Evolution of pathogenicity and sexual reproduction in eight Candida genomes

Geraldine Butler1, Matthew D. Rasmussen2, Michael F. Lin2,3, Manuel A. S. Santos4, Sharadha Sathkikumar5, Carol A. Munro6, Esther Rheinbay1,6, Manfred Grabherr7, Anja Forche8, Jennifer L. Reedy9, Ino Agrafioti10, Martha B. Arnaud10, Steven Bates11, Alistair J. P. Brown12, Sascha Brunke12, Maria C. Costanzo10, David A. Fitzpatrick11, Piet W. J. de Groot13, David Harris14, Lois L. Hoyer15, Bernhard Hube12, Frans M. Klis11, Chinnappa Kodira16, Nicola Lennard14, Mary E. Logue1, Ronny Martin12, Aaron M. Neiman16, Elissavet Nikolau5, Michael A. Quail14, Janet Quinn17, Maria C. Santos4, Florian F. Schmitzberger10, Gavin Sherlock10, Prachi Shah10, Kevin A. T. Silverstein18, Marek S. Skrzypek10, David Soll19, Rodney Stagg18, Ian Stansfield1, Michael P. H. Stumpf19, Peter E. Sudbery20, Thyagarajan Srikantha19, Qiandong Zeng3, Judith Berman17, Matthew Berriiman14, Joseph Heitman18, Neil A. R. Gow5, Michael C. Lorenz21, Bruce W. Birren3, Manolis Kellis3,2,3,5 & Christina A. Cuomo3,5

Candida species are the most common cause of opportunistic fungal infection worldwide. Here we report the genome sequences of six Candida species and compare these and related pathogens and non-pathogens. There are significant expansions of cell wall, secreted and transporter gene families in pathogenic species, suggesting adaptations associated with virulence. Large genomic tracts are homozygous in three diploid species, possibly resulting from recent recombination events. Surprisingly, key components of the mating and meiosis pathways are missing from several species. These include major differences at the mating-type loci (MTL); Lodderomyces elongisporus lacks MTL, and components of the a1/a2 cell identity determinant were lost in other species, raising questions about how mating and cell types are controlled. Analysis of the CUG leucine-to-serine genetic-code change reveals that 99% of ancestral CUG codons were erased and new ones arose elsewhere. Lastly, we revise the Candida albicans gene catalogue, identifying many new genes.

Four Candida species, C. albicans, C. glabrata, C. tropicalis and C. parapsilosis, together account for 95% of identifiable Candida infections1. Although C. albicans is still the most common causative agent, its incidence is declining and the frequency of other species is increasing. Of these, C. parapsilosis is a particular problem in neonates, transplant recipients and patients receiving parenteral nutrition; C. tropicalis is more commonly associated with neutropenia and malignancy. Other Candida species, including C. krusei, C. lusitaniae and C. guilliermondii, account for <5% of invasive candidiasis. Almost all Candida species, with the exceptions of C. glabrata and C. krusei, belong in a single Candida clade (Fig. 1) characterized by the unique translation of CUG codons as serine rather than leucine2. Within this, haploid and diploid species occupy two separate sub-clades (Fig. 1).

To determine the genetic features underlying their diversity of biology and pathogenesis, we sequenced six genomes from the Candida clade (Fig. 1). These include a second sequenced isolate of C. albicans (WO-1) characterized for white–opaque switching, a phenotypic change that correlates with host specificity and mating3,4. We also sequenced the major pathogens C. tropicalis and C. parapsilosis; L. elongisporus, a close relative of C. parapsilosis recently identified as a cause of bloodstream infection2; and two haploid emerging pathogens, C. guilliermondii and C. lusitaniae. We compared these with the previously sequenced C. albicans strain (SC5314)6,7, Debaromyces Hansenii8, a marine yeast rarely associated with disease, and nine species from the related Saccharomyces clade (Fig. 1). These species span a wide evolutionary range and show large phenotypic differences in pathogenicity and mating, allowing us to study the genomic basis for these traits.

