of motif 1 in *FLO1* and *FLO10*). These sequences are thought to have undergone genetic drift over time, resulting in many third-position nucleotide changes within codons.

in a *flo11*-mutant background, so that any adhesive phenotype was a consequence of the *FLO* gene being studied. The results showed that the silent genes are functional when they are expressed from the *GAL1* promoter. *FLO10* and *FLO11*, but not *FLO1*, promote both adhesion to agar and plastic, and filamentation. *FLO1* and, to a lesser extent, *FLO10*, promote cell–cell adhesion, but *FLO11* does not[11]. So, expression of each of the Flo proteins provides new cell-surface antigens, as well as new adhesion characteristics.

The analysis of adhesin gene function in *Candida* species is impeded at present by the large number of family members. Although the assembly of the *C. albicans* genome is not complete, there are at least eight members of the *ALS* family and there might be many more adhesin genes — such as *EAP1* (REFS 19,26,38), *INT1* (REF. 39) and *HWP1* (REFS 40–42) — that are not strictly considered to be family members.

Several laboratories have generated mutant strains in both *C. glabrata* and *C. albicans* in which specific members of the adhesin gene family have been knocked out. However, these strains often have weak phenotypes or no phenotype at all[18,23,43,44]. This is presumably because of overlapping functions of adhesin family members, so that a defect that is caused by one mutation is compensated for and masked by the remaining functional copies of other family members. Some information has been gleaned from overexpressing *Candida* genes in *Candida* mutants that are defective in the regulation of filamentation, such as the *efg1* mutant. For example, overexpression of *ALS1* enhanced filamentation, flocculation and adhesion in the *efg1* mutant[44]. However, the phenotypes that are conferred by overexpression can often be misleading.

Another method for studying the function of *Candida* genes is to transform them into *S. cerevisiae* and study the phenotypes of the resulting transformants. For example, a *Candida* gene might be transformed into an *S. cerevisiae* strain to determine whether heterologous expression confers the ability to adhere to animal cells. As *S. cerevisiae* does not adhere to animal cells, increased adhesion of a transformant provides evidence that the introduced *Candida* gene has a role in this tropism. Using this system, several *ALS* genes have been shown to enable *S. cerevisiae* to adhere to extracellular-matrix (ECM) proteins[17], human umbilical-vein endothelial cells[43] and various other substrates[45]. In a similar assay, expression of the *C. glabrata EPA1* gene was shown to enable *S. cerevisiae* to bind to epithelial cells[18]. Of course, the interpretation of these heterologous expression experiments must be tempered by the recognition that a gene that is expressed in a foreign host might not show the complete range of its functions or might show functions that it does not have in its natural host.

Generation of adhesin diversity by recombination

Sequence analysis shows that there are several DNA motifs that are conserved among different *S. cerevisiae* *FLO* genes (FIG. 3). The most striking conservation is in motif 4, which is a highly conserved sequence of about 100 nucleotides that is present as a tandem repeat in the regions that encode domain B in *FLO1* (17 copies), *FLO9* (13 copies) and *FLO5* (8 copies) (FIG. 4). Both *FLO10* and *FLO11* also have tandem nucleotide repeats in the regions that encode domain B, but these are all distinct from motif 4. There are other large motifs at both the 5′ and 3′ ends of each gene that are shared among several *FLO* genes. A comparison of the proteins that are encoded by these genes at the amino-acid level provides a further indication of the close relationship between the Flo proteins. There are several highly conserved amino-acid motifs that are not highly conserved at the DNA level (see motifs 3, 5 and 7 in FIG. 3). Presumably, these segments were retained because of their importance for the function of the protein.

The conservation of the DNA motifs within individual *FLO* genes and among the *FLO* family indicates that they have functional significance. Without strong selection for the conservation of these DNA motifs they would have degenerated over time, accumulating — at the very least — many changes at the third-nucleotide positions of codons. Such drift away from a conserved DNA motif to a conserved amino-acid

Figure 4 | **The conserved nucleotide sequence of motif 4 of the *FLO* genes.** The high level of nucleotide-sequence identity in motif 4 within each gene and between different *FLO* genes is shown. The sequences were derived using several bioinformatics tools (see FIG. 3). The numbers on the left of each sequence represent the location of the sequence in the open reading frame. The numbers at the bottom indicate the nucleotide position with respect to the start of the aligned sequences. Asterisks below the sequence indicate positions at which the same nucleotide is present in each *FLO* gene.

motif can be seen in DNA motif 1 of *FLO10* (FIG. 5). However, other motifs, such as motif 4, are not only highly conserved at the amino-acid level, but also at the nucleotide level. The simplest explanation for the conservation of the nucleotide motifs is that these sequences provide sites for homologous recombination, and that these exchange events provide some selective advantage. New adhesins might, for example, confer greater adhesion to solid surfaces or new adhesion properties that enable attachment to different surfaces or cells.

