



Genome Sequence of the Oleaginous Yeast *Yarrowia lipolytica* H222

Hugo Devillers,^a  Cécile Neuvéglise^a

^aMicalis Institute, INRA, AgroParisTech, Université Paris-Saclay, Jouy-en-Josas, France

ABSTRACT Here, we report the genome sequence of the oleaginous yeast *Yarrowia lipolytica* H222. *De novo* genome assembly shows three main chromosomal rearrangements compared to that of strain E150/CLIB122. This genomic resource will help integrate intraspecies diversity into synthetic biology projects that utilize *Yarrowia* as a biotechnological chassis for value-added chemical productions.

The yeast *Yarrowia lipolytica* belongs to the “basal” lineages of the subphylum *Saccharomycotina*. Its oleaginous capacities in hydrophobic environments, the development of genetic tools, and a first available genome sequence in 2004 (1) have made it an interesting candidate for biotechnological applications for more than 30 years. Recent developments in synthetic biology and metabolic engineering have contributed to increasing interest in this yeast, which now emerges as a major host for chemical production (2). Here, we sequenced the genome of the German strain H222, which is one of the most utilized strains for biotechnological applications, such as production of organic acids (3–5).

Total genomic DNA of H222 cells grown in complete medium to the stationary phase was used to construct a shotgun 400-bp insert library (PE) and a mate pair 8-kb insert library (MP). Both libraries were sequenced in paired-end (2×100 bp) using the Illumina HiSeq 2000 platform with chemistry v3 (PE) and v2 (MP), resulting in a raw sequencing depth of $275\times$ (28,182,153 reads) and $47\times$ (4,798,539 reads) for PE and MP, respectively. Sequencing reads were cleaned using Trimmomatic v0.32 (6) and Cutadapt (7) with the options ILLUMINACLIP:<adaptors.fasta>:2:15:5 LEADING:5 TRAILING:5 SLIDINGWINDOW:5:20 MINLEN:36 and $-\text{error-rate} = 0.2$, respectively. Note that Cutadapt was used only for 5' adaptor clipping. After trimming, 26,533,605 PE reads ($255\times$) and 1,325,112 MP reads ($13\times$) were used for *de novo* genome assembly using SOAPdenovo2 v2.04 (8), with a k-mer value of 77, as estimated with kmergenie version 1.67 (9). Gap closure was performed using GapCloser v1.12 (8). The final assembly comprised 17 scaffolds larger than 5 kb (N_{50} of 3.9 Mb, obtained with three scaffolds) for a cumulative length of 20,519,037 bp. A single scaffold of 48,435 bp corresponded to mitochondrial DNA. The remaining 16 scaffolds were suitable for automatic annotation using Rapid Annotation Transfer Tool (RATT) (10) with the *Y. lipolytica* E150 genome sequence as a reference (genome sequence available at <http://gryc.inra.fr>). Manual curation was performed with transcriptome sequencing (RNA-Seq) reads (BioProject accession number [PRJEB29941](https://www.ncbi.nlm.nih.gov/bioproject/PRJEB29941)) mapped to the assembly with TopHat2 (11). In total, 6,490 protein-coding genes were predicted, including 6,415 coding sequences (CDS) and 128 pseudogenes. A set of 510 nuclear tRNA genes were identified using tRNAscan-SE v1.3.1 (12). Transposable elements (TE) were identified by a BLAST search using different TE families from *Y. lipolytica* (13–18). A total of 88 solo long terminal repeats (LTR), mainly from Tyl5 (19), and 108 intact or remnant TE were annotated, including 97 copies of Ylli (13).

A draft genome sequence of H222 is already available (assembly ASM305430v1,

Citation Devillers H, Neuvéglise C. 2019. Genome sequence of the oleaginous yeast *Yarrowia lipolytica* H222. Microbiol Resour Announc 8:e01547-18. <https://doi.org/10.1128/MRA.01547-18>.

Editor Jason E. Stajich, University of California, Riverside

Copyright © 2019 Devillers and Neuvéglise. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Cécile Neuvéglise, cecile.neuveglise@inra.fr.