Genome sequence and comparative annotation

We found enormous variation in genome size and composition between the Candida genomes sequenced (Table 1). Each genome assembly displayed high continuity, ranging from nine to 27 scaffolds...
(Supplementary Table 1). Scaffold number and size largely match pulsed-field gel electrophoresis estimates for all genomes, and telomeric repeat arrays are linked to the ends of nearly all large scaffolds (Supplementary Information, section S2). Genome size ranges from 10.6 to 15.5 megabases (Mb), a difference of nearly 50%, with haploid species having smaller genomes. GC content ranges from 33% to 45% (Table 1). Transposable elements and other repetitive sequences vary in number and type between assemblies (Supplementary Information, section S6). Regions similar to the major repeat sequence (MRS) and meric repeat arrays are linked to the ends of nearly all large scaffolds (Supplementary Information, section S6). We also corrected existing annotations in C. albicans, revealing 222 dubious genes, and also identified 190 probable frame shifts and 36 nonsense sequencing errors in otherwise well-conserved genes (Supplementary Information, section S4). In each case, manual curation confirmed ~80% of these predictions.

### Polymorphism in diploid genomes

To gain insights into the recent history of C. albicans, we compared the two diploid strains, SC5314 and WO-1, which belong to different population subgroups59. Variation in the karyotype of these strains is primarily due to translocations at MRS sequences (ref. 12 and Supplementary Fig. 1). The two assemblies are largely co-linear with 12 inversions of 5–94 kilobases between them, except that in WO-1 some non-homologous chromosomes have recombined at the MRS (Supplementary Information, section S8). We found similar rates of single nucleotide polymorphisms (SNPs) within each strain (one SNP per 330–390 bases), and twice this rate between them, suggesting relatively recent divergence. Polymorphism rates in the other diploids range from one SNP per 222 bases in L. elongisporus and one SNP per 576 bases in C. tropicalis to a particularly low one SNP per 15,533 bases in C. parapsilosis, more than 70-fold lower than in the closely related L. elongisporus.

Notable regions of extended homozygosity are found in three of the four diploid genomes, which may reflect break-induced replication or recent passage through a parasexual or sexual cycle. Candida albicans, C. tropicalis and L. elongisporus each shows large chromosomal regions devoid of SNPs, extending up to ~1.2 Mb (Fig. 2 and Supplementary Figs 6–8). In contrast, the few SNPs in C. parapsilosis are randomly distributed across the genome (Supplementary Fig. 9a). A total of 4.3 Mb (30%) of the WO-1 assembly is homozygous for SNPs, approximately twice that found in SC5314 (Supplementary Information, section 7, and refs 7, 13, 14). There is at least one homozygous region per chromosome, none of which spans the predicted centromeres and only one of which starts at a MRS (Fig. 2). Whereas nearly all homogeneous regions are present at diploid levels and are therefore homozygous, WO-1 has lost one copy of a >300-kilobase region on chromosome 3 comprising nearly 200 genes (Supplementary Fig. 10). The pressure to maintain this region as homozygous in both strains is apparently high, as it is diploid but not homozygous in SC5314.

### Usage and evolution of CUG codons

All Candida clade species translate CUG codons as serine instead of leucine52. This genetic-code change altered the decoding rules of CUN codons in the Candida clade: whereas Saccharomyces cerevisiae uses two transfer RNAs, each of which translates two codons, Candida species use a dedicated tRNA CAG15. This genetic-code change altered the decoding rules of CUG codons for C. guilliermondii and D. hansenii (Supplementary Fig. 5). Although syntenic blocks have been shuffled by local inversions and rearrangements, these have been primarily intrachromosomal as chromosome boundaries have been largely preserved across the diploid genomes (Supplementary Fig. 5).

Given the high conservation of protein-coding genes across the Candida clade, we used multiple alignments of the related genomes to revise the annotation of C. albicans. We identified 91 new or updated genes, of which 80% are specific to the Candida clade (Supplementary Information, section S4). We also corrected existing annotations in C. albicans, revealing 222 dubious genes, and also identified 190 probable frame shifts and 36 nonsense sequencing errors in otherwise well-conserved genes (Supplementary Information, section S4). In each case, manual curation confirmed ~80% of these predictions.

