Genome Sequence of the Yeast *Clavispora lusitaniae* Type Strain CBS 6936

Pascal Durrens,[a,b,c] Christophe Klopp,[e] Nicolas Biteau,[b,d] Valérie Fitton-Ouhabi,[b,d] Karine Dementhon,[b,d] Isabelle Accoceberry,[b,d] David J. Sherman,[a,b,c] Thierry Noël[b,d]

Laboratoire Bordelais de Recherche en Informatique (LaBRI), UMR-CNRS 5800, Talence, France[a]; Université de Bordeaux, Bordeaux, France[b]; INRIA Joint Project Team PLEIADE, Talence, France[c]; Laboratoire de Microbiologie Fondamentale et Pathogénicité UMR-CNRS 5234, Bordeaux, France[d]; Genotoul Bioinfo, MIAT, INRA Toulouse, Castanet-Tolosan, France[e]

ABSTRACT *Clavispora lusitaniae*, an environmental saprophytic yeast belonging to the CTG clade of *Candida*, can behave occasionally as an opportunistic pathogen in humans. We report here the genome sequence of the type strain CBS 6936. Comparison with sequences of strain ATCC 42720 indicates conservation of chromosomal structure but significant nucleotide divergence.

Address correspondence to Pascal Durrens, pascal.durrens@labri.fr.

Clavispora lusitaniae, a teleomorph of *Candida lusitaniae*, is an environmentally ubiquitous ascomycetous yeast with no known specific ecological niche. It can be isolated from different substrates, such as soils, waters, plants, and gastrointestinal tracts of many animals including birds, mammals, and humans. In immunocompromised hosts, *C. lusitaniae* can be pathogenic and is responsible for about 1% of invasive candidiasis, particularly in pediatric and hematology-oncology patients (1).

So far, 2 strains have had their genomes sequenced: ATCC 42720, isolated from the blood of a patient with myeloid leukemia (2), and MTCC 1001, a self-fertile strain isolated from citrus (3). We report here the genome sequencing and assembly of the *C. lusitaniae* type strain CBS 6936 (4), isolated from citrus peel juice. Genomic DNA was isolated from a 50-mL yeast extract-peptone-dextrose culture, after spheroplast osmotic lysis and glass-rod purification of ethanol-precipitated nucleic acids. The DNA library was prepared from 1 μg according to the NEBNext DNA library prep master mix set for the Illumina (E6040) protocol with an insert size of 368 $\pm$ 122 nucleotides generated with a Covaris ultrasonicator.

The library was sequenced on the Illumina MiSeq version 1.18.54 platform, producing approximately 1.5 million paired-end reads, representing a 28$\times$ coverage. A *de novo* assembly of the reads was performed to compare the contigs to those of strain ATCC 42720. We used SPAdes version 3.9.0 with default parameters (5) to perform the assembly. Only the 53 scaffolds longer than 1 kb were retained, as the GC% of many shorter ones deviated from the average.

The scaffolds represent an overall length of 12.0 Mb, which is close to the 12.11 Mb of the ATCC 42720 genome. The GC contents amount to 44.5%, which is identical to that of strain ATCC 42720. The alignment of the CBS 6936 scaffolds against the ATCC 42720 contigs by Nucmer (6) shows almost colinearity between the two sets, apart from a 66-kb-long inversion. The number of single nucleotide polymorphisms (SNPs) between CBS 6936 and ATCC 42720 amounts to 132,141 with Phred quality higher than 30, according to Burrows–Wheeler alignment (7) of reads and FreeBayes detection (8). This SNP density of 1 SNP per 90 bp is twice the level observed between strains SC5314 and WO-1 of *Candida albicans*, which are members of different subgroups within the species and qualified as having diverged relatively recently (9).

Genes for 4 rRNAs, 197 tRNAs, and 5,539 proteins were predicted by alignment, TRNAscan-SE and a combination of Augustus (10), Snap (11), and GeneMark-ES (12), respectively. The average identity of proteins between the two strains amounts to 89.9% $\pm$ 23.2%, based on BLASTp (13) best-hit alignments.