Received 12 November 2018

Accepted 18 December 2018

Published 24 January 2019

submitted by Patrice Lubuta's laboratory). However, the assembly is probably incorrect, since it is completely colinear to the genome of E150/CLIB122 (1), even though large differences have been observed in their karyotypes (20). In our assembly, we found two major reciprocal translocations compared to E150 chromosomes, involving scaffolds H222S03/S06 and H222S04/S08, and a large inversion of 300 kb in H222S01 compared to chromosome YAL10E of E150. This explains the observed chromosome size differences and shows that chromosomal rearrangements occurred in *Y. lipolytica*. Consequently, *de novo* assembly should be preferred over reference-assisted scaffolding for this species.

Data availability. The draft genome sequence of *Yarrowia lipolytica* strain H222 has been deposited in DDBJ/ENA/GenBank under the accession number [GCA_900537225](https://doi.org/10.1093/bioinformatics/btt310). The version described in this paper is the first version. The accession number for the project is [PRJEB28424](https://doi.org/10.1093/bioinformatics/btt310), and for the reads, [ERR2767096](https://doi.org/10.1093/bioinformatics/btt310) (PE), [ERR2767094](https://doi.org/10.1093/bioinformatics/btt310) (MP), and [ERR2767095](https://doi.org/10.1093/bioinformatics/btt310) (MP). The accession numbers of the 17 scaffolds are [UTQH01000001](https://doi.org/10.1093/bioinformatics/btt310) to [UTQH01000017](https://doi.org/10.1093/bioinformatics/btt310). Genome sequences and annotations are also available at the GRYC server (<http://gryc.inra.fr>).

ACKNOWLEDGMENT

We thank François Brunel for his contribution to genome assembly.