Recombination within adhesin genes. The motifs within each adhesin gene, their presence in multiple copies and the chromosomal arrangement of duplicated genes provide the basis for a constantly shifting repertoire of cell-surface molecules (FIG. 6). Recombination between the tandem repeats of a motif within a single adhesin gene can provide an inexhaustible source of diversity. UNEQUAL CROSSOVER between homologues in a diploid cell, or during the G2 phase of the cell cycle in a haploid cell, might produce larger and smaller versions of one of the adhesins. Slippage during replication would have the same effect. Mechanisms that are similar to these have been proposed for variations in MINISATELLITE regions in the *S. cerevisiae* genome that occur during mitosis[46] and meiosis[47]. The loss of repeated motifs in one of the *FLO* genes has been associated with loss of flocculation in strains of *S. cerevisiae* that that are used in brewing, indicating that the number of repeats correlates with the strength of cell–cell adhesion[48,49].

In *C. albicans*, the evidence for such intragenic recombination events is compelling. In strain SC5314, the adhesin genes *ALS1, ALS5* and *ALS9* are present in a long form on one chromosome and in a short form on its homologue. The most notable difference between the long and short forms is the number of repeated nucleotide motifs: the two forms of *ALS1* differ by 12 repeats, those of *ALS5* by 1 repeat and those of *ALS9* by 3 repeats[26]. All of the long forms are linked on the same homologue. The variation in the number of repeats means that although *C. albicans* has three *ALS* genes on chromosome 6, this region can specify

UNEQUAL CROSSOVER
A recombination event between DNA sequences that are not correctly aligned. This often occurs in repetitive sequences when the repeat units (DNA motifs) are paired out of register.

MINISATELLITES
Also known as variable-number tandem repeats. These are DNA sequences of variable length that consist of many tandemly repeated DNA motifs of 5–35 base pairs. Minisatellite regions are unstable and often expand or contract during meiosis and mitosis, making them good targets for genotyping.

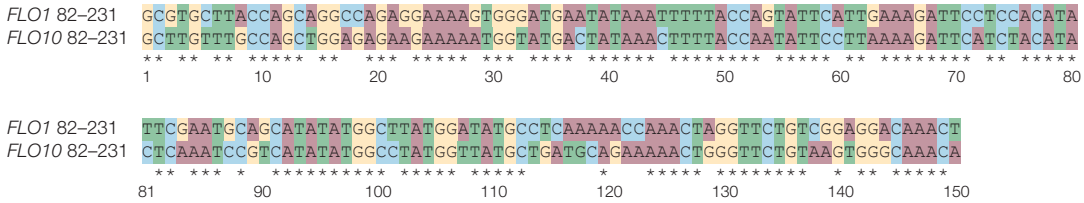

Figure 5 | **The low level of conservation of third-position nucleotides in nucleotide motif 1 of *FLO10*.** The amino-acid sequence corresponding to motif 1 is conserved in *FLO1*, *FLO5*, *FLO9* and *FLO10* (see FIG. 3). In *FLO1*, *FLO5* and *FLO9*, the motif is also conserved at the nucleotide level. In *FLO10*, however, many third-position changes have occurred. The numbers on the left of each sequence represent the location of the sequence in the open reading frame. The numbers below the sequence indicate the nucleotide position with respect to the start of each sequence. The asterisks below the nucleotides indicate positions at which there is identity between the two sequences. Statistical analysis (using a chi-squared test) indicates that the low level of conservation of the third-nucleotide position compared with that of the first and second positions is statistically significant ($p < 0.01$).

six different adhesins. This variation might be of medical importance: analysis of clinical isolates has identified as many as 60 different alleles of *ALS7* (REF. 25). The differences between alleles are largely caused by rearrangements between the repeat elements[25]. Clearly, recombination between the repeated nucleotide motifs is ammunition for the propagation of an endless supply of new cell-surface proteins.

Recombination between adhesin genes. The adhesin genes are often linked in tandem arrays. This close proximity provides further opportunities for the exchange of information between homologous sequences (FIG. 6). The *EPA1*, *EPA2* and *EPA3* genes of *C. glabrata* are closely linked, as are the *ALS1*, *ALS5* and *ALS9* genes of *C. albicans*. The sequence similarity between adjacent adhesin genes provides the possibility of unequal crossovers that could generate strains that have duplications and deletions, as well as new hybrid alleles. Indeed, *C. albicans* strains have been described that lack either *ALS1* (REF. 26) or *ALS5* (REF. 50). Although *C. albicans* has no known meiotic cycle, there is ample opportunity for these unequal crossover events to be produced by mitotic recombination. As this organism is diploid, the events that are described above could also take place between *ALS* genes on homologous chromosomes.

Although the *S. cerevisiae FLO* genes are not closely linked in the genome, the *FLO1*, *FLO5* and *FLO9* genes have adjacent, truncated, non-functional copies, which are annotated as pseudogenes in the database[51] (FIG. 1). These defective *FLO* genes are probably the relics of unequal crossover events rather than processed cDNAs, as none of the functional *FLO* genes has introns. Although they are non-functional, these pseudogenes provide a reservoir of sequences that could become incorporated into the adjacent functional *FLO* genes.

