Draft Genome Sequence of *Yarrowia lipolytica* Strain A-101 Isolated from Polluted Soil in Poland

Hugo Devillers,[a] François Brunel,[a] Xymena Połomska,[b] Véronique Sarilar,[a] Zbigniew Lazar,[a,b] Małgorzata Robak,[b] Cécile Neuvéglise[a]

Micalis Institute, INRA, AgroParisTech, Université Paris-Saclay, Jouy-en-Josas, France[a]; Department of Biotechnology and Food Microbiology, Wrocław University of Environmental and Life Sciences, Chelmonskiego, Wrocław, Poland[b]

Yarrowia lipolytica is an early diverging species of the *Saccharomycotina* subphylum, which is recognized as a valuable host for many biotechnological applications exploiting its oleaginous capacities. The 20.5-Mb genome of the Polish *Y. lipolytica* strain A-101 will greatly help decipher the genetic basis of the regulation of its lipid metabolism.

Yarrowia lipolytica is one of the most extensively studied oleaginous yeasts used as a cell factory, which shows great capacity to produce a number of biotechnologically important metabolites, such as organic acids, enzymes, polyols, fatty acids, and aromas (1–7). Due to its ability to degrade organic compounds, including aliphatic and aromatic hydrocarbons, *Y. lipolytica* is also used in bioremediation and environment protection (8, 9).

Up to now, the whole-genome sequences of three *Y. lipolytica* strains are available, E150/CLIB122 (10), WSH-Z06 (BioProject PRJEB5051), and PO1f (11), as well as a draft of the W29 genome sequence (12). W29 and its derivative PO1f originate from France, whereas E150 derives from a cross between W29 and the American strain CBS6124-2. WSH-Z06 is a Chinese strain used for alpha-ketoglutaric acid production (13). Here, we present the genome sequence of strain A-101 isolated from polluted soil at a car wash at Wrocław, Poland (14). A-101 has been investigated for citrate biosynthesis from various substrates in different bioreactor systems (15, 16) and was used for soil bioremediation (17, 18). It was intensively mutated and genetically engineered to produce strains improved for citrate (19) and erythritol (16) biosynthesis.

A-101 DNA was sequenced with the Illumina HiSeq DNA sequencing platform (paired-end [PE] 2 × 100 bp), with a shotgun library of 280-bp inserts and a mate-pair library of 7.2 kb on average. The raw reads were trimmed with Trimmomatic version 0.32 (20) and cutadapt version 1.8.3 (21). The assembly was done using SOAP*denovo*2 version 2.04 (22), with a k-mer of 59, as estimated with kmergenie version 1.67 (23). Two successive runs of GapCloser from the Short Oligonucleotide Analysis Package (http://soap.genomics.org.cn/index.html) were used to close gaps, and manual curation was performed in overlapping regions. The current draft genome sequence is composed of 29 scaffolds larger than 5 kb (28 nuclear and a mitochondrial one), for a total size of nuclear DNA of 20,581,016 bp, with an N_{50} of 2,267,247 bp (4 scaffolds) and a G+C content of 49.03%. Coding sequence (CDS) prediction was essentially performed using the Amadea automatic annotation transfer software (24, 25), with E150 as the reference genome. A total of 6,576 putative CDS (including 15

alternative isoforms) were identified after manual curation. tRNA genes were determined using tRNAscan-SE version 1.4 (26) and 5S rRNA genes by similarity to E150. Transposable elements (TE) were manually annotated by similarity to yeast TE, including those of strain E150 (27). In addition to copies of Tyl6, Ylli, and Mutyl, a relic of Fotyl, and a single long terminal repeat (LTR) of Ylt1, a new full-length TE of the *Ty1/Copia* superfamily, named Tyl5, was identified with LTR corresponding to the previously known LTRyl1 (28).

By mapping the reads to the E150 genome using BWA version 0.7.10 (29) and analyzing them using Samtools version 1.2 (30), a total of 38,802 single nucleotide polymorphisms and 3,353 short indels were called in the A-101 genome sequence (nucleotide frequency of the reference (QS), <0.1; read depth (DP), ≥30), i.e., 2,048 nucleotide variations per Mb. Further comparison of this genome against other strains of *Y. lipolytica* and species of the *Yarrowia* clade will bring additional insights into gene functions and evolutionary events.