### Table 1 | Candida genome features

| Species* | Genome size (Mb) | GC content (%) | No. of genes | Ave. gene size (bp) | Intergenic ave. (bp) | Ploidy | Pathogen† |
|----------|-----------------|----------------|-------------|---------------------|---------------------|--------|-----------|
| C. albicans | 14.4 | 33.5 | 6,159 | 1,444 | 921 | diploid | + |
| C. albicans | 14.3 | 33.5 | 6,107 | 1,468 | 858 | diploid | + |
| C. tropicalis | 14.5 | 33.1 | 6,258 | 1,454 | 902 | diploid | + |
| C. parapsilosis | 13.1 | 38.7 | 5,733 | 1,533 | 752 | diploid | + |
| L. elongisporus | 15.4 | 37.0 | 5,802 | 1,530 | 1,174 | diploid | + |
| C. guilliermondii | 10.6 | 43.8 | 5,920 | 1,402 | 426 | haploid | + |
| C. lusitaniae | 12.1 | 44.5 | 5,941 | 1,382 | 770 | haploid | + |
| D. hansenii | 12.2 | 36.3 | 6,318 | 1,382 | 550 | haploid | − |

* C. albicans SC5314 assembly 21 and gene set dated 28 January 2008 downloaded from the Candida Genome Database; D. hansenii assembly from GenBank®. The remaining assemblies are reported as part of this work, and are available in GenBank and at the Broad Institute Candida Database website.

† Relative level of pathogen strength: +, strong pathogen; +, moderate pathogen; −, rare pathogen.
except for those in genes with two CUA codons (P′sc 5, 9 including Pichia stipitis and excluding C. dubliniensis and Saccharomyces spp.)

We also examined the evolutionary fate of ancestral CUG codons and the origin of new CUG codons (Supplementary Table 12). CUG codons in C. albicans almost never (1%) align opposite CUG codons in S. cerevisiae. Instead, CUG serine codons in C. albicans align primarily to Saccharomyces codons for serine (20%) and other hydrophilic residues (49%). CUG leucine codons in S. cerevisiae align primarily to leucine codons in Candida (50%) and to other hydrophobic-residue codons (30%). This suggests a complete functional replacement of CUG codons in Candida.

Gene family evolution

To identify gene families likely to be associated with Candida pathogenicity and virulence, we used a phylogenomic approach across

(Fig. 3b). An additional pressure influencing codon usage may be the GC content, as usage of leucine codons in Candida species is correlated with the percentage GC composition (Supplementary Table 11).

We also examined the evolutionary fate of ancestral CUG codons and the origin of new CUG codons (Supplementary Table 12). CUG codons in C. albicans almost never (1%) align opposite CUG codons in S. cerevisiae. Instead, CUG serine codons in C. albicans align primarily to Saccharomyces codons for serine (20%) and other hydrophilic residues (49%). CUG leucine codons in S. cerevisiae align primarily to leucine codons in Candida (50%) and to other hydrophobic-residue codons (30%). This suggests a complete functional replacement of CUG codons in Candida.

Gene family evolution

To identify gene families likely to be associated with Candida pathogenicity and virulence, we used a phylogenomic approach across

(Fig. 3b). An additional pressure influencing codon usage may be the GC content, as usage of leucine codons in Candida species is correlated with the percentage GC composition (Supplementary Table 11).

We also examined the evolutionary fate of ancestral CUG codons and the origin of new CUG codons (Supplementary Table 12). CUG codons in C. albicans almost never (1%) align opposite CUG codons in S. cerevisiae. Instead, CUG serine codons in C. albicans align primarily to Saccharomyces codons for serine (20%) and other hydrophilic residues (49%). CUG leucine codons in S. cerevisiae align primarily to leucine codons in Candida (50%) and to other hydrophobic-residue codons (30%). This suggests a complete functional replacement of CUG codons in Candida.

Gene family evolution

To identify gene families likely to be associated with Candida pathogenicity and virulence, we used a phylogenomic approach across

(Fig. 3b). An additional pressure influencing codon usage may be the GC content, as usage of leucine codons in Candida species is correlated with the percentage GC composition (Supplementary Table 11).

We also examined the evolutionary fate of ancestral CUG codons and the origin of new CUG codons (Supplementary Table 12). CUG codons in C. albicans almost never (1%) align opposite CUG codons in S. cerevisiae. Instead, CUG serine codons in C. albicans align primarily to Saccharomyces codons for serine (20%) and other hydrophilic residues (49%). CUG leucine codons in S. cerevisiae align primarily to leucine codons in Candida (50%) and to other hydrophobic-residue codons (30%). This suggests a complete functional replacement of CUG codons in Candida.

Gene family evolution

To identify gene families likely to be associated with Candida pathogenicity and virulence, we used a phylogenomic approach across

(Fig. 3b). An additional pressure influencing codon usage may be the GC content, as usage of leucine codons in Candida species is correlated with the percentage GC composition (Supplementary Table 11).