The sequence similarity between different adhesin genes means that even *FLO* genes on different chromosomes can interact genetically. Recombination between adhesin genes on different chromosomes would result in translocations. Such interchromosomal events might have generated the Lg-*FLO1* gene, which is responsible for the increased flocculation of certain strains used in brewing. It has been proposed that the Lg-*FLO1* gene is a chimaera that results from a translocation between the 5′ non-coding region of the pseudogene *YHR211* on chromosome VIII and the 3′ non-coding region of pseudogene *YAL065* on chromosome I[29].

PSEUDOHYPHAL GROWTH
A form of cell division that results in a filament of elongated cells. The pseudohyphal filament differs from a hypha because each member of the pseudohyphal filament is a distinct cell, whereas the hypha is a long, multi-nucleate filament without separate cells.

Differential regulation generates diversity

The *FLO11* gene of *S. cerevisiae* is the only adhesin gene that is expressed in vegetative cells of laboratory strains; the other *FLO* genes are silent[27]. This makes it possible to study the phenotypic consequences of *FLO11* regulation without interference from the other family members. Nutritional signals control *FLO11* gene expression through both the mitogen-activated protein kinase and protein kinase A pathways[52]. Under conditions of nitrogen starvation, these pathways induce the expression of *FLO11*. Even when cells are grown under nutritional conditions that promote expression, some cells express *FLO11* and some do not[27]. This switching of *FLO11* between 'on' and 'off' states is due to reversible epigenetic repression by chromatin-binding proteins[27].

The morphogenetic consequences of *FLO11* switching are striking: the cells that express *FLO11* form a filament, whereas those that do not express *FLO11* continue to divide in the single-celled yeast form. The 'on' state persists through subsequent generations, producing a phalanx of PSEUDOHYPHAL cells that make up the intact filament[27]. This switching means that even a strain with a single *FLO11* gene has cells with two different cell surfaces: those that have Flo11 in their cell walls and those that do not.

The subtelomeric *FLO1*, *FLO5* and *FLO10* genes provide another reservoir of cell-surface variation. So far, no physiological conditions have been discovered that release the telomeric *FLO* genes from repression. However, the silent genes can be activated by *ira* mutations that occur at high frequency. In *ira* mutants, the *FLO10* gene is expressed and confers hyperfilamentation and hyperadhesion[27]. Even when the *FLO10* gene is on, its expression is epigenetically unstable and switches from the 'on' to the 'off' state at a high frequency (it is in each state for ~50% of the time). The 'off' state is maintained by Sir3 and other components

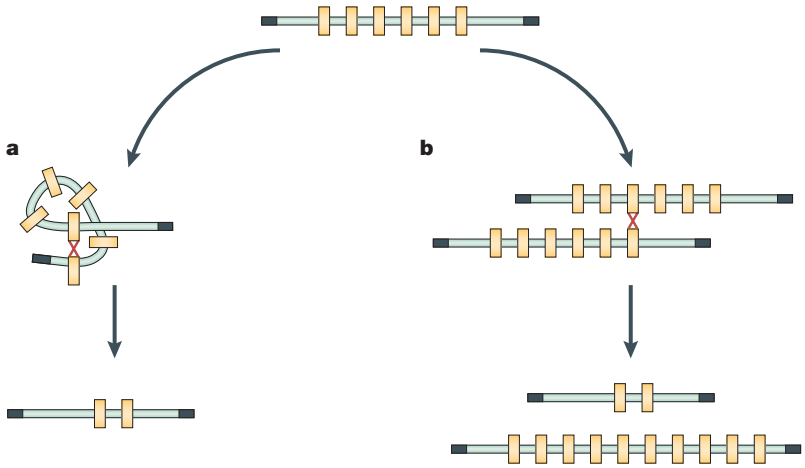

Figure 6 | **Recombination between repeated DNA motifs in adhesin genes generates new alleles.** Boxes indicate nucleotide motifs; the regions shown in black at the ends of the chromosomes represent telomeres. **a** | Intrachromosomal pairing coupled with a recombination event can generate a short gene with a reduced number of repeats. **b** | An unequal crossover between two identical *FLO* genes on homologous chromosomes that have not aligned perfectly is shown. Similarly, *FLO* genes with significant homology that are located on different chromosomes could recombine to produce new chimeric genes. This event would generate both a long and a short form of the gene. Although simple reciprocal recombination events are shown, the amplification and loss of repeat motifs could occur by many mechanisms, including slippage during replication and double-strand breaks. It is interesting to note that the *FLO* genes that are adjacent to telomeres on different chromosomes all have the same orientation with respect to the centromere. The fact that they are all transcribed towards the telomere means that interchromosomal recombination would reconstitute functional chromosomes with a single centromere, and not dicentrics (chromosomes with two centromeres).

of the complex that produces silent chromatin at the telomeres. As a consequence of unstable silencing at both *FLO10* and *FLO11*, a genetically homogeneous population of *S. cerevisiae* that carries de-silenced *FLO10* can consist of four types of cells with different cell-surface proteins: $Flo11^{+}Flo10^{+}$, $Flo11^{+}Flo10^{-}$, $Flo11^{-}Flo10^{+}$ and $Flo11^{-}Flo10^{-}$.