Accession number(s). This whole-genome shotgun analysis has been deposited at DDBJ/EMBL/GenBank under the accession no. FLLM00000000 (BioProject PRJEB14097 and scaffold sequences LT576309 to LT576337). The version described in this project is the first version, FLLM01000000. All data are also available at http://gryc.inra.fr.

REFERENCES

1. **Förster 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. http://dx.doi.org/10.1007/s00253-007-0958-0.
2. **Gonçalves FA, Colen G, Takahashi JA**. 2014. *Yarrowia lipolytica* and its multiple applications in the biotechnological industry. ScientificWorldJournal **2014**:476207. http://dx.doi.org/10.1155/2014/476207.
3. **Zinjarde SS**. 2014. Food-related applications of *Yarrowia lipolytica*. Food Chem **152**:1–10. http://dx.doi.org/10.1016/j.foodchem.2013.11.117.

4. **Zhu Q, Jackson EN**. 2015. Metabolic engineering of *Yarrowia lipolytica* for industrial applications. Curr Opin Biotechnol **36:**65–72. http://dx.doi.org/10.1016/j.copbio.2015.08.010.

5. **Morgunov IG, Kamzolova SV, Lunina JN**. 2013. The citric acid production from raw glycerol by *Yarrowia lipolytica* yeast and its regulation. Appl Microbiol Biotechnol **97:**7387–7397. http://dx.doi.org/10.1007/s00253-013-5054-z.

6. **Mirończuk AM, Furgała J, Rakicka M, Rymowicz W**. 2014. Enhanced production of erythritol by *Yarrowia lipolytica* on glycerol in repeated batch cultures. J Ind Microbiol Biotechnol **41:**57–64. http://dx.doi.org/10.1007/s10295-013-1380-5.

7. **Ledesma-Amaro R, Nicaud JM**. 2016. *Yarrowia lipolytica* as a biotechnological chassis to produce usual and unusual fatty acids. Prog Lipid Res **61:**40–50. http://dx.doi.org/10.1016/j.plipres.2015.12.001.

8. **Bankar AV, Kumar AR, Zinjarde SS**. 2009. Environmental and industrial applications of *Yarrowia lipolytica*. Appl Microbiol Biotechnol **84:**847–865. http://dx.doi.org/10.1007/s00253-009-2156-8.

9. **Zinjarde S, Apte M, Mohite P, Kumar AR**. 2014. *Yarrowia lipolytica* and pollutants: interactions and applications. Biotechnol Adv **32:**920–933. http://dx.doi.org/10.1016/j.biotechadv.2014.04.008.

10. **Dujon B, Sherman D, Fischer G, Durrens P, Casaregola S, Lafontaine I, De Montigny J, Marck C, Neuvéglise C, Talla E, Goffard N, Frangeul L, Aigle M, Anthouard V, Babour A, Barbe V, Barnay S, Blanchin S, Beckerich JM, Beyne E, Bleykasten C, Boisrame 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 JM, Nikolski M, Oztas S, Ozier-Kalogeropoulos O, Pellenz S, Potier S, Richard GF, Straub ML**. 2004. Genome evolution in yeasts. Nature **430:**35–44. http://dx.doi.org/10.1038/nature02579.

11. **Liu L, Alper HS**. 2014. Draft genome sequence of the oleaginous yeast *Yarrowia lipolytica* PO1f, a commonly used metabolic engineering host. Genome Announc 2(4):e00652-14. http://dx.doi.org/10.1128/genomeA.00652-14.

12. **Pomraning KR, Baker SE**. 2015. Draft genome sequence of the dimorphic yeast *Yarrowia lipolytica* strain W29. Genome Announc 3(6):e01211-15. http://dx.doi.org/10.1128/genomeA.01211-15.

13. **Yin X, Madzak C, Du G, Zhou J, Chen J**. 2012. Enhanced alpha-ketoglutaric acid production in *Yarrowia lipolytica* WSH-Z06 by regulation of the pyruvate carboxylation pathway. Appl Microbiol Biotechnol **96:**1527–1537. http://dx.doi.org/10.1007/s00253-012-4192-z.