We also examined the evolutionary fate of ancestral CUG codons and the origin of new CUG codons (Supplementary Table 12). CUG codons in C. albicans almost never (1%) align opposite CUG codons in S. cerevisiae. Instead, CUG serine codons in C. albicans align primarily to Saccharomyces codons for serine (20%) and other hydrophilic residues (49%). CUG leucine codons in S. cerevisiae align primarily to leucine codons in Candida (50%) and to other hydrophobic-residue codons (30%). This suggests a complete functional replacement of CUG codons in Candida.

Gene family evolution

To identify gene families likely to be associated with Candida pathogenicity and virulence, we used a phylogenomic approach across

(Fig. 3b). An additional pressure influencing codon usage may be the GC content, as usage of leucine codons in Candida species is correlated with the percentage GC composition (Supplementary Table 11).

We also examined the evolutionary fate of ancestral CUG codons and the origin of new CUG codons (Supplementary Table 12). CUG codons in C. albicans almost never (1%) align opposite CUG codons in S. cerevisiae. Instead, CUG serine codons in C. albicans align primarily to Saccharomyces codons for serine (20%) and other hydrophilic residues (49%). CUG leucine codons in S. cerevisiae align primarily to leucine codons in Candida (50%) and to other hydrophobic-residue codons (30%). This suggests a complete functional replacement of CUG codons in Candida.

Gene family evolution

To identify gene families likely to be associated with Candida pathogenicity and virulence, we used a phylogenomic approach across

(Fig. 3b). An additional pressure influencing codon usage may be the GC content, as usage of leucine codons in Candida species is correlated with the percentage GC composition (Supplementary Table 11).

We also examined the evolutionary fate of ancestral CUG codons and the origin of new CUG codons (Supplementary Table 12). CUG codons in C. albicans almost never (1%) align opposite CUG codons in S. cerevisiae. Instead, CUG serine codons in C. albicans align primarily to Saccharomyces codons for serine (20%) and other hydrophilic residues (49%). CUG leucine codons in S. cerevisiae align primarily to leucine codons in Candida (50%) and to other hydrophobic-residue codons (30%). This suggests a complete functional replacement of CUG codons in Candida.

Gene family evolution

To identify gene families likely to be associated with Candida pathogenicity and virulence, we used a phylogenomic approach across

(Fig. 3b). An additional pressure influencing codon usage may be the GC content, as usage of leucine codons in Candida species is correlated with the percentage GC composition (Supplementary Table 11).

We also examined the evolutionary fate of ancestral CUG codons and the origin of new CUG codons (Supplementary Table 12). CUG codons in C. albicans almost never (1%) align opposite CUG codons in S. cerevisiae. Instead, CUG serine codons in C. albicans align primarily to Saccharomyces codons for serine (20%) and other hydrophilic residues (49%). CUG leucine codons in S. cerevisiae align primarily to leucine codons in Candida (50%) and to other hydrophobic-residue codons (30%). This suggests a complete functional replacement of CUG codons in Candida.
We identified 64 families showing positive selection in the highly pathogenic Candida species (Supplementary Table 32). These are highly enriched for cell wall, hyphal, pseudohyphal, filamentous growth and biofilm functions (Supplementary Information, section S13). Six of the families have been previously associated with pathogenesis, including that of ERG3, a C-5 sterol desaturase essential for ergosterol biosynthesis, for which mutations can cause drug resistance.

Structure of the MTL locus

Pathogenic fungi may have limited their sexual cycles to maximize their virulence, and the sequenced Candida species show tremendous diversity in their apparent abilities to mate. Among the four diploids, C. albicans has a parasexual cycle (mating of diploid cells followed by mitosis and chromosome loss instead of meiosis), L. longisporus has been described as sexual and homothallic (self-mating), whereas C. tropicalis and C. parapsilosis have never been observed to mate. Among the three haploids, C. guilliermondii and C. lusitaniae are heterothallic (cross-mating only) and have a complete sexual cycle, whereas D. hansenii is haploid and homothallic (Supplementary Information, section 14c).

