

***Candida albicans* genome sequence: a platform for genomics in the absence of genetics**

Frank C Odds, Alistair JP Brown and Neil AR Gow

Address: Aberdeen Fungal Group, Institute of Medical Sciences, Aberdeen AB25 2ZD, UK.

Correspondence: Frank C Odds. E-mail: f.odds@abdn.ac.uk

Published: 11 June 2004

Genome **Biology** 2004, **5**:230

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2004/5/7/230>

© 2004 BioMed Central Ltd

Abstract

Publication of the complete diploid genome sequence of the yeast *Candida albicans* will accelerate research into the pathogenesis of *Candida* infections. Comparative genomic analysis highlights genes that may contribute to *C. albicans* survival and its fitness as a human commensal and pathogen.

For several years investigators studying the pathogenic yeast *Candida albicans* have had internet access to partial genomic sequence information, as the Stanford DNA Sequencing and Technology Center generously released data at several stages during their sequencing project [1]. The publication of the full diploid sequence of this fungus [2] represents a landmark in the history of *Candida* research and is the culmination of more than ten years of work. The drive for the *C. albicans* genome sequence originated at the University of Minnesota with the early interest of Stewart Scherer and Paul T. Magee in the molecular genetics of *C. albicans* [3]; the sequencing itself is the product of the Stanford Genome Technology Center, headed by Ron Davis. Davis and his team have succeeded in overcoming considerable computational hurdles to eliminate the problems of aligning sequence contigs for an organism with no known haploid state. Heterozygosity at numerous alleles originally resulted in single-copy genes being assigned to two distinct contigs. The now-completed diploid genome sequence, known as Assembly 19, is the result of novel alignment methods that make use of physical mapping data, paired plasmid clone sequences and archived GenBank sequences to assemble a set of supercontigs representing the diploid genome sequence.

C. albicans is unique among fungal pathogens in terms of the diversity of infections it can cause. The fungus is a

normal gut commensal in the majority of humans, but it is also able to infect mucosal surfaces, skin and nails when local antimicrobial defences are impaired, and it can spread via the bloodstream to infect deep tissues in severely immunocompromised individuals [4,5]. Comprehensive understanding of the pathogenesis of these many forms of *Candida* infection in terms of the molecular cross-talk between host and pathogen is an obvious prerequisite to progress in their diagnosis and treatment. The availability of a full diploid genome sequence provides an invaluable tool for researchers in the field.

The main facts and figures of the *C. albicans* genome sequence are as follows. Eight chromosomes (historically named 1-7 and R) constitute a haploid genome size of 14,851 kilobases (kb), containing 6,419 open reading frames (ORFs) longer than 100 codons, of which some 20% have no known counterpart in other available genome sequences. The codon CUG, which is translated abnormally by *C. albicans* as serine rather than leucine, is found at least once in approximately two-thirds of ORFs.

The *C. albicans* isolate used for the sequencing project turns out to have been an excellent representative choice. Strain SC5314 was used in the 1980s by scientists at the E.R. Squibb company (now Bristol-Myers Squibb) for their pioneering studies of *C. albicans* molecular biology. It was engineered by

Fonzi and Irwin [6] to provide the uridine autotrophic mutant that has been essential to most subsequent molecular genetic research into *C. albicans*. The strain is usually described merely as a 'clinical isolate', but it is worth setting on record that SC5314 was originally isolated from a patient with generalized *Candida* infection by Margarita Silva-Hutner at the Department of Dermatology, Columbia College of Physicians and Surgeons (New York, USA). The original isolate number was 1775 and the strain is identical with strain NYOH#4657 in the New York State Department of Health collection. (This information was provided by Joan Fung-Tomc at Bristol-Myers Squibb as a personal communication.) SC5314 belongs to the predominant clade of closely related *C. albicans* strains that represents almost 40% of all isolates worldwide, as determined by DNA fingerprinting [7] and multi-locus sequence typing (A. Tavanti, A.D. Davidson, N.A.R.G., M.C.J. Maiden and F.C.O., unpublished observations). It is highly susceptible to all clinically used antifungal agents (F.C.O., unpublished observations) and hence its genome sequence forms an excellent reference for comparison with drug-resistant isolates. Furthermore, this strain is highly virulent in animal models of *Candida* infection [8], and its genome sequence can therefore be presumed to encode most or all of the species' virulence factors.