14. **Wojtatowicz M, Rymowicz W, Kautola H**. 1991. Comparison of different strains of the yeast *Yarrowia lipolytica* for citric acid production from glucose hydrol. Appl Biochem Biotechnol **31:**165–174. http://dx.doi.org/10.1007/BF02921787.

15. **Rywińska A, Rymowicz W, Zarowska B, Skrzypiński A**. 2010. Comparison of citric acid production from glycerol and glucose by different strains of *Yarrowia lipolytica*. World J Microbiol Biotechnol **26:**1217–1224. http://dx.doi.org/10.1007/s11274-009-0291-0.

16. **Rywińska A, Juszczyk P, Wojtatowicz M, Robak M, Lazar Z, Tomaszewska L, Rymowicz W**. 2013. Glycerol as a promising substrate for *Yarrowia lipolytica* biotechnological applications. Biomass Bioenerg **48:**148–166. http://dx.doi.org/10.1016/j.biombioe.2012.11.021.

17. **Żogała B, Robak M, Rymowicz W, Wzientek K, Rusin M, Maruszczak J**. 2005. Geoelectrical observation of *Yarrowia lipolytica* bioremediation of petrol contaminated soil. Pol J Environ Stud **14:**665–669.

18. **Robak M, Boruczkowski T, Drozdz W, Lazar Z, Baranowska M, Przado D**. 2011. Application of the yeasts *Yarrowia lipolytica* for *in-situ* bioremediation of soil contaminated with creosote oil—a case study. Ochrona Srodowiska **33:**27–33.

19. **Lazar Z, Rossignol T, Verbeke J, Crutz-Le Coq AM, Nicaud JM, Robak M**. 2013. Optimized invertase expression and secretion cassette for improving *Yarrowia lipolytica* growth on sucrose for industrial applications. J Ind Microbiol Biotechnol **40:**1273–1283. http://dx.doi.org/10.1007/s10295-013-1323-1.

20. **Bolger AM, Lohse M, Usadel B**. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics **30:**2114–2120. http://dx.doi.org/10.1093/bioinformatics/btu170.

21. **Martin M**. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J **17:**10–12. http://dx.doi.org/10.14806/ej.17.1.200.

22. **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. SOAP*denovo*2: an empirically improved memory-efficient short-read *de novo* assembler. Gigascience **1:**18. http://dx.doi.org/10.1186/2047-217X-1-18.

23. **Chikhi R, Medvedev P**. 2014. Informed and automated k-mer size selection for genome assembly. Bioinformatics **30:**31–37. http://dx.doi.org/10.1093/bioinformatics/btt310.

24. **De Brevern AG, Meyniel JP, Fairhead C, Neuvéglise C, Malpertuy A**. 2015. Trends in IT innovation to build a next generation bioinformatics solution to manage and analyse biological big data produced by NGS technologies. Biomed Res Int **2015:**904541. http://dx.doi.org/10.1155/2015/904541.

25. **Vakirlis N, Sarilar V, Drillon G, Fleiss A, Agier N, Meyniel JP, Blanpain L, Carbone A, Devillers H, Dubois K, Gillet-Markowska A, Graziani S, Huu-Vang N, Poirel M, Reisser C, Schott J, Schacherer J, Lafontaine I, Llorente B, Neuvéglise C, Fischer G**. 2016. Reconstruction of ancestral chromosome architecture and gene repertoire reveals principles of genome evolution in a model yeast genus. Genome Res **26:**918–932. http://dx.doi.org/10.1101/gr.204420.116.

26. **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. http://dx.doi.org/10.1093/nar/25.5.0955.

27. **Bleykasten-Grosshans C, Neuvéglise C**. 2011. Transposable elements in yeasts. C R Biol **334:**679–686. http://dx.doi.org/10.1016/j.crvi.2011.05.017.

28. **Neuvéglise 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. http://dx.doi.org/10.1101/gr.219202.

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

30. **Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing Subgroup**. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics **25:**2078–2079.