This pattern — in which one family member is expressed and the others are silenced — might also apply to *C. glabrata*. For example, in this organism only *EPA1* is expressed, whereas *EPA2*, *EPA3*, *EPA4* and *EPA5* are not[23]. This transcriptional silencing seems to be due to the location of these genes in the subtelomeric regions, because the regional repression depends on functional Sir3 (REF. 23). In *C. albicans*, the situation is different; more than one of the *ALS* genes is expressed in the strains that have been studied[21]. Moreover, there is no evidence that the *ALS* genes are located in the telomeric regions.

It has been suggested that variation in *FLO* gene expression could also be controlled by the unstable yeast prion PSI^{+} (REF. 51). PSI^{+} is the heritable prion form of the *S. cerevisiae* Sup35 protein, which mediates translational termination at nonsense codons. PSI^{+} increases nonsense suppression, whereas the PSI^{-} non-prion form of Sup35 reduces it[53]. Many strains of *S. cerevisiae* have a nonsense mutation in the *FLO8* gene. Flo8 is a transcriptional regulator of the expression of *FLO* genes: *flo8* mutants are unable to express any of these genes[37]. It has been proposed that the *flo8* nonsense mutation might be suppressed by the PSI^{+} protein, thereby permitting the expression of the *FLO* genes and increasing adhesion. As the PSI^{+} and PSI^{-} states switch back and forth at high frequencies, the prion might contribute to the presence or absence of Flo proteins on the cell surface.

Adhesins and biofilms

The adhesive properties that are conferred by these fungal cell-surface molecules have prompted the investigation of their role in infection. Disseminated candidiasis is usually associated with plastic medical devices, such as catheters, that are thought to provide the vehicle for fungal invasion[54,55]. It is presumed that the cells attach to the device, hitch a ride as the device is implanted and that the device provides a continuing source of fungal cells for further infiltration and proliferation. The high rates of mortality that are associated with these infections (41%)[6] place the adhesins centre stage in modern medicine.

Studies using several experimental systems have indicated that the adhesins have a role in fungal adhesion to plastic. The catheter-disc model measures the increase over time in the number of *C. albicans* cells that adhere to a disk or strip that is cut from the catheter material (polyvinylchloride or polyurethane). Cells are allowed to adhere to the disk, which is washed to remove the non-adherent cells, and the cells are then incubated in growth medium. The cell mass on the disk increases with time. Analysis by northern blotting indicates that the *ALS1* gene, and perhaps other genes in this family, are upregulated in the adherent cells as compared with those that are grown in the absence of the disk[56].

A second method uses the heterologous expression of a *Candida* gene in a non-adherent mutant strain of *S. cerevisiae* to determine whether the candidate gene is capable of conferring the ability to adhere to plastic. A recent study that combined these two methods illustrates the key role of the *Candida* adhesins in adhesion to plastic, as well as in many other morphogenetic changes that are important for pathogenesis[19]. This study used a chamber in which cells on a plastic surface were subjected to a controlled flow of medium. An *S. cerevisiae flo8* mutant was transformed with a library of *Candida* genes and then screened for adherent cells in the flow chamber. Recycling the adherent population through the device identified *EAP1*, a *C. albicans* gene that increased the adhesion of *S. cerevisiae* to polystyrene[19]. Eap1 is a new member of the adhesin family that has some homology to Flo1 and other *S. cerevisiae* adhesins.

The properties of the *S. cerevisiae flo8* mutant expressing *EAP1* are extremely informative. The *flo8* mutant does not express any of the *S. cerevisiae* adhesins. It is therefore defective in the adhesion of the haploid form to agar and polystyrene, filamentation of the diploid form and film formation (see below). The transformant can be tested for the ability of the heterologous gene to restore these functions in the mutant. The *EAP1* gene not only confers adhesion to polystyrene, but also restores invasive and filamentous growth to *S. cerevisiae* mutants that are defective in these properties[19]. Remarkably, *EAP1* also enhances the

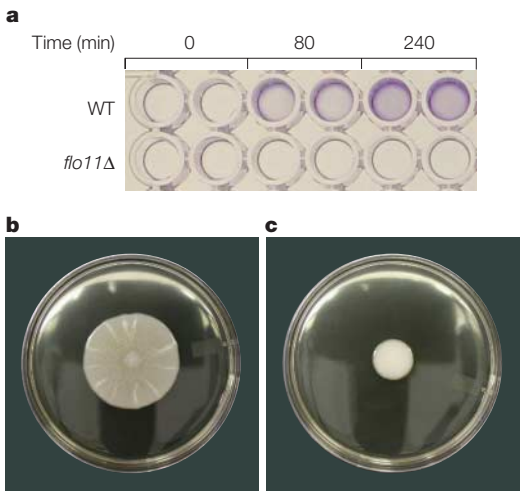

Figure 7 | ***Saccharomyces cerevisiae FLO11* is required for adhesion to plastic and film formation. a** | Wild-type (WT) and *flo11Δ* cells were incubated for predetermined lengths of time in polystyrene wells. The cells were stained with crystal violet and the wells were washed thoroughly to remove non-adherent cells. The test shows that wild-type cells adhere to the plastic, whereas *flo11Δ* cells do not. **b** | Formation of a film with extensive morphological differentiation when wild-type cells are inoculated onto a low-agar (0.3%) plate. **c** | *flo11Δ* cells fail to form a film on low-agar plates.

attachment of *S. cerevisiae* to HEK293 human embryonic kidney epithelial cells[19]. These attributes indicate that *EAP1* might provide many of the characteristics that are associated with the pathogenicity of *C. albicans*. However, the genetic repertoire that is responsible for the adherent properties of *C. albicans* is clearly more complex and includes not only the entire ensemble of *ALS* genes[57], but also genes such as *INT1* and *HWP1*, which have no homology to the *ALS* family[39].

