CrossMark
click for updates

Draft Genome Sequence of the Dimorphic Yeast *Yarrowia lipolytica* Strain W29

Kyle R. Pomraning, Scott E. Baker

Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, Washington, USA

Here, we present the draft genome sequence of the dimorphic ascomycete yeast *Yarrowia lipolytica* strain W29 (ATCC 20460). *Y. lipolytica* is a commonly employed model for the industrial production of lipases, small molecules, and more recently for its ability to accumulate lipids.

We sequenced genomic DNA from six mutants derived from UV mutagenesis of *Yarrowia lipolytica* strain W29 (ATCC 20460) using a whole-genome shotgun sequencing approach on an Illumina HiSeq instrument. Ten million paired-end 150-bp reads with an average insert size of 243 ± 28 bp (standard deviation) from each strain (60 million reads total) were assembled into a consensus sequence of 369 contigs using Velvet version 1.2.10 (1). A *k*-mer length of 87 nucleotides (nt) was chosen to optimize for the highest N_{50}. The resulting assembly had a size of 20.3 Mb (N_{50}, 125,743 bp; N_{max}, 356,851 bp; median coverage, 148.2×) and G+C content of 49.0%. The contigs were scaffolded to six chromosomes based on the reference strain CLIB122 (2) using CONTIGuator (3) and named YaliW29_A through YaliW29_F. Unscaffolded contigs include three high-coverage contigs that are predicted to be repetitive nuclear DNA. These are the 18S rRNA repeats (2,092 bp; ~182 copies), the 26S rRNA repeats (2,283 bp; ~177 copies), and a repeat (2,289 bp; ~20 copies) with high homology to the non-long terminal repeat (LTR) retrotransposon Ylli (4). Fifteen short AT-rich contigs (44,803-bp total, $22.6 \pm 1.7\%$ G+C content [standard deviation]) with high coverage ($24.0 \pm 1.8\times$ [standard deviation] that of the nuclear genome) were identified that aligned well to previously published *Y. lipolytica* mitochondrial sequence (5), against which they were scaffolded to generate YaliW29_M.

Strain W29 was isolated from wastewater in Paris, France, and was used in early inbreeding programs to develop *Y. lipolytica* as a model for genetic studies (6). This lineage has contributed to the genetic background of the widely used Po1 series of backcrossed auxotrophic strains (7, 8) that are employed by labs worldwide for genetic studies. The best annotated reference genome for *Y. lipolytica* was produced using strain CLIB122 (2). We mapped the W29-derived high-throughput sequencing reads to CLIB122 using Bowtie 2 (9) to identify differences between these strains. Analysis using SAMtools (10), Pindel (11), and custom Perl scripts identified 841 indels, 21,541 single-nucleotide polymorphisms (SNPs), and 20,206,880 bases conserved with CLIB122, giving an average SNP density between CLIB122 and W29 of 1,065 SNPs/Mb. Ninety-five percent of the predicted coding sequences

from CLIB122 align well to the W29 assembly (>90% full-length identity), while 287 align along less than half their length or not at all. The strain W29 genome sequence will aid future genetic and genomic studies using *Y. lipolytica*.

Nucleotide sequence accession numbers. This whole-genome shotgun project has been deposited at DDBJ/ENA/GenBank under the accession no. LJBI00000000. The version described in this paper is the first version, LJBI00000000.1.

REFERENCES

1. **Zerbino DR.** 2010. Using the Velvet *de novo* assembler for short-read sequencing technologies. Curr Protoc Bioinformatics Chapter 11:Unit 11.5.
2. **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 J, 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.
3. **Galardini M, Biondi EG, Bazzicalupo M, Mengoni A.** 2011. CONTIGuator: a bacterial genomes finishing tool for structural insights on draft genomes. Source Code Biol Med **6:**11. http://dx.doi.org/10.1186/1751-0473-6-11.
4. **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. http://dx.doi.org/10.1093/oxfordjournals.molbev.a004125.
5. **Kerscher S, Durstewitz G, Casaregola S, Gaillardin C, Brandt U.** 2001.

The complete mitochondrial genome of *Yarrowia lipolytica*. Comp Funct Genomics **2**:80–90. http://dx.doi.org/10.1002/cfg.72.

6. **Wolf K.** 1996. Nonconventional yeasts in biotechnology: a handbook. Springer Verlag, Berlin, Germany.

7. **Le Dall M, Nicaud J, Gaillardin C**. 1994. Multiple-copy integration in the yeast *Yarrowia lipolytica*. Curr Genet **26**:38–44. http://dx.doi.org/10.1007/BF00326302.

8. **Madzak C, Treton B, Blanchin-Roland S**. 2000. Strong hybrid promoters and integrative expression/secretion vectors for quasi-constitutive expression of heterologous proteins in the yeast *Yarrowia lipolytica*. J Mol Microbiol Biotechnol **2**:207–216.

9. **Langmead B, Salzberg SL**. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods **9**:357–359. http://dx.doi.org/10.1038/nmeth.1923.

10. **Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup**. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics **25**:2078–2079.

11. **Ye K, Schulz MH, Long Q, Apweiler R, Ning Z**. 2009. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. Bioinformatics **25**:2865–2871. http://dx.doi.org/10.1093/bioinformatics/btp394.