

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

Pascal Durrens, Christophe Klopp, Nicolas Biteau, Valérie Fitton-Ouhabi,
Karine Dementhon, Isabelle Accoceberry, David J. Sherman, Thierry Noël

► **To cite this version:**

Pascal Durrens, Christophe Klopp, Nicolas Biteau, Valérie Fitton-Ouhabi, Karine Dementhon, et al..
Genome Sequence of the Yeast *Clavispora lusitaniae* Type Strain CBS 6936. *Genome Announcements*,
American Society for Microbiology, 2017, 5 (31), pp.e00724-17. 10.1128/genomeA.00724-17. hal-
01583349

HAL Id: hal-01583349

<https://hal.inria.fr/hal-01583349>

Submitted on 7 Sep 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



1 The genome of the yeast *Clavispora lusitaniae* type strain CBS 6936

2

3 Pascal Durrens^{#abc}, Christophe Klopp^e, Nicolas Biteau^{bd}, Valérie Fitton-Ouhabi^{bd},

4 Karine Dementhon^{bd}, Isabelle Accoceberry^{bd}, David J. Sherman^{cab}, Thierry Noël^{bd}

5

6 Laboratoire Bordelais de Recherche en Informatique (LaBRI), UMR-CNRS 5800,

7 Talence, France^a; Université de Bordeaux, Bordeaux, France^b; INRIA joint project

8 team PLEIADE^c; Laboratoire de Microbiologie Fondamentale et Pathogénicité UMR-

9 CNRS 5234, Bordeaux, France^d; Genotoul Bioinfo, MIAT, INRA Toulouse, Castanet-

10 Tolosan, France^e

11

12

13 Running Head: Genome of *C. lusitaniae* type strain CBS 6936

14

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

16

17 **Abstract**

18

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

24

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

31

32 So far, two strains have had their genomes sequenced : ATCC 42720, isolated from the
33 blood of a patient with myeloid leukemia (2), and MTCC 1001, a self-fertile strain
34 isolated from citrus (3). We report here the genome sequencing and assembly of the *C.*
35 *lusitaniae* type strain CBS 6936 (4), isolated from citrus peel juice. Genomic DNA was
36 isolated from a 50 ml YPD yeast culture, after sphaeroplast osmotic lysis and glass rod
37 purification of ethanol-precipitated nucleic acids. The DNA library was prepared from
38 1 µg according to the NEBNext DNA Library Prep Master Mix Set for Illumina
39 (E6040) protocol with an insert size of 368±122nt generated with an ultrasonicator
40 Covaris.

41

42 The library was sequenced on a MiSeq Illumina v1.18.54 producing *circa* 1.5 million
43 paired end reads, representing a 28X coverage. A *de novo* assembly of the reads was
44 performed with the aim to compare the contigs to those of the ATCC 42720 strain. We
45 used SPAdes v3.9.0 with default parameters (5) to perform the assembly. Only the 53
46 scaffolds longer than 1kb were retained as the GC % of many shorter ones deviated
47 from the average.

48

49 The scaffolds represent an overall length of 12.0Mb which is close to the 12.11Mb of
50 the ATCC 42720 genome. The GC% amounts to 44.5% which is identical to that of
51 ATCC 42720 strain. The alignment of CBS 6936 scaffolds against ATCC 42720 contigs
52 by Nucmer (6) shows almost co-linearity between the two sets, apart from a 66kb long
53 inversion. The number of SNPs between CBS 6936 and ATCC 42720 amounts to
54 132141 SNPs with Phred quality higher than 30, according to BWA (7) alignment of
55 reads and Freebayes detection (8). This SNP density of 1 SNP per 90bp is twice the
56 level observed between the strains SC5314 and WO-1 of *Candida albicans*, members
57 of different subgroups in the species and qualified as relatively recent divergent (9).

58

59 Genes for 4 rRNA, 197 tRNA and 5539 proteins were predicted by alignment,
60 TRNAscan-SE and a combination of Augustus (10), Snap (11) and Genmark-ES (12),
61 respectively. The average identity of proteins between the two strains amounts to
62 $89.9\% \pm 23.2$, based on BLASTp (13) best hit alignments.

63

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

66

67 **Accession numbers.** This whole genome shotgun project has been deposited at
68 DDBJ/EMBL/GenBank under the accession numbers SRP075809 for reads and
69 LYUB00000000 for scaffolds. The second version of the assembly of CBS 6936 is
70 version LYUB02000000 and is described here.