Altogether, these results indicate that, in spite of a conserved genome structure, the sequences are different, meaning that the strains underwent a significant divergence.

Accession number(s). This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession numbers SRP075809 for reads and LYUB00000000 for scaffolds. The second version of the assembly of CBS 6936 is version LYUB02000000 and is described here.

REFERENCES

1. Favel A, Michel-Nguyen A, Peyron F, Martin C, Thomachot L, Datry A, Bouchara JP, Challier S, Noël T, Chastin C, Regli P. 2003. Colony morphology switching of *Candida lusitaniae* and acquisition of multidrug resistance during treatment of a renal infection in a newborn: case report and review of the literature. Diagn Microbiol Infect Dis 47:331–339. https://doi.org/10.1016/S0732-8893(03)00094-4.

2. Pappagianis D, Collins MS, Hector R, Remington J. 1979. Development of resistance to amphotericin B in *Candida lusitaniae* infecting a human. Antimicrob Agents Chemother 16:123–126. https://doi.org/10.1128/AAC.16.2.123.

3. Lachance MA, Daniel HM, Meyer W, Prasad GS, Gautam SP, Boundy-Mills K. 2003. The D1/D2 domain of the large-subunit rDNA of the yeast species *Clavispora lusitaniae* is unusually polymorphic. FEMS Yeast Res 4:253–258. https://doi.org/10.1016/S1567-1356(03)00113-2.

4. Rodrigues de Miranda L. 1979. *Clavispora*, a new yeast genus of the Saccharomycetales. Antonie Van Leeuwenhoek 45:479–483. https://doi.org/10.1007/BF00443285.

5. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021.

6. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. 2004. Versatile and open software for comparing large genomes. Genome Biol 5:R12. https://doi.org/10.1186/gb-2004-5-2-r12.

7. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25:1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

8. Garrison E, Marth G. 2012. Haplotype-based variant detection from short-read sequencing, preprint, arXiv:1207.3907 [q-bio.GN]. https://arxiv.org/abs/1207.3907v2.

9. Butler G, Rasmussen MD, Lin MF, Santos MAS, Sakthikumar S, Munro CA, Rheinbay E, Grabherr M, Forche A, Reedy JL, Agrafioti I, Arnaud MB, Bates S, Brown AJP, Brunke S, Costanzo MC, Fitzpatrick DA, de Groot PWJ, Harris D, Hoyer LL, Hube B, Klis FM, Kodira C, Lennard N, Logue ME, Martin R, Neiman AM, Nikolaou E, Quail MA, Quinn J, Santos MC, Schmitzberger FF, Sherlock G, Shah P, Silverstein KA, Skrzypek MS, Soll D, Staggs R, Stansfield I, Stumpf MPH, Sudbery PE, Thyagarajan S, Zeng Q, Berman J, Berriman M, Heitman J, Gow NAR, Lorenz MC, Birren BW, Kellis M, Cuomo CA. 2009. Evolution of pathogenicity and sexual reproduction in eight *Candida* genomes. Nature 459:657–662. https://doi.org/10.1038/nature08064.

10. Stanke M, Schöffmann O, Morgenstern B, Waack S. 2006. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. BMC Bioinformatics 7:62. https://doi.org/10.1186/1471-2105-7-62.

11. Korf I. 2004. Gene finding in novel genomes. BMC Bioinformatics 5:59. https://doi.org/10.1186/1471-2105-5-59.

12. Ter-Hovhannisyan V, Lomsadze A, Chernoff YO, Borodovsky M. 2008. Gene prediction in novel fungal genomes using an *ab initio* algorithm with unsupervised training. Genome Res 18:1979–1990. https://doi.org/10.1101/gr.081612.108.

13. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421. https://doi.org/10.1186/1471-2105-10-421.