Unlike most yeasts, *C. albicans* is a diploid organism with no known haploid phase, and for a long time it was considered to be asexual. But genome sequencing has profoundly altered our understanding of this organism. Early assemblies of the *C. albicans* genome sequence revealed a mating-type (*MAT*-like) locus [9] that led to the engineering of mating-competent strains [10,11]. Further work led to the identification of a natural mating-competent form that mates naturally at high frequency to give a tetraploid gamete [12]. So far, attempts to demonstrate meiosis, and thereby complete a sexual cycle, have been unsuccessful [2], although the *C. albicans* genome has revealed a nearly complete repertoire of genes homologous to those predicted to execute the essential stages of meiosis in the yeast *Saccharomyces cerevisiae* [13]. Nevertheless, a parasexual cycle has been completed following the description of *in vitro* conditions that promote concerted chromosome loss from tetraploids to generate diploid segregants [14], and this is likely to be a valuable experimental tool in the future.

The assembly of a complete diploid genome sequence for SC5314 has allowed a reliable estimate of the frequency of heterozygosities in *C. albicans* of 4.21 polymorphisms per kb, or 1 polymorphism per 237 bases [2]. These heterozygosities are distributed unevenly across the *C. albicans* genome, however, with the highest prevalence on chromosomes 5 and 6. Highly polymorphic loci include the mating type-like (*MTL*) locus and a region on chromosome 6 that encodes several genes in the agglutinin-like sequence (*ALS*) gene family, thought to be involved in adhesion to and interaction with host surfaces [15]. Nevertheless, over half of the

approximately 6,400 *C. albicans* genes contain allelic differences, and two-thirds of these polymorphisms are predicted to alter the protein sequence. Furthermore, considerable allelic variation in the *C. albicans* genome also results from tandem repeat sequences, with many trinucleotide tandem repeats located in coding regions of the genome [2]. This suggests that the frequency with which seemingly equivalent heterozygous mutants display phenotypic differences might be higher than expected. Indeed there are a number of reported cases of this (see, for example, [16]).

What can be gleaned from the genome sequences of a pathogen such as *C. albicans* (and from other related fungi)? *C. albicans* has rarely been isolated in nature away from an animal host and has probably co-evolved along with humans for millions of years. It is presumed, therefore, that the present-day *C. albicans* genome contains the information that enables this fungus to thrive in its human host in competition with the immune system and with other microflora. There are more than 1,000 *C. albicans* genes of unknown function that have no obvious ortholog in *S. cerevisiae* or the fission yeast *Schizosaccharomyces pombe*. These genes are of particular interest to those interested in fungus-host interactions, because many might play roles in the infection process. The genome of the closely related species *Candida dubliniensis* is now being sequenced. *C. dubliniensis* is the nearest known phylogenetic neighbour to *C. albicans* and infects humans but is less virulent in animal models [17]. Hence, comparisons of the two *Candida* genome sequences may provide important clues about *C. albicans* genes that contribute to its success as a human pathogen.

The genome sequence of the next most prevalent serious agent of systemic fungal disease, *Aspergillus fumigatus*, will also be released this year. This fungus primarily infects the lungs of immunocompromised patients [18], whereas the main focus of *C. albicans* and *C. dubliniensis* infections is the kidneys [5]. Also, *A. fumigatus* has evolved as a saprophyte, decomposing leaf litter, whereas *C. albicans* appears to have an obligate association with mammalian hosts. Hence, comparative analyses of the genome sequences of these fungi is likely to provide important insights into the evolution of niche-specific functions related to pathogenesis in humans. There are now more than forty fungal genome-sequencing projects underway, including representatives of almost all major groups pathogenic for humans [19]. The *C. albicans* genome sequence is likely to stimulate many new investigations that probe the nature of fungal pathogenesis and evolution.

For now, the *C. albicans* genome sequence offers clues about the means by which *C. albicans* thrives in its host. For example, *C. albicans* has numerous large gene families, some of which encode known virulence attributes - such as secreted aspartyl proteinase (*SAP*) genes, secreted lipase (*LIP*) genes, agglutinin (*ALS*) genes and genes involved in

iron assimilation. Other gene families identified by genome sequencing may also contribute to the fitness of *C. albicans* in at least one of the niches it occupies and/or to its pathogenicity. *C. albicans* also contains multiple copies of genes involved in the tricarboxylic acid cycle, oligopeptide transport and sphingomyelin degradation. These may contribute to the efficient assimilation of available carbon sources when the fungus is growing in different microenvironments within the host. Also, the increased emphasis upon sulfur metabolism, compared with *S. cerevisiae* [2], might reflect an increased reliance upon glutathione metabolism and the relative resistance of *C. albicans* to oxidative stresses [20]. Presumably these would help the fungus resist oxidative killing by the host's immune defences. These (and other) speculations that emerge from scrutiny of the genome sequence now need to be tested experimentally.