REFERENCES

- Dujon B, Sherman D, Fischer G, Durrens P, Casaregola S, Lafontaine I, De Montigny J, Marck C, Neuveglise C, Talla E, Goffard N, Frangeul L, Aigle M, Anthouard V, Babour A, Barbe V, Barnay S, Blanchin S, Beckerich J-M, Beyne E, Bleykasten C, Boisramé A, Boyer J, Cattolico L, Confanioleri F, De Daruvar A, Despons L, Fabre E, Fairhead C, Ferry-Dumazet H, Groppi A, Hantraye F, Hennequin C, Jauniaux N, Joyet P, Kachouri R, Kerrest A, Koszul R, Lemaire M, Lesur I, Ma L, Muller H, Nicaud J-M, Nikolski M, Oztas S, Ozier-Kalogeropoulos O, Pellenz S, Potier S, Richard G-F, Straub M-L, Suleau A, Swennen D, Tekaija F, Wésolowski-Louvel M, Westhof E, Wirth B, Zeniou-Meyer M, Zivanovic I, Bolotin-Fukuhara M, Thierry A, Bouchier C, Caudron B, Scarpelli C, Gaillardin C, Weissenbach J, Wincker P, Souciet J-L. 2004. Genome evolution in yeasts. *Nature* 430:35–44. <https://doi.org/10.1038/nature02579>.
- Markham KA, Alper HS. 2018. Synthetic biology expands the industrial potential of *Yarrowia lipolytica*. *Trends Biotechnol* 36:1085–1095. <https://doi.org/10.1016/j.tibtech.2018.05.004>.
- Forster A, Aurich A, Mauersberger S, Barth G. 2007. Citric acid production from sucrose using a recombinant strain of the yeast *Yarrowia lipolytica*. *Appl Microbiol Biotechnol* 75:1409–1417. <https://doi.org/10.1007/s00253-007-0958-0>.
- Jost B, Holz M, Aurich A, Barth G, Bley T, Muller RA. 2015. The influence of oxygen limitation for the production of succinic acid with recombinant strains of *Yarrowia lipolytica*. *Appl Microbiol Biotechnol* 99:1675–1686. <https://doi.org/10.1007/s00253-014-6252-z>.
- Yovkova V, Otto C, Aurich A, Mauersberger S, Barth G. 2014. Engineering the alpha-ketoglutarate overproduction from raw glycerol by overexpression of the genes encoding NADP⁺-dependent isocitrate dehydrogenase and pyruvate carboxylase in *Yarrowia lipolytica*. *Appl Microbiol Biotechnol* 98:2003–2013. <https://doi.org/10.1007/s00253-013-5369-9>.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
- Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *Embnet J* 17:10–12. <https://doi.org/10.14806/ej.17.1.200>.
- Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, Tang J, Wu G, Zhang H, Shi Y, Liu Y, Yu C, Wang B, Lu Y, Han C, Cheung DW, Yiu SM, Peng S, Xiaoqian Z, Liu G, Liao X, Li Y, Yang H, Wang J, Lam TW, Wang J. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. *Gigascience* 1:18. <https://doi.org/10.1186/2047-217X-1-18>.
- Chikhi R, Medvedev P. 2014. Informed and automated *k*-mer size selection for genome assembly. *Bioinformatics* 30:31–37. <https://doi.org/10.1093/bioinformatics/btt310>.
- Otto TD, Dillon GP, Degraeve WS, Berriman M. 2011. RATT: Rapid Annotation Transfer Tool. *Nucleic Acids Res* 39:e57. <https://doi.org/10.1093/nar/gkq1268>.
- Kim D, Perteza G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* 14:R36. <https://doi.org/10.1186/gb-2013-14-4-r36>.
- Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 25:955–964. <https://doi.org/10.1093/nar/25.5.955>.
- Casaregola S, Neuveglise C, Bon E, Gaillardin C. 2002. Ylli, a non-LTR retrotransposon L1 family in the dimorphic yeast *Yarrowia lipolytica*. *Mol Biol Evol* 19:664–677. <https://doi.org/10.1093/oxfordjournals.molbev.a004125>.
- Kovalchuk A, Senam S, Mauersberger S, Barth G. 2005. Tyl6, a novel Ty3/gypsy-like retrotransposon in the genome of the dimorphic fungus *Yarrowia lipolytica*. *Yeast* 22:979–991. <https://doi.org/10.1002/yea.1287>.
- Magnan C, Yu J, Chang I, Jahn E, Kanomata Y, Wu J, Zeller M, Oakes M, Baldi P, Sandmeyer S. 2016. Sequence assembly of *Yarrowia lipolytica* strain W29/CLIB89 shows transposable element diversity. *PLoS One* 11:e0162363. <https://doi.org/10.1371/journal.pone.0162363>.
- Neuveglise C, Chalvet F, Wincker P, Gaillardin C, Casaregola S. 2005. Mutator-like element in the yeast *Yarrowia lipolytica* displays multiple alternative splicings. *Eukaryot Cell* 4:615–624. <https://doi.org/10.1128/EC.4.3.615-624.2005>.
- Neuveglise C, Feldmann H, Bon E, Gaillardin C, Casaregola S. 2002. Genomic evolution of the long terminal repeat retrotransposons in hemiascomycetous yeasts. *Genome Res* 12:930–943. <https://doi.org/10.1101/gr.219202>.
- Schmid-Berger N, Schmid B, Barth G. 1994. Ylt1, a highly repetitive retrotransposon in the genome of the dimorphic fungus *Yarrowia lipolytica*. *J Bacteriol* 176:2477–2482. <https://doi.org/10.1128/jb.176.9.2477-2482.1994>.
- Devillers H, Brunel F, Polomska X, Sarilar V, Lazar Z, Robak M, Neuveglise C. 2016. Draft genome sequence of *Yarrowia lipolytica* strain A-101 isolated from polluted soil in Poland. *Genome Announc* 4:e01094-16. <https://doi.org/10.1128/genomeA.01094-16>.
- Casaregola S, Feynerol C, Diez M, Fournier P, Gaillardin C. 1997. Genomic organization of the yeast *Yarrowia lipolytica*. *Chromosoma* 106:380–390.