The ability of fungi to adhere to many inert substances has led to the suggestion that *Candida* spp. and other species form biofilms, by analogy to the adherent structures that are formed by bacteria[4,22,58]. Although there is no universally accepted definition of a biofilm, several authors have attempted to describe biofilms in a manner that distinguishes this type of association from other forms of bacterial growth. These descriptions include: "…communities of microorganisms that are attached to a surface"[59]; "…assemblages of microbes at an interface that are typically attached to a surface"[60]; and "…populations of adherent microbes enclosed by a matrix"[61]. Although many other phenotypes, such as new morphologies, have been attributed to bacterial biofilms, the most important feature of medical significance is their high level of resistance to antibiotics[7,56,62].

Given the diversity of microbial lifestyles, none of these attributes is common to all of the structures that are called biofilms. For example, the *Pseudomonas aeruginosa* biofilm is composed of cells that are suspended in the exopolysaccharide alginate[63]. By contrast, *Mycobacterium smegmatis* does not secrete an ECM, but it does form a film on wet agarose Petri plates and polyvinylchloride[64]. These properties are dependent on capsular glycopeptidolipids, and not secreted molecules, because mutants that are defective in the deposition of these glycopeptidolipids do not adhere to polyvinylchloride or form a film on agarose.

S. cerevisiae, like *M. smegmatis*, forms a film on low-agar Petri plates[16] (FIG. 7). The formation of this film, and adhesion to polystyrene, requires the adhesin Flo11. Mutants that lack functional Flo11 do not form this film or the elaborate structures that are associated with it. These properties seem to be unique to Flo11, as none of the other adhesins that have been tested form the film or increase adhesion to plastic. Cells that express Flo11 are much more hydrophobic than *flo11* mutant cells, indicating that some of these properties could be a consequence of the exclusion of water from the space between the cells and the surface to which they bind. So far, there is no evidence that *S. cerevisiae* secretes an extracellular material, either when growing as a film or when adhering to plastic.

C. albicans shows many of the features of bacterial biofilms in the assays that have been described earlier. *C. albicans* cells within the biofilm become highly resistant to antifungal agents and surround themselves with extracellular material[7,56,65,66]. Within this material, the cells develop into a morphologically diverse network, which consists of single-celled yeast, hyphae and pseudohyphae. Could the ECM consist of adhesin proteins that have been shed from the yeast cells? It has been shown that cells in *C. albicans* biofilms express several *ALS* genes[67]. So far, no specific proteins have been identified in the extracellular material, although it seems to consist of both proteins and carbohydrates[58,65]. The chemical composition of the extracellular material is important because it will determine whether the matrix is a mixture of cell-wall proteins that have been sloughed off from the surface or whether it consists of a specific protein that is associated with this new, communal form of growth. Moreover, it will show whether the repertoire of adhesins provides the diversity that enables *C. albicans* to form biofilms on such a variety of substrates.

Conclusions

Fungi, like many other microorganisms, can vary their cell-surface antigens. The diversity of adhesins that is produced by frequent recombination events has important implications for treating fungal infections, as it makes them poor targets for antifungal agents. Enzymes that are required for the localization of all these adhesins, such as those involved in the modification of the GPI anchor, might be better candidates for drug targeting.

The variation that exists among fungal adhesins seems to be caused by differential genetic and epigenetic regulation and frequent recombination events, both between and within adhesin genes. However, the mechanisms by which these events occur and their frequencies *in vitro* and in animal hosts are still not known. The tools that are needed to answer these questions are now available, and improving our knowledge of these processes should benefit both basic molecular biology and our understanding of fungal pathogenic strategies.