To understand the genomic basis for this diversity, we studied the Candida MTL locus, which determines mating type, similar to the MAT locus in S. cerevisiae. In both C. albicans and S. cerevisiae, the mating locus has two idiomorphs, a and a, encoding the regulators α1 and α2, respectively. Candida albicans MTLa also encodes a2, and both idiomorphs in this species contain alleles of three additional genes without known roles in mating: PAP, OBP and PIK. The MTLa and MTLa genes, alone or in combination, specify one of the three possible cell-type programs (a haploid, a haploid, a α diploid). In C. albicans, the alpha-domain protein α1 activates α-specific mating genes, the high-mobility group (HMG) factor a2 activates a-specific mating genes, and the a1/a2 homeodomain heterodimer represses mating genes in a/a cells.

Despite extended conservation of the genomic context flanking the mating-type locus, there is great variability in the MTLa gene content (Fig. 5). MTLa1 has become a pseudogene in C. parapsilosis, and is probably a recent loss because target genes retain predicted a1/a2 binding sites (Supplementary Information, section S14). MTLa2 is missing in both C. guilliermondii and C. lusitaniae (J. L. Reddy, A. Floyd and J. Heitman, submitted). A fused mating-type locus containing both α and α genes is found in D. hansenii and Pichia stipitis.

Most surprisingly, all four mating-type genes are missing in L. longisporus. It contains a site syntenic to MTLa in other species, but this contains only 508 base pairs of apparently non-coding DNA.
a length insufficient to encode a1 or a2 even if they were extensively divergent in sequence. We confirmed this finding in seven other L. elongisporus isolates (not shown). The sexual state of L. elongisporus has been assumed to be homothallic because asc.is are generated from identical cells but the absence of an MTL pseudogene. There is no MTL locus in L. elongisporus; gene order around the PAPA, OBPA and MTL-1/2, respectively.

**Figure 5 | Organization of MTL loci in the Candida clade.** a, MTLa-specific genes are shown in grey, MTLz-specific genes are shown in black and other orthologues are shown in colour. Two idiomorphs from C. albicans and C. tropicalis are shown. Arrows indicate inversions relative to C. albicans. Crosses show gene losses and the MTLa pseudogene. There are frequent losses of MTLa genes is shown. For C. guilliermondii and C. lusitaniae, the genome project sequenced one idiomorph; the second was obtained by J. Reedy. b, Placement of gene losses on the phylogenetic tree. HOM, homothallic; HET, heterothallic.

**Mating and meiosis**

To gain further insight into their diversity in sexual behaviour, we examined whether 227 genes required for meiosis in S. cerevisiae and other fungi have orthologues in the Candida species (Supplementary Information, section S14). A previous report that some components of meiosis, such as the major regulator IME1, are missing from C. albicans led us to propose that their loss could be correlated with lack of meiosis. Surprisingly, however, we find that these genes are missing in all Candida species, suggesting that sexual Candida species undergo meiosis without them. Conversely, even seemingly non-mating species showed highly conserved pheromone response pathways, suggesting that pheromone signalling plays an alternative role such as regulation of biofilm formation. These findings suggest considerable plasticity and innovation of meiotic pathways in Candida.

Moreover, we find that sexual Candida species have undergone a recent dramatic change in the pathways involved in meiotic recombination, with loss of the Dmc1-dependent pathway in the heterothallic species C. lusitaniae and C. guilliermondii (Supplementary Information, section 14c). We also found that mechanisms of chromosome pairing and crossover formation have changed recently in these two species, because they (and to a lesser extent D. hansenii) have lost several components of the synaptonemal complex.

The genome sequences reported here provide a resource that will allow current knowledge of C. albicans biology, the product of decades of research, to be applied with maximum effect to the other pathogenic species in the Candida clade. They also allow many of the unusual features of C. albicans—such as cell wall gene family amplifications, and its apparent ability to undergo mating and a parasexual cycle without meiosis—to be understood in an evolutionary context that shows that the genes involved in virulence and mating have highly dynamic rates of turnover and loss.

**METHODS SUMMARY**

The methods for this paper are described in Supplementary Information. Here we outline the resources generated by this project.
Assemblies, gene sets, and single nucleotide polymorphisms are available in GenBank and at the Broad Institute Candida Database website (http://www.broad.mit.edu/annotation/genome/candida_group/MultiHome.html). The Broad Institute website provides search, visualization, BLAST and download of assemblies and gene sets. Gene families can be accessed by searching either for individual genes or with family identifiers (CF#####). The revised annotation of C. albicans (SC5314) is available at the Candida Genome Database (www.candidagenome.org).

Received 22 February; accepted 15 April 2009. Published online 24 May 2009.