To summarize, the *C. albicans* genome sequence is a very important step forward for researchers working on this fungus or on other pathogenic fungi. Classical genetic approaches have not been feasible for *C. albicans* because it is diploid and there has been no exploitable sexual cycle. Hence the genome sequence now provides an invaluable platform for the genomic screens that are so vital in the absence of genetic screens. We in the *C. albicans* research community are very grateful to the Stanford DNA Sequencing and Technology Center for their efforts.

References

1. **Sequencing of *Candida Albicans* at the Stanford Genome Technology Center**
[http://www.sequence.stanford.edu/group/candida/index.html]
2. Jones T, Federspiel NA, Chibana H, Dungan J, Kalman S, Magee BB, Newport G, Thorstenson YR, Agabian N, Magee PT, et al.: **The diploid genome sequence of *Candida albicans***. *Proc Natl Acad Sci USA* 2004, **101**:7329-7334.
3. Scherer S, Magee PT: **Genetics of *Candida albicans***. *Microbiol Rev* 1990, **54**:226-241.
4. Odds FC: *Candida and Candidosis*. 2nd edn. London: Bailliere Tindall; 1988.
5. Calderone RA: *Candida and Candidiasis*. Washington, DC: ASM Press; 2002.
6. Fonzi W, Irwin M: **Isogenic strain construction and gene mapping in *Candida albicans***. *Genetics* 1993, **134**:717-728.
7. Soll DR, Pujol C: ***Candida albicans* clades**. *FEMS Immunol Med Microbiol* 2003, **39**:1-7.
8. Odds FC, Van Nuffel L, Gow NAR: **Survival in experimental *Candida albicans* infections depends on inoculum growth conditions as well as animal host**. *Microbiology* 2000, **146**:1881-1889.
9. Hull CM, Johnson AD: **Identification of a mating type-like locus in the asexual pathogenic yeast *Candida albicans***. *Science* 1999, **285**:1271-1275.
10. Hull CM, Raisner RM, Johnson AD: **Evidence for mating of the "asexual" yeast *Candida albicans* in a mammalian host**. *Science* 2000, **289**:307-310.
11. Magee BB, Magee PT: **Induction of mating in *Candida albicans* by construction of *MTLa* and *MTLa* strains**. *Science* 2000, **289**:310-313.
12. Lockhart SR, Daniels KJ, Zhao R, Wessels D, Soll DR: **Cell biology of mating in *Candida albicans***. *Eukaryot Cell* 2003, **2**:49-61.
13. Tzung KW, Williams RM, Scherer S, Federspiel N, Jones T, Hansen N, Bivolarevic V, Huizar L, Komp C, Surzycki R, et al.: **Genomic evidence for a complete sexual cycle in *Candida albicans***. *Proc Natl Acad Sci USA* 2001, **98**:3249-3253.
14. Bennett RJ, Johnson AD: **Completion of a parasexual cycle in *Candida albicans* by induced chromosome loss in tetraploid strains**. *EMBO J* 2003, **22**:2505-2515.
15. Zhao XM, Pujol C, Soll DR, Hoyer LL: **Allelic variation in the contiguous loci encoding *Candida albicans* *ALS5*, *ALS1* and *ALS9***. *Microbiology* 2003, **149**:2947-2960.
16. Kohler JR, Fink GR: ***Candida albicans* strains heterozygous and homozygous for mutations in mitogen-activated protein kinase signaling components have defects in hyphal development**. *Proc Natl Acad Sci USA* 1996, **93**:13223-13228.
17. Sullivan DJ, Moran GP, Pinjon E, Al-Mosaid A, Stokes C, Vaughan C, Coleman DC: **Comparison of the epidemiology, drug resistance mechanisms, and virulence of *Candida dubliniensis* and *Candida albicans***. *FEMS Yeast Res* 2004, **4**:369-376.
18. Lin SJ, Schranz J, Teutsch SM: **Aspergillosis case-fatality rate: systematic review of the literature**. *Clin Infect Dis* 2001, **32**:358-366.
19. Gow NAR: **New angles in mycology: studies in directional growth and directional motility**. *Mycol Res* 2004, **108**:5-13.
20. Jamieson DJ, Stephen DWS, Terriere EC: **Analysis of the adaptive oxidative stress response of *Candida albicans***. *FEMS Microbiol Lett* 1996, **138**:83-88.