1. Hazen, K. C. New and emerging yeast pathogens. *Clin. Microbiol. Rev.* **8**, 462–478 (1995).
2. Pfaller, M. A. *et al.* International surveillance of bloodstream infections due to *Candida* species: frequency of occurrence and *in vitro* susceptibilities to fluconazole, ravuconazole, and voriconazole of isolates collected from 1997 through 1999 in the SENTRY antimicrobial surveillance program. *J. Clin. Microbiol.* **39**, 3254–3259 (2001).
3. Cassone, M. *et al.* Outbreak of *Saccharomyces cerevisiae* subtype *boulardii fungemia* in patients neighboring those treated with a probiotic preparation of the organism. *J. Clin. Microbiol.* **41**, 5340–5343 (2003).
4. Douglas, L. J. *Candida* biofilms and their role in infection. *Trends Microbiol.* **11**, 30–36 (2003).
This paper and references 56 and 66 review the evidence for the formation of biofilms by *Candida* spp.
5. Richards, M. J., Edwards, J. R., Culver, D. H. & Gaynes, R. P. Nosocomial infections in medical intensive care units in the United States. *Crit. Care Med.* **27**, 887–892 (1999).
6. Nguyen, M. H. *et al.* Therapeutic approaches in patients with candidemia. *Arch. Intern. Med.* **155**, 2429–2435 (1995).
7. Baillie, G. S. & Douglas, L. J. *Candida* biofilms and their susceptibility to antifungal agents. *Methods Enzymol.* **310**, 644–656 (1999).
8. Herre, J., Gordon, S. & Brown, G. D. Dectin-1 and its role in the recognition of β-glucans by macrophages. *Mol. Immunol.* **40**, 869–876 (2004).
9. Taylor, P. R. *et al.* The role of SIGNR1 and the β-glucan receptor (Dectin-1) in the nonopsonic recognition of yeast by specific macrophages. *J. Immunol.* **172**, 1157–1162 (2004).
10. Fraser, I. P., Koziel, H. & Ezekowitz, R. A. The serum mannose-binding protein and the macrophage mannose receptor are pattern recognition molecules that link innate and adaptive immunity. *Semin. Immunol.* **10**, 363–372 (1999).
11. Guo, B., Styles, C. A., Feng, Q. & Fink, G. A *Saccharomyces* gene family involved in invasive growth, cell–cell adhesion, and mating. *Proc. Natl Acad. Sci. USA* **97**, 12158–12163 (2000).
12. Hoyer, L. L. The *ALS* gene family of *Candida albicans*. *Trends Microbiol.* **9**, 176–180 (2001).
13. Teunissen, A. W. R. H. & Steensma, H. Y. Review: the dominant flocculation genes of *Saccharomyces cerevisiae* constitute a new subtelomeric gene family. *Yeast* **11**, 1001–1013 (1995).
14. Lo, W. S. & Dranginis, A. M. The cell surface flocculin Flo11 is required for pseudohyphae formation and invasion by *Saccharomyces cerevisiae*. *Mol. Biol. Cell* **9**, 161–171 (1998).
15. Vyas, V. K., Kuchin, S., Berkey, C. D. & Carlson, M. Snf1 kinases with different β-subunit isoforms play distinct roles in regulating haploid invasive growth. *Mol. Cell. Biol.* **23**, 1341–1348 (2003).
16. Reynolds, T. B. & Fink, G. R. Bakers' yeast, a model for fungal biofilm formation. *Science* **291**, 878–881 (2001).
17. Gaur, N. K. & Klotz, S. A. Expression, cloning, and characterization of a *Candida albicans* gene, *ALA1*, that confers adherence properties upon *Saccharomyces cerevisiae* for extracellular matrix proteins. *Infect. Immun.* **65**, 5289–5294 (1997).
18. Cormack, B. P., Ghori, N. & Falkow, S. An adhesin of the yeast pathogen *Candida glabrata* mediating adherence to human epithelial cells. *Science* **285**, 578–582 (1999).
A clever approach to cloning a key adhesin gene from *Candida glabrata*. The paper demonstrates the advantage of assessing the functions of *Candida* adhesin genes in *Saccharomyces cerevisiae*.
19. Li, F. & Palecek, S. P. *EAP1*, a *Candida albicans* gene involved in binding human epithelial cells. *Eukaryot. Cell* **2**, 1266–1273 (2003).
20. Hazen, K. C., Brawner, D. L., Riesselman, M. H., Jutila, M. A. & Cutler, J. E. Differential adherence of hydrophobic and hydrophilic *Candida albicans* yeast cells to mouse tissues. *Infect. Immun.* **59**, 907–912 (1991).
21. Green, C. B. *et al.* RT–PCR detection of *Candida albicans ALS* gene expression in the reconstituted human epithelium (RHE) model of oral candidiasis and in model biofilms. *Microbiol.* **150**, 267–275 (2004).
22. Baillie, G. S. & Douglas, L. J. Role of dimorphism in the development of *Candida albicans* biofilms. *J. Med. Microbiol.* **48**, 671–679 (1999).
23. De Las Penas, A. *et al.* Virulence-related surface glycoproteins in the yeast pathogen *Candida glabrata* are encoded in subtelomeric clusters and subject to *RAP1*- and *SIR*-dependent transcriptional silencing. *Genes Dev.* **17**, 2245–2258 (2003).
24. Hoyer, L. L. *et al.* Characterization of agglutinin-like sequence genes from non-*albicans Candida* and phylogenetic analysis of the *ALS* family. *Genetics* **157**, 1555–1567 (2001).