1. Pfaffer, M.A. & Diekema, D.J. Epidemiology of invasive candidiasis: a persistent public health problem. Clin. Microbiol. Rev. 20, 133–163 (2007).
2. Santos, M.A. & Tuite, M.F. The CUG codon is decoded in vivo as serine and not leucine in Candida albicans. Nucleic Acids Res. 23, 1481–1486 (1995).
3. Lockhart, S.R. et al. In Candida albicans, white-opaque switchers are homoyogous for mating type. Genetics 162, 737–745 (2002).
4. Slutskey, B., Buffa, J. & Soll, D.R. High-frequency switching of colony morphology in Candida albicans. Science 230, 666–669 (1985).
5. Lockhart, S.R., Messer, S.A., Pfaffer, M.A. & Diekema, D.J. Lodderomyces elongisporus masquerading as Candida parapsilosis as a cause of bloodstream infections. J. Clin. Microbiol. 46, 374–376 (2008).
6. Braun, B.R. et al. A human-curated annotation of the Candida albicans genome. PLoS Genet. 1, e1 (2005).
7. Jones, T. et al. The diploid genome sequence of Candida albicans. Proc. Natl Acad. Sci. USA 101, 7329–7334 (2004).
8. van het Hoog, M., et al. Assembly of the Candida albicans genome into sixteen supercontigs aligned on the eight chromosomes. Genome Biol. 8, RS2 (2007).
9. Djouz, B. et al. Genome evolution in yeasts. Nature 430, 35–44 (2004).
10. Zhang, J., Hollis, R.J. & Pfaffer, M.A. Variations in DNA subtype and antifungal susceptibility among clinical isolates of Candida albicans. J. Clin. Microbiol. 37, 63–67 (1997).
11. Tavanti, A. et al. Population structure and properties of Candida albicans, as determined by multilocus sequence typing. J. Clin. Microbiol. 43, 5601–5613 (2005).
12. Chu, W.S., Magee, B.B. & Magee, P.T. Construction of an SfiI macrorestriction map of the Candida albicans genome. J. Bacteriol. 175, 6637–6651 (1993).
13. Forche, A., Magee, P.T., Magee, B.B. & May, G. Genome-wide single-nucleotide polymorphism map for Candida albicans. Eukaryot. Cell 3, 705–714 (2004).
14. Legrand, M. et al. Haplotype mapping of a diploid non-meiotic organism using existing and induced aneuplodids. PLoS Genet. 4, e1 (2008).
15. Massey, S.E. et al. Comparative evolutionary genomics unveils the molecular mechanism of reassignment of the CTG codon in Candida spp. Genome Res. 13, 544–557 (2003).
16. Bates, S. et al. Candida albicans HML, a secreted protein required for cell wall structure and virulence. Infect. Immun. 75, 2922–2928 (2007).
17. Hoyer, L.L., Green, C.B., Oh, S.H. & Zhao, X. Discovering the secrets of the Candida albicans agglutinin-like sequence (ALS) gene family--a sticky pursuit. Med. Mycol. 46, 1–15 (2008).
18. Yeater, K.M. et al. Temporal analysis of Candida albicans gene expression during biofilm development. Microbiology 153, 2373–2385 (2007).
19. Par, Q. et al. Als6 is a Candida albicans invasin that binds to cadherins and induces endocytosis by host cells. PLoS Biol. 5, e64 (2007).
20. Almeida, R.S. et al. The hyphal-associated adhesin and invasin Als3 of Candida albicans mediates iron acquisition from host ferritin. PLoS Pathog. 4, e1000217 (2008).
21. Verstrepen, K.J., Jansen, A., Lewitter, F. & Fink, G.R. Intraspecific tandem repeats generate functional variability. Nature Genet. 37, 986–990 (2005).
22. Kaur, R., Ma, B. & Cormack, B.P. A family of glycosylphosphatidylinositol-linked aspartyl proteases is required for virulence of Candida glabrata. Nat. Protoc. 2, 1064, 7628–7633 (2007).
23. Chau, A.S. et al. Inactivation of sterol Δ^54-desaturase attenuates virulence in Candida albicans. Antimicrob. Agents Chemother. 49, 3646–3651 (2005).
24. Nielsen, K. & Heitman, J. Sex and virulence of human pathogenic fungi. Adv. Genet. 57, 143–173 (2007).
25. Noble, S.M. & Johnson, A.D. Genetics of Candida albicans, a diploid human fungal pathogen. Annu. Rev. Genet. 41, 193–211 (2007).