71

72 **Acknowledgements**

73

74 The library and nucleotide sequencing were performed at the Genomic and Transcrip-
75 tomic Facility of Bordeaux (grants from the Conseil Regional d'Aquitaine
76 n°20030304002FA and 20040305003FA and from the European Union, FEDER
77 n°2003227 and from Investissements d'avenir, Convention attributive d'aide N°ANR-
78 10-EQPX-16-01). We would like to thank Christophe Hubert for its expertise and tech-
79 nical assistance.

80 The development of the bioinformatic pipeline was carried out using the PLAFRIM
81 experimental testbed, being developed under the Inria PlaFRIM development action
82 with support from Bordeaux INP, LABRI and IMB and other entities: Conseil
83 Régional d'Aquitaine, Université de Bordeaux and CNRS (and ANR in accordance to
84 the programme d'investissements d'Avenir (see <https://plafrim.fr/>)).

85

87

- 88 1. **Favel A, Michel-Nguyen A, Peyron F, Martin C, Thomachot L, Datry A,**
89 **Bouchara JP, Challier S, Noël T, Chastin C, Regli P.** 2003. Colony
90 morphology switching of *Candida lusitanae* and acquisition of multidrug
91 resistance during treatment of a renal infection in a newborn: case report and
92 review of the literature. *Diagn. Microbiol. Infect. Dis.* **47**:331-339.
- 93 2. **Pappagianis D, Collins MS, Hector R, Remington J.** 1979. Development of
94 resistance to amphotericin B in *Candida lusitanae* infecting a human.
95 *Antimicrob Agents Chemother.* **16**:123-126.
- 96 3. **Lachance M, Daniel H, Meyer W, Prasad G, Gautam S, Boundymills K.**
97 2003. The D1/D2 domain of the large-subunit rDNA of the yeast species is
98 unusually polymorphic. *FEMS Yeast Res* **4**:253–258.
- 99 4. **Rodrigues de Miranda L.** 1979. *Clavispora*, a new yeast genus of
100 Saccharomycetales. *Antonie van Leeuwenhoek.* **45**:479-483.
- 101 5. **Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS,**
102 **Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin**
103 **AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA.** 2012. SPAdes: A New
104 Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J*
105 *Comput Biol.* **5**:455-477.
- 106 6. **Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C,**
107 **Salzberg SL.** 2004. Versatile and open software for comparing large genomes.
108 *Genome Biology,* **5**:R12.
- 109 7. **Li H, Durbin R.** 2009. Fast and accurate short read alignment with Burrows-

- 110 Wheeler Transform. *Bioinformatics*, 25:1754-1760.
- 111 8. **Garrison E, Marth G.** 2012. Haplotype-based variant detection from short-
112 read sequencing. arXiv preprint arXiv:1207.3907 [q-bio.GN].
- 113 9. **Butler G, Rasmussen MD, Lin MF, , Santos MAS, Sakthikumar S, Munro**
114 **CA, Rheinbay E, Grabherr M, Forche A, Reedy JL, Agrafioti I, Arnaud**
115 **MB, Bates S, Brown AJP, Brunke S, Costanzo MC, Fitzpatrick DA, de**
116 **Groot PWJ, Harris D, Hoyer LL, Hube B, Klis FM, Kodira C, Lennard N,**
117 **Logue ME, Martin R, Neiman AM, Nikolaou E, Quail MA, Quinn J,**
118 **Santos MC, Schmitzberger FF, Sherlock G, Shah P, Silverstein K,**
119 **Skrzypek MS, Soll D, Staggs R, Stansfield I, Stumpf MPH, Sudbery PE,**
120 **Thyagarajan S, Zeng Q, Berman J, Berriman M, Heitman J, Gow NAR,**
121 **Lorenz MC, Birren BW, Kellis M, Cuomo CA.** 2009. Evolution of
122 pathogenicity and sexual reproduction in eight *Candida* genomes. *Nature*
123 459:657-662.
- 124 10. **Stanke M, Schöffmann O, Morgenstern B, Waack S.** 2006. Gene prediction
125 in eukaryotes with a generalized hidden Markov model that uses hints from
126 external sources. *BMC Bioinformatics* 7, 62.
- 127 11. **Korf I.** 2004. Gene finding in novel genomes. *BMC Bioinformatics* 5, 59.
- 128 12. **Ter-Hovhannisyan V, Lomsadze A, Chernoff YO, Borodovsky M.** 2008.
129 Gene prediction in novel fungal genomes using an *ab initio* algorithm with
130 unsupervised training. *Genome Res.* 18:1979-1990.
- 131 13. **Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K,**
132 **Madden TL.** 2008 BLAST+: architecture and applications. *BMC*
133 *Bioinformatics* 10:421.