25. Zhang, N. *et al.* Sixty alleles of the *ALS7* open reading frame in *Candida albicans: ALS7* is a hypermutable contingency locus. *J. Genome Res.* **13**, 2005–2017 (2003).
26. Zhao, X., Pujol, C., Soll, D. R. & Hoyer, L. L. Allelic variation in the contiguous loci encoding *Candida albicans ALS5*, *ALS1* and *ALS9*. *Microbiology* **149**, 2947–2960 (2003).
Shows that three of the *Candida albicans* adhesins are linked on chromosome 6 and that the two alleles for each gene have different numbers of repeats.
27. Halme, A., Bumgarner, S., Styles, C. A. & Fink, G. R. Genetic and epigenetic regulation of the *FLO* gene family generates cell-surface variation in yeast. *Cell* **116**, 405–415 (2004).
28. Hoyer, L. L., Payne, T. L. & Hecht, J. E. Identification of *Candida albicans ALS2* and *ALS4* and localization of Als proteins to the fungal cell surface. *J. Bacteriol.* **180**, 5334–5343 (1998).
29. Kobayashi, O., Hayashi, N., Kuroki, R. & Sone, H. Region of Flo1 proteins responsible for sugar recognition. *J. Bacteriol.* **180**, 6503–6510 (1998).
30. Bony, M., Thines-Sempoux, D., Barre, P. & Blondin, B. Localization and cell surface anchoring of the *Saccharomyces cerevisiae* flocculation protein Flo1p. *J. Bacteriol.* **179**, 4929–4936 (1997).
31. De Groot, P. W. J., Hellingwerf, K. J. & Klis, F. M. Genome-wide identification of fungal GPI proteins. *Yeast* **20**, 781–796 (2003).
32. Frieman, M. B. & Cormack, B. P. The Ω-site sequence of glycosylphosphatidyl-inositol-anchored proteins in *Saccharomyces cerevisiae* can determine distribution between the membrane and the cell wall. *Mol. Microbiol.* **50**, 883–896 (2003).
33. Udenfriend, S. & Kodukula, K. How glycosyl-phosphatidylinositol-anchored membrane proteins are made. *Annu. Rev. Biochem.* **64**, 563–591 (1995).
34. Lu, C. F. *et al.* Glycosyl phosphatidylinositol-dependent cross-linking of α-agglutinin and β-1,6-glucan in the *Saccharomyces cerevisiae* cell wall. *J. Cell. Biol.* **128**, 333–340 (1995).
35. Klis, F. M., Mol, P., Hellingwerf, K. & Brul, S. Dynamics of cell wall structure in *Saccharomyces cerevisiae*. *FEMS Microbiol. Rev.* **26**, 239–256 (2002).
Provides a superb guide to the molecular organization of the yeast cell wall.
36. Frieman, M. B., McCaffery, J. M. & Cormack, B. P. Modular domain structure in the *Candida glabrata* adhesin Epa1p, a β-1,6 glucan-cross-linked cell wall protein. *Mol. Microbiol.* **46**, 479–492 (2002).
37. Liu, H. P., Styles, C. A. & Fink, G. R. *Saccharomyces cerevisiae* S288C has a mutation in *FLO8*, a gene required for filamentous growth. *Genetics* **144**, 967–978 (1996).
38. Sundstrom, P. Adhesins in *Candida albicans*. *Curr. Opin. Microbiol.* **2**, 353–357 (1999).
39. Gale, C. A. *et al.* Linkage of adhesion, filamentous growth, and virulence in *Candida albicans* to a single gene, *INT1*. *Science* **279**, 1355–1358 (1998).
40. Staab, J. F. & Sundstrom, P. Genetic organization and sequence analysis of the hypha-specific cell wall protein gene *HWP1* of *Candida albicans*. *Yeast* **14**, 681–686 (1998).
41. Staab, J. F., Bradway, S. D., Fidel, P. L. & Sundstrom, P. Adhesive and mammalian transglutaminase substrate properties of *Candida albicans* Hwp1. *Science* **283**, 1535–1538 (1999).
Describes a unique mechanism by which a *Candida* cell-wall protein might become linked to the surface of human cells.
42. Sohn, K., Urban, C., Brunner, H. & Rupp, S. *EFG1* is a major regulator of cell wall dynamics in *Candida albicans* as revealed by DNA microarrays. *Mol. Microbiol.* **47**, 89–102 (2003).
43. Fu, Y. *et al.* Expression of the *Candida albicans* gene *ALS1* in *Saccharomyces cerevisiae* induces adherence to endothelial and epithelial cells. *Infect. Immun.* **66**, 1783–1786 (1998).
44. Fu, Y. *et al. Candida albicans* Als1p: an adhesin that is a downstream effector of the *EFG1* filamentation pathway. *Mol. Microbiol.* **44**, 61–72 (2002).
45. Sheppard, D. C. *et al.* Functional and structural diversity in the Als protein family of *Candida albicans*. *J. Biol. Chem.* (in the press).
46. Pâques, F., Richard, G.-F. & Haber, J. E. Expansions and contractions in 36-bp minisatellites by gene conversion in yeast. *Genetics* **158**, 155–166 (2001).
An excellent analysis of specific mechanisms by which the number of repeats can be amplified or reduced.
47. Bishop, A. J. R., Louis, E. J. & Borts, R. H. Minisatellite variants generated in yeast meiosis involve DNA removal during gene conversion. *Genetics* **156**, 7–20 (2000).
48. Sato, M., Watari, J. & Shinotsuka, K. Genetic instability in flocculation of bottom-fermenting yeast. *J. Am. Soc. Brew. Chem.* **59**, 130–134 (2001).