26. van der Walt, J.P. Lodderomyces, a new genus of the Saccharomycetaceae. Antonie Van Leeuwenhoek 32, 1–5 (1966).
27. van der Walt, J.P., Taylor, M.B. & Liebenberg, N.V.D. W. Ploidy, ascus formation and recombination in Tokulaspera (Debaryomyces) hansenii. Antonie Van Leeuwenhoek 43, 205–218 (1977).
28. Hull, C.M., Raisner, R.M. & Johnson, A.D. Evidence for mating of the “asexual” yeast Candida albicans in a mammalian host. Science 289, 307–310 (2000).
29. Tsong, A.E., Miller, M.G., Raisner, R.M. & Johnson, A.D. Evolution of a combinatorial transcriptional circuit: a case study in yeasts. Cell 115, 389–399 (2003).
30. Tsong, A.E., Tuch, B.B., Li, H. & Johnson, A.D. Evolution of alternative transcriptional circuits with identical logic. Nature 443, 415–420 (2006).
31. Logue, M.E., Wong, S., Wolfe, K.H. & Butler, G. A genome sequence survey shows that the pathogenic yeast Candida parapsilosis has a defective MTLα1 allele at its mating type locus. Eukaryot. Cell 4, 1009–1017 (2005).
32. Fabre, E. et al. Comparative genomics in hemiascomycete yeasts: evolution of sex, silencing, and subtelomeres. Mol. Biol. Evol. 22, 856–873 (2005).
33. Jeffries, T.W. et al. Genome sequence of the lignoncellulose-bioconverting and xyllose-fermenting yeast Pichia stipitis. Nature Biotechnol. 25, 319–326 (2007).
34. Zordan, R.E., Miller, M.G., Gaglory, J.D., Tuch, B.B. & Johnson, A.D. Interlocking transcriptional feedback loops control white-opaque switching in Candida albicans. PLoS Biol. 5, e256 (2007).
35. Tuch, B.B., Gaglory, D.J., Hernaday, A.D., Li, H. & Johnson, A.D. The evolution of combinatorial gene regulation in fungi. PLoS Biol. 6, e38 (2008).
36. Tzung, K.W. et al. Genomic evidence for a complete sexual cycle in Candida albicans. Proc. Natl Acad. Sci. USA 98, 3249–3253 (2001).
37. Daniels, K.J., Srikanta, T., Lockhart, S.R., Pujol, C. & Soll, D.R. Opaque cells signal white cells to form biofilms in Candida albicans. EMBO J. 25, 2240–2252 (2006).
38. Argueso, J.L., Wanat, J., Gemici, Z. & Alan, E. Competing crossover pathways act during meiosis in Saccharomyces cerevisiae. Genetics 168, 1805–1816 (2004).
39. de los Santos, T. et al. The Mus81/Mms4 endonuclease acts independently of double-Holliday junction resolution to promote a distinct subset of crossovers during meiosis in budding yeast. Genetics 164, 81–94 (2003).
40. Scannell, D.R., Byrne, K.P., Gordon, J.L., Wong, S. & Wolfe, K.H. Multiple rounds of speciation associated with reciprocal gene loss in polypliod yeasts. Nature 440, 341–345 (2006).
41. Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the US National Human Genome Research Institute (NHGRI) for support under the Fungal Genome Initiative at the Broad Institute. We thank C. Kurtzman for providing the sequenced strains of L. elongisporus, C. guilliermondii and C. lusitaniae, M. Koehrsen for Broad Institute website support, D. Park for Informatics support and K. Wolfe and A. Regew for comments on the manuscript. We acknowledge the contributions of the Broad Institute Sequencing Platform and A. Barron, L. Clark, C. Corton, D. Ormond, D. Saunders, K. Seeger and R. Squires from the Wellcome Trust Sanger Institute for the C. parapsilosis sequencing and assembly. N.A.R.G., A.P.B., M.B. and co-workers were supported by the Wellcome Trust; M.G., S.S., Q.Z., C.K., B.W.B. and C.A.C. were supported by NHGRI and the National Institute of Allergy and Infectious Disease, US National Institutes of Health (NIH), Department of Health and Human Services; G.B. and co-workers were supported by Science Foundation Ireland; J.B., A.F., J.H., A.M.N. and co-workers were supported by the NIH; and M.K., M.D.R. and M.F.L. by the NIH, the US National Science Foundation and the Sloan Foundation.