49. Jibiki, M., Ishibiki, T., Yuuki, T. & Kagami, N. Application of polymerase chain reaction to determine the flocculation properties of brewer's lager yeast. *J. Am. Soc. Brew. Chem.* **59**, 107–110 (2001).
50. Hoyer, L. L. & Hecht, J. E. The *ALS5* gene of *Candida albicans* and analysis of the Als5p N-terminal domain. *Yeast* **18**, 49–60 (2001).
51. Harrison, P. *et al.* A small reservoir of disabled ORFs in the yeast genome and its implications for the dynamics of proteome evolution. *J. Mol. Biol.* **316**, 409–419 (2002).
52. Rupp, S., Summers, E., Lo, H. J., Madhani, H. & Fink, G. MAP kinase and cAMP filamentation signaling pathways converge on the unusually large promoter of the yeast *FLO11* gene. *EMBO J.* **18**, 1257–1269 (1999).
53. Serio, T. R. & Lindquist, S. L. [PSI+]: an epigenetic modulator of translation termination efficiency. *Annu. Rev. Cell. Dev. Biol.* **15**, 661–703 (1999).
54. Eggimann, P. & Pittet, D. Overview of catheter-related infections with special emphasis on prevention based on educational programs. *Clin. Microbiol. Infect.* **8**, 295–309 (2002).
55. Hawser, S. P. & Douglas, L. J. Biofilm formation by *Candida* species on the surface of catheter materials *in vitro*. *Infect. Immun.* **62**, 915–921 (1994).
56. Chandra, J. *et al.* Biofilm formation by the fungal pathogen *Candida albicans*: development, architecture, and drug resistance. *J. Bacteriol.* **183**, 5385–5394 (2001).
57. Sundstrom, P. Adhesion in *Candida* spp. *Cell. Microbiol.* **4**, 461–469 (2002).
58. Hawser, S. P., Baillie, G. S. & Douglas, L. J. Production of extracellular matrix by *Candida albicans* biofilms. *J. Med. Microbiol.* **47**, 253–256 (1998).
59. O'Toole, G., Kaplan, H. B. & Kolter, R. Biofilm formation as microbial development. *Annu. Rev. Microbiol.* **54**, 49–79 (2000).
60. Davey, M. E. & O'Toole, G. A. Microbial biofilms: from ecology to molecular genetics. *Microbiol. Mol. Biol. Rev.* **64**, 847–867 (2000).
61. Darby, C., Hsu, J. W., Ghori, N. S. & Falkow, S. *Caenorhabditis elegans*: plague bacteria biofilm blocks food intake. *Nature* **417**, 243–244 (2002).
62. Mah, T. F. & O'Toole, G. A. Mechanisms of biofilm resistance to antimicrobial agents. *Trends Microbiol.* **9**, 34–39 (2001).
63. Wingender, J., Strathmans, M., Rode, A., Leis, A. & Flemming, H. C. Isolation and biochemical characterization of extracellular polymeric substances from *Pseudomonas aeruginosa*. *Methods Enzymol.* **336**, 302–314 (2001).
64. Martinez, A., Torello, S. & Kolter, R. Sliding motility in mycobacterium. *J. Bacteriol.* **181**, 7331–7338 (1999).
65. Baillie, G. S. & Douglas, L. J. Matrix polymers of *Candida* biofilms and their possible role in biofilm resistance to antifungal agents. *J. Antimicrob. Chemo.* **46**, 397–403 (2000).
66. Kumamoto, C. A. *Candida* biofilms. *Curr. Opin. Microbiol.* **5**, 608–611 (2002).
67. Garcia-Sanchez, S. *et al. Candida albicans* biofilms: a developmental state associated with specific and stable gene expression patterns. *Eukaryot. Cell* **3**, 536–545 (2004).
68. Kellis, M., Birren, B. W. & Lander, E. S. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* **428**, 617–624 (2004).
69. Bailey, T. L. & Elkan, C. Unsupervised learning of multiple motifs in biopolymers using expectation maximization. *Mach. Learn.* **21**, 51–80 (1995).
70. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
71. Rice, P., Longden, I. & Bleasby, A. EMBOSS: the European molecular biology open software suite. *Trends Genet.* **16**, 276–277 (2000).
72. Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).

Acknowledgements

The authors thank F. Lewitter for her assistance with sequence analysis and S. Bumgarner for valuable discussions and proof-reading of the manuscript. K.J.V. is a D. Collen Fellow of the Belgian American Educational Foundation and a beneficiary of a long-term travel grant from the Fund for Scientific Research — Flanders (F.W.O. Vlaanderen). G.R.F. is an American Cancer Society Professor of Genetics. Research in G.R.F.'s laboratory was supported by a National Institutes of Health grant.

Competing interests statement

The authors declare that they have no competing financial interests.

Online links

DATABASES

The following terms in this article are linked online to:

SwissProt: http://ca.expasy.org/sprot/

ALS1 | *FLO1* | *FLO5* | *FLO9* | Sir3 | Sup35

FURTHER INFORMATION

Sequencing of *Candida albicans* at the Stanford Genome Technology Center:

http://www-sequence.stanford.edu/group/candida

***Saccharomyces* genome database:**

http://www.yeastgenome.org/

Access to this interactive links box is free online.